



**VNiVERSiDAD
D SALAMANCA**
CAMPUS DE EXCELENCIA INTERNACIONAL



CSIC
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



CiC
CENTRO DE INVESTIGACIÓN
DEL CÁNCER

VNiVERSiDAD D SALAMANCA

**INSTITUTO DE BIOLOGÍA MOLECULAR Y CELULAR DEL
CÁNCER (CSIC-USAL)**

**CARACTERIZACIÓN DE LA
ORGANIZACIÓN ESTRUCTURAL DEL
FACTOR DE INTERCAMBIO DE
NUCLEÓTIDO DE GUANINA C3G**

MEMORIA PARA OPTAR AL GRADO DE DOCTOR PRESENTADA POR

María Gómez Hernández

Bajo la dirección de los Doctores

Carmen Guerrero Arroyo
José María de Pereda Vega

Salamanca, 2014

Dra. Carmen Guerrero Arroyo, Profesora Titular de la Universidad de Salamanca; y **Dr. José María de Pereda Vega**, Científico Titular del Consejo Superior de Investigaciones Científicas, miembros del Instituto de Biología Molecular y Celular del Cáncer (IBMCC, CSIC-USAL)

CERTIFICAN:

Que D^a. María Gómez Hernández, Licenciada en Química por la Universidad de Salamanca, ha realizado bajo su dirección conjunta el trabajo de Tesis Doctoral que lleva por título: ***“Caracterización de la organización estructural del factor de intercambio de nucleótido de guanina C3G”***; y consideran que éste reúne originalidad y contenidos suficientes para que sea presentado ante el tribunal correspondiente y optar al Grado de Doctor por la Universidad de Salamanca.

Y para que así conste, y a los efectos oportunos, expiden el presente certificado en Salamanca, a 26 de Mayo de 2014.

Dra. Carmen Guerrero Arroyo

Dr. José María de Pereda Vega

Esta tesis ha sido realizada siendo **María Gómez Hernández** beneficiaria de una ayuda predoctoral de la Universidad de Salamanca financiada por el Banco Santander durante el período de Diciembre de 2008 a Diciembre de 2012.

Este trabajo se ha enmarcado dentro de los proyectos con referencia CSI06A09 financiados por la Consejería de Educación de la Junta de Castilla y León y SAF2010-20918-C02-02 financiado por el Ministerio de Ciencia e Innovación y el Fondo Europeo de Desarrollo Regional.

AGRADECIMIENTOS

Voy a intentar resumir en unas líneas mis agradecimientos a todas las personas que han recorrido conmigo el camino hasta llegar aquí. Pueden resultar unos agradecimientos impersonales, pero aunque no aparezcan los nombres, cada una de las personas que han compartido conmigo los momentos buenos, que han sido muchos; y por qué no, también los malos, saben que forman parte de este trabajo.

En primer lugar, dar las gracias a mis directores de tesis: la Dra. Carmen Guerrero y el Dr. José María de Pereda, por darme esta oportunidad y por guiarme a lo largo de esta etapa. Por sus opiniones, correcciones, paciencia y ánimos, ya que sin su ayuda esto no habría sido posible.

Gracias al Dr. Carlos Alfonso por los experimentos de ultracentrifugación analítica y al Dr. José Manuel Andreu por el acceso al dicrógrafo, ambos del Centro de Investigaciones Biológicas (CSIC, Madrid). Al Dr. Rubén Martínez Buey por su colaboración en los experimentos de SAXS (SLS-PSI, Suiza) y por todos los consejos y opiniones sinceras; y al Dr. Arnoud Sonnenberg (NKI, Ámsterdam) por cedernos las proteínas fluorescentes.

Por otro lado, a todas las personas que de una manera u otra han estado ahí. A todos con los que he compartido muchas horas dentro y fuera del laboratorio. A los que estaban desde el principio, a los que fueron llegando, a los que se fueron y a los que siguen ahí y me preguntan día a día y me aguantan. Gracias de verdad.

A mis amigos de siempre, que no entendían muy bien qué es lo que hacía; a los que llegaron más tarde, pero que se quedarán mucho tiempo.

Y por último, y en este caso los más importantes, a mi familia, y en especial a mis padres y mi hermano, por estar SIEMPRE ahí.

A mis padres y a mi hermano

*Todos somos muy ignorantes.
Lo que ocurre es que no todos ignoramos las mismas cosas.*

Albert Einstein

ABREVIATURAS

ADN	Ácido desoxirribonucleico
ADNc	ADN complementario
Arf	Factor de ribosilación-ADP (<i>ADP-ribosylation factor</i>)
ARN	Ácido ribonucleico
BSA	Albúmina de suero bovino (<i>Bovine Serum Albumin</i>)
CAT	Dominio catalítico
CD	Dicroísmo circular (<i>Circular Dichroism</i>)
CDC25H	Dominio de homología CDC25
CC	<i>Coiled coil</i>
CFP	Proteína azul fluorescente (<i>Cyan Fluorescent Protein</i>)
cNBD	Dominio de unión a nucleótidos cíclicos (<i>cyclic Nucleotide Binding Domain</i>)
coIP	Co-inmunoprecipitación
c(s)	Distribución diferencial de coeficientes de sedimentación
C-terminal	Carboxilo terminal
C1	Región conservada de la proteína quinasa C1
C2	Región conservada de la proteína quinasa C2
C3G	Factor intercambiador de nucleótidos de guanina de unión a dominios SH3 de Crk (<i>Crk SH3 domain-binding guanine nucleotide exchange factor</i>)
C3Gh	C3G humano
C3Gopt	C3G optimizado
DAG	Diacilglicerol
DENN	Dominio expresado de manera diferencial en células normales y neoplásicas (<i>Differentially expressed in normal and neoplastic cells</i>)
DEP	Dominio <i>Dishevelled</i> , <i>Egl-10</i> y <i>Pleckstrin</i>
DH	Dominio de homología Dbl (<i>Dbl Homology</i>)
DHR2	Región de homología Dock (<i>Dock Homology Region</i>)
Dmax	Dimensión máxima
DMSO	Dimetilsulfóxido
DO	Densidad óptica
DTT	Ditiotreitrol
E	Coefficiente de extinción másico
EDTA	Ácido etilendiamino tetraacético
EF	Dominio de unión a Ca ²⁺
EGF	Factor de crecimiento epidermal (<i>Epidermal Growth Factor</i>)
Epac	Proteínas intercambiadoras directamente activadas por cAMP (<i>Exchange proteins directly activated by cAMP</i>)
Erk	Quinasa regulada por señales extracelulares
ϵ	Coefficiente de extinción molar

FBS	Suero fetal bovino (<i>Fetal Bovine Serum</i>)
FRET	Transferencia de energía de resonancia de Förster (<i>Förster Resonance Energy Transfer</i>)
GAP	Proteína activadora de la actividad GTPasa (<i>GTPase-Activating Proteins</i>)
GDP	Guanosín difosfato
GEF	Factor intercambiador de nucleótido de guanina (<i>Guanine nucleotide Exchange Factor</i>)
GH	Hormona de crecimiento (<i>Growth Hormone</i>)
GppNp	Guanilil imidodifosfato
GTP	Guanosín trifosfato
GTPasas	Trifosfatasas de guanosina
HA	Hemaglutinina
HGF <i>Factor</i>)	Factor de crecimiento de hepatocitos (<i>Hepatocyte Growth Factor</i>)
I	Intesidad
IMAC	Cromatografía de afinidad de metal inmovilizado (<i>Immobilized-Metal Affinity Chromatography</i>)
IPTG	Isopropil- β -D-tiogalactopiranosido
IQ	Dominio de unión a calmodulina
I(0)	Intensidad de dispersión a ángulo cero
K _{ap}	Constante aparente de dimerización
K _d	Constante de disociación
LMC	Leucemia mieloide crónica
λ_{exc}	Longitud de onda de excitación
MEF	Fibroblastos embrionarios de ratón
MM	Masa molecular
MMap	Masa molecular aparente
MRW	Masa por residuo medio
MS	Espectrometría de masas (<i>Mass Spectrometry</i>)
m/z	Relación masa/carga
NaPi	Fosfato sódio
N-C3G	Región amino terminal de C3G
NGF	Factor de crecimiento neuronal (<i>Neuronal Growth Factor</i>)
NSD	Discrepancia espacial normalizada (<i>Normalized Spatial Discrepancy</i>)
N-terminal	Amino terminal

OG	Oregon Green
OG-IA	Oregon Green 488 iodoacetamida
PBS	Tampón fosfato salino (<i>Phosphate Buffered Saline</i>)
PC	Fosfatidilcolina
PCR	Reacción en cadena de la polimerasa (<i>Polymerase Chain Reaction</i>)
PDB	Banco de datos de proteínas (<i>Protein Data Bank</i>)
PDGF	Factor de crecimiento derivado de plaquetas (<i>Platelet Derived Growth Factor</i>)
PDZ	Dominio presente en <i>PSD-95</i> , <i>Dlg</i> , y <i>ZO-1/2</i>
PH	Dominio de homología Pleckstrina (<i>Pleckstrin Homology</i>)
pI	Punto isoelectrico
PLCXY	Dominios catalíticos de la fosfolipasa C
PVDF	Polifluoruro de vinilideno
p/p	Relación peso/peso
P(r)	Función de distribución de distancias
PS	Fosfatidilserina
p-Tyr	Fosfo-tirosina
p/v	Relación peso/volumen
PI-3,4,5-P ₃	Fosfatidil-inositol 3,4,5-trisfosfato
PI-4,5-P ₂ = PIP ₂	fosfatidil-inositol 4,5-bifosfato
r	Anisotropía
R	Distancia
RA	Dominio de asociación a Ras (<i>Ras association</i>)
REM	Dominio de intercambio de Ras (<i>Ras Exchange Motif</i>)
Rg	Radio de giro
Rmsd	Desviación cuadrática media (<i>Root mean square Deviation</i>)
rpm	Revoluciones por minuto
RTK	Receptor tirosín-kinasa
s	Vector de dispersión
S _{20^o,w}	Coeficiente de sedimentación aparente corregido en agua a 20 °C
SAXS	Dispersión de rayos-X a bajo ángulo (<i>Small Angle X-ray Scattering</i>)
SB	Tampón de solubilización (<i>sample buffer</i>)
SDS-PAGE	Electroforesis en gel de poliacrilamida con dodecilsulfato sódico
SH3b	Motivo de unión a dominios SH3 (<i>SH3-binding</i>)
SOS	<i>Son of Sevenless</i>
SVD	Descomposición en valores singulares (<i>Singular Value Decomposition</i>)
TBST	<i>Tris-Buffered Saline and Tween-20</i>
TC-PTP	Tirosín fosfatasa de células T (<i>T-Cell Protein Tyrosine Phosphatase</i>)
TCR	Receptores de células T (<i>T-Cell Receptor</i>)

TEV	Virus del grabado del tabaco
u.a.	Unidades arbitrarias
UV	Ultravioleta
V_e	Volumen de elución
V_p	Volumen de Porod
v/v	Volumen/volumen
WB	<u>Western Blot</u>
WT	Forma silvestre (<u>wild type</u>)

INTRODUCCIÓN	9
1. La superfamilia de GTPasas Ras	12
1.1. Ciclo de regulación de las GTPasas	13
1.2. Estructura de las GTPasas de la familia Ras	14
2. GEFs: Factores intercambiadores de nucleótido de guanina	15
2.1. GEFs que actúan sobre GTPasas de la familia Ras	16
2.2. C3G: un GEF de Rap1 y R-Ras.....	20
2.2.1. C3G es una proteína multifuncional.....	21
2.2.2. Estructura de C3G	23
2.2.3. Activación y regulación de C3G.....	25
2.3. Mecanismos de autoinhibición de los GEFs: SOS1, RasGRP1 y Epac2	26
2.3.1. SOS1 presenta un mecanismo de regulación alostérica.....	27
2.3.2. RasGRP1 presenta un mecanismo de regulación por impedimento estérico y dimerización.....	31
2.3.3. Epac2 presenta un mecanismo de regulación estérico.....	34
OBJETIVOS	39
MATERIALES Y MÉTODOS	43
1. Secuencias de ADNc de C3G	45
2. Métodos generales de manipulación y caracterización de ác. nucleicos ...	45
2.1. Purificación y cuantificación de ADN.....	45
2.2. Electroforesis de ADN	46
2.3. Secuenciación de ADN.....	46
2.4. Mutagénesis dirigida	47
3. Métodos generales de manipulación y caracterización de proteínas	48
3.1. Cuantificación de proteínas	48
3.2. Electroforesis de proteínas.....	49
3.3. Inmunodetección de proteínas: Western blot.....	50
4. Métodos para estudios en cultivos celulares	51
4.1. Construcciones para la expresión de C3G en células de mamífero	51
4.1.1. Clonación en vectores de la familia pCEFL.....	51
4.1.2. Vectores con proteínas fluorescentes para FRET	53
4.1.2.1. Construcción de vectores con proteínas fluorescentes para FRET ..	54

4.1.2.2. Clonación de C3G en vectores para FRET.....	55
4.2. Línea celular COS-1	56
4.3. Ensayos de co-inmunoprecipitación.....	57
4.4. Ensayos de FRET en células COS-1	58
5. Producción heteróloga de proteínas en bacterias	61
5.1. Clonación en vectores de expresión en bacterias	61
5.2. Expresión y purificación de proteínas.....	63
6. Métodos biofísicos y bioquímicos de análisis de proteínas.....	64
6.1. Ensayos de cristalización.....	64
6.2. Dicroísmo circular en el UV lejano.....	65
6.3. Proteólisis limitada	66
6.4. Cromatografía de exclusión molecular analítica	66
6.5. Ultracentrifugación analítica.....	67
6.5.1. Velocidad de sedimentación	68
6.5.2. Equilibrio de sedimentación	68
6.6. Anisotropía de fluorescencia extrínseca.....	70
6.6.1. Marcaje de N-C3G con Oregon green 488 I-acetamida.....	71
6.6.2. Análisis de la homo-asociación de N-C3G mediante fluorescencia.....	71
6.7. Dispersión de rayos-X a bajo ángulo (SAXS)	72
6.7.1. Toma de datos de SAXS.....	73
6.7.2. Procesado y análisis de datos de SAXS.....	74
6.7.3. Modelado <i>ab initio</i> de estructuras a partir de datos de SAXS	75
6.8. Ensayos de unión a lípidos.....	76
RESULTADOS.....	77
1. Organización estructural de C3G.....	79
1.1. Interacción de N-C3G con otras regiones de C3G: la región N-C3G interacciona con el dominio REM	79
1.2. Interacción intramolecular de tipo cabeza-cola en C3G, análisis mediante FRET	82
1.3. Identificación de residuos del dominio REM importantes para la interacción con el dominio N-C3G	85

1.4. Efecto de la sustitución Y504E sobre la interacción cabeza-cola.....	92
2. Estructura de la región N-C3G	94
2.1. Análisis de la estructura secundaria de la región N-C3G	95
2.2. Mapas de proteólisis controlada de N-C3G.....	99
2.3. Homo-asociación de N-C3G	106
2.4. Análisis de la estructura de N-C3G en solución mediante dispersión de rayos-X a bajo ángulo (SAXS)	117
3. Unión de N-C3G a fosfolípidos	123
DISCUSIÓN.....	125
1. Organización estructural de C3G.....	127
1.1. Posible papel de la interacción cabeza-cola en la regulación de C3G.....	127
1.1.1. Relación entre la unión de la región amino de C3G a E-cadherina y la interacción cabeza-cola.....	131
2. Estructura de la región amino terminal de C3G.....	132
2.1. Homo-asociación de la región amino.....	134
3. Interacción de la región amino de C3G con lípidos.....	136
4. Modelo hipotético de regulación de C3G	137
CONCLUSIONES.....	139
BIBLIOGRAFÍA	143

INTRODUCCIÓN

Las células reciben señales de su entorno que controlan múltiples procesos como su proliferación, diferenciación, migración o muerte. Estas señales pueden tener diferentes orígenes, ya sea moléculas solubles, moléculas de la matriz extracelular o moléculas que establecen un contacto directo con otras células, y son recibidas por receptores celulares específicos. La activación de estos receptores inicia multitud de rutas de señalización. La transducción de señales no ocurre de manera lineal desde los receptores hasta las respuestas celulares específicas, sino que existe una gran red de conexiones entre diferentes rutas de señalización. Muchas de las moléculas que forman parte de esta red son proteínas modulares que tienen una regulación muy compleja, como es el caso de las proteínas Ras y sus reguladores, las proteínas RasGEF.

1. La superfamilia de GTPasas Ras

Las proteínas de la superfamilia Ras, también llamadas GTPasas pequeñas, forman un grupo independiente dentro de las GTP hidrolasas. Constituyen una familia de GTPasas muy conservadas a lo largo de la evolución de todos los organismos eucariotas, desde levaduras hasta el hombre, lo que sugiere que estas proteínas desempeñan funciones críticas en la célula (Lowy and Willumsen, 1993; Macara et al., 1996; Reuther and Der, 2000; Rojas and Santos, 2002; Santos and Nebreda, 1989). En humanos, la superfamilia de proteínas Ras está compuesta por unos 150 miembros que tienen en común un dominio G (~ 20 kDa) implicado en la unión e hidrólisis de GTP (Bourne et al., 1991; Wennerberg et al., 2005). Estas proteínas se pueden dividir en al menos cinco familias atendiendo a su similitud de secuencia y a su función:

- Familia Ras (*Ras sarcoma*). Involucradas en la regulación de la expresión génica, apoptosis, proliferación, diferenciación y morfología celular (Takai et al., 2001). Esta familia de proteínas Ras comprende unos 16-18 miembros entre los que se encuentran las proteínas Ras, Rap, Ral entre otras (Reuther and Der, 2000).
- Familia Rho (*Ras homologus*). Participan en la reorganización de los filamentos de actina, regulación del ciclo celular y de la expresión génica (Etienne-Manneville and Hall, 2002).
- Familia Rab (*Ras-like proteins in brain*). Controlan el transporte vesicular y el tráfico de proteínas entre distintos orgánulos (Zerial and McBride, 2001).
- Familia Arf (*ADP-ribosylation factor*). Regulan el tráfico vesicular (Memon, 2004).
- Familia Ran (*Ras-like nuclear*). Implicadas en el transporte nucleocitoplasmático de ARN y proteínas (Weis, 2003).

Desde la clonación del primer oncogén humano *H-ras* (Santos et al., 1982), numerosos estudios oncológicos han encontrado mutaciones activantes de Ras en un 30 % de la totalidad de tumores (Barbacid, 1990; Bos, 1989; Pulciani et al., 1985), lo que indica que la desregulación de la actividad de Ras es una de las causas más comunes involucradas en la aparición de cáncer, de ahí la importancia de su regulación.

1.1. Ciclo de regulación de las GTPasas

Una característica común de todas las GTPasas de la superfamilia Ras es su habilidad para experimentar cambios estructurales en respuesta a su unión a GDP o a GTP. La unión a GDP representa una forma inactiva, mientras que la unión a GTP representa una forma activa que interacciona con efectores y activa rutas que afectan a la morfología celular, el tráfico celular, crecimiento, diferenciación y apoptosis. Cada uno de estos estados reconoce diferentes proteínas, de modo que esto permite funcionar a las GTPasas como interruptores moleculares en la célula (Cherfils and Chardin, 1999) (Figura I1).

La afinidad de la mayoría de las GTPasas pequeñas por el nucleótido está en el rango de nanomolar a picomolar, siendo la media de tiempo que tarda un nucleótido en disociarse de entre una a varias horas. Los GEFs (*Guanine nucleotide Exchange Factors*), factores intercambiadores de nucleótido de guanina, son los responsables de catalizar esta reacción, de manera que la respuesta biológica se produzca en pocos minutos. Los GEFs reducen la afinidad por el nucleótido y de este modo aumentan la velocidad de disociación varios órdenes de magnitud. Posteriormente, debido a la mayor concentración celular de GTP sobre GDP, las proteínas Ras unen GTP, constituyendo así una forma activada e interaccionan con efectores, que son proteínas que se unen preferentemente a la forma unida a GTP. En el proceso en sentido contrario, el GTP es hidrolizado a GDP por la actividad GTPasa intrínseca de las proteínas Ras en colaboración con las proteínas GAP (*GTPase-Activating Proteins*) (Boguski and McCormick, 1993; Bos, 1998; Vetter and Wittinghofer, 2001).

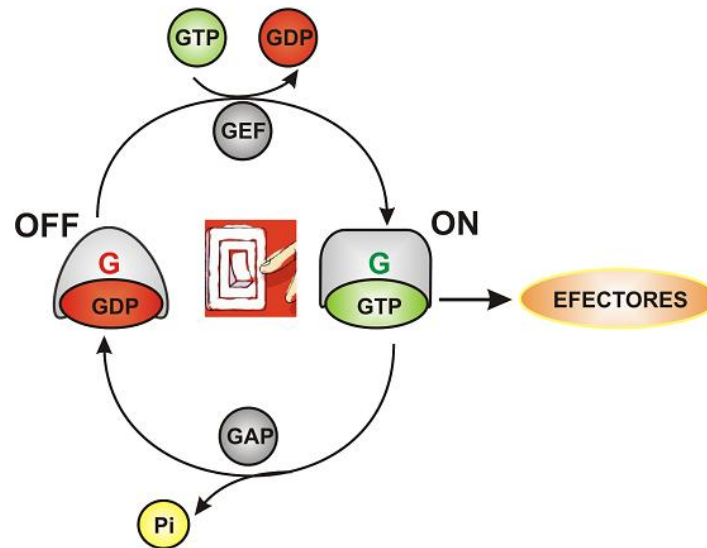


Figura I1. Ciclo de regulación de las GTPasas. Las GTPasas Ras funcionan como interruptores moleculares alternando entre un estado inactivo unido a GDP y un estado activo unido a GTP. En ambos estados el nucleótido se encuentra unido muy estrechamente a la GTPasa. El intercambio de GDP por GTP está regulado catalíticamente por dos tipos de moléculas de acciones opuestas, los factores activadores del intercambio de nucleótidos, GEFs; y las proteínas que estimulan la hidrólisis, GAPs, que aumentan la actividad GTPasa intrínseca de estas proteínas facilitando su inactivación.

1.2. Estructura de las GTPasas de la familia Ras

Las GTPasas Ras son proteínas monoméricas de pequeño tamaño, alrededor de los 20 kDa. Los miembros de la familia comparten entre un 30 y un 50 % de identidad de secuencia, y todos ellos poseen un motivo consenso, denominado dominio G, que es el responsable de los cambios conformacionales dependientes de nucleótido. El dominio G de la proteína Ras se considera como la unidad funcional básica de las GTPasas; el resto se describen como modificaciones sobre esta estructura canónica (Vetter and Wittinghofer, 2001; Wittinghofer and Vetter, 2011).

Las estructuras de GTPasas unidas a GDP (forma inactiva) y GTP (forma activa) revelan cambios conformacionales ligados a la unión de nucleótidos (Vetter and Wittinghofer, 2001). Las principales zonas de contacto de las GTPasas con el nucleótido son el motivo N/TKxD, el lazo P y las regiones denominadas *switch I* y *switch II* (Figura I2). El motivo N/TKxD (donde x puede ser cualquier aminoácido) interacciona con la base del nucleótido. El lazo P interacciona con los fosfatos β - γ y

con el ion Mg^{2+} . Las regiones *switch I* y *switch II* envuelven el fosfato- γ del GTP (Jurnak, 1985; la Cour et al., 1985), por lo que su estructura es sensible a la naturaleza del nucleótido unido. La hidrólisis del GTP y posterior liberación del fosfato- γ causa la relajación de los *switch I* y *II*.

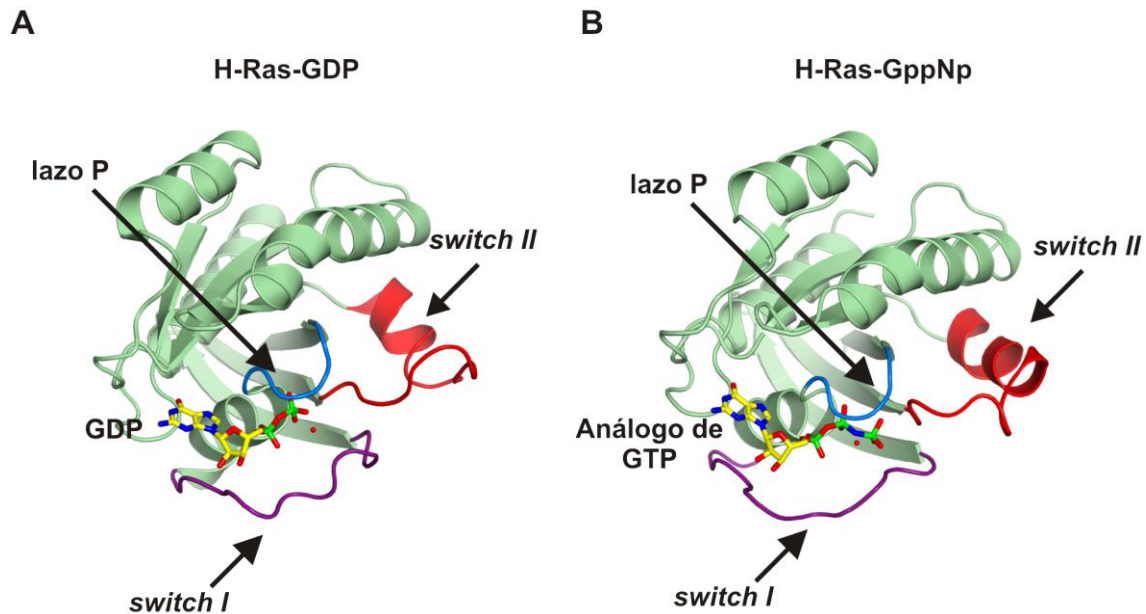


Figura 12. Estructura de las GTPasas de la familia Ras. (A) Estructura de H-Ras inactiva unida a GDP (código PDB: 4Q21) (Milburn et al., 1990). (B) Estructura de H-Ras activa unida a GppNp, un análogo no hidrolizable de GTP (código PDB: 5P21) (Pai et al., 1990). Los mayores cambios estructurales entre los dos estados se localizan en las regiones *switch I* y *switch II* (coloreadas en morado y en rojo respectivamente). El lazo P se muestra en azul.

Una característica importante de la mayoría de las GTPasas es que son modificadas post-traduccionalmente por lípidos (Casey et al., 1989). Estas modificaciones son esenciales para favorecer la asociación con la membrana y para determinar su localización subcelular.

2. GEFs: Factores intercambiadores de nucleótido de guanina

Los GEFs catalizan el intercambio del nucleótido unido a la GTPasa (ver antes). El número de GEFs supera considerablemente el número de GTPasas. En concreto, la familia de proteínas Ras con unos 16-18 miembros (Reuther and Der, 2000), es activada por unos 30 GEFs y además las distintas subfamilias de Ras comparten muchos de estos activadores. Esta mayor diversidad de los reguladores

respecto a sus sustratos posibilita una mayor especificidad en la activación de las GTPasas en respuesta a diferentes estímulos, ya que algunos, por ejemplo, son específicos de tejido (GRF1 de cerebro, CalDAG-GEF de tejido hematopoyético). La mayoría de GEFs son proteínas modulares en las que el dominio catalítico, responsable de la actividad GEF, está flanqueado por regiones que pueden incluir dominios que promueven la oligomerización, interacción con otras proteínas, interacción con membranas, así como otras regiones cuyas funciones son desconocidas. Esta diversidad de dominios hace que los diferentes GEFs puedan canalizar distintas rutas de activación de las proteínas Ras, lo que constituye otro factor de especificidad (Bos et al., 2007; Cherfils and Zeghouf, 2013; Stork and Dillon, 2005).

2.1. GEFs que actúan sobre GTPasas de la familia Ras

Las GTPasas de la familia Ras están reguladas mayoritariamente por GEFs que contienen un dominio catalítico con homología con la proteína CDC25 de *S. cerevisiae* (Camonis et al., 1986), por lo que se les denomina dominios CDC25H. Los dominios CDC25H son estructuralmente diferentes a los dominios catalíticos de GEFs de otras familias de GTPasas, los cuales incluyen dominios DH (*Dbl homology*), PH (*Pleckstrin homology*), DHR2 (*Dock Homology Region*) y PRONE (*Plant-specific Rop nucleotide exchanger*) en GEFs que actúan sobre la familia Rho; dominios Sec7 en GEFs de la familia Arf; un propulsor- β de RCC1 en el GEF de Ran y dominios DENN (*Differentially expressed in normal and neoplastic cells*), Vps9, Sec2 y Mss4 presentes en GEFs para proteínas de la familia Rab (Bos et al., 2007; Cherfils and Zeghouf, 2013).

Los GEFs de tipo CDC25H que actúan sobre proteínas de la familia Ras se pueden dividir en varios grupos o familias atendiendo a su estructura modular (Tabla I1, Figura I3). Así destacan: (i) las proteínas SOS (*Son of Sevenless*), caracterizadas por poseer un dominio DH-PH (con actividad RhoGEF) y un dominio rico en prolinas (PxxP) de unión a dominios SH3; (ii) RasGRFs (*Ras Guanine nucleotide Releasing Factor*), que además de poseer el dominio DH-PH, al igual que las proteínas SOS, también tienen un dominio de unión a calmodulina (IQ), por lo que participan en la señalización por Ca^{2+} ; (iii) RasGRPs (*Ras Guanine*

nucleotide Releasing Protein, también denominadas CalDAG-GEFs), cuya principal característica es la presencia de dominios EF de unión a Ca^{2+} y de un dominio C1 (región conservada 1 de la proteína quinasa C) de unión a diacilglicerol (DAG); (iv) Epac (cNBD-GEFs, *cyclic Nucleotide Binding Domain*), que poseen dominios DEP (*Dishevelled*, *Egl-10*, *Plekstrin*) y dominios cNBD de unión a nucleótidos cíclicos; (v) PDZ-GEFs, que además de dominios cNBD poseen dominios PDZ (*PSD-95*, *Dgl*, *ZO*); (vi) RalGEFs, como RalGPS, RalGDS, Rgl o Rlf, que se caracterizan por ser además efectores de las proteínas Ras, las cuales se unen a su dominio RA (*Ras association*), dominio que también está presente en las proteínas Epac y PDZ-GEF; (vii) PLC ϵ , que además de poseer también un dominio RA, contiene dominios PLCXY (dominios catalíticos de la fosfolipasa C) y C2 (región conservada 2 de la proteína quinasa C). Por último, C3G posee un dominio PxxP en su región central y un dominio de unión a E-cadherina en su extremo N-terminal. En todos estos GEFs, a excepción de PLC ϵ y RalGPS, el dominio CDC25H aparece en combinación con un dominio REM (*Ras Exchange Motif*) (Quilliam et al., 2002). El dominio REM es normalmente esencial para la actividad de las proteínas pero su función parece que no está conservada en los diferentes GEFs.

Tabla I1. GTPasas de la familia Ras y sus GEFs

Subfamilia	GTPasa	GEFs
Ras	H-Ras	CDC25/GRF 1/2, SOS 1/2, RasGRP1 (CalDAG-GEF II), PDZ-GEF1 (cNBD-RasGEF), RasGRP3 (CalDAG-GEF III), RasGRP4
	N-Ras	GRF 1/2, SOS 1/2, RasGRP1, RasGRP2 (CalDAG-GEF I)
	K-Ras A, B	smgGDS, RasGRP 1/2 (solo K-Ras4B)
Rap	Rap1 A, B	C3G , smgGDS, MR-GEF (GFR, Repac), PDZ-GEF 1/2 (RA-GEF 1/2), Epac 1/2 (cNBD-GEF I/II), RasGRP2, RasGRP3, AND-34 (NSP2, BCAR3), PLC ϵ , Dock-4, Chat (SHEP1, NSP3)
	Rap2 A, B	C3G , PDZ-GEF 1/2, RasGRP2, RasGRP3, MR-GEF
Ral	Ral A, B	smgGDS, RalGDS, Rgl, Rlf, RalGPS 1A/1B, AND-34
R-Ras	R-Ras	GRF1, C3G , RasGRP1, RasGRP2, RasGRP3, AND-34, Chat
	R-Ras2 (TC21)	GRF1, C3G , RasGRP1, RasGRP2, RasGRP3, AND-34, SOS1
	R-Ras3 (M-Ras)	SOS1, GRF1, RasGRP2, RasGRP3

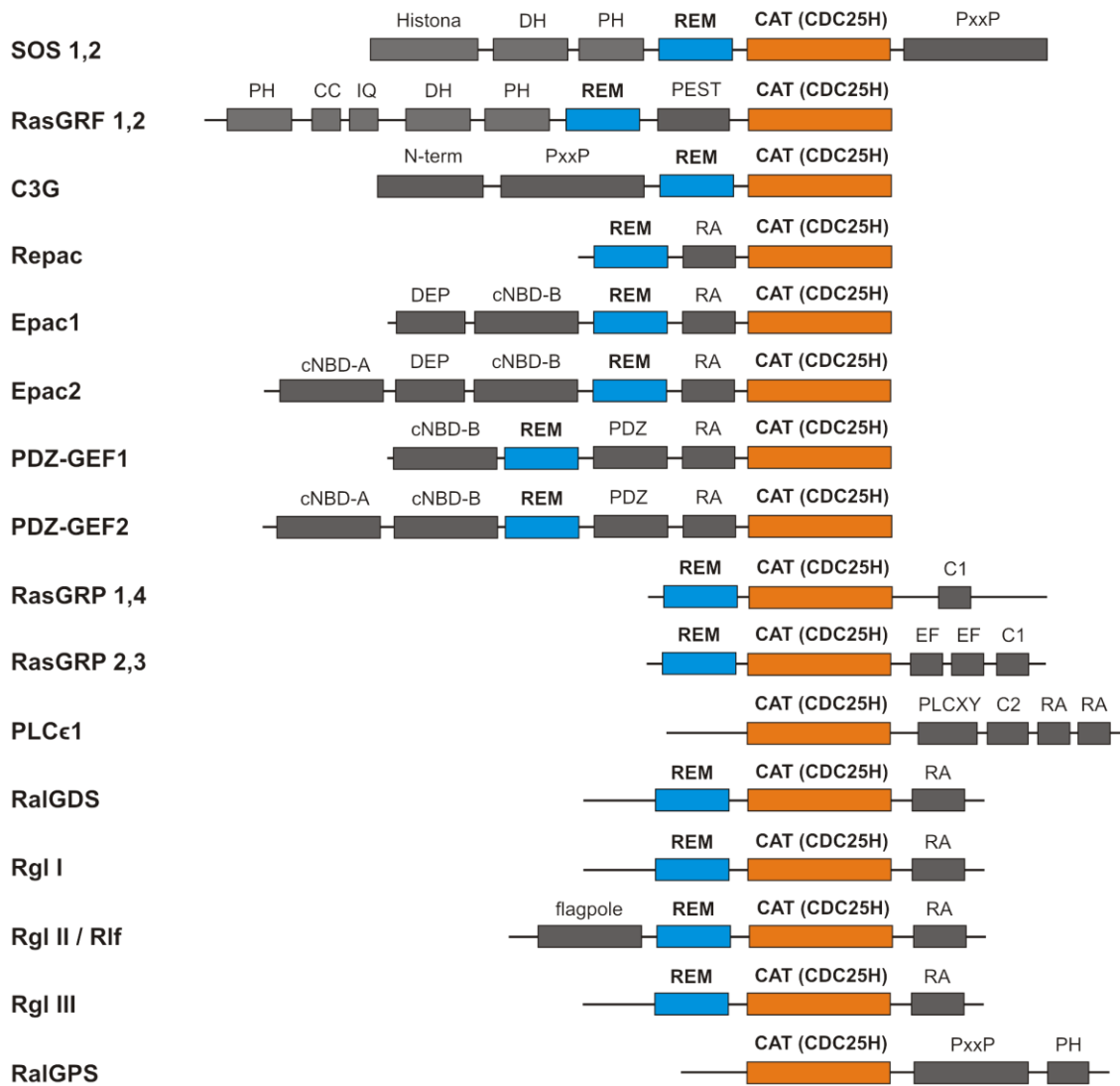


Figura 13. Estructura modular de GEFs de la familia de GTPasas Ras. La mayoría de los GEFs de la familia Ras contienen un dominio catalítico del tipo CDC25H (naranja) precedido de un dominio REM (azul). Los dominios variables se muestran en gris: *DH*, dominio de homología Dbl; *PH*, dominio de homología Pleckstrin; *PxxP*, dominio de unión a SH3; *CC*, región *coiled coil*; *IQ*, dominio de unión a calmodulina; *PEST*, región tipo PEST; *RA*, dominio de asociación a Ras; *DEP*, dominio *Dishevelled*, *Egl-10* y *Pleckstrin*; *cNBD-A/B*, dominio de unión a nucleótidos cíclicos; *PDZ*, dominio presente en *PSD-95*, *Dlg*, y *ZO-1/2*; *C1*, región conservada de la proteína quinasa C1; *EF*, dominio de unión a Ca^{2+} ; *PLCXY*, dominio fosfolipasa C; *C2*, región conservada de la proteína quinasa C2; *flagpole*, lámina- β que contiene una estructura específica que permite la unión de proteínas con dominios SH3. C3G presenta el dominio CDC25H junto con el dominio REM, y ambos se localizan en la región C-terminal y carece de dominios modulares de interacción con otras proteínas que se encuentran en la gran mayoría de otros GEFs de Rap.

El mecanismo de acción de los GEFs es muy parecido en todos ellos a pesar de que sus estructuras son muy diferentes. Los GEFs deforman la estructura del *switch I*, del *switch II* y del lazo P perturbando el sitio de unión de los fosfatos, pero dejando la región de unión a la base prácticamente inalterada. Como consecuencia,

los grupos fosfato son liberados en primer lugar tras la interacción con el GEF (Bos et al., 2007; Cherfils and Zeghouf, 2013) (Figura I4).

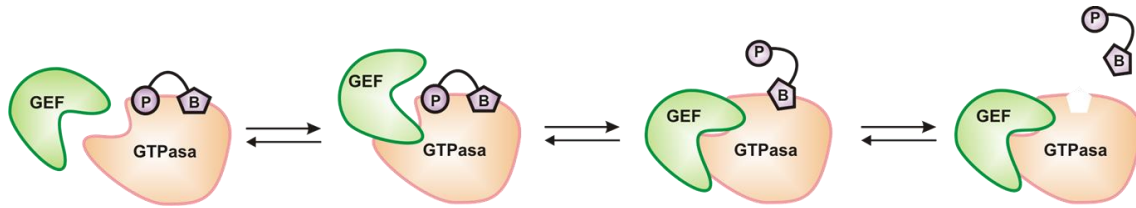


Figura I4. Esquema del mecanismo de acción de los GEFs. El GEF se une a la GTPasa formando un complejo ternario; esta interacción desestabiliza el sitio de unión de los grupos fosfato del nucleótido provocando su liberación. La reacción es completamente reversible, siendo la base nitrogenada la región del nucleótido que antes se une a la GTPasa. La unión del nuevo nucleótido produce el desplazamiento del GEF. Adaptado de (Bos et al., 2007).

La estructura de la región REM-CDC25H se describió inicialmente en el complejo de SOS1 unido a Ras (Boriack-Sjodin et al., 1998). El dominio CDC25H está formado por 11 α -hélices (α A- α K) que se organizan en una estructura en forma de cuenco (Figura I5). Las hélices α H- α I y dos pequeñas hebras- β localizadas entre estas dos hélices se proyectan hacia fuera del núcleo compacto del dominio formando una estructura antiparalela denominada horquilla helicoidal. La GTPasa se une a la cara cóncava del dominio catalítico y las regiones de la GTPasa que interaccionan con el dominio CDC25H incluyen el lazo P y los *switch I* y *II*. La horquilla helicoidal del dominio catalítico se inserta entre los *switch* desplazándolos y bloquea los sitios de unión al fosfato- α del Mg^{2+} .

El dominio REM está formado por 6 α -hélices (α 1- α 6) que forman un haz helicoidal. La principal zona de contacto entre los dominios se establece entre las hélices α 1 y α 2 del dominio REM y la hélice α I de la horquilla helicoidal del dominio catalítico. El dominio REM no contacta directamente con la GTPasa unida al dominio CDC25H. La horquilla helicoidal se debe encontrar en la conformación adecuada tanto para la activación como para la regulación de CDC25H (Freedman et al., 2006; Freedman et al., 2009).

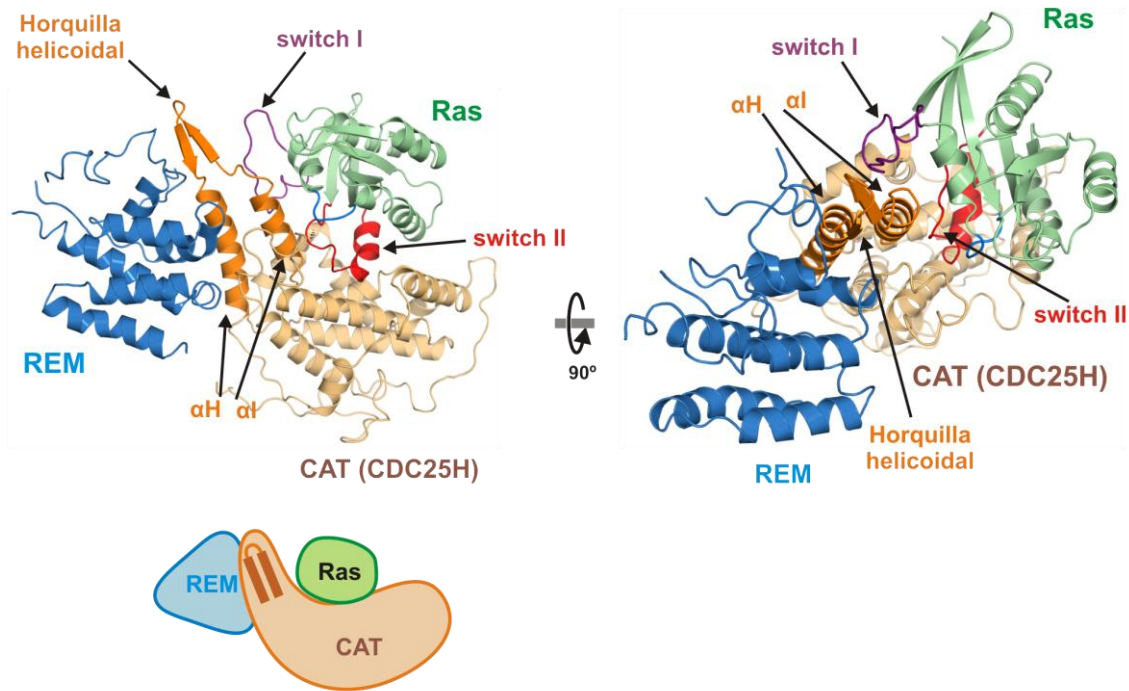


Figura 15. Estructura de la región catalítica de SOS1 unida a Ras. (A) Dos vistas ortogonales del complejo formado por los dominios REM (azul) y CDC25H (naranja) de SOS1 unidos a la GTPasa Ras (verde) (código PDB: 1BKD). La horquilla helicoidal del dominio catalítico, formada por las hélices αH y αI (naranja oscuro), se inserta entre los dos *switch* de Ras alterando la zona de unión al fosfato y así se promueve la liberación del nucleótido. El dominio REM establece contactos con la horquilla helicoidal del dominio catalítico.

2.2. C3G: un GEF de Rap1 y R-Ras

C3G (*Crk SH3 domain-binding guanine nucleotide exchange factor*) se clonó por su capacidad de unirse al dominio SH3 de la proteína adaptadora Crk (Tanaka et al., 1994), descubriéndose posteriormente su capacidad para actuar como GEF para Rap1 y R-Ras además de Rap2, TC21 y el miembro de la familia Rho TC10 (Chiang et al., 2001; Gotoh et al., 1995; Gotoh et al., 1997; Ling et al., 2003; Mochizuki et al., 2000; Ohba et al., 2000; Wang et al., 2006; York et al., 1998). Si bien, recientemente se ha descrito que el dominio CDC25H de C3G aislado no tiene actividad GEF *in vitro* sobre Rap2A-C ni R-Ras (Popovic et al., 2013a).

2.2.1 C3G es una proteína multifuncional

C3G participa a través de la activación de Rap1 en procesos de adhesión y migración mediados por integrinas (Arai et al., 1999; Arai et al., 2001; de Jong et al., 1998; Uemura and Griffin, 1999). La inactivación en ratón del gen *RAPGEF1* que codifica C3G es letal durante el desarrollo embrionario (Ohba et al., 2001), lo cual sugiere que su función no puede ser compensada por otros RapGEFs en estadios tempranos del desarrollo.

C3G interviene en las interacciones proteicas que se producen en los focos de adhesión, que son estructuras que conectan las integrinas extracelulares con los filamentos de actina (Feller, 2001), regulando así funciones celulares tales como adhesión, movilidad y arquitectura celular. La fosforilación de C3G favorece la activación y la formación de complejos con las proteínas Crk y p130Cas, lo que aumenta la habilidad de las células NIH3T3 para asociarse con estructuras del citoesqueleto (de Jong et al., 1998). Estos datos se complementan con los resultados obtenidos a partir de los estudios con células MEF deficientes en C3G, donde se ha relacionado la ausencia de C3G con la deficiencia de la activación de Rap1, lo que provoca una inhibición de la adhesión celular, desajustes de la expansión celular y aumento de la migración celular (Ohba et al., 2001). Recientemente se ha demostrado un papel de C3G en la regulación de la adhesión en leucemia mieloide crónica (LMC) a través de la interacción con p38 α MAPK y proteínas de las adhesiones focales (Maia et al., 2013). C3G también interviene en la formación de las uniones célula-célula, gracias a su interacción directa con E-cadherina. Esta unión es excluyente con la unión de β -catenina a E-cadherina, sugiriendo que C3G es relevante como iniciador de las uniones célula-célula y posteriormente sería desplazada por β -catenina, dando lugar a uniones maduras (Hogan et al., 2004).

C3G es capaz de inhibir la capacidad transformadora de numerosos oncogenes como *ras*, *sis*, *v-raf*, *dbl* y *R-ras* ya que impide la transformación celular a través de mecanismos independientes de su actividad GEF (Guerrero et al., 1998; Guerrero et al., 2004). La sobreexpresión de C3G en fibroblastos NIH3T3 inhibe la fosforilación de Erk1/2 y la expresión de la ciclina A, provocando la pérdida de la capacidad de anclaje al sustrato propia del crecimiento tumoral (Martin-Encabo et al., 2007). Por tanto C3G podría ser considerado como un gen supresor de tumores.

C3G colabora mediante su interacción con Hck en la inducción de la apoptosis, efecto también independiente de su dominio catalítico (Shivakrupa et al., 2003). Además, C3G participa en la regulación de la apoptosis en varios tipos celulares a través de la inhibición de p38 α MAPK (Gutiérrez-Uzquiza et al., 2010; Maia et al., 2009).

C3G está involucrada en la reorganización de la actina y su sobre-expresión resulta en la formación de filopodios en líneas celulares epiteliales y en células transformadas con v-Abl (Radha et al., 2007).

C3G es inducido durante la diferenciación neuronal y su sobre-expresión estimula el crecimiento de las neuritas. Además, la interacción de C3G con Crk es necesaria para la diferenciación de adipocitos. Por otra parte, los niveles de C3G aumentan durante la diferenciación de monocitos a macrófagos (Jin et al., 2000; Radha et al., 2008). C3G también tiene un papel importante en la ruta de diferenciación de megacariocitos a plaquetas inducida por la trombopoyetina (Stork and Dillon, 2005), lo que abre nuevas perspectivas de estudio de C3G como diana terapéutica en la enfermedad trombótica (Gutierrez-Herrero et al., 2012; Maia et al., 2013).

En cuanto a las rutas de señalización en las que participa, C3G regula la activación de ERK/MAPK a través de Rap1 en células que expresan B-Raf (Kao et al., 2001; Mizuchi et al., 2005; Wu et al., 2001), así como la ruta de activación de JNK en fibroblastos y células 293T a través de R-Ras (Tanaka and Hanafusa, 1998; Tanaka et al., 1997). Además, regula la vía p38 α MAPK en MEFs y células K562 a través de la activación de Rap1, modulando el balance apoptosis/supervivencia en estos tipos celulares (Gutiérrez-Uzquiza et al., 2010; Maia et al., 2009).

Otras funciones importantes de C3G incluyen un papel regulador en el desarrollo del córtex cerebral (Radha et al., 2008; Voss et al., 2008; Voss et al., 2006), así como en la maduración vascular (Voss et al., 2003) y en la maduración y expansión de los timocitos (Nolz et al., 2008). C3G también participa en el transporte de glucosa en adipocitos y células musculares a través de la activación de TC10 (Chiang et al., 2006; Gual et al., 2002).

2.2.2. Estructura de C3G

La proteína C3G humana consta de 1077 residuos (~ 140 kDa) y tiene una estructura modular en la que se distinguen 3 regiones (Figura I6).

La región N-terminal (aproximadamente residuos 1-250) no presenta similitud con otras proteínas. Dentro de esta región, el segmento 144-230 interacciona con el dominio citoplasmático de E-cadherina (aminoácidos 823-844) durante la formación de las uniones adherentes y se pierde en uniones maduras por competición con la unión de β -catenina (Hogan et al., 2004).

La región central contiene 5 motivos de secuencia ricos en Pro con el motivo Pro-x-x-Pro característico de sitios de unión a dominios SH3, por lo que nos referimos a esta región como SH3b (SH3-*binding*). Se han identificado cuatro pequeñas secuencias de unión a Crk dentro del dominio SH3b, residuos 282-291, 452-462, 539-549 y 607-614 (Chen et al., 1994; Knudsen et al., 1994). En posición amino con respecto a la región de unión a Crk se encuentra otra región rica en Pro, residuos 265-276, que se une exclusivamente al dominio SH3 de la proteína adaptadora p130Cas (Knudsen et al., 1994). La región central interacciona también con otras proteínas efectoras como Grb2 (Tanaka et al., 1994), Hck (Shivakrupa et al., 2003) y c-Abl (Radha et al., 2007) que presentan dominios SH3 en su estructura. Además, se ha descrito la interacción de la región central de C3G con actina (Martin-Encabo et al., 2007), la tirosín fosfatasa de células T, TC-PTP (Mitra et al., 2011) y β -catenina (Hogan et al., 2004), que son moléculas que carecen de dominios SH3.

La región C-terminal está formada por un dominio REM seguido de un dominio catalítico del tipo CDC25H responsable de la actividad GEF.

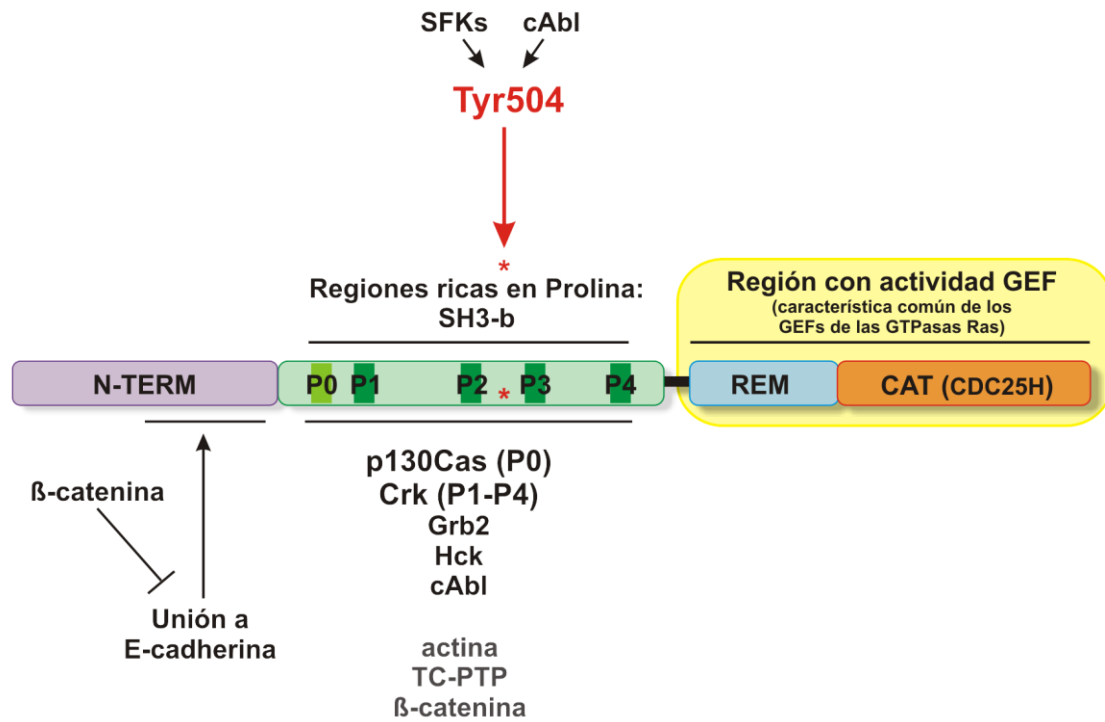


Figura 16. Estructura de C3G. Se indican los diferentes dominios así como su interacción con otras moléculas implicadas en su regulación.

En humanos, existen al menos 3 variantes de C3G (Figura 17). Mediante procesamiento alternativo se generan predominantemente dos isoformas que difieren en su región amino terminal. Los 3 primeros aminoácidos de la isoforma *a* (isoforma canónica) son reemplazados por 21 aminoácidos en la isoforma *b*. Ambas se expresan en niveles altos en músculo esquelético, placenta, corazón fetal y cerebro; y en niveles bajos en el hígado (Guerrero et al., 1998). Existe una tercera isoforma truncada de 87 kDa, denominada p87C3G, que carece de los 305 primeros aminoácidos de la proteína completa, región que incluye la zona de unión a E-cadherina y las dos primeras regiones ricas en Pro (Gutiérrez-Berzal et al., 2006). p87C3G interacciona con Bcr-Abl y de hecho se expresa en líneas celulares y en muestras de pacientes de LMC (Nowell and Hungerford, 1960) por lo que se ha sugerido que p87C3G juega un papel en la patogénesis de esta enfermedad (Gutiérrez-Uzquiza et al., 2010; Maia et al., 2013; Maia et al., 2009).

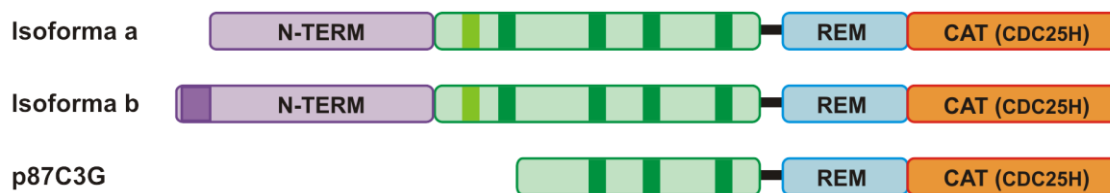


Figura 17. Isoformas de C3G. Estructura de las tres principales isoformas de C3G generadas por procesamiento alternativo.

2.2.3. Activación y regulación de C3G

La ruta C3G-Rap1 es activada en respuesta a diversos estímulos extracelulares que incluyen la activación de receptores de células T (TCR) (Reedquist et al., 1996), factor de crecimiento de hepatocitos (HGF) (Sakkab et al., 2000), hormona de crecimiento (GH) (Ling et al., 2003), factor de crecimiento derivado de plaquetas (PDGF) (Yokote et al., 1998), factor de crecimiento epidermal (EGF) (Okada and Pessin, 1997), factor de crecimiento neuronal (NGF) (York et al., 1998), interferón- γ (Alsayed et al., 2000), eritropoyetina (Ling et al., 2003), interleucina-3 (Arai et al., 2001; Nosaka et al., 1999) y señales de adhesión a través de integrinas (Buensuceso and O'Toole, 2000).

La activación de C3G ocurre principalmente a través del reclutamiento del complejo Crk-C3G a la membrana donde interacciona con Rap1 (Gotoh et al., 1995; van den Berghe et al., 1997). Esta traslocación está mediada por la unión del dominio SH2 de la proteína adaptadora Crk a proteínas fosforiladas en Tyr (Ichiba et al., 1997), como receptores tirosín-quinasa (Yokote et al., 1998). La activación de C3G va acompañada de la fosforilación en residuos de Tyr (Ichiba et al., 1999; Radha et al., 2004).

La delección de la mitad N-terminal, aminoácidos 1-579, que incluye la región de cabeza y parte de la región SH3b, aumenta la actividad GEF de C3G de forma independiente de Crk, lo que sugiere que esta región constituye un elemento regulador negativo en *cis* (Ichiba et al., 1999). El efecto inhibitorio de la región amino de C3G sobre su actividad catalítica es posible que esté ligado a interacciones intramoleculares en las que participa la región amino y que darían lugar a una conformación autoinhibida de la molécula.

C3G es fosforilado en la Tyr504 por quinasas de la familia Src (c-Src y Hck) (Radha et al., 2004; Shivakrupa et al., 2003) y c-Abl (Mitra and Radha, 2010) (Figura I6). La fosforilación de C3G en la Tyr504 es necesaria para la activación dependiente de Crk (Ichiba et al., 1999). Recientemente se ha descrito que Cdk5/p35 fosforila C3G en residuos de Ser y que esta modificación está ligada a la activación de Rap1 (Utreras et al., 2013).

2.3. Mecanismos de autoinhibición de los GEFs: SOS1, RasGRP1 y Epac2

La actividad intercambiadora de nucleótido de los GEFs debe estar estrechamente regulada de modo que ejerzan su función de forma controlada en respuesta a estímulos específicos y evitar así una estimulación desorganizada de las GTPasas, que podría resultar en la transformación celular por la existencia de niveles altos de Ras activado.

Los GEFs de la familia Ras contienen dominios externos a la región catalítica que regulan negativamente la actividad intercambiadora. Las regiones reguladoras de GEFs como SOS1, Epac2 y RasGRP1 no tienen elementos en común (Figura I8), pero de forma general en estos GEFs los dominios reguladores reprimen la región catalítica mediante la formación de interacciones intramoleculares que estabilizan conformaciones autoinhibidas que de forma directa o alostérica interfieren con la función del dominio catalítico (Pufall and Graves, 2002).

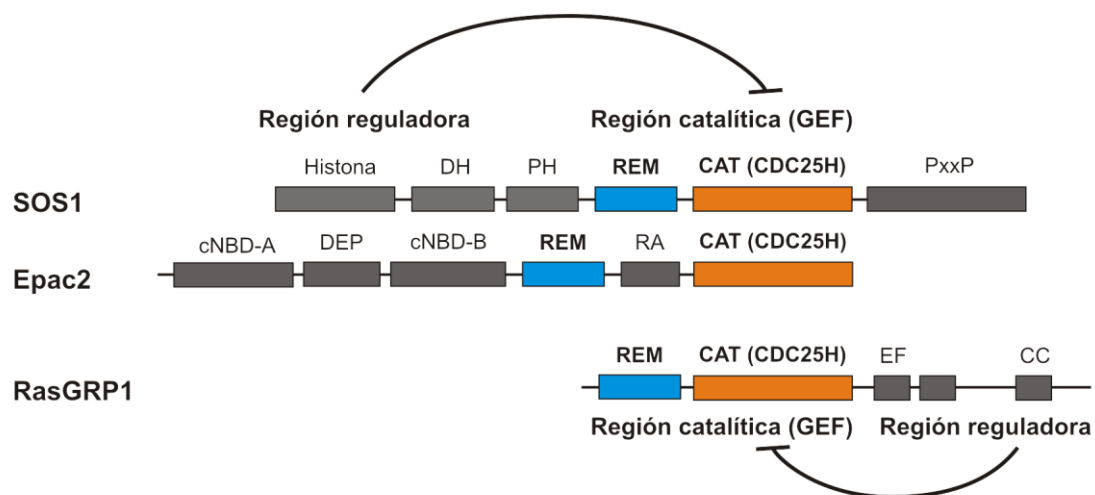


Figura 18. Mecanismos de autoinhibición en GEFs de tipo CDC25H. Estructura de los GEFs: SOS1, Epac2 y RasGRP1 (CalDAG-GEF II). En SOS1 y Epac2, regiones en posición amino a la región catalítica regulan la actividad GEF, mientras que en RasGRP1 la región reguladora se localiza en posición C-terminal. Las regiones reguladoras de estos GEFs no tienen similitud entre sí, pero de forma general las regiones reguladoras inhiben la actividad GEF de la región catalítica mediante contactos intramoleculares que estabilizan conformaciones inactivas en las que se impide o desfavorece la unión de la GTPasa al sitio activo.

2.3.1. SOS1 presenta un mecanismo de regulación alostérica

SOS es un GEF de Ras (Wang et al., 1995) que también puede estimular el intercambio de nucleótido de Rac (Nimnual et al., 1998). SOS se identificó en *D. melanogaster* como producto del gen corriente abajo de *sevenless* en la ruta de señalización Ras/MAPK (Simon et al., 1993). En mamíferos hay dos homólogos, SOS1 y SOS2, que se expresan de forma ubicua y están codificados por diferentes *loci* genéticos. SOS1 (1333 aminoácidos) y SOS2 (1332 aminoácidos) de humano muestran 74 % de identidad de secuencia.

Las proteínas SOS tienen una estructura modular en la que se distinguen 3 regiones (Figura 19). La región amino terminal, que actúa como un elemento regulador, está formada por un dominio Histona, un dominio DH y un dominio PH. El dominio Histona media la interacción con fosfolípidos como fosfatidil-inositol 4,5-bisfosfato (PI-4,5-P₂ o PIP₂) y ácido fosfatídico a través de una superficie conservada cargada positivamente (Sondermann et al., 2003; Yadav and Bar-Sagi, 2010). Los dominios DH normalmente se encuentran en GEFs de la familia Rho, lo que sugiere que SOS1 también podría tener función GEF para estas GTPasas (Nimnual et al., 1998; Soisson et al., 1998); si bien la activación de Rac1 por SOS podría estar mediada por su interacción con proteínas adaptadoras como Eps8 (Scita et al., 2001). El dominio PH une fosfolípidos como PI-4,5-P₂ y PI-3,4,5-P₃, y ácido fosfatídico (Chen et al., 1997; Koshiba et al., 1997; Kubiseski et al., 1997; Zhao et al., 2007; Zheng et al., 1997).

La región catalítica está formada por un dominio CDC25H precedido de un dominio REM. Por último, la región C-terminal tiene un dominio rico en Pro (PxxP) que interacciona con dominios SH3 de la proteína adaptadora Grb2 durante el reclutamiento de SOS1 desde el citosol hasta la membrana celular (Egan et al., 1993; Li et al., 1993; Pierre et al., 2011). Esta región también interacciona con la

subunidad p85 de PI3 kinasa (Innocenti et al., 2003); PLC γ 1 (Halupa et al., 2005; Kim et al., 2000; Scholler et al., 2000; Wu et al., 2007); y Abi1/E3b1 (Innocenti et al., 2002). Además este dominio contiene sitios de fosforilación por ERK y probablemente otras quinasas (Corbalan-Garcia et al., 1996a; Corbalan-Garcia et al., 1996b; Porfiri and McCormick, 1996; Rozakis-Adcock et al., 1995; Waters et al., 1995), que participan en la regulación de SOS inhibiendo la interacción con Grb2.

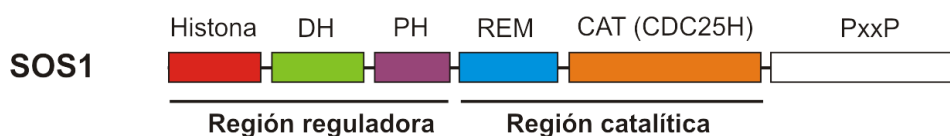


Figura I9. Estructura primaria de SOS1. La denominada región reguladora N-terminal está formada por un dominio Histona y el módulo en tándem DH-PH. La región catalítica contiene un dominio CDC25H precedido de un dominio REM. La región C-terminal tiene un dominio rico en Pro (PxxP).

La región catalítica de SOS1 alberga dos sitios de unión a Ras (Figura I10-A). El sitio catalítico, similar al observado en otros GEFs de tipo CDC25H, une de forma transitoria una molécula de Ras, de modo que la horquilla helicoidal se inserta entre las regiones *switch I* y *II* de la GTPasa desestabilizando la interacción con el nucleótido (Boriack-Sjodin et al., 1998) (ver antes). Además, SOS1 tiene un segundo sitio de unión a Ras formado por los dominios REM y CDC25H, que se localiza en posición distal con respecto al sitio activo (Margarit et al., 2003b). Este sitio distal de unión a Ras presenta mayor afinidad por RasGTP que por RasGDP (Sondermann et al., 2004). La posición de la horquilla helicoidal del dominio CDC25H se encuentra acoplada fuertemente a la orientación del dominio REM. La unión de RasGTP al sitio distal estabiliza a través del dominio REM una conformación activa de la horquilla helicoidal y estimula la actividad catalítica GEF, constituyendo un sitio de regulación alostérica (Freedman et al., 2006). La estimulación de la actividad intercambiadora de SOS1 por unión de RasGTP al sitio alostérico distal crea una retroalimentación positiva en la activación de Ras por SOS (Margarit et al., 2003a) (Figura I10-B).

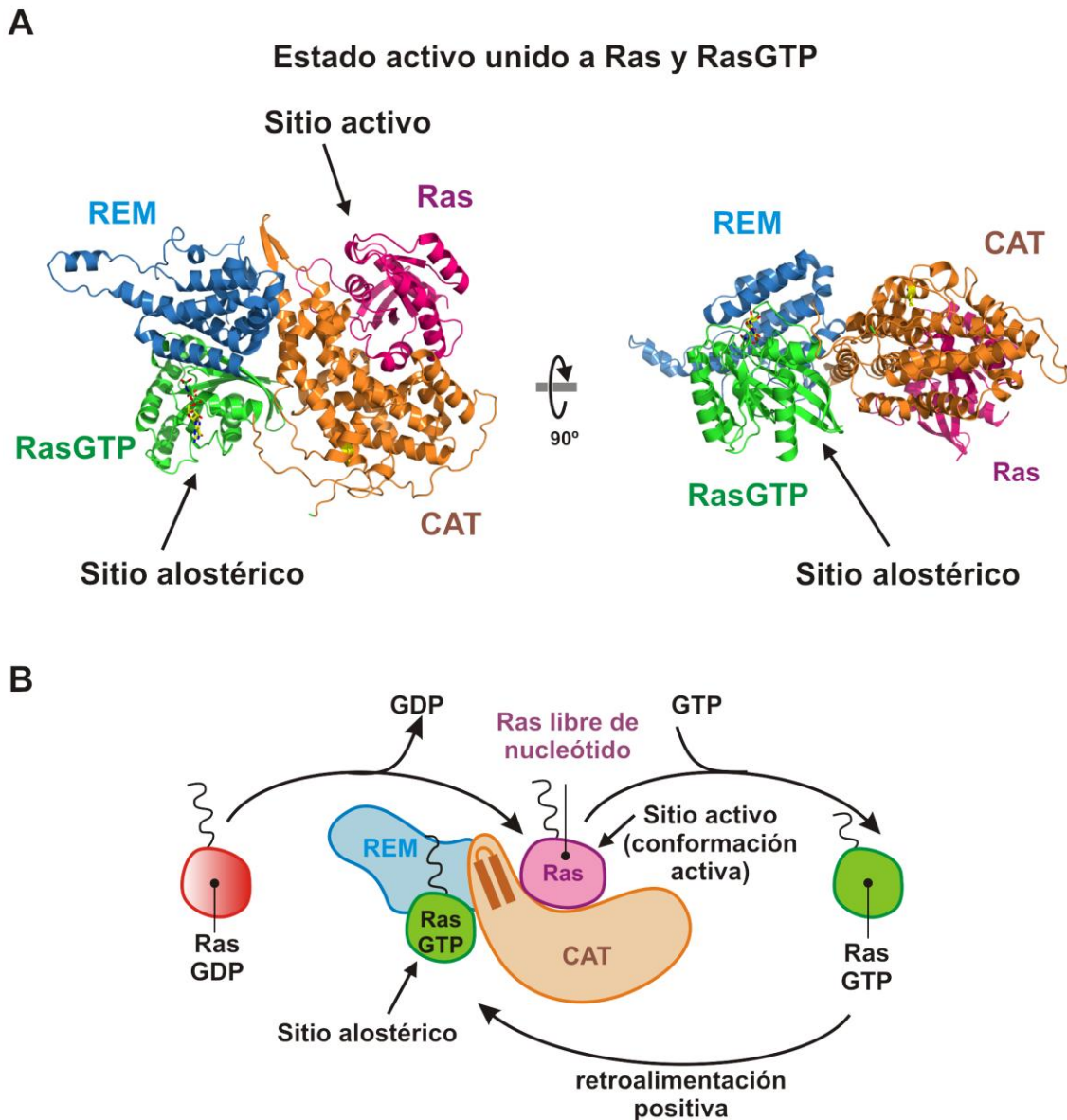


Figura 110. Activación alostérica de SOS1. (A) Estructura de la región REM-CDC25H de SOS1 unida a dos moléculas de Ras (código PDB: 1NVV) que representa el estado activo de SOS1. La molécula de Ras (magenta) unida en el sitio catalítico carece de nucleótido. La segunda molécula de RasGTP (verde) se une al sitio distal alostérico, que a través del dominio REM estabiliza la horquilla helicoidal del dominio CDC25H en una conformación activa, aumentando la actividad de Ras por el sitio catalítico de SOS1. (B) Ciclo de intercambio de nucleótido de SOS1. La unión de RasGTP al sitio distal alostérico de SOS1 estimula la actividad GEF sobre moléculas de Ras que se unen al sitio catalítico estableciendo un bucle de retroalimentación positiva. Se han omitido el resto de los dominios de SOS1 para simplificar el esquema.

La región amino terminal de SOS1, formada por los dominios Histona-DH-PH, regula negativamente la actividad GEF. En la conformación inactiva (Figura I11), la unidad DH-PH interactúa con el dominio REM bloqueando de forma estérica el acceso de RasGTP al sitio alostérico (Sondermann et al., 2004). El

dominio Histona tiene un papel dual, por un lado ocluye directamente el sitio alostérico y por otro lado estabiliza la conformación autoinhibitoria de la unidad DH-PH (Gureasko et al., 2010). En resumen, en la conformación autoinhibida de SOS1 el sitio catalítico está accesible, pero se encuentra atrapado en una conformación inactiva.

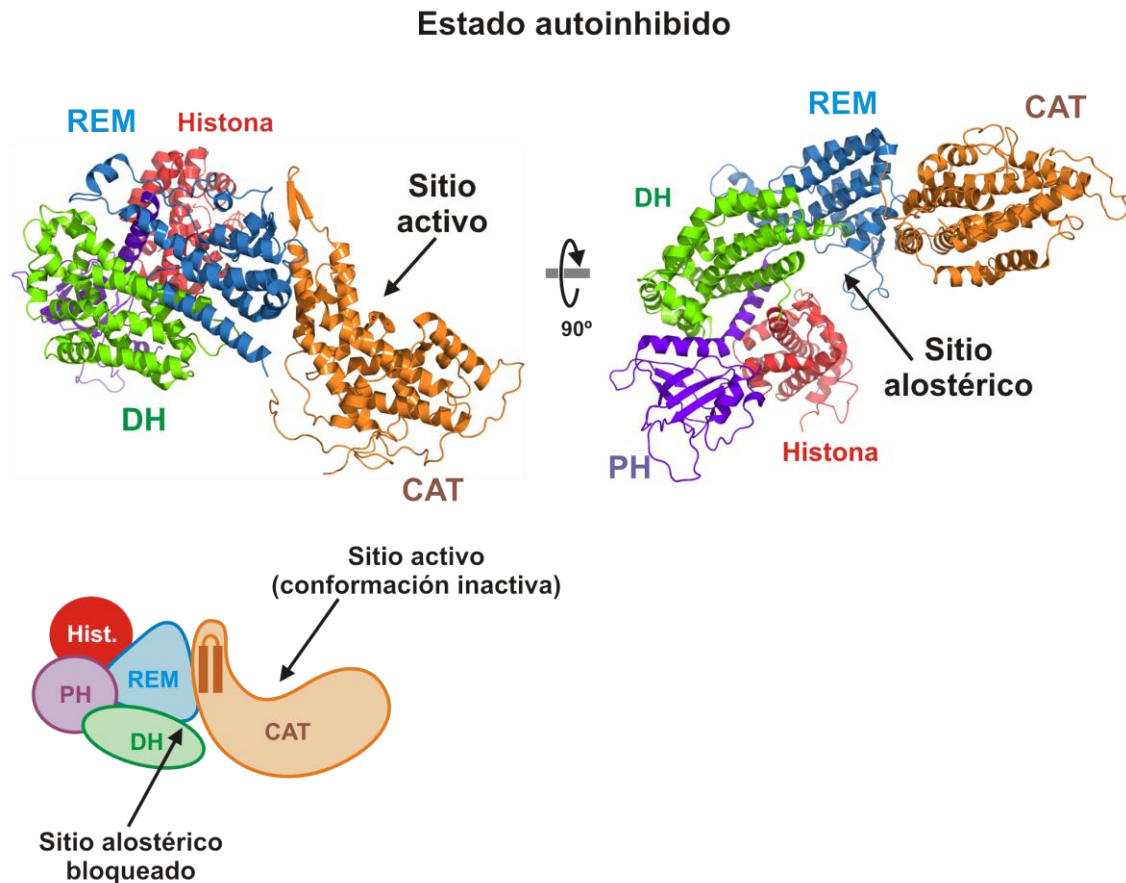


Figura I11. Mecanismo de autoinhibición de SOS1. Estructura del fragmento Histona-DH-PH-REM-CDC25H de SOS1 en una conformación autoinhibida (código del PDB: 3KSY) (Gureasko et al., 2010). El dominio PH está replegado sobre el dominio REM bloqueando el sitio alostérico aunque el dominio PH no interacciona con el dominio REM. El dominio Histona ocluye el sitio alostérico y además estabiliza la conformación inhibitoria del segmento DH-PH.

La autoinhibición puede ser liberada por distintas señales desde la membrana, procedentes de diferentes proteínas o lípidos. Por ejemplo, la interacción simultánea de SOS1 con la membrana a través de la unión del dominio PH a fosfolípidos como PI-4,5-P₂ y del dominio Histona a cargas negativas en la

membrana causaría una reorganización del segmento Histona-DH-PH que haría accesible el sitio alostérico en la región catalítica (Gureasko et al., 2010).

La relevancia fisiológica de la autoinhibición en SOS1 se pone de manifiesto en el síndrome de Noonan, en el cual hay una activación constitutiva de las rutas Ras/ERK (Roberts et al., 2013; Schubbert et al., 2007). La mayoría de las mutaciones en SOS1 asociadas con el síndrome de Noonan se localizan en la región reguladora (Histona-DH-PH) o en el dominio REM, que no se encuentran implicados en la reacción de intercambio pero sí en la función intrínseca inhibitoria de SOS1 (Roberts et al., 2007; Roberts et al., 2013; Tartaglia et al., 2007). Estas mutaciones causan un debilitamiento de contactos intramoleculares o disminuyen ciertas restricciones estructurales en la conformación de SOS1, permitiendo una mayor señalización en la ruta de Ras.

2.3.2. RasGRP1 presenta un mecanismo de regulación por impedimento estérico y dimerización

Los RasGRPs 1 a 4 son GEFs de la familia Ras que tienen patrones de expresión específicos de tejido (Jun et al., 2013). Por ejemplo, RasGRP1 se expresa en células T (Priatel et al., 2002), cerebro (Kawasaki et al., 1998) y en queratinocitos (Rambaratsingh et al., 2003). La desregulación de RasGRP1 se ha implicado en leucemia de células T (Oki et al., 2012) y lupus eritematoso sistémico (Yasuda et al., 2007).

En la secuencia de RasGRP1 se distinguen dos regiones (Figura I12). En posición N-terminal se localiza la región catalítica formada por un dominio REM y un dominio CDC25H. La región C-terminal contiene un dominio EF, un dominio C1 y una región *coiled coil*. El dominio EF está formado por dos motivos EF, de los cuales el primero une un ion Ca^{2+} , mientras que el segundo motivo no tiene capacidad de unir Ca^{2+} (Iwig et al., 2013). Tras el dominio EF se encuentra un dominio C1 de unión a DAG. Seguidamente se extiende una región potencialmente desestructurada de unos 140 residuos y por último una región *coiled coil* (Beaulieu et al., 2007; Ebinu et al., 1998). Parte del segmento C-terminal está implicado en el

reclutamiento a la membrana a través de interacciones electrostáticas con fosfoinosítidos (Zahedi et al., 2011).

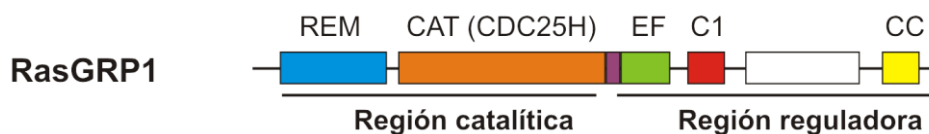


Figura I12. Estructura primaria de RasGRP1. La denominada región catalítica N-terminal contiene un dominio REM y un dominio CDC25H y se encuentra seguida por la región reguladora que contiene un dominio EF un dominio de unión a membrana C1 y una región *coiled coil*. El dominio CDC25H y el dominio EF se encuentran conectados entre sí por medio de un lazo.

Recientemente se ha descrito la estructura de un fragmento de RasGRP1 que contiene los dominios REM, CDC25H, EF y C1, que corresponde a una conformación autoinhibida (Iwig et al., 2013) (Figura I13-A). La región REM-CDC25H de RasGRP1 presenta una estructura muy similar a la de SOS1, de modo que se espera que Ras se una al dominio CDC25H de manera muy similar en ambos GEFs. El dominio REM de RasGRP1 interactúa con la horquilla helicoidal del dominio CDC25H de manera similar a la observada en SOS1, si bien, la horquilla helicoidal de RasGRP1 se encuentra rotada hacia el exterior del sitio activo unos 15 ° con respecto a la orientación en SOS1 activo, lo que es consistente con una conformación abierta que está preparada para la unión de Ras. El dominio EF contacta de manera extensa con CDC25H y la secuencia que une el dominio catalítico con el dominio EF se organiza sobre la superficie del dominio CDC25H ocupando parte del sitio de unión a Ras, de modo que impide la unión de la GTPasa e inhibe la actividad intercambiadora de nucleótidos. El dominio C1 contacta estrechamente con el dominio EF, pero no establece contactos intramoleculares directos con el dominio catalítico.

RasGRP1 es un dímero en solución y la estructura cristalográfica revela que el módulo regulador EF-C1 media la formación de un dímero en el cristal (Figura I13-B). El dominio C1 de cada molécula contacta con los dominios EF, C1 y CDC25H de la otra molécula del dímero y la dimerización oculta la superficie de los dominios C1 implicada en la unión a la membrana. La alteración mediante mutagénesis de la superficie de dimerización resulta en una activación de la

actividad GEF de RasGRP1, lo que sugiere que la formación del dímero mediada por el dominio C1 estabiliza la forma autoinhibida. La estructura del dominio *coiled coil* C-terminal de RasGRP1 forma un dímero paralelo (Iwig et al., 2013) muy similar al de otras proteínas *coiled coil* canónicas como GCN4 (O'Shea et al., 1991), que favorecería la dimerización de la región reguladora previniendo una activación inapropiada de Ras.

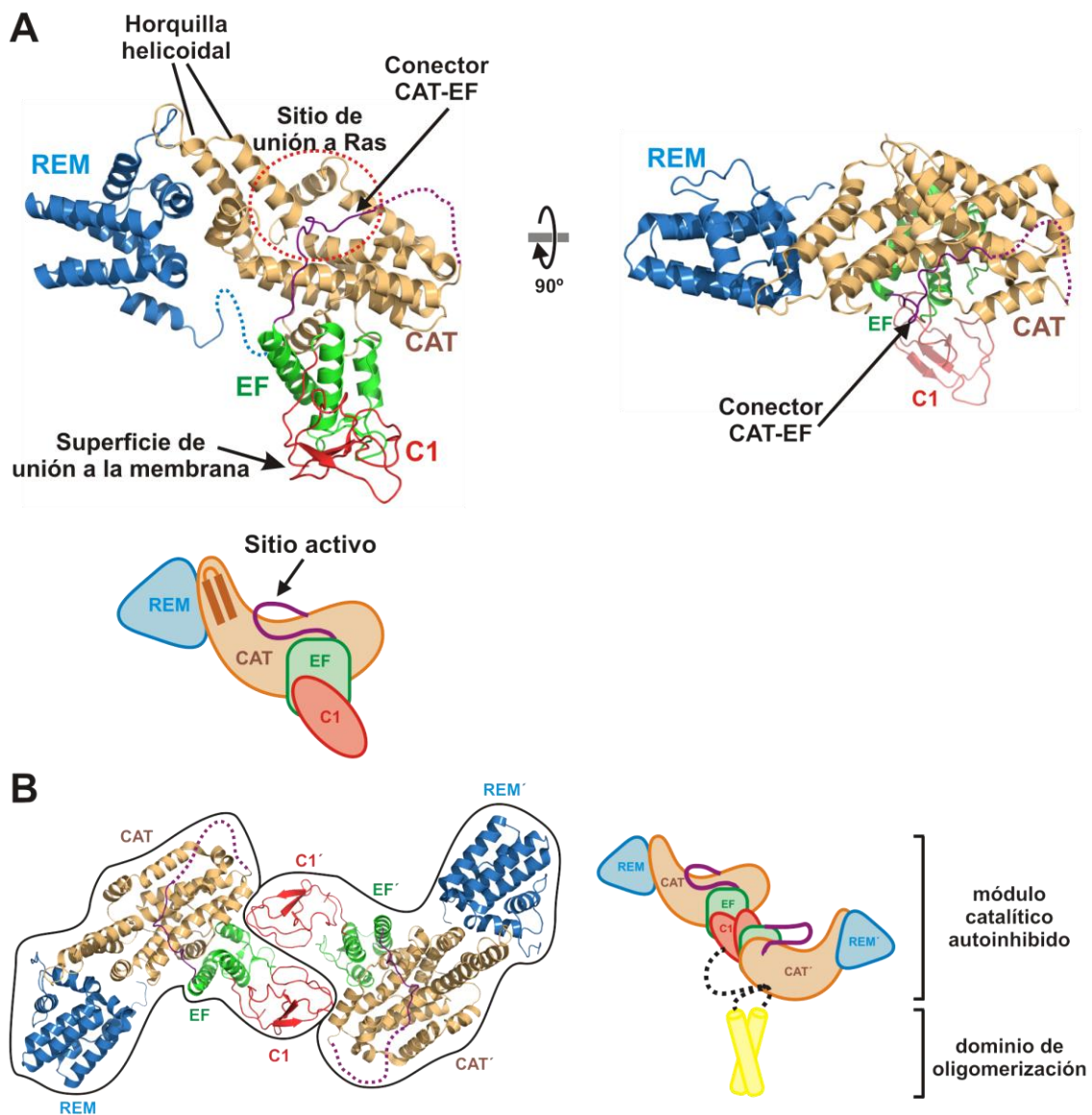


Figura I13. Estructura de la región REM-CDC25H-EF-C1 de RasGRP1 en una conformación autoinhibida. (A) Estructura cristalográfica de los 4 primeros dominios de RasGRP1 (código PDB: 4L9M). El dominio REM (azul) estabiliza la horquilla helicoidal del dominio CDC25H (naranja) en una conformación abierta que permitiría la unión de la GTPasa. El dominio EF (verde) contacta con el dominio CDC25H, mientras que el dominio C1 (rojo) no interacciona de manera directa con

CDC25H. El lazo (morado) que une los dominios CDC25H y EF se extiende sobre el CDC25H atravesando el sitio de unión a Ras, por lo que bloquea la unión de la GTPasa. (B) Estructura del dímero cristalográfico en la que se muestra cómo la dimerización a través del dominio C1 da lugar a contactos entre el dominio C1 de una molécula y el dominio CDC25H de la molécula vecina. Estos contactos favorecerían la estabilización de la conformación dimérica autoinhibida.

En resumen, en el estado autoinhibido el sitio activo del dominio CDC25H de RasGRP1 tiene una conformación aparentemente activa, pero está bloqueado por la secuencia que precede al dominio EF, la cual está estabilizada por contactos intramoleculares entre los dominios CDC25H y EF, y por contactos intermoleculares entre los dominios CDC25H, EF y C1 en el dímero.

RasGRP1 es activado por DAG (Ebinu et al., 1998) que recluta al dominio C1 a la membrana desestabilizando la dimerización de RasGRP1 y los contactos que mantiene la conformación autoinhibida. El dominio EF muestra relativamente baja afinidad por Ca^{2+} sugiriendo que RasGRP1 no une Ca^{2+} en células en reposo, con concentraciones submicromolares de Ca^{2+} intracelular. La unión de Ca^{2+} al dominio EF induce marcados cambios en la estructura de este dominio que probablemente desestabilizan los contactos que el dominio EF establece en la conformación autoinhibida, favoreciendo el desbloqueo del sitio activo.

2.3.3. Epac2 presenta un mecanismo de regulación estérico

Las proteínas Epac 1 y 2 (*Exchange proteins directly activated by cAMP*) son GEFs de Rap1 y Rap2 que están reguladas directamente por el segundo mensajero cAMP y participan en el control de diferentes procesos celulares entre los que se encuentran la adhesión celular y la secreción de insulina (de Rooij et al., 1998; Gloerich and Bos, 2010).

En las secuencias de Epac1 y Epac2, de 923 y 1011 aminoácidos respectivamente en humanos, se distinguen dos regiones (Figura I14). En Epac2 la región reguladora N-terminal contiene dos dominios de unión a nucleótidos cíclicos (cNBD-A y cNBD-B) que unen cAMP, si bien el cNBD-A tiene baja afinidad por cAMP. Entre los dominios cNBD se extiende un dominio DEP implicado en la asociación a membranas. La mitad C-terminal corresponde a la región catalítica que contiene un dominio REM y un dominio CDC25H. Además, esta región alberga

un dominio RA, con plegamiento de tipo ubiquitina, insertado entre los dominios REM y CDC25H. La principal diferencia entre Epac1 y Epac2 es que la primera carece del primer cNBD (de Rooij et al., 2000).

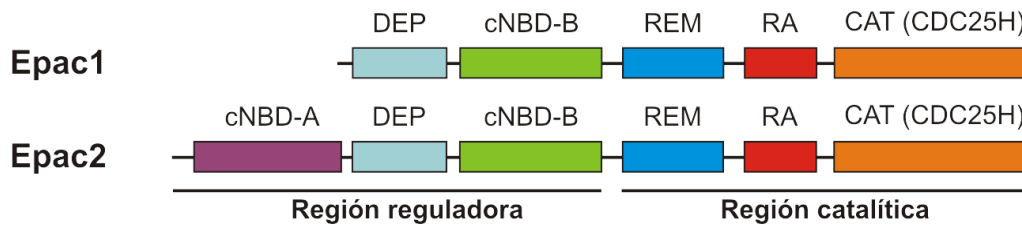


Figura I14. Estructura primaria de Epac1 y de Epac2. La región reguladora de Epac2 está formada por dos dominios cNBD y un dominio DEP. Epac1 carece del primer dominio cNBD. La región catalítica contiene dominios REM y CDC25H entre los que se sitúa un dominio RA.

La estructura cristalográfica de Epac2 en ausencia de cAMP revela una conformación autoinhibida (Rehmann et al., 2008) (Figura I15-A) en la que el cNBD-B contacta con la horquilla helicoidal del dominio CDC25H bloqueando parcialmente el sitio de unión de la GTPasa. El dominio cNBD-A establece extensos contactos con el cNBD-B y aunque no contacta directamente con el dominio CDC25H bloquea el acceso al sitio de unión de Rap. Por último, el dominio DEP media contactos adicionales entre la región reguladora y el dominio CDC25H, si bien no participa directamente en el bloqueo del sitio activo. El dominio CDC25H y en concreto la horquilla helicoidal tienen una estructura similar a la observada en CDC25H de SOS1 unido a Ras, lo que sugiere que el dominio catalítico de Epac2 mantiene una conformación activa. En resumen, en el estado autoinhibido la región reguladora de Epac2 interacciona con el dominio catalítico y bloquea el sitio activo por impedimento estérico, inhibiendo la actividad GEF.

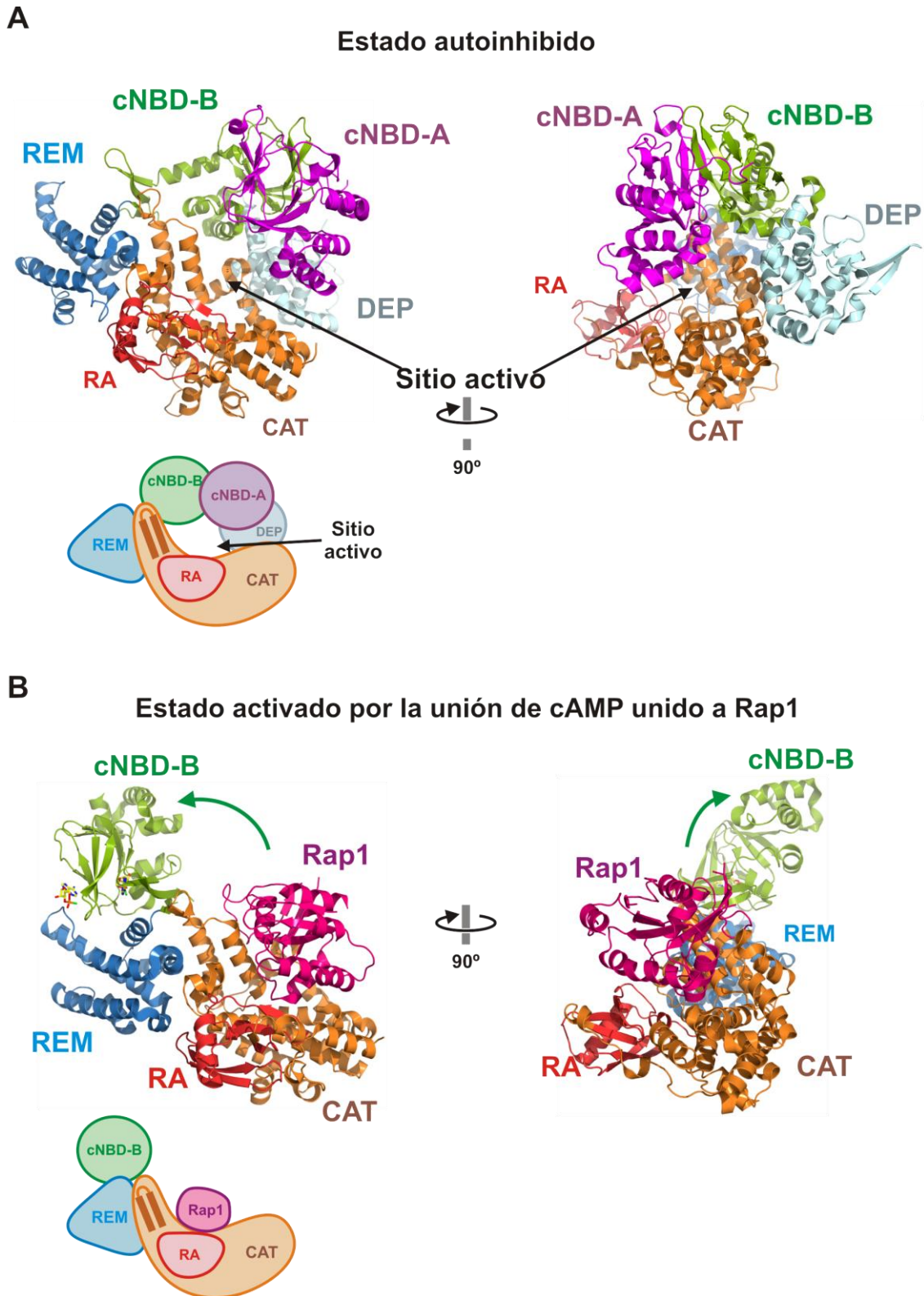


Figura I15. Mecanismo de regulación de Epac2. (A) Representación de la estructura de Epac2 en estado autoinhibido (código PDB: 2BYV). La región reguladora, formada por los dominios cNBD-A (magenta), DEP (cian) y cNBD-B (verde), está replegada sobre el dominio CDC25H (naranja) bloqueando el sitio activo. La inhibición de la actividad GEF se produce por impedimento estérico. (B) Estructura del fragmento cNBD-B-REM-RA-CDC25H de Epac2 unido a un análogo de cAMP y a Rap1 (magenta) (código PDB: 3CF6) correspondiente a un estado activo. La unión de cAMP induce

un cambio conformacional en el que la región reguladora se aleja del sitio activo como indica la flecha, permitiendo la unión de la GTPasa.

La unión de cAMP al segundo dominio cNBD-B resulta en la activación de Epac2 (Figura I15-B), tal como se ha observado en la estructura de un fragmento de Epac2 que contiene los dominios cNBD-B, REM, RA y CDC25H, unido a un análogo de cAMP y a Rap1 (Rehmann et al., 2008). La unión de cAMP al cNBD-B causa un cambio conformacional en este dominio que está acompañado de una rotación del cNBD-B de modo que se desplaza e interacciona con el dominio REM, alejándose del dominio CDC25H, dando lugar a una conformación abierta, de forma que el sitio catalítico queda accesible al sustrato y esto conduce a la activación (Rehmann et al., 2008). La activación de Epac2 por cAMP parece limitarse a la liberación del impedimento estérico ya que no hay diferencias significativas entre las estructuras del dominio CDC25H en los estados autoinhibido y activo; es decir, cAMP no parece inducir un cambio alostérico en el sitio activo de unión a Rap.

Epac2 también interacciona con RasGTP a través del dominio RA y esta interacción es necesaria para la activación de Rap1. La unión de RasGTP recluta a Epac2 hasta la membrana aumentando su concentración local y acelerando la activación de Rap1, pero no afecta a la autoinhibición ni induce cambios alostéricos en el dominio CDC25H (Liu et al., 2008). La interacción de Epac2 con Ras es independiente de cAMP, en consonancia con la accesibilidad del sitio de unión a Ras en el dominio RA en la estructura autoinhibida de Epac2.

En resumen, la actividad intercambiadora de SOS1, RasGRP1 y Epac2 está autoinhibida por regiones reguladoras que establecen interacciones intramoleculares con la región catalítica REM-CDC25H. Sin embargo, cada uno de estos GEFs muestra mecanismos de autoinhibición y regulación específicos que se resumen en la Figura I16.

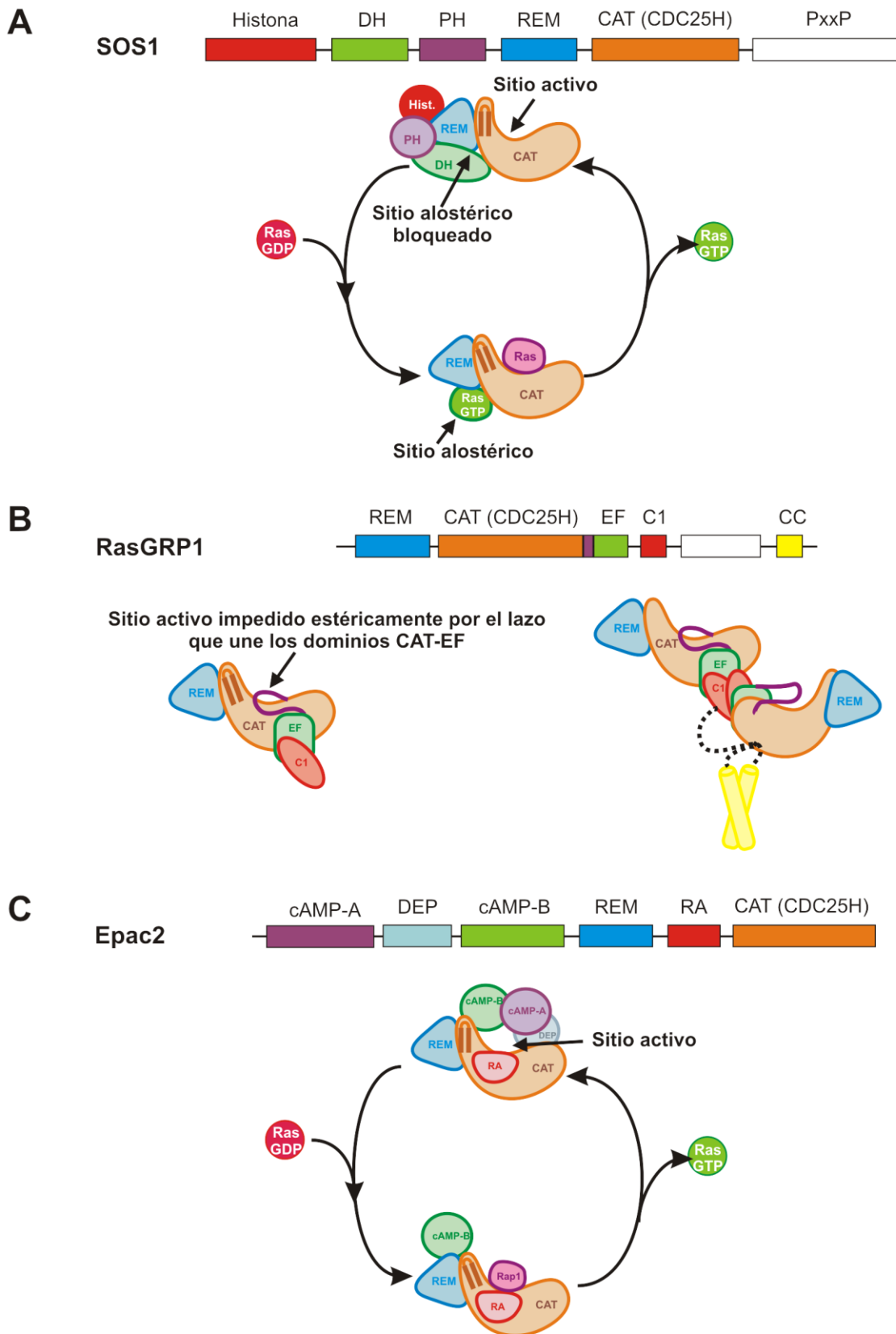


Figura I16. Mecanismos de regulación específicos de SOS1, RasGRP1 y Epac2. (A) SOS1 presenta un mecanismo de regulación alostérico, donde la región reguladora bloquea el sitio distal de RasGTP. (B) La forma autoinhibida de RasGRP1 presenta un mecanismo de regulación por impedimento estérico y dimerización. (C) En Epac2, la región reguladora bloquea el sitio catalítico, por lo que se trata de una regulación por impedimento estérico.

OBJETIVOS

El objetivo general de esta tesis es avanzar en el conocimiento de la estructura de la proteína C3G. Para ello se han abordado los siguientes objetivos específicos:

1. Estudio de la organización modular de C3G mediante el análisis de interacciones entre dominios.
2. Caracterización de la estructura de la región amino terminal de C3G.

MATERIALES Y MÉTODOS

1. Secuencias de ADNc de C3G

Se han empleado dos secuencias de ADNc que codifican la proteína C3G de humano (la numeración de aminoácidos corresponde al código Q13905-1 de Uniprot). Para crear las construcciones en vectores de expresión en células de mamífero se ha utilizado la secuencia de ADNc de C3G humano (Guerrero et al., 1998) que es prácticamente idéntica a la secuencia NM_005312 de NCBI, pero difiere de ésta en que el extremo amino de la proteína que codifica es MSGKIEKA en nuestra secuencia, en lugar de MDT en la secuencia de Uniprot. Por eso las construcciones que comienzan en el extremo amino lo hacen a partir del cuarto residuo de Q13905-1, que es la primera posición común en ambas secuencias. Además, la secuencia utilizada difiere de las descritas en NCBI y en Uniprot en que la posición 281 está ocupada por Gly en vez de Ser. Nos referimos a esta secuencia como C3G_h.

Por otro lado se ha empleado una secuencia optimizada de ADNc, diseñada para hacer uso de los codones más comunes empleados por *E. coli*. Este ADNc se obtuvo de la empresa GenScript Corporation en forma de una construcción en la que el ADNc de C3G estaba clonado en el vector pUC57 empleando los sitios de restricción EcoRI y HindIII. Denominamos a este ADNc sintético como C3G_{opt}, el cual se empleó en las construcciones de C3G realizadas en vectores de expresión en bacterias. El ADNc de C3G_{opt} codifica una Ser en la posición 281. Salvo esta diferencia, las secuencias de aminoácidos codificadas por C3G_h y C3G_{opt} son idénticas.

2. Métodos generales de manipulación y caracterización de ácidos nucleicos

2.1. Purificación y cuantificación de ADN

El aislamiento de los ADNs plasmídicos a partir de *E. coli* se realizó en dos escalas. La purificación rutinaria de ADN a pequeña escala, para las tareas de clonación, secuenciación, etc, se realizó empleando los sistemas comerciales Wizard Plus SV Minipreps DNA Purification System (Promega) o GeneJET Plasmid Miniprep Kit (Fermentas). Alternativamente, cuando se requirió obtener mayores cantidades de ADN, por ejemplo para la realización de transfecciones de cultivos

celulares, se emplearon los sistemas Genopure Plasmid Maxi Kit (Roche) o GeneJET Plasmid Maxiprep Kit (Fermentas). En todos los casos se siguieron las instrucciones del fabricante. La concentración de ADN se determinó espectrofotométricamente en un espectrofotómetro Nanodrop (Thermo). Se utilizó un coeficiente de extinción molar a 260 nm de $50 \text{ (ng/}\mu\text{L)}^{-1} \text{ cm}^{-1}$.

2.2. Electroforesis de ADN

Los plásmidos y los fragmentos de ADN se separaron mediante electroforesis en geles de agarosa D-1 (Pronadisa) entre 0.8-1.2 % (p/v). Los geles se prepararon en cubetas de 6 x 10 cm (BioRad) empleando tampón TAE 1x (40 mM Tris-Acetato, pH 8.0, 1 mM EDTA) y se visualizaron con tinción con bromuro de etidio (0.1 $\mu\text{g/mL}$) bajo luz ultravioleta. Como marcadores de peso molecular se utilizó el DNA 1kb ladder (Promega). Las electroforesis se realizaron en tampón TAE 1x a voltaje constante. Cuando fue necesario purificar fragmentos de ADN, las bandas correspondientes se cortaron del gel y el ADN se extrajo empleando el sistema comercial High Pure PCR Product Purification (Roche) siguiendo las instrucciones del fabricante.

2.3. Secuenciación de ADN

La secuenciación de ADN se realizó en la Unidad de Genómica del Centro de Investigación del Cáncer de Salamanca. Se emplearon por un lado oligonucleótidos que hibridan en secuencias de los vectores (Tabla M1). Además, cuando fue necesario se utilizaron cebadores que hibridan en regiones de la secuencia de C3G (Tabla M2).

Tabla M1. Oligonucleótidos de uso general utilizados para la secuenciación de ADNc.

OLIGONUCLEÓTI DO	SECUENCIA	VECTORES
T7 PROMOTOR	5' TAATACGACTCACTATAGGG 3'	pETEV15b, pcDNA3
T7 TERMINADOR	5' GCTAGTTATTGCTCAGCGG 3'	pETEV15b
SP6	5' GATTTAGGTGACACTATAG 3'	pcDNA3
CMV-F	5' CGCAATGGGCGGTAGGCGTG 3'	pCEFLAG, pcDNA3
pCEFLAG-F	5' GAGCTCTCTGGCTAACTAGAG 3'	pCEFLAG
pCEFLAU-HA-F	5' GAGTTTGGATCTTGGTTCATTCTC 3'	pCEFL-HA
BGH-R	5' TAGAAGGCACAGTCGAGGC 3'	pCEFLAG, pCEFL-HA

Tabla M2. Oligonucleótidos específicos para la secuenciación de ADNc de C3G.

OLIGONUCLEÓTI DO ^a	SECUENCIA ^b
C3Gh-SEQ004	5' GACTCTCAGCGTTCTCATCTC 3' D S Q R S H L 4
C3Gh-SEQ246	5' AAGACGACTGGGATGTCACAG 3' K T T G M S Q 246
C3Gh-SEQ470	5' GACGGCTCTGGCTGCAGG 3' D G S G C R 470
C3Gh-SEQ695	5' GGATCTGGGGACATCTTACTG 3' G S G D I L L 695
C3Gh-SEQ918	5' AAGCACTTGC GGAAGCTGAATAAC 3' K H L R K L N N 918

^a El nombre de los oligonucleótidos se corresponde con la numeración del primer residuo codificado por la secuencia de nucleótidos.

^b Bajo la secuencia de nucleótidos se muestra la secuencia de aminoácidos que codifica esa región del ADNc de C3G.

2.4. Mutagénesis dirigida

La introducción de sustituciones puntuales en el ADNc de C3G se realizó mediante mutagénesis dirigida empleando el método QuikChange (Stratagene). El método consiste en amplificar mediante PCR las dos cadenas de ADN del vector empleando dos oligonucleótidos complementarios, directo e inverso, que contienen la mutación deseada (Tablas M3 y M4). En estas reacciones se utilizó la ADN polimerasa Pfu Turbo (Stratagene). Los ciclos empleados en las PCR fueron típicamente: un ciclo de desnaturalización a 95 °C durante 30 s, 18 ciclos que consistían en: 30 s a 95 °C, 1 min a 55 °C y extensión a 68 °C durante 2 min por

cada 1000 pares de bases amplificadas. El ADN parental se eliminó mediante digestión con DpnI (2 h a 37 °C), ya que esta enzima sólo digiere el ADN metilado y no afecta al nuevo ADN amplificado. El producto de la digestión se transformó en la cepa DH5α de *E. coli*. La correcta introducción de las mutaciones se verificó mediante secuenciación completa del ADNc de interés.

Tabla M3. Oligonucleótidos empleados para introducir la mutación C154A en C3Gopt.

MUTACIÓN	OLIGONUCLEÓTIDO	
	NOMBRE	SECUENCIA ^a
Cys-154-Ala	C3G-C154A-F	5' CTAGCGCACTGAGCAGC <u>GC</u> CTATAGCCGTGTGTATC 3'
	C3G-C154A-R	5' GATACACACGGCTATAG <u>GC</u> GCTGCTCAGTGCCTAG 3'

^a Se subraya la secuencia del codón cambiado

Tabla M4. Oligonucleótidos empleados para introducir mutaciones en C3Gh.

MUTACIÓN	OLIGONUCLEÓTIDO	
	NOMBRE	SECUENCIA
Tyr-504-Glu	C3G-Y504E-F	5' ATCCCATCCGTCCCC <u>GAA</u> GCGCCCTTTGCTGC 3'
	C3G-Y504E-R	5' GCAGCAAAGGGCGC <u>TTC</u> GGGGACGGATGGGAT 3'
Leu-714-Arg	C3G-L714R-F	5' CAGGAAAGATTGGTG <u>CGG</u> TACTGCGAGGCATTCCTG 3'
	C3G-L714R-R	5' CAGGAATGCCTCGCAGTAC <u>CG</u> CACCAAATCTTTCCTG 3'
Glu-731-Arg	C3G-E731R-F	5' CATCTCCCCAGAG <u>CGG</u> CTCATCAAGAAGCTGCAGTAC 3'
	C3G-E731R-R	5' GTACTGCAGCTTCTTGATGAGC <u>CG</u> CTCTGGGGAGATG 3'
Phe-760-Arg	C3G-F760R-F	5' GTCAGCAAGAACACGTT <u>CGC</u> TGCTGGTACGGGTGG 3'
	C3G-F760R-R	5' CCACCCGTACCAGCACG <u>CG</u> GAACGTGTTCTTGCTGAC 3'
Leu-774-Arg	C3G-L774R-F	5' CTGCCTGGTGGAG <u>CGG</u> ACAGAAGAGATCCTGAAGC 3'
	C3G-L774R-R	5' GCTTCAGGATCTCTTCTGT <u>CG</u> CTCCACCAGGCAG 3'
Glu-784-Arg	C3G-E784R-F	5' GAAGCTGCTGATG <u>AGA</u> CTGGTCTTCCGCCTGGTGTGC 3'
	C3G-E784R-R	5' GCACACCAGGCGGAAGACCAGT <u>CT</u> CATCAGCAGCTTC 3'
Leu-803-Glu	C3G-K803E-F	5' GTGTGCTCCGG <u>GAG</u> AACATCCTGGACAAGGTGG 3'
	C3G-K803E-R	5' CCACCTTGTCAGGATGTT <u>CT</u> CCGGAGCACAC 3'

3. Métodos generales de manipulación y caracterización de proteínas

3.1. Cuantificación de proteínas

La concentración de proteína en las muestras de fragmentos de C3G expresados en bacteria y purificados se determinó espectrofotométricamente

utilizando un espectrofotómetro U-2001 (HITACHI) a partir de la absorbancia a 280 nm. Los coeficientes de extinción molar (Tabla M5) se calcularon a partir de la secuencia de aminoácidos (Pace et al., 1995) empleando el servidor ExPASy (<http://www.expasy.org>).

Tabla M5. Parámetros de la colección de fragmentos de C3G.

FRAGMENTO		$\epsilon_{280\text{nm}}$ ($\text{M}^{-1} \text{cm}^{-1}$) ^b	$\epsilon_{280\text{nm}}$ ($(\text{mg mL}^{-1})^{-1} \text{cm}^{-1}$) ^b	MM (Da) ^b	pI ^b
NOMBRE	REGIÓN ^a				
N-C3G	(4 – 245)	16960	0.619	27406.6	9.13
N-C3G	(65 – 245)	15470	0.755	20488.5	6.22
N-C3G-C154A	(65 – 245)	15470 ^b	0.756	20456.5	6.22

^a Además de las regiones indicadas, las proteínas contienen el tetrapéptido GSHM en el extremo amino.

^b Valores calculados a partir de la secuencia teórica de aminoácidos.

Cuando no fue posible la medida de la concentración de proteína espectrofotométricamente a 280 nm, la cuantificación se realizó por el ensayo de Bradford (Bradford, 1976) usando albúmina de suero bovino (BSA) como patrón.

3.2. Electroforesis de proteínas

Las muestras de proteínas se analizaron mediante electroforesis desnaturizante en geles de poliacrilamida (SDS-PAGE) entre 7.5 y 13 % (p/v) siguiendo el método de Laemmli (Laemmli, 1970) utilizando sistemas electroforéticos Mini-Protean (Bio-Rad). Las muestras se prepararon mezclándolas con un volumen igual de tampón de solubilización 2x (SB) que consiste en: 125 mM Tris-HCl pH 6.8, 4 % (p/v) SDS, 2 % (v/v) β -mercaptoetanol, 0.02 % (p/v) azul de bromofenol y 55 % (v/v) glicerol. Las proteínas se desnaturalizaron a 95 °C durante 5 min. Las electroforesis se realizaron a intensidad constante en tampón 25 mM Tris, 200 mM glicina y 0.05 % (p/v) SDS. Tras la separación de las muestras, las proteínas se visualizaron mediante tinción con azul de Coomassie o se transfirieron a membranas para su inmunodetección. Como marcadores de peso molecular se utilizaron los patrones preteñidos Precision Plus (Bio-Rad).

3.3. Inmunodetección de proteínas: *Western blot*

Las proteínas se transfirieron a membranas Immobilon-P (Millipore) de PVDF utilizando un equipo de transferencia húmeda Mini Trans-Blot (Bio-Rad), empleando como tampón 25 mM Tris pH 8.5, 192 mM glicina, 20 % (v/v) metanol. La transferencia se realizó a intensidad constante (300 mA) típicamente durante 2 h a 4 °C.

Una vez transferidas las proteínas, la membrana se bloqueó mediante incubación con 5 % (p/v) de leche en polvo desnatada en tampón TBST (25 mM Tris-HCl pH 8.0, 50 mM NaCl, 2.5 mM KCl, 0.05 % Tween-20) durante 2 h a temperatura ambiente con agitación suave, o alternativamente durante 12 h a 4 °C; tras lo cual las membranas se lavaron 3 veces con TBST. Seguidamente las membranas se incubaron con el anticuerpo primario diluido en TBST suplementado con 2 % (p/v) de BSA, según se indica en la Tabla M6. Tras lo cual se realizaron 3 lavados con TBST de 5 min cada uno. Seguidamente la membrana se incubó con el anticuerpo secundario (Tabla M7) diluido en TBST con 5 % de leche en polvo desnatada, durante 1 h a temperatura ambiente con agitación. Tras la incubación con el anticuerpo secundario se realizaban 3 lavados con TBST de 5 min cada uno.

Cuando se emplearon anticuerpos secundarios conjugados con peroxidasa, éstos se detectaron empleando los sistemas ECL y ECL Plus (GE Healthcare) basados en un compuesto quimioluminiscente que al ser procesado por la peroxidasa emite una señal de luminiscencia que se detectó empleando películas Super RX (Fujifilm). Alternativamente se utilizaron anticuerpos secundarios conjugados con cromóforos que se detectaron mediante fluorescencia en el infrarrojo utilizando un lector Odyssey Infrared Imaging System (Li-Cor).

Tabla M6. Anticuerpos primarios utilizados para *Western blot*.

Anticuerpo	Isotipo	Casa comercial	Referencia	Dilución	Incubación	
					Tiempo	Temp.
HA (11)	ratón	Covance	mms-101r	1 : 1000	12 h	4 °C
HA (Y11)	conejo	Santa Cruz Biotechnology	sc-805	1 : 1000	12 h	4 °C
FLAG-M2	ratón	SIGMA	F1804	1 : 1000	12 h	4 °C
β-ACTIN	ratón	SIGMA	A5441	1 : 2000	12 h	4 °C
GFP B-2)	ratón	Santa Cruz Biotechnology	sc-9996	1 : 500	1 – 2 h	20 °C
C3G (C-19)	conejo	Santa Cruz Biotechnology	sc-869	1 : 500	1 – 2 h	20 °C
C3G(H-300)	conejo	Santa Cruz Biotechnology	sc-15359	1 : 500	1 – 2 h	20 °C

Tabla M7. Anticuerpos secundarios utilizados para *Western blot*.

Anticuerpo	Isotipo	Casa comercial	Referencia	Dilución
anti-mouse IgG HRP	oveja	GE Healthcare	NXA931	1 : 5000
anti-rabbit IgG HRP	cabra	Santa Cruz Biotechnology	sc-2004	1 : 5000
anti-mouse IgG DyLight 680	cabra	Pierce	35518	1 : 5000
anti-rabbit IgG DyLight 680	cabra	Pierce	35568	1 : 5000

4. Métodos para estudios en cultivos celulares

4.1. Construcciones para la expresión de C3G en células de mamífero

4.1.1. Clonación en vectores de la familia pCEFL

Para la expresión en células de mamífero de fragmentos de *C3Gh* que incorporan una etiqueta del antígeno de hemaglutinina del virus de la gripe (HA) o una etiqueta FLAG, se emplearon los vectores pCEFLHA y pCEFLAG (Chiariello et al., 2000). Estos vectores contienen genes de resistencia a ampicilina y a neomicina. La zona de clonación contiene sitios únicos de corte, entre otros para las enzimas BglII, EcoRI y NotI (Figura M1). Las proteínas de fusión codificadas contienen la secuencia HA (Figura M1-A) o FLAG (Figura M1-B) en el extremo amino. La transcripción del gen de interés está en ambos vectores bajo el control del promotor del factor de elongación 1 α .

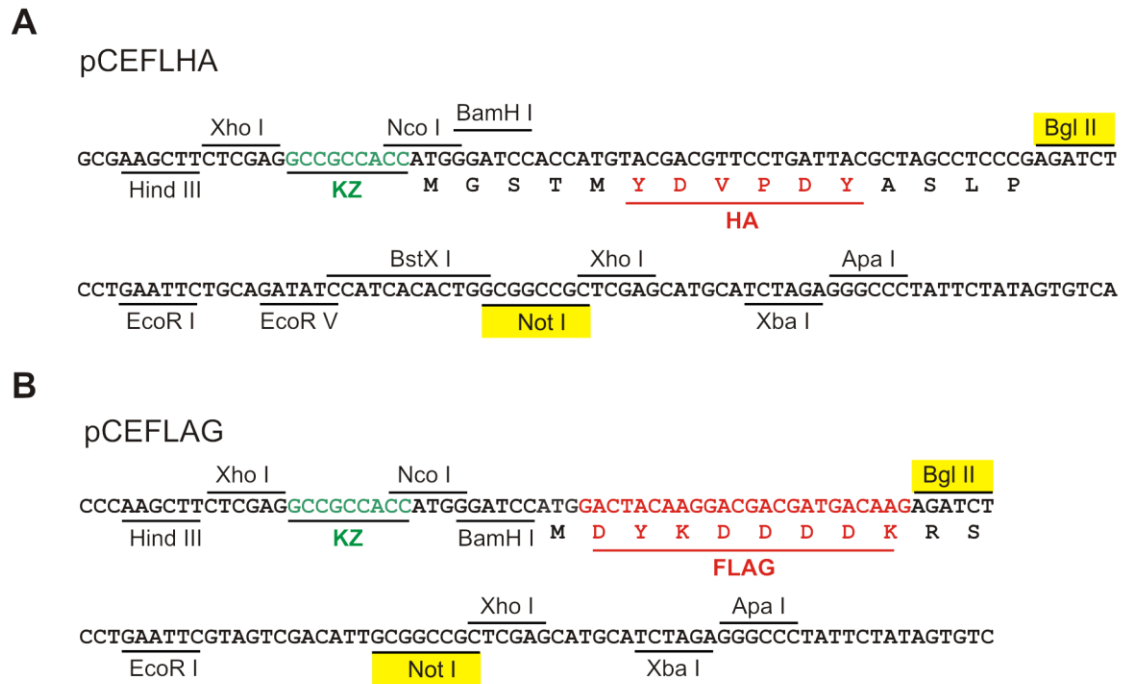


Figura M1. Regiones de clonación y expresión de los vectores pCEFLHA y pCEFLAG. Se destacan en amarillo los sitios de restricción empleados, en verde la secuencia Kozak y en rojo las etiquetas HA (A) y FLAG (B) respectivamente.

Los ADNc de diferentes regiones de C3G humano se amplificaron mediante PCR usando como molde una construcción de C3Gh (4-1077) en el vector pETEV15bNco de expresión en bacterias (de Pereda y Guerrero, resultados sin publicar). Se utilizó la ADN polimerasa Pfu-Turbo (Stratagene, Agilent) y cebadores específicos de C3Gh. Los oligonucleótidos directos contienen un sitio BglII, mientras que los oligonucleótidos reversos contienen un codón de terminación y un sitio NotI (Tabla M8). Los productos de la PCR se digirieron con las enzimas BglII y NotI (Fermentas) y se clonaron en el vector pCEFLAG previamente digerido con las mismas enzimas y desfosforilado con la fosfatasa alcalina Fast-AP (Fermentas). La ausencia de errores en los ADNc se verificó mediante secuenciación. Una vez confirmadas las secuencias de C3Gh en el vector pCEFLAG, éstas se clonaron en el vector pCEFLHA por digestión con BglII y NotI y ligación en dicho vector empleando esos mismos sitios de restricción.

Tabla M8. Oligonucleótidos utilizados para amplificar fragmentos de *C3Gh* para su clonación en los vectores pCEFLAG o pCEFLHA.

OLIGONUCLEÓTIDO	SECUENCIA
C3Gh-004F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> GAC TCT CAG CGT TCT CAT C 3' EcoR I Bgl II D S Q R S H 4
C3Gh-065F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> TTG GAT CTG GAG CAG CAG GC 3' EcoR I Bgl II L D L E Q Q 65
C3Gh-246F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> AAG ACG ACT GGG ATG TCA CAG 3' EcoR I Bgl II K T T G M S Q 246
C3Gh-277F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> GTG GTT GAT AAT GGT CCT CC 3' EcoR I Bgl II V V D N G P 277
C3Gh-292F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> CAG TCG GCG CCG TCC 3' EcoR I Bgl II Q S A P S 292
C3Gh-465F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> GCC TCC CAG ACG GCG G 3' EcoR I Bgl II A S Q T A 465
C3Gh-550F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> CAC ATG CTG GCC TAC ATG C 3' EcoR I Bgl II H M L A Y M 550
C3Gh-670F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> CTC ATT GAC CAC AAC GAA ATT ATG 3' EcoR I Bgl II L I D H N E I M 670
C3Gh-815F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> CTC AGG TGT GCC ACC TCC 3' EcoR I Bgl II L R C A T S 815
C3Gh-245R-Not	5' <u>GAA TTC</u> <u>G CGG CCG CTA</u> GTT CAG GAT CTC TAC CTC GCG 3' EcoRI NotI STOP N L I E V E R 245
C3Gh-669R-Not	5' <u>GAA TTC</u> <u>G CGG CCG CTA</u> GGA CAG CTC GTC CAC TTC CTC 3' EcoRI NotI STOP S L E D V E E 669
C3Gh-814R-Not	5' <u>GAA TTC</u> <u>G CGG CCG CTA</u> TAG CTT CTT CTG GTC CAC C 3' EcoRI NotI STOP L K K Q D V 814
C3Gh-1077R-Not	5' <u>GAA TTC</u> <u>G CGG CCG CTA</u> GGT CTT CTC TTC CCG GTC 3' EcoRI NotI STOP T K E E R D 1077

4.1.2. Vectores con proteínas fluorescentes para FRET

En los ensayos de FRET se emplearon construcciones que codificaban fragmentos de *C3Gh* fusionados a las proteínas fluorescentes CFP, Venus o a ambas. Para ello, en primer lugar, se crearon vectores que codifican para estas proteínas fluorescentes.

4.1.2.1. Construcción de vectores con proteínas fluorescentes para FRET

A partir del vector pcDNA3 (Invitrogen) se crearon tres vectores que codifican las proteínas fluorescentes CFP y Venus.

- pcDNA3-CFP, codifica la proteína CFP en la posición N-terminal.
- pcDNA3-Venus, codifica la proteína Venus en la posición C-terminal.
- pcDNA3-CFP-Venus, codifica la proteína CFP en la posición N-terminal y Venus en posición C-terminal.

Para crear el vector pcDNA3-CFP, el ADNc de CFP se amplificó mediante PCR usando como molde el vector pCFP-N1 (BD Biosciences Clontech) y empleando la polimerasa Pfu Turbo. Se utilizó el cebador directo CFP-Hind-For que contiene un sitio de reconocimiento de la enzima HindIII e incorpora una secuencia Kozak, y el oligonucleótido reverso CFP-Bam-Rev que contiene un sitio para la enzima BamHI (Tabla M9). El producto de la PCR se digirió con HindIII y BamHI (Fermentas) y se clonó en el vector pcDNA3 previamente digerido con las mismas enzimas y desfosforilado.

Para crear el vector pcDNA3-Venus, el ADNc que codifica para Venus se amplificó a partir del vector pVenus-N1 (una variante del vector pEYFD de BD Biosciences Clontech). Se utilizó el oligonucleótido directo Venus-Eco-For que contiene sitios de reconocimiento para EcoRI y NotI, y el oligonucleótido reverso Venus-Xho-Rev que contiene un codón de terminación y un sitio XhoI (Tabla M9). El producto de la PCR se digirió con EcoRI y XhoI (Fermentas) y se clonó en el vector pcDNA3 previamente digerido con las mismas enzimas y desfosforilado.

Tabla M9. Oligonucleótidos utilizados para amplificar los ADNc de las proteínas fluorescentes CFP y Venus.

OLIGONUCLEÓTIDO	SECUENCIA
CFP-Hind-For	5' TGA <u>GAA TTC</u> <u>AAG CTT</u> <u>GCC</u> <u>GCC ACC ATG G</u> TG AGC AAG GGC 3'
	EcoRI HindIII KZ M V S K G
CFP-Bam-Rev	5' GCC <u>GAA TTC</u> <u>GGA TCC</u> CTT GTA CAG CTC GTC CAT GC 3'
	EcoRI BamHI K Y L E D M G
Venus-Eco-For	5' GCC <u>GAA TTC</u> <u>GCG GCC GCT</u> ATG GTG AGC AAG GGC GAG GAG 3'
	EcoRI Not I M V S K G E E
Venus-Xho-Rev	5' GCC <u>GAT ATC</u> <u>CTC GAG</u> <u>TTA</u> CTT GTA CAG CTC GTC CAT GCC 3'
	EcoRV XhoI Stop K Y L E D M G

Para generar la construcción pcDNA3-CFP-Venus la secuencia de ADNc de Venus se amplificó con los oligonucleótidos Venus-Eco-For y Venus-Xho-Rev. El ADNc amplificado se digirió con EcoRI y XhoI, y se ligó en el vector pcDNA3-CFP previamente digerido con las mismas enzimas y desfosforilado. En todos los casos, la ausencia de errores en el ADNc se verificó mediante secuenciación.

4.1.2.2. Clonación de C3G en vectores para FRET

C3G completo, residuos 4-1077, se clonó en los vectores pcDNA3-CFP, pcDNA3-Venus y pcDNA3-CFP-Venus.

Para crear la construcción de C3Gh (4-1077) en pcDNA3-CFP a partir de la construcción en pCEFLAG (ver antes) se escindió con BglII y NotI el fragmento que contiene el ADNc de C3G y un codón de terminación. Dicho fragmento se clonó en pcDNA3-CFP digerido con BamHI y NotI y desfosforilado. Las cadenas sencillas producidas por BglII y BamHI son complementarias y en la ligación se destruyen ambos sitios de restricción.

Para clonar C3Gh (4-1077) en los vectores pcDNA3-Venus y pcDNA3-CFP-Venus, se amplificó el ADNc de C3G empleando los cebadores C3Gh-004F-Bgl-KZ y C3Gh-1077R-noStop (Tabla M10). El oligonucleótido directo contiene una secuencia reconocida por BglII y una secuencia Kozak. El oligonucleótido reverso contiene un sitio NotI. El producto de la PCR se digirió con BglII y NotI y se clonó en pcDNA3-Venus o pcDNA3-CFP-Venus previamente digeridos con BamHI y NotI y desfosforilados.

Los fragmentos C3G- Δ N (246-1077), C3G- Δ CAT (4-814), y C3G- Δ REM-CAT (4-669) se clonaron en el vector pcDNA3-CFP-Venus de forma análoga a como se clonó C3G (4-1077). Los cDNAs de las diferentes regiones de C3Gh se amplificaron mediante PCR empleando oligonucleótidos directos que contienen un sitio BglII y oligonucleótidos reversos que contienen un sitio NotI (Tabla M10). La región C3G- Δ N (246-1077) se amplificó con los cebadores C3Gh-246F-Bgl y C3G-1077R-noStop. La región C3G- Δ CAT (4-814) se amplificó con los oligonucleótidos C3Gh-004F-Bgl y C3G-814R-noStop. La región C3G- Δ REM-CAT (4-669) se amplificó con los cebadores C3Gh-004F-Bgl y C3G-669R-noStop. Los ADNc así amplificados se digirieron con BglII y NotI y se clonaron en los sitios BamHI y NotI del vector

pcDNA3-CFP-Venus. En todos los casos, la ausencia de errores en el ADNc se verificó mediante secuenciación.

Tabla M10. Oligonucleótidos utilizados para amplificar fragmentos de C3Gh y clonarlos en pcDNA3-Venus o pcDNA3-CFP-Venus.

OLIGONUCLEÓTIDO	SECUENCIA
C3Gh-004F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> GAC TCT CAG CGT TCT CAT C 3' EcoR I Bgl II D S Q R S H L 4
C3Gh-246F-Bgl	5' TGA <u>GAA TTC</u> <u>AGA TCT</u> AAG ACG ACT GGG ATG TCA CAG 3' EcoR I Bgl II K T T G M S Q 246
C3G-669R-noStop	5' <u>GAA TTC</u> <u>G CGG CCG C</u> GGA CAG CTC GTC CAC TTC CTC 3' EcoRI NotI S L E D V E E 669
C3G-814R-noStop	5' <u>GAA TTC</u> <u>G CGG CCG C</u> TAG CTT CTT CTG GTC CAC C 3' EcoRI NotI L K K Q D V 814
C3G-1077R-noStop	5' <u>GAA TTC</u> <u>G CGG CCG C</u> GGT CTT CTC TTC CCG GTC 3' EcoRI NotI T K E E R D 1077

4.2. Línea celular COS-1

La línea celular COS-1 de fibroblastos de riñón de mono se mantuvo en cultivo en medio DMEM (Sigma) completo, es decir, suplementado con 10 % de suero bovino fetal, 1 % glutamina, 1 % estreptomycin/penicilina (todos los suplementos de Life Technologies); esporádicamente se incluyó en el medio 0.2 % ciprofloxacino (Sigma). Los cultivos se crecieron a 37 °C con una atmósfera de 5 % de CO₂. Cuando fue necesario las células se despegaron de las placas de cultivo mediante incubación con 0.25 % Tripsina, 0,02 % EDTA (Gibco), la cual se neutralizó con medio DMEM completo, por la acción del suero bovino fetal.

Para criopreservar los cultivos, las células procedentes de una placa de 10 cm de diámetro se resuspendieron en 1.5 mL de medio DMEM completo suplementado con 10 % de DMSO en frío, se congelaron lentamente hasta -80 °C durante 12 a 24 h empleando una cubeta con isopropanol (Nalgene) y se almacenaron a temperaturas inferiores a -135 °C en un tanque de nitrógeno líquido. La descongelación se realizó de forma rápida, pasando los viales directamente del nitrógeno a un baño de 37 °C durante unos segundos,

transfiriéndose seguidamente la suspensión de células a placas con las diluciones oportunas según el tiempo estimado para su confluencia.

Para realizar transfecciones transitorias, se sembraron $\sim 4.0 \times 10^5$ células COS-1 por placa de cultivo de 6 cm de diámetro y se crecieron durante ~ 12 h hasta alcanzar una confluencia del 70-80 %. Seguidamente, se cambió el medio y las células se transfectaron con los plásmidos de interés empleando el reactivo jetPEI (Polyplus Transfection) siguiendo las instrucciones del fabricante. Para cada transfección se emplearon 5 μg totales de ADN, es decir, 5 μg de un único plásmido en el caso de transfecciones sencillas, y 2.5 μg de cada plásmido para transfecciones dobles.

4.3. Ensayos de co-inmunoprecipitación

Para el análisis de interacciones mediante ensayos de co-inmunoprecipitación (coIP) se realizaron dobles transfecciones de células COS-1. Se co-transfectó la construcción N-C3G (4-245) en pCEFLAG, o el vector pCEFLAG vacío, con construcciones de diferentes fragmentos de C3G en el vector pCEFLHA.

Pasadas de 36 a 48 horas desde la transfección, las placas de cultivo se colocaron sobre hielo y se lavaron dos veces con 1 mL de PBS frío. Las células se recogieron mediante raspado en 200 μL de tampón de lisis: 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5 % Triton X-100, 1 mM Na_3VO_4 , 25 mM NaF, 1 mM PMSF y 1x cóctel inhibidor de proteasas (Complete, Roche). La lisis se completó mediante agitación vigorosa en vórtex, 3 pulsos de 30 s separados por pausas de 10 s. Las muestras se clarificaron mediante centrifugación durante 15 min, a 13200 rpm a 4 $^\circ\text{C}$, tras la cual el sobrenadante se transfirió a un nuevo tubo. Se tomó una muestra del sobrenadante (50 μL), correspondiente al lisado celular total, que se mezcló con un volumen igual de SB 2x para su posterior análisis.

Para la IP se utilizó gel de afinidad anti-FLAG M2 (Sigma), que consiste en el anticuerpo monoclonal anti-FLAG M2 unido covalentemente a agarosa. Se empleó una proporción de 20 μL de la suspensión comercial de la resina (en 10 mM NaPi, 150 mM NaCl, pH 7.4, 50 % glicerol) por cada muestra de sobrenadante. La resina se preparó mediante 4 lavados con tampón de lisis, entre los que la resina se sedimentó mediante centrifugación suave (1000 g , 30 s, 4 $^\circ\text{C}$). Tras el último

lavado, el gel se resuspendió en tampón de lisis en una proporción de 100 μ L para cada reacción. El sobrenadante del lisado celular se incubó con 100 μ L de la suspensión de la resina anti-FLAG durante 1 h a 4 °C con agitación orbital suave. Pasado este tiempo, se decantó el sobrenadante y se lavó el gel 3 veces como se ha descrito antes. Tras el último lavado se aspiró el resto de tampón, se añadieron 20 μ L de SB 2x a la resina y se incubaron las muestras 5 min a 95 °C para extraer las proteínas unidas.

Las muestras correspondientes a los lisados totales y las proteínas inmunoprecipitadas se analizaron mediante SDS-PAGE e inmunodetección. La expresión de las proteínas con etiquetas FLAG y HA se analizó en las muestras de lisados totales empleando anticuerpos monoclonales HA (11) y anti-FLAG M2 (Tabla M6). Además se analizaron los niveles de actina en los lisados empleando un anticuerpo monoclonal frente a β -actina. La presencia de proteínas con la etiqueta HA en las muestras inmunoprecipitadas se realizó empleando un anticuerpo policlonal anti-HA (Tabla M6) con el fin de evitar la interferencia con el anticuerpo monoclonal M2 procedente del gel de inmunoafinidad.

4.4. Ensayos de FRET en células COS-1

La transferencia de energía de resonancia Förster (FRET) es un proceso en el que cuando un fluoróforo donador es excitado disipa el exceso de energía a través de un mecanismo no radiativo de acoplamiento dipolo-dipolo hasta un aceptor cercano. Esto resulta en una disminución de la fluorescencia del donador y un incremento de la del aceptor (Förster, 1948; Stryer, 1978; Van Der Meer, 1994). La eficiencia de la transferencia de energía es inversamente proporcional a la sexta potencia de la distancia entre los dos fluoróforos (Förster, 1948):

$$E = \frac{R_0^6}{R_0^6 + r^6}$$

donde r es la distancia entre los dos fluoróforos y R_0 (distancia de Förster) es la distancia a la cual el 50 % de la transferencia de energía tiene lugar (típicamente entre los 20 y los 60 Å). R_0 depende del solapamiento de los espectros entre el donador y el aceptor, el rendimiento cuántico del donador y la orientación relativa de los fluoróforos. Esta técnica proporciona información sobre distancias inter- o

intra-moleculares en un rango de 10 a 100 Å, lo que permite su aplicación al análisis de estructura de proteínas, como cambios conformacionales (Figura M2), y de interacciones proteína-proteína.

En los ensayos de FRET se ha utilizado como par donador-aceptor las proteínas fluorescentes CFP (*cian*) y Venus (*yellow*) (una variante de la proteína fluorescente amarilla) (Heim and Tsien, 1996; Nagai et al., 2002). CFP muestra un máximo de excitación a ~ 435 nm y un pico de emisión a ~ 475 nm; mientras que Venus tiene un máximo de absorción a ~ 515 nm y un máximo de emisión a ~ 527 nm.

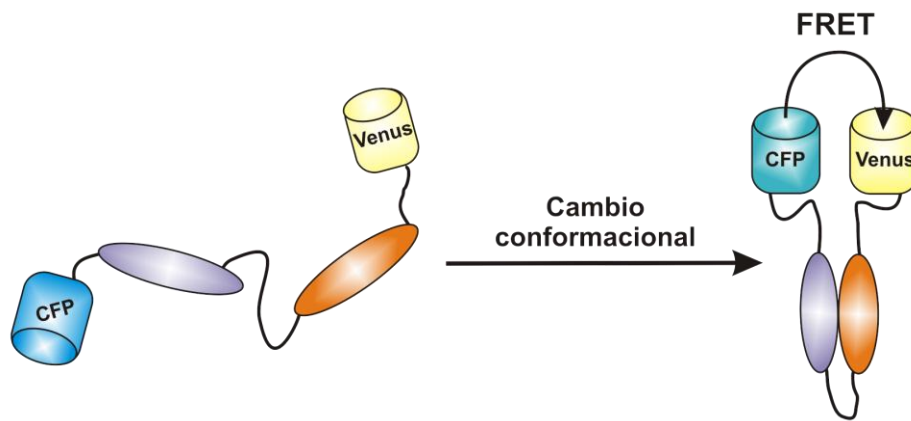


Figura M2. Esquema de un biosensor de FRET intramolecular para el análisis de cambios conformacionales. Se han creado proteínas de fusión con CFP y Venus en los extremos opuestos de la cadena polipeptídica. La transferencia de energía entre donador y aceptor permite monitorizar cambios conformacionales que conlleven un cambio en la distancia entre los fluoróforos. Adaptado de (Zhang et al., 2002).

Para los ensayos de fluorescencia se transfectaron células COS-1 con construcciones de C3G que incorporaban las proteínas fluorescentes CFP y Venus en los extremos amino y carboxilo respectivamente, por lo que la estequiometría del donador y del aceptor es fija. También se realizaron ensayos cotransfectando dos construcciones de C3G completo (4-1077), una de ellas con CFP en el extremo amino y la otra con Venus en el extremo carboxilo. Como blanco para sustraer la autofluorescencia celular se usaron cultivos de COS-1 no transfectados.

Los ensayos de FRET se realizaron tanto con células vivas en suspensión como con lisados celulares. Para los experimentos con células en suspensión, las células de una placa de 6 cm de diámetro se despegaron por tripsinización, y la acción de la tripsina se paró añadiendo medio DMEM completo. Seguidamente las

células se recogieron mediante centrifugación a 1280 rpm, 5 min a temperatura ambiente y se resuspendieron en 500 μ L de PBS precalentado a 37 °C. Las suspensiones de células se mantuvieron a 37 °C en un baño hasta su medida. Las medidas de fluorescencia se registraron en un espectrofluorímetro FluoroMax-3 (HORIBA-Jobin-Yvon) equipado con un soporte para muestras termostaticado. Se empleó una cubeta de cuarzo de 5 mm de paso óptico tanto en la dirección de excitación como de emisión (Hellma), en la que se incorporó un imán para mantener las células en suspensión mediante agitación. Las medidas de fluorescencia se realizaron a 37 °C. Se utilizaron rendijas de 2 nm de ancho de banda en excitación y 5 nm en emisión; el tiempo de integración fue de 0.5 s y se registraron datos espaciados 1 nm.

Se midieron espectros utilizando una longitud de onda de excitación (λ_{ex}) de 435 nm, que corresponde al máximo de excitación de la CFP, y un rango de longitud de onda de emisión (λ_{em}) de 450 a 600 nm. El máximo de emisión de la CFP (donador) se localiza a 475 nm. La transferencia de energía se detectó como la aparición de un nuevo pico en el espectro de emisión a 527 nm, correspondiente al máximo de emisión de Venus (aceptor). Se restó la contribución de la autofluorescencia de las células sustrayendo el espectro de una muestra con un número similar de células que no expresan proteínas con CFP o Venus. Por último los espectros de emisión se normalizaron por la emisión a 475 nm. La señal de FRET se cuantificó como la relación entre la fluorescencia del aceptor y del donador calculando el cociente de la intensidad a λ_{em} 527 nm dividido entre la intensidad a λ_{em} 475 nm. Por otro lado se midieron espectros en los que se excitó directamente Venus empleando λ_{ex} de 490 nm y se registraron datos a λ_{em} de 505 a 600 nm.

Para las medidas de fluorescencia de lisados celulares, las células crecidas en una placa de 6 cm de diámetro se resuspendieron en 450-500 μ L de tampón de lisis tal como se ha descrito para los ensayos de coIP. Las medidas de fluorescencia de lisados celulares se realizaron de forma similar a las de células en suspensión con pequeñas diferencias: la temperatura de la muestra fue de 25 °C, se emplearon rendijas de 3 nm tanto en excitación como en emisión y se utilizó un tiempo de integración de 0,5 s. Se restó la contribución de lisados de células COS-1 sin

transfectar. En general la intensidad de fluorescencia en las muestras de lisados celulares fue muy superior a la observada con células en suspensión.

Después de realizar las medidas de fluorescencia, las muestras se recuperaron y se analizó la presencia de las proteínas de interés mediante SDS-PAGE e inmunodetección con un anticuerpo anti-GFP que reconoce globalmente las proteínas fluorescentes CFP y Venus, además de anticuerpos anti-C3G (Tabla M6). En el caso de las muestras medidas como células en suspensión, previo a la inmunodetección se lisaron las células.

5. Producción heteróloga de proteínas en bacterias

Las proteínas recombinantes de C3G empleadas para su estudio *in vitro* se produjeron de forma heteróloga en *E. coli* a partir de la secuencia de ADNc optimizada.

5.1. Clonación en vectores de expresión en bacterias

Para la expresión de proteínas en *E. coli* se utilizó el vector pETEV15b (Alonso-García et al., 2009), vector derivado del pET15b (Novagen). Este vector contiene (Figura M3):

1. Un gen de resistencia a ampicilina.
2. El promotor T7, específico de la T7 ARN polimerasa, bajo el control del operón *lac* inducible con isopropil- β -D-tiogalactopiranosido (IPTG).
3. Una secuencia que codifica para una cola de 8 His seguida por una secuencia reconocida por la proteasa del virus del grabado del tabaco (TEV) en la región N-terminal (secuencia de aminoácidos MGSSHHHHHHSSGENLYFQ *GSHM, el asterisco indica el sitio de corte de la proteasa). Cuando el producto expresado se digiere con la correspondiente proteasa, el polipéptido GSHM permanece en el extremo N-terminal de la proteína.
4. Una secuencia con sitios de corte único para las enzimas de restricción NdeI, XhoI y BamHI, utilizados para insertar los ADNc de interés.

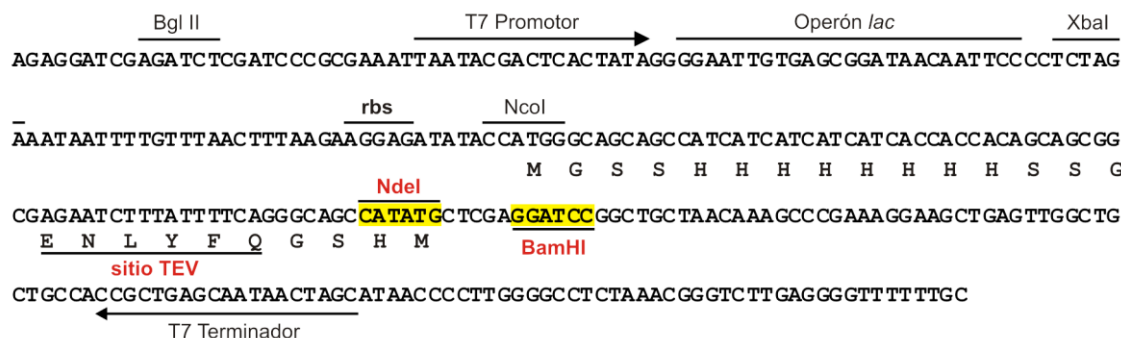


Figura M3. Secuencia de nucleótidos del vector pETEV15b. Se representan los sitios de reconocimiento para las enzimas de restricción NdeI y BamHI, el sitio de reconocimiento para la proteasa TEV y las secuencias del vector que hibridan con los oligonucleótidos T7 Promotor y T7 Terminador.

Los ADNc correspondientes a la región N-C3G (4-210), (65-210), (4-245) y (65-245) de C3G humano se amplificaron mediante PCR con la polimerasa Pfu (Stratagene, Agilent) usando como molde pUC57-C3Gopt. La secuencia C3Gopt carece de sitios de reconocimiento para BamHI y NdeI. Los oligonucleótidos directos contienen secuencias de reconocimiento para las enzimas de restricción EcoRI y NdeI, mientras que los reversos contienen un codón de terminación y sitios para las enzimas de restricción BamHI y EcoRI (Tabla M11). Los ADNc de C3Gopt amplificados se digirieron con BamHI y NdeI y se clonaron en el vector pETEV15b previamente digerido con esas mismas enzimas y desfosforilado. Las regiones de los plásmidos correspondientes a C3Gopt se secuenciaron para comprobar la ausencia de errores.

Tabla M11. Oligonucleótidos utilizados para amplificar fragmentos de C3Gopt.

OLIGONUCLEÓTIDO	SECUENCIA
C3G-U004	5' TGA <u>GAA TTC</u> <u>CAT ATG</u> GAT AGC CAG CGC AGT CAT C 3' EcoR I Nde I D S Q R S H L 4
C3G-U065	5' TGA <u>GAA TTC</u> <u>CAT ATG</u> CTG GAT CTG GAG CAG CAG 3' EcoR I Nde I L D L E Q Q 65
C3G-L210	5' GCC <u>GAA TTC</u> <u>GGA TCC</u> <u>CTA</u> CTG TTT TTC AAT TGT CAG ACG 3' EcoR I BamH I STOP Q K E I T L R 210
C3G-L245	5' GCC <u>GAA TTC</u> <u>GGA TCC</u> <u>CTA</u> ATT CAG AAT TTC CAC TTC GCG 3' EcoR I BamH I STOP N L I E V E R 245

5.2. Expresión y purificación de proteínas

Las construcciones de C3G en el vector pETEV15b se transformaron en células de la cepa BL21(DE3)T1 de *E. coli*. Como paso previo a la producción a gran escala de los fragmentos de C3G, se analizó su expresión con el fin de determinar si las proteínas se producían de manera soluble. Para ello, 25 mL de medio Terrific Broth (TB) (Sambrook et al., 1989) suplementado con 100 µg/mL de ampicilina se inocularon con una colonia de BL21(DE3)T1 que contenía el plásmido de interés. El cultivo se creció a 37 °C en agitación, hasta alcanzar una DO₆₀₀ de 0.4-0.6. En este punto, se indujo la expresión de la proteína de interés mediante la adición de 0.2 mM IPTG; la expresión se mantuvo durante 3 h a 37 °C; o alternativamente, alcanzada la DO, el cultivo se enfrió a 15 °C y se indujo la expresión a esta temperatura durante 12 h. Seguidamente, las células se recogieron por centrifugación a 1800 *g* (3000 rpm, rotor Eppendorf A-4-62) durante 15 min a 4 °C y se resuspendieron en 1 mL de 20 mM Tris-HCl (pH 7.9), 500 mM NaCl, 5 mM imidazol (solución A) suplementado con 0.1 % (v/v) Triton-X100. Las células se lisaron mediante sonicación (3 pulsos de 15 s). El homogeneizado resultante se centrifugó a 16000 *g* a 4 °C durante 30 min, y el sobrenadante se decantó. El precipitado se resuspendió en 1 mL de la solución A suplementada con 6 M urea. La presencia de la proteína de interés en la fracción soluble (sobrenadante) o insoluble (precipitado) se analizó mediante SDS-PAGE.

La expresión de proteínas a gran escala se realizó de forma similar, pero empleando un volumen de 3 L de medio de cultivo. Tras recoger las bacterias por centrifugación, éstas se resuspendieron en 50 mL de solución A con 0.1 % (v/v) Triton-X100 por cada litro de cultivo.

Los fragmentos que corresponden a la región N-terminal de C3G (4-245), (65-245), y el mutante C154A de la versión corta de N-C3G (65-245) aparecen en la fracción soluble y se purificaron en condiciones nativas mediante cromatografía de afinidad IMAC (Block et al., 2009) utilizando una columna quelante de Ni²⁺ His-trap HP de 5 ml (GE-Healthcare) y un equipo de cromatografía ÄKTA-Prime (GE-Healthcare). Antes de cargar en la columna los sobrenadantes provenientes de la lisis bacteriana, se redujo su viscosidad mediante sonicación. Las proteínas se eluyeron con un gradiente de 5 mM a 500 mM de imidazol en 20 mM Tris-HCl (pH 7.9), 500 mM NaCl. Las fracciones obtenidas se analizaron por SDS-PAGE para ver

el grado de pureza de las mismas. Las fracciones positivas se juntaron y se dializaron frente a 20 mM Tris-HCl (pH 7.5), 150 mM NaCl a 4 °C. La cola de octa-His se eliminó mediante digestión con la proteasa TEV a 30 °C durante unas ocho horas, empleando una forma recombinante de TEV que contiene una cola de poli-His. La muestra se dializó frente a 20 mM Tris-HCl (pH 7.9), 150 mM NaCl a 4 °C y la proteína digerida se separó de la cola de poli-His y de la TEV-His mediante un segundo paso de cromatografía de afinidad. Finalmente la proteína se dializó frente a la solución deseada y se concentró en células de ultrafiltración Amicon (Millipore). Cuando las muestras no se utilizaron inmediatamente, se dividieron en alícuotas, se congelaron mediante inmersión en nitrógeno líquido y se almacenaron a -80 °C.

En la purificación del fragmento N-C3G-C154A (65-245) para el análisis mediante SAXS, se incluyó un paso de purificación adicional. Tras la segunda cromatografía de afinidad, la muestra se sometió a una cromatografía de exclusión molecular en una columna Sephacryl S300 HiPrep 26/60 (26 x 600 mm) (GE-Healthcare). La columna se equilibró en 20 mM NaPi (pH 7.5), 150 mM NaCl, 5% (v/v) glicerol, 3mM DTT, siendo el flujo 1.2 mL/min. Antes de ser aplicada a la columna, la muestra se concentró hasta un volumen inferior a 5 mL y se centrifugó a 4 °C durante 30 min a 16000 *g*. Se recogieron fracciones de 4 mL que se analizaron mediante SDS-PAGE. Las fracciones de interés se juntaron, se concentraron mediante ultrafiltración y se almacenaron como se ha descrito.

6. Métodos biofísicos y bioquímicos de análisis de proteínas

6.1. Ensayos de cristalización

Los ensayos de cristalización se llevaron a cabo por el método de difusión de vapor en gota sentada, utilizando placas de 96 pocillos de MRC (Swissci). Se empleó una colección de soluciones de cristalización comerciales: Wizard I y II (Emerald BioSystems, WA), Cryo I y II (Emerald BioSystems, WA), Salt Rx (Hampton Research, CA), Hampton Crystal Screen (Hampton Research, CA) y PACT (Newman et al., 2005), que incluyen un total de 432 soluciones de cristalización. Los ensayos se realizaron a temperatura ambiente y a 4 °C. Se usaron gotas compuestas por 1 µL de solución de proteína y 1 µL de solución de cristalización, y

se equilibraron frente a 80 μ L de la solución de cristalización que se encontraba en el reservorio.

6.2. Dicroísmo circular en el UV lejano

El dicroísmo circular (CD) en el UV lejano es una técnica espectroscópica sensible a la estructura secundaria de las proteínas (Woody, 1995).

Para los experimentos de CD, las proteínas se equilibraron en 20 mM NaPi (pH 7.5), 150 mM NaCl mediante diálisis. Los espectros de CD se midieron en un dicrógrafo J-720 (Jasco) en el Centro de Investigaciones Biológicas (CSIC, Madrid), empleando cubetas de cuarzo con un paso óptico de 1 mm (Hellma). Se tomaron datos entre 185 y 260 nm a 25 °C. Se empleó un ancho de banda de 1 nm, y una velocidad de barrido de 20 nm/min. Se midieron 5 acumulaciones de cada espectro y se calculó el espectro promedio. Tras restar la contribución del tampón, los valores de elipticidad se convirtieron en elipticidad molar media por residuo según la ecuación:

$$[\theta]_{\lambda} = \frac{\theta_{obs} \cdot MRW}{10 \cdot d \cdot c}$$

siendo $[\theta]_{\lambda}$ el valor de la elipticidad molar a la longitud de onda λ ; θ_{obs} , el valor de la elipticidad observado; MRW , la masa por residuo medio; d , el paso óptico; y c , la concentración de proteína.

La estimación del contenido de estructura secundaria a partir de los espectros de CD se realizó empleando el servidor K2D3 (Andrade et al., 1993; Louis-Jeune et al., 2011; Perez-Iratxeta and Andrade-Navarro, 2008); y los programas Selcon3 (Sreerama et al., 1999; Sreerama and Woody, 1993), CONTIN (Provencher and Glockner, 1981; van Stokkum et al., 1990) y CDSSTR (Compton and Johnson, 1986; Manavalan and Johnson, 1987; Sreerama and Woody, 2000) del servidor Dichroweb (Lobley et al., 2002; Whitmore and Wallace, 2004, 2008).

6.3. Proteólisis limitada

Las proteínas correspondientes a la región N-C3G (4-245) y (65-245) se sometieron a proteólisis controlada. La digestión se realizó a temperatura ambiente y a una concentración de proteína de ~ 1 mg/mL en presencia de 0.5 % (p/p) de tres proteasas con especificidades químicas diferentes: tripsina, quimotripsina y endoproteasa V8 (todas las diluciones realizadas, tanto de las proteínas como de las enzimas, se hicieron en el tampón 20 mM Tris-HCl (pH 7.5), 150 mM NaCl). Se tomó una muestra antes de añadir la proteasa ($t = 0$), y a varios tiempos hasta los 30 min; la reacción se paró en hielo seco. Las muestras se analizaron mediante SDS-PAGE. Alternativamente los productos de la digestión se analizaron mediante espectrometría de masas (MS) IP-MALDI-TOF en la Unidad de Proteómica del Centro de Investigación del Cáncer. Las muestras para MS se prepararon mediante extracción con metanol y cloroformo (Wessel and Flugge, 1984): se añadieron a cada muestra primero 4 volúmenes de metanol, luego 2 volúmenes de cloroformo y por último 3 volúmenes de agua; tras cada adición la muestra se agitó vigorosamente con un vórtex. Seguidamente se centrifugó a 13200 rpm durante 5 min, de modo que las proteínas formaron un disco en la interfase. Se retiró la fase superior, se añadieron 3 volúmenes de metanol, se mezcló bien y se centrifugó a 13200 rpm durante 5 min. Se eliminó el sobrenadante y se dejó secar el precipitado que contiene las proteínas. Los resultados obtenidos en MS se analizaron con el servidor Protein Prospector (<http://prospector.ucsf.edu>), en el apartado de MS-Digest (UCSF) para la elaboración de los mapas peptídicos.

6.4. Cromatografía de exclusión molecular analítica

Para el estudio por medio de cromatografía de exclusión molecular las proteínas N-C3G (65-245) silvestre y N-C3G-C154A (65-245) se dializaron frente al tampón 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. Los experimentos se realizaron a temperatura ambiente empleando un sistema ÄKTA-Purifier (GE Healthcare) y una columna Superdex 200-HR 10/300 (GE-Healthcare) de 10 x 300 mm equilibrada

en 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. El volumen de inyección fue de 100 μ L y se utilizó un flujo de 0.5 mL/min.

La masa molecular aparente (MMap) y el radio de Stokes (Rh, radio hidrodinámico) se estimó a partir del volumen de elución (Ve) mediante las expresiones:

$$Ve = A - B \log_{10} MMap$$

$$Ve = A' - B' \log_{10} Rh$$

Para la determinación de los parámetros A, B, A' y B' la columna se calibró empleando un juego con las siguientes proteínas patrón (BioRad):

Proteína	MM (kDa)	Ro (Å)
Tiroglobulina	670	85.0
γ -globulina	158	48.0
ovoalbúmina	44	30.5
mioglobina	17	19.0

6.5. Ultracentrifugación analítica

Las técnicas de centrifugación analítica permiten obtener información sobre la masa molecular, el estado de agregación y las propiedades hidrodinámicas de una macromolécula biológica en solución (Lebowitz et al., 2002; Waxman et al., 1994). Además, posibilitan la detección y cuantificación de las interacciones macromoleculares reversibles, que incluyen homo- y hetero-asociaciones proteína-proteína o proteína-ligando (Rivas et al., 1999).

Se aplicaron dos métodos de ultracentrifugación analítica: velocidad de sedimentación y equilibrio de sedimentación. Estos experimentos se realizaron en Servicio de Ultracentrifugación Analítica y Dispersión de Luz del Centro de Investigaciones Biológicas (CSIC, Madrid) por el Dr. Carlos Alfonso, utilizando una ultracentrífuga analítica Beckman Optima XLI (Beckman Coulter Instrument, Inc.).

6.5.1. Velocidad de sedimentación

Los experimentos de velocidad de sedimentación se realizaron empleando celdas de doble sector de 12 mm de paso óptico en un rotor AN50Ti. 400 μ L de soluciones de proteína N-C3G (65-245) a 0.25, 0.50, 1.00 y 2.00 mg/mL, en 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, se centrifugaron a 45000 rpm y 20 °C. La distribución radial de la proteína se midió por absorción a 295 nm cada 5 min. A partir de los perfiles de sedimentación obtenidos a intervalos de tiempo regulares se obtuvieron distribuciones de los coeficientes de sedimentación aplicando el método $c(s)$ (Schuck et al., 2002) mediante el ajuste directo de las soluciones a la ecuación de Lamm con el programa SEDFIT (Schuck, 2010; Schuck and Rossmanith, 2000).

Los valores de coeficiente de sedimentación, s , se corrigieron a condiciones estándar $s_{20,w}$ (agua, 20 °C) empleando la ecuación:

$$s_{20,w} = s \frac{(1 - \bar{v}\rho)_{20,w} \eta_{T,b}}{(1 - \bar{v}\rho)_{T,b} \eta_{20,w}}$$

donde los subíndices se refieren a la temperatura experimental T, o 20 °C, y las condiciones de la solución, agua (w) o solución tampón (b). Para el agua se emplearon valores de viscosidad, $\eta_{20,w}$, de 0.01002 P y densidad, $\rho_{20,w}$, de 0.9982 g/mL. Para el tampón 20 mM NaPi (pH 7.5), 150 mM NaCl se emplearon valores de viscosidad, $\eta_{20,b}$, de 0.01023 P, y densidad, $\rho_{20,w}$, de 1.0062 g/mL. El volumen específico parcial a 20 °C, \bar{v} , de N-C3G 65-245 (0.742 cm³/g) se calculó a partir de la composición de aminoácidos empleando el programa SEDNTERP (<http://bitcwiki.sr.unh.edu/>).

6.5.2. Equilibrio de sedimentación

Mediante esta técnica podemos obtener información termodinámica acerca de la masa molecular promedio, estequiometría y afinidad de unión de las proteínas estudiadas. Dada una disolución uniforme, que se somete a una fuerza centrífuga moderada se establece un equilibrio entre el proceso de sedimentación de las macromoléculas hacia el fondo de la celda, y el proceso de difusión que se opone al anterior. El equilibrio se alcanza, cuando, transcurrido un tiempo, los dos

procesos se igualan y como resultado el transporte neto se anula ($J_i = 0$); llegado este punto, las macromoléculas estarán distribuidas formando un gradiente de concentración creciente hacia el fondo de la celda que no variará con el tiempo.

Para una situación de equilibrio $J_i = 0$, se obtiene la ecuación:

$$\frac{d \ln w_i}{dr^2} = \frac{M_i^* w^2}{2RT}$$

donde w es la concentración y r es el radio de la celda de medida. Si esta expresión es integrada respecto a r^2 , obtendremos la ecuación de equilibrio para la sedimentación de un soluto ideal.

$$w_i(r) = w_i(r_0) \left(\frac{M_i^* w^2}{2RT} (r^2 - r_0^2) \right)$$

donde r_0 indica la posición de referencia donde empieza el gradiente. El gradiente en el equilibrio, se caracteriza fundamentalmente porque es invariante con el tiempo, es independiente de las propiedades hidrodinámicas del soluto, y depende únicamente de la masa molecular de flotación (M_i^*). En los experimentos de equilibrio de sedimentación, el tiempo necesario para alcanzar el equilibrio depende de la velocidad utilizada, del tamaño de la macromolécula y de la longitud de la columna que es la distancia desde el menisco hasta el fondo de la celda.

En los experimentos de equilibrio de sedimentación de C3G la muestra se centrifugó a 10000 rpm a 20 °C. Se midió la absorbancia de la muestra a 250, 280 o 295 nm en función de la distancia radial a las 12, 16 y 18 h, obteniéndose los mismos resultados, lo que confirmaba que la muestra estaba en equilibrio. Seguidamente se repitieron las medidas a 12000 rpm y 14000 rpm. Por último, la línea de base correspondiente a la aportación del tampón a la señal, se determinó midiendo la absorbancia cerca del menisco tras centrifugar la muestra durante 8 h a 35000 rpm.

Los resultados experimentales se analizaron con el programa HeteroAnalysis (Cole, 2004), con el que se obtuvo la masa molecular aparente de flotación. El volumen específico parcial de las proteínas y la densidad de la disolución de N-C3G se estimaron con el programa SEDNTERP (Laue et al., 1992) a partir de la composición de aminoácidos.

6.6. Anisotropía de fluorescencia extrínseca

La anisotropía de fluorescencia es la polarización producida por una muestra que emite luz después de ser sometida a una excitación con luz polarizada. La causa predominante de la despolarización de fluorescencia es la difusión rotacional de los fluoróforos que está relacionada con el coeficiente de difusión rotacional, que a su vez depende del tamaño y de la forma de las especies. Por lo tanto la anisotropía es sensible a cambios en el estado de asociación de las proteínas. Las muestras se excitaron con luz linealmente polarizada vertical y se midieron las componentes de la intensidad de fluorescencia polarizada en las direcciones paralela (I_{VV}) y perpendicular (I_{VH}) al plano de polarización de la luz de excitación.

En condiciones ideales, la anisotropía de fluorescencia en estado estacionario viene dada por la siguiente expresión (Jablonski, 1957):

$$r = \frac{I_{VV} - I_{VH}}{I_{VV} + 2I_{VH}}$$

donde I_{VV} y I_{VH} son las componentes polarizadas vertical y horizontalmente de la intensidad de fluorescencia emitida cuando la excitación está polarizada verticalmente. Sin embargo, en la expresión anterior es necesario introducir una corrección debida a la diferente sensibilidad del canal de emisión a la luz polarizada horizontal y verticalmente. Para esto se emplea el factor G , que se puede definir como la relación de las sensibilidades del sistema de detección para la luz polarizada vertical y horizontalmente. El factor G se determinó empleando luz de excitación polarizada horizontalmente:

$$G = \frac{S_V}{S_H} = \frac{S_V I_{VH}}{S_H I_{VH}} = \frac{I_{VH}}{I_{HH}}$$

donde S_V y S_H son las sensibilidades del canal de emisión para la luz polarizada vertical y horizontal respectivamente; e I_{VH} e I_{HH} son las intensidades observadas con el polarizador de emisión vertical y horizontal, respectivamente, cuando la excitación está polarizada horizontalmente.

Por lo tanto la anisotropía se calcula como cociente de intensidades con la corrección por la sensibilidad del sistema según la siguiente ecuación:

$$r = \frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}}$$

6.6.1. Marcaje de N-C3G con *Oregon green 488 I-acetamida*

N-C3G (65-245) se marcó de forma covalente en la Cys 154 (que es la única Cys presente en la región N-terminal de C3G) con la sonda fluorescente Oregon Green 488 iodoacetamida (OG-IA) (Invitrogen) que reacciona con grupos tiol a través del grupo I-acetamida. A una solución $\sim 250 - 300 \mu\text{M}$ de la proteína en 20 mM Tris-HCl (pH 8.0), 150 mM NaCl se le añadió un exceso de 10x de OG-IA sobre la concentración molar de proteína. Previamente la sonda se disolvió en DMSO de modo que la concentración final de DMSO añadida a la muestra de proteína fue menor del 10 %. La reacción de marcaje se prolongó durante 12 horas a 4 °C, con agitación y en la oscuridad. La reacción se paró añadiendo DTT en un exceso de 10 veces superior a la concentración molar de fluoróforo, y se dejó incubar durante una hora a 4 °C. Posteriormente la proteína marcada se separó del exceso de reactivo mediante cromatografía de exclusión molecular en una columna de Sephadex G25 (10 x 300 mm) equilibrada en 20 mM NaPi (pH 7.5), 150 mM NaCl. Las fracciones positivas se juntaron y las concentraciones de proteína y de fluoróforo se determinaron espectrofotométricamente utilizando los coeficientes de extinción molar del OG $\epsilon_{496} = 71000 \text{ M}^{-1} \text{ cm}^{-1}$ y de la proteína $\epsilon_{280} = 15470 \text{ M}^{-1} \text{ cm}^{-1}$. Para la estimación de la concentración de proteína se tuvo en cuenta la absorbancia del fluoróforo a 280 nm, siendo $\epsilon_{280} = 20300 \text{ M}^{-1} \text{ cm}^{-1}$. El grado de marcaje, definido como el número promedio de moléculas de fluoróforo unidas a cada molécula de proteína, se determinó como el cociente entre la concentración de OG y la concentración de proteína resultando ser de 1.2 ± 0.2 moléculas de fluoróforo por molécula de N-C3G. Las muestras de proteína marcada se congelaron en nitrógeno líquido y se guardaron hasta su utilización a -80 °C.

6.6.2. Ensayo de la homo-asociación de N-C3G basado en fluorescencia

Las medidas de anisotropía de fluorescencia de OG en estado estacionario se realizaron a 25 °C en un espectrofluorímetro FluoroMax-3 equipado con polarizadores (HORIBA-Jobin-Yvon). Se utilizó una longitud de onda de excitación de 496 nm (2 nm de ancho de banda) y se midió la emisión a 520 nm (10 nm de ancho de banda); se empleó un tiempo de integración de 1 s y se realizaron 5

repeticiones por medida. Se utilizó una cubeta de cuarzo de 3 x 3 mm (3 mm de paso óptico en las direcciones de excitación y de emisión) (Hellma). Se midió la anisotropía de fluorescencia de una solución 20 nM de N-C3G (65-245) marcado con OG (N-C3G-OG) en 20 mM Tris-HCl (pH 7.5), 150 mM NaCl suplementado con 1 mg/mL de BSA, en presencia de diversas concentraciones de N-C3G sin marcar. Para preparar las muestras se partió de una solución 20 nM N-C3G-OG y 1 mM N-C3G y se realizaron diluciones seriadas en 20 nM N-C3G-OG, de esta forma la concentración de proteína marcada fue constante en todas las muestras.

A partir de los valores de anisotropía se estimó la constante aparente de dimerización (K_{ap}) ajustando la siguiente ecuación que corresponde a un modelo de equilibrio monómero-dímero (Zorrilla et al., 2004):

$$r = r_d + \left[(r_d - r_m) \left(\frac{-1 + \sqrt{1 + 8K_{ap}[C3G]_T}}{4K_{ap}[C3G]_T} \right) \right]$$

donde, r_m y r_d son la anisotropía del estado monomérico y dimérico respectivamente, y $[C3G]_T$ es la concentración total de proteína. Los valores de r_m , r_d y K_{ap} se estimaron ajustando este modelo a los datos empleando el programa SIGMAPLOT (Systat Software).

6.7. Dispersión de rayos-X a bajo ángulo (SAXS)

La dispersión de rayos-X a bajo ángulo o SAXS (*Small angle X-ray scattering*) es una técnica que permite analizar la estructura de macromoléculas en solución. En un experimento de SAXS se hace incidir un haz de rayos-X colimado y monocromático sobre una disolución de proteína u otra macromolécula y se mide la radiación dispersada a ángulos muy bajos (típicamente entre 0.1 ° y 10 °). SAXS es una técnica de contraste, por lo cual para conocer la dispersión producida por las macromoléculas es necesario sustraer la señal de dispersión debida al solvente. A partir del análisis de los perfiles de dispersión es posible estimar diversos parámetros estructurales de las proteínas objeto de estudio, como el radio de giro, R_g ; la función de probabilidad de distribución de distancias, $P(r)$; la distancia máxima intramolecular, D_{max} o el volumen de la partícula. Asimismo, a partir del

perfil de dispersión es posible calcular modelos a baja resolución de la proteína utilizando métodos de modelado *ab initio*.

6.7.1. Toma de datos de SAXS

Las medidas de SAXS de N-C3G (65-245) se realizaron en la estación de dispersión coherente de rayos-X (cSAXS) de la fuente de radiación sincrotrón Swiss Light Source (SLS) del Paul Scherrer Institut (PSI, Villigen, Suiza). Se utilizó un detector Pilatus 2_M (Dectris Ltd.), que se situó a una distancia de 2.15 m de la muestra, de tal forma que se cubrió un rango de vectores de dispersión (s) desde aproximadamente 0.015 hasta 0.5 Å⁻¹. Siendo el vector de dispersión:

$$s = \frac{4\pi \sin\theta}{\lambda}$$

donde 2θ es el ángulo de dispersión entre el rayo incidente y el detector, y λ es la longitud de onda de los rayos-X que fue de 1 Å (correspondiente a una energía de 12 KeV).

Las muestras de proteína a diferentes concentraciones se midieron empleando capilares de borosilicato de 1 mm de diámetro (Hilgenberg GMBH) situados en un soporte de cobre termostaticado a 10 ± 1 °C. De la misma manera se midió el tampón en el que estaba disuelta la proteína: 20 mM NaPi (pH 7.5), 150 mM NaCl, 5 % (v/v) glicerol, 3 mM DTT. Tanto el glicerol como el DTT minimizan los daños de la muestra por radiación, removiendo las especies reactivas del oxígeno generadas en la disolución acuosa por los rayos-X (Kuwamoto et al., 2004). Con el fin de reducir los daños por radiación se tomaron datos en 10 posiciones de cada capilar (espaciados 0.5 mm entre sí) realizando 60 exposiciones de 0.5 s, con un tiempo total de 30 s. Los perfiles de dispersión se promediaron radialmente, y tras comprobar que no existía daño por radiación, los diversos perfiles tomados a lo largo del tiempo para una misma muestra se promediaron con objeto de maximizar la relación señal/ruido.

6.7.2. Procesado y análisis de datos de SAXS

El procesado de los datos de dispersión se llevó a cabo utilizando el conjunto de programas del paquete ATSAS 2.3 (Konarev et al., 2006). Las curvas de intensidad, $I(s)$, en función del ángulo de dispersión, s , obtenidas para cada una de las concentraciones medidas, así como el tampón, se procesaron con el programa PRIMUS (Konarev et al., 2003). En primer lugar, se restó la contribución del tampón a las muestras de proteína. Se normalizó por la concentración de proteína, se superpusieron los perfiles de las diferentes concentraciones de la muestra y se extrapolaron los datos a dilución infinita.

El radio de giro, R_g , definido como la distancia cuadrática media de los átomos de la proteína al centro de masas de la proteína, se calculó con el programa PRIMUS mediante la aproximación de Guinier (Rice, 1956) según la cual a bajo ángulo la dispersión se describe como:

$$\ln[I(s)] = \ln[I(0)] - \frac{s^2 R_g^2}{3}$$

donde $I(0)$ es la intensidad de dispersión extrapolada a ángulo cero y R_g es el radio de giro. El gráfico de Guinier, en el que se representa $\ln[I(s)]$ frente a s^2 , ha de ser lineal en el rango $0.7 < sR_g < 1.3$, para una proteína plegada, pura y monodispersa.

La función de distribución de pares de distancia, $P(r)$, que se puede considerar como la probabilidad de la presencia de distancias inter-atómicas específicas en la muestra, se obtuvo mediante métodos de transformada de Fourier indirecta con el programa GNOM (Svergun, 1992) a partir del perfil de dispersión. A partir de esta función se calculó la distancia máxima de la partícula, D_{\max} y el R_g .

La descomposición de los perfiles de SAXS en valores singulares (SVD) se realizó con el programa SVDPLOT (Konarev et al., 2003). Este método permite estimar el número de especies presentes en la muestra sin requerir un conocimiento previo de la estructura de dichos componentes. Siendo A una matriz formada por los datos de dispersión de un conjunto de curvas, según la teoría de SVD se puede expresar como:

$$A = U (S V^T)$$

donde U es una matriz en la que las columnas son un conjunto de vectores básicos que se pueden combinar linealmente para representar las curvas de SAXS a las diferentes concentraciones, S es una matriz diagonal que contiene los valores

singulares de los vectores básicos y V^T es una matriz que contiene los vectores de amplitud de cada vector básico (Figura M4).

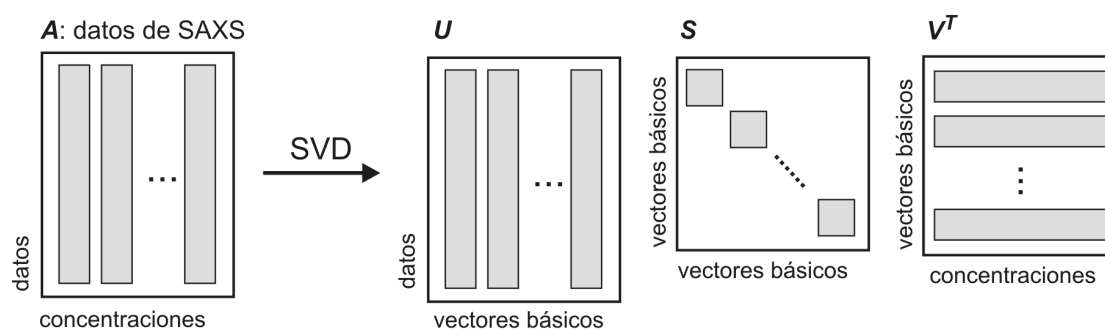


Figura M4. Representación esquemática de la descomposición de datos de SAXS en valores singulares. A partir de un conjunto de perfiles de dispersión medidos a varias concentraciones de proteína se obtienen los vectores básicos, sus valores singulares y los vectores de amplitud. Adaptado de (Williamson et al., 2008).

6.7.3. Modelado *ab initio* de estructuras a partir de datos de SAXS

A partir de los perfiles de dispersión experimentales, se calcularon estructuras tridimensionales a baja resolución mediante métodos *ab initio* utilizando el programa de modelado DAMMIF (Franke and Svergun, 2009) aplicando las simetrías P1 y P2 de manera independiente. Este programa crea modelos formados por esferas que simulan pseudo-átomos (*dummy atom model*) densamente empaquetados dentro de un volumen esférico de diámetro igual a la distancia máxima de la molécula (D_{\max}). El programa va modificando los modelos hasta que la curva teórica de dispersión se ajusta a la experimental. Se generaron 15 modelos independientes para cada una de las simetrías.

Cada grupo de 15 modelos se compararon entre sí mediante superposición y alineamiento con el programa SUPCOMB (Kozin and Svergun, 2001). El parámetro NSD (Kozin and Svergun, 2001) (discrepancia real normalizada) se utilizó para caracterizar las diferencias entre modelos que presentan la misma resolución. En modelos idénticos, el valor de NSD es cero. El modelo que presentó valor más bajo de NSD se utilizó como referencia, mientras que los modelos que presentaron un valor de NSD superior a la media y 2 veces la desviación estándar se eliminaron del análisis. A partir de los modelos alineados se empleó el programa

DAMAVER (Volkov and Svergun, 2003) para calcular una estructura promedio que representa el volumen más poblado dentro de cada grupo de modelos. La representación volumétrica de las envueltas modeladas de SAXS se realizó mediante el programa SITUS 2.5 (Wriggers et al., 1999).

6.9. Ensayos de unión a lípidos

En los ensayos de unión a lípidos se emplearon vesículas de fosfatidilcolina (PC) o de una mezcla 1:1 de PC y fosfatidilserina (PS) (Avanti Polar Lipids). Para preparar las vesículas se depositaron en tubos de vidrio las cantidades necesarias de soluciones de los lípidos en cloroformo y se secaron, primero bajo una corriente de nitrógeno y después mediante vacío durante 2 h. Posteriormente los lípidos se resuspendieron en 20 mM Tris-HCl (pH 7.5), 150 mM NaCl a 37 °C durante 2h, seguido de sonicación en un baño durante 30 min.

La unión de fragmentos de N-C3G a vesículas de lípidos se realizó mediante ensayos de cosedimentación. Las vesículas, a una concentración de lípidos de 0.5 o 1 mM, se incubaron con los fragmentos de N-C3G (4-245) o (65-245), a una concentración final de 20 µM, durante 30 min a temperatura ambiente; el volumen de cada reacción fue de 100 µL. Seguidamente las vesículas se sedimentaron mediante centrifugación 100000 *g* en un rotor TLA-100 (Beckman Coulter), durante 15 min a 4 °C. Se recogieron los sobrenadantes y los precipitados se resuspendieron en 100 µL de 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. La presencia de las proteínas en las fracciones libre y la unida a las vesículas se analizó mediante SDS-PAGE.

RESULTADOS

1. Organización estructural de C3G

1.1. Interacción de N-C3G con otras regiones de C3G: la región N-C3G interacciona con el dominio REM

La delección de la mitad N-terminal de C3G aumenta la actividad GEF, lo que sugiere que esta región constituye un elemento regulador negativo en *cis* (Ichiba et al., 1999). En nuestra hipótesis de partida es posible que el efecto represor de la región amino de C3G sobre su actividad catalítica se deba a la existencia de interacciones intramoleculares en las que participa esta región amino y que den lugar a una conformación cerrada, posiblemente autoinhibida.

La posible interacción de la región amino terminal de C3G (N-C3G) con otras regiones de esta proteína, se ha analizado mediante la expresión de fragmentos de C3G en células de mamífero (COS-1) combinada con ensayos de co-inmunoprecipitación (coIP). Para ello se ha diseñado una amplia colección de fragmentos que incluyen un número variable de los dominios que componen C3G (Figura R1-A). En el diseño de las construcciones se ha intentado respetar la integridad estructural de los diferentes dominios. La región N-C3G, residuos 4-245, se clonó en el vector de expresión en mamíferos pCEFLAG, que codifica una proteína de fusión que incorpora el epítipo FLAG en el extremo amino. Por otro lado se clonó C3G completo y diferentes fragmentos de delección en el vector pCEFLHA que dirige la expresión de proteínas con la secuencia HA en el extremo amino.

Se realizaron dobles transfecciones en células COS-1 en las que se combinó el fragmento FLAG-N-C3G (4-245) con diferentes construcciones de C3G con la etiqueta HA; además de los respectivos controles en los que las construcciones HA-C3G se cotransfectaron con el vector pCEFLAG vacío. La correcta expresión de las proteínas se comprobó mediante inmunoblot de los lisados celulares. La proteína FLAG-N-C3G se inmunoprecipitó con el anticuerpo monoclonal M2 anti-FLAG, mientras que las proteínas co-inmunoprecipitadas se detectaron con un anticuerpo policlonal anti-HA (Figura R1-B).

C3G completo (4-1077) no interacciona con N-C3G (4-245). Sin embargo el fragmento C3G- Δ N (246-1077) que carece del dominio amino terminal, sí interacciona con N-C3G, lo que sugiere que en la molécula intacta el sitio de

interacción con el dominio amino se encuentra bloqueado. Cuatro fragmentos, 277-1077, 292-1077, 465-1077 y 550-1077, que corresponden a deleciones progresivas de los cuatro primeros motivos ricos en Pro de la región SH3b, así como la región catalítica aislada, formada por los dominios REM-CAT (670-1077), interaccionan con N-C3G. Por otro lado, la región central SH3b aislada, que incluye los cinco motivos ricos en Pro, no interacciona, lo que indica que la región central SH3b no es necesaria ni parece contribuir significativamente a la interacción. Dentro de la región carboxilo terminal, N-C3G (4-245) interacciona con el dominio REM pero no lo hace con el dominio CAT. Considerando en conjunto todos estos datos, se deduce que el dominio REM (670-814) es necesario y suficiente para la interacción con N-C3G (4-245).

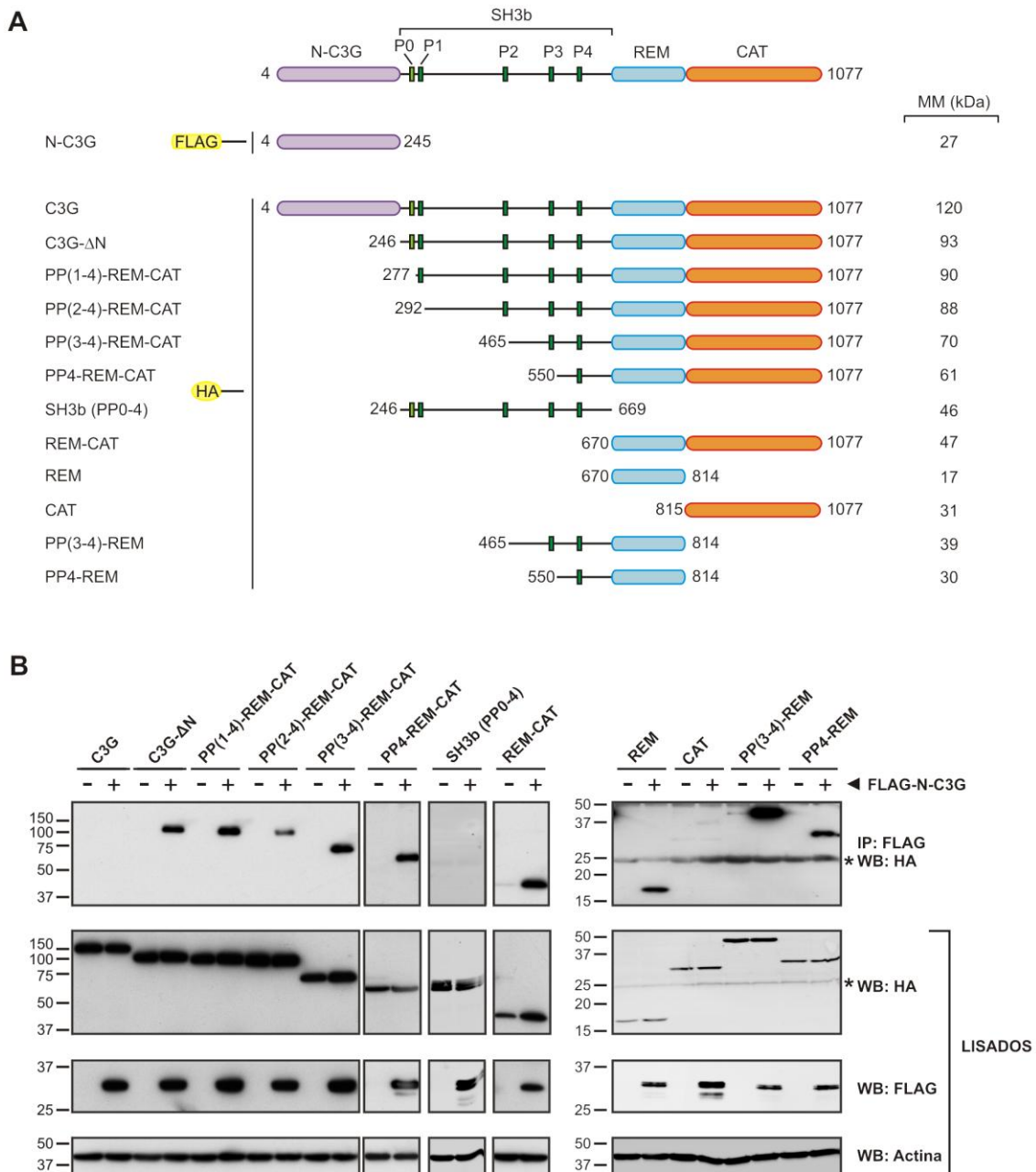


Figura R1. Análisis de la interacción de N-C3G con otras regiones de C3G mediante coIP. (A) Representación de los fragmentos de C3G empleados en los ensayos de coIP. La estructura de C3G se muestra en la parte superior, bajo ésta se muestra la construcción FLAG-N-C3G y los fragmentos de C3G con la etiqueta HA. A la derecha se indica la masa molecular de cada fragmento calculada a partir de la secuencia teórica de aminoácidos. (B) Ensayos de coIP donde se analiza la interacción de FLAG-N-C3G (4-245) con diversos fragmentos de C3G con la etiqueta HA. Para cada fragmento de HA-C3G se analizaron dos co-transfecciones: una con el vector pCEFLAG vacío (-) y otra con pCEFLAG-N-C3G (+). Aquellos fragmentos de HA-C3G que co-precipitan con el anticuerpo anti-FLAG sólo lo hacen en presencia de FLAG-N-C3G. A la izquierda de cada inmunoblot (WB) se muestra la posición y la masa molecular de los marcadores. En los inmunoblots con el anticuerpo anti-HA se observa una banda inespecífica con una masa molecular ~ 25 kDa que se indica con un asterisco.

1.2. Interacción intramolecular de tipo cabeza-cola en C3G, análisis mediante FRET

Con el fin de analizar si la interacción entre la región amino y el dominio REM de C3G es un proceso inter- o intramolecular se aplicaron métodos de fluorescencia empleando la técnica de FRET.

Para el estudio de la organización estructural de C3G se generó un conjunto de sensores conformacionales de C3G en un derivado del vector pcDNA3 que codifica proteínas de fusión que incorporan la proteína fluorescente *cian* (CFP) en el extremo amino y la proteína fluorescente amarilla Venus en el extremo carboxilo (Figura R2-A). Se crearon sensores de C3G completo (4-1077), C3G- Δ N (246-1077), C3G- Δ CAT (4-814) y C3G- Δ REM-CAT (4-669). También se creó la construcción CFP-C3G (4-1077) que codifica C3G con sólo CFP en el extremo amino y C3G-Venus (4-1077) que sólo contiene Venus en el extremo carboxilo. Las construcciones se transfectaron y expresaron en células COS-1 y se confirmó su correcta expresión mediante inmunodetección en lisados celulares (Figura R2-B).

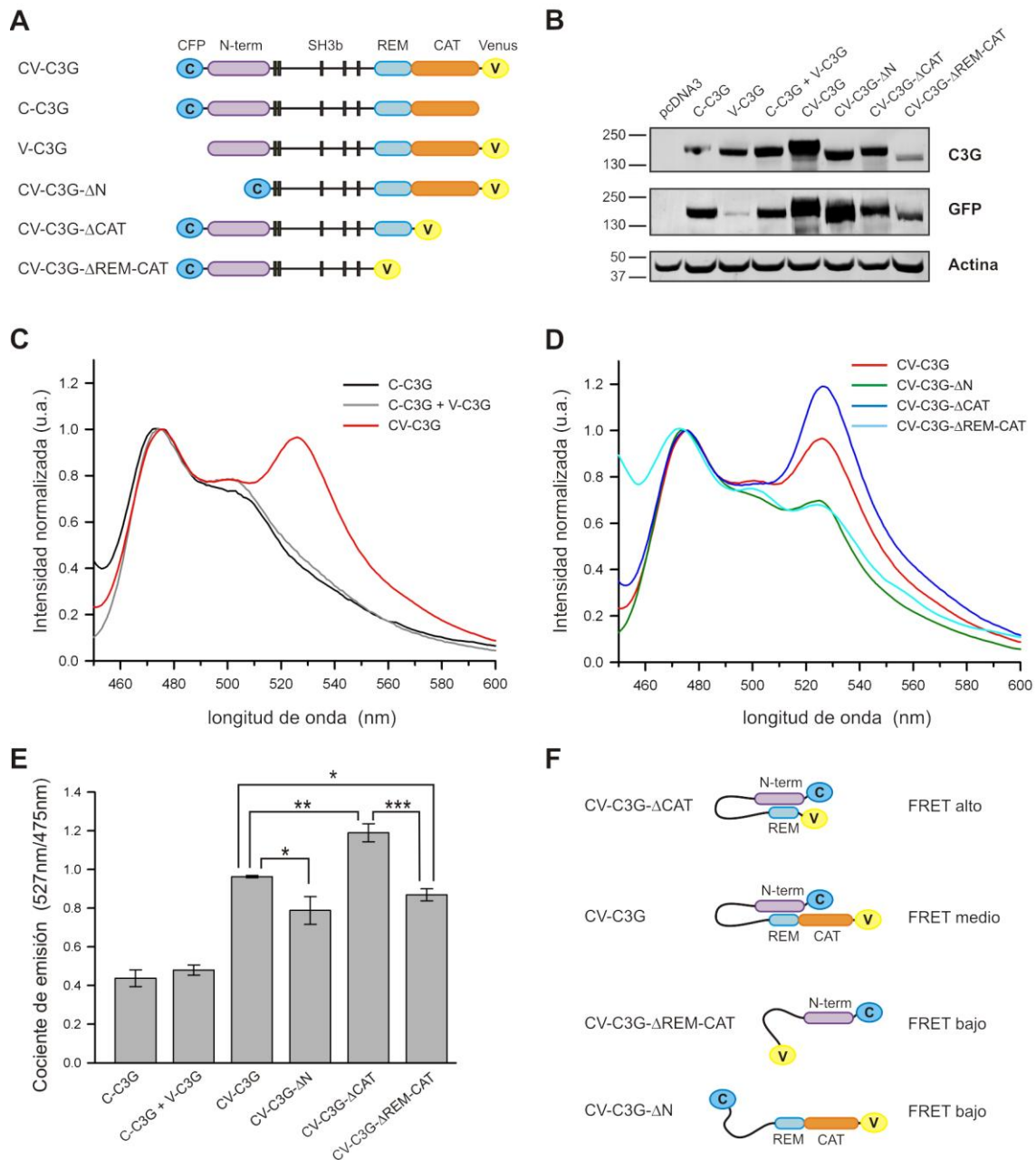


Figura R2. Análisis de la organización estructural de C3G mediante FRET. (A) Estructura esquemática de los sensores conformacionales de C3G empleados en los ensayos de FRET: CFP-C3G-Venus (4-1077), se ha simplificado como CV-C3G en la figura; CFP-C3G (4-1077, C-C3G) con sólo CFP en el extremo amino; C3G-Venus (4-1077, V-C3G), con sólo Venus en el extremo carboxilo. Y los mutantes de deleción CFP-C3G-ΔN-Venus (246-1077, CV-C3G-ΔN); CFP-C3G-ΔCAT-Venus (4-814, CV-C3G-ΔCAT) y CFP-C3G-ΔREM-CAT-Venus (4-669, CV-C3G-ΔREM-CAT). (B) Inmunodetección de los diferentes fragmentos de C3G en lisados de células COS-1 con anticuerpos anti-C3G (H-300) y anti-GFP. Se observa la correcta expresión de los fragmentos. Como control de carga se analizaron mediante inmunoblot los niveles de actina (panel inferior). (C) Espectros de emisión (λ_{exc} de 435 nm, específica para CFP) normalizados por la intensidad a 475nm, medidos en células en suspensión que expresan C-C3G (4-1077) (negro), coexpresión de C-C3G (4-1077) y V-C3G (4-1077) (gris) o CV-C3G (4-1077) (rojo). (D) Espectros de emisión medidos en las mismas condiciones, de CV-C3G (4-1077) (rojo) y los mutantes de deleción CV-C3G-ΔCAT (4-814) (azul oscuro), CV-C3G-ΔN (246-1077) (verde) y de CV-C3G-ΔREM-CAT (4-669) (cian). (E) Cocientes de la intensidad emisión (527nm/475nm) de las construcciones de C3G. Se muestran la media \pm desviación estándar de al menos 3 experimentos independientes. * $p < 0.01$, ** $p < 0.0001$ y *** $p < 0.0001$.

0.000001. (F) Posibles conformaciones que adoptan los sensores conformacionales de FRET para C3G. CV-C3G- Δ CAT mantendría una conformación cerrada en la que el extremo amino de la región N-terminal y el extremo carboxilo del dominio REM se encuentran próximos, lo cual resulta en un FRET alto. CV-C3G, también presentaría una conformación cerrada, pero muestra un nivel de FRET medio posiblemente por una mayor separación entre las sondas a la observada en CV-C3G- Δ CAT. Por último, los mutantes CV-C3G- Δ REM-CAT y CV-C3G- Δ N, corresponderían a conformaciones abiertas, en las que las sondas fluorescentes se encuentran alejadas entre sí y pueden rotar con mayor libertad, dando lugar a señales de FRET bajas.

En los ensayos de FRET se midió la fluorescencia en lisados celulares o en células en suspensión, utilizando una longitud de onda de excitación (λ_{exc}) de 435 nm, específica para CFP. El espectro de emisión de CFP se caracteriza por un máximo a ~ 475 nm, tal como se observa en muestras de células transfectadas sólo con CFP-C3G (4-1077) (Figura R2-C). Cuando se expresó la construcción CFP-C3G-Venus (4-1077), el espectro de emisión mostró un nuevo máximo a ~ 527 nm, máximo de emisión de Venus, debido a la transferencia de energía del donador al aceptor, lo cual indica que las proteínas fluorescentes situadas en los extremos amino y carboxilo están espacialmente próximas. Con el fin de evaluar si la proximidad entre las regiones de cabeza y cola se produce de forma intra- o intermolecular se coexpresaron en COS-1 las construcciones CFP-C3G (4-1077) y C3G-Venus (4-1077); en este caso no se observó el pico a ~ 527 nm en el espectro de emisión, indicando que no hay proximidad entre las regiones amino y carboxilo de diferentes moléculas de C3G. En resumen, estos datos indican que los extremos de C3G están espacialmente próximos y revelan una interacción intramolecular de tipo cabeza-cola que mantiene a la molécula en una conformación cerrada. En estos experimentos se obtuvieron resultados similares empleando células en suspensión o lisados celulares (datos no mostrados), si bien la eficiencia de FRET (estimada como el cociente de la intensidad entre 527 nm y 475 nm) parece ligeramente superior en las células en suspensión. Por lo tanto, el resto de los experimentos de FRET se realizaron empleando células en suspensión.

Una vez que se puso de manifiesto la existencia de una interacción cabeza-cola en C3G, y con el fin de evaluar la contribución de diversos dominios a la interacción, se analizaron tres construcciones del sensor conformacional FRET de C3G, con CFP y Venus en los extremos amino y carboxilo respectivamente, en las que se eliminaron dominios de C3G (Figura R2-D). La amplitud del pico a 527 nm de CFP-C3G- Δ CAT-Venus (4-814) es mayor que la observada para CFP-C3G-Venus

(4-1077), lo cual revela una mayor eficiencia de FRET al eliminar el dominio CAT y sugiere que el extremo amino de N-C3G está más próximo del extremo carboxilo del dominio REM que del extremo carboxilo del dominio CAT. Por otro lado, en el espectro de emisión de CFP-C3G- Δ N-Venus (246-1077), que carece de la región N-terminal, se observa un descenso marcado de la intensidad del máximo a 527 nm. De forma similar en el espectro de emisión de CFP-C3G- Δ REM-CAT-Venus (4-669) se observa una señal a 527 nm relativamente baja. La menor intensidad de fluorescencia a 527 nm revela una reducción de FRET y sugiere un distanciamiento de las sondas por pérdida de la interacción cabeza-cola en estas dos construcciones. El hombro a 527 nm que se observa en los espectros de CFP-C3G- Δ N-Venus y CFP-C3G- Δ REM-CAT-Venus puede deberse a FRET residual debido a que la separación máxima entre donador y aceptor está limitada por formar parte de una misma cadena polipeptídica. En conjunto, estos datos indican que el dominio CAT (CDC25H) es dispensable para la interacción cabeza-cola, la cual se debe a la proximidad espacial de la región N-C3G y del dominio REM.

1.3. Identificación de residuos del dominio REM importantes para la interacción con el dominio N-C3G

Con el fin de identificar residuos específicos del dominio REM que participan en la interacción con el dominio N-C3G se diseñaron mutaciones puntuales de residuos potencialmente expuestos en la superficie del dominio REM y se analizó su efecto sobre la interacción con N-C3G mediante ensayos de coIP.

No se conoce la estructura tridimensional del dominio REM de C3G, pero se han descrito las estructuras de los dominios REM de las proteínas SOS1 (Boriack-Sjodin et al., 1998), Epac2 (Rehmann et al., 2006), RasGRP1 (Iwig et al., 2013) y Rlf (Popovic et al., 2013a). El dominio REM de C3G muestra un 20 % de identidad de secuencia con el dominio REM de SOS1, 9 % con el REM de Epac2, 15 % con el REM de RasGRP1 y 16 % con el REM de Rlf. Con el fin de identificar residuos en C3G potencialmente expuestos en la superficie del dominio, hemos utilizado la estructura del dominio REM de SOS1 y un alineamiento basado en perfiles de secuencia (Figura R3).

Los dominios REM de SOS1 y otros RasGEFs están formados por un núcleo de 6 α -hélices (α 1 a α 6). Con el fin de explorar ampliamente la superficie del dominio REM de C3G se seleccionaron 6 residuos de la superficie de la estructura del dominio REM de SOS1 (Figura R3-A) y mediante el alineamiento de secuencia se identificaron los residuos equivalentes en C3G (Figura R3-B). Cinco de estos residuos de C3G (L714, E731, F760, E784 y K803) se localizan potencialmente en cada una de las α -hélices α 2, α 3, α 4, α 5 y α 6, y un sexto residuo (L774) forma parte del lazo que une las hélices α 4 y α 5. Estos residuos se sustituyeron por Arg a excepción de K803 que se sustituyó por Glu para introducir un cambio de carga en la presente cadena lateral. De esta manera se crearon los mutantes puntuales: L714R, E731R, F760R, L774R, E784R y K803E; posteriormente se creó el mutante doble E731R-E784R.

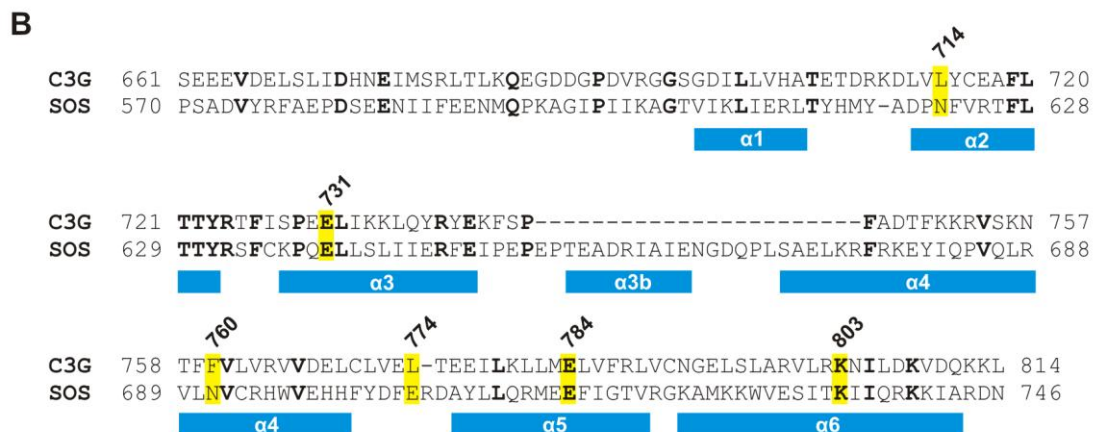
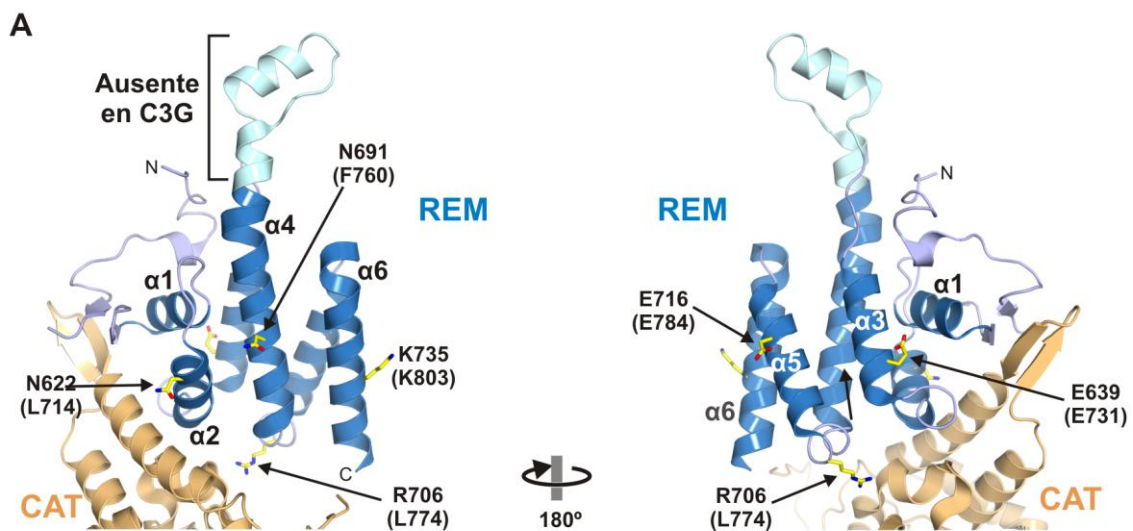


Figura R3. Identificación de residuos de C3G expuestos en la superficie del dominio REM por comparación con la estructura de la región REM-CAT de SOS1. (A) Estructura del dominio REM (azul) y el dominio CAT (CDC25H, naranja) de SOS1 tal como aparece en el complejo ternario con Ras unido en el sitio activo y una segunda molécula de Ras-GppNp (análogo de GTP no hidrolizable) en el sitio alostérico (código PDB: 1NVV). Por claridad no se han representado las moléculas de Ras. Se muestra la cadena lateral de los residuos de SOS1 equivalentes a los residuos de C3G que se han mutado. Los números entre paréntesis indican los residuos equivalentes en C3G. (B) Alineamiento de las secuencias de los dominios REM de C3G y SOS1, calculado con el servidor FFAS (<http://ffas.burnham.org/>) (Jaroszewski et al., 2011). Los residuos de C3G mutados y los aminoácidos equivalentes de SOS1 se indican con recuadros amarillos.

Inicialmente estas mutaciones se introdujeron en el fragmento correspondiente a los dominios REM-CAT (670-1077) en el vector pCEFLHA. Los niveles de expresión del fragmento HA-REM-CAT, silvestre y de los mutantes puntuales, son marcadamente dependientes de la co-expresión de FLAG-N-C3G, siendo la expresión de HA-REM-CAT generalmente baja cuando se cotransfecta con vector pCEFLAG vacío (Figura R4). Empleando esta construcción no se observaron diferencias entre los mutantes en su interacción con N-C3G en ensayos de coIP, además se observa inmunoprecipitación inespecífica de algunas proteínas HA-REM-CAT en los ensayos control en los que se cotransfectaron con pCEFLAG vacío.

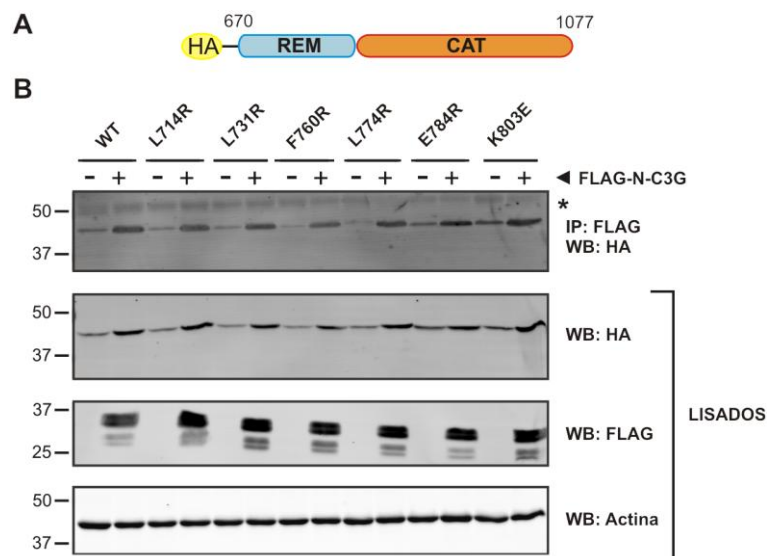


Figura R4. Análisis del efecto de mutaciones puntuales en el dominio REM sobre la interacción de N-C3G con la región REM-CAT. (A) Esquema de la estructura del fragmento HA-REM-CAT. (B) Ensayo de coIP para los mutantes puntuales de C3G creados en el fragmento HA-C3G-REM-CAT (670-1077). Los niveles de expresión del fragmento REM-CAT son más altos cuando se co-expresa con FLAG-N-C3G (+). (-) indica coexpresión con el vector pCEFLAG vacío. En el inmunoblot con el anticuerpo anti-HA se observa una banda inespecífica con una masa molecular ~ 50 kDa que se indica con un asterisco.

Seguidamente se analizó el efecto de las mutaciones puntuales empleando el dominio REM aislado. La construcción HA-REM (670-814) presenta niveles de expresión bajos y muy variables en células COS-1, lo cual dificulta la correcta comparación del efecto de las mutaciones (Figura R5). Los resultados parecen sugerir que las mutaciones L714R, E731R y E784R comprometen la interacción, si bien la falta de reproducibilidad en la expresión del dominio REM impidió un análisis riguroso.

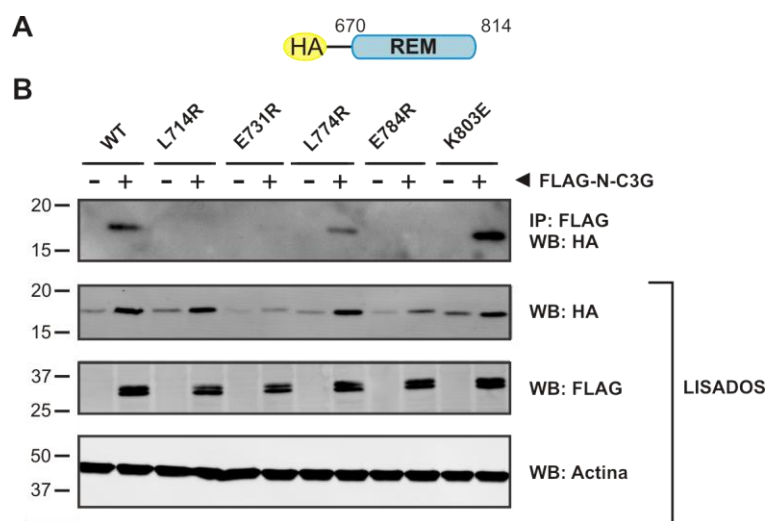


Figura R5. Análisis del efecto de mutaciones puntuales en el dominio REM aislado sobre la interacción con N-C3G. (A) Esquema de la estructura del fragmento HA-REM. (B) Ensayo de coIP con el fragmento FLAG-N-C3G (+) o vector vacío (-) para los mutantes puntuales de C3G diseñados en el fragmento C3G-REM (670-814). En este caso no se analizó el mutante F760R. Individualmente, las sustituciones L714R, E731R y E784R comprometen la interacción con N-C3G.

La dificultad para analizar el efecto de las mutaciones en el dominio REM parecía estar ligada a la baja expresión del dominio REM y del fragmento REM-CAT en células COS-1. Por otro lado, habíamos observado que la inclusión de parte de la región SH3b en construcciones que también contienen la región REM-CAT mejoraba la expresión hasta niveles adecuados para los ensayos de coIP (Figura R1). Por lo tanto se trabajó con las dos construcciones en el vector pCEFLHA que incluían el dominio REM y parte del SH3b, PP(3-4)-REM (465-814) y PP4-REM (550-814). De éstas, PP(3-4)-REM presenta los mejores y más constantes niveles de expresión en COS-1 y es el fragmento que se utilizó para evaluar el efecto de las mutaciones en el dominio REM sobre la interacción con N-C3G mediante ensayos

de coIP (Figura R6). Ni la forma silvestre ni los mutantes puntuales de HA-PP(3,4)-REM inmunoprecipitan cuando se co-expresan con el vector pCEFLAG vacío. Las mutaciones L714R, F760R, L774R y K803E no afectan apreciablemente a la interacción con FLAG-N-C3G. Por otro lado, las sustituciones puntuales E731R y E784R comprometen la interacción de PP(3-4)-REM con N-C3G (4-245), observándose tan sólo niveles marginales de estos mutantes de PP(3-4)-REM en el inmunoprecipitado. Se creó también un doble mutante en el que se combinaron las sustituciones E731R-E784R, el cual tampoco interaccionó con FLAG-N-C3G, si bien el ensayo de coIP no permitió discernir si E731R y E784R tienen un efecto aditivo.

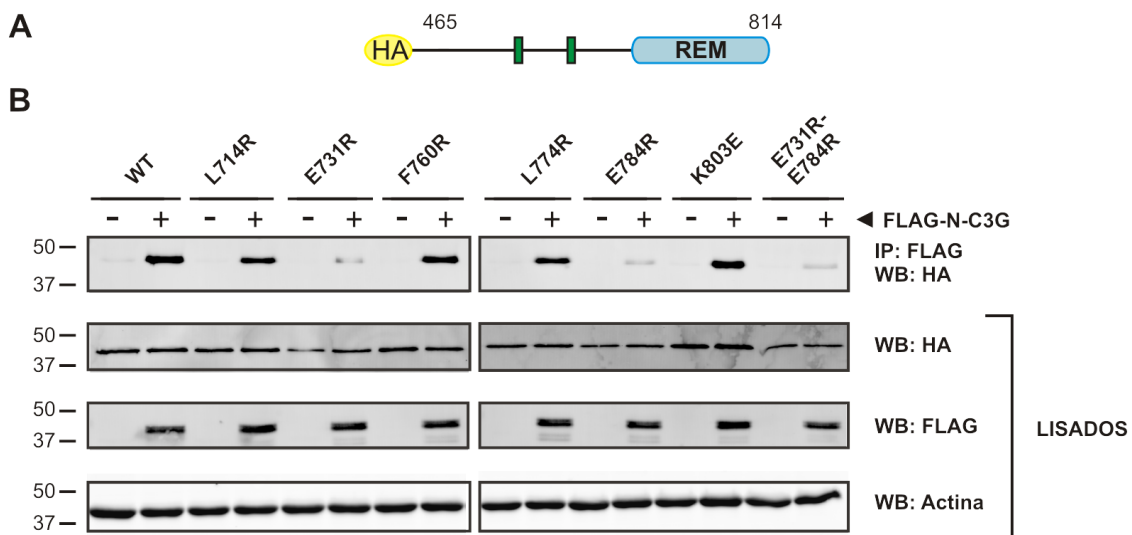


Figura R6. Análisis del efecto de mutaciones puntuales en el dominio REM sobre la interacción de N-C3G con la región PP(3-4)-REM. (A) Esquema de la estructura del fragmento HA-PP(3-4)-REM (465-814). (B) Ensayo de coIP con el fragmento FLAG-N-C3G (+) o vector vacío (-) y mutantes puntuales del fragmento C3G-PP(3-4)-REM. Las mutaciones E731R y E784R, así como el doble mutante que combina estas dos sustituciones, comprometen la interacción con N-C3G.

Por comparación con las estructuras de SOS1 y Epac2, los residuos equivalentes a E731 y E784 se localizan próximos en una zona de la superficie del dominio REM formada por las hélices $\alpha 3$ y $\alpha 5$ que es la potencial zona de interacción con el dominio N-C3G (Figura R7).

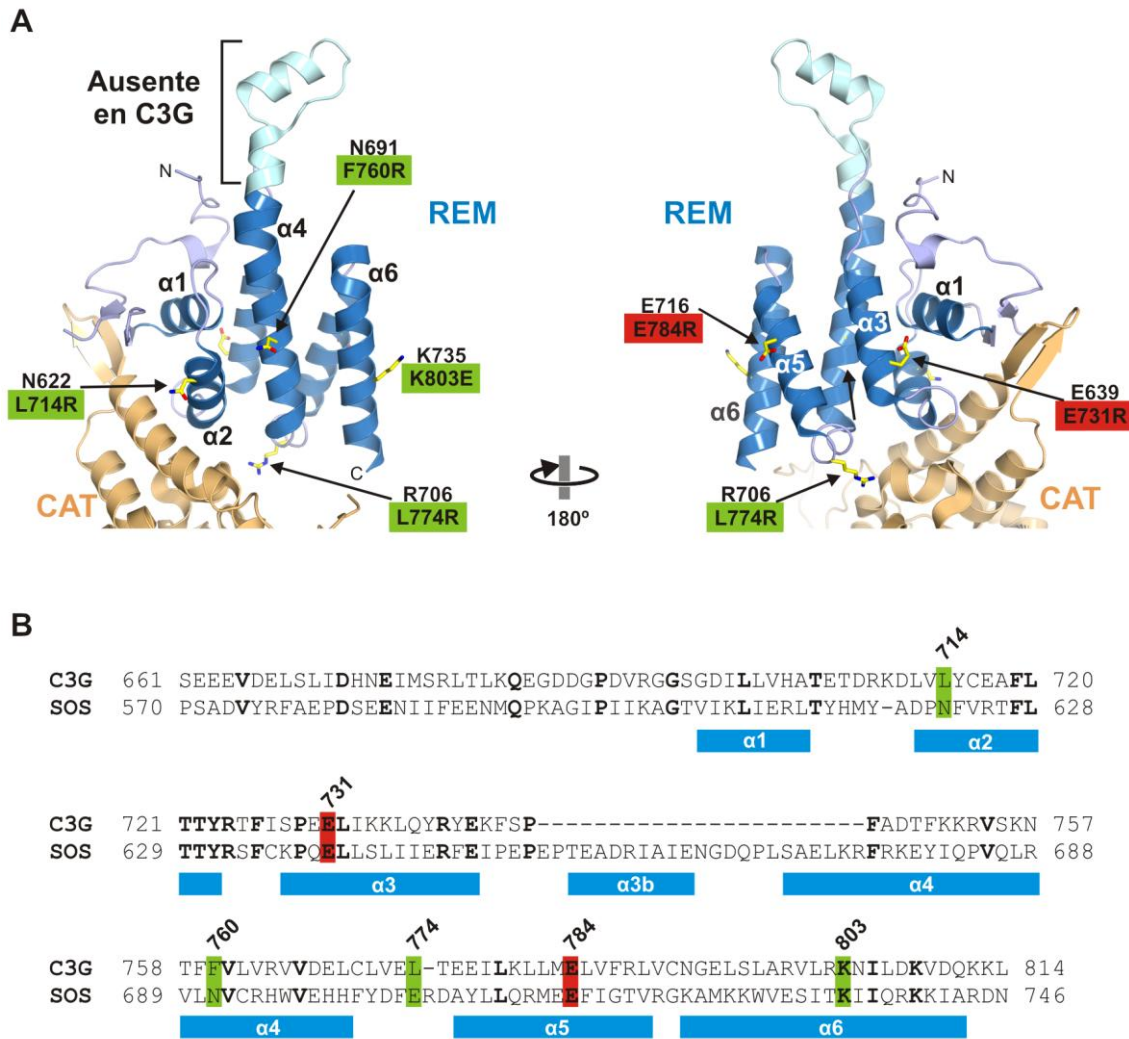


Figura R7. Los residuos E731 y E784 se localizan próximos en la superficie del dominio REM. (A) Estructura de la región REM-CAT de SOS1 (código PDB: 1NVV). En rojo se indican las mutaciones de los residuos equivalentes de C3G que comprometen la interacción con N-C3G, y en verde las que no lo hacen. Los residuos equivalentes a E731 y E784 se localizan en las hélices α_3 y α_5 respectivamente, y definen una potencial zona de interacción con el dominio N-C3G. (B) Alineamiento de las secuencias de los dominios REM de C3G y SOS1. Los residuos mutados en C3G y el efecto de las mutaciones sobre la interacción con N-C3G se indica mediante rectángulos de colores tal como se muestra en (A).

En resumen, los experimentos de coIP revelaron que el dominio N-C3G interactúa con el dominio REM y que los residuos E731 y E784 de este dominio están implicados en dicha interacción.

Seguidamente se analizó el efecto de las sustituciones E731R y E784R, localizadas en el dominio REM, sobre la organización de C3G empleando sensores conformacionales y métodos de FRET. Las mutaciones individuales E731R y

E784R, y la sustitución doble E731R-E784R se introdujeron en la construcción CFP-C3G- Δ CAT-Venus (4-814) ya que el dominio CAT (CDC25H) es dispensable para mantener la conformación cerrada y esta construcción mostró mayor eficiencia de FRET. En el espectro de emisión de los mutantes sencillos E731R y E784R se observa una ligera pero significativa reducción en la señal de FRET (proporción de los picos a 475 nm y 527 nm), que indican un debilitamiento parcial de la interacción cabeza-cola (Figura R8). Por otro lado, en el doble mutante E731R-E784R se observa un descenso más pronunciado de la señal de FRET. El efecto sinérgico de la combinación de estas dos mutaciones sugiere que estos dos residuos participan en dos puntos de interacción proteína-proteína singulares.

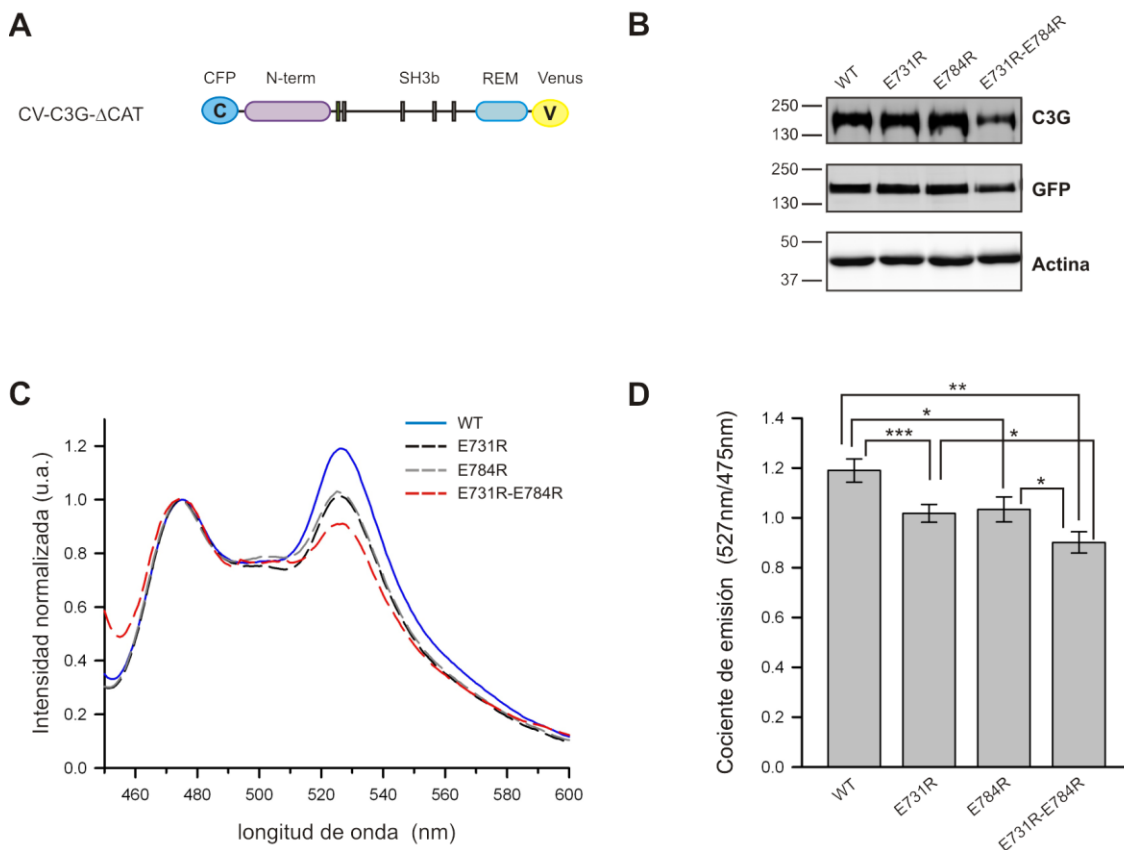


Figura R8. Papel de los residuos E731 y E784 del dominio REM sobre la organización estructural de C3G. (A) Estructura esquemática de la construcción CFP-C3G- Δ CAT-Venus (4-814) en la que se generaron los mutantes sencillos E731R y E784R y el doble mutante E731R-E784R. (B) Inmunodetección de los diferentes mutantes de C3G- Δ CAT (4-814) en lisados de células COS-1 con anticuerpos anti-C3G (H-300) y anti-GFP. Se observa la correcta expresión de los fragmentos. Como control de carga se analizaron mediante inmunoblot los niveles de actina (panel inferior). (C) Espectros de emisión (λ_{exc} de 435 nm, específica para CFP) normalizados por la intensidad a 475nm, medidos en células en suspensión que expresan la forma silvestre de CFP-C3G- Δ CAT-Venus (4-814) (azul) y los mutantes E731R (negro) y E784R (gris) y el doble mutante E731R-E784R

(rojo). (E) Cocientes de la intensidad emisión (527nm/475nm) de los mutantes de C3G-ΔCAT (4-814). *p <0.01, **p <0.001 y ***p <0.00001.

En resumen, los resultados de co-IP y de FRET revelan que los residuos E731 y E784 del dominio REM son importantes para la interacción con el dominio N-C3G y sugieren que pueden participar en una superficie implicada en la interacción intramolecular cabeza-cola.

1.4. Efecto de la sustitución Y504E sobre la interacción cabeza-cola

La fosforilación de la Y504 favorece la activación de C3G mediada por Crk (Ichiba et al., 1999). Con el fin de estudiar el posible papel de Y504 en la regulación de la interacción cabeza-cola analizamos el efecto de la sustitución por Glu (Y504E), que mimetiza la fosforilación de forma aproximada, sobre la interacción entre los dominios N-C3G y REM.

En ensayos de coIP el dominio N-C3G no interacciona con C3G completo (Figura R1), debido probablemente a que el dominio REM está bloqueado por una interacción intramolecular con la región N-C3G. El mutante Y504E tampoco co-precipita con N-C3G (Figura R9), sugiriendo que esta sustitución no inhibe la interacción cabeza-cola.

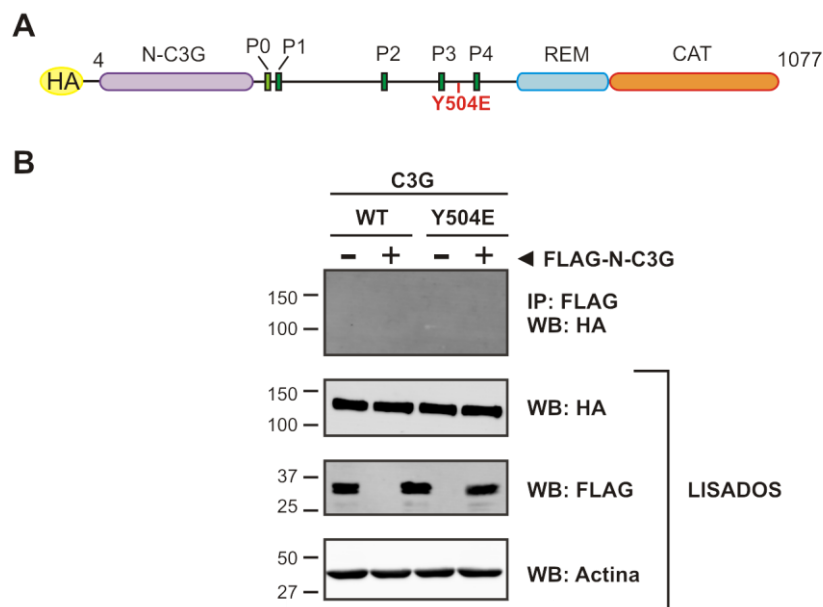


Figura R9. Efecto de la sustitución Y504E sobre la interacción de C3G con N-C3G mediante coIP. (A) Esquema de la estructura de C3G, la Tyr504 se localiza entre las regiones P3 y P4 de la región SH3b. (B) Ensayo comparativo de la posible interacción de C3G silvestre (WT) y del mutante Y504E con la región N-C3G (4-245) expresada con una secuencia FLAG. C3G-HA (WT o Y504E) se co-expresó con N-C3G-FLAG (+) o con el vector pCEFLAG vacío (-) en células COS-1. N-C3G-FLAG se inmunoprecipitó con el anticuerpo M2 anti-FLAG y se analizó la presencia de C3G-HA mediante inmunoblot en el precipitado. No se detectó interacción de N-C3G-FLAG con C3G-HA ni con el mutante Y504E. Los niveles de las proteínas expresadas se analizaron en los lisados celulares totales mediante inmunodetección.

De forma complementaria se analizó el efecto de la sustitución Y504E sobre la estructura de C3G mediante FRET empleando el sensor conformacional CFP-C3G- Δ CAT-Venus (4-814) (Figura R10). No se observan cambios significativos entre los espectros de emisión de la forma silvestre y del mutante Y504E, lo que indica que esta sustitución no es suficiente para inhibir la interacción intramolecular de C3G.

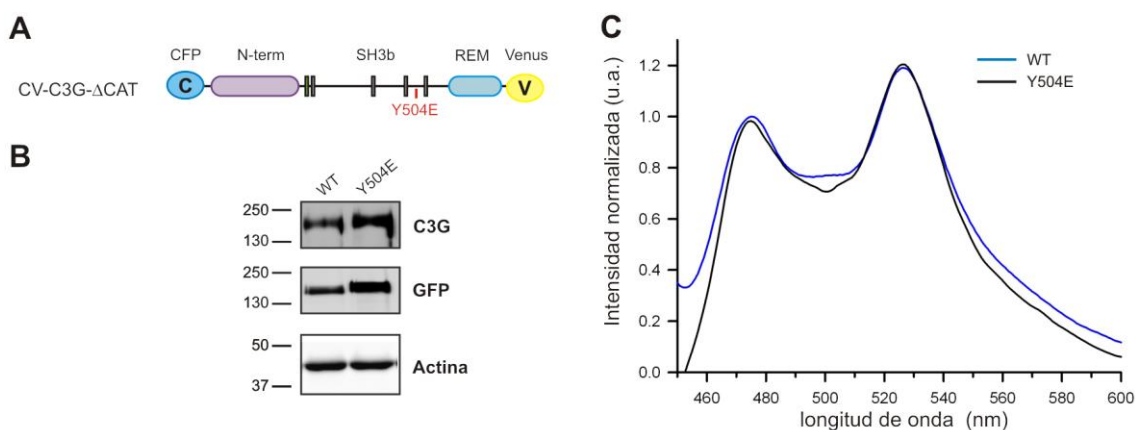


Figura R10. Análisis mediante FRET del efecto de la sustitución Y504E sobre la interacción cabeza-cola. (A) Esquema de la estructura del sensor conformacional CFP-C3G- Δ CAT-Venus (4-814) donde se indica la posición de la mutación Y504E. (B) Inmunodetección de la expresión de CFP-C3G- Δ CAT-Venus (4-814) silvestre (WT) y del mutante Y504E en lisados de células COS-1 con anticuerpos anti-C3G (H-300) y anti-GFP. Como control de carga se analizaron mediante inmunoblot los niveles de actina (panel inferior). (C) Espectros de emisión (λ_{exc} de 435 nm, específica para CFP) normalizados por la intensidad a 475nm, medidos en células en suspensión que expresan la forma silvestre de CFP-C3G- Δ CAT-Venus (4-814) (azul) y el mutante Y504E (negro).

2. Estructura de la región N-C3G

Con el fin de caracterizar la estructura de la región N-terminal de C3G se crearon cuatro construcciones que codifican las regiones 4-210, 4-245, 65-210 y 65-245 en el vector de expresión en bacterias pETEV15b (Figura R11). Los límites de estos fragmentos se diseñaron en función de la predicción de estructura secundaria. Los fragmentos 4-245 y 65-245 se produjeron solubles en *E. coli* como proteínas de fusión con una cola de poli-His en posición amino terminal y se purificaron mediante cromatografía de afinidad. En cambio, no se detectó expresión de los fragmentos 4-210 y 65-210. Esto sugiere que la región 211-245 es necesaria para la estabilidad de los fragmentos correspondientes a la región N-C3G, mientras que los primeros 65 residuos no son imprescindibles para dicha estabilidad.

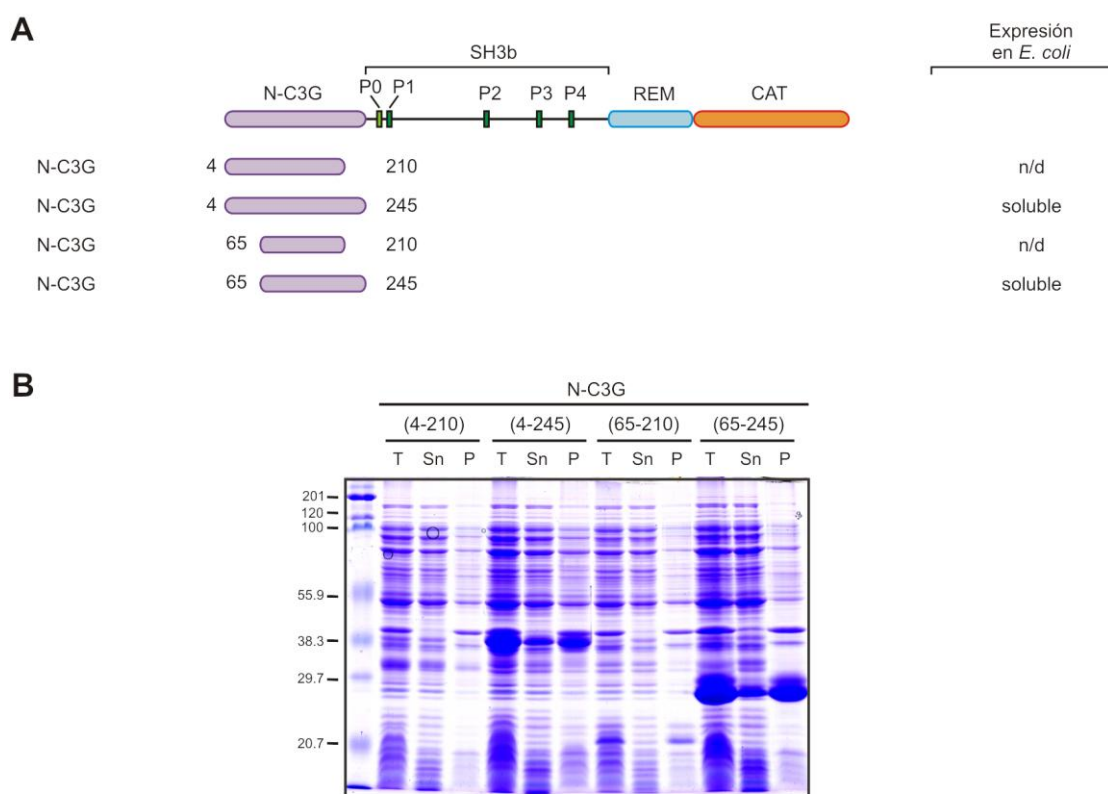


Figura R11. Construcciones de la región amino terminal de C3G clonadas en un vector de expresión en bacterias. (A) Representación esquemática de la estructura primaria de C3G bajo la cual se muestran los fragmentos de C3G clonados en el vector pETEV15b de expresión en *E. coli*. A la derecha se indica el resultado de la expresión en la cepa BL21(DE3) de *E. coli*. (B) Análisis mediante SDS-PAGE de los diferentes fragmentos de N-C3G; siendo (T) lisado total; (Sn) sobrenadante con lo fracción soluble y (P) precipitado o fracción insoluble. No se detectó expresión de los fragmentos 4-210 y 65-210 (n/d).

Se realizaron ensayos de cristalización de los fragmentos 4-245 y 65-245, no obteniéndose cristales. La región N-terminal de C3G contiene una única Cys en la posición 154. Con el fin de evitar la posible reactividad del grupo tiol, se generó el mutante C154A del fragmento 65-245. Este mutante puntual también se sometió a ensayos de cristalización sin éxito.

2.1. Análisis de la estructura secundaria de la región N-C3G

Para estudiar la estructura secundaria se midieron espectros de dicroísmo circular (CD) en el UV-lejano de los fragmentos recombinantes de N-C3G (4-245) y (65-245) (Figura R12). Los espectros se caracterizan por tener un máximo a 192 nm y dos bandas negativas ~ 208 nm y ~ 223 nm. Estas bandas son características de estructuras secundarias de α -hélice (Holzwarth and Doty, 1965) lo que sugiere que esta es la estructura predominante en N-C3G.

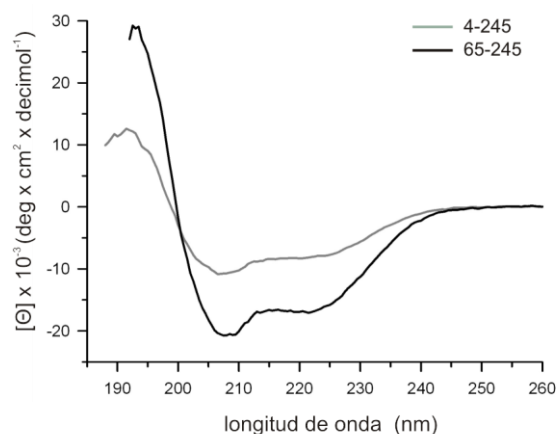


Figura R12. Espectros de dicroísmo circular de los fragmentos 4-245 y 65-245 de la región N-C3G. Las proteínas se equilibraron en 20 mM NaPi (pH 7.5), 150 mM NaCl y se midieron a 20 °C. Los espectros se caracterizan por un máximo a 192 nm y dos mínimos a 208 y 223 nm, con valores de elipticidad molar de 29200, -20500 y -17000 deg cm²/decimol para el fragmento corto 65-245 (línea negra); y de 12500, -10000 y -7800 para el fragmento largo 4-245 (línea gris).

A partir de los espectros de CD se estimó el contenido de los distintos tipos de estructura secundaria empleando cuatro métodos diferentes, SELCON3 (Sreerama et al., 1999; Sreerama and Woody, 1993), CONTIN (Provencher and Glockner, 1981; van Stokkum et al., 1990), CDSSTR (Compton and Johnson, 1986;

Manavalan and Johnson, 1987; Sreerama and Woody, 2000) y K2D3 (Andrade et al., 1993; Louis-Jeune et al., 2011; Perez-Iratxeta and Andrade-Navarro, 2008) (Tabla R1). La estimación del contenido en α -hélice y lámina- β es muy similar con los cuatro métodos empleados. El fragmento 4-245 tendría ~ 24 % de estructura de α -hélice (equivalente a ~ 60 aminoácidos) y un contenido similar de lámina- β , ~ 13 % de giros y ~ 37 % de estructura desordenada. En cambio, para el fragmento corto 65-245 se estimó que el contenido en α -hélice supone más de la mitad de la estructura secundaria, que equivaldría a ~ 97 residuos, siendo el contenido en lámina- β inferior al 10 %. A pesar de las discrepancias en la estimación del contenido en α -hélice de los fragmentos 4-245 y 65-245, parece razonable que el contenido en α -hélice se concentra mayoritariamente en esta segunda región, mientras que el segmento 4-64 carecería de un contenido significativo en estructura helicoidal. Por otro lado, la estimación del contenido en lámina- β del fragmento 4-245 podría estar sobre-estimada.

Tabla R1. Estimación del contenido de estructura secundaria de N-C3G a partir de los espectros de dicroísmo circular.

Fragmento	Método	Rmsd ^b	Tipo de estructura secundaria			
			α -hélice	lámina- β	giros	desordenada
4-245	SELCON3	0.139	23.9	26.4	13.0	36.7
	CONTIN	0.061	25.5	23.9	13.2	37.4
	CDSSTR	0.022	24.0	25.0	13.0	38.0
	K2D3 ^a	0.088	23.5	23.3	n.a.	n.a.
	promedio			24 \pm 1	25 \pm 1	13
65-245	SELCON3	0.162	51.0	4.0	13.6	31.4
	CONTIN	0.093	51.1	4.0	13.2	31.7
	CDSSTR	0.020	59.0	6.0	10.0	25.0
	K2D3 ^a	0.097	53.9	8.5	n.a.	n.a.
	promedio			54 \pm 4	6 \pm 2	12 \pm 2

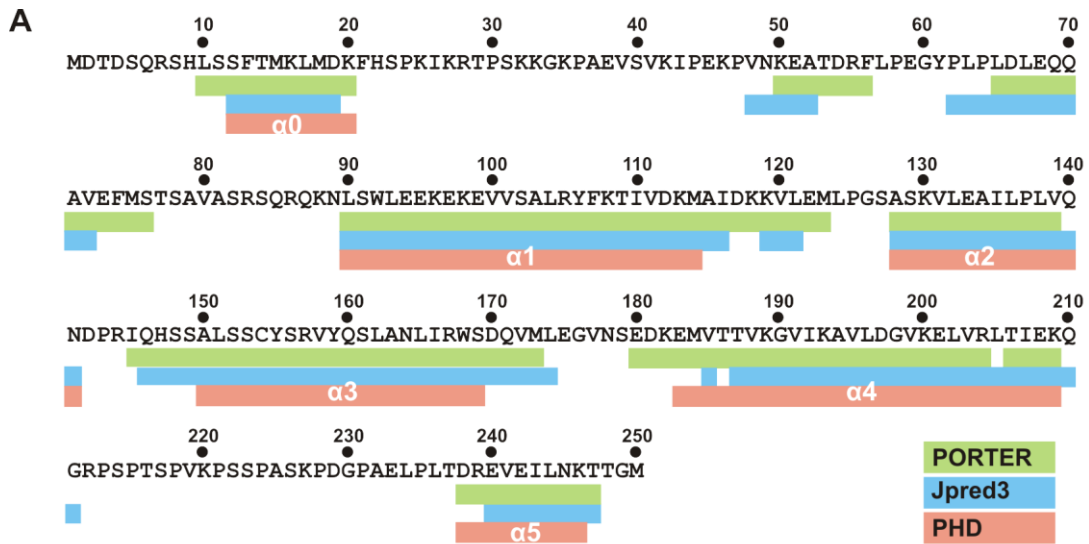
^a, el método K2D3 sólo estima el contenido en α -hélice y lámina- β , y no distingue entre giros y estructura desordenada (n.a.: no aplica). Los valores están expresados en tanto por ciento.

^b, Rmsd: desviación cuadrática media

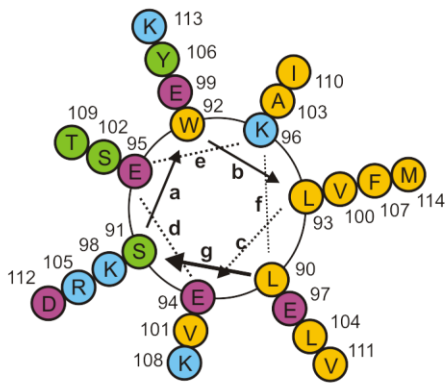
De manera paralela, se realizó una predicción de estructura secundaria a partir de la secuencia con los programas PORTER (Pollastri and McLysaght, 2005), Jpred3 (Cole et al., 2008) y PHD (Rost and Sander, 1993, 1994) (Figura R13-A). Los

tres métodos predicen 4 α -hélices en la región 90-210 (α 1- α 4) y dos hélices cortas en las regiones 10-20 (α 0) y 240-250 (α 5). Dos de los métodos, PORTER y Jpred3, indican la presencia de otras dos hélices aproximadamente en la región 50-75. Ninguno de los métodos predice la presencia de hebras- β en la región amino terminal de C3G. En conjunto el contenido en α -hélice de la región 4-245 predicho por PORTER, Jpred3 y PHD fue de 58 %, 53 % y 46 % respectivamente; mientras que para la región 65-245, el contenido predicho fue mayor: 68 %, 64 % y 57 %. En general se observa una concordancia entre el contenido de estructura secundaria del fragmento 65-245 estimado a partir del espectro de CD y a partir de la secuencia. Si bien, el contenido en α -hélice de la región 4-245 determinado a partir del espectro de CD parece estar sub-estimado.

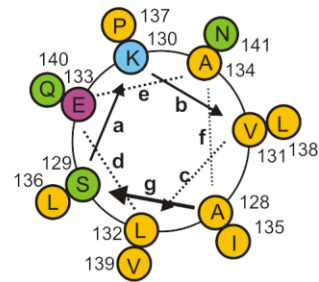
Con el fin de analizar en más detalle la región 90-210 se realizaron diagramas de rueda helicoidal (Schiffer and Edmundson, 1967) de las cuatro potenciales hélices de esta región, α 1 (90-114), α 2 (128-141), α 3 (150-169) y α 4 (183-209) (Figura R13-B). Se observa una distribución asimétrica de los residuos polares y no polares, lo que sugiere que pueden ser hélices anfipáticas que podrían formar una estructura de tipo haz de hélices.



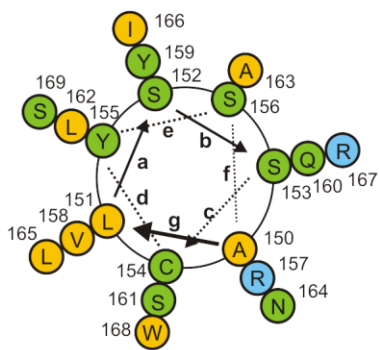
B



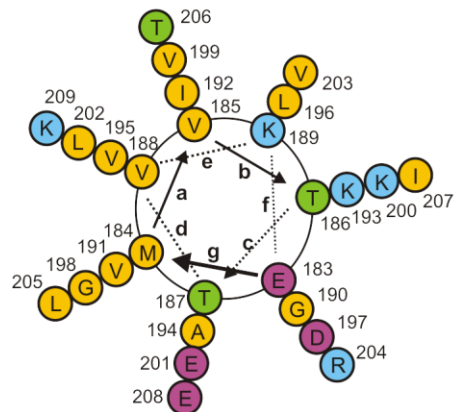
Hélice $\alpha 1$: 90-114



Hélice $\alpha 2$: 128-141



Hélice $\alpha 3$: 150-169



Hélice $\alpha 4$: 183-209

No polares: G, A, V, L, I, M, P, F, W
 Polares (sin carga): Q, N, S, T, Y, C

Ácidos: E, D
 Básicos: K, R, H

Figura R13. Predicción de la estructura secundaria de la región N-C3G. (A) Secuencia de la región amino terminal de C3G y predicción de estructura secundaria. Las cajas de colores bajo la secuencia indican la predicción de α -hélices por los programas PORTER, Jpred3 y PHD. (B) Representación de ruedas helicoidales de los residuos de las 4 potenciales hélices de la región 90-210. Los residuos se han clasificado y coloreado en cuatro categorías según su carácter. Se observa una distribución asimétrica de los residuos polares y no polares.

2.2. Mapas de proteólisis controlada de N-C3G

Ante la falta de éxito en la cristalización de los fragmentos 4-245 y 65-245, éstos se analizaron mediante proteólisis controlada con el fin de identificar posibles núcleos resistentes a proteasas dentro de esta región, que pudieran corresponder a dominios estructuralmente compactos que estarían más indicados para la obtención de cristales. Ambos fragmentos se digirieron de forma independiente con tripsina, quimotripsina o la proteasa V8. Estas enzimas tienen diferentes especificidades químicas: tripsina digiere tras residuos básicos; quimotripsina hidroliza enlaces tras aminoácidos aromáticos y V8 digiere enlaces peptídicos tras residuos ácidos.

El análisis mediante SDS-PAGE reveló que el fragmento N-C3G 4-245 es rápidamente digerido por las tres proteasas empleadas (Figura R14-A). Tripsina produce incluso a tiempos cortos (5 min) dos fragmentos de masa molecular aparente (MMap) ~ 21 kDa y ~ 17 kDa. A tiempos de incubación más largos el fragmento de 21 kDa desaparece y se acumula el de 17 kDa sugiriendo que el segundo se produce por procesamiento del primero. Quimotripsina a tiempos cortos produce un fragmento transitorio de MMap ligeramente inferior a la del fragmento intacto; además da lugar a un patrón de fragmentos de menor tamaño en el que destacan bandas con MMap de ~ 23 kDa, ~ 15 kDa y ~ 12 kDa, que son estables hasta 20 min de digestión. V8 produce un patrón de proteólisis complejo caracterizado por múltiples bandas con MMap en la región de ~ 22 kDa y por un fragmento que migra en la posición del marcador de 7 kDa.

Con el fin de localizar los sitios de corte y de identificar fragmentos proteolíticos estables, se analizaron mediante espectrometría de masas (MS) muestras de C3G (4-245) digeridas con tripsina a los 20 min y quimotripsina a los 10 min (Figura R14-B). No se analizaron muestras digeridas con V8 ya que esta proteasa no da lugar a fragmentos estables de tamaño medio (> 10 kDa). En el

espectro de MS de los fragmentos digeridos por tripsina se detectaron dos fragmentos de 11022 Da y 10324 Da que se asignaron a los fragmentos 145-245 y 145-239, y que podrían corresponder a la banda de MMap ~ 17 kDa observada mediante SDS-PAGE. Además se detectaron otros 6 fragmentos con valores de m/z 8443 (asignado a la región 168-245), 7746 (168-239 o 131-200), 7272 (145-209), 6687 (145-204), 6094 (34-88) y 5851 (158-209 o 34-86). En resumen, la región 1-144 parece ser muy sensible a tripsina, mientras que la región 145-245 es más estable, si bien se han detectado 4 sitios de corte por tripsina en este fragmento.

En el espectro de MS de la muestra digerida con quimotripsina (10 min de tiempo de incubación) se observa un pico a 25806 Da que corresponde a la región 15-245 y otro a 18899 Da que corresponde a la región 75-245. Además se observan fragmentos de 9842 Da y 9336 Da que se asignaron a las regiones 156-245 y 160-245; así como un pico a 9453 Da que parece corresponder a una especie doblemente cargada del fragmento 75-245. En resumen, el enlace entre F74-M75 parece ser muy sensible a la digestión por quimotripsina, pero además dentro de la región 75-245 la zona 155-160 también es sensible a esta proteasa.

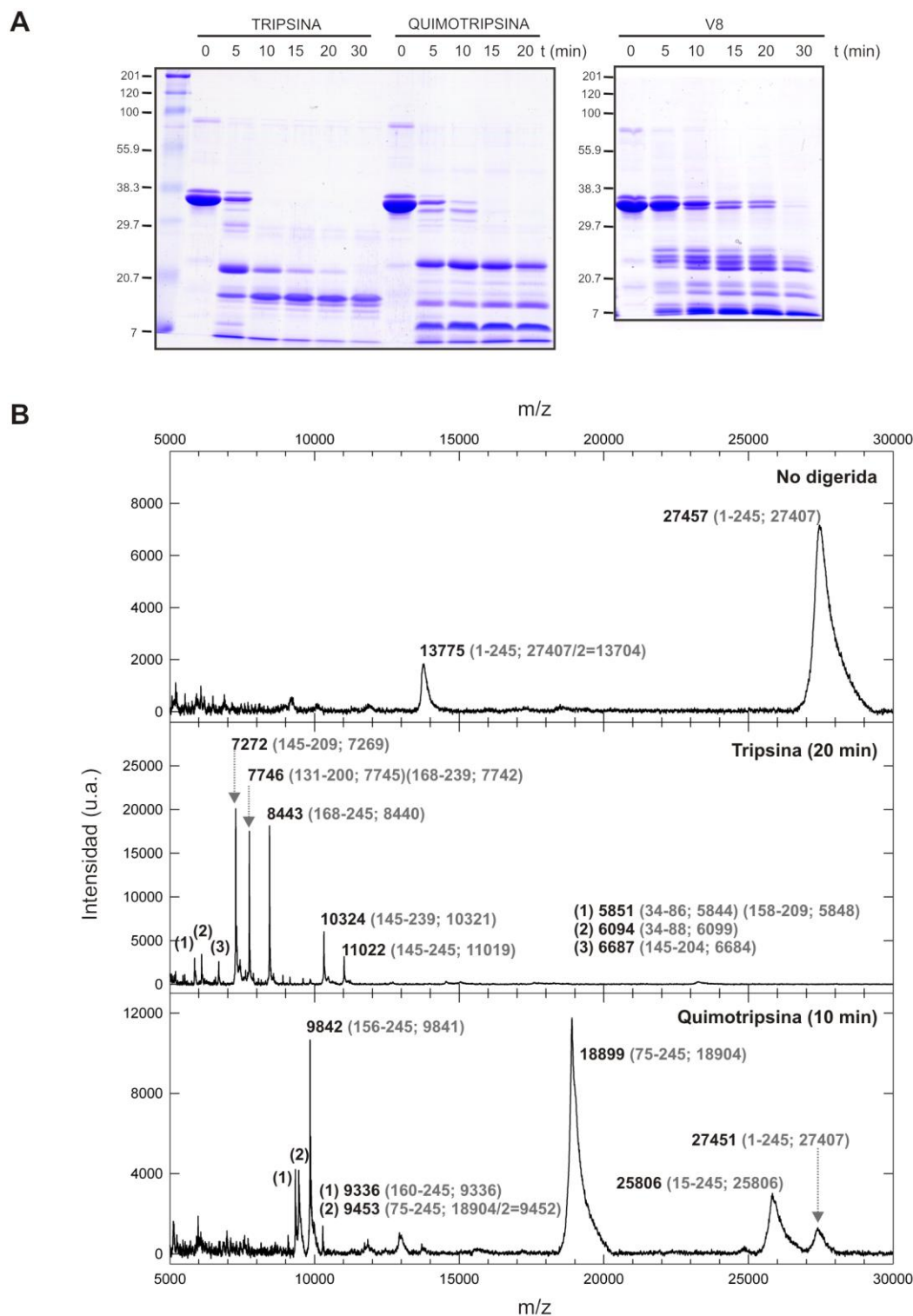


Figura R14. Proteólisis limitada de N-C3G (4-245). (A) Curso de tiempo de la digestión a temperatura ambiente de N-C3G (4-245) (1 mg/mL) en presencia de 0.5 % (p/p) de tripsina, quimotripsina y V8. Se tomaron muestras a diversos tiempos y se analizaron mediante SDS-PAGE. (B) Espectros representativos de N-C3G (4-245) intacto (MM teórica = 27407 Da) y digerido con tripsina (20 min) y quimotripsina (10 min). Se muestran los valores de m/z y junto a ellos se indica entre paréntesis los límites del fragmento correspondiente y su masa teórica.

En la Figura R15-A se muestra un análisis mediante SDS-PAGE de las digestiones con tripsina, quimotripsina o con la proteasa V8 del fragmento N-C3G 65-245. Tripsina produce de forma transitoria y a tiempos de incubación cortos (≤ 10 min) una banda de MMap ligeramente superior a 20 kDa. A tiempos más largos parece acumularse un fragmento de MMap ~ 12 kDa. El análisis mediante MS de los fragmentos obtenidos tras 20 min de incubación con tripsina muestra picos con relaciones m/z de 11030 (asignado al fragmento 145-245), 10330 (145-239), 9607 (97-182), 8446 (168-245), 7276 (145-209) y 4689 (168-209) (Figura R15-B). En resumen, tripsina digiere en múltiples posiciones de la región 97-245.

La digestión del fragmento 65-245 con quimotripsina genera de forma transitoria un fragmento de MMap ligeramente inferior al de la proteína intacta, si bien a tiempos de incubación largos se observan múltiples bandas de menor MMap (Figura R15-A). En el espectro de MS de N-C3G (65-245) digerido durante 15 min con 0.5 % de quimotripsina (Figura R15-B) se observa un pico con valor m/z 15232 que se asignó a la región 108-245, además se detectaron picos a valores de m/z : 9767 (156-245), 9140 (75-155), 7526 (93-159), 5809 (108-159), 5481 (61-107) y 5394 (107-155).

La digestión de N-C3G (65-245) con V8 produce un patrón complejo, con múltiples bandas agrupadas en tres zonas (Figura R15-A), unas con MMap ligeramente superior a 20 kDa, otras con MMap entre 18 kDa y 15 kDa y una banda con MMap < 7 kDa. El análisis mediante MS de los fragmentos producidos por V8 reveló 4 picos en la región entre 18.5 y 20.5 kDa que se asignaron de la siguiente manera: 20482 (61-245), 19916 (61-240), 19045 (74-245 o 69-240) y 18477 (74-240). Además se observaron varios picos de menor MMap a 7161 (fragmento 61-122), 6277 (69-122) y 5720 (74-122); estos fragmentos estarían causados por el procesamiento secundario en E122-M123 de los fragmentos de mayor MMap. También se observó un fragmento de 8275 Da que se asignó a la región 61-133.

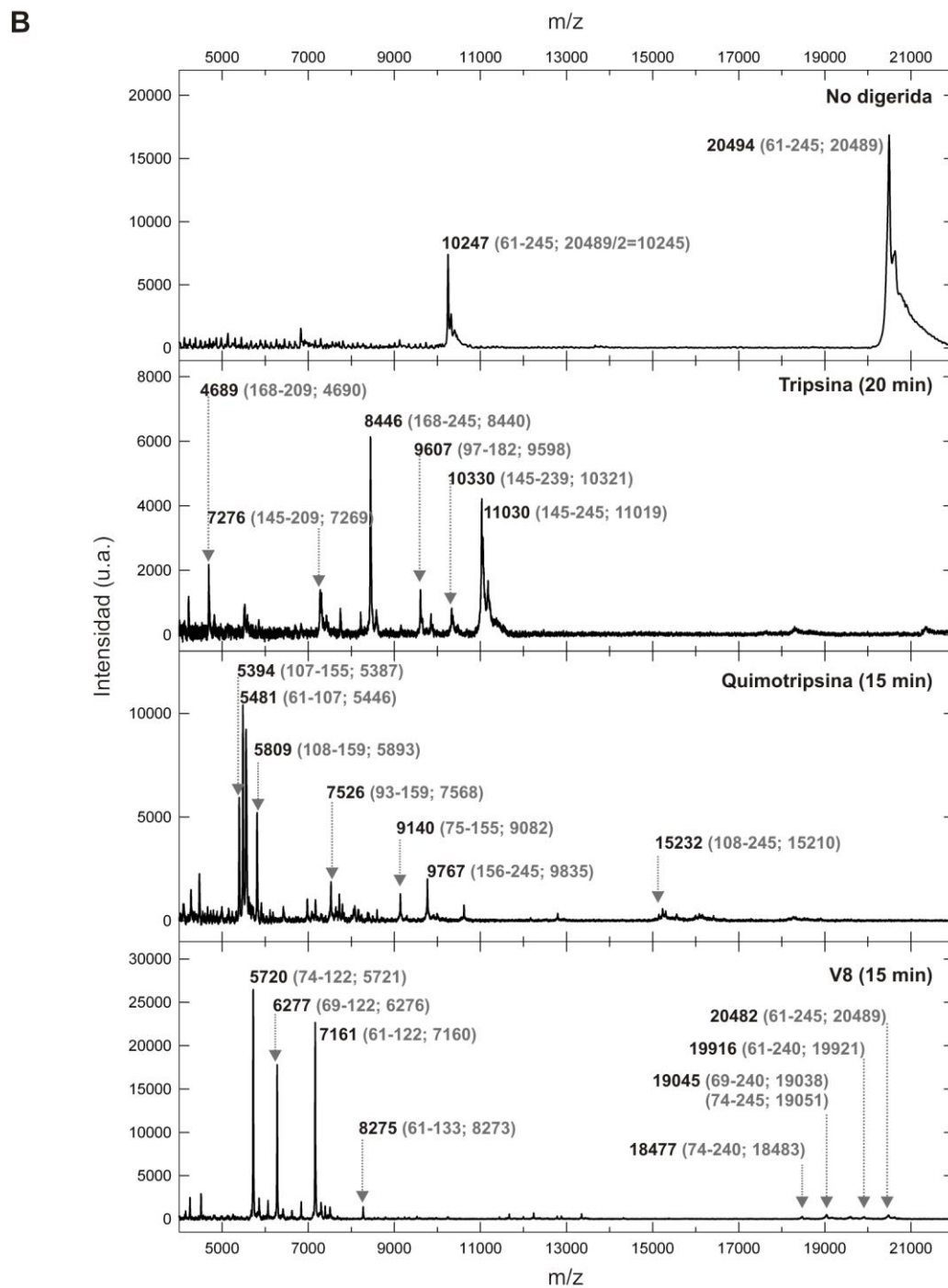
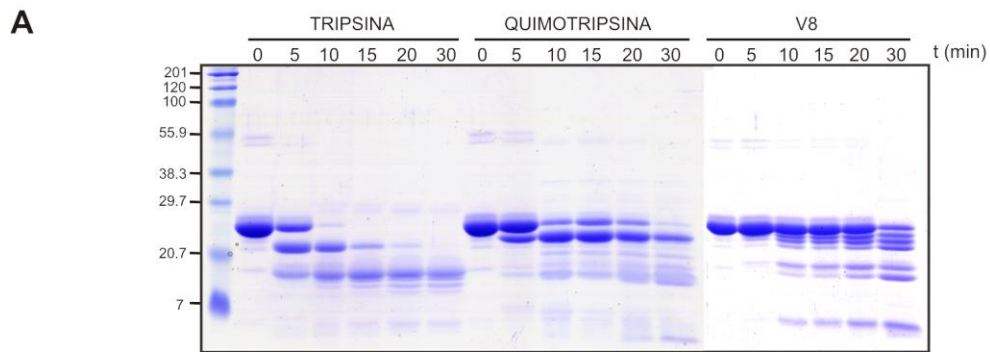


Figura R15. Proteólisis limitada de N-C3G (65-245). (A) Curso de tiempo de la digestión a temperatura ambiente de N-C3G (65-245) (1 mg/mL) en presencia de 0.5 % (p/p) de tripsina, quimotripsina y V8. Se tomaron muestras a diversos tiempos y se analizaron mediante SDS-PAGE. (B) Espectros representativos de N-C3G (65-245) intacto (MM teórica = 20489 Da) y digerido con tripsina (20 min), quimotripsina (15 min) y V8 (15 min). Se muestran los valores de m/z y junto a ellos se indica entre paréntesis los límites del fragmento correspondiente y su masa teórica. En el caso del espectro de quimotripsina las relaciones m/z identificadas presentan una mayor desviación con respecto a la MM teórica que en el resto de los casos.

En la figura R16 se muestra la posición de los sitios de corte por tripsina, quimotripsina y la proteasa V8 en los fragmentos de N-C3G (4-245) y (65-245). De forma individual algunas proteasas muestran preferencia por digerir ambos fragmentos en regiones preferentes, por ejemplo la región 155-160 es sensible a quimotripsina. Sin embargo no se observa un patrón claro de zonas bien definidas que sean particularmente sensibles a varias proteasas. Por otro lado, no se han detectado sitios de digestión en los segmentos centrales de las hélices predichas $\alpha 1$, $\alpha 2$ y $\alpha 4$, con excepción de dos sitios de corte por tripsina (K200-E201 y K204-L204) en $\alpha 4$, si bien éstos se detectaron en fragmentos de menor tamaño en los que la pérdida de otras regiones podría causar una desestabilización de $\alpha 4$. Por otro lado, se han identificado sitios de corte por tripsina y quimotripsina en la hélice $\alpha 3$, lo que sugiere que esta hélice podría corresponder a una región menos rígida que $\alpha 1$, $\alpha 2$ y $\alpha 4$.

En resumen, la región N-terminal de C3G no parece contener ninguna región resistente a proteasas. Esto no implica que esta región no tenga estructura terciaria definida ya que la susceptibilidad a proteasas podría deberse a la presencia de regiones internas flexibles, como por ejemplo lazos que conectan las α -hélices.

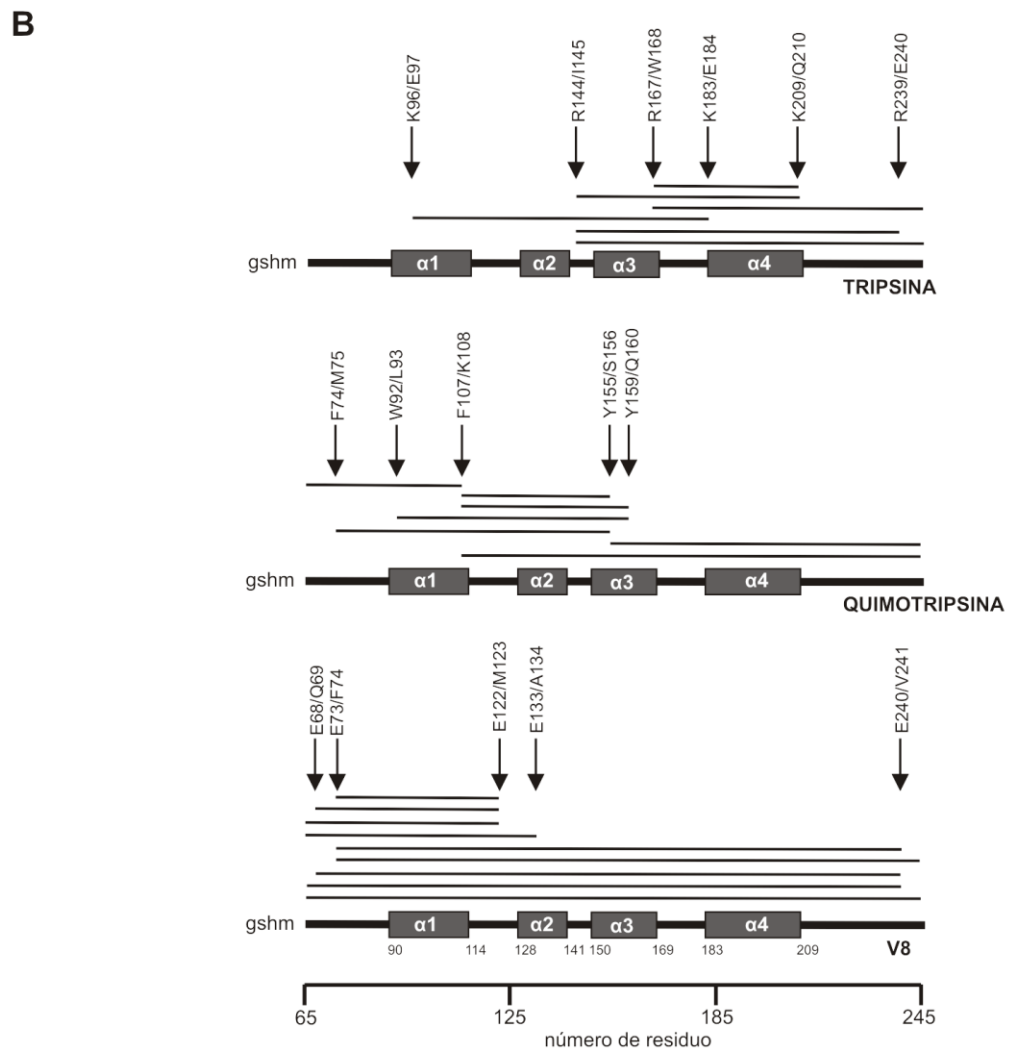
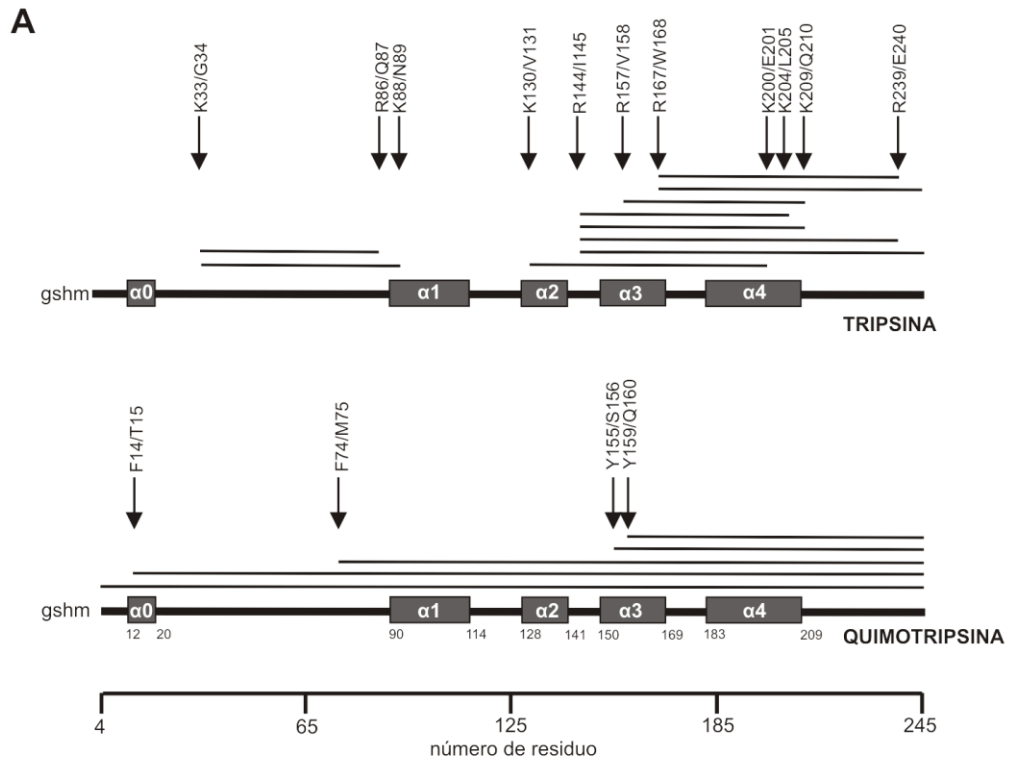


Figura R16. Localización de puntos de digestión en N-C3G. (A) Sitios de digestión en el fragmento N-C3G (4-245) producidos por tripsina (superior) y quimotripsina (inferior). La estructura lineal de esta región se muestra en la parte inferior y en ella se indican las α -hélices predichas por el programa PHD. Los fragmentos detectados mediante espectrometría de masas se muestran como líneas horizontales. Los puntos de corte deducidos a partir de los mapas peptídicos se indican por flechas verticales sobre las que se indican los pares de residuos entre los que se produce la digestión. (B) Sitios de digestión en N-C3G (65-245) producidos por tripsina (superior), quimotripsina (intermedio) y la proteasa V8 (inferior).

2.3. Homo-asociación de N-C3G

Durante la purificación del fragmento N-C3G (65-245) expresado en *E. coli*, en la etapa final se realizó una cromatografía de exclusión molecular en escala preparativa, en la que se observó que el volumen de elución de esta proteína correspondía a una MMap muy superior a la masa del monómero. Dada la participación de la región N-C3G en la interacción cabeza-cola con el dominio REM y su posible implicación en la regulación de C3G, analizamos en detalle la posible homo-asociación de esta región amino terminal.

Inicialmente se analizó el fragmento N-C3G (65-245) mediante cromatografía de exclusión molecular empleando una columna analítica Superdex 200 HR 10/300 (GE Healthcare) equilibrada en tampón 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT, inyectando muestras en un rango de concentraciones de proteína entre 10 μ M y 500 μ M (Figura R17). Cuando se inyectó una solución 10 μ M de N-C3G (65-245) el perfil de elución mostró un pico mayoritario a 15.6 mL que corresponde con una MMap de \sim 38 kDa, un segundo pico a 13.9 mL y picos minoritarios de elución a volúmenes menores. Cuando se cargaron soluciones más concentradas se observó un desplazamiento del pico mayoritario hacia volúmenes de elución menores, hasta 14.6 mL para la muestra cargada a 500 μ M. Sin embargo, la posición del segundo pico permaneció constante (13.91 ± 0.03 mL); esto sugiere que este pico corresponde a una especie definida, la MMap estimada para este pico fue de \sim 84 kDa por lo que podría corresponder a un tetrámero.

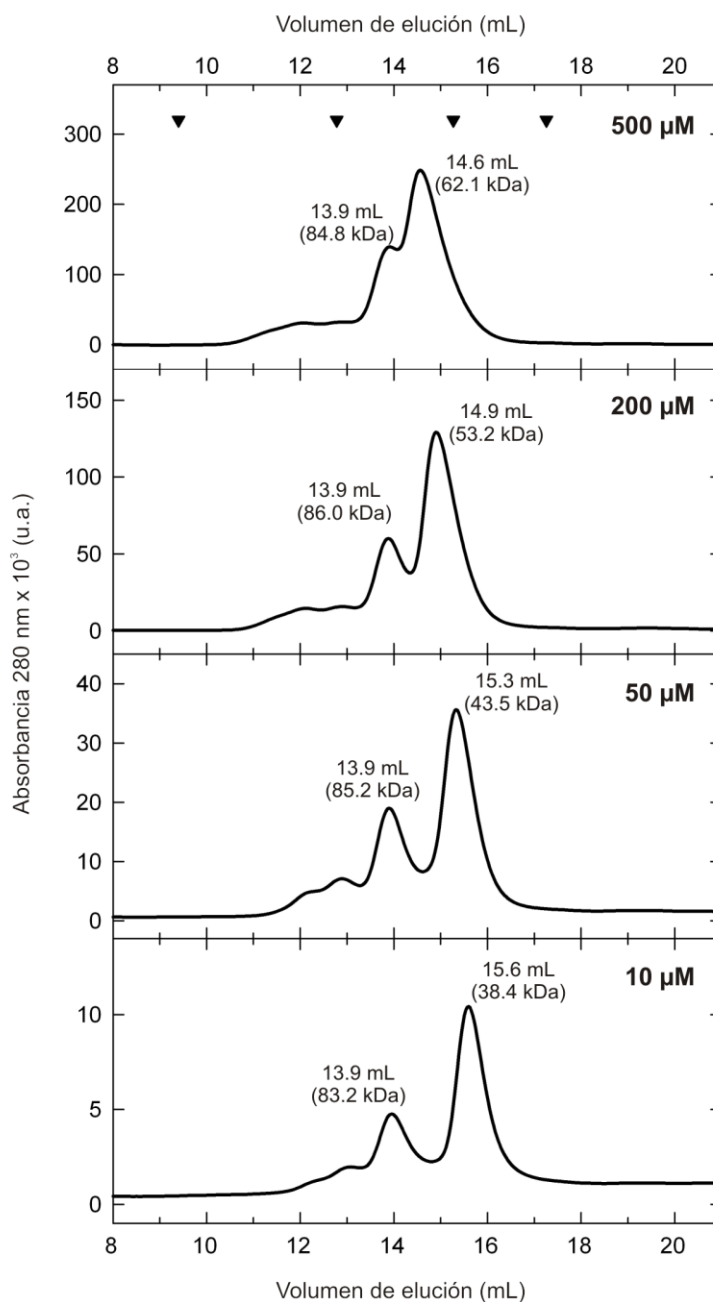


Figura R17. Análisis de N-C3G (65-245) mediante cromatografía de exclusión molecular. Perfiles de elución en una columna Superdex 200 HR 10/300. Se inyectaron 100 μ L de soluciones de proteína entre 10 μ M y 500 μ M. La columna se equilibró en 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT; el flujo fue de 0,5 mL/min. En los picos se indica el volumen medio de elución y la MMap estimada (entre paréntesis). La posición de elución de los patrones de peso molecular se indican por triángulos invertidos en la parte superior, como proteínas patrón se emplearon: tiroglobulina (670 kDa), γ -globulina (158 kDa), ovalbúmina (44 kDa) y mioglobina (17 kDa).

Es decir, en el rango de concentraciones analizadas, el radio de Stokes aumentó de manera significativa para el pico mayoritario (Figura R18), indicando

que el tamaño promedio de las especies aumenta con la concentración; mientras que para el segundo pico el radio de Stokes se mantuvo sin variaciones.

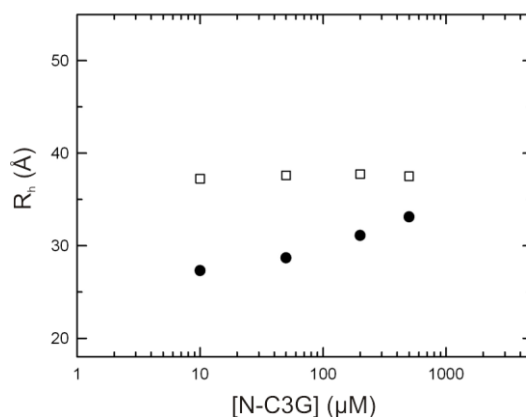


Figura R18. Análisis del radio de Stokes de N-C3G (65-245) en función de la concentración. Variación del radio de Stokes en el rango de concentración entre 10 μM y 500 μM. Los puntos negros corresponden al pico mayoritario de la cromatografía de exclusión molecular y los cuadrados vacíos al pico minoritario.

La región N-C3G (65-245) contiene una Cys (C154). Con el fin de evitar que la formación de dímeros covalentes por formación de un puente disulfuro intramolecular pudiera contribuir al proceso de homo-asociación, se analizó el mutante libre de Cys, C154A. Se inyectaron soluciones de proteína entre 0.5 μM y 1 mM y se analizaron los perfiles de elución (Figura R19). Cuando se cargaron soluciones de N-C3G-C154A (65-245) entre 0.5 y 2 μM la proteína eluyó en un único pico. A concentraciones de carga entre 5 y 20 μM se observó la aparición de un segundo pico minoritario a volúmenes de elución menores. Cuando se inyectaron concentraciones ≥ 50 μM se observó un tercer pico minoritario que precede a los dos anteriores. De forma análoga a lo observado con la proteína silvestre (Figura R17), la posición del pico mayoritario se fue desplazando desde 16.9 mL para una concentración de carga 0.5 μM hasta 14.6 mL cuando se inyectó N-C3G-C154A 65-245 a 1 mM. En cambio el segundo pico eluyó en un volumen prácticamente constante (13.94 ± 0.03 mL) para concentraciones de carga entre 20 μM y 1 mM. Por último, el pico minoritario de elución que precede a los anteriores se desplazó desde 12.7 mL a 12.3 mL para un rango de concentraciones de inyección entre 50 μM y 1 mM. La proporción del área del pico de elución en ~ 13.9 mL con respecto al pico mayoritario es claramente menor en el mutante C154A

que en el fragmento silvestre, lo que sugiere que C154 podría favorecer la estabilización del oligómero que eluye en esa posición y que podría corresponder a un tetrámero.

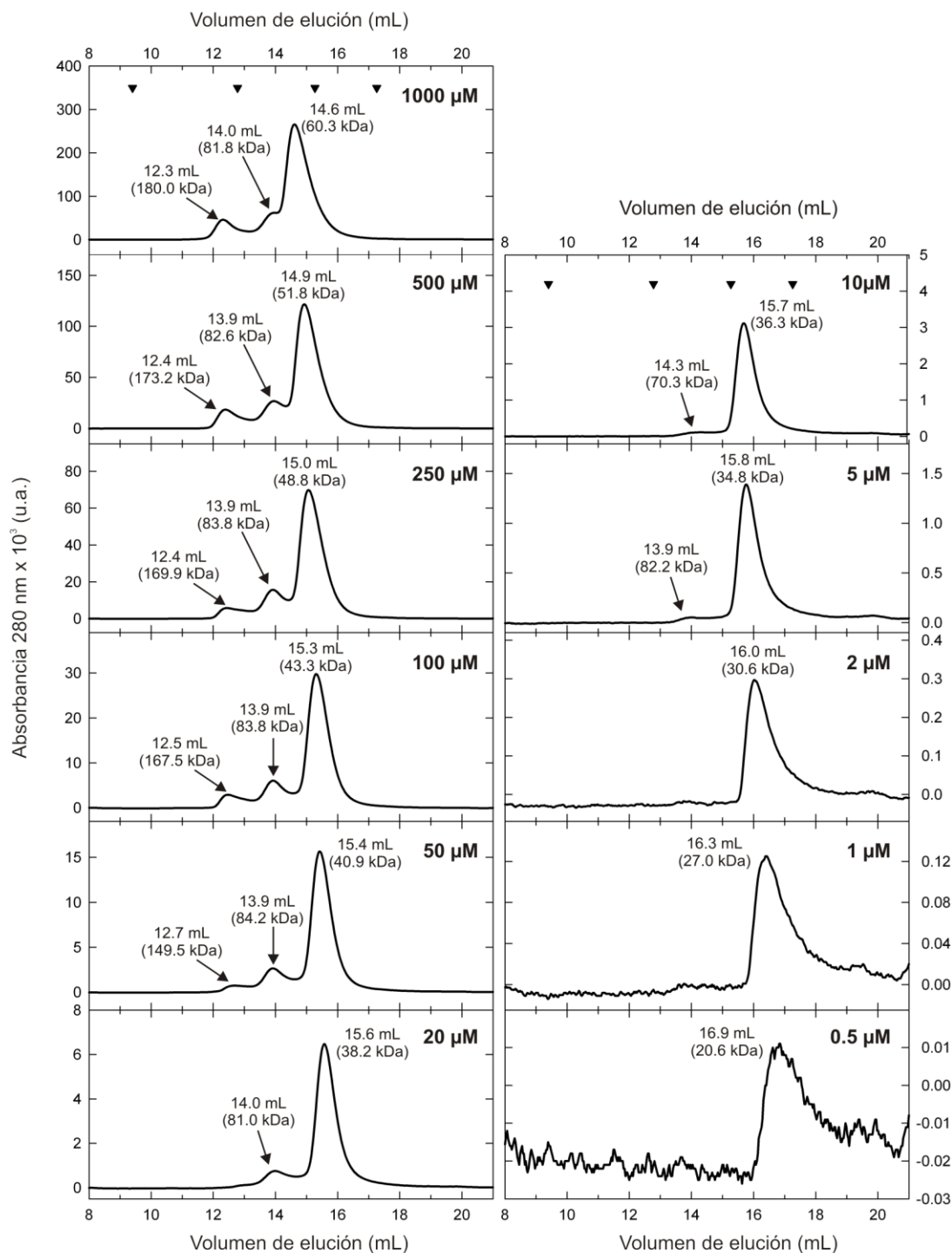


Figura R19. Análisis de N-C3G-C154A (65-245) mediante cromatografía de exclusión molecular. Perfiles de elución en una columna Superdex 200 HR 10/300. Se inyectaron 100 μL de soluciones de proteína entre 0.5 μM y 1 mM. La columna se equilibró en 20 mM Tris-HCl (pH 7.5), 150 mM NaCl; el flujo fue de 0,5 mL/min.

En el caso del mutante, N-C3G-C154A, el radio de Stokes aumentó de manera significativa para el pico mayoritario (Figura R20) y para el pico con menor volumen de elución, lo que indica que en estos casos el tamaño promedio de las especies aumenta con la concentración; mientras que para el pico con volumen de elución intermedio el radio de Stokes se mantuvo sin variaciones.

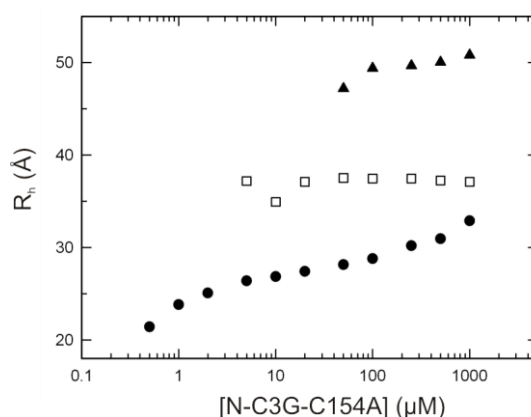


Figura R20. Análisis del radio de Stokes de N-C3G-C154A (65-245) en función de la concentración. Variación del radio de Stokes en el rango de concentración entre 0.5 μM y 1 mM. Los puntos negros corresponden al pico que presenta mayor volumen de elución en la cromatografía de exclusión molecular; los cuadrados vacíos al pico con volumen de elución intermedio; y los triángulos negros al pico con menor volumen de elución en la cromatografía de exclusión molecular.

En conjunto el análisis mediante cromatografía de exclusión molecular sugirió que N-C3G sufre un proceso de homo-asociación con múltiples especies definidas que podrían corresponder a oligómeros con simetría cerrada.

Con el fin de caracterizar en más detalle el grado de asociación de N-C3G en solución se realizaron experimentos de ultracentrifugación analítica, tanto en equilibrio de sedimentación como en velocidad de sedimentación, en un rango de concentración desde 0.25 hasta 2.0 mg/mL, que corresponde a 12-98 μM. Los experimentos de velocidad de sedimentación muestran la presencia de dos especies (Figura R21): una mayoritaria con un coeficiente de sedimentación entre 2,3-2,5 S y otra minoritaria de 3,6-4,2 S. La proporción de la especie de mayor coeficiente de sedimentación aumenta a mayor concentración, desde ~ 14 % a 12

μM hasta $\sim 22\%$ a $98\ \mu\text{M}$. Esto sugiere que en este rango de concentraciones N-C3G sufre un proceso de homo-asociación entre dos estados.

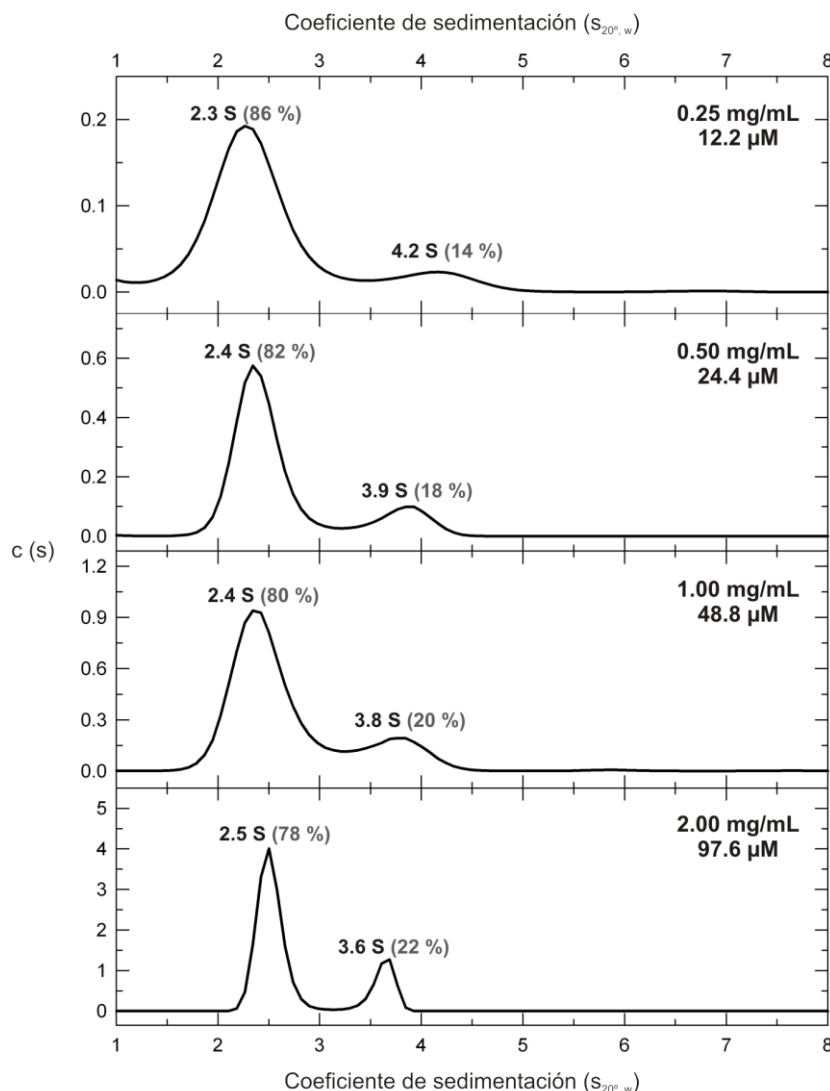


Figura R21. Análisis de N-C3G (65-245) mediante velocidad de sedimentación. Distribución de los coeficientes de sedimentación de N-C3G (65-245) en 20 mM NaPi (pH 7.5), 150 mM NaCl a concentraciones iniciales de 0.25, 0.50, 1.00 y 2.00 mg/mL (que corresponden a 12.2, 24.2, 48.4 y 97.6 μM). Se representa la concentración de proteína en función del coeficiente de sedimentación aparente corregido en agua a 20 °C ($s_{20,w}$) obtenido de la integración de las curvas de velocidad de sedimentación obtenidas a 45000 rpm y 20 °C. Se indica el coeficiente de sedimentación de cada pico y entre paréntesis su abundancia relativa.

Seguidamente se analizó el grado de asociación de N-C3G (65-245) en solución mediante equilibrio de sedimentación. Se analizaron gradientes de equilibrio de sedimentación de N-C3G (65-245) a 20 °C para 4 concentraciones

iniciales entre 12-98 μM a 3 velocidades. Inicialmente se analizaron los gradientes empleando un modelo de una única especie ideal (Figura R22).

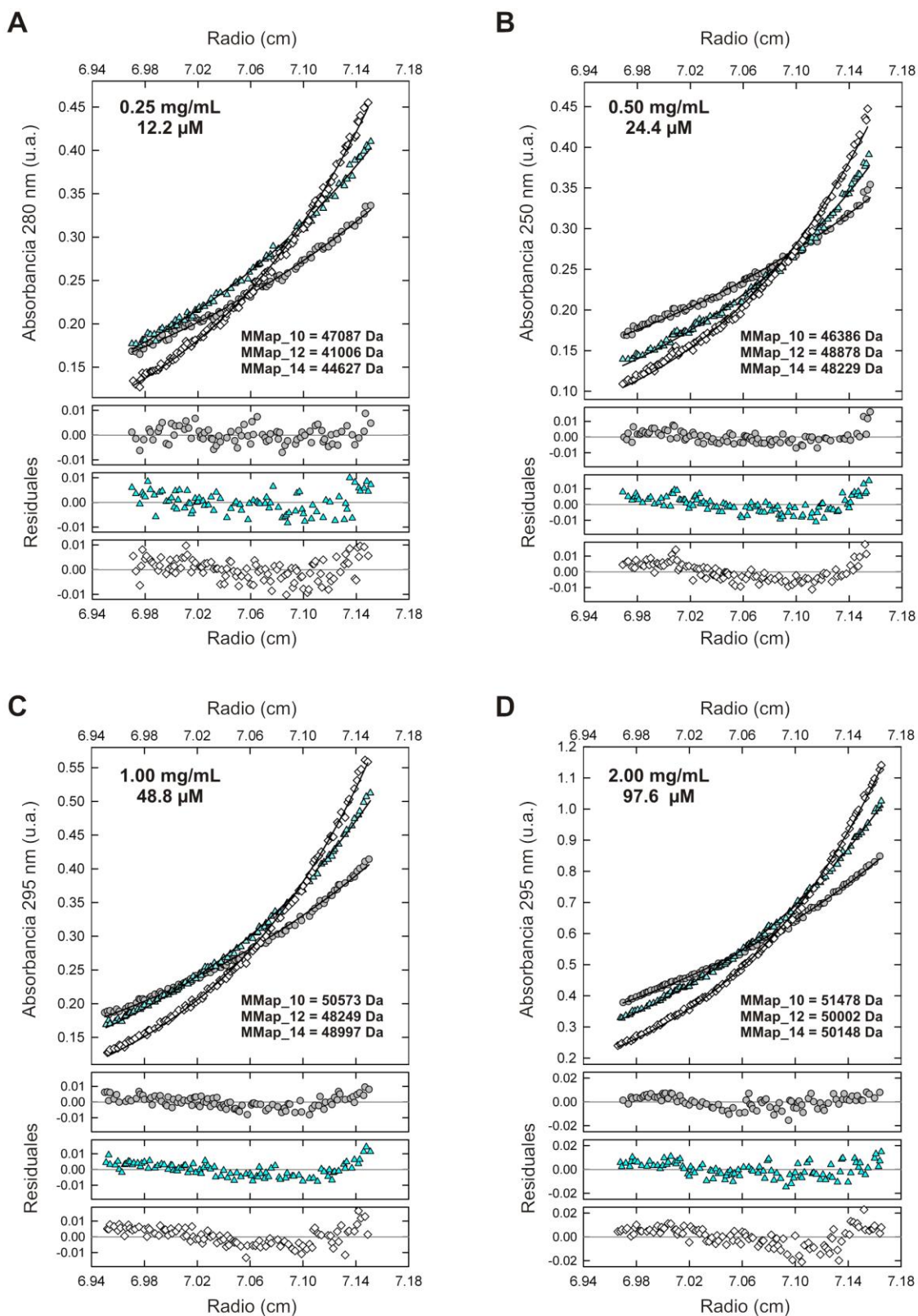


Figura R22. Análisis de N-C3G (65-245) mediante equilibrio de sedimentación, ajuste de modelos con una única especie. Gradientes de equilibrio de sedimentación de N-C3G (65-245) en 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, a 20 °C obtenidos a 10000 rpm (círculos), 12000 rpm (triángulos) y 14000 rpm (rombos). Se representa la absorbancia de N-C3G (65-245) en función de la distancia radial. Las líneas representan el ajuste usando un modelo de una única especie. En cada panel se indica la concentración inicial así como las MMap estimadas a partir de los ajustes a las diferentes velocidades. Bajo los gráficos con los gradientes se muestran los residuales del ajuste del modelo a los datos a cada velocidad.

La masa molecular promedio estimada a partir de los gradientes obtenidos a 10000 rpm correspondió aproximadamente a la masa de un dímero de N-C3G (MM teórica monómero 20.5 kDa) para la muestra a 12 μM , y aumentó con la concentración de proteína (Figura R23). Por otro lado, los residuales del ajuste del modelo de una única especie mostraron desviaciones sistemáticas que eran más acentuadas a alta velocidad, lo que sugería la presencia de múltiples especies.

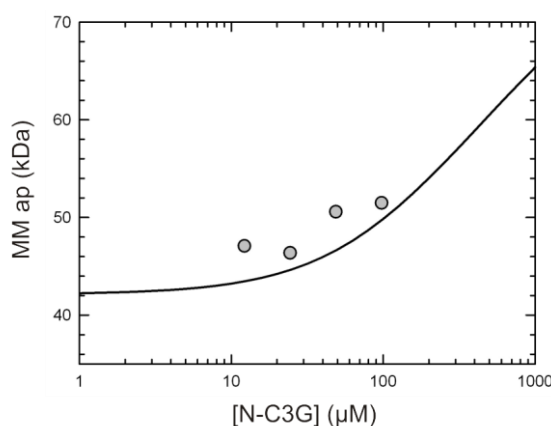


Figura R23. Dependencia de la masa molecular promedio de N-C3G (65-245) en función de la concentración de proteína. Se representan los valores de MMap promedio estimados ajustando un modelo de una única especie a los gradientes de equilibrio de sedimentación obtenidos a 10000 rpm y 20 °C. La línea de trazo continuo corresponde a un modelo de equilibrio dímero-tetrámero (MMap del dímero 42.1 kDa y Kd 0.73 mM).

En conjunto, este primer análisis de los datos de equilibrio de sedimentación sugirió la existencia de un proceso de oligomerización en el que se producía un enriquecimiento de especies de mayor tamaño a alta concentración. Puesto que los experimentos de velocidad de sedimentación habían indicado la presencia de dos especies mayoritarias se empleó un modelo de homo-asociación dímero-tetrámero para el análisis de los datos de equilibrio de sedimentación (Figura R24). El ajuste simultáneo de los 12 gradientes a dicho modelo resultó en

un mejor ajuste a los datos experimentales, que correspondió con una masa molecular del dímero de 42.1 kDa y una Kd del equilibrio dímero-tetrámero de ~ 0.73 mM. Este valor de Kd ha de considerarse solamente como una estimación aproximada ya que en el rango de concentraciones analizadas ($\leq 98 \mu\text{M}$) la fracción de tetrámero sería inferior al 20 %.

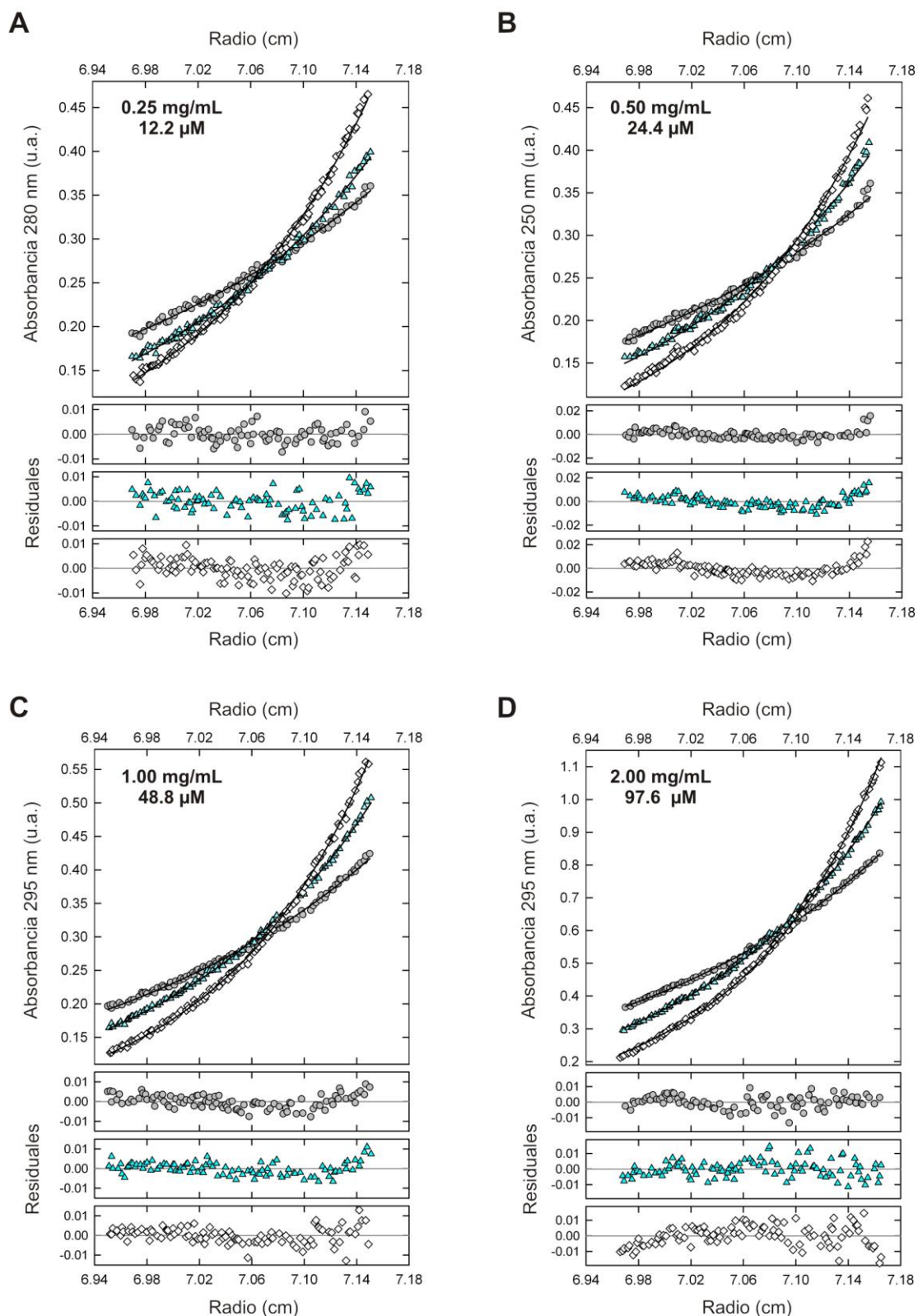


Figura R24. Análisis de N-C3G (65-245) mediante equilibrio de sedimentación, ajuste de un modelo de equilibrio dímero-tetrámero. Gradientes de equilibrio de sedimentación de N-C3G (65-245) en 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, a 20 °C obtenidos a 10000 rpm (círculos), 12000 rpm (triángulos) y 14000 rpm (rombos). Se representa la absorbancia de N-C3G (65-245) en función de la distancia radial. Las líneas representan el ajuste usando un modelo homoasociación dímero-tetrámero. Bajo los gráficos con los gradientes se muestran los residuales del ajuste del modelo a los datos a cada velocidad.

Con el fin de explorar la homo-asociación de N-C3G en un rango de concentración de proteína más amplio al analizado mediante ultracentrifugación analítica se aplicó un método basado en anisotropía de fluorescencia. Para estos experimentos se empleó N-C3G (65-245) marcado selectivamente con la sonda fluorescente *Oregon Green 488* en la Cys154 (única Cys en esta región). En un experimento típico se monitorizó la anisotropía de la sonda en presencia de concentraciones crecientes de N-C3G no marcado (Figura R25). La anisotropía de fluorescencia está relacionada con el coeficiente de difusión rotacional, que a su vez depende del tamaño y de la forma de las especies, y por lo tanto es sensible a cambios en el estado de asociación de las proteínas. Se observó un incremento de la anisotropía a concentraciones de N-C3G (65-245) superiores a 10 μM lo cual se interpreta como la formación de oligómeros. No se observó saturación del valor de anisotropía incluso a 1 mM de N-C3G, lo cual es compatible con una baja afinidad para los equilibrios de formación de especies de mayor tamaño (como posibles octámeros). Por otro lado, el valor de anisotropía fue aproximadamente constante en un rango de concentración total de proteína entre 20 nM y 5 μM lo que sugirió que no se produciría una disociación del dímero en ese rango.

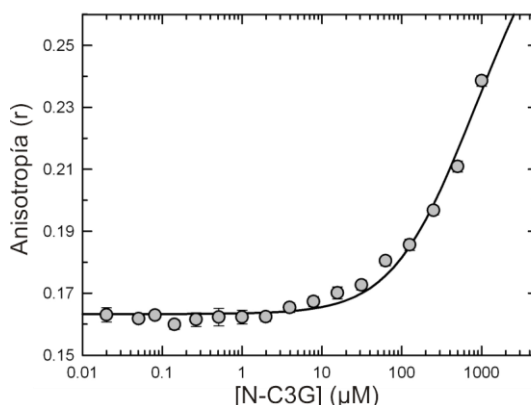


Figura R25. Análisis de la homo-asociación de N-C3G (65-245) mediante anisotropía de fluorescencia extrínseca. Se representa la anisotropía de fluorescencia a 25 $^{\circ}\text{C}$ de soluciones 20 nM de N-C3G-Oregon Green en 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mg/mL BSA en presencia de diversas concentraciones de N-C3G (65-245) sin marcar. La longitud de onda de excitación fue 496 nm y la de emisión 520 nm. La línea de trazo continuo corresponde a un modelo de equilibrio 1n:2n (K_d 1.4 mM) ajustado a los datos experimentales.

También se analizó la posible homo-asociación de la región N-C3G mediante coIP. Para ello combinamos el fragmento N-C3G (4-245) marcado con la etiqueta

FLAG, con ese mismo fragmento marcado con la etiqueta HA. No se detectó homo-asociación del dominio N-C3G en estos ensayos (Figura R26). Tampoco se detectó homo-asociación de un fragmento de la región amino, N-C3G (65-245), que carece de los primeros 64 aminoácidos.

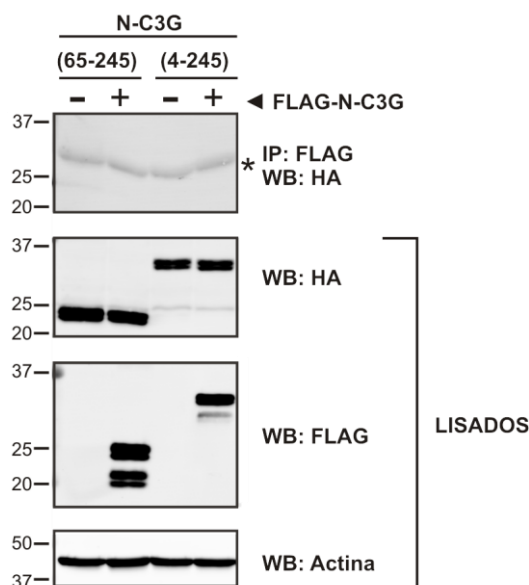


Figura R26. Análisis de la homo-asociación de N-C3G mediante coIP. Ensayo de coIP donde se muestra que los fragmentos FLAG-N-C3G (65-245) y FLAG-N-C3G (4-245) no interactúan con HA-N-C3G (65-245) y HA-N-C3G (4-245) respectivamente. Para cada fragmento de HA-C3G se analizaron dos co-transfecciones: una con el vector pCEFLAG vacío (-) y otra con pCEFLAG-N-C3G (+). A la izquierda de cada inmunoblot (WB) se muestra la posición y la masa molecular de los marcadores. En el inmunoblot de la coIP se observa una banda inespecífica con una masa molecular ~ 25 kDa que se indica con un asterisco.

2.4. Análisis de la estructura de N-C3G en solución mediante dispersión de rayos-X a bajo ángulo (SAXS)

De forma complementaria a los estudios biofísicos anteriores, el fragmento N-C3G (65-245) se analizó mediante dispersión de rayos-X a bajo ángulo (SAXS). Esta técnica permite analizar macromoléculas en un amplio rango de concentraciones proporcionando información sobre el tamaño, es decir, el grado de asociación y la forma de macromoléculas en solución. Además, a partir de los datos de SAXS es posible reconstruir a baja resolución la estructura de macromoléculas. El estudio de N-C3G (65-245) mediante SAXS fue particularmente

indicado debido a la falta de éxito en la obtención de cristales que permitieran elucidar la estructura cristalográfica de C3G.

Se midieron los perfiles de dispersión de N-C3G (65-245) a 9 concentraciones de proteína entre 0.25 y 64 mg/mL (de $\sim 12 \mu\text{M}$ a $\sim 3 \text{mM}$) (Figura R27-A). Mediante análisis de Guinier, en el que se emplean datos de dispersión a muy bajo ángulo, se determinó el radio de giro (R_g) (Figura R27-B) y la intensidad de dispersión a ángulo cero ($I(0)$) (Figura R27-C) para cada concentración de proteína. Se observó que el valor de R_g era aproximadamente constante en el rango entre $\sim 10 \mu\text{M}$ y $\sim 100 \mu\text{M}$ ($R_g \sim 34 \text{Å}$). Sin embargo, a concentraciones superiores a $\sim 100 \mu\text{M}$ el R_g aumentó significativamente. De forma similar, la $I(0)$ era aproximadamente constante hasta $100 \mu\text{M}$, mientras que por encima de $100 \mu\text{M}$ la $I(0)$ aumentó progresivamente con la concentración de proteína. En conjunto, la variación de R_g e $I(0)$ con la concentración de proteína sugirió la presencia significativa de especies de mayor tamaño a concentraciones $> 100 \mu\text{M}$. A partir de los datos de $I(0)$ frente a concentración se ajustó a un modelo de equilibrio $1n:2n$, estimándose una K_d aparente de $1.0 \pm 0.4 \text{mM}$, que era similar a la estimada en el análisis de los datos de equilibrio de sedimentación y de anisotropía de fluorescencia.

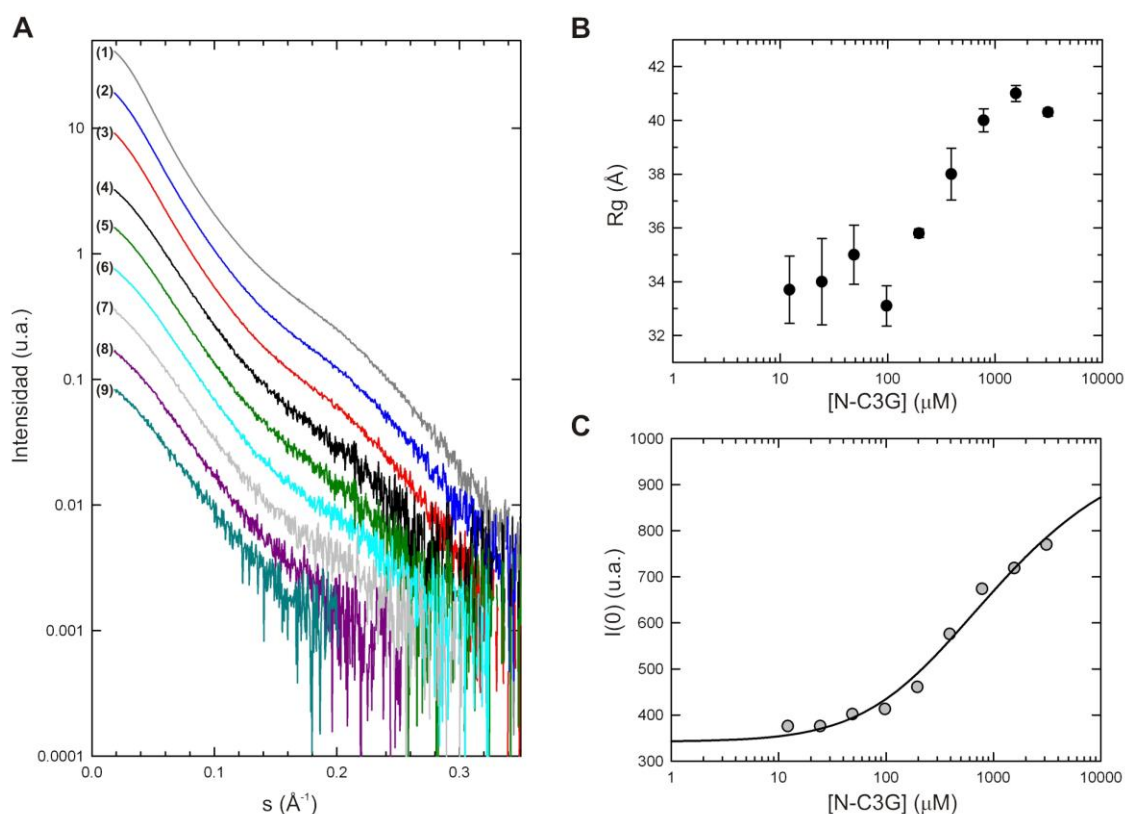


Figura R27. Análisis de N-C3G (65-245) mediante SAXS. (A) Perfiles de dispersión de N-C3G a 10 °C en 20 mM NaPi (pH 7.5), 150 mM NaCl, 5 % (v/v) glicerol, 3 mM DTT determinados a 9 concentraciones diferentes de proteína: 64 mg/mL (1), 32 mg/mL (2), 16 mg/mL (3), 8 mg/mL (4), 4 mg/mL (5), 2 mg/mL (6), 1 mg/mL (7), 0.50 mg/mL (8) y 0.25 mg/mL (9). (B) Dependencia del R_g , estimado mediante análisis de Guinier de los datos de dispersión, en función de la concentración de proteína. (C) Intensidad de la dispersión a ángulo cero, $I(0)$, en función de la concentración de N-C3G (65-245). La línea corresponde a un modelo de equilibrio $1n:2n$ ajustado a los datos experimentales ($K_d 1.0 \pm 0.4$ mM) suponiendo la ausencia de interacciones interpartículas significativas.

Con el fin de estimar el número de formas oligoméricas de N-C3G (65-245) se analizó el conjunto de las 9 curvas de SAXS a diferentes concentraciones de proteína mediante descomposición en valores singulares (SVD). La descomposición de las 9 curvas en otros tantos componentes reveló que los valores propios de los dos primeros componentes eran significativamente superiores a los del resto (Figura R28-A). De forma paralela, excepto para los dos primeros componentes, el resto de los vectores básicos contenían esencialmente ruido (Figura R28-B). En conjunto, el análisis mediante SVD indicó que las curvas de dispersión contenían dos componentes principales, que sugirió la contribución significativa de dos especies oligoméricas en el rango de concentraciones analizado.

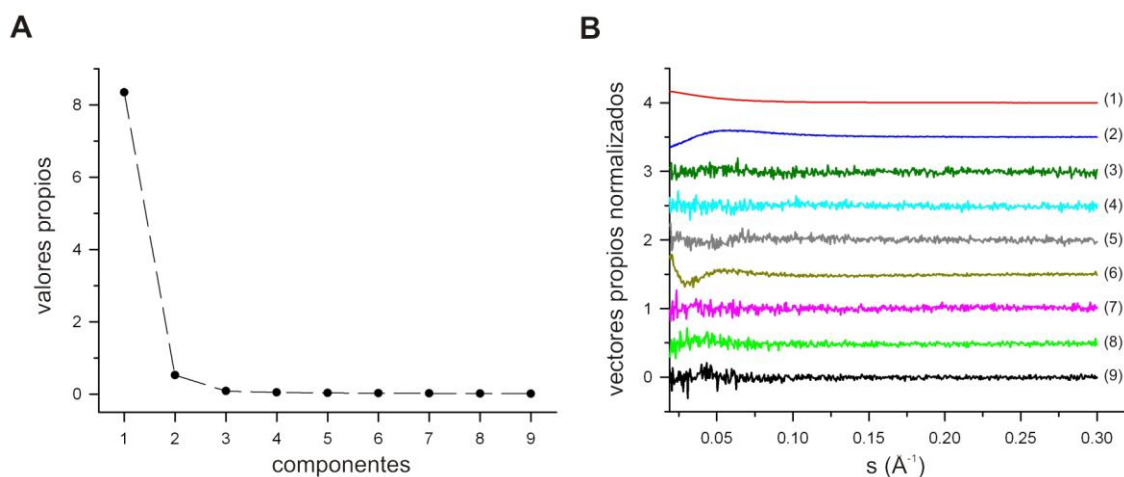


Figura R28. Análisis de descomposición en valores singulares (SVD) de los perfiles de dispersión de N-C3G (65-245). (A) Gráfico de valores propios como función del número ordinal de componente. El número de valores propios significativos corresponde con el número de especies que contribuyen de manera significativa a los datos de la dispersión. (B) Gráfico de los 9 vectores propios normalizados por el valor propio correspondiente.

Con el fin de obtener información sobre la estructura de N-C3G (65-245) de los datos de SAXS, se emplearon los datos de dispersión obtenidos para las cuatro concentraciones más bajas (entre 2.00 y 0.25 mg/mL), ya que en este rango de concentración se observó una mínima variación del tamaño de las partículas, estimado por el R_g y la $I(0)$, con la concentración de proteína. A partir de esas cuatro curvas de dispersión se estimó el perfil de dispersión de N-C3G (65-245) a dilución infinita (Figura R29-A). El R_g calculado mediante análisis de Guinier empleando datos en el rango $0.64 < sR_g < 1.20$ fue de $33.9 \pm 1.4 \text{ \AA}$. La representación de Kratky de los datos de dispersión a dilución infinita se aproximó a cero a valores altos del vector de dispersión, s (Figura R29-B), lo cual sugirió que N-C3G tiene una estructura mayoritariamente compacta sin presencia significativa de regiones desordenadas. Sin embargo, este gráfico de Kratky presenta una serie de limitaciones, como es la dificultad de extraer información sobre proteínas parcialmente desplegadas pero que sin embargo presentan cierto grado de compactación debido a la presencia de regiones con estructura de tamaño significativo. Para obviar estos problemas, se empleó la representación de Kratky adimensional (Durand et al., 2010), donde al dividir $I(s)$ entre $I(0)$ se convierte en independiente de la masa molecular de la proteína; y al multiplicar s por R_g la escala angular es independiente del tamaño. El patrón de dispersión de una proteína globular obedece la ley de Guinier, de modo que un gráfico de Kratky

adimensional muestra un valor máximo de 1.104 para $sRg = \sqrt{3}$ sea cual sea el tamaño de la proteína. En el gráfico de Kratky adimensional para N-C3G (65-245) (Figura R29-C) este máximo se encontró ligeramente desplazado hacia valores mayores de las coordenadas, lo que se identificó con una estructura compacta y con una flexibilidad muy moderada.

Empleando datos de SAXS en un rango entre 0.02 y 0.30 \AA^{-1} se calculó la función de distribución de distancias, $P(r)$ (Figura R29-D), que reveló un Rg de 35 \AA y una dimensión máxima (D_{max}) de 140 \AA . La forma de la $P(r)$ es ligeramente asimétrica con una pequeña cola hacia valores de distancia más largos, lo que es característico de estructuras moderadamente alargadas.

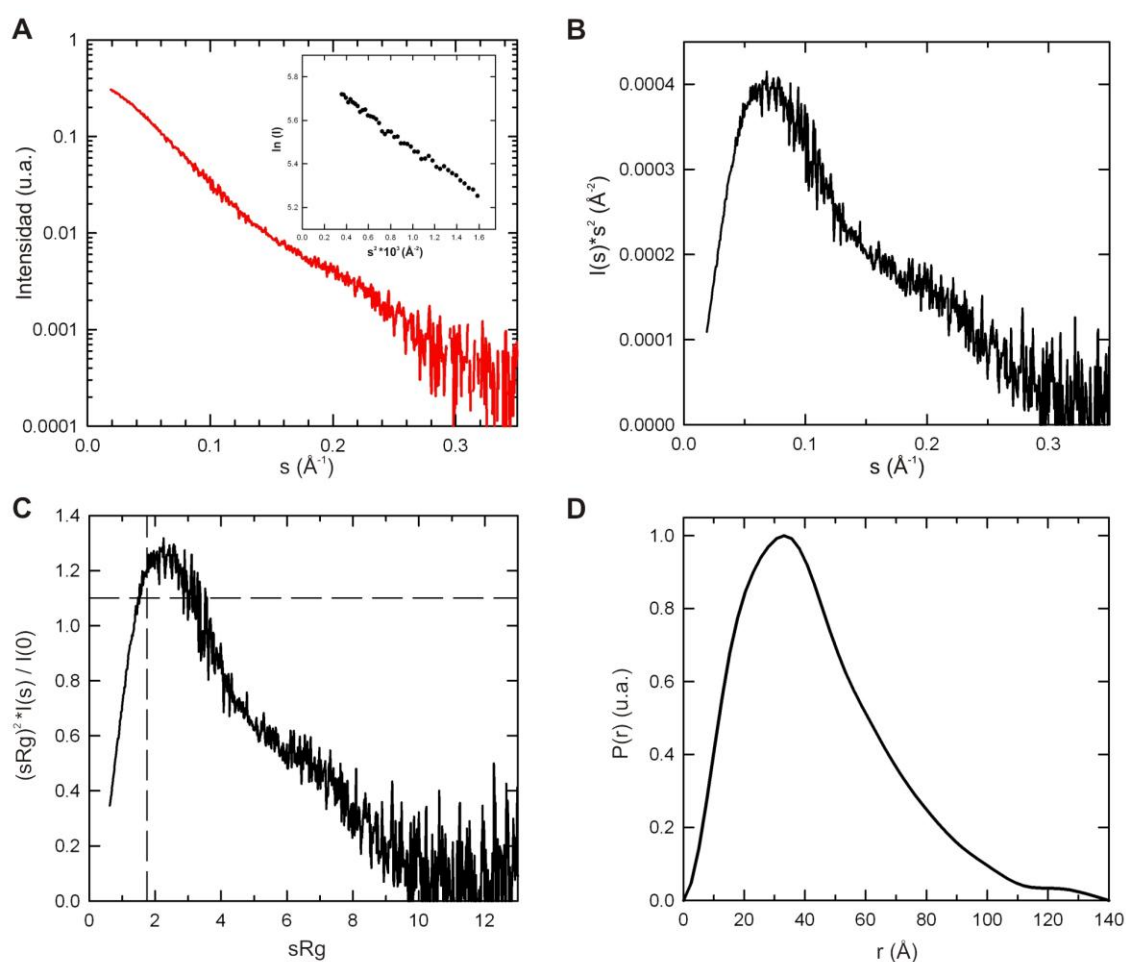


Figura R29. Análisis de N-C3G (65-245) mediante SAXS. (A) Perfil de dispersión experimental extrapolado a dilución infinita. (B) Gráfico de Kratky calculado a partir de los datos de SAXS. (C) Gráfico de Kratky adimensional. (D) Función de distribución de distancias, $P(r)$, que representa el histograma de las distancias dentro de la molécula.

A partir del perfil de SAXS a dilución infinita se reconstruyeron estructuras de N-C3G (65-245) a baja resolución mediante métodos *ab initio* empleando el programa DAMMIF. Inicialmente se realizaron 15 reconstrucciones independientes de la estructura sin imponer simetría interna (simetría P1). Las estructuras se superpusieron y se calculó la discrepancia espacial normalizada (NSD), que es un parámetro que da idea de la similitud entre estructuras (siendo 0 para estructuras idénticas). La NSD del conjunto de los modelos fue de 0.58 ± 0.02 , lo cual indica que las estructuras fueron muy similares entre sí y que la reconstrucción es reproducible. A partir de las estructuras superpuestas se calculó una estructura promedio que representa el volumen más probable (Figura R30-A). La estructura tiene forma alargada y aplanada con unas dimensiones aproximadas de $133 \text{ \AA} \times 48 \text{ \AA} \times 35 \text{ \AA}$, que presenta una región central ancha y prolongaciones estrechas en los extremos del eje longitudinal. La estructura presentó una aparente simetría interna con un eje binario. Ante la aparente presencia de simetría interna se realizaron nuevas reconstrucciones *ab initio* imponiendo simetría P2 a las partículas (Figura R30-B), que dieron lugar a estructuras muy similares a las modeladas sin simetría interna.

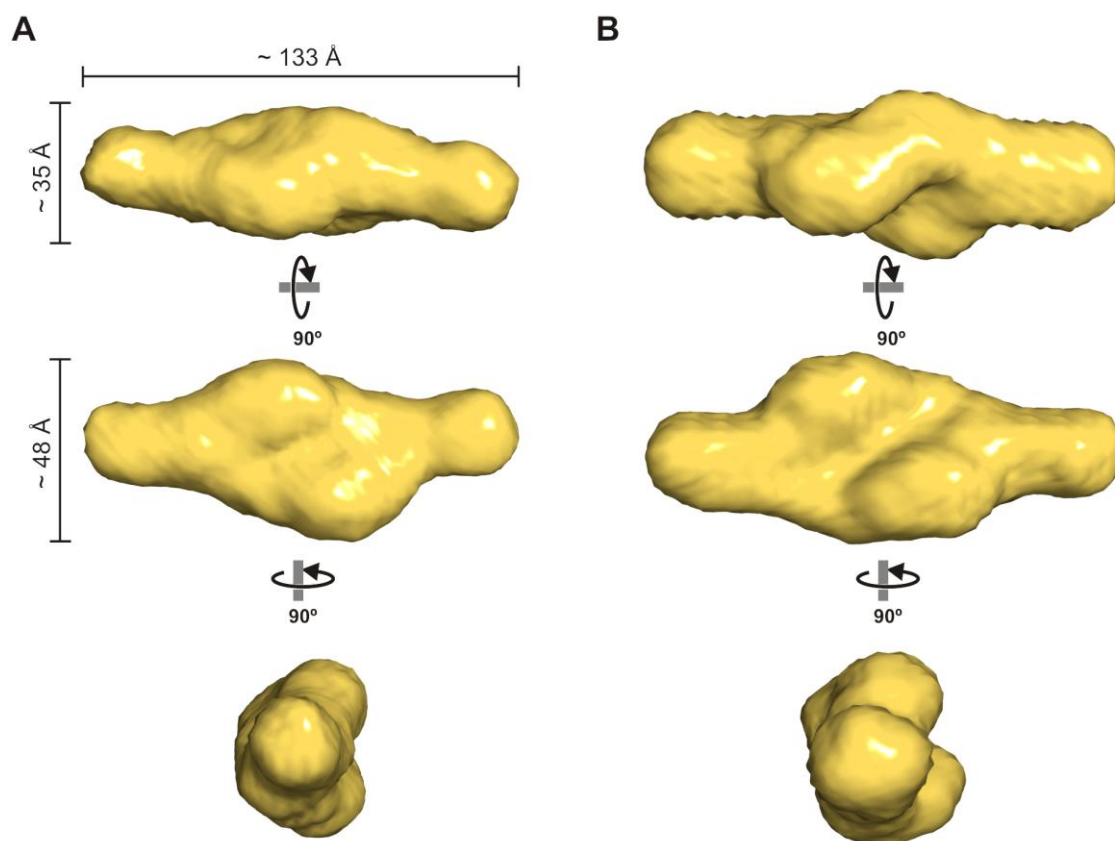


Figura R30. Reconstrucciones de la estructura de N-C3G (65-245) a partir de perfiles de SAXS. Tres vistas ortogonales de las estructuras promedio determinadas a partir de dos conjuntos de 15 modelos generados con el programa DAMMIF aplicando una simetría P1 (A) y P2 (B). Las tres representaciones en A y B se muestran en orientaciones equivalentes y en la misma escala.

3. Unión de N-C3G a fosfolípidos

Mediante los estudios de estructura secundaria hemos determinado que la región de N-C3G es especialmente rica en α -hélices con carácter anfipático que podrían formar un haz helicoidal. Los dominios formados por haces de hélices se observan en proteínas que interactúan con lípidos, como la cola de vinculina (Bakolitsa et al., 1999). Esto sugiere que esta región podría unirse a membranas. En concreto, la secuencia más próxima al extremo amino, residuos 4-64, es rica en residuos básicos ($pI = 9.9$), lo que sugiere que podría interactuar con fosfolípidos aniónicos (Figura R31-A).

Para explorar esta posible unión a fosfolípidos se realizaron experimentos de cosedimentación de los fragmentos de N-C3G (4-245) y (65-245) con vesículas de lípidos formadas por 100 % de PC y 50 % PS:PC a 0.5mM y a 1mM (Figura R31-B). No se observó interacción de ninguna de las 2 proteínas de N-C3G con vesículas formadas por 100 % de PC; sin embargo, sí se observó interacción de la proteína 4-245 con vesículas formadas por 50 % PS:PC tanto a 0.5mM como a 1mM, pero dicha interacción se perdió con la eliminación del extremo amino. En resumen, N-C3G se une a lípidos aniónicos y dicha interacción requiere los primeros 64 aminoácidos.

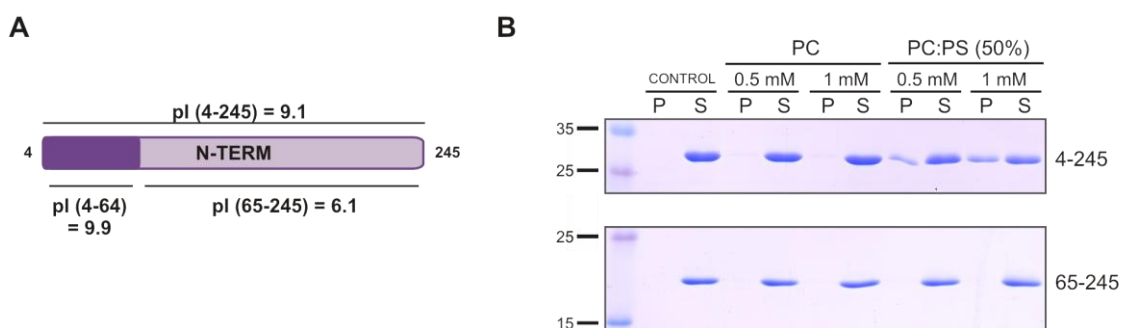


Figura R31. Ensayos de cosedimentación de N-C3G (4-245) y (65-245) con fosfolípidos. (A) Estudio del punto isoeléctrico de diferentes segmentos de N-C3G. (B) Ensayo de unión a lípidos de N-C3G con vesículas formadas por 100 % de PC y 50 % PS:PC. Las muestras corresponden a los precipitados (P) y a los sobrenadantes (S).

DISCUSIÓN

1. Organización estructural de C3G

C3G tiene una estructura modular formada por varios dominios organizados en tres regiones diferenciadas. Hemos caracterizado una interacción intramolecular entre la región N-terminal y el dominio REM de la región catalítica de C3G, los cuales están separados por la región SH3b de ~ 425 residuos. Por lo tanto C3G adopta una estructura compacta o cerrada.

La región central SH3b no interviene en la interacción cabeza-cola. La predicción de estructura secundaria sugiere que en el SH3b predomina la estructura desordenada y flexible, que permitiría que el dominio de cabeza se repliegue sobre el dominio REM. Por otro lado, es posible que la interacción cabeza-cola reduzca la libertad conformacional de SH3b y contribuya a que se organice la estructura de la región central, la cual podría interaccionar al menos en alguna parte con los otros dominios de C3G (Figura D1).

Los residuos E731 y E784 del dominio REM son importantes para la interacción con la región de cabeza. El efecto aditivo de la combinación de las mutaciones E731R y E784R sobre la conformación de C3G sugiere que E731 y E784 median contactos independientes. Estos residuos se localizan próximos en la superficie del dominio REM en las hélices $\alpha 3$ y $\alpha 5$, y definen una posible zona de interacción con el dominio N-terminal.

1.1. Posible papel de la interacción cabeza-cola en la regulación de C3G

Evolutivamente el desarrollo de proteínas modulares se asocia con una mayor complejidad filogenética. La combinación de múltiples dominios en una cadena polipeptídica es frecuente en proteínas reguladoras y permite integrar múltiples funciones y desarrollar mecanismos de regulación (Bhattacharyya et al., 2006; Pawson and Nash, 2003). Las interacciones intramoleculares entre elementos separados dentro de la misma molécula es una estrategia de regulación muy común para modular la función de las proteínas. Así, en un estado cerrado o autoinhibido, una región de la proteína interacciona con otra para regular de forma negativa su actividad (Pufall and Graves, 2002; Trudeau et al., 2013). Los contactos intramoleculares tienen lugar frecuentemente entre superficies pequeñas y son

interacciones relativamente débiles que se encuentran reforzadas por la alta concentración efectiva local de las partes que interactúan, de modo que previenen la activación constitutiva al mismo tiempo que permiten la desestabilización de la autoinhibición en respuesta a estímulos celulares. Frecuentemente los mecanismos de apertura y activación incluyen unión de ligandos a los dominios reguladores, introducción de modificaciones post-traduccionales o combinación de ambos mecanismos.

Hay multitud de ejemplos de proteínas con funciones muy diversas que establecen interacciones intramoleculares. Por ejemplo, muchas quinasas se encuentran autoinhibidas por interacciones intramoleculares de regiones reguladoras con el dominio catalítico, como es el caso de FAK (quinasa de adhesiones focales) (Lietha et al., 2007), Src (Sicheri and Kuriyan, 1997; Xu et al., 1999) y Abl (Nagar et al., 2003) entre otras. También se han descrito interacciones intramoleculares en diversos GEFs. Por ejemplo, las proteínas Vav (GEFs de GTPasas de la familia Rho) adoptan un estado autoinhibido en el que la región N-terminal interactúa con el dominio catalítico DH bloqueando la unión a la GTPasa (Aghazadeh et al., 2000; Llorca et al., 2005). También se han descrito mecanismos de autoinhibición en GEFs de tipo CDC25H. En SOS1 los dominios DH-PH de la región amino interactúan con el dominio REM bloqueando el acceso de Ras-GTP al sitio alostérico (Sondermann et al., 2004). En RasGRP1 el dominio EF de la región C-terminal interactúa con el dominio CDC25H bloqueando de manera indirecta el sitio de unión de la GTPasa (Iwig et al., 2013). En Epac2 la interacción del dominio cNBD-B con el dominio CDC25H estabiliza un estado autoinhibido (Rehmann et al., 2008).

Diversas evidencias sugieren la presencia en C3G de elementos reguladores fuera de la región catalítica. La expresión de la región catalítica aislada induce niveles de Rap1-GTP superiores a la expresión de C3G completo (Ichiba et al., 1997). De manera más específica, la expresión en células COS-1 de una forma truncada de C3G que carece de la mitad N-terminal, aminoácidos 1-579, que incluye la región de cabeza y parte de la región SH3b, aumenta la actividad GEF de C3G de manera independiente de Crk, lo que sugiere que la región 1-579 constituye un elemento regulador negativo en *cis* (Ichiba et al., 1999).

Dado el papel de la región N-terminal de C3G en su regulación y su participación en la interacción cabeza-cola, es razonable que exista una relación entre ambos procesos y que la conformación cerrada estabilizada por la interacción entre la región N-C3G y el dominio REM corresponda a un estado autoinhibido. A pesar de que el dominio CDC25H es dispensable para la interacción cabeza-cola, sería posible que en la conformación cerrada la actividad GEF esté inhibida por alguno de los siguientes mecanismos (Figura D1). (i) En las estructuras de las regiones REM-CDC25H de otros GEFs las hélices $\alpha 3$ y $\alpha 5$ del dominio REM se localizan próximas a la horquilla helicoidal del dominio catalítico. Por lo tanto sería posible que la región N-C3G unida al dominio REM se extienda hacia el dominio catalítico y pueda bloquear la entrada de la GTPasa al sitio activo sin llegar a establecer contactos significativos con el dominio CDC25H (Figura D1-A). (ii) Por otro lado, la unión de N-C3G al dominio REM podría transmitir un cambio conformacional al dominio catalítico, por ejemplo induciendo una distorsión en la horquilla helicoidal, de modo que éste quede bloqueado en una conformación inactiva (Figura D1-B). (iii) La inhibición también podría producirse de forma indirecta, por ejemplo, la interacción cabeza-cola podría estabilizar una conformación de la región SH3b que bloqueara el sitio activo del dominio catalítico; esto sería similar al mecanismo de autoinhibición observado en RasGRP1, en el cual el dominio EF se une a una superficie del CDC25H que no está implicada en la unión de la GTPasa, pero facilita que parte de la secuencia que une los dominios CDC25H y EF se organice sobre la superficie del dominio catalítico bloqueando el sitio activo (Iwig et al., 2013) (Figura D1-C). (iv) Además, es posible una combinación de varios de los anteriores mecanismos, por ejemplo que las regiones N-C3G y SH3b contribuyan directamente al bloqueo estérico del sitio activo (Figura D1-D).

En SOS1 la interacción de la región reguladora (dominio Histona-DH-PH) con el dominio REM bloquea el acceso de Ras-GTP al sitio distal alostérico, manteniendo el dominio catalítico en una conformación inactiva (Gureasko et al., 2010). No hay evidencias de regulación alostérica de la región catalítica de C3G, por lo tanto parece improbable que la interacción cabeza-cola en C3G pueda regular la actividad GEF mediante un mecanismo similar al descrito en SOS1.

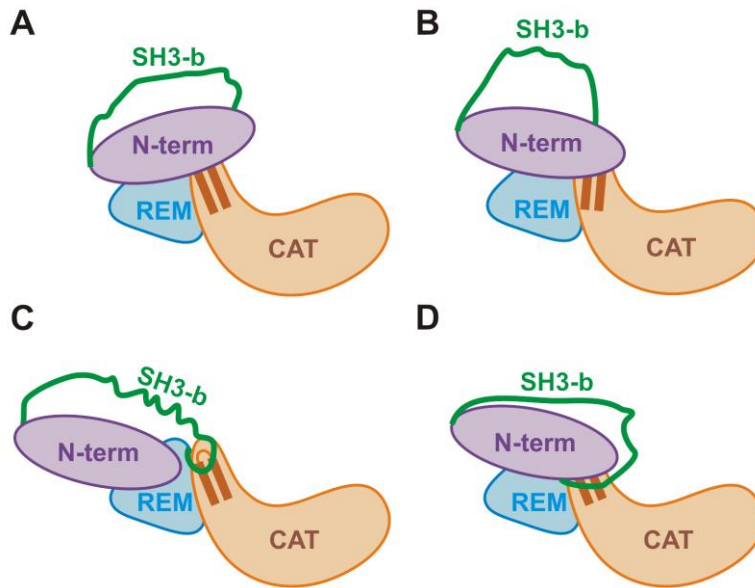


Figura D1. Modelos de la organización de la conformación cerrada de C3G. (A) La región N-C3G bloquea el sitio activo. (B) N-C3G induce un cambio conformacional en la horquilla helicoidal estabilizando una conformación inactiva. (C) La región SH3b bloquea el sitio activo del dominio catalítico. (D) Las regiones N-C3G y SH3b bloquean el sitio activo por impedimento estérico.

La fosforilación de C3G en la Y504 estimula su actividad GEF sobre Rap1 *in vivo* (Ichiba et al., 1999). Con el fin de analizar el posible papel de la Y504 en la regulación de la interacción cabeza-cola, se analizó el efecto de la sustitución Y504E, que mimetiza de forma aproximada la fosforilación introduciendo un grupo con carga negativa. Los resultados de coIP y FRET indican que Y504E no inhibe la interacción intramolecular de C3G. Por lo tanto Y504 no parece ser necesaria para mantener la interacción entre la región N-terminal y el dominio REM, lo cual concuerda con el aparente papel accesorio de la región SH3b en esta interacción. Sin embargo, nuestros datos no descartan que la fosforilación de Y504 pueda tener un papel en la regulación de la conformación de forma indirecta. Por ejemplo la fosforilación podría crear un sitio de unión para alguna proteína con dominios que reconocen p-Tyr (por ejemplo dominios SH2), cuya unión pudiera modular la conformación de C3G. En este sentido, la sustitución por Glu no recrea un sitio de unión a dominios que reconocen p-Tyr. Además, tampoco se puede descartar que la regulación de la interacción cabeza-cola pueda requerir la fosforilación de otras Tyr. Por ejemplo Hck y c-Abl fosforilan el mutante C3G-Y504F en otras Tyr (Mitra and Radha, 2010; Shivakrupa et al., 2003). La fosforilación en Y504 también podría contribuir a la regulación de C3G; en este sentido, se ha descrito que dicha

fosforilación participa en el control de la localización subcelular de C3G (Radha et al., 2007).

Recientemente se ha descrito que la actividad GEF *in vitro* del fragmento de C3G 544-1077, que contiene un motivo rico en Pro de la región SH3b y la región catalítica, es muy inferior a la del dominio catalítico CDC25H aislado, lo que sugiere que la región 544-670 del SH3b es suficiente para mediar la autoinhibición de C3G (Popovic et al., 2013b). Además el fragmento 535-1077, que incluye dos motivos ricos en Pro, es activado *in vitro* de forma aditiva por la unión del primer dominio SH3 de Crk y por la fosforilación por Src. Dado que en este trabajo se han utilizado fragmentos que carecen de la región N-terminal, será importante verificar si la región 544-670 autoinhibe la actividad GEF en C3G completo *in vivo* y si la unión de Crk a la región C-terminal de la región SH3b, ligada a la fosforilación por Src en condiciones fisiológicas, es suficiente para activar C3G.

1.1.1. Relación entre la unión de la región amino de C3G a E-cadherina y la interacción cabeza-cola

C3G interacciona con el dominio citoplasmático de E-cadherina durante la formación de uniones adherentes y esta interacción se pierde en uniones maduras por competición con β -catenina. La interacción se establece entre el extremo de la región citoplasmática de E-cadherina, residuos 823-884 (669-730 en la proteína madura), y el segmento 144-230 de la región N-terminal de C3G (Hogan et al., 2004). El reclutamiento de C3G a las uniones adherentes contribuye a la activación local de Rap1 por lo que es posible que C3G se active al interactuar con E-cadherina. Puesto que la región N-terminal de C3G media tanto la interacción con el dominio REM como con E-cadherina, es posible que ambos procesos estén acoplados.

La región citoplasmática de E-cadherina tiene una estructura desordenada cuando no está unida a otras proteínas (Huber et al., 2001). Por otro lado, la región 731-882 de E-cadherina (residuos 577-728 de la proteína madura) se une a la región armadillo de β -catenina en una conformación mayoritariamente extendida que se dispone a lo largo de las 12 repeticiones armadillo de β -catenina (Huber

and Weis, 2001). De forma análoga es previsible que E-cadherina se una a C3G en una conformación extendida que cubra una superficie relativamente grande de la región N-terminal y por lo tanto sería probable que la superficie de unión a E-cadherina se solape, al menos parcialmente, con la superficie de interacción con el dominio REM. En resumen, la unión de E-cadherina podría competir con la interacción intramolecular estabilizando una conformación activa, de modo que la potencial apertura de C3G inducida por E-cadherina estaría ligada a un aumento de la actividad GEF sobre Rap1 en las uniones adherentes.

2. Estructura de la región amino terminal de C3G

La región N-terminal de C3G no muestra similitud de secuencia con dominios de otras proteínas. La estimación de la estructura secundaria a partir de los espectros de CD de los fragmentos 4-245 y 65-245, expresados en bacteria, sugiere un alto contenido en α -hélice. Esto concuerda con la predicción de estructura secundaria a partir de la secuencia.

En la secuencia de la región N-terminal se distinguen tres segmentos. La primera parte, ~ 60 residuos, se caracteriza por una abundancia de aminoácidos con grupos básicos y una predicción de estructura mayoritariamente desordenada. En el segmento central, residuos $\sim 90-210$, la predicción de estructura secundaria sugiere la presencia de cuatro α -hélices ($\alpha 1-\alpha 4$) anfipáticas relativamente largas, de en torno a 20-30 residuos a excepción de $\alpha 2$ que se extendería ~ 13 residuos. Por último se extiende un segmento, secuencia 210-245, rico en residuos de Pro y Ser.

El carácter anfipático de las potenciales hélices de la región 90-210 sugiere que éstas pueden disponerse formando un haz de hélices hacia cuyo núcleo se orienten los residuos hidrofóbicos. Los haces de cuatro hélices son motivos estructurales comunes que están presentes en diversas proteínas, por ejemplo vinculina (Bakolitsa et al., 2004), α -catenina (Yang et al., 2001), talina (Goult et al., 2013), FAK (Hayashi et al., 2002), GIT1 (proteína que interacciona con la quinasa de receptores acoplados a proteínas-G) (Zhang et al., 2008) y apolipoproteína E (Wilson et al., 1991) entre otras.

Empleando el servidor FFAS03, que compara perfiles de secuencias y es más sensible para identificar similitudes entre proteínas que métodos basados en la comparación de secuencias (Jaroszewski et al., 2005), no se detectó ninguna similitud claramente significativa entre N-C3G y proteínas con dominios formados por haces de hélices (FFAS considera que existe similitud significativa cuando el valor del parámetro Z es ≤ -9.5). Sin embargo, FFAS03 sugiere cierta similitud ($Z > -9.5$) entre la región 90-210 de C3G y varios dominios de talina, GIT1 y vinculina (Tabla D1), siendo llamativo que se trata en todos los casos de estructuras formadas por haces de hélices (Figura D2). Además, curiosamente, se trata en su mayoría de proteínas que, al igual que C3G, se localizan en las adhesiones focales (Maia et al., 2013). En resumen, es posible que la estructura de la región 90-210 de C3G esté relacionada y sea similar a la de dominios de haces de hélices.

Tabla D1. Resumen de alineamientos de la región 90-210 de C3G frente a la base de datos del PDB empleando el servidor FFAS03.

Proteína	Código PDB	Región	Valor Z
Talina 1	1U89	765-874	-8.84
Talina 1	2L7A	796-911	-8.14
GIT1	2JX0	652-766	-8.14
Talina 1	2KBB	1700-1816	-7.53
Talina 1	2KVP	1855-1970	-7.45
Talina 1	2L7N	1089-1201	-7.36
Talina 1	2LQG	914-1040	-7.15
Vinculina	1U6H	130-250	-6.45

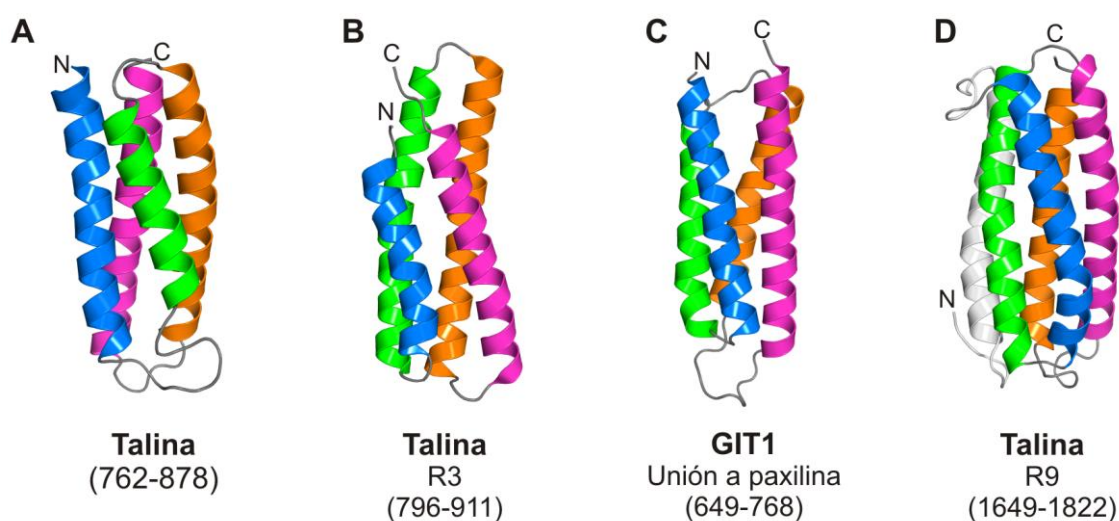


Figura D2. Estructura de dominios con estructura de haz de hélices. Representaciones de cintas de las cuatro estructuras identificadas por el servidor FFAS03 con mejor puntuación de

similitud con la región 90-210 de C3G. Se consideran estructuras con similitud fiable (menos de 3% de falsos positivos) aquellas que obtiene un valor $Z < -9.5$. (A) Región 762-878 de talina ($Z = -8.8$) (código PDB: 1U89) (Fillingham et al., 2005). (B) Dominio R3 de talina ($Z = -8.1$) (código PDB: 2L7A) (Goult et al., 2013). (C) Dominio de unión a paxilina de GIT1 ($Z = -8.1$) (código PDB: 2JX0) (Zhang et al., 2008). (D) Dominio R9 de talina ($Z = -7.5$) (código PDB: 2KBB) (Goult et al., 2009). En todas las estructuras se han coloreado las hélices en azul (α_1), verde (α_2), naranja (α_3) y magenta (α_4). La región 762-878 de talina forma un haz de hélices levógiro, mientras que en las otras tres estructuras las hélices están dispuestas en haces dextrógiros. El dominio R9 de talina es un haz de cinco hélices, la quinta hélice precede la región de potencial homología con C3G y se muestra en color blanco.

Aunque la región 90-210 de C3G pueda constituir un núcleo estructural, el segmento adyacente 211-245 también parece ser necesario para la estabilidad de este dominio, ya que no se detectó expresión en bacteria de las construcciones 4-210 y 65-210. La predicción de estructura secundaria sugiere la presencia de una α -hélice en torno a las posiciones 238-246, que podría estar interactuando con el conjunto de hélices en el segmento 90-210.

Los fragmentos 4-245 y 65-245 son relativamente sensibles a la digestión por proteasas, existiendo múltiples puntos de corte a lo largo de la región N-terminal. Sin embargo es posible que algunos de los sitios de digestión identificados puedan ser puntos de procesamiento secundario, cuya accesibilidad a las proteasas requiera un relajamiento de la estructura. Esto se debe a que nuestros experimentos se encaminaron a identificar potenciales zonas resistentes a la digestión, que pudieran corresponder a zonas compactas adecuadas para el estudio cristalográfico.

2.1. Homo-asociación de la región amino

Fragmentos recombinantes de la región N-terminal de C3G expresados en *E. coli* muestran homo-asociación dependiente de la concentración. Hemos caracterizado en detalle el estado de asociación del fragmento 65-245 mediante técnicas biofísicas complementarias que incluyen cromatografía de exclusión molecular, ultracentrifugación analítica, SAXS y técnicas de fluorescencia. A partir de los datos de equilibrio de sedimentación y SAXS se observó que en un rango de concentraciones entre $\sim 10 \mu\text{M}$ y $\sim 100 \mu\text{M}$ existe un predominio aparente de dímeros de N-C3G con una pequeña fracción de especies de mayor tamaño,

posiblemente tetrámeros. Esto concuerda con la presencia de dos especies (2.3-2.5 S y 3.6-4.2 S respectivamente) observadas mediante velocidad de sedimentación. A concentraciones $> 100 \mu\text{M}$ se detecta la presencia de oligómeros de mayor masa, siendo esta transición de menor afinidad. Por otro lado los datos de cromatografía de exclusión molecular sugieren disociación de los dímeros de N-C3G a concentraciones sub-micromolares.

Es significativo que no se detectó homo-asociación de la región N-C3G mediante ensayos de coIP empleando proteínas expresadas en células COS-1. Una posibilidad es que la interacción sea sensible a las condiciones experimentales empleadas en los ensayos de coIP, por ejemplo la presencia de Triton X-100 podría desestabilizar la interacción. Alternativamente, la homo-asociación de la región N-C3G podría estar bloqueada en células COS-1 por modificaciones postraduccionales o por la interacción de esta región con una segunda proteína.

Será importante determinar si la región N-terminal media la homo-asociación de C3G en el contexto de la proteína completa. No se detectó ninguna señal de FRET cuando CFP-C3G (CFP en el extremo amino) y C3G-Venus (Venus en el extremo carboxilo) se co-expresaron en células COS-1, lo que sugiere que ambas construcciones no interaccionan. Además, un análisis preliminar mediante cromatografía de exclusión molecular combinado con dispersión estática de luz de C3G expresado en bacterias indica que C3G es un monómero en solución (datos no mostrados); en este experimento se empleó una columna Superdex 200 (300 mm x 10 mm) en la que se cargaron 100 μl de una solución 12 μM de C3G, por lo que se puede estimar que la muestra se diluyó entre 5 y 10 veces, de modo que la concentración efectiva sería ~ 1 o 2 μM . Por lo tanto no se puede descartar que C3G forme oligómeros *in vitro* a concentraciones más elevadas. Sin embargo, también puede ser posible que la interacción de la región amino con el dominio REM compita con la homo-asociación, de modo que la formación de oligómeros esté bloqueada en la conformación cerrada. Esto implicaría la posible existencia de un acoplamiento entre una potencial apertura de C3G y la oligomerización.

Se ha descrito oligomerización de otros GEFs. Por ejemplo Dbl, RasGRF1 y RasGRF2 dimerizan a través de los dominios DH (Anborgh et al., 1999; Zhu et al., 2001); DOCK2 y DOCK9 homo-dimerizan a través de sus dominios DHR2 (también denominado CZH2) (Kulkarni et al., 2011; Yang et al., 2009) y esta dimerización es

esencial para la activación de Rac por DOCK2 (Terasawa et al., 2012). RasGRP1 presenta dos regiones de dimerización; por una lado, una hélice en su extremo C-terminal forma un coiled coil dimérico paralelo; por otro lado los dominios CDC25H, EF y C1 crean una segunda zona de dimerización que estabiliza un estado autoinhibido (Iwig et al., 2013).

La oligomerización de C3G podría contribuir a aumentar la concentración localizada de C3G y a estabilizar su interacción con otras moléculas, por ejemplo en el entorno de la membrana, al aumentar su avidéz. Además, dada la capacidad de C3G de interactuar con múltiples proteínas, por ejemplo en adhesiones focales (Maia et al., 2013), la oligomerización permitiría que C3G actúe como nucleador de complejos multiproteína en los que se aglutinen diversas moléculas señalizadoras. En resumen, la homoasociación de C3G podría tener implicaciones para su función.

3. Interacción de la región amino de C3G con lípidos

Los lípidos tienen un papel importante en el control de las funciones de los GEFs de Ras. Después de la activación de los receptores de membrana se generan productos lipídicos que se unen a diversos dominios en los diferentes tipos de GEFs para facilitar la activación de las GTPasas Ras que se encuentran ancladas a la membrana.

El primer segmento de la región N-terminal de C3G, residuos 4-64, es un sitio de unión a fosfolípidos aniónicos. Esta región no contiene dominios reconocibles de unión a fosfolípidos (por ejemplo dominios PH) y probablemente se trate de una secuencia con una estructura mayoritariamente desordenada (ver antes), por lo que la interacción con las membranas parece deberse a interacciones electrostáticas. Esto es reminiscente de los modos de unión de otros dominios, como el dominio FERM de FAK (Cai et al., 2008) o el dominio Histona de SOS1 (Gureasko et al., 2010), en los que la interacción con fosfolípidos requiere regiones ricas en residuos con cadenas laterales básicas.

La interacción de N-C3G con fosfolípidos de membrana podría contribuir a la regulación de C3G. De forma análoga al papel de la interacción del dominio Histona de SOS1 con PIP₂ o ácido fosfatídico, la cual estabiliza una conformación catalíticamente activa, la interacción del extremo N-terminal de C3G con la

membrana podría estabilizar una conformación activa. La participación de la región amino de C3G en la interacción con el dominio REM sugiere que la unión a fosfolípidos podría modular la interacción cabeza-cola. Por ejemplo, en linfocitos la traslocación de RIAM a la membrana plasmática se produce tras la activación de Rap1 por C3G y la unión de RIAM a Rap1-GTP requiere la unión del dominio PH de RIAM a PIP₂, por el que muestra moderada afinidad (Wynne et al., 2012); por lo tanto, es posible que la activación de Rap1 por C3G también esté favorecida por la presencia de fosfoinosítidos fosfato. Por lo tanto, será importante identificar los fosfolípidos fisiológicos a los que se une el extremo N-terminal de C3G in vivo y evaluar su efecto sobre la actividad de C3G.

4. Modelo hipotético de regulación de C3G

A partir de los resultados obtenidos en este trabajo se puede proponer el siguiente modelo genérico de activación de C3G (Figura D3). En ausencia de estímulo, C3G se encuentra en una conformación cerrada y se localiza en el citoplasma asociado mayoritariamente a Crk. En respuesta a la activación de receptores tirosín quinasa, el complejo C3G-Crk se trasloca a la membrana donde C3G se pone en contacto con la GTPasa y además se activa en un proceso ligado a la fosforilación de C3G en Tyr por quinasas de la familia Src. La activación de C3G en la membrana posiblemente se produce por un cambio conformacional en el cual se podría romper la interacción entre la región N-terminal y el dominio REM. La interacción del extremo N-terminal con fosfolípidos aniónicos en la membrana no sería responsable de la traslocación a la membrana si no que cooperaría para estabilizar la conformación catalíticamente activa. La activación localizada de C3G podría estar favorecida por el aglutinamiento a través de interacciones homotípicas de la región N-terminal.

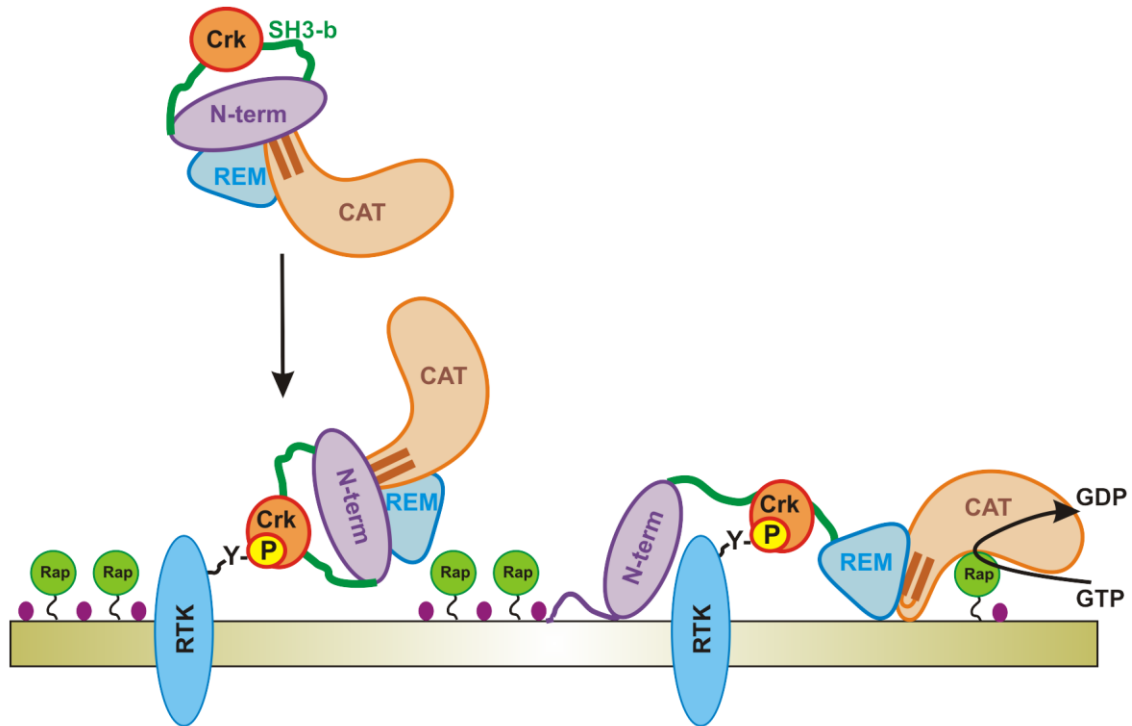


Figura D3. Modelo hipotético de activación de C3G. En el citosol C3G se encuentra en una conformación cerrada asociado a Crk. C3G se trasloca a la membrana en respuesta a la activación de RTKs donde establece contactos con la GTPasa. En el proceso de activación en la membrana C3G sufre un cambio conformacional.

En resumen, las herramientas desarrolladas en este trabajo permitirán analizar el papel de la interacción cabeza-cola en la regulación de C3G. Los sensores conformacionales de FRET de C3G serán de utilidad para evaluar el estado conformacional de C3G en respuesta a estímulos. Las sustituciones puntuales que modifican la conformación de C3G (E731R, E784R y el doble mutante E731R-E784R) permitirán analizar el papel de la interacción cabeza-cola en la regulación de C3G. En este sentido y dado que algunas funciones de C3G son independientes de su actividad GEF el uso de sustituciones puntuales para analizar el efecto de la conformación de C3G sobre su función, es a priori más adecuado que la utilización de mutantes de delección en los que se podrían eliminar simultáneamente múltiples funciones.

CONCLUSIONES

1. La región amino terminal de C3G (residuos 4-245) establece una interacción intramolecular con el dominio REM de la región catalítica, siendo el dominio CDC25H dispensable para esta interacción.
2. Los residuos E731 y E784 del dominio REM son importantes para la interacción con la región amino terminal, se encuentran potencialmente próximos en la superficie de este dominio y definen una posible zona de interacción con la región amino terminal.
3. La sustitución Y504E en la región central (dominio SH3b) no es suficiente para inhibir la interacción intramolecular de C3G.
4. La región amino terminal de C3G tiene una estructura secundaria rica en α -hélices que podrían formar un haz de hélices.
5. El fragmento 65-245 de la región amino terminal homoasocia *in vitro* siendo el dímero la especie predominante a concentraciones entre 10 y 100 μ M.
6. El extremo amino terminal, residuos 4-65, se une a fosfolípidos aniónicos.

BIBLIOGRAFÍA

Aghazadeh, B., Lowry, W.E., Huang, X.Y., and Rosen, M.K. (2000). Structural basis for relief of autoinhibition of the Dbl homology domain of proto-oncogene Vav by tyrosine phosphorylation. *Cell* *102*, 625-633.

Alonso-Garcia, N., Ingles-Prieto, A., Sonnenberg, A., and de Pereda, J.M. (2009). Structure of the Calx-beta domain of the integrin beta4 subunit: insights into function and cation-independent stability. *Acta Crystallogr D Biol Crystallogr* *65*, 858-871.

Alsayed, Y., Uddin, S., Ahmad, S., Majchrzak, B., Druker, B.J., Fish, E.N., and Platanius, L.C. (2000). IFN-gamma activates the C3G/Rap1 signaling pathway. *J Immunol* *164*, 1800-1806.

Anborgh, P.H., Qian, X., Papageorge, A.G., Vass, W.C., DeClue, J.E., and Lowy, D.R. (1999). Ras-specific exchange factor GRF: oligomerization through its Dbl homology domain and calcium-dependent activation of Raf. *Molecular and cellular biology* *19*, 4611-4622.

Andrade, M.A., Chacon, P., Merelo, J.J., and Moran, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Eng* *6*, 383-390.

Arai, A., Nosaka, Y., Kanda, E., Yamamoto, K., Miyasaka, N., and Miura, O. (2001). Rap1 is activated by erythropoietin or interleukin-3 and is involved in regulation of beta1 integrin-mediated hematopoietic cell adhesion. *J Biol Chem* *276*, 10453-10462.

Arai, A., Nosaka, Y., Kohsaka, H., Miyasaka, N., and Miura, O. (1999). CrkL activates integrin-mediated hematopoietic cell adhesion through the guanine nucleotide exchange factor C3G. *Blood* *93*, 3713-3722.

Bakolitsa, C., Cohen, D.M., Bankston, L.A., Bobkov, A.A., Cadwell, G.W., Jennings, L., Critchley, D.R., Craig, S.W., and Liddington, R.C. (2004). Structural basis for vinculin activation at sites of cell adhesion. *Nature* *430*, 583-586.

Bakolitsa, C., de Pereda, J.M., Bagshaw, C.R., Critchley, D.R., and Liddington, R.C. (1999). Crystal structure of the vinculin tail suggests a pathway for activation. *Cell* *99*, 603-613.

Barbacid, M. (1990). ras oncogenes: their role in neoplasia. *Eur J Clin Invest* *20*, 225-235.

Beaulieu, N., Zahedi, B., Goulding, R.E., Tazmini, G., Anthony, K.V., Omeis, S.L., de Jong, D.R., and Kay, R.J. (2007). Regulation of RasGRP1 by B cell antigen receptor requires cooperativity between three domains controlling translocation to the plasma membrane. *Molecular biology of the cell* *18*, 3156-3168.

Bhattacharyya, R.P., Remenyi, A., Yeh, B.J., and Lim, W.A. (2006). Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits. *Annual review of biochemistry* *75*, 655-680.

Block, H., Maertens, B., Spriestersbach, A., Brinker, N., Kubicek, J., Fabis, R., Labahn, J., and Schafer, F. (2009). Immobilized-metal affinity chromatography (IMAC): a review. *Methods in enzymology* *463*, 439-473.

Boguski, M.S., and McCormick, F. (1993). Proteins regulating Ras and its relatives. *Nature* *366*, 643-654.

- Boriack-Sjodin, P.A., Margarit, S.M., Bar-Sagi, D., and Kuriyan, J. (1998). The structural basis of the activation of Ras by Sos. *Nature* 394, 337-343.
- Bos, J.L. (1989). ras oncogenes in human cancer: a review. *Cancer Res* 49, 4682-4689.
- Bos, J.L. (1998). All in the family? New insights and questions regarding interconnectivity of Ras, Rap1 and Ral. *EMBO J* 17, 6776-6782.
- Bos, J.L., Rehmann, H., and Wittinghofer, A. (2007). GEFs and GAPs: critical elements in the control of small G proteins. *Cell* 129, 865-877.
- Bourne, H.R., Sanders, D.A., and McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. *Nature* 349, 117-127.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* 72, 248-254.
- Buensuceso, C.S., and O'Toole, T.E. (2000). The association of CRKII with C3G can be regulated by integrins and defines a novel means to regulate the mitogen-activated protein kinases. *J Biol Chem* 275, 13118-13125.
- Cai, X., Lietha, D., Ceccarelli, D.F., Karginov, A.V., Rajfur, Z., Jacobson, K., Hahn, K.M., Eck, M.J., and Schaller, M.D. (2008). Spatial and temporal regulation of focal adhesion kinase activity in living cells. *Molecular and cellular biology* 28, 201-214.
- Camonis, J.H., Kalekine, M., Gondre, B., Garreau, H., Boy-Marcotte, E., and M., J. (1986). Characterization, cloning and sequence analysis of the CDC25 gene which controls the cyclic AMP level of *Saccharomyces cerevisiae*. *EMBO J* 5, 375-380.
- Casey, P.J., Solski, P.A., Der, C.J., and Buss, J.E. (1989). p21ras is modified by a farnesyl isoprenoid. *Proceedings of the National Academy of Sciences of the United States of America* 86, 8323-8327.
- Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36, W197-201.
- Cole, J.L. (2004). Analysis of heterogeneous interactions. *Methods in enzymology* 384, 212-232.
- Compton, L.A., and Johnson, W.C., Jr. (1986). Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Analytical biochemistry* 155, 155-167.
- Corbalan-Garcia, S., Degenhardt, K.R., and Bar-Sagi, D. (1996a). Insulin-induced dissociation of Sos from Grb2 does not contribute to the down regulation of Ras activation. *Oncogene* 12, 1063-1068.
- Corbalan-Garcia, S., Yang, S.S., Degenhardt, K.R., and Bar-Sagi, D. (1996b). Identification of the mitogen-activated protein kinase phosphorylation sites on human Sos1 that regulate interaction with Grb2. *Molecular and cellular biology* 16, 5674-5682.

Chen, J., Parsons, S., and Brautigan, D.L. (1994). Tyrosine phosphorylation of protein phosphatase 2A in response to growth stimulation and v-src transformation of fibroblasts. *J Biol Chem* 269, 7957-7962.

Chen, R.H., Corbalan-Garcia, S., and Bar-Sagi, D. (1997). The role of the PH domain in the signal-dependent membrane targeting of Sos. *The EMBO journal* 16, 1351-1359.

Cherfils, J., and Chardin, P. (1999). GEFs: structural basis for their activation of small GTP-binding proteins. *Trends Biochem Sci* 24, 306-311.

Cherfils, J., and Zeghouf, M. (2013). Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol Rev* 93, 269-309.

Chiang, S.H., Baumann, C.A., Kanzaki, M., Thurmond, D.C., Watson, R.T., Neudauer, C.L., Macara, I.G., Pessin, J.E., and Saltiel, A.R. (2001). Insulin-stimulated GLUT4 translocation requires the CAP-dependent activation of TC10. *Nature* 410, 944-948.

Chiang, S.H., Chang, L., and Saltiel, A.R. (2006). TC10 and insulin-stimulated glucose transport. *Methods Enzymol* 406, 701-714.

Chiariello, M., Marinissen, M.J., and Gutkind, J.S. (2000). Multiple mitogen-activated protein kinase signaling pathways connect the cot oncoprotein to the c-jun promoter and to cellular transformation. *Molecular and cellular biology* 20, 1747-1758.

de Jong, R., van Wijk, A., Heisterkamp, N., and Groffen, J. (1998). C3G is tyrosine-phosphorylated after integrin-mediated cell adhesion in normal but not in Bcr/Abl expressing cells. *Oncogene* 17, 2805-2810.

de Rooij, J., Rehmann, H., van Triest, M., Cool, R.H., Wittinghofer, A., and Bos, J.L. (2000). Mechanism of regulation of the Epac family of cAMP-dependent RapGEFs. *The Journal of biological chemistry* 275, 20829-20836.

de Rooij, J., Zwartkruis, F.J., Verheijen, M.H., Cool, R.H., Nijman, S.M., Wittinghofer, A., and Bos, J.L. (1998). Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature* 396, 474-477.

Durand, D., Vives, C., Cannella, D., Perez, J., Pebay-Peyroula, E., Vachette, P., and Fieschi, F. (2010). NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J Struct Biol* 169, 45-53.

Ebinu, J.O., Bottorff, D.A., Chan, E.Y., Stang, S.L., Dunn, R.J., and Stone, J.C. (1998). RasGRP, a Ras guanyl nucleotide- releasing protein with calcium- and diacylglycerol-binding motifs. *Science* 280, 1082-1086.

Egan, S.E., Giddings, B.W., Brooks, M.W., Buday, L., Sizeland, A.M., and Weinberg, R.A. (1993). Association of Sos Ras exchange protein with Grb2 is implicated in tyrosine kinase signal transduction and transformation. *Nature* 363, 45-51.

Etienne-Manneville, S., and Hall, A. (2002). Rho GTPases in cell biology. *Nature* 420, 629-635.

Feller, S.M. (2001). Crk family adaptors-signalling complex formation and biological roles. *Oncogene* 20, 6348-6371.

Fillingham, I., Gingras, A.R., Papagrigoriou, E., Patel, B., Emsley, J., Critchley, D.R., Roberts, G.C., and Barsukov, I.L. (2005). A vinculin binding domain from the talin rod unfolds to form a complex with the vinculin head. *Structure* 13, 65-74.

Förster, T. (1948). Zwischenmolekulare Energiewanderung und Fluoreszenz. *Annalen der Physik* 437, 55-75.

Franke, D., and Svergun, D.I. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *Journal of Applied Crystallography* 42, 342-346.

Freedman, T.S., Sondermann, H., Friedland, G.D., Kortemme, T., Bar-Sagi, D., Marqusee, S., and Kuriyan, J. (2006). A Ras-induced conformational switch in the Ras activator Son of sevenless. *Proceedings of the National Academy of Sciences of the United States of America* 103, 16692-16697.

Freedman, T.S., Sondermann, H., Kuchment, O., Friedland, G.D., Kortemme, T., and Kuriyan, J. (2009). Differences in flexibility underlie functional differences in the Ras activators son of sevenless and Ras guanine nucleotide releasing factor 1. *Structure* 17, 41-53.

Gloerich, M., and Bos, J.L. (2010). Epac: defining a new mechanism for cAMP action. *Annual review of pharmacology and toxicology* 50, 355-375.

Gotoh, T., Hattori, S., Nakamura, S., Kitayama, H., Noda, M., Takai, Y., Kaibuchi, K., Matsui, H., Hatase, O., Takahashi, H., *et al.* (1995). Identification of Rap1 as a target for the Crk SH3 domain-binding guanine nucleotide-releasing factor C3G. *Mol Cell Biol* 15, 6746-6753.

Gotoh, T., Niino, Y., Tokuda, M., Hatase, O., Nakamura, S., Matsuda, M., and Hattori, S. (1997). Activation of R-Ras by Ras-guanine nucleotide-releasing factor. *J Biol Chem* 272, 18602-18607.

Goult, B.T., Bate, N., Anthis, N.J., Wegener, K.L., Gingras, A.R., Patel, B., Barsukov, I.L., Campbell, I.D., Roberts, G.C., and Critchley, D.R. (2009). The structure of an interdomain complex that regulates talin activity. *The Journal of biological chemistry* 284, 15097-15106.

Goult, B.T., Zacharchenko, T., Bate, N., Tsang, R., Hey, F., Gingras, A.R., Elliott, P.R., Roberts, G.C., Ballestrom, C., Critchley, D.R., *et al.* (2013). RIAM and vinculin binding to talin are mutually exclusive and regulate adhesion assembly and turnover. *The Journal of biological chemistry* 288, 8238-8249.

Gual, P., Shigematsu, S., Kanzaki, M., Gremeaux, T., Gonzalez, T., Pessin, J.E., Le Marchand-Brustel, Y., and Tanti, J.F. (2002). A Crk-II/TC10 signaling pathway is required for osmotic shock-stimulated glucose transport. *J Biol Chem* 277, 43980-43986.

Guerrero, C., Fernandez-Medarde, A., Rojas, J.M., Font de Mora, J., Esteban, L.M., and Santos, E. (1998). Transformation suppressor activity of C3G is independent of its CDC25-homology domain. *Oncogene* 16, 613-624.

Guerrero, C., Martin-Encabo, S., Fernandez-Medarde, A., and Santos, E. (2004). C3G-mediated suppression of oncogene-induced focus formation in fibroblasts involves inhibition of ERK activation, cyclin A expression and alterations of anchorage-independent growth. *Oncogene* 23, 4885-4893.

Gureasko, J., Kuchment, O., Makino, D.L., Sondermann, H., Bar-Sagi, D., and Kuriyan, J. (2010). Role of the histone domain in the autoinhibition and activation of the Ras activator Son of Sevenless. *Proceedings of the National Academy of Sciences of the United States of America* 107, 3430-3435.

Gutierrez-Berzal, J., Castellano, E., Martin-Encabo, S., Gutierrez-Cianca, N., Hernandez, J.M., Santos, E., and Guerrero, C. (2006). Characterization of p87C3G, a novel, truncated C3G isoform that is overexpressed in chronic myeloid leukemia and interacts with Bcr-Abl. *Exp Cell Res* 312, 938-948.

Gutierrez-Herrero, S., Maia, V., Gutierrez-Berzal, J., Calzada, N., Sanz, M., Gonzalez-Manchon, C., Pericacho, M., Ortiz-Rivero, S., Gonzalez-Porras, J.R., Arechederra, M., *et al.* (2012). C3G transgenic mouse models with specific expression in platelets reveal a new role for C3G in platelet clotting through its GEF activity. *Biochimica et biophysica acta* 1823, 1366-1377.

Gutiérrez-Uzquiza, A., Arechederra, M., Molina, I., Baños, R., Maia, V., Benito, M., Guerrero, C., and Porras, A. (2010). C3G down-regulates p38 MAPK activity in response to stress by Rap-1 independent mechanisms: Involvement in cell death. *Cellular signalling* 22, 533-542.

Halupa, A., Chohan, M., Stickle, N.H., Beattie, B.K., Miller, B.A., and Barber, D.L. (2005). Erythropoietin receptor Y479 couples to ERK1/2 activation via recruitment of phospholipase Cgamma. *Exp Cell Res* 309, 1-11.

Hayashi, I., Vuori, K., and Liddington, R.C. (2002). The focal adhesion targeting (FAT) region of focal adhesion kinase is a four-helix bundle that binds paxillin. *Nat Struct Biol* 9, 101-106.

Heim, R., and Tsien, R.Y. (1996). Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol* 6, 178-182.

Hogan, C., Serpente, N., Cogram, P., Hosking, C.R., Bialucha, C.U., Feller, S.M., Braga, V.M., Birchmeier, W., and Fujita, Y. (2004). Rap1 regulates the formation of E-cadherin-based cell-cell contacts. *Mol Cell Biol* 24, 6690-6700.

Holzwarth, G., and Doty, P. (1965). The Ultraviolet Circular Dichroism of Polypeptides. *J Am Chem Soc* 87, 218-228.

Huber, A.H., Stewart, D.B., Laurents, D.V., Nelson, W.J., and Weis, W.I. (2001). The cadherin cytoplasmic domain is unstructured in the absence of beta-catenin. A possible mechanism for regulating cadherin turnover. *The Journal of biological chemistry* 276, 12301-12309.

Huber, A.H., and Weis, W.I. (2001). The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. *Cell* 105, 391-402.

Ichiba, T., Hashimoto, Y., Nakaya, M., Kuraishi, Y., Tanaka, S., Kurata, T., Mochizuki, N., and Matsuda, M. (1999). Activation of C3G guanine nucleotide exchange factor for Rap1 by phosphorylation of tyrosine 504. *J Biol Chem* 274, 14376-14381.

Ichiba, T., Kuraishi, Y., Sakai, O., Nagata, S., Groffen, J., Kurata, T., Hattori, S., and Matsuda, M. (1997). Enhancement of guanine-nucleotide exchange activity of C3G for Rap1 by the expression of Crk, CrkL, and Grb2. *J Biol Chem* 272, 22215-22220.

- Innocenti, M., Frittoli, E., Ponzanelli, I., Falck, J.R., Brachmann, S.M., Di Fiore, P.P., and Scita, G. (2003). Phosphoinositide 3-kinase activates Rac by entering in a complex with Eps8, Abi1, and Sos-1. *J Cell Biol* 160, 17-23.
- Innocenti, M., Tenca, P., Frittoli, E., Faretta, M., Tocchetti, A., Di Fiore, P.P., and Scita, G. (2002). Mechanisms through which Sos-1 coordinates the activation of Ras and Rac. *J Cell Biol* 156, 125-136.
- Iwig, J.S., Vercoulen, Y., Das, R., Barros, T., Limnander, A., Che, Y., Pelton, J.G., Wemmer, D.E., Roose, J.P., and Kuriyan, J. (2013). Structural analysis of autoinhibition in the Ras-specific exchange factor RasGRP1. *Elife* 2, e00813.
- Jablonski, A. (1957). Decay of photoluminescence of solutions. *Acta Phys Polon* 26: 471.
- Jaroszewski, L., Li, Z., Cai, X.H., Weber, C., and Godzik, A. (2011). FFAS server: novel features and applications. *Nucleic Acids Res* 39, W38-44.
- Jaroszewski, L., Rychlewski, L., Li, Z., Li, W., and Godzik, A. (2005). FFAS03: a server for profile-profile sequence alignments. *Nucleic Acids Res* 33, W284-288.
- Jin, S., Zhai, B., Qiu, Z., Wu, J., Lane, M.D., and Liao, K. (2000). c-Crk, a substrate of the insulin-like growth factor-1 receptor tyrosine kinase, functions as an early signal mediator in the adipocyte differentiation process. *J Biol Chem* 275, 34344-34352.
- Jun, J.E., Rubio, I., and Roose, J.P. (2013). Regulation of Ras Exchange Factors and Cellular Localization of Ras Activation by Lipid Messengers in T Cells. *Front Immunol* 4, 239.
- Jurnak, F. (1985). Structure of the GDP domain of EF-Tu and location of the amino acids homologous to ras oncogene proteins. *Science* 230, 32-36.
- Kao, S., Jaiswal, R.K., Kolch, W., and Landreth, G.E. (2001). Identification of the mechanisms regulating the differential activation of the mapk cascade by epidermal growth factor and nerve growth factor in PC12 cells. *J Biol Chem* 276, 18169-18177.
- Kawasaki, H., Springett, G.M., Toki, S., Canales, J.J., Harlan, P., Blumenstiel, J.P., Chen, E.J., Bany, I.A., Mochizuki, N., Ashbacher, A., *et al.* (1998). A Rap guanine nucleotide exchange factor enriched highly in the basal ganglia. *Proc Natl Acad Sci U S A* 95, 13278-13283.
- Kim, M.J., Chang, J.S., Park, S.K., Hwang, J.I., Ryu, S.H., and Suh, P.G. (2000). Direct interaction of SOS1 Ras exchange protein with the SH3 domain of phospholipase C-gamma1. *Biochemistry* 39, 8674-8682.
- Knudsen, B.S., Feller, S.M., and Hanafusa, H. (1994). Four proline-rich sequences of the guanine-nucleotide exchange factor C3G bind with unique specificity to the first Src homology 3 domain of Crk. *J Biol Chem* 269, 32781-32787.
- Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. (2006). ATSAS 2.1, a program package for small-angle scattering data analysis. *Journal of Applied Crystallography* 39, 277-286.
- Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *Journal of Applied Crystallography* 36, 1277-1282.

- Koshihara, S., Kigawa, T., Kim, J.H., Shirouzu, M., Bowtell, D., and Yokoyama, S. (1997). The solution structure of the pleckstrin homology domain of mouse Son-of-sevenless 1 (mSos1). *Journal of molecular biology* 269, 579-591.
- Kozin, M.B., and Svergun, D.I. (2001). Automated matching of high- and low-resolution structural models. *Journal of Applied Crystallography* 34, 33-41.
- Kubiseski, T.J., Chook, Y.M., Parris, W.E., Rozakis-Adcock, M., and Pawson, T. (1997). High affinity binding of the pleckstrin homology domain of mSos1 to phosphatidylinositol (4,5)-bisphosphate. *The Journal of biological chemistry* 272, 1799-1804.
- Kulkarni, K., Yang, J., Zhang, Z., and Barford, D. (2011). Multiple factors confer specific Cdc42 and Rac protein activation by dedicator of cytokinesis (DOCK) nucleotide exchange factors. *The Journal of biological chemistry* 286, 25341-25351.
- Kuwamoto, S., Akiyama, S., and Fujisawa, T. (2004). Radiation damage to a protein solution, detected by synchrotron X-ray small-angle scattering: dose-related considerations and suppression by cryoprotectants. *J Synchrotron Radiat* 11, 462-468.
- la Cour, T.F., Nyborg, J., Thirup, S., and Clark, B.F. (1985). Structural details of the binding of guanosine diphosphate to elongation factor Tu from *E. coli* as studied by X-ray crystallography. *The EMBO journal* 4, 2385-2388.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Laue, T., Shah, B., Ridgeway, T., and Pelletier, S. (1992). *Analytical Ultracentrifugation in Biochemistry and Polymer Science*. Edited by Harding S, Rowe A, and Horton J. Royal Society of Chemistry, Cambridge, UK 90-125.
- Lebowitz, J., Lewis, M.S., and Schuck, P. (2002). Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci* 11, 2067-2079.
- Li, N., Batzer, A., Daly, R., Yajnik, V., Skolnik, E., Chardin, P., Bar-Sagi, D., Margolis, B., and Schlessinger, J. (1993). Guanine-nucleotide-releasing factor hSos1 binds to Grb2 and links receptor tyrosine kinases to Ras signalling. *Nature* 363, 85-88.
- Lietha, D., Cai, X., Ceccarelli, D.F., Li, Y., Schaller, M.D., and Eck, M.J. (2007). Structural basis for the autoinhibition of focal adhesion kinase. *Cell* 129, 1177-1187.
- Ling, L., Zhu, T., and Lobie, P.E. (2003). Src-CrkII-C3G-dependent activation of Rap1 switches growth hormone-stimulated p44/42 MAP kinase and JNK/SAPK activities. *J Biol Chem* 278, 27301-27311.
- Liu, C., Takahashi, M., Li, Y., Song, S., Dillon, T.J., Shinde, U., and Stork, P.J. (2008). Ras is required for the cyclic AMP-dependent activation of Rap1 via Epac2. *Molecular and cellular biology* 28, 7109-7125.
- Lobley, A., Whitmore, L., and Wallace, B.A. (2002). DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* 18, 211-212.
- Louis-Jeune, C., Andrade-Navarro, M.A., and Perez-Iratxeta, C. (2011). Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins*.

Lowy, D.R., and Willumsen, B.M. (1993). Function and regulation of *ras*. *Ann Rev Biochem* 62, 851-891.

Llorca, O., Arias-Palomo, E., Zugaza, J.L., and Bustelo, X.R. (2005). Global conformational rearrangements during the activation of the GDP/GTP exchange factor Vav3. *The EMBO journal* 24, 1330-1340.

Macara, I.G., Lounsbury, K.M., Richards, S.A., McKiernan, C., and Bar-Sagi, D. (1996). The Ras superfamily of GTPases. *FASEB J* 10, 625-630.

Maia, V., Ortiz-Rivero, S., Sanz, M., Gutierrez-Berzal, J., Alvarez-Fernandez, I., Gutierrez-Herrero, S., de Pereda, J.M., Porras, A., and Guerrero, C. (2013). C3G forms complexes with Bcr-Abl and p38alpha MAPK at the focal adhesions in chronic myeloid leukemia cells: implication in the regulation of leukemic cell adhesion. *Cell communication and signaling : CCS* 11, 9.

Maia, V., Sanz, M., Gutierrez-Berzal, J., de Luis, A., Gutierrez-Uzquiza, A., Porras, A., and Guerrero, C. (2009). C3G silencing enhances STI-571-induced apoptosis in CML cells through p38 MAPK activation, but it antagonizes STI-571 inhibitory effect on survival. *Cell Signal* 21, 1229-1235.

Manavalan, P., and Johnson, W.C., Jr. (1987). Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Analytical biochemistry* 167, 76-85.

Margarit, S.M., Sondermann, H., Hall, B.E., Nagar, B., Hoelz, A., Pirruccello, M., Bar-Sagi, D., and Kuriyan, J. (2003a). Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. *Cell* 112, 685-695.

Margarit, S.M., Sondermann, H., Hall, B.E., Nagar, B., Hoelz, A., Pirruccello, M., Bar-Sagi, D., and Kuriyan, J. (2003b). Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. *Cell* 112, 685-695.

Martin-Encabo, S., Santos, E., and Guerrero, C. (2007). C3G mediated suppression of malignant transformation involves activation of PP2A phosphatases at the subcortical actin cytoskeleton. *Exp Cell Res* 313, 3881-3891.

Memon, A.R. (2004). The role of ADP-ribosylation factor and SAR1 in vesicular trafficking in plants. *Biochimica et biophysica acta* 1664, 9-30.

Milburn, M.V., Tong, L., deVos, A.M., Brunger, A., Yamaizumi, Z., Nishimura, S., and Kim, S.H. (1990). Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* 247, 939-945.

Mitra, A., Kalayarasan, S., Gupta, V., and Radha, V. (2011). TC-PTP Dephosphorylates the Guanine Nucleotide Exchange Factor C3G (RapGEF1) and Negatively Regulates Differentiation of Human Neuroblastoma Cells. *PLoS ONE* 6, e23681.

Mitra, A., and Radha, V. (2010). F-actin-binding domain of c-Abl regulates localized phosphorylation of C3G: role of C3G in c-Abl-mediated cell death. *Oncogene* 29, 4528-4542.

- Mizuchi, D., Kurosu, T., Kida, A., Jin, Z.H., Jin, A., Arai, A., and Miura, O. (2005). BCR/ABL activates Rap1 and B-Raf to stimulate the MEK/Erk signaling pathway in hematopoietic cells. *Biochem Biophys Res Commun* 326, 645-651.
- Mochizuki, N., Ohba, Y., Kobayashi, S., Otsuka, N., Graybiel, A.M., Tanaka, S., and Matsuda, M. (2000). Crk activation of JNK via C3G and R-Ras. *J Biol Chem* 275, 12667-12671.
- Nagai, T., Ibata, K., Park, E.S., Kubota, M., Mikoshiba, K., and Miyawaki, A. (2002). A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nat Biotechnol* 20, 87-90.
- Nagar, B., Hantschel, O., Young, M.A., Scheffzek, K., Veach, D., Bornmann, W., Clarkson, B., Superti-Furga, G., and Kuriyan, J. (2003). Structural basis for the autoinhibition of c-Abl tyrosine kinase. *Cell* 112, 859-871.
- Newman, J., Egan, D., Walter, T.S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J.L., Stuart, D.I., and Perrakis, A. (2005). Towards rationalization of crystallization screening for small-to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr D Biol Crystallogr* 61, 1426-1431.
- Nimnual, A.S., Yatsula, B.A., and Bar-Sagi, D. (1998). Coupling of Ras and Rac guanosine triphosphatases through the Ras exchanger Sos. *Science* 279, 560-563.
- Nolz, J.C., Nacusi, L.P., Segovis, C.M., Medeiros, R.B., Mitchell, J.S., Shimizu, Y., and Billadeau, D.D. (2008). The WAVE2 complex regulates T cell receptor signaling to integrins via Abl- and CrkL-C3G-mediated activation of Rap1. *J Cell Biol* 182, 1231-1244.
- Nosaka, Y., Arai, A., Miyasaka, N., and Miura, O. (1999). CrkL mediates Ras-dependent activation of the Raf/ERK pathway through the guanine nucleotide exchange factor C3G in hematopoietic cells stimulated with erythropoietin or interleukin-3. *J Biol Chem* 274, 30154-30162.
- Nowell, P.C., and Hungerford, D.A. (1960). A minute chromosome in human chronic granulocytic leukemia. *Science* 132, 1497-1502.
- O'Shea, E.K., Klemm, J.D., Kim, P.S., and Alber, T. (1991). X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254, 539-544.
- Ohba, Y., Ikuta, K., Ogura, A., Matsuda, J., Mochizuki, N., Nagashima, K., Kurokawa, K., Mayer, B.J., Maki, K., Miyazaki, J., *et al.* (2001). Requirement for C3G-dependent Rap1 activation for cell adhesion and embryogenesis. *EMBO J* 20, 3333-3341.
- Ohba, Y., Mochizuki, N., Yamashita, S., Chan, A.M., Schrader, J.W., Hattori, S., Nagashima, K., and Matsuda, M. (2000). Regulatory proteins of R-Ras, TC21/R-Ras2, and M-Ras/R-Ras3. *J Biol Chem* 275, 20020-20026.
- Okada, S., and Pessin, J.E. (1997). Insulin and epidermal growth factor stimulate a conformational change in Rap1 and dissociation of the CrkII-C3G complex. *J Biol Chem* 272, 28179-28182.
- Oki, T., Kitaura, J., Watanabe-Okochi, N., Nishimura, K., Maehara, A., Uchida, T., Komeno, Y., Nakahara, F., Harada, Y., Sonoki, T., *et al.* (2012). Aberrant expression of RasGRP1 cooperates with gain-of-function NOTCH1 mutations in T-cell leukemogenesis. *Leukemia* 26, 1038-1045.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to measure and predict the molar absorption coefficient of a protein. *Protein Sci* 4, 2411-2423.

Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W., and Wittinghofer, A. (1990). Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 Å resolution: implications for the mechanism of GTP hydrolysis. *The EMBO journal* 9, 2351-2359.

Pawson, T., and Nash, P. (2003). Assembly of cell regulatory systems through protein interaction domains. *Science* 300, 445-452.

Perez-Iratxeta, C., and Andrade-Navarro, M.A. (2008). K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC Struct Biol* 8, 25.

Pierre, S., Bats, A.S., and Coumoul, X. (2011). Understanding SOS (Son of Sevenless). *Biochemical pharmacology* 82, 1049-1056.

Pollastri, G., and McLysaght, A. (2005). Porter: a new, accurate server for protein secondary structure prediction. *Bioinformatics* 21, 1719-1720.

Popovic, M., Jakobi, A.J., Rensen-de Leeuw, M., and Rehmann, H. (2013a). The guanine nucleotide exchange factor Rlf interacts with SH3 domain-containing proteins via a binding site with a preselected conformation. *J Struct Biol* 183, 312-319.

Popovic, M., Rensen-de Leeuw, M., and Rehmann, H. (2013b). Regulation and selectivity of exchange factors for G-proteins of the Ras family

Porfiri, E., and McCormick, F. (1996). Regulation of epidermal growth factor receptor signaling by phosphorylation of the ras exchange factor hSOS1. *The Journal of biological chemistry* 271, 5871-5877.

Priatel, J.J., Teh, S.J., Dower, N.A., Stone, J.C., and Teh, H.S. (2002). RasGRP1 transduces low-grade TCR signals which are critical for T cell development, homeostasis, and differentiation. *Immunity* 17, 617-627.

Provencher, S.W., and Glockner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* 20, 33-37.

Pufall, M.A., and Graves, B.J. (2002). Autoinhibitory domains: modular effectors of cellular regulation. *Annual review of cell and developmental biology* 18, 421-462.

Pulciani, S., Santos, E., Long, L.K., Sorrentino, E., and Barbacid, M. (1985). *ras* gene amplification and malignant transformation. *Mol Cell Biol* 5, 2836-2841.

Quilliam, L.A., Rebhun, J.F., and Castro, A.F. (2002). A growing family of guanine nucleotide exchange factors is responsible for activation of Ras-family GTPases. *Prog Nucleic Acid Res Mol Biol* 71, 391-444.

Radha, V., Rajanna, A., Gupta, R., Dayma, K., and Raman, T. (2008). The guanine nucleotide exchange factor, C3G regulates differentiation and survival of human neuroblastoma cells. *J Neurochem* 107, 1424-1435.

Radha, V., Rajanna, A., Mitra, A., Rangaraj, N., and Swarup, G. (2007). C3G is required for c-Abl-induced filopodia and its overexpression promotes filopodia formation. *Exp Cell Res* 313, 2476-2492.

Radha, V., Rajanna, A., and Swarup, G. (2004). Phosphorylated guanine nucleotide exchange factor C3G, induced by pervanadate and Src family kinases localizes to the Golgi and subcortical actin cytoskeleton. *BMC Cell Biol* 5, 31.

Rambaratsingh, R.A., Stone, J.C., Blumberg, P.M., and Lorenzo, P.S. (2003). RasGRP1 represents a novel non-protein kinase C phorbol ester signaling pathway in mouse epidermal keratinocytes. *The Journal of biological chemistry* 278, 52792-52801.

Reedquist, K.A., Fukazawa, T., Panchamoorthy, G., Langdon, W.Y., Shoelson, S.E., Druker, B.J., and Band, H. (1996). Stimulation through the T cell receptor induces Cbl association with Crk proteins and the guanine nucleotide exchange protein C3G. *J Biol Chem* 271, 8435-8442.

Rehmann, H., Arias-Palomo, E., Hadders, M.A., Schwede, F., Llorca, O., and Bos, J.L. (2008). Structure of Epac2 in complex with a cyclic AMP analogue and RAP1B. *Nature* 455, 124-127.

Rehmann, H., Das, J., Knipscheer, P., Wittinghofer, A., and Bos, J.L. (2006). Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state. *Nature* 439, 625-628.

Reuther, G.W., and Der, C.J. (2000). The Ras branch of small GTPases: Ras family members don't fall far from the tree. *Curr Opin Cell Biol* 12, 157-165.

Rice, S.A. (1956). Small angle scattering of X-rays. A. Guinier and G. Fournet. Translated by C. B. Wilson and with a bibliographical appendix by K. L. Yudowitch. Wiley, New York, 1955. 268 pp. *Journal of Polymer Science* 19, 594-594.

Rivas, G., Stafford, W., and Minton, A.P. (1999). Characterization of heterologous protein-protein interactions using analytical ultracentrifugation. *Methods* 19, 194-212.

Roberts, A.E., Allanson, J.E., Tartaglia, M., and Gelb, B.D. (2013). Noonan syndrome. *Lancet* 381, 333-342.

Roberts, A.E., Araki, T., Swanson, K.D., Montgomery, K.T., Schiripo, T.A., Joshi, V.A., Li, L., Yassin, Y., Tamburino, A.M., Neel, B.G., *et al.* (2007). Germline gain-of-function mutations in SOS1 cause Noonan syndrome. *Nature genetics* 39, 70-74.

Rojas, J.M., and Santos, E. (2002). *ras* genes and human cancer: different implications and different roles. *Curr Genom* 3, 295-311.

Rost, B., and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *Journal of molecular biology* 232, 584-599.

Rost, B., and Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 19, 55-72.

Rozakis-Adcock, M., van der Geer, P., Mbamalu, G., and Pawson, T. (1995). MAP kinase phosphorylation of mSos promotes dissociation of mSos1-Shc and mSos1-EGF receptor complexes. *Oncogene* 11, 1417-1426.

- Sakkab, D., Lewitzky, M., Posern, G., Schaeper, U., Sachs, M., Birchmeier, W., and Feller, S.M. (2000). Signaling of hepatocyte growth factor/scatter factor (HGF) to the small GTPase Rap1 via the large docking protein Gab1 and the adapter protein CRKL. *J Biol Chem* 275, 10772-10778.
- Sambrook, J., Fritsch, E., and Maniatis, T. (1989). *Molecular cloning: A laboratory manual*. 2nd edn New York: Cold Spring Harbor Laboratory Press.
- Santos, E., and Nebreda, A.R. (1989). Structural and functional properties of *ras* proteins. *FASEB J* 3, 2151-2163.
- Santos, E., Tronick, S.R., Aaronson, S.A., Pulciani, S., and Barbacid, M. (1982). T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB-and harvey-MSV transforming genes. *Nature* 298, 343-347.
- Scita, G., Tenca, P., Areces, L.B., Tocchetti, A., Frittoli, E., Giardina, G., Ponzanelli, I., Sini, P., Innocenti, M., and Di Fiore, P.P. (2001). An effector region in Eps8 is responsible for the activation of the Rac-specific GEF activity of Sos-1 and for the proper localization of the Rac-based actin-polymerizing machine. *The Journal of cell biology* 154, 1031-1044.
- Schiffer, M., and Edmundson, A.B. (1967). Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys J* 7, 121-135.
- Scholler, J.K., Perez-Villar, J.J., O'Day, K., and Kanner, S.B. (2000). Engagement of the T lymphocyte antigen receptor regulates association of son-of-sevenless homologues with the SH3 domain of phospholipase Cgamma1. *European journal of immunology* 30, 2378-2387.
- Schubbert, S., Shannon, K., and Bollag, G. (2007). Hyperactive Ras in developmental disorders and cancer. *Nat Rev Cancer* 7, 295-308.
- Schuck, P. (2010). Some statistical properties of differencing schemes for baseline correction of sedimentation velocity data. *Analytical biochemistry* 401, 280-287.
- Schuck, P., Perugini, M.A., Gonzales, N.R., Howlett, G.J., and Schubert, D. (2002). Size-distribution analysis of proteins by analytical ultracentrifugation: strategies and application to model systems. *Biophys J* 82, 1096-1111.
- Schuck, P., and Rossmanith, P. (2000). Determination of the sedimentation coefficient distribution by least-squares boundary modeling. *Biopolymers* 54, 328-341.
- Shivakrupa, R., Radha, V., Sudhakar, C., and Swarup, G. (2003). Physical and functional interaction between Hck tyrosine kinase and guanine nucleotide exchange factor C3G results in apoptosis, which is independent of C3G catalytic domain. *J Biol Chem* 278, 52188-52194.
- Sicheri, F., and Kuriyan, J. (1997). Structures of Src-family tyrosine kinases. *Curr Opin Struct Biol* 7, 777-785.
- Simon, M.A., Dodson, G.S., and Rubin, G.M. (1993). An SH3-SH2-SH3 protein is required for p21Ras1 activation and binds to sevenless and Sos proteins in vitro. *Cell* 73, 169-177.

- Soisson, S.M., Nimnual, A.S., Uy, M., Bar-Sagi, D., and Kuriyan, J. (1998). Crystal structure of the Dbl and pleckstrin homology domains from the human Son of sevenless protein. *Cell* *95*, 259-268.
- Sondermann, H., Soisson, S.M., Bar-Sagi, D., and Kuriyan, J. (2003). Tandem histone folds in the structure of the N-terminal segment of the ras activator Son of Sevenless. *Structure* *11*, 1583-1593.
- Sondermann, H., Soisson, S.M., Boykevisch, S., Yang, S.S., Bar-Sagi, D., and Kuriyan, J. (2004). Structural analysis of autoinhibition in the Ras activator Son of sevenless. *Cell* *119*, 393-405.
- Sreerama, N., Venyaminov, S.Y., and Woody, R.W. (1999). Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci* *8*, 370-380.
- Sreerama, N., and Woody, R.W. (1993). A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Analytical biochemistry* *209*, 32-44.
- Sreerama, N., and Woody, R.W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical biochemistry* *287*, 252-260.
- Stork, P.J., and Dillon, T.J. (2005). Multiple roles of Rap1 in hematopoietic cells: complementary versus antagonistic functions. *Blood* *106*, 2952-2961.
- Stryer, L. (1978). Fluorescence energy transfer as a spectroscopic ruler. *Annual review of biochemistry* *47*, 819-846.
- Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* *25*, 495-503.
- Takai, Y., Sasaki, T., and Matozaki, T. (2001). Small GTP-binding proteins. *Physiol Rev* *81*, 153-208.
- Tanaka, S., and Hanafusa, H. (1998). Guanine-nucleotide exchange protein C3G activates JNK1 by a ras-independent mechanism. JNK1 activation inhibited by kinase negative forms of MLK3 and DLK mixed lineage kinases. *J Biol Chem* *273*, 1281-1284.
- Tanaka, S., Morishita, T., Hashimoto, Y., Hattori, S., Nakamura, S., Shibuya, M., Matuoka, K., Takenawa, T., Kurata, T., Nagashima, K., *et al.* (1994). C3G, a guanine nucleotide-releasing protein expressed ubiquitously, binds to the Src homology 3 domains of CRK and GRB2/ASH proteins. *Proc Natl Acad Sci U S A* *91*, 3443-3447.
- Tanaka, S., Ouchi, T., and Hanafusa, H. (1997). Downstream of Crk adaptor signaling pathway: activation of Jun kinase by v-Crk through the guanine nucleotide exchange protein C3G. *Proc Natl Acad Sci U S A* *94*, 2356-2361.
- Tartaglia, M., Pennacchio, L.A., Zhao, C., Yadav, K.K., Fodale, V., Sarkozy, A., Pandit, B., Oishi, K., Martinelli, S., Schackwitz, W., *et al.* (2007). Gain-of-function SOS1 mutations cause a distinctive form of Noonan syndrome. *Nature genetics* *39*, 75-79.

- Terasawa, M., Uruno, T., Mori, S., Kukimoto-Niino, M., Nishikimi, A., Sanematsu, F., Tanaka, Y., Yokoyama, S., and Fukui, Y. (2012). Dimerization of DOCK2 is essential for DOCK2-mediated Rac activation and lymphocyte migration. *PLoS One* 7, e46277.
- Trudeau, T., Nassar, R., Cumberworth, A., Wong, E.T., Woollard, G., and Gsponer, J. (2013). Structure and intrinsic disorder in protein autoinhibition. *Structure* 21, 332-341.
- Uemura, N., and Griffin, J.D. (1999). The adapter protein Crkl links Cbl to C3G after integrin ligation and enhances cell migration. *J Biol Chem* 274, 37525-37532.
- Utreras, E., Henriquez, D., Contreras-Vallejos, E., Olmos, C., Di Genova, A., Maass, A., Kulkarni, A.B., and Gonzalez-Billault, C. (2013). Cdk5 regulates Rap1 activity. *Neurochem Int* 62, 848-853.
- van den Berghe, N., Cool, R.H., Horn, G., and Wittinghofer, A. (1997). Biochemical characterization of C3G: an exchange factor that discriminates between Rap1 and Rap2 and is not inhibited by Rap1A(S17N). *Oncogene* 15, 845-850.
- Van Der Meer, B.W., Coker, G. III, and Simon Chen, S.-Y (1994). *Resonance Energy Transfer: Theory and Data* New York: VCH Publishers, Inc.
- van Stokkum, I.H., Spoelder, H.J., Bloemendal, M., van Grondelle, R., and Groen, F.C. (1990). Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Analytical biochemistry* 191, 110-118.
- Vetter, I.R., and Wittinghofer, A. (2001). The guanine nucleotide-binding switch in three dimensions. *Science* 294, 1299-1304.
- Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *Journal of Applied Crystallography* 36, 860-864.
- Voss, A.K., Britto, J.M., Dixon, M.P., Sheikh, B.N., Collin, C., Tan, S.S., and Thomas, T. (2008). C3G regulates cortical neuron migration, preplate splitting and radial glial cell attachment. *Development* 135, 2139-2149.
- Voss, A.K., Gruss, P., and Thomas, T. (2003). The guanine nucleotide exchange factor C3G is necessary for the formation of focal adhesions and vascular maturation. *Development* 130, 355-367.
- Voss, A.K., Krebs, D.L., and Thomas, T. (2006). C3G regulates the size of the cerebral cortex neural precursor population. *EMBO J* 25, 3652-3663.
- Wang, W., Fisher, E.M., Jia, Q., Dunn, J.M., Porfiri, E., Downward, J., and Egan, S.E. (1995). The Grb2 binding domain of mSos1 is not required for downstream signal transduction. *Nature genetics* 10, 294-300.
- Wang, Z., Dillon, T.J., Pokala, V., Mishra, S., Labudda, K., Hunter, B., and Stork, P.J. (2006). Rap1-mediated activation of extracellular signal-regulated kinases by cyclic AMP is dependent on the mode of Rap1 activation. *Mol Cell Biol* 26, 2130-2145.
- Waters, S.B., Holt, K.H., Ross, S.E., Syu, L.J., Guan, K.L., Saltiel, A.R., Koretzky, G.A., and Pessin, J.E. (1995). Desensitization of Ras activation by a feedback disassociation of the SOS-Grb2 complex. *The Journal of biological chemistry* 270, 20883-20886.

- Waxman, E., Laws, W., Laue, T., and Ross, J.B.A. (1994). Refining Hydrodynamic Shapes of Proteins: The Combination of Data From Analytical Ultracentrifugation and Time-Resolved Fluorescence Anisotropy Decay. In *Modern Analytical Ultracentrifugation*, T. Schuster, and T. Laue, eds. (Birkhäuser Boston), pp. 189-205.
- Weis, K. (2003). Regulating access to the genome: nucleocytoplasmic transport throughout the cell cycle. *Cell* *112*, 441-451.
- Wennerberg, K., Rossman, K.L., and Der, C.J. (2005). The Ras superfamily at a glance. *Journal of cell science* *118*, 843-846.
- Wessel, D., and Flugge, U.I. (1984). A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Analytical biochemistry* *138*, 141-143.
- Whitmore, L., and Wallace, B.A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* *32*, W668-673.
- Whitmore, L., and Wallace, B.A. (2008). Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* *89*, 392-400.
- Wilson, C., Wardell, M.R., Weisgraber, K.H., Mahley, R.W., and Agard, D.A. (1991). Three-dimensional structure of the LDL receptor-binding domain of human apolipoprotein E. *Science* *252*, 1817-1822.
- Williamson, T.E., Craig, B.A., Kondrashkina, E., Bailey-Kellogg, C., and Friedman, A.M. (2008). Analysis of self-associating proteins by singular value decomposition of solution scattering data. *Biophys J* *94*, 4906-4923.
- Wittinghofer, A., and Vetter, I.R. (2011). Structure-function relationships of the G domain, a canonical switch motif. *Annual review of biochemistry* *80*, 943-971.
- Woody, R.W. (1995). Circular dichroism. *Methods in enzymology* *246*, 34-71.
- Wriggers, W., Milligan, R.A., and McCammon, J.A. (1999). Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J Struct Biol* *125*, 185-195.
- Wu, C., Lai, C.F., and Mobley, W.C. (2001). Nerve growth factor activates persistent Rap1 signaling in endosomes. *J Neurosci* *21*, 5406-5416.
- Wu, C., Ma, M., Brown, K., Geisler, M., Li, L., Tzeng, E., Jia, C.Y., Jurisica, I., and Li, S.S. (2007). Systematic identification of SH3 domain-mediated human protein-protein interactions by peptide array target screening. *Proteomics* *7*, 1775-1785.
- Wynne, J.P., Wu, J., Su, W., Mor, A., Patsoukis, N., Boussiotis, V.A., Hubbard, S.R., and Philips, M.R. (2012). Rap1-interacting adapter molecule (RIAM) associates with the plasma membrane via a proximity detector. *The Journal of cell biology* *199*, 317-330.
- Xu, W., Doshi, A., Lei, M., Eck, M.J., and Harrison, S.C. (1999). Crystal structures of c-Src reveal features of its autoinhibitory mechanism. *Mol Cell* *3*, 629-638.

- Yadav, K.K., and Bar-Sagi, D. (2010). Allosteric gating of Son of sevenless activity by the histone domain. *Proceedings of the National Academy of Sciences of the United States of America* 107, 3436-3440.
- Yang, J., Dokurno, P., Tonks, N.K., and Barford, D. (2001). Crystal structure of the M-fragment of alpha-catenin: implications for modulation of cell adhesion. *The EMBO journal* 20, 3645-3656.
- Yang, J., Zhang, Z., Roe, S.M., Marshall, C.J., and Barford, D. (2009). Activation of Rho GTPases by DOCK exchange factors is mediated by a nucleotide sensor. *Science* 325, 1398-1402.
- Yasuda, S., Stevens, R.L., Terada, T., Takeda, M., Hashimoto, T., Fukae, J., Horita, T., Kataoka, H., Atsumi, T., and Koike, T. (2007). Defective expression of Ras guanyl nucleotide-releasing protein 1 in a subset of patients with systemic lupus erythematosus. *Journal of immunology* 179, 4890-4900.
- Yokote, K., Hellman, U., Ekman, S., Saito, Y., Ronnstrand, L., Saito, Y., Heldin, C.H., and Mori, S. (1998). Identification of Tyr-762 in the platelet-derived growth factor alpha-receptor as the binding site for Crk proteins. *Oncogene* 16, 1229-1239.
- York, R.D., Yao, H., Dillon, T., Ellig, C.L., Eckert, S.P., McCleskey, E.W., and Stork, P.J. (1998). Rap1 mediates sustained MAP kinase activation induced by nerve growth factor. *Nature* 392, 622-626.
- Zahedi, B., Goo, H.J., Beaulieu, N., Tazmini, G., Kay, R.J., and Cornell, R.B. (2011). Phosphoinositide 3-kinase regulates plasma membrane targeting of the Ras-specific exchange factor RasGRP1. *The Journal of biological chemistry* 286, 12712-12723.
- Zerial, M., and McBride, H. (2001). Rab proteins as membrane organizers. *Nature reviews Molecular cell biology* 2, 107-117.
- Zhang, J., Campbell, R.E., Ting, A.Y., and Tsien, R.Y. (2002). Creating new fluorescent probes for cell biology. *Nature reviews Molecular cell biology* 3, 906-918.
- Zhang, Z.M., Simmerman, J.A., Guibao, C.D., and Zheng, J.J. (2008). GIT1 paxillin-binding domain is a four-helix bundle, and it binds to both paxillin LD2 and LD4 motifs. *The Journal of biological chemistry* 283, 18685-18693.
- Zhao, C., Du, G., Skowronek, K., Frohman, M.A., and Bar-Sagi, D. (2007). Phospholipase D2-generated phosphatidic acid couples EGFR stimulation to Ras activation by Sos. *Nat Cell Biol* 9, 706-712.
- Zheng, J., Chen, R.H., Corblan-Garcia, S., Cahill, S.M., Bar-Sagi, D., and Cowburn, D. (1997). The solution structure of the pleckstrin homology domain of human SOS1. A possible structural role for the sequential association of diffuse B cell lymphoma and pleckstrin homology domains. *The Journal of biological chemistry* 272, 30340-30344.
- Zhu, K., Debreceni, B., Bi, F., and Zheng, Y. (2001). Oligomerization of DH domain is essential for Dbl-induced transformation. *Molecular and cellular biology* 21, 425-437.
- Zorrilla, S., Rivas, G., Acuna, A.U., and Lillo, M.P. (2004). Protein self-association in crowded protein solutions: a time-resolved fluorescence polarization study. *Protein Sci* 13, 2960-2969.

