







# **UNIVERSIDAD DE SALAMANCA**

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

## **Mechanisms of auto-regulation and activation of the guanine nucleotide exchange factor C3G**

---

**MEMORIA PARA OPTAR AL GRADO DE DOCTOR  
PRESENTADA POR**

**Arturo Carabias del Rey**

Bajo la dirección de los Doctores

**José M de Pereda Vega**

**Carmen Guerrero Arroyo**

Salamanca, 2019

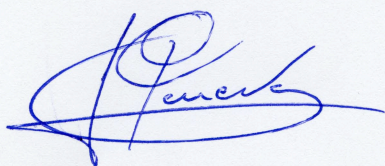


**Dr. JOSÉ MARÍA DE PEREDA VEGA**, Científico Titular del Consejo Superior de Investigaciones Científicas (CSIC), y **Dra. CARMEN GUERRERO ARROYO**, Profesora Titular del Departamento de Medicina en el Instituto de Biología Molecular y Celular del Cáncer (IBMCC, CSIC-Universidad de Salamanca).

**CERTIFICAN:**

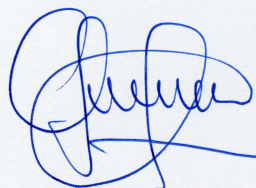
Que D. Arturo Carabias del Rey, graduado en Biología por la Universidad de Salamanca, ha realizado bajo su dirección el trabajo de Tesis Doctoral que lleva por título "***Mechanisms of auto-regulation and activation of the guanine nucleotide exchange factor C3G***", y considera que éste reúne originalidad y contenidos suficientes para que sea presentada ante el Tribunal correspondiente y optar al Grado de Doctor por la Universidad de Salamanca.

Y para que así conste a los efectos oportunos, expide el presente certificado en Salamanca a 24 de Junio de 2019



Dr. José María de Pereda Vega

Director de la tesis



Dra. Carmen Guerrero Arroyo

Co-directora de la tesis



Arturo Carabias del Rey ha realizado esta tesis doctoral siendo beneficiario de una ayuda del programa de la Fundación Dr. Moraza de retención de talento (BEC-MORAZA-14), durante el periodo de noviembre de 2014 a diciembre de 2014; un contrato de Personal Investigador en Formación de la Universidad de Salamanca, durante el periodo de enero de 2015 a agosto de 2015; una ayuda predoctoral del programa de Formación de Profesorado Universitario (FPU014/0625) del Ministerio de Ciencia, Innovación y Universidades, durante el periodo de septiembre de 2015 a enero de 2019; y una ayuda del programa de la Fundación Dr. Moraza de retención de talento (BEC-MORAZA-18), durante el periodo de enero de 2019 hasta la actualidad.

Este trabajo se ha enmarcado dentro de los proyectos del Plan Estatal I+D+i “*Bases estructurales y mecánicas de la regulación del activador de Rap1 C3G*” (BFU2015-69499-P) y “*Función de C3G en el desarrollo tumoral y en la patofisiología del hígado. Implicación del C3G plaquetario en la angiogénesis y en enfermedades hepáticas y cardiovasculares*” (SAF2016-76588-C2-2-R) financiados por el Ministerio de Ciencia, Innovación y Universidades; y del proyecto del Programa de Apoyo a Proyectos de Investigación de la Consejería de Educación de la Junta de Castilla y León titulado “*Papel de C3G en la regulación de la función plaquetaria: Implicaciones en angiogénesis y aplicación al diagnóstico y tratamiento de la enfermedad trombotica*” (SA017U16). Estos proyectos han sido co-financiados por el Fondo Europeo de Desarrollo Regional (FEDER). El Centro de Investigación del Cáncer (IBMCC-CIC), dónde se ha realizado este trabajo, recibe financiación del Programa de Apoyo a Planes Estratégicos de Investigación de Estructuras de Investigación de Excelencia co-financiado por la Junta de Castilla y León y FEDER (CLC-2017-01).

Arturo Carabias del Rey realizó parte del trabajo de esta tesis durante dos estancias de tres meses de duración total en el laboratorio de la Dra. Sandra de Macedo Ribeiro en el Instituto de Investigação e Inovação em Saúde (i3S) (Oporto, Portugal). Parte de esa estancia estuvo financiada por una ayuda de Estancia Breve (referencia EST17/00446) del Ministerio de Ciencia, Innovación y Universidades.





*A mi padre*

*"No te olvides de ser feliz"*

(Javier Carabias)



## **Acknowledgements**

First, I would like to thank my mentors, Dr. José M de Pereda and Dra. Carmen Guerrero, for giving me the opportunity to learn in their lab during these years. I would also like to extend my thanks to the rest of the members of their lab.

An important part of this work was performed in two stays at the Dra. Sandra de Macedo Ribeiro lab. I would like to offer my special thanks to her, for supporting me during my time at i3S, the warm welcome, and hospitality during my stay. The acknowledgement should be extended to the rest of the members of her lab, specially to Dr. Pedro J.B. Pereira, and Dr. Jorge Ripoll-Rozada, for the critical reading of the manuscript and evaluation; and Dra. Zsuzsa Sárkány, Dr. José Antonio Manso, and Joana Fraga for their help with the experimental work.

I wish to acknowledge the assistance given by Dr. Carlos Alfonso Botello and Dr. Juan Ramon Luque-Ortega with the SEC-MALS and analytical ultracentrifugation experiments. Particularly, I would like to express my very great appreciation to the analysis of the ITC and AUC data performed by Dr. JR Luque-Ortega, and his willingness to collaborate with us in this project.

## **Agradecimientos**

Quiero mostrar mi agradecimiento a toda la gente que me ha acompañado durante estos años. Inicialmente a mis directores, Chema y Carmen, por acogerme en su laboratorio. Por enseñarme a pensar más claramente, apoyarme cuando al principio del proyecto nada funcionaba, por hacer del laboratorio un lugar de trabajo agradable, y por revisar todos los experimentos y manuscritos como si fueran suyos. Sois mis padres científicos, y espero no perder la amistad que hemos forjado.

Continuo agradeciendo a mis compañeros de juegos, que además de ser muy competentes, son grandes personas. A Jam, por ser mi apoyo desde el principio, discutir de ciencia y resolver mis dudas, y darme casa cuando la necesité. También a Jamming por hacer de mi cama un arenero y tratar de inundar el baño mientras me duchaba. A mis compañeros de lucha contra C3G, Sergio y Antonio. Sergio, gracias por escucharme siempre, apaciguarme, no rendirte nunca, darme tu opinión sincera, recomendarme grandes películas, tus postres en los seminarios, calentarme los pies

en las noches frías, ... , y enseñarme como trabajar y vivir, sin que una cosa no quite la otra. Antonio, gracias por tu apoyo incondicional, tu frescura, tu predisposición a hacerlo todo, y por atreverte a seguir.

También agradezco al resto de compañeros. A Sara G, por ser siempre amable, a Sara O, por no ser siempre amable, a Víctor por acompañarme desde el principio, a Luís por escuchar mis chistes malos, a Cristina por su sonrisa y su calma aunque fueran momentos difíciles, y Alba por darme a probar la cocina de "mama". A los chicos "especiales" del L19. A Curro, por debatir incansablemente, a Oscar por disfrutar de valiosas conversaciones y momentos musicales, a Santi por ser mi novio durante el máster, a Elena por ser tan luchadora y a Alberto por ser un buen chaval y enseñarme la palabra "troleo". Por supuesto, no puedo olvidar a la pandilla del sincrotón. Especialmente a Rubén, por enseñarme que para ser "presumía" hay que ponerse el traje de los domingos "to los días".

A la señorita Morejón, por hacerme moratones por las noches, por enseñarme a bailar jota, por preocuparse por mis uñas, por enseñarme como se dice "foco" en inglés, y ser tan valiente como inconsciente. Al resto de amigos del L4, Helena por enseñarme a bailar salsa, y a Eva, Nacho y Raúl, por las noches divertidas. También a Patri, Sonia, y Carlos por las conversaciones de los pasillos.

Muchas gracias a la Perrina, por aguantar mis desvaríos, mis largas conversaciones en la ducha y mis episodios de sonambulismo, también gracias por discutirme todo (que lo haces muy bien), no rendirme, ayudarme, apoyarme y quererme incondicionalmente. Has sido mi compañera desde el principio. Si esto no nos ha matado es que estamos más gordos.

Finalmente, gracias a mis padres, Javi y Caty. He tenido una gran suerte. Por darme vuestro amor siempre. Por criarme rodeado de cultura. Por dejarme elegir y ser yo mismo. Por creer en mí. Por enseñarme que la única forma de conseguir las cosas es hacerlas con amor y esmero, y que lo más importante en la vida es ser feliz.

## List of Abbreviations

a.c.s.	Average Conservation Score
$A_{280}$	Absorbance at 280 nm
AUC	Analytical Ultracentrifugation
BSA	Bovine Serum Albumin
C3G-BR	Cdc25H-Binding Region
C3G-IT	Cdc25H-Inhibitory Tail
CML	Chronic Myeloid Leukemia
cNBD	cyclic Nucleotide-Binding Domain
$D$	Diffusion coefficient
DAG	Diacylglycerol
$\Delta H$	Binding enthalpy
DEP	Disheveled, Egl-10, Pleckstrin
DH	Dbl Homology
DLS	Dynamic Light Scattering
DTT	Dithiothreitol
EC50	Half maximal effective concentration
EDTA	Ethylenediaminetetraacetic acid
EF1	Elongation Factor-1
GEF	Guanine nucleotide Exchange Factor
GST	Glutathione S-transferase
HCC	Hepatocellular carcinoma
HMM	Hidden Markov Model
HPLC	High-Performance Liquid Chromatography
IC50	The half maximal inhibitory concentration
IMAC	Immobilized Metal Affinity Chromatography
ITC	Isothermal Titration Calorimetry
$k_d$	Dissociation Constant
$k_{obs}$	Nucleotide dissociation rate constant from the GTPase
$k_{off}$	Dissociation rate
Kozak	Kozak consensus sequence
MALS	Multiangle light scattering
MCS	Multi-cloning site
mEGFP	Monomeric Enhanced Green Fluorescent Protein
MSA	Multiple Sequence Alignment
MW	Molecular Weight
N	The stoichiometry constant
NaPi	sodium phosphate buffer
NCBI	National Center for Biotechnology Information
NTD	N-Terminal Domain
$OD_{600nm}$	Optical density 600 nm
PD	Pull Down
PDB	Protein Data Bank

PH	Pleckstrin Homology
PIP <sub>2</sub>	Phosphatidylinositol biphosphate
PIP <sub>3</sub>	Phosphatidylinositol triphosphate
RA	Ras Association
<i>Rbs</i>	Ribosome binding site
REM	Ras exchange motif
$R_g$	Radius of gyration
RTK	Receptor tyrosine
<i>s</i>	Sedimentation coefficient
$s_{(20,w)}$	Corrected Sedimentation coefficient in water at 20 °C
SEC	Size-Exclusion Chromatography
SFKs	Src Family Kinases
SH2	Src Homology 2
SH3	Src Homology 3
SH3b	SH3-binding
SNVs	Single Nucleotide Variants
SrcKD	Src Kinase Domain
TCR	T-Cell Receptor
TEV	Tobacco Etch Virus
WB	Western Blot
WCL	Whole Cell Lysates

## Table of contents

<b>ACKNOWLEDGEMENTS</b> .....	<b>i</b>
<b>LIST OF ABBREVIATIONS</b> .....	<b>iii</b>
<b>TABLE OF CONTENTS</b> .....	<b>v</b>
<b>INTRODUCTION</b> .....	<b>1</b>
1. Ras GTPase family and their GEFs .....	3
The GTPase switch .....	3
GEFs of the Ras family .....	5
Mechanisms of regulation of GEFs .....	7
SOS .....	7
RasGRP1.....	9
Epac.....	10
2. C3G .....	12
Primary structure and interactions .....	12
C3G isoforms and tissue expression .....	13
C3G functions.....	15
Regulation of cell adhesion .....	15
Role in haematopoietic cells.....	15
C3G pathological functions .....	16
Cancer .....	16
Participation in the liver physiopathology .....	17
Other diseases.....	18
Adaptor proteins .....	18
Regulation and activation of C3G .....	21
Summary .....	23
<b>OBJECTIVES</b> .....	<b>25</b>
<b>METHODS</b> .....	<b>29</b>
1. cDNA .....	31
2. Constructs overview .....	31
3. Vectors and cloning .....	33
Cloning into pET22b-x2.....	34
Cloning into pETEV15b and pETEV15b-NcoI.....	34
Cloning into pGEX-TEV and pGEX-2xTEV-cHis.....	34
Cloning into pETEV15b-Halo-Avi .....	36

Cloning into pETDuet .....	36
Cloning into pEF1-mEGFP and pEF1-mEGFP-CAAX .....	36
Cloning into pCEFLHA .....	37
4. C3G mutants .....	37
5. Bacterial protein expression and purification .....	39
Bacterial protein expression .....	39
Purification of His-tag proteins .....	39
Purification of GST-tag proteins .....	40
Optimization of C3G full-length purification .....	41
GST-C3G-His two-step affinity purification .....	42
His-Halo-C3G purification .....	43
Protein concentration measurement .....	43
6. Protein modification .....	43
Biotinylation .....	43
Src-mediated phosphorylation .....	43
7. Nucleotide exchange experiments .....	44
Mant-dGDP loading of Rap1b .....	44
Fluorescence measurements and analysis of dissociation kinetics .....	45
Analysis of dose response experiments .....	45
8. Sequence and structural analysis .....	46
Structural analysis .....	46
C3G orthologs search with HMMER .....	46
Evolutionary conservation analysis .....	47
9. Size exclusion chromatography coupled to multiple angle light scattering .....	47
10. Analytical ultracentrifugation and dynamic light scattering .....	47
11. Isothermal titration calorimetry .....	49
12. Global analysis of the C3G-CrkL interaction with SEDPHAT software .....	50
13. Protein electrophoresis, Western blot and antibodies .....	50
14. Pull Down assays for the analysis of protein-protein interactions <i>in vitro</i> .....	51
15. Mammalian cell culture and transfection .....	51
16. Mammalian cell lysis, GST-pull downs and Rap1 activation assays .....	52
Rap1 activation assays in cells .....	52
<b>RESULTS .....</b>	<b>53</b>
<i>RESULTS I: Mechanisms of auto-regulation of C3G .....</i>	<b>53</b>
1. C3G is autoinhibited .....	55



2. An interaction between the SH3b and Cdc25H domains of C3G inhibits the GEF activity	56
3. The C-terminal segment of the SH3b is sufficient to inhibit the catalytic activity of the Cdc25H domain	58
4. Mapping residues in the SH3b important for binding to the Cdc25H domain	61
5. Disruption of the SH3b/Cdc25H interaction induces constitutive activation of C3G <i>in vitro</i> and in cell cultures	64
6. Mapping of the SH3b binding surface in the Cdc25H domain	67
7. CrkL binding and phosphorylation by Src increase the GEF activity of C3G	71
8. CrkL binding to SH3b disrupts the SH3b/Cdc25H inhibitory interaction while phosphorylation of SH3b does not	74
9. The NTD/REM interaction positively regulates the GEF activity and contributes to the stability of recombinant C3G	77
10. Analysis of C3G activation in cell cultures	84
<b>RESULTS II: Analysis of C3G/CrkL interaction</b>	<b>87</b>
1. Only one molecule of CrkL binds to C3G by SEC-MALS	89
2. Study of C3G-CrkL complex by Isothermal Titration Calorimetry (ITC)	92
3. Analysis of CrkL binding to C3G by sedimentation velocity	94
4. CrkL binds preferentially to the P1 and P2 motifs of C3G	97
5. C3G is differentially activated by CrkL and CrkII proteins <i>in vitro</i>	102
<b>DISCUSSION</b>	<b>107</b>
1. Autoinhibition of C3G: a lid-lock mechanism	109
2. Crk binding and phosphorylation of C3G	110
3. NTD-REM interaction positively regulates the GEF activity	115
4. Multilayered regulation of C3G	117
5. Missense variants of C3G in cancer	118
6. Concluding remarks: model of activation of C3G	121
<b>CONCLUSIONS</b>	<b>123</b>
<b>BIBLIOGRAPHY</b>	<b>127</b>
<b>APPENDICES</b>	<b>139</b>
<i>APPENDIX I: Primers</i>	<b>141</b>
<i>APPENDIX II: Vectors</i>	<b>149</b>
<i>APPENDIX III: C3G orthologs</i>	<b>153</b>
<b>LIST OF FIGURES</b>	<b>157</b>
<b>LIST OF TABLES</b>	<b>159</b>



# *INTRODUCTION*



## 1. RAS GTPase family and their GEFs

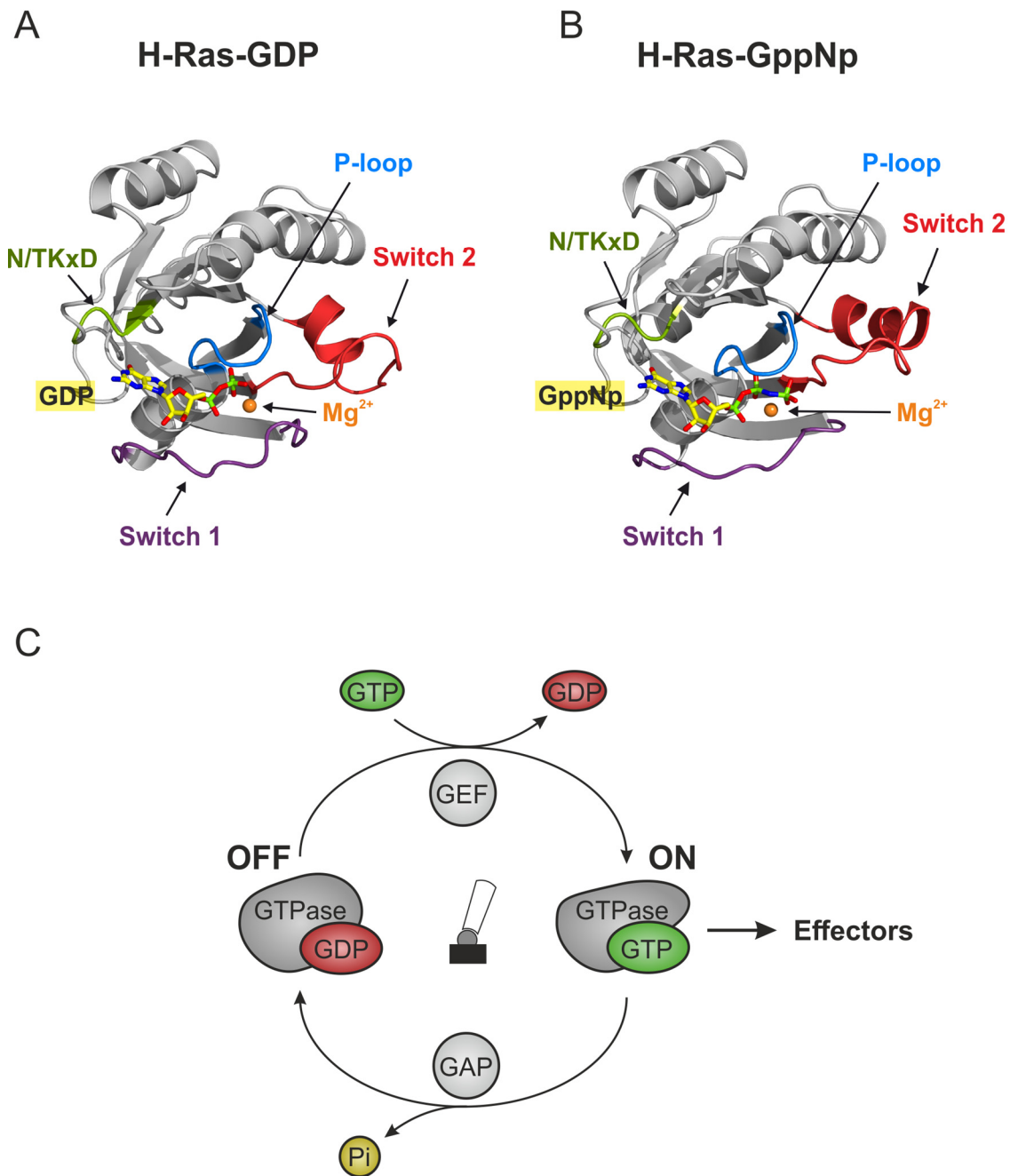
Cells within the human body must grow, divide and specialize to perform specific tasks. These processes are usually regulated by small GTPases of Ras superfamily, which integrate input signals and trigger different output responses to determine the cell fate (Iwig *et al.* 2013).

Within the Ras superfamily, the Ras family comprises four main subfamilies (Ras, Rap, R-Ras and Ral proteins). GTPases of the Ras family regulate cell growth, differentiation, and survival [reviewed in (Reuther *et al.* 2000, Cherfils *et al.* 2013)]. Since the discovery of Ras as the first human oncogene (Santos *et al.* 1982), activating mutations affecting Ras have been found in 30% of all tumors. Afterwards, other GTPases were also found to contribute to the development of cancer and other diseases, such as developmental diseases [reviewed in (Simanshu *et al.* 2017)]. This reinforces the idea that GTPases must be accurately regulated.

### The GTPase switch

All members of the Ras family share 30-50% of sequence identity and contain a G-domain (~20 kDa), which is responsible for the GTP hydrolysis (Bourne *et al.* 1991). Since most of the work about the GTPase switching cycle has been described for the Ras protein (H, N and K-Ras), for the rest of the members it is described as modifications of the canonical mechanism (Wittinghofer *et al.* 2011).

A common feature of all GTPases is that they can bind the nucleotides GDP and GTP. These nucleotides bind to the G-domain with very high affinity (usually in the picomolar or nanomolar range) (Klebe *et al.* 1995, Lenzen *et al.* 1998). This is determined by the dual interaction of the guanine base of the nucleotide with a N/TKxD motif of the GTPase and the  $\beta$ - $\gamma$ -phosphate of the nucleotide with a  $Mg^{2+}$  ion and the P-loop of the GTPase (Figure I1A and B) (Cherfils *et al.* 2013). Other regions of the GTPases, called switch 1 and switch 2 regions, also interact with the nucleotide, experiencing conformational changes between the GDP- and GTP-bound forms. Additionally, they also participate in the GTP hydrolysis. Since most of the effectors bind to the GTP-bound form of the GTPase, this is considered the active form; on the other hand, the GDP-bound form is considered the inactive form. Therefore, GTPases function as molecular switches (Figure I1C).



**Figure I1. The GTPase switch.** (A) Structure of inactive H-Ras bound to GDP (PDB ID: 4Q21) (Milburn *et al.* 1990). (B) Structure of active H-Ras bound to the GTP analogue GppNp (PDB ID: 5P21) (Pai *et al.* 1990). (C) The GTPase switch cycle: the exchange of GDP by GTP is catalyzed by GEFs. GAPs increase the intrinsic GTPase activity leading to GTP hydrolysis and inactivation of the GTPase. Pi, inorganic phosphate.

Ras GTPases have very low intrinsic GTPase activity, for this reason they are also called "GTP-binding proteins". In a similar manner, the exchange of the GDP and GTP nucleotides from the GTPase is also very slow (several hours). So that, GTPases need other proteins, known as guanine nucleotide exchange factors (GEFs), and GTPase activating proteins (GAPs), to catalyze both processes (Figure I1C).

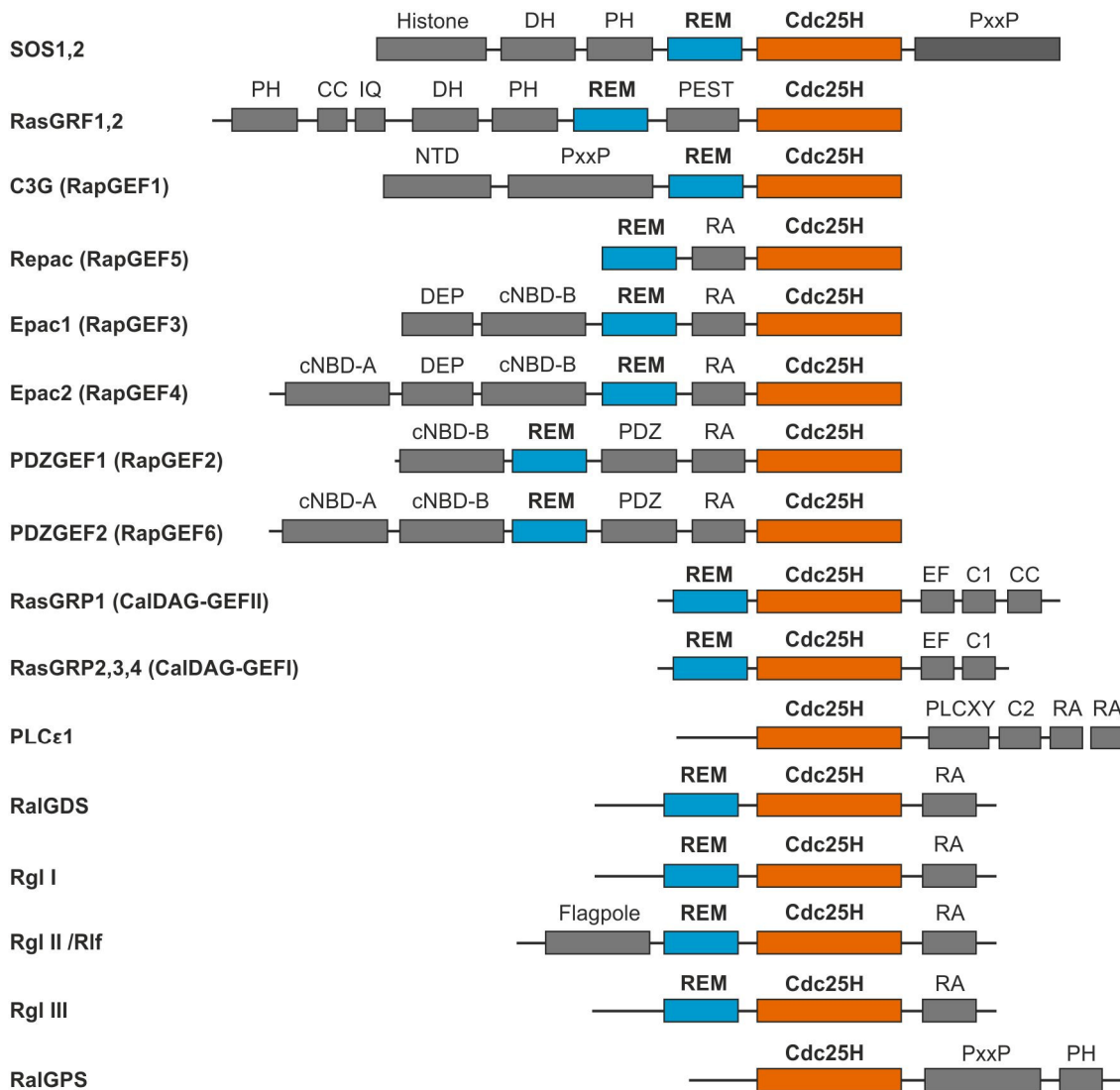
GEFs are proteins that interact with the GTPase and induce the destabilization of the bound nucleotide, accelerating the speed of the dissociation reaction. In the cell, upon the binding of the GTPase to the GEFs and the release of the GDP, the more abundant nucleotide in the cytosol, GTP, binds to the GTPase, switching it to the active form. On the other hand GAPs help the GTPase to acquire the right arrangement for the GTP hydrolysis. Thus, GEFs and GAPs have opposing roles in the control of the nucleotide switch.

## **GEFs of the Ras family**

GEFs of the Ras family contain a domain homologous to the Cdc25 gene of *S. cerevisiae* (Cdc25H). This domain is responsible for the binding to the GTPase and the destabilization of the bound nucleotide. This event has been described at the molecular level as a push-and-pull mechanism: the Cdc25H domain pulls the switch 2 region to a position close to the Mg<sup>2+</sup> ion, and the Mg<sup>2+</sup> is expelled by repulsion forces. This weakens the interaction of the nucleotide with the GTPase. Simultaneously, the switch 1 region is pushed by a Cdc25H helix, allowing the release of the nucleotide from the GTPase (Cherfils *et al.* 1999, Vetter *et al.* 2001). Therefore, the Cdc25H/GTPase interaction induces the sequential release of the phosphate groups and the nucleotide base, in a first step of a complex multistep reaction that ends up with the binding of GTP (Klebe *et al.* 1995, Lenzen *et al.* 1998).

Several GEF families have been described in the Ras family, such as SOS, RasGRF, C3G, Epac, RasGRP, PDZ-GEF, PLC $\epsilon$  and RalGEF families (Figure I2). These proteins have a modular structure with several domains. A common feature of almost all members is the presence of Ras exchange motif (REM) and Cdc25H domains. The Cdc25H domain of SOS is considered a canonical Cdc25H domain. It contains two  $\alpha$ -helices ( $\alpha$ H- $\alpha$ I) with an antiparallel arrangement projected out of the Cdc25H core, which are named helical-hairpin (Boriack-Sjodin *et al.* 1998). The orientation of the helical hairpin respect the Cdc25H core is essential for the interaction with the GTPase, and, therefore, for the GEF activity (Freedman *et al.* 2006). The REM does not contact the GTPase, but establishes a close-packed interaction with the helical-hairpin of the Cdc25H domain in all GEFs of known structure. It has been demonstrated for SOS and proposed for other Cdc25H-containing GEFs that the REM domain controls the GEF activity of the Cdc25H domain by modulating the orientation of the helical hairpin (Margarit *et al.* 2003, Boykevisch *et al.* 2006, Freedman *et al.*

2006). In addition to the REM and Cdc25H, a wide variety of domains with different functions are present in particular families. The non-catalytic domains frequently participate in intermolecular or intramolecular regulatory interactions (Figure I2) (Cherfils *et al.* 2013).



**Figure I2. Domain structure of GEFs of the Ras family.** The catalytic region of these GEFs is characterized by a Cdc25H domain that is frequently accompanied by a REM domain. Outside the REM-Cdc25H region, these GEFs contain different types of domains that in most cases are involved in the regulation of these proteins. DH, Dbl homology domain; PH, Pleckstrin homology domain; PxxP, Proline-rich domain, which binds to SH3 domains; CC, coiled-coil; IQ, Calmodulin binding domain; PEST, domain enriched in P,E,S and T residues; NTD, N-terminal domain; RA, Ras association domain; DEP, Dishevelled, Egl-10 and Pleckstrin domain; cNBD-A/B, cyclic nucleotide binding domains; PDZ, domain shared by PSD95, DlgA and ZO-1 proteins; C1, homologous to the DAG binding domain of protein kinase C1 (PKC1); PLCY, Phospholipase C domain; C2, Conserved domain of PKC2; flagpole,  $\beta$ -sheet enriched domain binding to SH3 domains. Adapted from (Vigil *et al.* 2010).



## Mechanisms of regulation of GEFs

Similarly to GTPases, the activity of the GEFs must be regulated in order to prevent uncontrolled signaling. Cdc25H GEFs are maintained in low GEF activity states via intramolecular interactions between catalytic and non catalytic regions that repress the GEF activity (Pufall *et al.* 2002). While this is a common feature of the mechanisms of regulation, the specific types of domains involved in the regulation, and the contacts that they establish with the catalytic region are highly specific. Here we describe the regulatory mechanisms of SOS, RasGRP1 and Epac to illustrate the diversity of the specific mechanisms and the common features.

### SOS

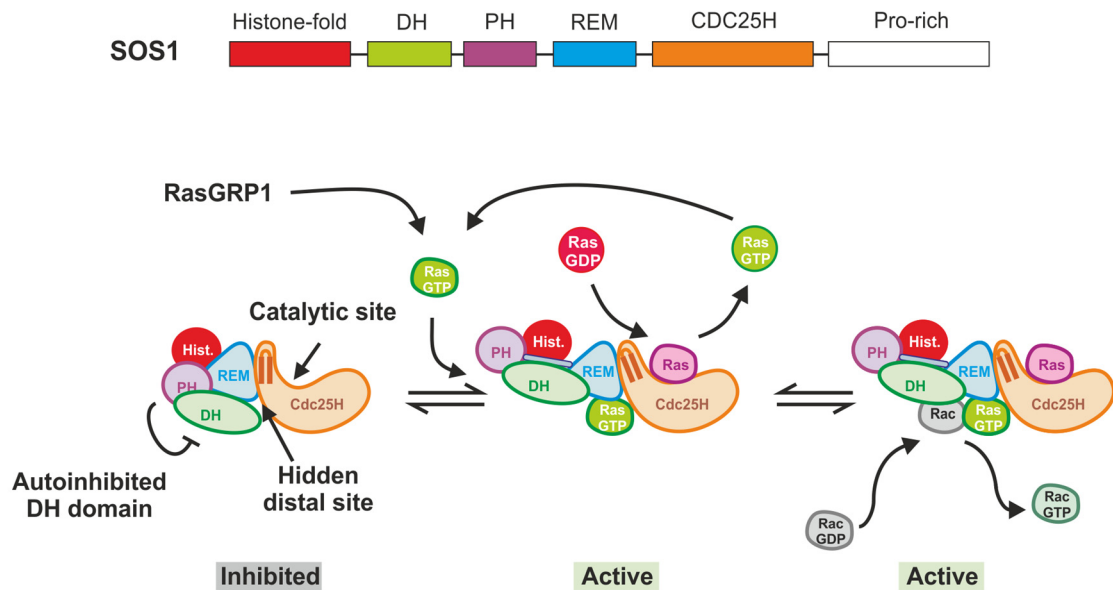
SOS is a GEF for the Ras subfamily of GTPases (Wang *et al.* 1995). SOS proteins have a modular structure with a histone-fold domain, a Dbl Homology (DH)-Pleckstrin Homology (PH) domain cassette, a helical linker, a REM domain, a Cdc25H domain and a C-terminal Proline-rich region (Figure I3). The histone-fold and PH domains bind phospholipids (PIP<sub>2</sub>, PIP<sub>3</sub> and phosphatidic acid). The DH domain is homologous to the catalytic domain of the Rho GTPase family and has been proposed to mediate the activation of Rac (Nimnual *et al.* 1998).

SOS has two binding sites for Ras molecules: a catalytic site and a distal site. Binding of Ras to the catalytic site (catalytic Ras) produces the release of the nucleotide from the GTPase (Boriack-Sjodin *et al.* 1998). On the other hand, the crystal structure of a SOS-Ras complex revealed a second Ras molecule bound to a distal site formed by the REM and Cdc25H domains (allosteric Ras) (Margarit *et al.* 2003). Ras-GTP binds to this site with higher affinity than Ras-GDP, and it is implicated in the activation feedback-loop of SOS through the modification of the relative position of the REM domain and the Cdc25H helical-hairpin (Margarit *et al.* 2003, Boykevisch *et al.* 2006). In the autoinhibited structure of SOS, the DH-PH tandem interacts with the REM domain and blocks the distal site for the Ras allosteric molecule; and the DH domain is inhibited by the helical linker between the PH and REM domains. The other domains establish additional interdomain contacts that stabilize the autoinhibited conformation (Sondermann *et al.* 2004). Therefore, in the autoinhibited state, the binding site for the catalytic Ras molecule in the Cdc25H domain is accessible, but kept in an incompetent form.

Although the first observations showed that Ras-GTP induces a moderate increase in the GEF activity of SOS in solution (Margarit *et al.* 2003), it increases over 2 orders of magnitude when the GTPase is cross-linked to artificial liposomes (Gureasko *et al.* 2008). SOS is enriched at the liposomes through its phospholipid-interacting domains. At the membrane, the high concentration of Ras molecules would be sufficient to displace the inhibition of the distal site for Ras-GTP, which would be the initial step of the activation mechanism (Findlay *et al.* 2008, Groves *et al.* 2010, Cherfils *et al.* 2013, Jun *et al.* 2013).

The activation of SOS has been proposed to be initiated with the production of Ras-GTP by other GEFs, such as RasGRP1 (Roose *et al.* 2007). Afterwards, inter-domain rearrangements would cause (i) the release of the autoinhibitory interactions, (ii) the reorientation of all the domains to bind properly to the GTPases and the phospholipids, and therefore (iii) the strengthening of the interaction with the membrane [reviewed in (Cherfils *et al.* 2013)]. In the final active form, the DH domain would be accessible to Rac, which is in agreement with the activation of Rac downstream of Ras activation. Recently it has been proposed that the Proline-rich domain could regulate the GEF activity of SOS through two mechanisms: (i) in the presence of the Proline-rich domain, SOS is recruited inefficiently to the membrane in a SH3-independent manner, which suggests that the C-terminal part of the protein could occlude the lipid-binding sites (Lee *et al.* 2017), (ii) the Rac GEF activity of the DH domain is inhibited by the Proline-rich region and it is released upon binding of the protein E3b1/Abi-1 (Scita *et al.* 1999, Innocenti *et al.* 2002).

The importance of the regulation of SOS is illustrated by several mutations that cause the Noonan syndrome. These mutations mainly affect the interfaces of the regulatory domains, while usually they do not affect catalytic regions (Lepri *et al.* 2011, El Bouchikhi *et al.* 2016).



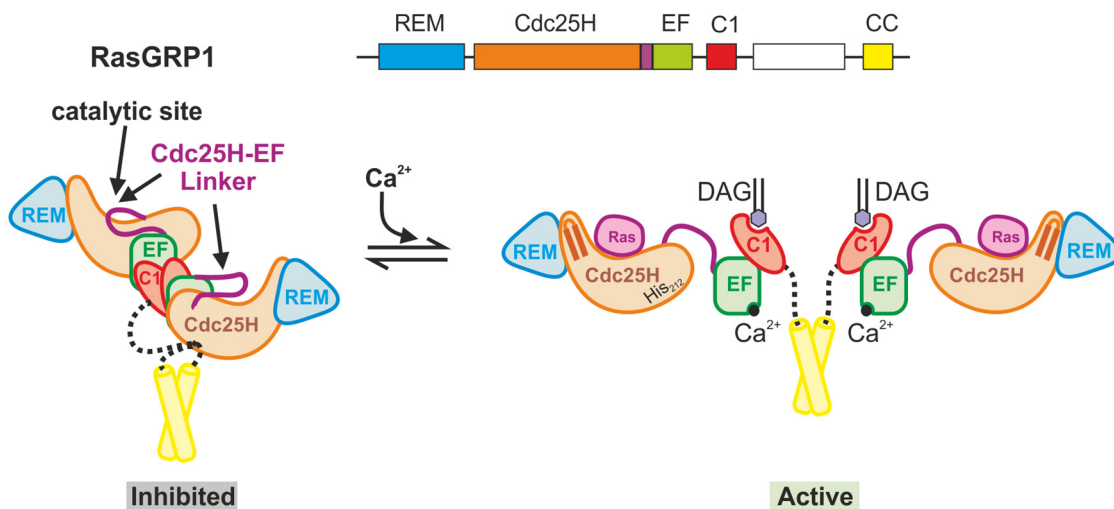
**Figure I3. Mechanism of autoinhibition and activation of SOS1.** In the autoinhibited structure both the distal binding site for Ras-GTP in the REM-Cdc25H domains and the Rac binding site in the DH domain are hidden. Upon the binding of Ras-GTP to the distal site the protein acquires an active conformation, in which the Cdc25H and DH domains are catalytically competent.

## RasGRP1

RasGRP1, also known as CalDAG-GEFII, contains a catalytic REM-Cdc25H tandem, an EF domain (which binds  $\text{Ca}^{2+}$ ), a C1 domain (which binds diacylglycerol, DAG) and a coiled-coil region (CC), responsible for the dimerization of the protein (Figure I4).

The Cdc25H domain of RasGRP1 is inhibited by direct blockage of its GTPase binding site. The EF domain binds to the Cdc25H domain and the linker between the two domains occupies the binding site for the switch 2 region of the Ras molecule (Iwig *et al.* 2013). The EF domain also blocks the binding of the C1 domain to DAG. In this manner, the EF domain has a dual function inhibiting the Cdc25H domain and the C1 domain. The activation of the protein is driven by three events: (i) the binding of one  $\text{Ca}^{2+}$  ion to the EF1 domain, which produces a conformational change of this domain leading to a rearrangement of the molecule. This rearrangement affects to the positions of the inhibitory linker and the C1 domain, which results in the exposure of the binding sites for the GTPase and DAG, respectively; (ii) the recruitment of the protein to the membrane fraction mediated by the C1/DAG interaction. Finally, (iii) the protein is maintained in a stable active conformation by phosphorylation of residue Thr184 in the linker between Cdc25H and REM domains (Vercoulen *et al.* 2017). Recently, it has

been described a regulatory His residue in the Cdc25H (Vercoulen *et al.* 2017). This residue is deprotonated when the pH increases and helps in the reorganization of the regulatory elements to acquire the active conformation. In this manner the intracellular pH has been proposed to modulate RasGRP1 signaling; for instance, it could contribute to the activation of Ras in tumoral cells in which the intracellular pH is usually higher than in normal cells (Webb *et al.* 2011, Vercoulen *et al.* 2017)



**Figure I4. Mechanism of autoinhibition and activation of RasGRP1.** In the autoinhibited state, the GTPase-binding site in the Cdc25H domain is blocked by the linker between the Cdc25H and EF domains. Upon binding of  $Ca^{2+}$  to the EF domain, the autoinhibition of the Cdc25H is released. The protein is recruited to the membrane via C1/DAG interactions. Representations are based in the 3D structures, PDB ID: 4L9M, 4L9U and 2MA2.

## Epac

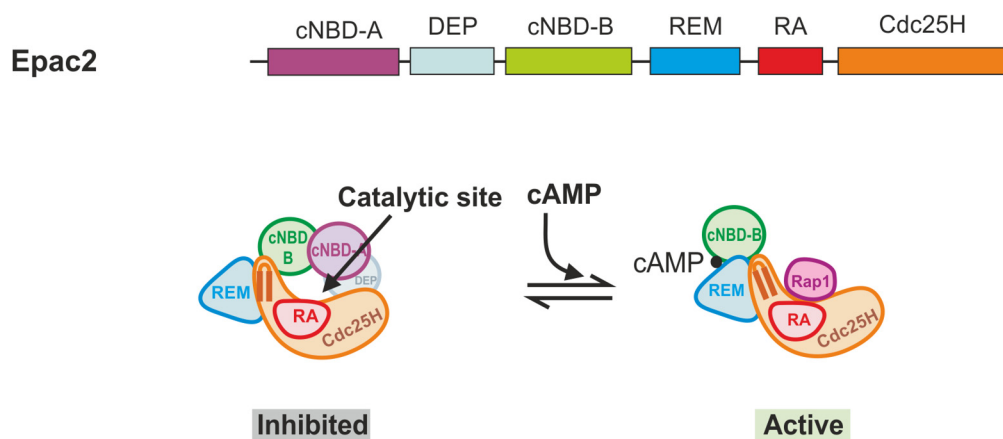
Epac proteins are GEFs for Rap1 and Rap2 GTPases. There are two Epac isoforms, Epac1 and Epac2, which differs by the presence of a N-terminal regulatory domain (cyclic nucleotide binding domain A; cNBD-A) in Epac2. The larger isoform, Epac2, contains a cNBD-A domain, a DEP (disheveled, Egl-10, pleckstrin) domain, a cNBD-B domain, a REM domain, a RA (Ras association) domain and a Cdc25H domain. The cNBD domains interact with cAMP (the cNBD-B interacts with higher affinity than the cNBD-A), the DEP domain interacts with lipids and the RA domain is an effector domain for Ras-GTP (Figure I5).

The structure of Epac2 in autoinhibited state shows extensive contacts between the cNBD-A and -B domains. Additionally, the DEP domain interacts with the Cdc25H

domain in a surface not shared with the GTPase binding site (Rehmann *et al.* 2003, Rehmann *et al.* 2006). In this conformation, the binding site for Rap GTPase is sterically occluded, but the conformation of the helical-hairpin is similar to that of other competent Cdc25H domains, suggesting that the inhibition is only mediated by steric hindrance.

Epac proteins are activated by the second messenger cAMP (de Rooij *et al.* 1998), which binds to the interfacial site formed by the cNBD-B and REM domains (Rehmann *et al.* 2006). This produces a three-dimensional reorganization of the regulatory domains, which ends up with the exposure of the GTPase binding site. Additionally, the binding of Ras-GTP to the RA domain would contribute to the activation by recruiting the protein to the membrane (Figure I5).

In summary, the catalytic activities of SOS, RasGRP1 and Epac are regulated by intramolecular inhibitory interactions, which involve non catalytic domains. However, their molecular mechanisms of autoinhibition are not conserved.



**Figure I5. Mechanism of autoinhibition and activation of Epac2.** In the autoinhibited structure the GTPase-binding site in the Cdc25H domain is blocked by the arrangement of the domains cNBD-B, DEP and cNBD-A. Binding of cAMP to the cNBD-B induces a conformational change that exposes the GTPase binding site of the Cdc25H domain.

## 2. C3G

C3G (Crk SH3 domain-binding guanine nucleotide exchange factor also known as RapGEF1) was first identified as a ubiquitously expressed Crk and Grb2 SH3-binding protein (Tanaka *et al.* 1994). Then, its GEF activity for the GTPases Rap1, Rap2, R-Ras, TC21 and TC10 was described (Gotoh *et al.* 1995, Gotoh *et al.* 1997). However, only the specificity for Rap1 has been corroborated *in vitro* (van den Berghe *et al.* 1997, Popovic *et al.* 2013).

### Primary structure and interactions

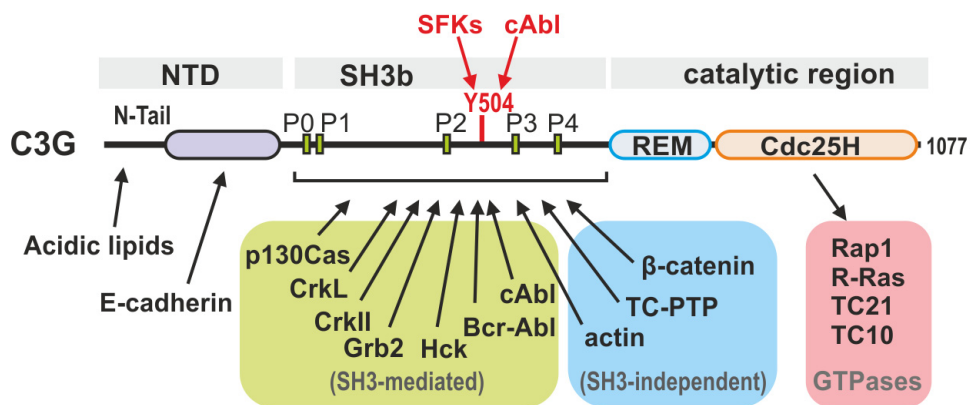
Human C3G (NCBI: NM\_005312) is a 1077-residue protein that has a modular structure consisting of three structurally and functionally different regions (Figure I6):

(i) The N-terminal domain (residues 4-245; hereafter NTD) of C3G has no-homology to other domains. A previous work from our lab had demonstrated that the NTD of C3G has a  $\alpha$ -helical rich sequence (residues 90-245), predicted to contain 4  $\alpha$ -helices that might form a helical bundle (Gómez-Hernández 2014). Within the NTD, the segment 144-230 interacts with E-cadherin and this interaction is relevant for the engagement of the adherens junctions (Hogan *et al.* 2004, Asuri *et al.* 2008). Previously, we found that the NTD binds to the REM domain intramolecularly, and this interaction was proposed to participate in the regulation of the GEF activity (Gómez-Hernández 2014). Finally, the most N-terminal segment (residues 4-64) is rich in basic residues ( $pI = 9.9$ ) and interacts with anionic lipids (Gómez-Hernández 2014), a feature also described in other GEFs (Cherfils *et al.* 2013, Karandur *et al.* 2017).

(ii) The central region of C3G or SH3-binding domain (residues 246-670, hereafter SH3b) is mainly involved in protein-protein interactions. It is predicted to be intrinsically disordered and contains 5 Proline-rich motifs (residues 265-276, 282-291, 452-562, 539-549 and 607-614; named P0-P4 motifs) that interact with SH3-domain containing proteins. Crk proteins were the first characterized binding partners of the SH3b region; they bind directly to P1-P4 motifs with high affinity ( $k_d = 2-4 \mu M$ ) through its N-terminal SH3 domain (SH3N) (Knudsen *et al.* 1994). The specificity and directionality of this interaction is controlled by the presence of a final Lys residue in the Proline-rich motifs (PPXLPXK) (Wu *et al.* 1995). On the other hand, the P0 motif binds directly to the SH3

domain of p130Cas (Kirsch *et al.* 1998). Other SH3 domain-containing proteins have been identified to bind to the C3G SH3b domain, such as Grb2 (Tanaka *et al.* 1994), Hck (Shivakrupa *et al.* 2003) and c-Abl (Radha *et al.* 2007), as well as Bcr-Abl oncoprotein (Gutierrez-Berzal *et al.* 2006). Finally, other proteins lacking SH3 domains have been found to associate to the SH3b, such as Actin (Martin-Encabo *et al.* 2007), the specific T-cell phosphatase TC-PTP (Mitra *et al.* 2011) and  $\beta$ -catenin (Dayma *et al.* 2012), although their direct binding have not been proved. The SH3b domain also harbors the residue Tyr504, whose phosphorylation is important for C3G activation (Ichiba *et al.* 1999).

(iii) The C-terminal region, also known as catalytic region, consists of a REM and Cdc25H domains, this last being responsible for the binding to the GTPase and the destabilization of the bound nucleotide (see before).



**Figure 16. C3G primary structure.** The different domains composing C3G and their interactions with binding partners are indicated.

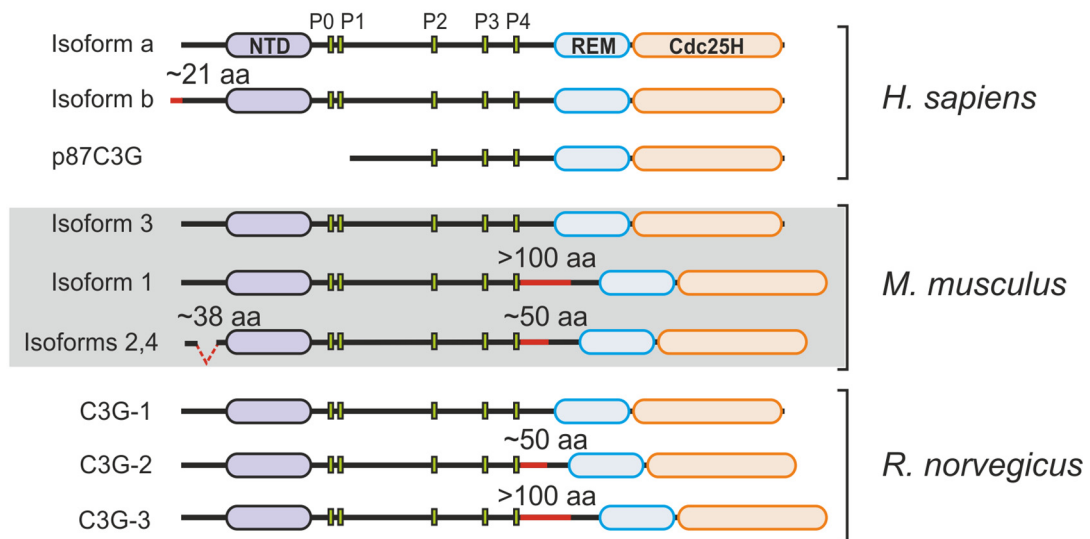
### C3G isoforms and tissue expression

Several isoforms of C3G have been described in different tissues and species (Figure 17). Variants arise from the alternative splicing of the RNA from a single locus, which in human correspond to *RAPGEF1* gene (Radha *et al.* 2011). Three isoforms have been described in humans: a, b, and p87C3G. The first three residues of C3G-a are substituted by a 21-residue segment in C3G-b. Isoforms a and b results in proteins weighting ~140 kDa (~7.5 kb transcript, p140-C3G). The p87 isoform arises from a ~4.5 kb transcript present in chronic myeloid leukemia (CML) cell lines (Gutierrez-

Berzal *et al.* 2006). It has been proposed that p87C3G might participate in the development of the disease (Maia *et al.* 2009, Maia *et al.* 2013).

Another two isoforms, containing an insertion variable in length (~50-170 residues) immediately after the P4 motif, has been found in mouse and rat, brain and testis (Shivakrupa *et al.* 1999, Kawai *et al.* 2001). A third isoform, lacking a 38-residue segment in the NTD has been described in mouse (Zhai *et al.* 2001). Yet, the specific functions associated to them have not been studied.

In summary, despite the expression of C3G is ubiquitous, some variants are associated with specific tissues, which suggests that the alternative splicing of C3G RNA could confer specific functional properties.



**Figure I7. C3G isoforms.** Schematic representation of C3G isoforms in human, mouse and rat. Reference sequences are reported below: human isoform a (NM\_005312.3), isoform b (NM\_198679.1, NM\_001304275.1), p87C3G (Gutierrez-Berzal *et al.* 2006); mouse isoform 3 (NM\_054050.2), 1 (NM\_001039087.1), 2 (NM\_001039086.1) and 4 (NM\_001362702.1); and rat isoform C3G-1 and -2 (Shivakrupa *et al.* 1999), and C3G-3 (UniProt ID: F1M8L9\_RAT).



## **C3G functions**

C3G, through its GEF-dependent and independent role in signaling, participates in the regulation of multiple cellular functions, such as actin remodeling, filopodia formation, cell junction integrity, adhesion, migration, proliferation, differentiation, suppression of malignant transformation, apoptosis and cell survival [reviewed in (Radha *et al.* 2011)].

### Regulation of cell adhesion

Rap GTPases are essential regulators of cell adhesion [reviewed in (Bos 2018)]. C3G knockout mice die before embryonic day 7.5, which suggest that C3G-function can not be compensated by other Rap1 GEFs in early developmental stages (Ohba *et al.* 2001). This phenotype is probably due to defects in cell-cell adhesion, since embryonic fibroblasts (MEFs) derived from these mice display reduced adhesion and increased migration.

C3G participates in the formation of adherens junctions (AJs) (Ohba *et al.* 2001, Hogan *et al.* 2004, Bos 2005, Fukuyama *et al.* 2005, Asuri *et al.* 2008). Particularly, C3G localizes at the AJs and activates Rap1 in the initial stages of the formation of adhesions, a process mediated by their interaction with E-cadherin (Hogan *et al.* 2004). Since C3G competes with  $\beta$ -catenin for the binding to the E-cadherin,  $\beta$ -catenin would substitute C3G in the mature junctions. On the other hand, an interaction between the SH3b domain of C3G and the armadillo repeats of  $\beta$ -catenin has been described and overexpression of  $\beta$ -catenin negatively regulates C3G expression (Dayma *et al.* 2012). Additionally, further research has shown that C3G associates with p130Cas in response to cell adhesion (de Jong *et al.* 1998), and activates Rap1 GTPase in response to stretching (Sawada *et al.* 2001, Tamada *et al.* 2004).

### Role in haematopoietic cells

Rap1 is a critical regulator of the activation of integrins upon TCR ligation in T-cells (Sebzda *et al.* 2002, Duchniewicz *et al.* 2006). C3G-dependent Rap1 activation participates in the "inside-out" clustering and activation of integrins, such as LFA1, in response to TCR activation. This process is mediated by the WAVE2 complex, which

orchestrates the recruitment to the membrane of the CrkL-C3G complex, and the Abl-dependent phosphorylation of C3G (Nolz *et al.* 2008). Additionally, C3G has been proposed to participate in the B-cell receptor-induced signaling (Smit *et al.* 1996).

Among all Rap family proteins, Rap1b is expressed at higher levels in mature megakaryocytes and platelets (Torti *et al.* 1994). Rap1b is a key regulator of the activation of platelets through the modulation of the secretion of platelet granules and the activation of platelet integrin  $\alpha\text{IIb}\beta\text{3}$ , in a process mediated by the interaction with Rap1 interacting adaptor molecule (RIAM), talin and kindlin (Lafuente *et al.* 2004, Gingras *et al.* 2019).

The more abundant GEF for Rap1 in platelets is CalDAG-GEFI, which mediates the  $\text{Ca}^{2+}$ -dependent activation (Stefanini *et al.* 2009). Our group has described that C3G participates in a  $\text{Ca}^{2+}$  independent activation of Rap1, suggesting that both proteins complement the regulation of platelet physiology (Gutierrez-Herrero *et al.* 2012, Gutierrez-Herrero 2018). Overexpression of C3G in platelets favors platelet-induced angiogenesis and tumor metastasis (Martin-Granado *et al.* 2017). Additionally, C3G plays a role in the differentiation of megakaryocytes (Ortiz-Rivero 2017).

## **C3G pathological functions**

### Cancer

The role of Rap1 in the development of cancer is controversial. Active Rap1 has been found to inhibit invasion and metastasis in lung, bladder and brain cancer. On the other hand, it promotes the development of the disease in melanoma, leukemia, head and neck squamous cell carcinoma (HNSCC), breast cancer, esophageal squamous cell carcinoma, non-small cell lung carcinoma and pancreatic carcinoma [reviewed in (Zhang *et al.* 2017)].

Similar to Rap1, the role of C3G in the development of cancer is also controversial. Studies have shown that C3G contributes to the transformation mediated by v-Crk oncogenes and the RET-PTC genetic rearrangements (Tanaka *et al.* 1997, De Falco *et al.* 2007). On the other hand, C3G was found to down-regulate the transformation induced by Ras, Sis, R-Ras and Dbl oncogenes, and this function only required the

SH3b domain (Guerrero *et al.* 1998, Guerrero *et al.* 2004). Recently, it has been shown that transgenic expression of C3G in platelets increases the metastatic properties of mouse melanoma cells (Martin-Granado *et al.* 2017), and that C3G knock-down exacerbates the migratory and invasive properties of MEFs and HCT116 cells through the activation of p38 $\alpha$  (Priego *et al.* 2016). Additionally, C3G levels are upregulated in hepatocellular carcinoma and lung cancer cell lines [reviewed in (Radha *et al.* 2011, Sequera *et al.* 2018)].

C3G was proposed to be a potential oncogene in the B-cell line Ba/F3. Using a high-throughput screening, in which the random insertion of the *sleeping beauty* transposon was used to test the ability of Ba/F3 cells to grow in an IL3-independent manner, recurrent transposon insertions were found in *RAPGEF1* gene (Guo *et al.* 2016). Depending on the position of the gene in which the transposon is inserted, it can result in a (i) gain-of-function effect, by enhancing the expression of the gene under the strong promoter of the transposon; or (ii) a loss-of-function effect, by producing a truncated form of the protein. Yet the causal association of C3G with the development of leukaemia has not been proved. It was described that active Rap1 promotes the development of leukaemia. Notably, the p87C3G isoform is directly downstream the Bcr-Abl oncogene in CML cells (Gutierrez-Berzal *et al.* 2006, Maia *et al.* 2013), and down-regulation of C3G expression has been detected in chronic lymphocytic leukaemia (Fernandez *et al.* 2008). In conclusion, it is likely that C3G plays pro-tumorigenic and anti-tumorigenic roles in different scenarios.

### Participation in the liver physiopathology

Rap proteins are important regulators of liver physiology [reviewed in (Sequera *et al.* 2018)]. For instance, Rap2B promotes the development and progression of hepatocellular carcinoma (HCC) (Zhang *et al.* 2017), while Rap1 could have a dual role (Cruise *et al.* 1997). Rap activation by Epac1 produces pro-survival signals in liver cells (Gates *et al.* 2009). On the other hand, the liver specific Epac isoform Epac2C, which lacks the first cNBD-A and DEP domains, has been described to promote fibrosis in a model of alcoholic liver fibrosis (Ueno *et al.* 2001, Yang *et al.* 2016). The short isoform p87C3G is expressed in the liver of mice at embryonic day 13.5. The precise functions of C3G in the liver are unknown; yet, the expression levels of C3G are higher in non-metastatic HCC and are reduced during metastasis. In addition, several single nucleotide coding variants (SNVs), affecting the *RAPGEF1* gene correlate with lower survival of HCC patients (Sequera *et al.* 2018).

## Other diseases

Although in other GEFs of Ras family, such as SOS, single mutations have been associated with the development of diseases (e.g. Noonan syndrome), no coding-SNVs have been linked to the *RAPGEF1* gene to date. However, different non-coding SNVs in the *RAPGEF1* locus, probably affecting protein expression, have been associated to both high risk and low risk Type 2 diabetes (Gaulton *et al.* 2008, Hong *et al.* 2009). On the other hand, a C3G hypomorphic mutant mouse model, expressing less than 5% of the protein, showed defects in neural migration and glial attachment during development, which resembles lissencephaly disease in humans (Voss *et al.* 2008).

## **Adaptor proteins**

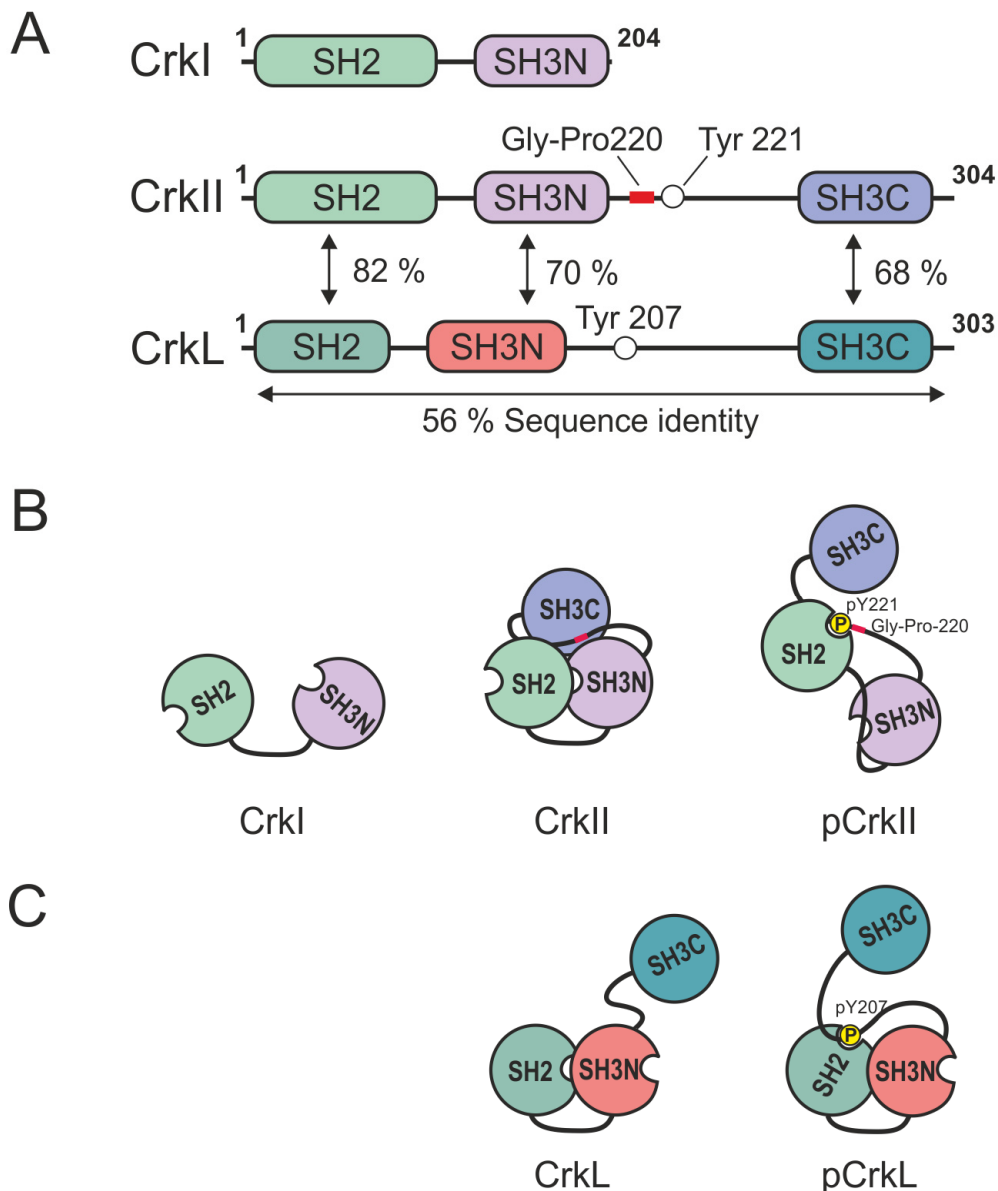
Several SH3-containing proteins have been described to bind to C3G to date (see before). Among all of them, Crk proteins bind to C3G with high affinity and perform the adaptor function in the majority of the processes (Radha *et al.* 2011).

The human Crk family of adaptor proteins is formed by three members (CrkI, CrkII and CrkL) arising from two independent *loci* (*CRK* and *CRKL*) (Matsuda *et al.* 1992, ten Hoeve *et al.* 1993, Birge *et al.* 2009). They are composed of three domains (Figure I8A). (i) A N-terminal Src homology-2 (SH2) domain that binds pTyr consensus sequences (pYXXP); (ii) a central SH-3 (SH3N) domain that binds Proline-rich motifs with the consensus sequence PPXXPXK/R; and (iii) a C-terminal SH3 (SH3C) domain that lacks Proline-rich motif-binding ability (Kobashigawa *et al.* 2012). The adaptor function is performed by recruiting effector proteins such as C3G (binding them via their SH3N domain), to scaffolding proteins on the membrane, such as p130Cas or receptor tyrosine kinases (binding them via SH2-pTyr interactions).

CrkI and CrkII result from the alternative splicing of the *CRK* gene RNA, and differ from each other by the presence at the C-terminal end of CrkII of an additional SH3 domain (CrkI: SH2-SH3N; CrkII: SH2-SH3N-SH3C) (Figure I8B). CrkL is encoded by a different gene (*CRKL*) and shares 56% of protein sequence identity with CrkII (Kobashigawa *et al.* 2012).

Despite the similarities in primary structure, Crk proteins differ largely in their 3D inter-domain arrangement [reviewed in (Kobashigawa *et al.* 2012)]. The SH2-SH3N domains of CrkI do not form a compact structure, and the binding sites for pTyr and Proline-rich motifs are accessible. However, in CrkII the C-terminal region stabilizes a compact conformation of the SH2-SH3N domains in which the binding site of the SH3N is blocked (Kobashigawa *et al.* 2007). On the other hand, the SH2-SH3N domains of CrkL form a compact structure stabilized by hydrophilic interactions between the two domains, which blocks the pTyr-binding site of the SH2 domain (Jankowski *et al.* 2012). In summary, CrkL and CrkII differ in the relative orientation of their domains and in the accessibility to bind to pTyr or Proline-rich motifs (Figures 18B and C).

CrkII and CrkL are regulated via phosphorylation by the Abl kinase at Tyr221 in CrkII and Tyr207 in CrkL (de Jong *et al.* 1997, Kobashigawa *et al.* 2007, Peterson *et al.* 2008) ; these two Tyr residues occupy an equivalent position in the linker that connects the SH3N and SH3C domains. Upon phosphorylation of CrkII-Tyr221 or CrkL-Tyr207 they bind to their companion SH2 domains in an intramolecular manner. Yet, these interactions have quite distinct effects in CrkII and CrkL. In phospho-CrkII (pCrkII) the SH3N-SH3C linker blocks the Proline-rich-binding site of the SH3N domain. In contrast, the interaction between the SH2 and pTyr207 in pCrkL does not block the SH3N binding site, which remains available for binding to Proline-rich ligands. In conclusion, the phosphorylation of CrkII by Abl kinase results in the inhibition of SH2- and SH3-dependent bindings, while the phosphorylation of CrkL only compromises the SH2-dependent binding. In both cases, the phosphorylation of Crk results in the detachment of the protein from pTyr motifs, which eventually causes the translocation of the protein to the cytosol.



**Figure I8. Structure and regulation of Crk proteins.** (A) Domain structure and sequence identity of Crk proteins. The Tyr-phosphorylation sites in CrkII and CrkL and the isomerization site Gly219-Pro220, for CrkII are highlighted. Adapted from (Kobashigawa *et al.* 2012). (B) Domain arrangement of the proteins CrkI, CrkII and pCrkII. (C) Domain arrangement of CrkL and pCrkL. The relative positions of the domains are based in the NMR structures of CrkI (PDB ID: 2EYY), CrkII (PDB ID: 2EYZ), CrkL (PDB ID: 2LQN), pCrkII (PDB ID: 2DVJ) and pCrkL (PDB ID: 2LQW) (Kobashigawa *et al.* 2007, Jankowski *et al.* 2012).

CrkII is also regulated by the *cis-trans* isomerization of a Proline switch (Gly237-Pro238 in chicken CrkII and Gly219-Pro220 in human) in the linker connecting the SH3N and SH3C domains (Sarkar *et al.* 2007, Isakov 2008, Sarkar *et al.* 2011, Schmidpeter *et al.* 2014, Saleh *et al.* 2016). The CrkII *cis*-isomer adopts a closed conformation in which the binding site of the SH3N domain is blocked by the SH3C domain, hence it is considered inactive. The CrkII *trans*-isomer adopts an open

conformation in which the SH3 domain is competent for the binding to Proline-rich motifs. The cycling between the *cis*- to *trans*- isomers happens very slowly and is reversibly catalyzed by peptidyl-prolyl *cis-trans* isomerases (PPIases), such as the cyclophilin A (CypA) and FK506 immunophilins. The expression of these proteins enhances CrkII-C3G association in T-cells, illustrating the functional relevance of this mechanism (Nath *et al.* 2014, Braiman *et al.* 2015). However, since the Gly-Pro220 is close to the regulatory Tyr221 in human CrkII, it has been proposed that this effect is mediated by the steric inhibition of the phosphorylation when the PPIases are bound to CrkII (Saleh *et al.* 2016).

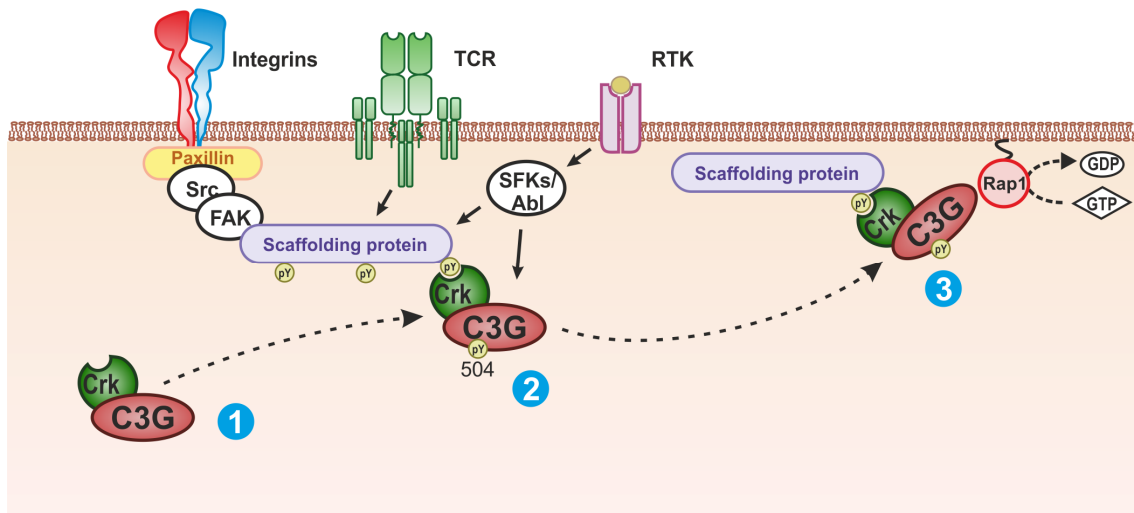
## Regulation and activation of C3G

At the cellular level, the C3G-Rap1 pathway is activated upon several stimuli, such as T-cell receptor (TCR) ligation (Medeiros *et al.* 2005, Nolz *et al.* 2008), hepatocyte growth factor (HGF) (Sakkab *et al.* 2000), growth Hormone (GH) (Ling *et al.* 2003), platelet-derived growth factor (PDGF) (Yokote *et al.* 1998), epidermal growth factor (EGF) and insulin (Okada *et al.* 1997), neuronal growth factor (NGF) (York *et al.* 1998), interferon- $\gamma$  (Alsayed *et al.* 2000), erythropoietin (EPO) and Interleukin-3 (Nosaka *et al.* 1999, Arai *et al.* 2001), as well as adhesion signals through the activation of integrins (Arai *et al.* 1999, Buensuceso *et al.* 2000).

Very little is known about the mechanism of regulation of C3G at the molecular level. The removal of the first half of the molecule (residues 1-579) results in the activation of Rap1 in cell cultures (Ichiba *et al.* 1999). In this regard, an intramolecular interaction between the NTD and REM domains was described in our lab and we hypothesized that it could negatively regulate C3G GEF activity (Gómez-Hernández 2014). However, another work suggested that a construct lacking the first 544 residues remains inhibited (Popovic 2013). Noteworthy, the structures of the isolated domains and their arrangement in the autoinhibited protein are not known to date.

A constitutive association of the Crk-C3G complex in the cytosol has been described (Okada *et al.* 1998, Buensuceso *et al.* 2000). It is accepted that C3G activation occurs through (i) the translocation of the CrkL-C3G complex to the membrane, where it interacts with Rap1 (Ichiba *et al.* 1997), and (ii) the phosphorylation of C3G in Tyr residues (Figure I9) (Ichiba *et al.* 1999, Shivakrupa *et al.* 2003, Radha *et al.* 2004, Mitra *et al.* 2011). There is evidence supporting that Crk proteins can activate directly

the GEF activity of C3G (Ichiba *et al.* 1999, Popovic 2013). Several kinases are described to phosphorylate C3G, such as the Src family kinases (SFK) Src and Hck (Shivakrupa *et al.* 2003, Radha *et al.* 2004), or c-Abl (Gutierrez-Berzal *et al.* 2006, Mitra *et al.* 2011). The phosphorylation of Y504 in the SH3b residue is essential for the CrkL-dependent C3G activation in cells (Ichiba *et al.* 1999), and it is commonly used as a reporter of the activation state of the protein.



**Figure 19. Current general model of C3G activation in cells.** (1) Crk-C3G complex is constitutive in the cytosol. (2) Upon stimulation (e.g. Receptor Tyrosin kinase (RTK) activation, TCR ligation, or Integrin activation), the complex is recruited to the membrane via Crk SH2 binding to pTyr residues of scaffolding proteins (e.g. p130Cas, CasL or ZAP-70). (3) Then, C3G is activated by phosphorylation in Tyr residues by Src family kinases (SFKs) or Abl kinase, promoting Rap1 activation on the membrane.



## Summary

Several C3G, GEF-dependent and independent, functions have been described to date. Yet, the structural organization of the protein, and the underlying mechanisms of autoinhibition and activation remain unknown.

This work provides the first systematic characterization of the mechanisms of autoregulation of C3G, including the precise identification of both, inhibitory and positive-regulatory interactions, involving regions out of the catalytic domains. Residues important for these interactions have been mapped, which, in some cases are affected by single nucleotide variants (SNVs) found in cancer patients. Therefore this work could be the first to link the mechanisms of autoregulation of C3G with the development of this disease. We also present a comprehensive analysis of the effects of CrkL binding and phosphorylation status on C3G activation. Collectively, this work sheds light on how C3G is inhibited and responds to different stimuli when getting activated.



## *OBJECTIVES*



The overall scope of this work is to understand the mechanisms of autoregulation and activation of C3G and to explore potential alterations linked to diseases. To achieve this goal, we have addressed the following specific objectives:

1. To identify and characterize intramolecular interactions that regulate both positively and negatively the GEF activity of C3G.
2. To understand the effect of the binding of Crk adaptor proteins to C3G and tyrosine-phosphorylation of C3G on the mechanisms of autoregulation, i.e. over its GEF activity.
3. To characterize in detail and quantitatively the interaction of the adaptor protein CrkL with C3G at multiple sites.



## *METHODS*





## 1. cDNA

The reference sequences in NCBI and UniProt databases for the proteins used in this work are listed in Table M1. The cDNA of C3G corresponds to the sequence NCBI ID: NM\_005312 with minor differences: (i) a polymorphism in position 281 (S281G), (ii) the start of our sequence is slightly different and includes 5 additional amino acids (MSGKIEKA instead of MDT in the NCBI sequence). Accordingly, all constructs begin in residue number four of the canonical sequence, which is the first common residue between both sequences.

**Table M1.** cDNAs used in this work

Protein (Gene)	NCBI ID	UniProt ID	Provided by
C3G ( <i>RAPGEF1</i> )	NM_005312	Q13905	Our laboratory
CrkL ( <i>CRKL</i> )	NM_005207.3	P46109	Our laboratory
CrkII ( <i>CRK</i> )	NM_016823.3	P46108	Our laboratory
Rap1b ( <i>RAP1B</i> )	NM_015646.5	P61224	Our laboratory
c-Src ( <i>SRC</i> )	NM_005417.3	P12931	Dr. Robert Lefkowitz, Addgene #42202
BirA ( <i>birA</i> )	NC_002695.1	A0A069FJV6	Dr. Alice Ting, Addgene #20857
YopH ( <i>yopH</i> )	NC_004564.1	O68720	Dr. Andrés Alonso

## 2. Constructs overview

The full list of constructs used in this work is summarized in Tables M2-M4 (M2 and M3 contains bacterial expression constructs; M4 contains mammalian expression constructs). They include information about the names, boundaries, characteristics of the construct, type of vector, restriction enzymes used to clone and the use of the constructs.

**Table M2.** C3G bacterial expression constructs

Name	Limits	Description	Vector	Sites	Purpose
GST-C3G-His (WT & mutants)	4-1077	Full-length	pGEX-2xTEV-cHis	NcoI/XhoI	SEC-MALS/ AUC / NEK / ITC
His-C3G	4-1077	Full-length	pETEV15b-NcoI	NcoI/BamHI	Purification test
GST-C3G	4-1077	Full-length	pGEX-TEV	NcoI/BamHI	Purification test
C3G-His	4-1077	Full-length	pETEV22b-x2	NcoI/XhoI	Purification test
His-Halo-C3G	4-1077	Full-length	pETEV15b-Halo	NdeI/BamHI	Purification test
C3G-PAAA	4-1077	Full-length PAAA	pGEX-2xTEV-cHis	NcoI/XhoI	PD / NEK / ITC
C3G-APAA	4-1077	Full-length APAA	pGEX-2xTEV-cHis	NcoI/XhoI	PD / NEK / ITC
C3G-AAPA	4-1077	Full-length AAPA	pGEX-2xTEV-cHis	NcoI/XhoI	PD / NEK / ITC
C3G-AAAP	4-1077	Full-length AAAP	pGEX-2xTEV-cHis	NcoI/XhoI	PD / NEK / ITC
C3G-ΔNTD	246-1077	P0-P4-REM-Cdc25H	pGEX-2xTEV-cHis	NcoI/XhoI	NEK
GST-454-1077-His	454-1077	P3-P4-REM-Cdc25H	pGEX-2xTEV-cHis	NdeI/XhoI	NEK
GST-530-1077-His	530-1077	P3-P4-REM-Cdc25H	pGEX-2xTEV-cHis	NdeI/XhoI	NEK
GST-602-1077-His	602-1077	P4-REM-Cdc25H	pGEX-2xTEV-cHis	NdeI/XhoI	SEC, Aggregated
His-REM-Cdc25H	670-1077	REM-Cdc25H	pETEV15b-NcoI	NcoI/BamHI	Not soluble
His-Cdc25H (WT & Mutants)	815-1077	Cdc25H	pETEV15b-NcoI	NcoI/BamHI	NEK / PD
GST-SH3b	274-646	P1-P4	pGEX-TEV	NcoI/BamHI	PD / NEK
GST-274-578	274-578	P1-P3	pGEX-TEV	NcoI/BamHI	PD
GST-274-500	274-500	P1-P2	pGEX-TEV	NcoI/BamHI	PD / NEK
GST-274-371	274-371	P1	pGEX-TEV	NcoI/BamHI	PD / NEK
GST-372-646	372-646	P2-P4	pGEX-TEV	NcoI/BamHI	PD / NEK
GST-501-646	501-646	P3-P4	pGEX-TEV	NcoI/BamHI	PD / NEK
GST-537-646 WT & Mutants	537-646	P3-P4	pGEX-TEV	NdeI/BamHI	PD / NEK
His-537-646	537-646	P3-P4	pETEV15b	NdeI/BamHI	NEK
GST-537-646-P4A	537-646	P3-P4A	pGEX-TEV	NdeI/BamHI	PD / NEK
GST-579-646	579-646	P4	pGEX-TEV	NcoI/BamHI	PD / NEK
GST-501-578	501-578	P3	pGEX-TEV	NcoI/BamHI	PD / NEK
GST-501-536	501-536	Upstream P3	pGEX-TEV	NcoI/BamHI	PD
GST-537-588	537-588	P3	pGEX-TEV	NdeI/BamHI	PD / NEK
GST-537-578	537-578	P3	pGEX-TEV	NdeI/BamHI	PD / NEK
GST-537-569	537-569	P3	pGEX-TEV	NdeI/BamHI	PD / NEK
GST-537-560	537-560	P3	pGEX-TEV	NdeI/BamHI	PD / NEK
GST-545-646	545-646	Downstream P3 + P4	pGEX-TEV	NdeI/BamHI	PD / NEK
GST-545-646-P4A	545-646	Downstream P3 + P4A	pGEX-TEV	NdeI/BamHI	PD / NEK
GST-545-569	545-569	Downstream P3	pGEX-TEV	NdeI/BamHI	PD / NEK
GST-545-560	545-560	Downstream P3	pGEX-TEV	NdeI/BamHI	PD

NTD, N-terminal domain; SEC-MALS, size exclusion chromatography coupled to multiangle light scattering; AUC, Analytical ultracentrifugation; NEK, Nucleotide exchange kinetics; PD, Pull down; ITC, Isothermal titration calorimetry.

**Table M3.** CrkL, CrkII, Rap1b and Src bacterial expression constructs

Name	Limits	Description	Vector	Sites	Purpose
His-CrkL	1-303	Full-length	pETEV15b	NcoI/BglIII	PD / ITC / NEK / SEC-MALS / AUC
CrkL-Avi	1-303	Full-length	pETEV15b-Avi	NdeI/BglIII	PD
His-CrkL-SH2-SH3N	1-182	SH2-SH3N	pETEV15b	NdeI/BamHI	PD / ITC / NEK
His-CrkL-SH3N-SH3C	125-303	SH3N-SH3C	pETEV15b	NdeI/BglIII	PD / ITC / NEK
His-CrkL-SH3N	125-182	SH3N	pETEV15b	NdeI/BamHI	PD / ITC / NEK
His-CrkII	1-304	Full-length	pETEV15b	NdeI/BamHI	PD / ITC / NEK
His-Rap1b	1-167	G-domain	pETEV15b	NdeI/BamHI	NEK
His-c-SrcKD / YopH	254-536 / -	Kinase domain / Phosphatase domain	pETDUET-HisTEV	EcoRI/HindIII	Phosphorylation

SrcKD, Src kinase domain; SEC-MALS, size exclusion chromatography coupled to multiangle light scattering; AUC, Analytical ultracentrifugation; NEK, Nucleotide exchange kinetics; PD, Pull down; ITC, isothermal titration calorimetry.

**Table M4.** C3G mammalian expression constructs

Name	Limits	Description	Vector	Sites	Purpose
C3G-mEGFP WT & mutants	4-1077	Full-length	pEF1-mEGFP	BglIII/NotI	Rap Act.
C3G -mEGFP-CAAX	4-1077	Full length + K-Ras CAAX box	pEF1-mEGFP	BglIII/NotI	Rap Act.
REM-Cdc25H-mEGFP	670-1077	REM-Cdc25H	pEF1-mEGFP	BglIII/NotI	PD
REM-mEGFP	670-814	REM	pEF1-mEGFP	BglIII/NotI	PD
Cdc25H-mEGFP	815-1077	Cdc25H	pEF1-mEGFP	BglIII/NotI	PD
HA-REM-Cdc25H	670-1077	REM-Cdc25H	pCEFHA	BglIII/NotI	PD
HA-REM	670-814	REM	pCEFHA	BglIII/NotI	PD
HA-Cdc25H	815-1077	Cdc25H	pCEFHA	BglIII/NotI	PD
CrkL-HA	1-303	Full length	pcDNA3-HA	-	Rap Act.

Rap Act., Rap activation assay; PD, Pull down; ITC, isothermal titration calorimetry.

### 3. Vectors and cloning

All constructs were generated by amplification of the cDNA fragments with specific primers and subsequent cloning in the desired vectors applying standard molecular biology techniques. The reference table, containing the information about the primers used to generate each construct, and the full list of primers, is described in Appendix I (Tables A1-A5). Schematic figures of the multicloning sites (MCSs) for each vector are shown in Appendix II. Next, an abbreviated description of the cloning is presented.

## Cloning into pET22b-x2

pET22b-x2 is a modified version of pET22b vector (Novagen) which lacks the *pelB* leader sequence for periplasmic localization. When cloned into this vector, proteins are produced with a non-cleavable C-terminal His-tag. cDNA encoding C3G full length (residues 4-1077) was cloned into pET22b-x2 with NcoI and XhoI restriction sites. In this case the reverse primer did not include stop codon to preserve the translation of the C-terminal His-tag.

## Cloning into pETEV15b and pETEV15b-NcoI

pETEV15b (Alonso-Garcia *et al.* 2009) and pETEV15b-NcoI vectors are derived from the pET15b vector (Novagen), which produce proteins with N-terminal cleavable His-tags. They encode the sequence recognized by the TEV protease, located between the His-tag and the MCS. pETEV15b has NdeI and BamHI sites in the MCS whereas pETEV15b-NcoI displays NcoI and BamHI sites. C3G full-length was cloned into pETEV15b-NcoI using NcoI and BamHI sites, including a stop codon in the reverse primer. C3G full-length has an internal NdeI site (affecting codon 340), therefore we avoided using NdeI for cloning. C3G REM-Cdc25H (residues 670-1077) and Cdc25H (residues 815-1077) were cloned in pETEV15b using NdeI and BamHI sites. CrkL, CrkII and Rap1b G-domain (1-167) constructs were cloned into pETEV15b digested with NdeI and BamHI. CrkL contains an internal BamHI site (affecting codon 198), therefore, we used BglII in the reverse primer, which was ligated into the BamHI site to clone CrkL full-length (residues 1-303) and CrkL-SH3N-SH3C (residues 125-303).

## Cloning into pGEX-TEV and pGEX-2xTEV-cHis

pGEX-TEV is a derivative of the pGEX4T3 vector (GE Life Sciences) in which the multicloning site encodes NdeI, NcoI, BamHI, EcoRI and NotI sites. Additionally, the thrombin protease recognition sequence has been substituted by a TEV site. This vector produce proteins with N-terminal GST-tags. We engineered a new version of this vector, pGEX-2xTEV-cHis (Figure M1). First, we introduced a short linker encoding a hexa-His-tag preceded by a XhoI site using EcoRI and NotI sites (Figure M1B). Then, a second linker encoding the TEV protease recognition sequence was introduced between EcoRI and XhoI sites (Figure M1C). The primers designed allowed to destroy the original XhoI site and reconstitute a new one upstream the TEV sequence.



C3G constructs full-length,  $\Delta$ NTD, 454-1077, 530-1077, and 602-1077 were cloned into pGEX-2xTEV-cHis vector using NcoI and XhoI sites or NdeI and XhoI sites (see Table M2). SH3b constructs were cloned into pGEX-TEV vector using NdeI and BamHI or NcoI and BamHI sites, depending on whether they contained the internal NdeI site or not.

### **Cloning into pETEV15b-Halo-Avi**

pETEV15b-Halo-Avi is a derivative version of pETEV15b (Novagen) created in our laboratory that produces fusion proteins with a N-terminal TEV-cleavable Halo-tag and a C-terminal non-cleavable Avi-tag. The Halo-tag allows the covalent immobilization of proteins by binding to ligands. In our case we used the Halo-tag to boost the expression of proteins in the bacteria. The C-terminal Avi tag allows the site specific biotinylation of proteins using BirA. CrkL full-length was cloned into pETEV15b-Halo-Avi vector using NcoI and BamHI sites. In this form, the sequence of the Halo-tag was substituted by the cDNA of CrkL. In consequence, we refer to the empty form of this vector as pETEV15b-Avi. CrkL cDNA was amplified with a reverse primer containing a BglII instead of a BamHI site, and BglII into BamHI ligation was performed. C3G full-length was cloned into pETEV15b-Halo-Avi vector using NdeI and BamHI sites. In order to do so, the internal NdeI site was previously mutated using specific primers. The reverse primer contained a stop codon. Accordingly, the C-terminal Avi-tag is not expressed and the vector is named pETEV15b-Halo.

### **Cloning into pETDuet**

We used pETDuet vector (Novagen) to coexpress the kinase domain of Src (SrcKD) and YopH phosphatase in bacteria (Seeliger *et al.* 2005). The vector encodes two MCSs each one preceded by a T7 promoter, lac operator and ribosome binding sites. SrcKD (251–533) was cloned in the first MCS using EcoRI and NotI sites, so the final protein contains a N-terminal non-cleavable His-tag. YopH cDNA was cloned without any tag into the second MCS using NdeI and XhoI sites.

### **Cloning into pEF1-mEGFP and pEF1-mEGFP-CAAX**

pEF1-mEGFP and pEF1-mEGFP-CAAX are derivatives of pEF1/V5-His vector (Invitrogen) in which a mEGFP (monomeric enhanced GFP) has been introduced using

NotI and XbaI sites (see appendix). pEF1-mEGFP-CAAX also harbors the C-terminal sequence of K-Ras (169-188), which includes the final CAAX-box. cDNA constructs in these vectors are under the control of the elongation factor 1 promoter (EF1), which allows high-levels of expression of proteins in mammalian cells. The vector also contains a neomycin resistance gene that can be used for selection. Constructs C3G-mEGFP, REM-Cdc25H-mEGFP, REM-mEGFP and Cdc25H-mEGFP were cloned into pEF1-mEGFP vector using BamHI and NotI sites. The reverse primers did not include stop codons to preserve the expression of the mEGFP protein. C3G was cloned into pEF1-mEGFP-CAAX vector using a similar strategy to generate C3G-mEGFP-CAAX.

### **Cloning into pCEFLHA**

pCEFLHA vector (Chiariello *et al.* 2000) displays a Kozak consensus sequence followed by a HA epitope (Wilson *et al.* 1984). Protein expression is under the control of the strong EF1 promoter. REM-Cdc25H, HA-REM and HA-Cdc25H cDNAs were cloned into pCEFLHA vector using BglII and NotI sites.

## **4. C3G mutants**

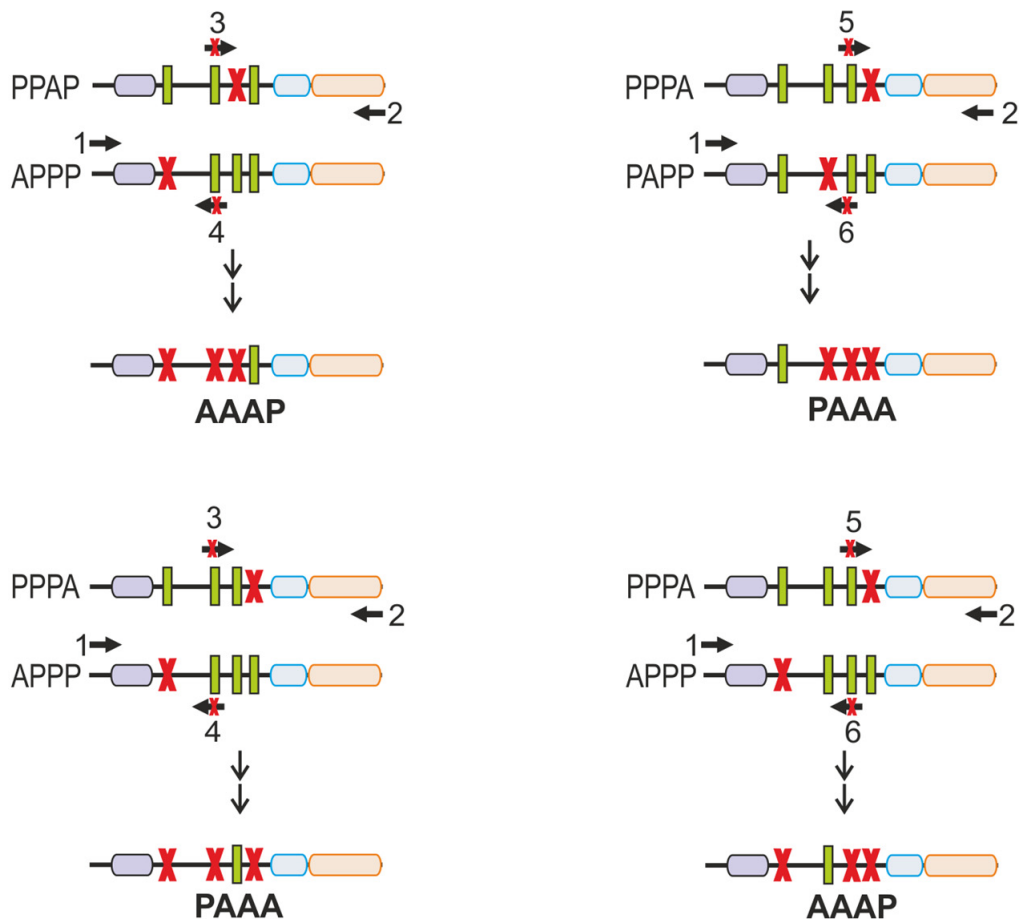
The list of the mutants generated in this work is shown in Table M6. The sequences of all primers used to perform mutagenesis are shown in Appendix 1 (Table A5).

Mutants were generated using the Quick-Change site-directed mutagenesis method (Stratagene) or the overlap extension method (Hussain *et al.* 2016). The first strategy was used when only a few bases were changed. The overlap extension method was used to introduce multiple mutations in the Proline-rich motifs. This method also allowed creating in a simple manner mutants with three Proline-rich motifs altered starting from mutants with a single motif modified. The combination of templates and primers used to generate the triple mutants from the single mutants are shown in Figure M2, and the primers used to mutate the Proline-rich motifs are shown in Table M7. Mutation of Proline-rich motifs consisted in substituting Pro and basic residues by Ala. Therefore, mutants with a single motif mutated were named APPP, PAPP, PPAP and PPPA, when they were introduced in constructs that contain the four motifs, or P3A and P4A when mutations were introduced in constructs that only contain one or two motifs. Similarly, mutants in which three motifs were altered at the time were named as PAAA, APAA, AAPA and AAAP

**Table M6** Mutants generated in this work

Construct / Vector	Mutations
C3G full-length (pGEX-2xTEV-cHis)	M551R, Y554R, M555K, PAAA, APAA, AAPA & AAAP
C3G 537-646 (pGEX-TEV)	K546E/K547E, H550E/M551R, Y554R, M555R/Q556R, E559R/D560R, Y561R/S562R, E563R, P566R, F569R/Y570R, Q571R/T572R, Q574R, E576R, Y579R, Q581R/K582E, E563R/P564R, H550E, M551R, M555R, Q556R, H550A, M551A, Y554A, M555A, Q556A, Y554H, M555K, P4A
C3G 545-646 (pGEX-TEV)	P4A
C3G Cdc25H (pETEV15b)	K922E, E956K, T959A, A973V, T883M, S890F, R894L, E863K, E873K, R1023C, N878R, Q881R, E884R, N887R, N888R, Y891R, S934R, F967R, Y894R, N1043R, L855R, K858E, E860R, E874R, D1044R, K915E, K918E, R921E, D962R, S966R, R968E
C3G 454-1077 (pETEV15b)	E731/784R and M551R
C3G-mEGFP (pEF1-mEGFP)	M551R, Y554H, M555K, E731/784R

T<sub>m</sub> calculation corresponds to the region matching C3G sequence; The limits of the construct are shown in bold; The Kozak sequence is highlighted in yellow. P4A refers to the mutant in which 5 essential residues of the Proline-rich sequence 4 (P4) of C3G have been mutated to Ala.



**Figure M2.** Strategy used to generate the triple-site mutants AAAP, APAA, AAPA and AAPP. Primers 1 and 2 were the ones used for amplification of C3G full-length. Primers 3 and 4 correspond to C3Gh-PP2A-F and -R. Primers 5 and 6 correspond to C3Gh-PP3A-F and -R.



**Table M7.** Primers used to mutate the Proline-rich motifs

Name	Sequence Prot. and DNA (5' to 3')
<b>P1A</b>	Original --> D N G P <b>P P</b> A <b>L P</b> P <b>K K</b> R Q S A
	Mutated --> D N G P <b>A A</b> A <b>A A</b> P <b>A</b> K R Q S A
	C3Gh-PP1A-F GATAATGGTCCT <b>G</b> CAGCAGCA <b>G</b> CGGCACCC <b>G</b> CGAAAAAGACAGTCGGCGC
	C3Gh-PP1A-R GCGCCGACTGTCTTTT <b>C</b> GGGGT <b>G</b> CC <b>G</b> CTGCT <b>G</b> CT <b>G</b> CAGGACCATTATC
<b>P2A</b>	Original --> Q T D T <b>P P</b> A <b>L P</b> E <b>K K</b> R R
	Mutated --> Q T D T <b>A A</b> A <b>A A</b> E <b>A</b> K R R
	C3Gh-PP2A-F GCAGACAGATAC <b>G</b> CGAGCTGCT <b>G</b> CCGCCGAG <b>G</b> CGAAGCGCAGGAG
	C3Gh-PP2A-R CTCTGCGCTT <b>C</b> CTCGG <b>C</b> GCAGCAG <b>C</b> T <b>G</b> CCGTATCTGTCTGC
<b>P3A</b>	Original --> D P E K <b>P P</b> P <b>P P</b> E <b>P</b> K N K H M L A
	Mutated --> D P E K <b>A A</b> P <b>A A</b> E <b>A</b> K N K H M L A
	C3Gh-PP3A-F GACCCAGAAAA <b>G</b> CAGCTCCT <b>G</b> CAGCAGAG <b>G</b> CGAAAAACAAACACATGCTGGCC
	C3Gh-PP3A-R GGCCAGCATGTGTTTGT <b>T</b> TTT <b>T</b> CGCCTCT <b>G</b> CT <b>G</b> CAGGAG <b>C</b> T <b>G</b> CTTTTCTGGGTC
<b>P4A</b>	Original --> A P <b>P P</b> A <b>L P</b> P <b>K</b> Q R Q
	Mutated --> A P <b>A A</b> A <b>A A</b> P <b>A</b> Q R Q
	C3Gh-PP4A-F GGCCCC <b>G</b> CGGCCCG <b>C</b> AGCCCC <b>G</b> CGCAGCGGCAG
	C3Gh-PP4A-R CTGCCGCT <b>G</b> CGGGGG <b>C</b> T <b>G</b> CGGGCG <b>C</b> CGCGGGGCC

The nucleotides changed respect the original sequence are shown in red.

## 5. Bacterial protein expression and purification

### Bacterial protein expression

All proteins were produced in the BL21(DE3)T1 bacterial strain. Rap1b was co-expressed with GroES/L chaperones (gift from Bernd Bukau, Addgene plasmid #27394) as described before (Noguchi *et al.* 2015). Cell cultures were started from single colonies or derived frozen bacterial stocks and were grown at 37° C in 1-5 L of Terrific Broth medium (Tartof *et al.* 1987) supplemented with 100 µg/ml ampicillin. Protein expression was induced in log phase ( $OD_{600nm} = 0.6$ ) by addition of 0.2 mM IPTG. Bacteria were allowed to express the proteins for 3-5 hours at 37 °C or 15-20 hours at 15 °C. In general, higher amounts of soluble recombinant proteins were obtained at 15 °C. After induction, cells were harvested by centrifugation and resuspended in buffers for purification supplemented with 0.1% (v/v) Triton X-100. Bacteria lysis was performed by one freezing-thawing cycle and subsequent sonication. Soluble content of the bacteria was separated from cell debris by centrifugation.

### Purification of His-tag proteins

The proteins Halo-C3G, Cdc25H, Rap1b, CrkL and CrkII were purified as described before (Manso *et al.* 2016). Briefly, recombinant proteins were purified from bacterial

supernatants by immobilized metal ion affinity chromatography (IMAC) using a 5 ml Hi-Trap Ni<sup>2+</sup>-chelating column (GE Healthcare). His-tag proteins were eluted in an imidazole gradient from 5 to 500 mM in 20 mM Tris-HCl (pH 7.9), 500 mM NaCl buffer. Fractions of the chromatography were analyzed by SDS-PAGE followed by Coomassie staining and those containing the protein of interest were pooled and dialyzed against buffer 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1 mM DTT, 0.1 mM EDTA. During the dialysis, the His-tag was cleaved by incubation with rTEV-His protease at RT for 3-5 hours. A reverse IMAC was used to separate the protein of interest from the non-digested sample and the rTEV-His (the His-tag of the rTEV is not cleavable). A final step of size exclusion chromatography (SEC) was performed using a Superdex 200 (10/300) column (GE Healthcare) equilibrated in buffer 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. Pure proteins were concentrated up to ~20 g/L by ultra-filtration in Amicon cells (Millipore) using YM3 and YM10 membranes (Millipore) or centrifugal filter units (cut-off 3-10 kDa, Merck Millipore). All proteins were flash-frozen in liquid nitrogen and stored at -80 °C.

In some cases minor modifications were performed. (i) For the purification of Rap1b all buffers were supplemented with 10 mM MgCl<sub>2</sub>. (ii) For C3G-Cdc25H, HEPES (pH 7.0) was used instead of Tris-HCl (pH 7.9) for IMAC and the buffer was changed to 20 mM citrate (pH 6.0), 250 mM NaCl using Sephadex G25 resin (GE Healthcare) before cleavage with rTEV. Finally, the pure C3G-Cdc25H was separated from rTEV-His by SEC using a Superdex 200(10/300) column pre-equilibrated in the same buffer.

### **Purification of GST-tag proteins**

This procedure was applied for GST-SH3b and GST-RaIGDS constructs. Supernatants of the bacterial lysates containing GST fusion proteins were loaded into 6 ml glutathione agarose columns (ABT). Bound material was washed in two sequential washing steps with buffer phosphate-buffered saline (PBS) supplemented with 0.1% (v/v) Triton X-100 and PBS alone. GST fusion proteins were eluted with PBS supplemented with 20 mM reduced glutathione and samples were extensively dialyzed against 20 mM Tris-HCl (pH7.5), 150 mM NaCl. Finally proteins were concentrated to 5-10 g/L by ultrafiltration.

## Optimization of C3G full length purification

Initial attempts to purify C3G multidomain constructs resulted in low rates of purity and yield. Several bands of lower molecular weight recognized by N-terminal and C-terminal C3G specific antibodies were detected. This indicates that the protein is sensitive to proteolysis. These lower MW contaminants appeared before cell lysis, indicating that protein degradation starts during expression. To solve this problem, we performed an expression and purification condition screening.

C3G full-length cDNA was cloned into five vectors encoding GST and/or His-tags in N- and C-terminal positions (see Table M8). Purification of C3G with a single His-tag in C-terminal (vector pET22b-x2) or N-terminal position (pETEV15b) resulted in low purity. When the protein was produced with a N-terminal GST, we detected stochastic contamination with DNA. The purification of GST-C3G-His (pGEX-2xTEV-cHis) using a two-step affinity purification protocol (GST- and His- tags) resulted in small amounts of pure protein. Initially, we used this strategy to produce C3G samples. Later, we found that the Halo-tag stabilizes the expression of the His-Halo-C3G construct, which after purification shows similar purity to the obtained with the two-step affinity purification strategy, but with much higher yield. Since both methodologies yielded C3G samples with indistinguishable purity and activity, the origin of the protein is not specified. Finally, a representative gel showing the final purity of the samples obtained by single C-His purification (1), single GST purification (2), GST-His two step affinity purification (3), and His-Halo purification (4) is showed in Figure M3.

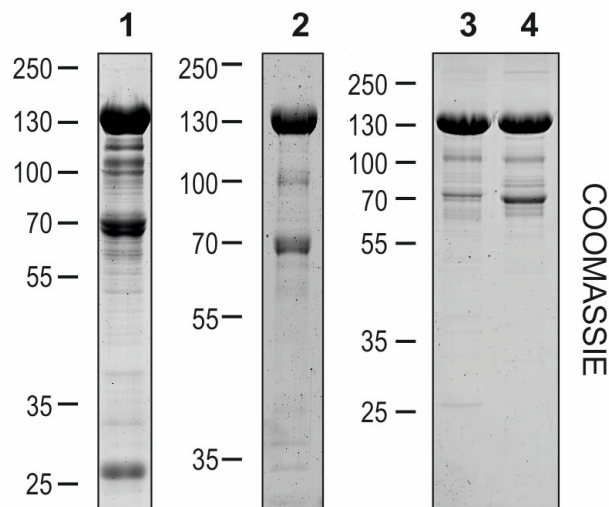
**Table M8.** Optimization of the purification conditions for C3G

Protein	Vector	Tag <sup>a</sup>	TEV cleavable?	Description
C3G-His	pET22	C-His	NO	Low MW contaminants
His-C3G	pETEV15b	N-His	YES	Low MW contaminants
GST-C3G	pGEX-TEV	N-GST	YES	Contamination with nucleic acids
GST-C3G-His	pGEX-2xTEV-cHis	N-GST / C-His	YES (both)	High purity <sup>b</sup> (~90%), low yield <sup>c</sup> (~0.2 g/L)
His-Halo-C3G	pETEV15b-Halo	N-His-Halo	YES	High purity <sup>b</sup> (~90%), high yield <sup>c</sup> (~4 g/L)

<sup>a</sup> C- and N- refer to the position of the tag

<sup>b</sup> The purity has been estimated by Coomassie staining

<sup>c</sup> The yield is given in grams per liter of initial culture



**Figure M3. Purity of C3G full-length samples produced from different vectors. (1) C3G-His, (2) GST-C3G, (3) GST-C3G-His, (4) His-Halo-C3G.** ~3  $\mu$ g of total protein was loaded in each lane. All samples show are ~130 kDa proteins because the GST and His-tags have been removed by rTEV-His cleavage.

### GST-C3G-His two-step affinity purification

Clarified bacterial lysates containing GST-C3G-His protein were loaded into a 5 ml Hi-Trap Ni<sup>2+</sup>-chelating column (GE Healthcare). The protein was eluted in an imidazole gradient (20 mM-500 mM) in buffer 20 mM Tris-HCl (pH 7.9), 500 mM NaCl. Subsequently, it was loaded into a 6 ml glutathione agarose column (ABT). Most of the C3G sample did not bind to the GST column. Bound material was washed in two sequential steps using buffer 20 mM Tris-HCl (pH 7.5), 300 mM NaCl with and without 0.1% Triton X-100. The first wash allowed the removal of a prominent 70 kDa impurity. The bound fraction (GST-C3G-His protein) was eluted with buffer 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, 20 mM reduced glutathione. The GST- and His-tags were removed by rTEV cleavage (3 hours, RT) during a dialysis step against buffer 20 mM Tris-HCl (pH 7.5), 300 mM NaCl which also allowed to remove the free glutathione also. A reverse GST chromatography was used to separate the free GST and C3G samples. The flow-through containing C3G was concentrated and subjected to SEC with a Superdex 200 (10/300) column (GE Healthcare) equilibrated in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl. C3G eluted as a single peak. Proteins were concentrated by ultrafiltration to the desired extent, flash-frozen in liquid nitrogen and stored at -80 °C until they were used.

## His-Halo-C3G purification

The protocol for His-Halo-C3G purification is similar to the IMAC general protocol with two small variations: (i) in the second IMAC (reverse), C3G is separated from the His-Halo and rTEV-His proteins, (ii) instead of 150 mM NaCl, buffers were supplemented with 300 mM NaCl for the final steps of the purification (dialysis and SEC).

## Protein concentration measurement

Protein concentration was determined spectrophotometrically by UV absorption at 280nm ( $A_{280}$ ) in an U-2001 UV-Visible double beam spectrophotometer (Hitachi) using quartz cuvettes with a 10 mm light path (Hellma GmbH). The concentration was calculated applying the Lambert–Beer law in which the molar ( $\epsilon_{280}$ ) and mass ( $E_{280}$ ) extinction coefficients were estimated from the theoretical sequences using ProtParam server (<https://web.expasy.org/protparam/>). The concentration of nucleotide-loaded Rap1b was estimated by Bradford assay using bovine serum albumin (BSA) as standard.

## 6. Protein modification

### Biotinylation

The *E. coli* biotin ligase BirA was used to conjugate biotin to the Avi-tag of the CrkL-Avi construct. 200  $\mu$ M CrkL-Avi was incubated with 1  $\mu$ M BirA, 1 mM biotin, and 4 mM ATP in buffer 50 mM Bicine (pH 8.3), 5 mM  $Mg(OAc)_2$  for ~4 hours at 30 °C. Biotinylation was confirmed by Fluorescent Streptavidin blotting (Streptavidin DyLight 680, Invitrogen) using Odyssey Infrared Imaging System (LI-COR).

### Src-mediated phosphorylation

The kinase domain of Src (SrcKD, residues 354-536) was used for C3G phosphorylation. A typical phosphorylation reaction was performed by diluting C3G proteins to a concentration, ranging between 20 to 100  $\mu$ M, in a buffer containing 50 mM Tris-HCl (pH7.5), 300 mM NaCl, 10 mM  $MgCl_2$ , 1mM DTT and 1 mM ATP. SrcKD was added at a final concentration of 1  $\mu$ M and the reaction was carried out for 20 min at 30 °C. In all cases the levels of phosphorylation of the samples were checked by

anti-pTyr immunoblotting. When required, p-C3G samples were purified from SrcKD by size exclusion chromatography using Superdex 200 (10/300) column equilibrated in buffer containing 20 mM Tris-HCl (pH7.5), 300 mM NaCl. The GEF activity of the phosphorylated constructs at 1  $\mu$ M was measured by nucleotide exchange experiments before and after the SEC, and no differences were found. Accordingly, the activity of the rest of the constructs was measured without purifying the samples by SEC.

## 7. Nucleotide exchange experiments

Nucleotide dissociation rates from Rap1b were studied by fluorescence spectroscopy using the GDP derivative mant-dGDP (2'-Deoxy-3'-O-(N'-methylantraniloyl) guanosine-5'-O-diphosphate sodium salt (Biolog) (Margarit *et al.* 2003, Rehmann 2006). Mant-fluorescence is sensitive to the chemical environment; when the nucleotide is bound to the GTPase (hydrophobic environment), the fluorescence intensity is approximately twice that of the free nucleotide (hydrophilic environment). Accordingly, mant-dGDP release from the GTPase can be monitored as a reduction of the fluorescence intensity along the time in the presence of an excess of free GDP or GTP nucleotides.

### Mant-dGDP loading of Rap1b

Rap1b (1-167) at 200  $\mu$ M was incubated with 2 mM mant-dGDP in 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 4 mM EDTA, 1mM DTT in a final volume of 500  $\mu$ L (1.5 hours, 4  $^{\circ}$ C). The loading reaction was stopped by addition of 10 mM  $MgCl_2$  (30 min, 4  $^{\circ}$ C). After addition of  $MgCl_2$ , precipitation was noticed and the sample was clarified by centrifugation at 16,000  $\times g$  (30 min, 4  $^{\circ}$ C). The free nucleotide was removed by SEC with a Superdex 200 (10/300) column pre-equilibrated in 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM  $MgCl_2$ . Fractions containing the protein were pooled and concentrated to  $\sim$ 200  $\mu$ M using centrifugal ultra-filtration cells (10kDa cut-off). Protein concentration was determined by Bradford assay using BSA for standard curve calibration.

## Fluorescence measurements and analysis of dissociation kinetics

Fluorescence measurements were done with a FLuoromax-3 Spectrofluorometer (Horiba-Jobin Yvon) using a Hellma 105.251-QS cuvette with a 3 by 3 mm light paths. Mant fluorophore was excited at 496 nm (1 nm bandwidth) and the light emitted was collected at 518 nm (10 nm bandwidth). Nucleotide exchange experiments were performed at 25 °C in 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM MgCl<sub>2</sub>. A maximum number of 1200 measurements per experiment were collected to prevent photobleaching. Rap1b loaded with mant-dGDP was used at 200 nM. The reaction was triggered by adding 40 μM GDP (200-fold molar excess with respect to Rap1b); manual addition of GDP resulted in initial delays of 20-30 seconds that were taken into account. C3G constructs were used at 1 μM. When indicated, reactions were supplemented with additional proteins. Data were recorded until the fluorescent signal reached a plateau (typically 15-120 min). Experiments were performed in triplicate unless specified. A single exponential decay function was fitted to the data:

$$I_t = A_0^{-k_{obs}t} + B$$

Where  $I_t$  is the fluorescence intensity at a given time,  $A_0$  is the amplitude of the change,  $B$  is the fluorescence intensity at infinite time,  $k_{obs}$  is the apparent nucleotide exchange rate and  $t$  is the time. In general, the single exponential decay fitted well to the data ( $R^2 > 0.9$ ). For simplicity, the maximum change ( $A_0$ ) in the fluorescence was normalized to 1 and used for representation.

## Analysis of dose response experiments

The dose response experiments were analyzed by fitting the following four parameter logistic curve:

$$k_{[bp]} = k_{min} + \frac{(k_{max} - k_{min})}{1 + \left(\frac{EC_{50}}{[bp]}\right)^{HillSlope}}$$

Where,  $k_{[bp]}$  is the  $k_{obs}$  observed at a concentration of binding partner;  $k_{min}$  is the minimum exchange rate;  $k_{max}$  is the maximum exchange rate;  $EC_{50}$  is the concentration

of the binding partner which produces the half maximal activation or inhibition;  $[bp]$  is the concentration of the binding partner (i.e. CrkL or C3G 537-646); and *HillSlope* is the parameter that defines the slope of the curve. In the case of the activation of C3G by CrkL the *HillSlope* was constrained to 1.

## 8. Sequence and structural analysis

### Structural analysis

The structure of the Cdc25H domain of C3G was modeled by homology using SWISS-MODEL server (<https://swissmodel.expasy.org/>). Structural models shown in the figures were obtained from the Protein Data Bank (PDB; <http://www.rcsb.org>). PyMOL Molecular Graphics System, Version 1.7.3.0 (Schrödinger) were used for analysis and representation of 3D structures. Secondary structure prediction was performed with the programs PORTER4.0 (<http://distillf.ucd.ie/porterpaleale/>), PHD ([https://npsa-prabi.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_phd.html](https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_phd.html)) and PSIPRED 4.0 (<http://bioinf.cs.ucl.ac.uk/psipred/>).

### C3G orthologs search with HMMER

The HMMER web server (<https://www.ebi.ac.uk/Tools/hmmer/>) (Potter *et al.* 2018) was used to search for C3G orthologs. HMMER uses profile hidden Markov models (HMM) to detect sequence similarities among protein sequences. HMM profiles contain information of the conservation, and the insertion and deletion rates for every position based on a multiple sequence alignment (MSA). In an iterative process, the initial identification of homologous sequences allows the finding of more distantly related sequences.

Initially, the search for C3G orthologs was done in the Reference Proteomes database (UniProt) using as bait the N-terminal domain (NTD, residues 89-250) of human C3G. Sequences from 119 species containing a C3G-like domain architecture, that is, containing NTD, REM and Cdc25h domains, were found and used to generate a HMM profile. Next, using that profile a HMM search identified 425 C3G-like sequences for 209 different species. For each species, the sequence with the best coverage of C3G sequence and the highest e-value was chosen. The final list of C3G orthologs of 209 species is shown in Appendix 3.



## Evolutionary conservation analysis

MSAs were performed with Clustal Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). Conservation scores based on the MSA of C3G full length (residues 4-1077) or the Cdc25H domain (residues 815-1077) were calculated with the ConSurf server (<http://consurf.tau.ac.il/2016/>) (Ashkenazy *et al.* 2016). In the case of the Cdc25H domain, conservation was color-coded on the solvent accessible surface of the 3D structural model.

## 9. Size exclusion chromatography coupled to multiple angle light scattering

Multiple Angle Light Scattering (MALS) is a technique that allows the determination of the absolute molar mass and the radius of gyration ( $R_g$ ) of proteins in solution (Wyatt 1993). Combined with size exclusion chromatography (SEC-MALS), it allows the determination of the mass of different species that may coexist in a polydisperse solution (Tarazona *et al.* 2003). SEC-MALS experiments were performed at the Molecular Interaction Facility at CIB, Madrid in collaboration with Dr. Carlos Alfonso Botello. Protein samples were analyzed using a Superdex 200 (10/300) Increase column (GE Healthcare) connected to an HPLC apparatus and coupled to a Static Multi Angle Light Scattering detector (DAWN-EOS, Wyatt Inc.) and a refractive index flow detector. The column was equilibrated in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl buffer, the flow was 0.5 ml/min, and the injection volume was 100  $\mu$ l. C3G, CrkL and C3G-CrkL samples were injected at concentrations specified in the results section. MWs were calculated from the light scattering and the refractive index in the center of the peak.

## 10. Analytical ultracentrifugation and dynamic light scattering

Analytical ultracentrifugation (AUC) techniques allow the determination of the MW, the hydrodynamic properties and the aggregation state of proteins in solution (Rivas *et al.* 1999). Sedimentation velocity experiments were performed at the Molecular Interaction Facility at CIB (CSIC-Madrid) in collaboration with Dr. Juan Roman Luque-Ortega using a Beckman Optima XLI centrifuge (Beckman Coulter Instrument, Inc) and 12 mm optical pass double sector cells. All experiments were done in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl buffer. Two experiments were performed:

(i) In the first experiment C3G and CrkL proteins were loaded independently and together at concentrations: C3G, 1.7  $\mu\text{M}$ ; and CrkL, 3, 6, 12, 24  $\mu\text{M}$ ; and the radial distribution of proteins was estimated by measuring the absorbance at 280 nm ( $A_{280}$ ) periodically every 5 min.

(ii) Next, we performed a similar experiment at lower concentrations of C3G and CrkL: C3G, 0.3  $\mu\text{M}$ ; and CrkL, 0.2, 0.4, 0.8, 1.5, and 3  $\mu\text{M}$ ; and the radial distributions were measured at 230 nm.

Proteins were subjected to centrifugation (48,000 rpm, 20 °C) and sedimentation coefficient distributions were obtained from the periodic sedimentation profiles by applying the  $c(s)$  method with SEDFIT software (Schuck *et al.* 2002). Partial specific volumes for proteins: C3G,  $v_{\text{C3G}} = 0.733 \text{ cm}^3/\text{g}$ , and CrkL,  $v_{\text{CrkL}} = 0.728 \text{ cm}^3/\text{g}$ , were calculated from the amino acidic composition of the proteins using the program SEDNTERP (<http://www.jphilo.mailway.com/>). The partial specific volume of C3G-CrkL mix,  $v_{\text{C3G-CrkL}} = 0.730 \text{ cm}^3/\text{g}$  was calculated as an average value of the estimated for the individual species. Finally, normalized sedimentation coefficients at 20 °C in water were calculated by applying the next equation:

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

In which  $\eta_{20,w}$  is the viscosity of water at 20 °C (0.0100 P);  $\rho_{20,w}$  is the water density at 20 °C (0.9982 g/mL);  $\eta_{20,b}$  is the viscosity of the buffer at 20 °C (0.0103 P); and  $\rho_{20,b}$  is the density of the buffer at 20 °C (1.0111 g/mL)

The binding of CrkL to C3G behaved as a rapid reversible equilibrium ( $K_{\text{off}} > 10^{-3} \text{ s}^{-1}$ ) displaying a "reaction boundary" (Gilbert *et al.* 1956). Three isotherms were generated from the coefficient sedimentation distribution ( $s_{\text{fast}}$ ,  $s_{\text{W}}$  and  $s_{\text{amplitude}}$ ) and a model assuming three symmetric binding sites was fitted using the SEDPHAT software (<http://www.analyticalultracentrifugation.com/sedphat/>) (Dam *et al.* 2005).

Dynamic light scattering (DLS) allows the estimation of the diffusion coefficient (Stetefeld *et al.* 2016). Experiments were performed with a DLS instrument DynaPro MS/X (Wyatt, Inc) equipped with a Peltier temperature regulation system. DLS signal was recorded every 20 sec at 97% laser power at 20 °C.

MWs of C3G and CrkL proteins were estimated by substituting the values of the diffusion and sedimentation coefficients in the terms of the Svedberg equation (Svedberg *et al.* 1940).

$$s = \frac{MD(1 - \bar{v}\rho)}{RT}$$

In which  $s$  is the sedimentation coefficient ( $s_{\text{C3G}} = 4.5 \text{ S}$ ;  $s_{\text{CrkL}} = 2.2 \text{ S}$ );  $M$  is the molar mass;  $D$  is the diffusion coefficient ( $D_{\text{C3G}} = 3.37 \times 10^{-7} \text{ cm}^2\text{s}^{-1}$ ;  $D_{\text{CrkL}} = 5.94 \times 10^{-7} \text{ cm}^2\text{s}^{-1}$ );  $v$  is the partial specific volume ( $v_{\text{C3G}} = 0.7328 \text{ ml/g}$ ;  $v_{\text{CrkL}} = 0.7283 \text{ ml/g}$ );  $\rho$  is the density of the solvent;  $R$  is the Gas constant; and  $T$  is the temperature.

## 11. Isothermal titration calorimetry

Isothermal Titration Calorimetry (ITC) is a powerful label-free technique used to determine the affinity and thermodynamic parameters of a binding reaction by measuring the heat evolved during the association of a ligand with its binding partner. ITC experiments were performed with a MicroCal VP-ITC machine (Malvern Panalytical). Protein samples were extensively dialyzed against the same buffer (20 mM Tris-HCl (pH 7.5), 300 mM NaCl or 20 mM NaPi (pH 7.5), 300 mM NaCl), flash-frozen in liquid nitrogen and stored at  $-80 \text{ }^\circ\text{C}$  until the experiment. Proteins were degassed with a ThermoVac machine (Malvern, 5 min,  $23 \text{ }^\circ\text{C}$ ) right before the experiment. The cell was filled with a 5 to 20  $\mu\text{M}$  solution of C3G and the syringe was filled with a 100 to 200  $\mu\text{M}$  solution of CrkL or CrkII constructs (the precise concentrations are specified in the results). Typically, ITC experiments were performed in triplicate with 36 injections of 8  $\mu\text{L}$  at  $\sim 7$  minutes intervals at  $25 \text{ }^\circ\text{C}$  with low feedback mode/gain mode. In some cases, the injection volume and the total number of injections were changed to improve the signal to noise ratio. Heat signals were automatically integrated with the program Nitpic (Keller *et al.* 2012) (<http://biophysics.swmed.edu/MBR/software.html>). Data analysis was performed individually and globally with the programs MicroCal Origin (version 7.0) and SEDPHAT (Zhao *et al.* 2015). In the case of SEDPHAT a model assuming three symmetric binding sites was used for the analysis of C3G WT interactions, and a 1:1 heteroassociation model was used for the C3G mutants displaying single Proline-rich motifs.

## 12. Global analysis of the C3G-CrkL interaction with SEDPHAT software

Isotherms derived from AUC and ITC experiments were fitted globally to a three symmetric binding site model with SEDPHAT program. This analysis was performed by Dr. Juan Ramon Luque Ortega in collaboration with the group of Dr. Peter Schuck.

## 13. Protein electrophoresis, Western blot and antibodies

Protein samples were denatured by addition of Laemli sample buffer 2x (125 mM Tris-HCl (pH 6.8), 4%SDS, 55% glycerol, 2%  $\beta$ -mercaptoethanol, 0.02 % bromophenol blue) and heating to 95 ° for 2-10 min. Next, they were loaded in SDS-polyacrylamide gels (9 to 14%) and separated electrophoretically (SDS-PAGE). Proteins were blotted to Immobilon-P membranes (Milipore) and detected by chemiluminescence (ECL-reagents, Bio-rad) or infra-red fluorescence (Odyssey Infrared Imaging System, LI-COR, Inc.) according to standard protocols. Table M9 contains the list of primary and secondary antibodies used in this work, respectively.

**Table M9.** Antibodies used in this work.

### Primary antibodies

Antibody	Host	Supplier	Reference number	dilution
Anti-His	Mouse	SIGMA	H1029	1:1000
Anti-C3G (G9)	Mouse	Santa Cruz Biotechnology	sc-393836	1:1000
Anti-HA (11)	Mouse	Covance	mms-101r	1:1000
Anti-FLAG-M2	Mouse	SIGMA	F1804	1:1000
Anti-GFP (B-2)	Mouse	Santa Cruz Biotechnology	sc-9996	1:1000
Anti-Rap1 (121)	Rabbit	Santa Cruz Biotechnology	sc-65	1:1000
Anti-pTyr (4G10)	Mouse	Merck Millipore	05-321	1:1000

pY, phospho-Tyrosine; HA, Internal region of the influenza hemagglutinin (HA); FLAG, Flag tag (DYKDDDDK). All antibodies were incubated >12 hours at 4 °C or 1 hour at RT.

### Secondary antibodies

Antibody	Detection	Host	Supplier	Reference number
Anti-mouse IgG HRP	ECL	Sheep	GE Healthcare	NXA931
Anti-rabbit IgG HRP	ECL	Goat	Santa Cruz Biotechnology	Sc-2004
Anti-mouse IgG DyLight 680	Odyssey	Goat	Pierce	35518
Anti-rabbit IgG DyLight 680	Odyssey	Goat	Pierce	35568
Anti-mouse IgG DyLight 800	Odyssey	Goat	Invitrogen	SA5-10176
Anti-rabbit IgG DyLight 800	Odyssey	Goat	Invitrogen	SA5-10036
Streptavidin DyLight 680	Odyssey	-	Invitrogen	21848

HRP, horseradish peroxidase; ECL, Chemiluminescent detection; Odyssey, Odyssey Infrared Imaging System. Secondary antibodies were incubated 1 hour at RT.

#### **14. Pull Down assays for the analysis of protein-protein interactions *in vitro***

Pull downs (PD) were performed with pure proteins or with bacterial lysates containing recombinant proteins. Typically, ~100  $\mu$ L of the lysates containing GST- or GST-His-tag proteins (~30-50  $\mu$ g of fusion protein) were incubated with 20  $\mu$ L glutathione agarose resin (20 min, 4  $^{\circ}$ C). The levels of expression of the proteins in the lysate were estimated by Coomassie staining for normalization. In case of using pure proteins, 30  $\mu$ g of GST-, His-tagged proteins, and CrkL-Avi were used in the presence of 0.01% (w/v) BSA to block unspecific binding. Bound material was washed 4 times with 500  $\mu$ L 20 mM Tris-HCl (pH7.5), 150 mM NaCl, 0.1% (v/v), Triton x-100 by centrifugation at 16.000  $xg$  (10 sec, 4  $^{\circ}$ C). Proteins were released from the resin by adding Laemli sample buffer 2x and heating to 95  $^{\circ}$ C for 2-10 min. The presence of GST- and His-proteins in the PD fraction and/or the initial lysates was detected by SDS-PAGE (7 to 14% polyacrylamide) followed by Coomassie staining (in the case of GST-proteins) or western blot (WB, in the case of His-proteins and CrkL-Avi). In general, 5% (v/v) of the PD sample was loaded in SDS-PAGE for detection.

#### **15. Mammalian cell culture and transfection**

HEK293T cell line was cultured in Dulbecco's Modified Eagle Medium (DMEM; Gibco or Sygma-Aldrich) supplemented with 10% fetal bovine serum (FBS; Life Technologies) and 1% penicillin/streptomycin (Life Technologies). Cells were grown at 37  $^{\circ}$ C in a humid atmosphere with 5% of CO<sub>2</sub>. For maintenance, cells grown to 90% confluence, in 10 cm plates, were detached by addition of trypsin (ThermoFisher) and split to 2-4 plates. For storage, cells were frozen in DMEM media supplemented with 10% dimethyl sulfoxide (DMSO) in liquid nitrogen. Cells were transfected at 60-70% confluence with 3-5  $\mu$ g of DNA pre-incubated with 2.5-ratio (w/w) of polyethylenimine (PEI; Polysciences Inc) in 150 mM NaCl for 30 min at RT. The expression of mEGFP-constructs was checked by fluorescent microscopy, and cells were processed 24-48 hours post-transfection.

## 16. Mammalian cell lysis, GST-pull downs and Rap1 activation assays

Typically,  $\sim 5 \times 10^6$  HEK293T cells (from a 10 cm plate) expressing the proteins of interest were washed twice with ice-cold phosphate buffered saline (PBS). Cell lysis was performed with lysis buffer: 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.5% Triton X-100, 1 mM  $\text{Na}_3\text{VO}_4$ , 25 mM NaF, 1mM phenylmethylsulfonyl fluoride (PMSF) and 1x protease inhibitor cocktail (Roche). Cell lysates were clarified by centrifugation at 16.000 xg (15 min, 4 °C) and the total protein content was estimated by Bradford assay, using BSA as standard. Typically, 1-2 mg of total protein (from a 10 cm plate) was incubated with 30  $\mu\text{g}$  of GST-SH3b. GST pull down was performed as described before (see *in vitro* pull downs). HA- or mEGFP- tagged proteins were detected in the cell lysates and in the PD fraction by using WB with specific antibodies (Table M9). The presence of the GST-SH3b construct in the PD sample was detected by Ponceau S staining of the same membrane used in WB.

### Rap1 activation assays in cells

Detection of Rap1-GTP in HEK293T cells was done with the pull down assay using GST-RalGDS-RBD proteins, following the protocol described in the Rap activation kit (Jena Bioscience,) with minor modifications. Cells were lysed in buffer MLB (25 mM HEPES (pH 7.5), 150 mM NaCl, 1% Igepal CA-630, 10 mM  $\text{MgCl}_2$ , 1 mM EDTA, 2% (v/v) glycerol, 1 mM  $\text{Na}_3\text{VO}_4$ , 1 mM PMSF, 1x protease inhibitor cocktail (Roche) and 1 mM DTT) supplemented with 30  $\mu\text{g}$  of purified GST-RalGDS-RBD protein. The pull down was then carried out as described above with the exception that MLB buffer was used to wash the material bound to the resin. Proteins were denatured by addition of Laemli sample buffer 2x, supplemented with 10 mM DTT for 12-20 hours at RT. Samples were loaded in 12% SDS-polyacrylamide gels. Total Rap1 in the cell lysates and Rap1-GTP in the GST-RalGDS pull downs were detected by immunoblot (Table M9). GST-RalGDS was detected in the pull downs by staining the same membrane used for the WB with Ponceau S. Quantification of the bands was performed with LICOR Image Studio Software, for Odyssey scans; and imageJ program (<https://imagej.nih.gov/>) for ECL scans.

## *BIBLIOGRAPHY*





- 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**(Pt 8): 858-871.
- Alsayed, Y., Uddin, S., Ahmad, S., Majchrzak, B., Druker, B. J., Fish, E. N. and Plataniias, L. C. (2000). "IFN-gamma activates the C3G/Rap1 signaling pathway." *J Immunol* **164**(4): 1800-1806.
- 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**(13): 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**(11): 3713-3722.
- Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. and Ben-Tal, N. (2016). "ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules." *Nucleic Acids Res* **44**(W1): W344-350.
- Asuri, S., Yan, J., Paranavitana, N. C. and Quilliam, L. A. (2008). "E-cadherin dis-engagement activates the Rap1 GTPase." *J Cell Biochem* **105**(4): 1027-1037.
- Barreira, M., Fabbiano, S., Couceiro, J. R., Torreira, E., Martinez-Torrecuadrada, J. L., Montoya, G., ... Bustelo, X. R. (2014). "The C-terminal SH3 domain contributes to the intramolecular inhibition of Vav family proteins." *Sci Signal* **7**(321): ra35.
- Barreira, M., Rodriguez-Fdez, S. and Bustelo, X. R. (2018). "New insights into the Vav1 activation cycle in lymphocytes." *Cell Signal* **45**: 132-144.
- Birge, R. B., Kalodimos, C., Inagaki, F. and Tanaka, S. (2009). "Crk and CrkL adaptor proteins: networks for physiological and pathological signaling." *Cell Commun Signal* **7**: 13.
- 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**(6691): 337-343.
- Bos, J. L. (2005). "Linking Rap to cell adhesion." *Curr Opin Cell Biol* **17**(2): 123-128.
- Bos, J. L. (2018). "From Ras to Rap and Back, a Journey of 35 Years." *Cold Spring Harb Perspect Med* **8**(2).
- Bourne, H. R., Sanders, D. A. and McCormick, F. (1991). "The GTPase superfamily: conserved structure and molecular mechanism." *Nature* **349**(6305): 117-127.
- Boykevisch, S., Zhao, C., Sondermann, H., Philippidou, P., Halegoua, S., Kuriyan, J. and Bar-Sagi, D. (2006). "Regulation of ras signaling dynamics by Sos-mediated positive feedback." *Curr Biol* **16**(21): 2173-2179.
- Braiman, A. and Isakov, N. (2015). "The Role of Crk Adaptor Proteins in T-Cell Adhesion and Migration." *Front Immunol* **6**: 509.
- 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**(17): 13118-13125.
- Cowburn, D. (2007). "Moving parts: how the adaptor protein CRK is regulated, and regulates." *Nat Struct Mol Biol* **14**(6): 465-466.
- Cruise, J. L., Rafferty, M. P. and Riehle, M. M. (1997). "Cell-cycle regulated expression of Rap1 in regenerating liver." *Biochem Biophys Res Commun* **230**(3): 578-581.
- Cherfils, J. and Chardin, P. (1999). "GEFs: structural basis for their activation of small GTP-binding proteins." *Trends Biochem Sci* **24**(8): 306-311.

- Cherfils, J. and Zeghouf, M. (2013). "Regulation of small GTPases by GEFs, GAPs, and GDIs." Physiol Rev **93**(1): 269-309.
- 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." Mol Cell Biol **20**(5): 1747-1758.
- Dam, J. and Schuck, P. (2005). "Sedimentation velocity analysis of heterogeneous protein-protein interactions: sedimentation coefficient distributions  $c(s)$  and asymptotic boundary profiles from Gilbert-Jenkins theory." Biophys J **89**(1): 651-666.
- Dayma, K., Ramadhas, A., Sasikumar, K. and Radha, V. (2012). "Reciprocal Negative Regulation between the Guanine Nucleotide Exchange Factor C3G and beta-Catenin." Genes Cancer **3**(9-10): 564-577.
- De Falco, V., Castellone, M. D., De Vita, G., Cirafici, A. M., Hershman, J. M., Guerrero, I., ... Santoro, M. (2007). "RET/papillary thyroid carcinoma oncogenic signaling through the Rap1 small GTPase." Cancer Res **67**(1): 381-390.
- de Jong, R., ten Hoeve, J., Heisterkamp, N. and Groffen, J. (1997). "Tyrosine 207 in CRKL is the BCR/ABL phosphorylation site." Oncogene **14**(5): 507-513.
- 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**(21): 2805-2810.
- 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**(6710): 474-477.
- DiNitto, J. P., Delprato, A., Gabe Lee, M. T., Cronin, T. C., Huang, S., Guilherme, A., ... Lambright, D. G. (2007). "Structural basis and mechanism of autoregulation in 3-phosphoinositide-dependent Grp1 family Arf GTPase exchange factors." Mol Cell **28**(4): 569-583.
- Duchniewicz, M., Zemojtel, T., Kolanczyk, M., Grossmann, S., Scheele, J. S. and Zwartkruis, F. J. (2006). "Rap1A-deficient T and B cells show impaired integrin-mediated cell adhesion." Mol Cell Biol **26**(2): 643-653.
- El Bouchikhi, I., Belhassan, K., Moufid, F. Z., Iraqui Houssaini, M., Bouguenouch, L., Samri, I., ... Ouldin, K. (2016). "Noonan syndrome-causing genes: Molecular update and an assessment of the mutation rate." Int J Pediatr Adolesc Med **3**(4): 133-142.
- Fernandez, V., Jares, P., Salaverria, I., Gine, E., Bea, S., Aymerich, M., ... Campo, E. (2008). "Gene expression profile and genomic changes in disease progression of early-stage chronic lymphocytic leukemia." Haematologica **93**(1): 132-136.
- Findlay, G. M. and Pawson, T. (2008). "How is SOS activated? Let us count the ways." Nat Struct Mol Biol **15**(6): 538-540.
- 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." Proc Natl Acad Sci U S A **103**(45): 16692-16697.
- Fukuyama, T., Ogita, H., Kawakatsu, T., Fukuhara, T., Yamada, T., Sato, T., ... Takai, Y. (2005). "Involvement of the c-Src-Crk-C3G-Rap1 signaling in the nectin-induced activation of Cdc42 and formation of adherens junctions." J Biol Chem **280**(1): 815-825.
- Gates, A., Hohenester, S., Anwer, M. S. and Webster, C. R. (2009). "cAMP-GEF cytoprotection by Src tyrosine kinase activation of phosphoinositide-3-kinase p110 beta/alpha in rat hepatocytes." Am J Physiol Gastrointest Liver Physiol **296**(4): G764-774.

- Gaulton, K. J., Willer, C. J., Li, Y., Scott, L. J., Conneely, K. N., Jackson, A. U., ... Mohlke, K. L. (2008). "Comprehensive association study of type 2 diabetes and related quantitative traits with 222 candidate genes." *Diabetes* **57**(11): 3136-3144.
- Gilbert, G. A. and Jenkins, R. C. (1956). "Boundary problems in the sedimentation and electrophoresis of complex systems in rapid reversible equilibrium." *Nature* **177**(4514): 853-854.
- Gingras, A. R., Lagarrigue, F., Cuevas, M. N., Valadez, A. J., Zorovich, M., McLaughlin, W., ... Ginsberg, M. H. (2019). "Rap1 binding and a lipid-dependent helix in talin F1 domain promote integrin activation in tandem." *J Cell Biol*.
- Gómez-Hernández, M. (2014). Caracterización de la organización estructural del factor de intercambio de nucleótido de guanina C3G (Thesis), Salamanca.
- Gotoh, T., Hattori, S., Nakamura, S., Kitayama, H., Noda, M., Takai, Y., ... et al. (1995). "Identification of Rap1 as a target for the Crk SH3 domain-binding guanine nucleotide-releasing factor C3G." *Mol Cell Biol* **15**(12): 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**(30): 18602-18607.
- Green, M. R., Gentles, A. J., Nair, R. V., Irish, J. M., Kihira, S., Liu, C. L., ... Alizadeh, A. A. (2013). "Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma." *Blood* **121**(9): 1604-1611.
- Groves, J. T. and Kuriyan, J. (2010). "Molecular mechanisms in signal transduction at the membrane." *Nat Struct Mol Biol* **17**(6): 659-665.
- 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**(5): 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**(28): 4885-4893.
- Guo, Y., Updegraff, B. L., Park, S., Durakoglugil, D., Cruz, V. H., Maddux, S., ... O'Donnell, K. A. (2016). "Comprehensive Ex Vivo Transposon Mutagenesis Identifies Genes That Promote Growth Factor Independence and Leukemogenesis." *Cancer Res* **76**(4): 773-786.
- Gureasko, J., Galush, W. J., Boykevisch, S., Sondermann, H., Bar-Sagi, D., Groves, J. T. and Kuriyan, J. (2008). "Membrane-dependent signal integration by the Ras activator Son of sevenless." *Nat Struct Mol Biol* **15**(5): 452-461.
- 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**(6): 938-948.
- Gutierrez-Herrero, S. (2018). Función de C3G en la señalización plaquetaria utilizando modelos de ratón transgénicos, Salamanca.
- Gutierrez-Herrero, S., Maia, V., Gutierrez-Berzal, J., Calzada, N., Sanz, M., Gonzalez-Manchon, C., ... Guerrero, C. (2012). "C3G transgenic mouse models with specific expression in platelets reveal a new role for C3G in platelet clotting through its GEF activity." *Biochim Biophys Acta* **1823**(8): 1366-1377.

- Hogan, C., Serpente, N., Cogram, P., Hosking, C. R., Bialucha, C. U., Feller, S. M., ... Fujita, Y. (2004). "Rap1 regulates the formation of E-cadherin-based cell-cell contacts." Mol Cell Biol **24**(15): 6690-6700.
- Hong, K. W., Jin, H. S., Lim, J. E., Ryu, H. J., Go, M. J., Lee, J. Y., ... Oh, B. (2009). "RAPGEF1 gene variants associated with type 2 diabetes in the Korean population." Diabetes Res Clin Pract **84**(2): 117-122.
- Hussain, H. and Chong, N. F. (2016). "Combined Overlap Extension PCR Method for Improved Site Directed Mutagenesis." Biomed Res Int **2016**: 8041532.
- Ichiba, T., Hashimoto, Y., Nakaya, M., Kuraishi, Y., Tanaka, S., Kurata, T., ... Matsuda, M. (1999). "Activation of C3G guanine nucleotide exchange factor for Rap1 by phosphorylation of tyrosine 504." J Biol Chem **274**(20): 14376-14381.
- Ichiba, T., Kuraishi, Y., Sakai, O., Nagata, S., Groffen, J., Kurata, T., ... 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**(35): 22215-22220.
- 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**(1): 125-136.
- Isakov, N. (2008). "A new twist to adaptor proteins contributes to regulation of lymphocyte cell signaling." Trends Immunol **29**(8): 388-396.
- Iwig, J. S., Vercoulen, Y., Das, R., Barros, T., Limnander, A., Che, Y., ... Kuriyan, J. (2013). "Structural analysis of autoinhibition in the Ras-specific exchange factor RasGRP1." Elife **2**: e00813.
- Jankowski, W., Saleh, T., Pai, M. T., Sriram, G., Birge, R. B. and Kalodimos, C. G. (2012). "Domain organization differences explain Bcr-Abl's preference for CrkL over CrkII." Nat Chem Biol **8**(6): 590-596.
- 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.
- Karandur, D., Nawrotek, A., Kuriyan, J. and Cherfils, J. (2017). "Multiple interactions between an Arf/GEF complex and charged lipids determine activation kinetics on the membrane." Proc Natl Acad Sci U S A **114**(43): 11416-11421.
- Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., ... the, F. C. (2001). "Functional annotation of a full-length mouse cDNA collection." Nature **409**(6821): 685-690.
- Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C. A. and Schuck, P. (2012). "High-precision isothermal titration calorimetry with automated peak-shape analysis." Anal Chem **84**(11): 5066-5073.
- Kirsch, K. H., Georgescu, M. M. and Hanafusa, H. (1998). "Direct binding of p130(Cas) to the guanine nucleotide exchange factor C3G." J Biol Chem **273**(40): 25673-25679.
- Klebe, C., Prinz, H., Wittinghofer, A. and Goody, R. S. (1995). "The kinetic mechanism of Ran-nucleotide exchange catalyzed by RCC1." Biochemistry **34**(39): 12543-12552.
- 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**(52): 32781-32787.
- Kobashigawa, Y. and Inagaki, F. (2012). "Structural biology: CrkL is not Crk-like." Nat Chem Biol **8**(6): 504-505.

- Kobashigawa, Y., Sakai, M., Naito, M., Yokochi, M., Kumeta, H., Makino, Y., ... Inagaki, F. (2007). "Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK." Nat Struct Mol Biol **14**(6): 503-510.
- Lafuente, E. M., van Puijenbroek, A. A., Krause, M., Carman, C. V., Freeman, G. J., Berezovskaya, A., ... Boussiotis, V. A. (2004). "RIAM, an Ena/VASP and Profilin ligand, interacts with Rap1-GTP and mediates Rap1-induced adhesion." Dev Cell **7**(4): 585-595.
- Lee, Y. K., Low-Nam, S. T., Chung, J. K., Hansen, S. D., Lam, H. Y. M., Alvarez, S. and Groves, J. T. (2017). "Mechanism of SOS PR-domain autoinhibition revealed by single-molecule assays on native protein from lysate." Nat Commun **8**: 15061.
- Lenzen, C., Cool, R. H., Prinz, H., Kuhlmann, J. and Wittinghofer, A. (1998). "Kinetic analysis by fluorescence of the interaction between Ras and the catalytic domain of the guanine nucleotide exchange factor Cdc25Mm." Biochemistry **37**(20): 7420-7430.
- Lepri, F., De Luca, A., Stella, L., Rossi, C., Baldassarre, G., Pantaleoni, F., ... Tartaglia, M. (2011). "SOS1 mutations in Noonan syndrome: molecular spectrum, structural insights on pathogenic effects, and genotype-phenotype correlations." Hum Mutat **32**(7): 760-772.
- Lin, K. B., Tan, P., Freeman, S. A., Lam, M., McNagny, K. M. and Gold, M. R. (2010). "The Rap GTPases regulate the migration, invasiveness and in vivo dissemination of B-cell lymphomas." Oncogene **29**(4): 608-615.
- 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**(29): 27301-27311.
- Maia, V., Ortiz-Rivero, S., Sanz, M., Gutierrez-Berzal, J., Alvarez-Fernandez, I., Gutierrez-Herrero, S., ... 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 Commun Signal **11**(1): 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**(7): 1229-1235.
- Manso, J. A., Garcia Rubio, I., Gomez-Hernandez, M., Ortega, E., Buey, R. M., Carballido, A. M., ... de Pereda, J. M. (2016). "Purification and Structural Analysis of Plectin and BPAG1e." Methods Enzymol **569**: 177-196.
- Margarit, S. M., Sondermann, H., Hall, B. E., Nagar, B., Hoelz, A., Pirruccello, M., ... Kuriyan, J. (2003). "Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS." Cell **112**(5): 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**(18): 3881-3891.
- Martin-Granado, V., Ortiz-Rivero, S., Carmona, R., Gutierrez-Herrero, S., Barrera, M., San-Segundo, L., ... Guerrero, C. (2017). "C3G promotes a selective release of angiogenic factors from activated mouse platelets to regulate angiogenesis and tumor metastasis." Oncotarget **8**(67): 110994-111011.
- Matsuda, M., Tanaka, S., Nagata, S., Kojima, A., Kurata, T. and Shibuya, M. (1992). "Two species of human CRK cDNA encode proteins with distinct biological activities." Mol Cell Biol **12**(8): 3482-3489.

- McLeod, S. J., Li, A. H., Lee, R. L., Burgess, A. E. and Gold, M. R. (2002). "The Rap GTPases regulate B cell migration toward the chemokine stromal cell-derived factor-1 (CXCL12): potential role for Rap2 in promoting B cell migration." *J Immunol* **169**(3): 1365-1371.
- Medeiros, R. B., Dickey, D. M., Chung, H., Quale, A. C., Nagarajan, L. R., Billadeau, D. D. and Shimizu, Y. (2005). "Protein kinase D1 and the beta 1 integrin cytoplasmic domain control beta 1 integrin function via regulation of Rap1 activation." *Immunity* **23**(2): 213-226.
- 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**(4945): 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**(8): e23681.
- Morin, R. D., Mendez-Lago, M., Mungall, A. J., Goya, R., Mungall, K. L., Corbett, R. D., ... Marra, M. A. (2011). "Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma." *Nature* **476**(7360): 298-303.
- Nath, P. R., Dong, G., Braiman, A. and Isakov, N. (2014). "Immunophilins control T lymphocyte adhesion and migration by regulating CrkII binding to C3G." *J Immunol* **193**(8): 3966-3977.
- 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**(5350): 560-563.
- Noguchi, H., Ikegami, T., Nagadoi, A., Kamatari, Y. O., Park, S. Y., Tame, J. R. and Unzai, S. (2015). "The structure and conformational switching of Rap1B." *Biochem Biophys Res Commun* **462**(1): 46-51.
- 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**(6): 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**(42): 30154-30162.
- Ohba, Y., Ikuta, K., Ogura, A., Matsuda, J., Mochizuki, N., Nagashima, K., ... Matsuda, M. (2001). "Requirement for C3G-dependent Rap1 activation for cell adhesion and embryogenesis." *EMBO J* **20**(13): 3333-3341.
- Okada, S., Matsuda, M., Anafi, M., Pawson, T. and Pessin, J. E. (1998). "Insulin regulates the dynamic balance between Ras and Rap1 signaling by coordinating the assembly states of the Grb2-SOS and CrkII-C3G complexes." *EMBO J* **17**(9): 2554-2565.
- 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**(45): 28179-28182.
- Ortiz-Rivero, S. (2017). Role of C3G in the differentiation and maturation of megakaryocytes, Universidad de Salamanca.
- 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." *EMBO J* **9**(8): 2351-2359.
- Peterson, M. E. and Long, E. O. (2008). "Inhibitory receptor signaling via tyrosine phosphorylation of the adaptor Crk." *Immunity* **29**(4): 578-588.

- Popovic, M. (2013). Regulation and Selectivity of Exchange Factors for G-proteins of the Ras-family (ISBN: 978-90-8891-615-1).
- Popovic, M., Rensen-de Leeuw, M. and Rehmann, H. (2013). "Selectivity of CDC25 homology domain-containing guanine nucleotide exchange factors." *J Mol Biol* **425**(15): 2782-2794.
- Potter, S. C., Luciani, A., Eddy, S. R., Park, Y., Lopez, R. and Finn, R. D. (2018). "HMMER web server: 2018 update." *Nucleic Acids Res* **46**(W1): W200-W204.
- Priego, N., Arechederra, M., Sequera, C., Bragado, P., Vazquez-Carballo, A., Gutierrez-Uzquiza, A., ... Porras, A. (2016). "C3G knock-down enhances migration and invasion by increasing Rap1-mediated p38alpha activation, while it impairs tumor growth through p38alpha-independent mechanisms." *Oncotarget* **7**(29): 45060-45078.
- Pufall, M. A. and Graves, B. J. (2002). "Autoinhibitory domains: modular effectors of cellular regulation." *Annu Rev Cell Dev Biol* **18**: 421-462.
- Radha, V., Mitra, A., Dayma, K. and Sasikumar, K. (2011). "Signalling to actin: role of C3G, a multitasking guanine-nucleotide-exchange factor." *Biosci Rep* **31**(4): 231-244.
- 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**(11): 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.
- Rehmann, H. (2006). "Characterization of the activation of the Rap-specific exchange factor Epac by cyclic nucleotides." *Methods Enzymol* **407**: 159-173.
- 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**(7076): 625-628.
- Rehmann, H., Prakash, B., Wolf, E., Rueppel, A., de Rooij, J., Bos, J. L. and Wittinghofer, A. (2003). "Structure and regulation of the cAMP-binding domains of Epac2." *Nat Struct Biol* **10**(1): 26-32.
- 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**(2): 157-165.
- Rivas, G., Stafford, W. and Minton, A. P. (1999). "Characterization of heterologous protein-protein interactions using analytical ultracentrifugation." *Methods* **19**(2): 194-212.
- Roose, J. P., Mollenauer, M., Ho, M., Kurosaki, T. and Weiss, A. (2007). "Unusual interplay of two types of Ras activators, RasGRP and SOS, establishes sensitive and robust Ras activation in lymphocytes." *Mol Cell Biol* **27**(7): 2732-2745.
- 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**(15): 10772-10778.
- Saleh, T., Jankowski, W., Sriram, G., Rossi, P., Shah, S., Lee, K. B., ... Kalodimos, C. G. (2016). "Cyclophilin A promotes cell migration via the Abl-Crk signaling pathway." *Nat Chem Biol* **12**(2): 117-123.
- 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**(5872): 343-347.

- Sarkar, P., Reichman, C., Saleh, T., Birge, R. B. and Kalodimos, C. G. (2007). "Proline cis-trans isomerization controls autoinhibition of a signaling protein." *Mol Cell* **25**(3): 413-426.
- Sarkar, P., Saleh, T., Tzeng, S. R., Birge, R. B. and Kalodimos, C. G. (2011). "Structural basis for regulation of the Crk signaling protein by a proline switch." *Nat Chem Biol* **7**(1): 51-57.
- Sawada, Y., Nakamura, K., Doi, K., Takeda, K., Tobiume, K., Saitoh, M., ... Ichijo, H. (2001). "Rap1 is involved in cell stretching modulation of p38 but not ERK or JNK MAP kinase." *J Cell Sci* **114**(Pt 6): 1221-1227.
- Scita, G., Nordstrom, J., Carbone, R., Tenca, P., Giardina, G., Gutkind, S., ... Di Fiore, P. P. (1999). "EPS8 and E3B1 transduce signals from Ras to Rac." *Nature* **401**(6750): 290-293.
- Schmidpeter, P. A. and Schmid, F. X. (2014). "Molecular determinants of a regulatory prolyl isomerization in the signal adapter protein c-CrkII." *ACS Chem Biol* **9**(5): 1145-1152.
- Schrödinger, L. "The PyMOL Molecular Graphics System, Version 1.7.3.0".
- 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**(2): 1096-1111.
- Sebzda, E., Bracke, M., Tugal, T., Hogg, N. and Cantrell, D. A. (2002). "Rap1A positively regulates T cells via integrin activation rather than inhibiting lymphocyte signaling." *Nat Immunol* **3**(3): 251-258.
- Seeliger, M. A., Young, M., Henderson, M. N., Pellicena, P., King, D. S., Falick, A. M. and Kuriyan, J. (2005). "High yield bacterial expression of active c-Abl and c-Src tyrosine kinases." *Protein Sci* **14**(12): 3135-3139.
- Sequera, C., Manzano, S., Guerrero, C. and Porras, A. (2018). "How Rap and its GEFs control liver physiology and cancer development. C3G alterations in human hepatocarcinoma." *Hepat Oncol* **5**(1): HEP05.
- Shivakrupa, Singh, R. and Swarup, G. (1999). "Identification of a novel splice variant of C3G which shows tissue-specific expression." *DNA Cell Biol* **18**(9): 701-708.
- 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**(52): 52188-52194.
- Simanshu, D. K., Nissley, D. V. and McCormick, F. (2017). "RAS Proteins and Their Regulators in Human Disease." *Cell* **170**(1): 17-33.
- Smit, L., van der Horst, G. and Borst, J. (1996). "Sos, Vav, and C3G participate in B cell receptor-induced signaling pathways and differentially associate with Shc-Grb2, Crk, and Crk-L adaptors." *J Biol Chem* **271**(15): 8564-8569.
- 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**(3): 393-405.
- Stefanini, L., Roden, R. C. and Bergmeier, W. (2009). "CalDAG-GEFI is at the nexus of calcium-dependent platelet activation." *Blood* **114**(12): 2506-2514.
- Stetefeld, J., McKenna, S. A. and Patel, T. R. (2016). "Dynamic light scattering: a practical guide and applications in biomedical sciences." *Biophys Rev* **8**(4): 409-427.
- Svedberg, T. and Pedersen, K. O. (1940). *The Ultracentrifuge*, Oxford University Press.



- Sztul, E., Chen, P. W., Casanova, J. E., Cherfils, J., Dacks, J. B., Lambright, D. G., ... Kahn, R. A. (2019). "ARF GTPases and their GEFs and GAPs: concepts and challenges." Mol Biol Cell **30**(11): 1249-1271.
- Tamada, M., Sheetz, M. P. and Sawada, Y. (2004). "Activation of a signaling cascade by cytoskeleton stretch." Dev Cell **7**(5): 709-718.
- Tanaka, S., Morishita, T., Hashimoto, Y., Hattori, S., Nakamura, S., Shibuya, M., ... 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**(8): 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**(6): 2356-2361.
- Tarazona, M. P. and Saiz, E. (2003). "Combination of SEC/MALS experimental procedures and theoretical analysis for studying the solution properties of macromolecules." J Biochem Biophys Methods **56**(1-3): 95-116.
- Tartof, K. D. and Hobbs, C. A. (1987). "Improved media for growing plasmid and cosmid clones." Bethesda Res Lab Focus **9**: 12.
- ten Hoeve, J., Morris, C., Heisterkamp, N. and Groffen, J. (1993). "Isolation and chromosomal localization of CRKL, a human crk-like gene." Oncogene **8**(9): 2469-2474.
- Torti, M. and Lapetina, E. G. (1994). "Structure and function of rap proteins in human platelets." Thromb Haemost **71**(5): 533-543.
- Ueno, H., Shibasaki, T., Iwanaga, T., Takahashi, K., Yokoyama, Y., Liu, L. M., ... Seino, S. (2001). "Characterization of the gene EPAC2: structure, chromosomal localization, tissue expression, and identification of the liver-specific isoform." Genomics **78**(1-2): 91-98.
- Utreras, E., Henriquez, D., Contreras-Vallejos, E., Olmos, C., Di Genova, A., Maass, A., ... Gonzalez-Billault, C. (2013). "Cdk5 regulates Rap1 activity." Neurochem Int **62**(6): 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**(7): 845-850.
- Vercoulen, Y., Kondo, Y., Iwig, J. S., Janssen, A. B., White, K. A., Amini, M., ... Roose, J. P. (2017). "A Histidine pH sensor regulates activation of the Ras-specific guanine nucleotide exchange factor RasGRP1." Elife **6**.
- Vetter, I. R. and Wittinghofer, A. (2001). "The guanine nucleotide-binding switch in three dimensions." Science **294**(5545): 1299-1304.
- Vigil, D., Cherfils, J., Rossman, K. L. and Der, C. J. (2010). "Ras superfamily GEFs and GAPs: validated and tractable targets for cancer therapy?" Nat Rev Cancer **10**(12): 842-857.
- 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**(12): 2139-2149.
- 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." Nat Genet **10**(3): 294-300.
- Webb, B. A., Chimenti, M., Jacobson, M. P. and Barber, D. L. (2011). "Dysregulated pH: a perfect storm for cancer progression." Nat Rev Cancer **11**(9): 671-677.

- Wilson, I. A., Niman, H. L., Houghten, R. A., Cherenson, A. R., Connolly, M. L. and Lerner, R. A. (1984). "The structure of an antigenic determinant in a protein." Cell **37**(3): 767-778.
- Wittinghofer, A. and Vetter, I. R. (2011). "Structure-function relationships of the G domain, a canonical switch motif." Annu Rev Biochem **80**: 943-971.
- Wu, X., Knudsen, B., Feller, S. M., Zheng, J., Sali, A., Cowburn, D., ... Kuriyan, J. (1995). "Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk." Structure **3**(2): 215-226.
- Wyatt, P. J. (1993). "Light Scattering and the Absolute Characterization of Macromolecules." Analytica Chimica Acta **272**(1): 1-40.
- Yan, W., Li, S. X., Wei, M. and Gao, H. (2018). "Identification of MMP9 as a novel key gene in mantle cell lymphoma based on bioinformatic analysis and design of cyclic peptides as MMP9 inhibitors based on molecular docking." Oncol Rep **40**(5): 2515-2524.
- Yang, Y., Yang, F., Wu, X., Lv, X. and Li, J. (2016). "EPAC activation inhibits acetaldehyde-induced activation and proliferation of hepatic stellate cell via Rap1." Can J Physiol Pharmacol **94**(5): 498-507.
- Yokote, K., Hellman, U., Ekman, S., Saito, Y., Ronnstrand, L., Saito, Y., ... Mori, S. (1998). "Identification of Tyr-762 in the platelet-derived growth factor alpha-receptor as the binding site for Crk proteins." Oncogene **16**(10): 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**(6676): 622-626.
- Yu, B., Martins, I. R., Li, P., Amarasinghe, G. K., Umetani, J., Fernandez-Zapico, M. E., ... Rosen, M. K. (2010). "Structural and energetic mechanisms of cooperative autoinhibition and activation of Vav1." Cell **140**(2): 246-256.
- Zhai, B., Huo, H. and Liao, K. (2001). "C3G, a guanine nucleotide exchange factor bound to adapter molecule c-Crk, has two alternative splicing forms." Biochem Biophys Res Commun **286**(1): 61-66.
- Zhang, L., Duan, H. B. and Yang, Y. S. (2017). "Knockdown of Rap2B Inhibits the Proliferation and Invasion in Hepatocellular Carcinoma Cells." Oncol Res **25**(1): 19-27.
- Zhang, Y. L., Wang, R. C., Cheng, K., Ring, B. Z. and Su, L. (2017). "Roles of Rap1 signaling in tumor cell migration and invasion." Cancer Biol Med **14**(1): 90-99.
- Zhao, H., Piszczek, G. and Schuck, P. (2015). "SEDPHAT--a platform for global ITC analysis and global multi-method analysis of molecular interactions." Methods **76**: 137-148.
- Zugaza, J. L., Lopez-Lago, M. A., Caloca, M. J., Dosil, M., Movilla, N. and Bustelo, X. R. (2002). "Structural determinants for the biological activity of Vav proteins." J Biol Chem **277**(47): 45377-45392.

## *APPENDICES*



## APPENDIX I

Tables A1, A2, and A3 include the information of the cDNA constructs used in this work, and the name of the primers used to clone them. Table A4 and A5 contain the list of primers used for amplification of cDNA fragments and mutagenesis respectively.

**Table A1.** Crk, Rap and SrcKD constructs for bacterial expression and primers used to clone them

Name	Forward primer	Reverse primer	Vector
CrkL	hCRKL-U001-NdeI	hCRKL-L303-BglII	pETEV15b
CrkL-Biotin	hCRKL-U001-NdeI	hCRKL-L303-BglII-noStop	pETEV15b-Avi
CrkL-SH2-SH3N	hCRKL-U001-NdeI	hCRKL-L182_BamHI	pETEV15b
CrkL-SH3N-SH3C	hCRKL-U125-NdeI	hCRKL-L303_BglII	pETEV15b
CrkL-SH3N	hCRKL-U125-NdeI	hCRKL-L182_BamHI	pETEV15b
CrkII	hCRKII-U001-NdeI	hCRKII-L304-BamHI	pETEV15b

**Table A2.** Constructs for mammalian expression and primers used to clone them

Name	Forward primer	Reverse primer	Vector
C3G-mEGFP WT & mutants	C3Gh-004-KKB-F (BamHI)	C3G-1077-noStop-R (NotI)	pEF1-mEGFP
C3G-CAAX-mEGFP	C3Gh-004-KKB-F (BamHI)	C3G-1077-noStop-R (NotI)	pEF1-mEGFP
REM-Cdc25H-mEGFP	C3Gh-670-F-BglII <sup>a</sup>	C3G-1077-noStop-R (NotI)	pEF1-mEGFP
REM-mEGFP	C3Gh-670-F-BglII <sup>a</sup>	C3Gh-814-noStop-R (NotI)	pEF1-mEGFP
Cdc25H-mEGFP	C3Gh-815-F-BglII <sup>a</sup>	C3G-1077-noStop-R (NotI)	pEF1-mEGFP
HA-REM-Cdc25H	C3Gh-670-F-BglII	C3G-1077-R (NotI)	pCEFHA
HA-REM	C3Gh-670-F-BglII	C3Gh-814-R (NotI)	pCEFHA
HA-Cdc25H	C3Gh-815-F-BglII	C3G-1077-R (NotI)	pCEFHA

<sup>a</sup> In these constructs BglII into BamHI ligation was performed.

**Table A3.** C3G constructs for bacterial expression and primers used to clone them

Name	Forward primer	Reverse primer	Vector
C3G WT & Mutants	C3G-004F (NcoI)	C3Gh-1077R-noStop-XhoI	pGEX-2xTEV-cHis
C3G-His	C3G-004F (NcoI)	C3Gh-1077R-noStop-XhoI	pET22b-x2
His-C3G	C3G-004F (NcoI)	C3G-1077R (BamHI)	pETEV15b-NcoI
GST-C3G	C3G-004F (NcoI)	C3G-1077R (BamHI)	pGEX-TEV
GST-C3G-His	C3G-004F (NcoI)	C3Gh-1077R-noStop-XhoI	pGEX-2xTEV-cHis
His-Halo-C3G	C3G-004F (NdeI)	C3G-1077R (BamHI)	pETEV15b-Halo
C3G-ΔNTD	C3G-246F (NcoI)	C3Gh-1077R-noStop-XhoI	pGEX-2xTEV-cHis
454-1077	C3Gh-454F (NdeI)	C3Gh-1077R-noStop-XhoI	pGEX-2xTEV-cHis
530-1077	C3Gh-530F (NdeI)	C3Gh-1077R-noStop-XhoI	pGEX-2xTEV-cHis
602-1077	C3Gh-602F (NdeI)	C3Gh-1077R-noStop-XhoI	pGEX-2xTEV-cHis
REM-Cdc25H	C3G-670F (NcoI)	C3Gh-1077R-noStop-XhoI	pETEV15b-NcoI
Cdc25H (WT & Mutants)	C3G-815F (NcoI)	C3Gh-1077R-noStop-XhoI	pETEV15b-NcoI
GST-SH3b	C3Gh-274F-NcoI	C3Gh-646R-Stop (BamHI)	pGEX-TEV
GST-274-578	C3Gh-274F-NcoI	C3Gh-578-Stop (For and Rev) <sup>a</sup>	pGEX-TEV
GST-274-500	C3Gh-274F-NcoI	C3Gh-500-Stop (For and Rev) <sup>a</sup>	pGEX-TEV
GST-274-371	C3Gh-274F-NcoI	C3Gh-371-Stop (For and Rev) <sup>a</sup>	pGEX-TEV
GST-372-646	C3Gh-372F-NcoI	C3Gh-646R-Stop (BamHI)	pGEX-TEV
GST-501-646	C3Gh-501F-NcoI	C3Gh-646R-Stop (BamHI)	pGEX-TEV
GST-537-646 (WT & Mutants)	C3Gh-537F-NdeI	C3Gh-646R-Stop (BamHI)	pGEX-TEV
537-646	C3Gh-537F-NdeI	C3Gh-646R-Stop (BamHI)	pETEV15b
GST-579-646	C3Gh-579F-NcoI	C3Gh-646R-Stop (BamHI)	pGEX-TEV
GST-501-578	C3Gh-501F-NcoI	C3Gh-578-Stop (For and Rev) <sup>a</sup>	pGEX-TEV
GST-501-536	C3Gh-501F-NcoI	C3Gh-536-Stop (For and Rev) <sup>a</sup>	pGEX-TEV
GST-537-588	C3Gh-537F-NdeI	C3Gh-588-Stop (For and Rev) <sup>a</sup>	pGEX-TEV
GST-537-578	C3Gh-537F-NcoI	C3Gh-578-Stop (For and Rev) <sup>a</sup>	pGEX-TEV
GST-537-569	C3Gh-537F-NdeI	C3Gh-569-Stop (For and Rev) <sup>a</sup>	pGEX-TEV
GST-537-560	C3Gh-537F-NdeI	C3Gh-560-Stop (For and Rev) <sup>a</sup>	pGEX-TEV
GST-545-646	C3Gh-545-NdeI	C3Gh-646R-Stop (BamHI)	pGEX-TEV
GST-545-569	C3Gh-545-NdeI	C3Gh-646R-Stop (BamHI)	pGEX-TEV
GST-545-560	C3Gh-545-NdeI	C3Gh-646R-Stop (BamHI)	pGEX-TEV

<sup>a</sup> Stop codons were introduced by mutagenesis in these constructs.

**Table A4.** Primers used to amplify cDNA fragments

Name	Sequence (5' to 3')
C3G-004F (NcoI)	CGGGGTACCATGGACTCTCAGCGTTCTCAT
C3G-246F (NcoI)	CGGGGTACCATGGGCAAGACGACTGGGATGTCACAGTC
C3Gh-454F (NdeI)	TGAGAATTCCATATGCCTGCTCTCCCCGAGAAG
C3Gh-530F (NdeI)	TGAGAATTCCATATGGCTCCTGAGTCAACCGGT
C3Gh-602F (NdeI)	TGAGAATTCCATATGGTGCAGGAGCTGGCCCC
C3G-670F (NcoI)	CGGGGTACCATGGCCCTCATTGACCACAACG
C3G-815F (NcoI)	CGGGGTACCATGGCACTCAGGTGTGCCACCTCC
C3Gh-274F-NcoI	CGGGGTACCATGGGCATTCGGGTGGTTGATAATGG
C3Gh-372F-NcoI	CATATGGGCAGCTCCATGGGGCAGTGCTCCCGG
C3Gh-501F-NcoI	CATATGGGCAGCTCCATGGTCCGTCCCCTACGCGC
C3Gh-537F-NdeI	CGGGGTACCCATATGCCAGAAAAACCACCTCCTCTACC
C3Gh-579F-NcoI	CATATGGGCAGCTCCATGGTACCAGCAGAAGAACAAGCTC
C3Gh-545F-NdeI	TTTTTCAGGGATCACATATGGAGAAGAAAAACAAACACATG
C3Gh-1077R-noStop-XhoI	GCGAATTCCTCGAGGGTCTTCTCTTCCCGGTC
C3G-1077R (BamHI)	CGGGGTACCGGATCCTAGGTCTTCTCTTCCCGG
C3Gh-646R-Stop (BamHI)	GCCGAATTCGGATCCCTAACTGCCGTCTCTGCTGTCCCTTC
C3Gh-578-Stop-F	GAACGAGCACATCTAGCAGCAGAAGAACAAG
C3Gh-578-Stop-R	CTTGTTCTTCTGCTGCTAGATGTGCTCGTTC
C3Gh-500-Stop-F	CAGCCCCGATCCCATAGGTCCCCTACGCGC
C3Gh-500-Stop-R	GCGCGTAGGGGACCTATGGGATCGGGGCTG
C3Gh-371-Stop-F	GGACAGGGACAGTTAGCAGTGCTCCCGG
C3Gh-371-Stop-R	CCGGGAGCACTGCTAACTGTCCCTGTCC
C3Gh-560-Stop-F	CAGTTGCTGGAGGACTAGTCGGAGCCGCAGCCCTC
C3Gh-560-Stop-R	GAGGGCTGCGGCTCCGACTAGTCCTCCAGCAACTG
C3Gh-569-Stop-F	GCAGCCCTCTATGTTCTAGCAGACGCCACAGAACGAG
C3Gh-569-Stop-R	CTCGTTCTGTGGCGTCTGCTAGAACATAGAGGGCTGC
C3Gh-536-Stop-F	GAGTCAACCGGTGACTAAGAAAAACCACCTCCTC
C3Gh-536-Stop-R	GAGGAGGTGGTTTTTCTTAGTCACCGGTTGACTC
hCRKL-U001-NdeI	TGACCATGGCATATGTCCTCCGCCAGGTTC
hCRKL-U125-NdeI	TGACCATGGCATATGCTGGAATATGTACGGACTCTG
hCRKL-L303-BglII	GCCGTGACAGATCTACTCGTTTTTCATCTGGGTTTTGAG
hCRKL-L303-BglII-noStop	GCCGTGACAGATCTCTCGTTTTTCATCTGGGTTTTGAG
hCRKL-L182-BamHI	GCCGTGACGGATCCCTACACAAGCTTTTCGACATAAGGG

**Table A4.** Primers used to amplify cDNA fragments

Name	Sequence (5' to 3')
C3Gh-004-KKB-F (BamHI)	TGAGGTACCGCCGCCACCATGGGATCCGACTCTCAGCGTTCTCATC
C3Gh-670-F-BglII	TGAGAATTCAGATCTCTCATTGACCACAACGAAATTATG
C3Gh-815-F-BglII	TGAGAATTCAGATCTCTCAGGTGTGCCACCTCC
C3G-1077-noStop-R (NotI)	GAATTCGCGGCCGCGGTCTTCTCTTCCCGGTC
C3Gh-814-noStop-R (NotI)	GAATTCGCGGCCGCTAGCTTCTTCTGGTCCAC
C3G-1077-R (NotI)	GAATTCGCGGCCGCTAGGTCTTCTCTTCCCGGTC
C3Gh-814-R (NotI)	GAATTCGCGGCCGCTATAGCTTCTTCTGGTCCACC

All reverse primers include stop codons unless specified.

**Table A5.** Mutagenesis primers

Name	Sequence (5' to 3')
C3Gh-NdeIX-F	GGAGGCAGCCACTCCTATGGTGGAGAGTCGC
C3Gh-NdeIX-R	GCGACTCTCCACCATAGGAGTGGCTGCCTCC
C3Gh-PP1A-F	GATAATGGTCCTGCAGCAGCAGCGGCACCCGCGAAAAGACAGTCGGCGC
C3Gh-PP1A-R	GCGCCGACTGTCTTTTCGCGGGTGCCGCTGCTGCTGCAGGACCATTATC
C3Gh-PP2A-F	GCAGACAGATACGGCAGCTGCTGCCGCCGAGGCGAAGCGCAGGAG
C3Gh-PP2A-R	CTCCTGCGCTTCGCTCGCCGCGGCGAGCAGCTGCCGTATCTGTCTGC
C3Gh-PP3A-F	GACCCAGAAAAAGCAGCTCCTGCAGCAGAGGCGAAAAACAAACACATGCTGGCC
C3Gh-PP3A-R	GGCCAGCATGTGTTTGTTCCTCTGCTGCAGGAGCTGCTTTTCTGGGTC
C3Gh-PP4A-F	GGCCCCGCGCGCCGCCGCAGCCCCCGCGCAGCGGCAG
C3Gh-PP4A-R	CTGCCGCTGCGCGGGGGCTGCGGCGGCCGCCGGGGCC
C3Gh-K546E/K547E-For	CACCTCCTCTACCAAGAGGAGGAAAACAAACACATGCTG
C3Gh-K546E/K547E-Rev	CAGCATGTGTTTGTTCCTCCTCTGGTAGAGGAGGTG
C3Gh-H550E/M551R-For	CCAGAGAAGAAAAACAAAAGAGAGGCTGGCCTACATGCAG
C3Gh-H550E/M551R-Rev	CTGCATGTAGGCCAGCCTCTCTTTGTTTTTCTTCTCTGG
C3Gh-Y554R-For	CAAACACATGCTGGCCCGCATGCAGTTGCTGGAG
C3Gh-Y554R-Rev	CTCCAGCAACTGCATGCGGGCCAGCATGTGTTTG
C3Gh-M555R/Q556R-For	CATGCTGGCCTACAAGCGGTTGCTGGAGGAC
C3Gh-M555R/Q556R-Rev	GTCTCCAGCAACCGCCTGTAGGCCAGCATG
C3Gh-E559R/D560R-For	CTACATGCAGTTGCTGCGGCGTACTCGGAGCCGCAG
C3Gh-E559R/D560R-Rev	CTGCGGCTCCGAGTAGCGCCGAGCAACTGCATGTAG
C3Gh-Y561R/S562R-For	GTTGCTGGAGGACCGCCGGGAGCCGCAGCCC
C3Gh-Y561R/S562R-Rev	GGGCTGCGGCTCCCGGCGGTCTCCAGCAAC



**Table A5.** Mutagenesis primers

Name	Sequence (5' to 3')
C3Gh-E563R/P564R-For	GGAGGACTACTCGCGGC <sup>C</sup> GCAGCCCTCTATG
C3Gh-E563R/P564R-Rev	CATAGAGGGCTGCGGCCGCGAGTAGTCCTCC
C3Gh-P566R-For	CTCGGAGCCGCAGC <sup>G</sup> CTCTATGTTCTACC
C3Gh-P566R-Rev	GGTAGAACATAGAGCGCTGCGGCTCCGAG
C3Gh-F569R/Y570R-For	CCGCAGCCCTCTATG <sup>CGCCG</sup> CCAGACGCCACAGAAC
C3Gh-F569R/Y570R-Rev	GTTCTGTGGCGTCTGGCGGCATAGAGGGCTGCGG
C3Gh-Q571R/T572R-For	CTCTATGTTCTACC <sup>GGA</sup> GCCACAGAACGAGC
C3Gh-Q571R/T572R-Rev	GCTCGTTCTGTGGCCTCCGGTAGAACATAGAG
C3Gh-Q574R-For	CTACCAGACGCCAC <sup>G</sup> GAACGAGCACATC
C3Gh-Q574R-Rev	GATGTGCTCGTTCCGTGGCGTCTGGTAG
C3Gh-E576R-For	CAGACGCCACAGAAC <sup>CG</sup> GCACATCTACCAGC
C3Gh-E576R-Rev	GCTGGTAGATGTGCCGTTCTGTGGCGTCTG
C3Gh-Y579R-For	CAGAACGAGCACATC <sup>CG</sup> CCAGCAGAAGAACAAG
C3Gh-Y579R-Rev	CTTGTTCCTTCTGCTGGCGGATGTGCTCGTTCTG
C3Gh-Q581R/K582E-For	GAGCACATCTACCAGC <sup>GGG</sup> GAGAACAAGCTCCTCATG
C3Gh-Q581R/K582E-Rev	CATGAGGAGCTTGTCTCCCGCTGGTAGATGTGCTC
C3Gh-E563R/P564R-For	GGAGGACTACTCGCGGC <sup>G</sup> GCAGCCCTCTATG
C3Gh-E563R/P564R-Rev	CATAGAGGGCTGCGGCCGCGAGTAGTCCTCC
C3Gh-H550E-For	CCAGAGAAGAAAAACAAA <sup>GAG</sup> ATGCTGGCCTACATGCAG
C3Gh-H550E-Rev	CTGCATGTAGGCCAGCATCTCTTTGTTTTTCTTCTCTGG
C3Gh-M551R-For	CCAGAGAAGAAAAACAAACACA <sup>G</sup> GCTGGCCTACATGCAG
C3Gh-M551R-Rev	CTGCATGTAGGCCAGCCTGTGTTTGTGTTTTCTTCTCTGG
C3Gh-M555R-For	CATGCTGGCCTACA <sup>G</sup> GCAGTTGCTGGAGGAC
C3Gh-M555R-Rev	GTCCCTCAGCAACTGCCTGTAGGCCAGCATG
C3Gh-Q556R-For	CATGCTGGCCTACATGC <sup>G</sup> GTTGCTGGAGGAC
C3Gh-Q556R-Rev	GTCCCTCAGCAACCGCATGTAGGCCAGCATG
<sup>o</sup> C3Gh-H550A-For	CCAGAGAAGAAAAACAAA <sup>GCC</sup> ATGCTGGCCTACATGCAG
C3Gh-H550A-Rev	CTGCATGTAGGCCAGCATGGCTTTGTTTTTCTTCTCTGG
C3Gh-M551A-For	CCAGAGAAGAAAAACAAACAC <sup>G</sup> CGCTGGCCTACATGCAG
C3Gh-M551A-Rev	CTGCATGTAGGCCAGCGGTGTTTGTGTTTTCTTCTCTGG
C3Gh-Y554A-For	CAAACACATGCTGGCC <sup>GC</sup> CATGCAGTTGCTGGAG
C3Gh-Y554A-Rev	CTCCAGCAACTGCATGGCGGCCAGCATGTGTTTG
C3Gh-M555A-For	CATGCTGGCCTACA <sup>G</sup> GCAGTTGCTGGAGGAC

**Table A5.** Mutagenesis primers

Name	Sequence (5' to 3')
C3Gh-M555A-Rev	GTCCTCCAGCAACTGCGCGTAGGCCAGCATG
C3Gh-Q556A-For	CATGCTGGCCTACATGGCGTTGCTGGAGGAC
C3Gh-Q556A-Rev	GTCCTCCAGCAACGCCATGTAGGCCAGCATG
C3Gh-Y554H-For	CAAACACATGCTGGCCACATGCAGTTGCTGGAG
C3Gh-Y554H-Rev	CTCCAGCAACTGCATGTGGGCCAGCATGTGTTTG
C3Gh-M555K-For	CATGCTGGCCTACAAGCAGTTGCTGGAGGAC
C3Gh-M555K-Rev	GTCCTCCAGCAACTGCTTGTAGGCCAGCATG
C3Gh-K922E-For	CATGAAGCACTTGCGGGAGCTGAATAACTTCAAC
C3Gh-K922E-Rev	GTTGAAGTTATTCAGCTCCCGCAAGTGCTTCATG
C3Gh-E956K-For	CTTCAGAGGGCCTGGCCAGTACTGCACACTGATCG
C3Gh-E956K-Rev	CGATCAGTGTGCAGTACTTGGCCAGGCCCTCTGAAG
C3Gh-T959A-For	CCTGGCCGAGTACTGCGCACTGATCGACAGCTCG
C3Gh-E959A-Rev	CGAGCTGTTCGATCAGTGCAGTACTCGGCCAGG
C3Gh-A973V-For	GAGCCTACCGGGCCGTCTCTCAGAGGTGG
C3Gh-A973V-Rev	CCACCTCTGAGAGGACGGCCCGGTAGGCTC
C3Gh-T883M-For	CCAACTTGACCCAGTTCAATGGAGCACTTCAACAACATG
C3Gh-T883M-Rev	CATGTTGTTGAAGTGCTCCATGAACTGGGTCAAGTTGG
C3Gh-S890F-For	GCACTTCAACAACATGTTCTACTGGGTCCGGTCC
C3Gh- S890F -Rev	GGACCGGACCCAGTAGAACATGTTGTTGAAGTGC
C3Gh-R894L-For	CATGTCCTACTGGGTCCGTCCATAATCATGTTAC
C3Gh-R894L-Rev	GTAACATGATTATGGACAGGACCCAGTAGGACATG
C3Gh-E863K-For	CTATAAAATAGAGATTCCTAAGGTTTTGCTTTGGGC
C3Gh-E863K-Rev	GCCCAAAGCAAACCTTAGGAATCTCTATTTTATAG
C3Gh-E873K-For	GGGCAAAGAGCAGAATAAGGAGAAGAGCCCCAAC
C3Gh-E873K-Rev	GTTGGGGCTCTTCTCCTTATTCTGCTCTTTTGCCC
C3Gh-R1023C-For	CATCCTCGACAGCATGTGCTGCTTCCAGCAGGCG
C3Gh-R1023C-Rev	CGCCTGCTGGAAGCAGCACATGCTGTTCGAGGATG
C3Gh-N878R-For	GAGGAGAAGAGCCCCAGATTGACCCAGTTCACGG
C3Gh-N878R-Rev	CCGTGAACTGGGTCAATCTGGGGCTCTTCTCCTC
C3Gh-Q881R-For	CCCAACTTGACCCGGTTCACGGAGCAC
C3Gh-Q881R-Rev	GTGCTCCGTGAACCGGGTCAAGTTGGG
C3Gh-E884R-For	CTTGACCCAGTTCACGAGGCACTTCAACAACATGTC
C3Gh-E884R-Rev	GACATGTTGTTGAAGTGCCTCGTGAAGTGGGTCAAG

**Table A5.** Mutagenesis primers

Name	Sequence (5' to 3')
C3Gh-N887R-For	G TTCACGGAGCACTTCA <b>GG</b> AACATGTCCTACTGGG
C3Gh-N887R-Rev	CCCAGTAGGACATGTTCCCTGAAGTGCTCCGTGAAC
C3Gh-N888R-For	CGGAGCACTTCAACA <b>GG</b> ATGTCCTACTGGGTCC
C3Gh-N888R-Rev	GGACCCAGTAGGACATCCTGTTGAAGTGCTCCG
C3Gh-Y891R-For	CAACAACATGTCC <b>CG</b> CTGGGTCCGGTCC
C3Gh-Y891R-Rev	GGACCGGACCCAGCGGGACATGTTGTTG
C3Gh-S934R-For	CTTGGCCATCCTC <b>CG</b> TGCCCTGGACTCGG
C3Gh-S934R-Rev	CCGAGTCCAGGGCACGGAGGATGGCCAAG
C3Gh-F967R-For	CAGCTCGTCCTCC <b>CG</b> CCGAGCCTACCGG
C3Gh-F967R-Rev	CCGGTAGGCTCGGC GGGAGGACGAGCTG
C3Gh-Y984R-For	CCGTGCATCCCG <b>CG</b> CCTGGGGCTGATC
C3Gh-Y984R-Rev	GATCAGCCCCAGGCGGGGATGCACGG
C3Gh-N1043R-For	GACATTATAAACTTCTT <b>CG</b> TGACTTCAGTGACCACCTG
C3Gh-N1043R-Rev	CAGGTGGTCACTGAAGTCACGGAAGAAGTTTATAATGTC
C3Gh-L855R-For	CTGCTGGATGCTGAGC <b>G</b> CTTCTATAAAAATAGAG
C3Gh-L855R-Rev	CTCTATTTTATAGAAGCGCTCAGCATCCAGCAG
C3Gh-K858E-For	GATGCTGAGCTCTTCTAT <b>G</b> AAATAGAGATTCTTGAGG
C3Gh-K858E-Rev	CCTCAGGAATCTCTATTTTCATAGAAGAGCTCAGCATC
C3Gh-E860R-For	GAGCTCTTCTATAAAAATA <b>CG</b> GATTCTGAGGTTTTGC
C3Gh-E860R-Rev	GCAAAACCTCAGGAATCCGTATTTTATAGAAGAGCTC
C3Gh-E874R-For	GCAAAAGAGCAGAATGAG <b>CG</b> GAAGAGCCCCAACTTGAC
C3Gh-E874R-Rev	GTCAAGTTGGGGCTCTTCCGCTCATTCTGCTCTTTTGC
C3Gh-D1044R-For	CATTATAAACTTCTTCAAT <b>CG</b> CTTCAGTGACCACCTGGC
C3Gh-D1044R-Rev	GCCAGGTGGTCACTGAAGCGATTGAAGAAGTTTATAATG
C3Gh-K915E-For	CTGCTCTTGAAGTTCATC <b>GA</b> GATCATGAAGCACTTGC
C3Gh-K915E-Rev	GCAAGTGCTTCATGATCTCGATGAACTTCAAGAGCAG
C3Gh-K918E-For	GAAGTTCATCAAGATCATG <b>G</b> AGCACTTGCGGAAGCTG
C3Gh-K918E-Rev	CAGCTTCCGCAAGTGCTCCATGATCTTGATGAACTTC
C3Gh-R921E-For	CAAGATCATGAAGCACTTG <b>GA</b> GAAGCTGAATAACTTCAAC
C3Gh-R921E-Rev	GTTGAAGTTATTAGCTTCTCCAAGTGCTTCATGATCTTG
C3Gh-D962R-For	GTACTGCACACTGATC <b>CG</b> CAGCTCGTCCTCCTTC
C3Gh-D962R-Rev	GAAGGAGGACGAGCTGCGGATCAGTGTGCAGTAC
C3Gh-S966R-For	GATCGACAGCTCGTCC <b>CG</b> CTTCCGAGCCTACCG

---

**Table A5.** Mutagenesis primers

---

<b>Name</b>	<b>Sequence (5' to 3')</b>
C3Gh-S966R-Rev	CGGTAGGCTCGGAAGCGGGACGAGCTGTCGATC
C3Gh-R968E-For	GCTCGTCCTCCTTCGAAGCCTACCGGGCCG
C3Gh-R968E-Rev	CGGCCCGGTAGGCTTCGAAGGAGGACGAGC

---

Nucleotides changed respect the original sequence are highlighted in red in the sequence of the forward primer

---



## pGEX-TEV

pGEX-5' →

...TGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATGAGAATCTTTATTTTCAGGGATCACAATATGGG  
 W Q A T F G G D H P P K S D E N L Y F Q G S H M G  
 GST TEV site

NcoI BamHI EcoRI NotI NdeI

CAGCTCCATGGGCGGGGATCCCCGAATTCCCGGGTTCGACTCGAGCGGCCGCATCGTGACTGACTGACGATCTGCCTCGC  
 S S M

CGTTTTCGGTGATGACGGTGAAAACCTCTGACACATG  
 ← pGEX-3'

## pGEX-2xTEV-cHis

pGEX-5' →

...TGGCAAGCCACGTTTGGTGGTGGCGACCATCCTCCAAAATCGGATGAGAATCTTTATTTTCAGGGATCACAATATGGG  
 W Q A T F G G D H P P K S D E N L Y F Q G S H M G  
 GST TEV site

NcoI BamHI EcoRI XhoI XhoI

CAGCTCCATGGGCGGGGATCCCCGAATTCCCGGGACTCGAGGAGAATCTTTATTTTCAGATCGAGCACCACCACCACCA  
 S S M L E E N L Y F Q I E H H H H H  
 TEV site His-tag

NotI

CCACTGAGCGGCCGCATCGTGACTGACTGACGATCTGCCTCGCGGTTTCGGTGATGACGGTGAAAACCTCTGACACATG  
 H Stop ← pGEX-3'

## pETEVI5b-Halo-Avi

T7 promoter → lac operator

BglIII XbaI

AGATCTCGATCCCAGCAAATTAATACGACTCACTATAGGGGAATTTGTGAGCGGATAACAATTCCTAGAAATAATTT  
 Rbs NcoI

TGTTTAACTTTAAGAAAGGAGATATAGCATGGGCAGCAGCCATCATCATCATCACCACCACAGCAGCGGCGAGAATCT  
 M G S S H H H H H H H S S G E N L

NcoI

TTCTTCTCAGGGCTCCATGGCAGAAATCGGTACTGGC...AGCGATAACGCGCATAGCAGCAGCGGCGAGAATCTTTATT  
 S S Q G S M A E I G T G S D N A H S S S G E N L Y  
 TEV site Halo tag TEV site

NdeI KpnI NheI SacI NotI BamHI

TTCAGGGCTCCCATATGGGAGGTACCGGTAGCGGTGAGCTCGCGGCCGCTGGATCCGGCGGTCTGAATGATATCTTCGAA  
 F Q G S H M G S G G L N D I F E

XhoI

GCCAGAAAATCGAGTGGCATGAAGATACCTAGCTCGAGATCCGGTGCTAACAAAGCCCAGAAAGGAGCTGAGTGTGG  
 A O K I E W H E D T Stop

Avi tag

CTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAACGGGTCTTGAGGGTTTTTTGCTGAAAG  
 ← T7 terminator primer

## pEF1-mEGFP and pEF1-mEGFP-CAAX

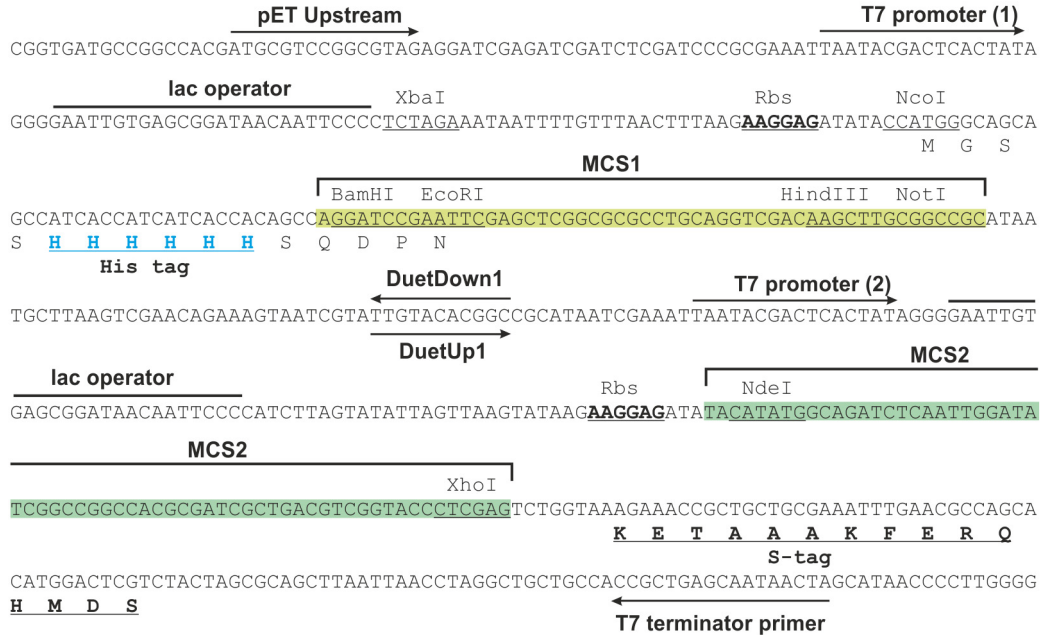
KpnI NcoI BamHI NdeI SpeI EcoRI NotI

GGT ACC GCC ACC ATG GGA TCC CAT ATG ACT AGT GAA TTC AGC GCGGCCGCT ATG GTG AGC AAG  
 Kozak G S H M T S E F S A A A M V S K  
 mEGFP

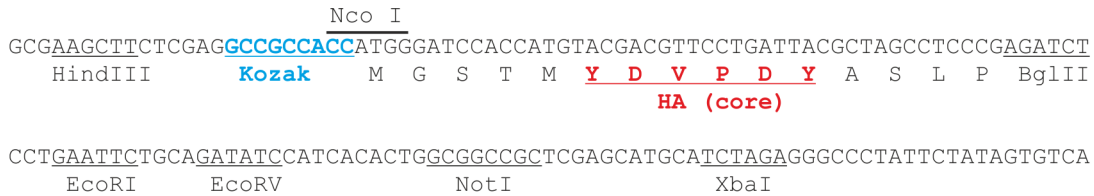
XbaI

...GGC ATG GAC GAG CTG TAC AAG TAA TCT AGA GGG CCC  
 G M D E L Y K Stop

## pETDUET



## pCEFLHA







## APPENDIX III

**Table A6.** Reference of the 209 sequences identified as C3G orthologs with HHMER

Clade	Subclade	Species	E-value	UniProt ID
Chordata	Mammalian	<i>Sarcophilus harrisii</i>	9.30E-72	G3WZS0_SARHA
Chordata	Mammalian	<i>Loxodonta africana</i>	3.00E-71	G3SSW6_LOXAF
Chordata	Mammalian	<i>Bos taurus</i>	6.10E-71	F1MJY2_BOVIN
Chordata	Mammalian	<i>Ovis aries</i>	6.30E-71	W5P986_SHEEP
Chordata	Mammalian	<i>Equus caballus</i>	7.80E-71	F6V778_HORSE
Chordata	Mammalian	<i>Equus caballus</i>	7.80E-71	F7D9U2_HORSE
Chordata	Mammalian	<i>Ursus maritimus</i>	9.40E-71	A0A384CRI6_URSM
Chordata	Mammalian	<i>Physeter catodon</i>	1.20E-70	A0A2Y9S3E4_PHYCD
Chordata	Mammalian	<i>Delphinapterus leucas</i>	1.40E-70	A0A2Y9NJ55_DELLE
Chordata	Mammalian	<i>Leptonychotes weddellii</i>	1.90E-70	A0A2U3XTA8_LEPWE
Chordata	Mammalian	<i>Ailuropoda melanoleuca</i>	2.00E-70	G1MDI0_AILME
Chordata	Mammalian	<i>Erinaceus europaeus</i>	2.00E-70	A0A1S3A6Y2_ERIEU
Chordata	Mammalian	<i>Tursiops truncatus</i>	2.00E-70	A0A2U4A417_TURTR
Chordata	Mammalian	<i>Balaenoptera acutorostrata scammoni</i>	2.80E-70	A0A384B489_BALAS
Chordata	Mammalian	<i>Odobenus rosmarus divergens</i>	2.90E-70	A0A2U3ZI88_ODORO
Chordata	Mammalian	<i>Felis catus</i>	3.30E-70	A0A2I2U2J0_FELCA
Chordata	Mammalian	<i>Mus musculus</i>	4.20E-70	Q91ZZ2_MOUSE
Chordata	Mammalian	<i>Cricetulus griseus</i>	4.90E-70	G3IFG2_CRIGR
Chordata	Mammalian	<i>Tarsius syrichta</i>	5.20E-70	A0A3Q0DU8_TARSY
Chordata	Mammalian	<i>Trichechus manatus latirostris</i>	9.90E-70	A0A2Y9QVW7_TRIMA
Chordata	Mammalian	<i>Canis lupus familiaris</i>	1.00E-69	F1PH97_CANLF
Chordata	Mammalian	<i>Mustela putorius furo</i>	1.10E-69	M3YMC6_MUSPF
Chordata	Mammalian	<i>Enhydra lutris kenyoni</i>	1.30E-69	A0A2Y9J7C9_ENHLU
Chordata	Mammalian	<i>Rattus norvegicus</i>	1.40E-69	F1M8L9_RAT
Chordata	Mammalian	<i>Dipodomys ordii</i>	1.50E-69	A0A1S3FLW8_DIPOR
Chordata	Mammalian	<i>Ornithorhynchus anatinus</i>	1.50E-69	F7EGC8_ORNAN
Chordata	Mammalian	<i>Cavia porcellus</i>	1.90E-69	H0VFR1_CAVPO
Chordata	Mammalian	<i>Ictidomys tridecemlineatus</i>	2.00E-69	A0A287D916_ICTTR
Chordata	Mammalian	<i>Monodelphis domestica</i>	3.20E-69	F6QKH4_MONDO
Chordata	Mammalian	<i>Cebus capucinus imitator</i>	4.40E-69	A0A2K5RV62_CEBCA
Chordata	Mammalian	<i>Myotis lucifugus</i>	4.50E-69	G1PC22_MYOLU
Chordata	Mammalian	<i>Saimiri boliviensis boliviensis</i>	4.90E-69	A0A2K6UXE9_SAIBB
Chordata	Mammalian	<i>Callithrix jacchus</i>	5.00E-69	F7HFD4_CALJA
Chordata	Mammalian	<i>Fukomys damarensis</i>	5.00E-69	A0A091DLG7_FUKDA
Chordata	Mammalian	<i>Aotus nancymae</i>	5.10E-69	A0A2K5DVJ3_AOTNA
Chordata	Mammalian	<i>Homo sapiens</i>	5.30E-69	RPGF1_HUMAN
Chordata	Mammalian	<i>Gorilla gorilla gorilla</i>	7.10E-69	G3QYV5_GORGO
Chordata	Mammalian	<i>Pan paniscus</i>	7.10E-69	A0A2R9C0A4_PANPA
Chordata	Mammalian	<i>Nomascus leucogenys</i>	7.40E-69	G1RQQ6_NOMLE
Chordata	Mammalian	<i>Propithecus coquereli</i>	7.50E-69	A0A2K6EGH8_PROCO
Chordata	Mammalian	<i>Pan troglodytes</i>	8.10E-69	A0A2I3SBV7_PANTR
Chordata	Mammalian	<i>Pongo abelii</i>	9.10E-69	H2PTR4_PONAB
Chordata	Mammalian	<i>Pteropus alecto</i>	9.90E-69	L5K9X4_PTEAL
Chordata	Mammalian	<i>Cercocebus atys</i>	1.40E-68	A0A2K5L7S3_CERAT
Chordata	Mammalian	<i>Macaca fascicularis</i>	1.40E-68	G7PRA8_MACFA
Chordata	Mammalian	<i>Macaca mulatta</i>	1.40E-68	F7HBH8_MACMU
Chordata	Mammalian	<i>Macaca nemestrina</i>	1.40E-68	A0A2K6BHW4_MACNE
Chordata	Mammalian	<i>Mandrillus leucophaeus</i>	1.40E-68	A0A2K5Z0K2_MANLE
Chordata	Mammalian	<i>Papio anubis</i>	1.40E-68	A0A2I3LVR3_PAPAN
Chordata	Mammalian	<i>Chlorocebus sabaeus</i>	1.80E-68	A0A0D9RQV8_CHLSB
Chordata	Mammalian	<i>Heterocephalus glaber</i>	7.50E-68	G5AT49_HETGA
Chordata	Mammalian	<i>Oryctolagus cuniculus</i>	7.80E-68	G1SCW1_RABIT
Chordata	Mammalian	<i>Sus scrofa</i>	1.70E-67	F1S0W0_PIG
Chordata	Mammalian	<i>Colobus angolensis palliatus</i>	1.80E-67	A0A2K5HW20_COLAP
Chordata	Mammalian	<i>Rhinopithecus bieti</i>	3.20E-67	A0A2K6LIA8_RHIBE
Chordata	Mammalian	<i>Rhinopithecus roxellana</i>	3.30E-67	A0A2K6RLQ3_RHIRO
Chordata	Mammalian	<i>Mesocricetus auratus</i>	2.50E-54	A0A3Q0DGG2_MESAU
Chordata	Mammalian	<i>Otolemur garnettii</i>	1.70E-46	H0WPA5_OTOGA
Chordata	Aves	<i>Columba livia</i>	1.10E-71	A0A2I0M3R2_COLL
Chordata	Aves	<i>Egretta garzetta</i>	1.50E-71	A0A091JHC9_EGRGA
Chordata	Aves	<i>Ficedula albicollis</i>	1.50E-71	U3JFG8_FICAL
Chordata	Aves	<i>Manacus vitellinus</i>	1.50E-71	A0A093S2Q0_9PASS
Chordata	Aves	<i>Patagioenas fasciata monilis</i>	1.50E-71	A0A1V4J7G7_PATFA
Chordata	Aves	<i>Corvus brachyrhynchos</i>	1.60E-71	A0A091F9Y5_CORBR
Chordata	Aves	<i>Lonchura striata domestica</i>	1.60E-71	A0A218V6J9_9PASE
Chordata	Aves	<i>Charadrius vociferus</i>	1.90E-71	A0A0A0AYN1_CHAVO
Chordata	Aves	<i>Opisthocomus hoazin</i>	2.10E-71	A0A091V8B4_OPIHO
Chordata	Aves	<i>Anas platyrhynchos</i>	2.30E-71	U3J7C7_ANAPL

**Table A6.** Reference of the 209 sequences identified as C3G orthologs with HHMER

Clade	Subclade	Species	E-value	UniProt ID
Chordata	Aves	<i>Aptenodytes forsteri</i>	2.30E-71	A0A087RGZ7_APTFO
Chordata	Aves	<i>Struthio camelus australis</i>	2.30E-71	A0A093HB10_STRCA
Chordata	Aves	<i>Cuculus canorus</i>	2.60E-71	A0A091GJC1_9AVES
Chordata	Aves	<i>Colinus virginianus</i>	3.10E-71	A0A226P0F5_COLVI
Chordata	Aves	<i>Calypte anna</i>	3.20E-71	A0A091HS62_CALAN
Chordata	Aves	<i>Nipponia nippon</i>	3.20E-71	A0A091UPS4_NIPNI
Chordata	Aves	<i>Dryobates pubescens</i>	7.10E-71	A0A093GB51_DRYPU
Chordata	Aves	<i>Callipepla squamata</i>	7.80E-71	A0A226NP31_CALSU
Chordata	Aves	<i>Tinamus guttatus</i>	9.80E-71	A0A099ZFB2_TINGU
Chordata	Aves	<i>Meleagris gallopavo</i>	1.60E-70	G1N2I3_MELGA
Chordata	Aves	<i>Taeniopygia guttata</i>	1.70E-70	H0Z3S0_TAEGU
Chordata	Aves	<i>Gallus gallus</i>	1.80E-70	E1BS61_CHICK
Chordata	Aves	<i>Amazona aestiva</i>	2.90E-35	A0A0Q3X9J0_AMAAE
Chordata	Crocodylia	<i>Alligator mississippiensis</i>	2.70E-67	A0A151MFQ4_ALLMI
Chordata	Crocodylia	<i>Alligator sinensis</i>	3.40E-67	A0A3Q0GRB5_ALLSI
Chordata	Testudines	<i>Pelodiscus sinensis</i>	4.60E-67	K7FMX2_PELSI
Chordata	Amphibia	<i>Xenopus tropicalis</i>	9.40E-65	F6YE36_XENTR
Chordata	Amphibia	<i>Xenopus laevis</i>	4.30E-64	Q6GP24_XENLA
Chordata	Actinopteri	<i>Pygocentrus nattereri</i>	3.20E-65	A0A3B4ED60_PYGNA
Chordata	Actinopteri	<i>Paramormyrops kingsleyae</i>	6.80E-65	A0A3B3QYD7_9TELE
Chordata	Actinopteri	<i>Ictalurus punctatus</i>	1.10E-64	A0A2D0T177 ICTPU
Chordata	Actinopteri	<i>Lepisosteus oculatus</i>	1.60E-63	W5M9X1_LEPOC
Chordata	Actinopteri	<i>Stegastes partitus</i>	3.80E-63	A0A3B4ZZA1_9TELE
Chordata	Actinopteri	<i>Salmo salar</i>	4.00E-63	A0A1S3M1U1_SALSA
Chordata	Actinopteri	<i>Danio rerio</i>	4.10E-63	E7F4B2_DANRE
Chordata	Actinopteri	<i>Oreochromis niloticus</i>	8.40E-62	I3JKA7_ORENI
Chordata	Actinopteri	<i>Pundamilia nyererei</i>	1.10E-61	A0A3B4F2N7_9CICH
Chordata	Actinopteri	<i>Scophthalmus maximus</i>	1.40E-61	A0A2U9CIG2_SCOMX
Chordata	Actinopteri	<i>Seriola dumerili</i>	1.50E-61	A0A3B4TXG4_SERDU
Chordata	Actinopteri	<i>Seriola lalandi dorsalis</i>	1.50E-61	A0A3B4XWR2_SERLL
Chordata	Actinopteri	<i>Takifugu rubripes</i>	1.50E-61	H2VCG7_TAKRU
Chordata	Actinopteri	<i>Tetraodon nigroviridis</i>	2.20E-61	H3CPY5_TETNG
Chordata	Actinopteri	<i>Austrofundulus limnaeus</i>	1.70E-60	A0A2I4B2N0_9TELE
Chordata	Actinopteri	<i>Xiphophorus maculatus</i>	2.70E-60	A0A3B5R6T1_XIPMA
Chordata	Actinopteri	<i>Poecilia latipinna</i>	3.20E-60	A0A3B3UA39_9TELE
Chordata	Actinopteri	<i>Poecilia formosa</i>	3.60E-60	A0A087XL47_POEFO
Chordata	Actinopteri	<i>Poecilia mexicana</i>	1.10E-59	A0A3B3XWA7_9TELE
Chordata	Actinopteri	<i>Xiphophorus couchianus</i>	2.20E-59	A0A3B5L1T6_9TELE
Chordata	Actinopteri	<i>Oryzias melastigma</i>	4.00E-59	A0A3B3DDP5_ORYME
Chordata	Actinopteri	<i>Oryzias latipes</i>	7.30E-59	H2LGG7_ORYLA
Chordata	Actinopteri	<i>Oncorhynchus mykiss</i>	3.50E-57	A0A060WGT3_ONCMY
Chordata	Actinopteri	<i>Periophthalmus magnuspinnatus</i>	1.70E-55	A0A3B3ZLX1_9GOBI
Chordata	Actinopteri	<i>Scleropages formosus</i>	1.80E-46	A0A0P7V340_9TELE
Chordata	Actinopteri	<i>Astyanax mexicanus</i>	6.10E-41	A0A3B1KBI6_ASTMX
Chordata	Actinopteri	<i>Gasterosteus aculeatus</i>	1.90E-40	G3P9F6_GASAC
Chordata	Coelacanthiformes	<i>Latimeria chalumnae</i>	2.90E-58	H3BDJ7_LATCH
Arthropoda	Insecta	<i>Ooceraea biroi</i>	1.70E-58	A0A026WC94_OOCBI
Arthropoda	Insecta	<i>Cyphomyrmex costatus</i>	2.20E-58	A0A195CM95_9HYME
Arthropoda	Insecta	<i>Trachymyrmex zeteki</i>	2.80E-58	A0A151XAU3_9HYME
Arthropoda	Insecta	<i>Acromyrmex echinator</i>	3.00E-58	F4WKD0_ACREC
Arthropoda	Insecta	<i>Trachymyrmex cornetzi</i>	3.00E-58	A0A195DIY2_9HYME
Arthropoda	Insecta	<i>Atta cephalotes</i>	3.10E-58	A0A158NQI2_ATTCE
Arthropoda	Insecta	<i>Atta colombica</i>	3.10E-58	A0A195BSS5_9HYME
Arthropoda	Insecta	<i>Trachymyrmex septentrionalis</i>	3.30E-58	A0A195EQG7_9HYME
Arthropoda	Insecta	<i>Dufourea novaeangliae</i>	4.30E-58	A0A154PJF2_9HYME
Arthropoda	Insecta	<i>Nasonia vitripennis</i>	4.30E-58	K7IZN6_NASVI
Arthropoda	Insecta	<i>Melipona quadrifasciata</i>	5.40E-58	A0A0N0BBJ6_9HYME
Arthropoda	Insecta	<i>Camponotus floridanus</i>	8.20E-58	E2AFG9_CAMFO
Arthropoda	Insecta	<i>Apis mellifera</i>	1.00E-57	A0A088ACI4_APIME
Arthropoda	Insecta	<i>Habropoda laboriosa</i>	3.20E-57	A0A0L7QYX3_9HYME
Arthropoda	Insecta	<i>Culex quinquefasciatus</i>	2.20E-55	B0XIC5_CULQU
Arthropoda	Insecta	<i>Zootermopsis nevadensis</i>	4.10E-55	A0A067QTZ8_ZOONE
Arthropoda	Insecta	<i>Aedes aegypti</i>	1.10E-54	Q16I23_AEDAE
Arthropoda	Insecta	<i>Pediculus humanus subsp. corporis</i>	7.20E-54	E0VHU1_PEDHC
Arthropoda	Insecta	<i>Anopheles albimanus</i>	2.50E-53	A0A182FT62_ANOAL
Arthropoda	Insecta	<i>Anopheles maculatus</i>	3.40E-53	A0A182SYB4_9DIPT
Arthropoda	Insecta	<i>Anopheles melas</i>	6.00E-53	A0A182U6F6_9DIPT
Arthropoda	Insecta	<i>Anopheles arabiensis</i>	7.80E-53	A0A182IDK1_ANOAR
Arthropoda	Insecta	<i>Anopheles gambiae</i>	1.20E-52	Q7PSV2_ANOGA
Arthropoda	Insecta	<i>Anopheles funestus</i>	1.30E-52	A0A182RVX9_ANOFN
Arthropoda	Insecta	<i>Anopheles minimus</i>	1.30E-52	A0A182W8D2_9DIPT
Arthropoda	Insecta	<i>Anopheles epiroticus</i>	2.40E-52	A0A182PL47_9DIPT

**Table A6.** Reference of the 209 sequences identified as C3G orthologs with HHMER

Clade	Subclade	Species	E-value	UniProt ID
Arthropoda	Insecta	<i>Anopheles merus</i>	2.70E-52	A0A182USQ8_ANOME
Arthropoda	Insecta	<i>Anopheles quadriannulatus</i>	4.30E-52	A0A182WRA8_ANOQN
Arthropoda	Insecta	<i>Anopheles culicifacies</i>	6.20E-52	A0A182MK97_9DIPT
Arthropoda	Insecta	<i>Anopheles dirus</i>	6.90E-52	A0A182NVF5_9DIPT
Arthropoda	Insecta	<i>Anopheles atroparvus</i>	7.20E-52	A0A182ZS8_9DIPT
Arthropoda	Insecta	<i>Anopheles stephensi</i>	1.20E-51	A0A182Y4W6_ANOST
Arthropoda	Insecta	<i>Drosophila sechellia</i>	4.40E-51	B4I052_DROSE
Arthropoda	Insecta	<i>Acyrtosiphon pisum</i>	6.20E-51	J9K728_ACYPI
Arthropoda	Insecta	<i>Drosophila virilis</i>	3.00E-50	B4M6Z9_DROVI
Arthropoda	Insecta	<i>Anopheles farauti</i>	3.10E-50	A0A182QIF9_9DIPT
Arthropoda	Insecta	<i>Drosophila mojavensis</i>	7.70E-50	A0A0Q9XPE3_DROMO
Arthropoda	Insecta	<i>Drosophila melanogaster</i>	2.10E-49	C3G_DROME
Arthropoda	Insecta	<i>Dendroctonus ponderosae</i>	5.20E-49	U4UA26_DENPD
Arthropoda	Insecta	<i>Drosophila ananassae</i>	8.10E-49	A0A0P9BUM9_DROAN
Arthropoda	Insecta	<i>Rhodnius prolixus</i>	8.50E-49	T1HVL0_RHOPR
Arthropoda	Insecta	<i>Drosophila willistoni</i>	3.00E-48	B4NCC9_DROWI
Arthropoda	Insecta	<i>Glossina austeni</i>	6.70E-48	A0A1A9V0K4_GLOAU
Arthropoda	Insecta	<i>Drosophila ficusphila</i>	8.70E-48	A0A1W4VEY8_DROFC
Arthropoda	Insecta	<i>Glossina pallidipes</i>	1.70E-47	A0A1B0A8G7_GLOPL
Arthropoda	Insecta	<i>Drosophila busckii</i>	1.80E-47	A0A0M4ESF0_DROBS
Arthropoda	Insecta	<i>Glossina morsitans morsitans</i>	1.80E-47	A0A1B0G205_GLOMM
Arthropoda	Insecta	<i>Tribolium castaneum</i>	2.80E-47	A0A139WN71_TRICA
Arthropoda	Insecta	<i>Glossina fuscipes fuscipes</i>	3.60E-47	A0A1A9XVI3_GLOFF
Arthropoda	Insecta	<i>Glossina palpalis gambiensis</i>	3.60E-47	A0A1B0B3W8_9MUSC
Arthropoda	Insecta	<i>Phlebotomus papatasi</i>	1.80E-45	A0A1B0D3V9_PHLPP
Arthropoda	Insecta	<i>Anopheles coluzzii</i>	2.20E-45	A0A182LAB5_9DIPT
Arthropoda	Insecta	<i>Glossina brevipalpis</i>	2.80E-45	A0A1A9W973_9MUSC
Arthropoda	Insecta	<i>Drosophila pseudoobscura</i>	1.40E-44	A0A0R3NYS3_DROPS
		<i>pseudoobscura</i>		
Arthropoda	Insecta	<i>Megaselia scalaris</i>	3.10E-44	T1GPF1_MEGSC
Arthropoda	Insecta	<i>Stomoxys calcitrans</i>	1.20E-43	A0A1I8PIT5_STOCA
Arthropoda	Insecta	<i>Drosophila persimilis</i>	1.30E-43	B4H2P5_DROPE
Arthropoda	Insecta	<i>Musca domestica</i>	1.40E-42	A0A1I8M2D7_MUSDO
Arthropoda	Insecta	<i>Diaphorina citri</i>	2.80E-41	A0A1S3D9U7_DIACI
Arthropoda	Insecta	<i>Clunio marinus</i>	5.90E-41	A0A1J1ILN1_9DIPT
Arthropoda	Insecta	<i>Aedes albopictus</i>	3.20E-29	A0A182H652_AEDAL
Arthropoda	Insecta	<i>Heliothis virescens</i>	5.30E-28	A0A2A4K8S8_HELVI
Arthropoda	Insecta	<i>Danaus plexippus plexippus</i>	9.20E-28	A0A212F3Y3_DANPL
Arthropoda	Insecta	<i>Papilio machaon</i>	6.70E-27	A0A0N1I6L6_PAPMA
Arthropoda	Insecta	<i>Papilio xuthus</i>	1.00E-26	A0A194PKK0_PAPXU
Arthropoda	Insecta	<i>Operophtera brumata</i>	9.00E-21	A0A0L7L86_9NEOP
Arthropoda	Insecta	<i>Harpegnathos saltator</i>	7.30E-18	E2C8C0_HARSA
Arthropoda	Insecta	<i>Drosophila simulans</i>	7.10E-09	B4R5P5_DROSI
Arthropoda	Insecta	<i>Cryptotermes secundus</i>	4.60E-07	A0A2J7PJ17_9NEOP
Arthropoda	Arachnida	<i>Stegodyphus mimosarum</i>	4.70E-49	A0A087TED5_9ARAC
Arthropoda	Arachnida	<i>Ixodes scapularis</i>	1.10E-37	B7QKE9_IXOSC
Arthropoda	Arachnida	<i>Tetranychus urticae</i>	1.30E-31	T1JUM2_TETUR
Arthropoda	Arachnida	<i>Tropilaelaps mercedesae</i>	2.30E-28	A0A1V9XBN6_9ACAR
Arthropoda	Branchiopoda	<i>Daphnia magna</i>	1.30E-20	A0A164RHE6_9CRUS
Arthropoda	Chilopoda	<i>Strigamia maritima</i>	9.90E-53	T1J7M1_STRMM
Nematoda	Nematoda	<i>Soboliphyme baturini</i>	4.70E-12	A0A183J2N6_9BILA
Nematoda	Nematoda	<i>Trichuris muris</i>	2.90E-07	A0A0N5E0P3_TRIMR
Nematoda	Nematoda	<i>Trichinella pseudospiralis</i>	2.60E-06	A0A0V0XNF5_TRIPS
Nematoda	Nematoda	<i>Trichinella britovi</i>	2.90E-06	A0A0V1CRX8_TRIBR
Nematoda	Nematoda	<i>Trichinella murrelli</i>	3.10E-06	A0A0V0TPS4_9BILA
Nematoda	Nematoda	<i>Trichinella spiralis</i>	3.30E-06	E5S397_TRISP
Nematoda	Nematoda	<i>Trichinella patagoniensis</i>	3.40E-06	A0A0V0ZN21_9BILA
Nematoda	Nematoda	<i>Trichinella nativa</i>	3.90E-06	A0A0V1KZS7_9BILA
Nematoda	Nematoda	<i>Trichinella nelsoni</i>	7.50E-06	A0A0V0S5D3_9BILA
Nematoda	Nematoda	<i>Trichinella papuae</i>	2.70E-05	A0A0V1M5Z5_9BILA
Nematoda	Nematoda	<i>Trichinella zimbabwensis</i>	4.30E-05	A0A0V1HAM0_9BILA
Brachiopoda	Brachiopoda	<i>Lingula unguis</i>	3.30E-35	A0A1S3KDD6_LINUN
Mollusca	Gastropoda	<i>Biomphalaria glabrata</i>	2.80E-27	A0A2C9JWL9_BIOGL
Mollusca	Bivalvia	<i>Mizuhopecten yessoensis</i>	2.30E-44	A0A210QQ99_MIZYE
Mollusca	Bivalvia	<i>Crassostrea gigas</i>	2.80E-41	K1RBH7_CRAGI
Mollusca	Cephalopoda	<i>Octopus bimaculoides</i>	6.80E-43	A0A0L8FY11_OCTBM
Cnidaria	Cnidaria	<i>Stylophora pistillata</i>	9.60E-11	A0A2B4RQI3_STYPI
Echinodermata	Echinodermata	<i>Strongylocentrotus purpuratus</i>	7.00E-41	W4YR90_STRPU
Echinodermata	Echinodermata	<i>Stichopus japonicus</i>	7.30E-28	A0A2G8KM08_STIJA
Tardigrada	Tardigrada	<i>Ramazzottius varieornatus</i>	2.30E-17	A0A1D1W6U7_RAMVA



## List of Figures

<b>INTRODUCTION</b> .....	<b>1</b>
<b>Figure I1.</b> The GTPase switch .....	4
<b>Figure I2.</b> Domain structure of GEFs of the Ras family .....	6
<b>Figure I3.</b> Mechanism of autoinhibition and activation of SOS1 .....	9
<b>Figure I4.</b> Mechanism of autoinhibition and activation of RasGRP1 .....	10
<b>Figure I5.</b> Mechanism of autoinhibition and activation of Epac2 .....	11
<b>Figure I6.</b> C3G primary structure .....	13
<b>Figure I7.</b> C3G isoforms.....	14
<b>Figure I8.</b> Structure and regulation of Crk proteins .....	20
<b>Figure I9.</b> Current general model of C3G activation in cells.....	22
<b>METHODS</b> .....	<b>31</b>
<b>Figure M1.</b> Engineering a TEV and His-tag in the pGEX-2xTEV-cHis vector .....	35
<b>Figure M2.</b> Strategy used to generate the triple-site mutants AAAP, APAA, AAPA and AAAP. .....	38
<b>Figure M3.</b> Purity of C3G full-length samples produced from different vectors.....	42
<b>RESULTS</b> .....	<b>53</b>
<b>Figure R1.</b> Nucleotide exchange kinetics of C3G constructs .....	56
<b>Figure R2.</b> The C3G SH3b domain binds to the Cdc25H domain and inhibits the GEF activity .....	57
<b>Figure R3.</b> The C-terminal segment of the SH3b region binds to the Cdc25H and blocks its catalytic GEF activity .....	60
<b>Figure R4.</b> Identification of residues in the SH3b essential for the binding to the Cdc25H....	62
<b>Figure R5.</b> Disruption of the SH3b/Cdc25H interaction produces constitutive activation of C3G in vitro and in cell cultures.....	65
<b>Figure R6.</b> Conservation analysis of the Cdc25H domains of SOS and RasGRP1.....	68
<b>Figure R7.</b> Mapping of the SH3b binding site on the Cdc25H domain.....	70
<b>Figure R8.</b> CrkL and phosphorylation of C3G stimulates GEF activity .....	73
<b>Figure R9.</b> CrkL competes with the Cdc25H for binding to SH3b .....	76
<b>Figure R10.</b> C3G constructs that lack the NTD have lower basal GEF activity .....	79
<b>Figure R11.</b> Mutants lacking the NTD show defective responses to CrkL and Src phosphorylation .....	81
<b>Figure R12.</b> Role of the NTD/REM interaction in the activation of C3G.....	83
<b>Figure R13.</b> Analysis of C3G activation by CrkL expression in cell cultures.....	85

<b>Figure R14.</b> Analysis of the stoichiometry of the C3G-CrkL complex by SEC-MALS .....	91
<b>Figure R15.</b> ITC analysis of CrkL binding to C3G full-length.....	93
<b>Figure R16.</b> Sedimentation velocity analysis of the C3G/CrkL interaction.....	95
<b>Figure R17.</b> Sedimentation boundary analysis of C3G-CrkL complex.....	97
<b>Figure R18.</b> Interaction of CrkL with C3G full-length containing individual Proline-rich motifs .....	99
<b>Figure R19.</b> Effect of CrkL binding to individual Proline-rich motifs on the activation of C3G .....	101
<b>Figure R20.</b> C3G is differentially activated by CrkL domains and CrkII .....	103
<b>Figure R21.</b> Binding of CrkL constructs and CrkII to C3G full-length .....	104
<b>DISCUSSION</b> .....	<b>107</b>
<b>Figure D1.</b> "Lid-lock" inhibitory mechanisms found in C3G, Arf GEFs and Vav1.....	111
<b>Figure D2.</b> Model of activation of C3G by CrkL-binding and phosphorylation .....	114
<b>Figure D3.</b> Model of the differential activation of C3G by CrkL and CrkII .....	115
<b>Figure D4.</b> Model of the allosteric activation of the REM-Cdc25H by the NTD.....	117
<b>Figure D5.</b> Different levels of C3G activation .....	118
<b>Figure D6.</b> Distribution of C3G missense SNVs in different cancer types .....	120
<b>Figure D7.</b> Model of activation of C3G .....	122

## List of Tables

<b>METHODS</b> .....	<b>31</b>
<b>Table M1.</b> cDNAs used in this work .....	31
<b>Table M2.</b> C3G bacterial expression constructs .....	32
<b>Table M3.</b> CrkL, CrkII, Rap1b and Src bacterial expression constructs .....	33
<b>Table M4.</b> C3G mammalian expression constructs .....	33
<b>Table M5.</b> Primers and process used to introduce the 6xHis and TEV sites in pGEX-2xTEV-cHis. The loss of restriction sites is indicated with strikethrough text. ....	35
<b>Table M6</b> Mutants generated in this work .....	38
<b>Table M7.</b> Primers used to mutate the Proline-rich motifs .....	39
<b>Table M8.</b> Optimization of the purification conditions for C3G .....	41
<b>Table M9.</b> Antibodies used in this work .....	50
<b>RESULTS</b> .....	<b>53</b>
<b>Table R1.</b> SNVs found in c-BioPortal and COSMIC databases (updated on June 2019) .....	64
<b>Table R2.</b> CrkL/C3G binding and thermodynamic parameters .....	93
<b>Table R3.</b> Hydrodynamic parameters of C3G and CrkL .....	94
<b>Table R4.</b> Binding parameters for CrkL/C3G interaction .....	97
<b>Table R5.</b> Thermodynamic parameters of the binding of CrkL to single-site C3G mutants .....	100
<b>Table R6.</b> Thermodynamic parameters of the binding of Crk constructs to C3G .....	105
<b>APPENDICES</b> .....	<b>139</b>
<b>Table A1.</b> Crk, Rap and SrcKD constructs for bacterial expression and primers used to clone them .....	141
<b>Table A2.</b> Constructs for mammalian expression and primers used to clone them .....	141
<b>Table A3.</b> C3G constructs for bacterial expression and primers used to clone them .....	142
<b>Table A4.</b> Primers used to amplify cDNA fragments .....	143
<b>Table A5.</b> Mutagenesis primers .....	144
<b>Table A6.</b> Reference of the 209 sequences identified as C3G orthologs with HHMER .....	153







