# UNIVERSIDAD DE SALAMANCA

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


# Mechanisms of activation of the guanine nucleotide exchange factor C3G

---

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR


**Antonio Rodríguez Blázquez**

Bajo la dirección del Doctor

**José María de Pereda Vega**



Salamanca, 2023

**Dr. JOSÉ MARÍA DE PEREDA VEGA**, Investigador Científico del Consejo Superior de Investigaciones Científicas (CSIC) en el Instituto de Biología Molecular y Celular del Cáncer (IBMCC, CSIC-Universidad de Salamanca).

**CERTIFICA**:

Que D. Antonio Rodríguez Blázquez, graduado en Biología por la Universidad de Salamanca, ha realizado bajo su dirección el trabajo de Tesis Doctoral que lleva por título "***Mechanisms of 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 04 de octubre de 2023

Dr. José María de Pereda Vega

Director de la tesis

Antonio Rodríguez Blázquez ha realizado esta tesis doctoral siendo beneficiario de un contrato de Personal Investigador en Formación de la Universidad de Salamanca, Programa III, cofinanciada por el Banco de Santander, durante el periodo de enero de 2019 a junio de 2023; y de un contrato financiado por un proyecto de la convocatoria de Ayudas Especial para la Preparación de Proyectos del Consejo Superior de Investigaciones Científicas, entre julio y septiembre de 2023.

Este trabajo es parte de los proyectos del I+D+i "*Structural and mechanistic basis for the activation of Crk-C3G-Rap1 signaling and its regulation in diseases*" (PID2019-105763GB-I00) y "*Structural and mechanistic basis of the regulation of Rap1 signaling*" (PID2022-136322NB-I00) financiados por el la Agencia Estatal de Investigación (AEI), Ministerio de Ciencia e Innovación, MCIN/ AEI/10.13039/501100011033. El trabajo también se encuadra en el proyecto "*Papel de C3G en tumores hematopoyéticos y en angiogénesis mediada por plaquetas. Evaluación de su uso como diana terapéutica*" (SA078P20) del Programa de Apoyo a Proyectos de Investigación de la Consejería de Educación de la Junta de Castilla y León. Estos proyectos han sido co-financiados por el Fondo Europeo de Desarrollo Regional "*FEDER Una manera de hacer Europa*". El Centro de Investigación del Cáncer (IBMCC-CIC), dónde se ha realizado este trabajo, ha recibido financiación del Programa de Apoyo a Planes Estratégicos de Investigación de Estructuras de Investigación de Excelencia cofinanciado por la Junta de Castilla y León y FEDER (CLC–2017–01).

Antonio Rodríguez Blázquez realizó parte del trabajo de esta tesis durante una estancia de tres meses de duración, entre septiembre y diciembre de 2021, 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 del Programa IV de la Universidad de Salamanca para la Movilidad del Personal Investigador.

**A mis padres y mi hermana**

*“Mientras el corazón tiene deseo,*
*la imaginación conserva ilusiones”*

*René de Chateaubriand*

**THESIS SUMMARY**

C3G (also known as RapGEF1) is a ubiquitously expressed guanine nucleotide exchange factor (GEF) that activates mainly the small GTPase Rap1 to promote integrin-mediated cell adhesion. C3G also regulates cell migration, actin remodeling, proliferation, apoptosis, differentiation, and exocytosis. C3G is a modular protein with three structurally and functionally distinct regions. The N-terminal domain (NTD) interacts with E-cadherin and participates in the self-regulation of C3G. The central or SH3b region is mostly flexible and is involved in protein-protein interactions. This region contains five proline-rich motifs (PRMs), P0 to P4, which are binding sites for SH3 domains. The Crk and CrkL adaptor proteins bind to four of these sites (P1, P2, P3 and P4). The C-terminal part contains a REM and a Cdc25H domain that form the catalytic region. The GEF activity of C3G is regulated by two intramolecular interactions. First, binding of the NTD to the REM domain contributes positively to the activity of C3G. Secondly, the C-terminal segment of the SH3b downstream the P3 is an autoinhibitory region (AIR) that binds to the Cdc25HD and blocks the GEF activity. C3G is physiologically activated in response to stimuli that operate through the activation of tyrosine kinases. These signals induce the recruitment of C3G to signaling sites at the membrane, which is mediated by Crk proteins. Activation of the GEF activity of C3G involves tyrosine phosphorylation by Src family and other kinases, and the interaction with Crk proteins. Despite the general processes involved in C3G activation were known, the detailed mechanisms were poorly understood.

In this Thesis, we have combined biochemical, biophysical and cell biology approaches to gain understanding of the structural and mechanistic basis of the autoinhibition and physiological activation mechanism of C3G activation within the C3G-Rap1 pathway. We have demonstrated that CrkL binds differentially to the PRMs P1, P2, P3 and P4 in full-length C3G. In resting conditions, sites P1 and P2 are constitutively fully accessible. Exposure of the P3 site is linked to the activation state of C3G. The P4 site is partially accessible independently of the activation state of C3G. The sites P1 and P2 do not participate in the direct activation of C3G. Instead, these sites participate in the constitutive interaction with Crk proteins in HEK293T and Jurkat cell lines, and are required for the recruitment of C3G to the plasma membrane upon TCR stimulation of Jurkat cells. Binding of CrkL to the P3 and P4 is required and sufficient for the direct activation of C3G by CrkL through the release of the autoinhibitory interaction. Binding of CrkL to the P3 is the main activation event. Collectively, P1 and P2 are recruitment sites and P3 and P4 are activation sites. Tyrosine phosphorylation of C3G alone

causes minimal activation, yet it is essential for the Crk-mediated stimulation. Phosphorylation by Src sensitizes C3G, which is activated at lower concentrations of CrkL and to higher activity levels than unphosphorylated C3G. We have mapped the main Src phosphorylation sites in C3G at Y329, Y504, Y579 and Y590. Despite CrkL interacts with the P3 and P4 primarily by the SH3N domain, the SH2 and SH3C domains of CrkL also contribute to the activation of C3G. In particular, the SH2 domain plays a key role in the activation of phosphorylated-C3G through a non-canonical low affinity interaction with phospho-Y590, which apparently stabilizes the interaction of CrkL with the activation sites. We have also shown that CrkL induces a stronger activation of C3G than CrkII, and these differences are apparently due to the different inter-domain arrangements of Crk proteins. We have also produced an experimentally validated structural model of autoinhibited C3G, which revealed additional contacts of the AIR with the REM domain. Based on the results, propose a multi-step cycle for the physiological activation and de-activation of C3G. Finally, we have applied the structural and functional data to identify three cancer somatic missense mutations that cause constitutive activation of C3G: Y570N and Y590N were found in non-Hodgkin lymphoma patients, and Y579C was described in a thyroid carcinoma. These mutations expand the repertoire of acquired alterations that deregulate C3G-Rap1 in cancers. In summary, the results of this Thesis contribute to understand the mechanisms of C3G-Rap1 signaling in healthy tissues and in diseases.

**RESUMEN**

C3G (RapGEF1) es un factor de intercambio de nucleótidos de guanina (GEF) expresado de forma ubicua que activa principalmente la pequeña GTPasa Rap1 promoviendo la adhesión celular mediada por integrinas. C3G también regula migración celular, remodelación de actina, proliferación, apoptosis, diferenciación y exocitosis. C3G es una proteína modular con tres regiones estructural y funcionalmente diferenciadas. El dominio N-terminal (NTD) interacciona con E-cadherina y participa en la autorregulación de C3G. La región central o SH3b es mayoritariamente flexible y participa en las interacciones proteína-proteína. Esta región contiene cinco motivos ricos en prolina (PRM), P0 a P4, que son sitios de unión para dominios SH3. Las proteínas adaptadoras Crk y CrkL se unen a cuatro de estos sitios (P1, P2, P3 y P4). La zona C-terminal contiene un dominio REM y un dominio Cdc25H que forman la región catalítica. La actividad GEF de C3G está regulada por dos interacciones intramoleculares. En primer lugar, la unión del NTD al dominio REM contribuye positivamente a la actividad de C3G. En segundo lugar, el segmento C-terminal de SH3b tras el P3 es una región autoinhibitoria (AIR) que se une al Cdc25HD y bloquea la actividad del GEF. C3G se activa fisiológicamente en respuesta a estímulos que operan mediante la activación de tirosina quinasas. Estas señales inducen el reclutamiento de C3G a sitios de señalización en la membrana, este proceso está mediado por proteínas Crk. La activación de la actividad GEF de C3G implica su fosforilación en tirosina quinasas de la familia Src y otras quinasas, y la interacción con las proteínas Crk. A pesar de que se conocían los procesos generales implicados en la activación de C3G, los mecanismos detallados no se comprendían bien.

En esta Tesis, hemos combinado enfoques bioquímicos, biofísicos y de biología celular para caracterizar las bases estructurales y mecanísticas de los mecanismos de autoinhibición y activación fisiológica de la activación de C3G dentro de la vía C3G-Rap1. Hemos demostrado que CrkL se une de manera diferencial a los PRMs P1, P2, P3 y P4 en C3G completo. En condiciones de reposo, los sitios P1 y P2 están constitutiva y totalmente accesibles. La exposición del sitio P3 está relacionada con el estado de activación de C3G. El sitio P4 está parcialmente accesible independientemente del estado de activación de C3G. Los sitios P1 y P2 no participan en la activación directa de C3G. En cambio, estos sitios median la unión constitutiva con las proteínas Crk en las líneas celulares HEK293T y Jurkat, y son necesarios para el reclutamiento de C3G a la membrana plasmática tras la estimulación de TCR en

células Jurkat. La unión de CrkL a P3 y P4 es necesaria y suficiente para la activación directa de C3G por CrkL causando la liberación de la interacción autoinhibitoria. La unión de CrkL al P3 es el principal evento de activación. Colectivamente, P1 y P2 son sitios de reclutamiento y P3 y P4 son sitios de activación. La fosforilación de C3G en tirosinas por sí sola causa una activación mínima, pero es esencial para la estimulación mediada por Crk. La fosforilación por Src sensibiliza a C3G, que se activa a concentraciones más bajas de CrkL y a niveles de actividad más altos que cuando C3G no está fosforilado. Hemos mapeado los principales sitios de fosforilación de Src en C3G, resultando en Y329, Y504, Y579 e Y590. A pesar de que CrkL interactúa con P3 y P4 principalmente mediante el dominio SH3N, los dominios SH2 y SH3C de CrkL también contribuyen a la activación de C3G. En particular, el dominio SH2 juega un papel clave en la activación de C3G fosforilado a través de una interacción no canónica de baja afinidad con fosfo-Y590, la cual aparentemente estabiliza la interacción de CrkL con los sitios de activación. También hemos demostrado que CrkL induce una activación más fuerte de C3G que CrkII, y estas diferencias parecen deberse a la diferente organización entre-dominios de las proteínas Crk. También hemos producido un modelo estructural validado experimentalmente de C3G autoinhibido, que reveló contactos adicionales del AIR con el dominio REM. En base a los resultados, proponemos un ciclo multi-etapa de la activación y desactivación fisiológica de C3G. Finalmente, hemos aplicado los datos estructurales y funcionales para identificar tres mutaciones somáticas de cambio de sentido en cáncer que causan la activación constitutiva de C3G: Y570N e Y590N están descritas en pacientes con linfoma no Hodgkin, y Y579C se había detectado en un carcinoma de tiroides. Estas mutaciones amplían el repertorio de alteraciones adquiridas que desregulan C3G-Rap1 en tumores. En resumen, los resultados de esta Tesis contribuyen a comprender los mecanismos de señalización de C3G-Rap1 en tejidos sanos y en enfermedades.

## LIST OF ABBREVIATIONS

| | |
|---|---|
| $A_{280}$ | Absorbance at 280 nm |
| $AC_{50}$ | Half maximal activation concentration |
| AIR | Autoinhibitory region |
| AIR-CBR | AIR Cdc25H-binding region |
| AIR-IT | AIR Cdc25H-inhibitory tail |
| AMP | Adenosine monophosphate |
| ANIS | Anisotropy |
| ANOVA | Analysis of variance |
| ATP | Adenosine triphosphate |
| BSA | Bovine serum albumin |
| BSD | Blasticidin S deaminase |
| CDS | Coding sequence |
| CML | Chronic myeloid leukemia |
| cNBD | Cyclic nucleotide-binding domain |
| co-IP | Co-immunoprecipitation |
| COSMIC | Catalogue of somatic mutations in cancer |
| DEP | Disheveled, Egl-10, Pleckstrin |
| DH | Dbl homology |
| DTT | Dithiothreitol |
| $E_{280}$ | Mass extinction coefficients |
| $EC_{50}$ | Half maximal effective concentration |
| EDTA | Ethylenediaminetetraacetic acid |
| $\varepsilon_{280}$ | Molar extinction coefficients |
| FAK | Focal adhesion kinases |
| FL | Full-length |
| GAP | GTPase activating proteins |
| GDI | Guanine nucleotide dissociation inhibitors |
| GDP | guanosine diphosphate |
| GEF | Guanine exchange factor |
| GFP | Green fluorescent protein |
| GST | Glutathione-S-transferase |
| GTP | Guanosine triphosphate |
| $IC_{50}$ | Half maximal inhibitor concentration |

| | |
|---|---|
| IF | Immunofluorescence |
| IMAC | Immobilized-metal affinity chromatography |
| IP | Immunoprecipitation |
| IPH | *In vitro* phosphorylation assay |
| IPTG | Isopropyl β-D-1-thiogalactopyranoside |
| ITC | Isothermal titration calorimetry |
| $k_a$ | Equilibrium association constant |
| $k_d$ | Equilibrium dissociation constant |
| $k_{free}$ | Minimum value of $k_{obs}$ |
| $k_{max}$ | Maximum value of $k_{obs}$ |
| $k_{min}$ | Minimum value of $k_{obs}$ |
| $k_{obs}$ | Apparent nucleotide exchange rate |
| MCS | Multi-cloning site |
| mEGFP | Mono Enhanced GFP |
| MLB | Magnesium lysis buffer |
| MOC | Manders´ overlap coefficient |
| MS | Mass spectrometry |
| MW | Molecular weight |
| N | Stoichiometry constant |
| NCBI | National center for biotechnology information |
| NEK | Nucleotide exchange kinetics |
| NTD | N-terminal domain |
| $OD_{600}$ | Optical density 600 nm |
| PBS | Phosphate-buffered saline |
| PCR | Polymerase chain reaction |
| PD | Pull-down |
| PDB | Protein data bank |
| PFA | Paraformaldehyde |
| PH | Pleckstrin homology |
| pI | Isoelectric point |
| PIP2 | Phosphatidylinositol biphosphate |
| PIP3 | Phosphatidylinositol triphosphate |
| pLDDT | Per-residue local distance difference test |
| PMSF | Phenylmethylsulphonyl fluoride |
| PRMs/PRM | Proline-rich motif |
| PSMs | Peptide spectrum matches |

| | |
|---|---|
| pY | Phosphorylated tyrosine |
| REM | Ras exchange motif |
| rmsd | Root-mean-square difference |
| SEC | Size exclusion chromatography |
| SFKs | Src family kinases |
| SH2 | Src homology-2 domain |
| SH3 | Src homology-3 domain |
| SH3b | SH3-binding |
| SH3C | C-terminal SH3 domain |
| SH3N | N-terminal SH3 domain |
| SH4 | Src homology-4 domain |
| SNVs | Single nucleotide variants |
| Src-KD | Src kinase domain |
| TCR | T cell receptor |
| TEV | Tobacco etch virus |
| WB | Western blot |
| WT | Wild type |
| ΔH | Binding enthalpy |

## TABLE OF CONTENTS

# INTRODUCTION

## 1. Small G-proteins

Cells constitute the fundamental building blocks of life, serving as structural and functional units of living organisms. They are incredibly diverse in form and function, yet they share common characteristics, being one the ability to communicate and response to their environment through intricate signaling processes. Inside cells, a complex network of molecular interactions regulates essential life functions such as growth, division, differentiation, and survival. These interactions are leaded by signaling pathways that involve proteins, lipids, and nucleic acids.

The Ras superfamily of small GTPases, consists of a group of around 167 human proteins that regulate highly diverse signaling pathways. These proteins are part of a high conserved group of proteins in the evolution from yeast to human, indicating their involvement in critical function for the cells (Lowy and Willumsen, 1993, Reuther and Der, 2000, Santos and Nebreda, 1989, Macara et al., 1996, Rojas and Santos, 2002). Together these proteins control a vast variety of cellular processes by the control of cell signaling pathways (Bos, 1997). Disruptions in their regulation have been associated with the development of numerous diseases, including cancer (Santos et al., 1982, Alan and Lundquist, 2013, Rauen, 2013). The Ras superfamily is divided in five major groups, according to sequence homologies, the Ras, Arf, Rho, Rab and Ran families (Rojas et al., 2012).

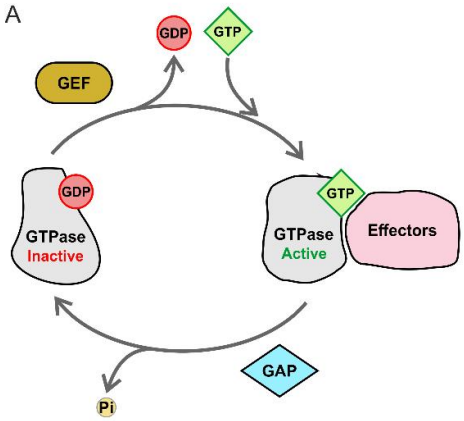
### The Ras family

In mammals, Ras protein family includes about 35 members that are grouped in four main branches or subfamilies Ras, Rap, R-Ras and Ral proteins (Colicelli, 2004). Ras proteins have a central function in cellular processes such as migration, adhesion, cell-polarity, proliferation and transcription (DeGeer and Lamarche-Vane, 2013, Heasman and Ridley, 2008, Karnoub and Weinberg, 2008, Kashatus, 2013, Loirand et al., 2013, Diekmann et al., 2011, Thumkeo et al., 2013).

### The GTPase molecular switch

All members of the Ras superfamily share a common 20-kDa G-domain, which exhibits a high degree (30 to 50%) of sequence identity, a universal structure and switch mechanism. This G-domain is responsible for nucleotide binding and hydrolysis (Bourne et al., 1991).

Small GTPases are characterized by their ability to bind to nucleotides GDP and GTP and to change their shape depending on the bound nucleotide, which enables them to function as molecular switches (Figure I1). These GTPases bind GDP and GTP with very high affinity, typically in the picomolar or nanomolar range (Klebe et al., 1995, Lenzen et al., 1998). The G-domain fold consists of five α-helices and a six-stranded β-sheet, which is typical for nucleotide-binding proteins. The G-domain of members of the Ras family contains four highly conserved elements that are involved in the binding of the nucleotide. Two of these elements interact directly with the nucleotide: the N/TKXDD motif recognizes the base of the nucleotide, while the P-loop ($GX_4GKS/T$ motif) recognizes the β-γ-phosphates of the nucleotide and a $Mg^{2+}$ ion (Saraste et al., 1990). In addition, two flexible and more variable motifs known as switch I and switch II also play a significant role in nucleotide interaction. These motifs adopt different conformations depending on whether GDP or GTP is bound, and they are involved in the hydrolysis of GTP (Figure IB and C) (Milburn et al., 1990, Vetter and Wittinghofer, 2001). The GTP-bound form of the GTPase is considered the active state because it is able to bind to most effectors, facilitating downstream signaling pathways. Conversely, the GDP-bound form is considered the inactive state.



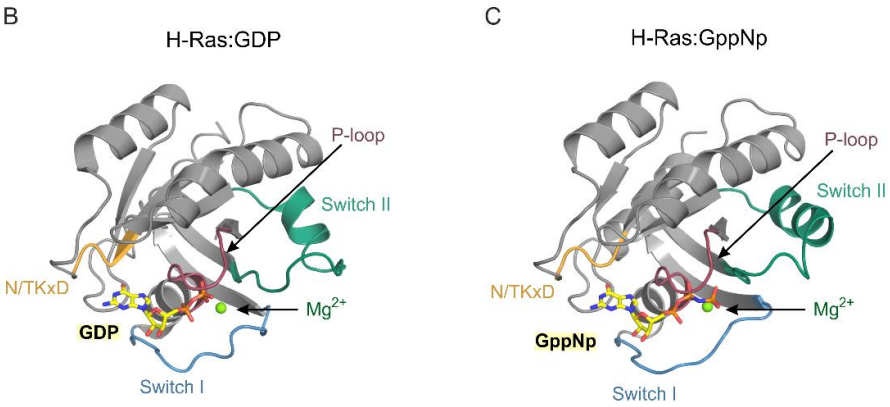



**Figure I1. The small GTPase molecular switch cycle. (A)** Small GTPase cycle between inactive (GDP-bound) and active (GTP-bound) states. GEFs catalyze the exchange of GDP for GTP. GAPs accelerate the intrinsic GTPase activity leading to the hydrolysis of GTP to GDP and subsequence release of inorganic phosphate (Pi), inducing the inactive state of the GTPase. **(B)** Structure of H-Ras in the inactive form bound to GDP (PDB: 4Q21) (Milburn et al., 1990). **(C)** Structure of H-Ras in the active form bound to the non-hydrolysable GTP analogue GppNp (PDB: 5P21) (Pai et al., 1990). The main structural elements involved in the nucleotide binding and the conformational changes linked to the nature of the bound nucleotide are indicated in B and C.

Ras GTPases have a low rate of guanine nucleotide release, whether it is GDP or GTP, primarily because the nucleotides are tightly bound. As with the nucleotide release, the GTP hydrolysis activity of Ras GTPases is intrinsically inefficient and slow; hence, these proteins are often referred to as GTP-binding proteins. The slow nucleotide exchange and hydrolysis are not suitable for the rapid biological signal transduction processes that require quick activation or inactivation mechanisms. To accelerate the necessary conformation changes during signaling, GTPases require the involvement of two other type of proteins. Guanine nucleotide exchange factors (GEFs) facilitate the dissociation of GDP from the GTPases, significantly accelerating this process by several orders of magnitude. Conversely, GTPase activating proteins (GAPs) stimulate the intrinsic hydrolase activity of the GTPases (Vigil et al., 2010, Cherfils and Chardin, 1999).

### Regulation of Ras proteins activity

Control of the activation and inactivation of Ras GTPases in time and space is crucial for regulating their activity in various cellular processes. The two main mechanisms of regulation are: (i) cellular compartmentalization of the GTPases, and (ii) regulation of the switch cycle by the activity of the GEFs, GAPs and GDIs (guanine nucleotide dissociation inhibitors) proteins, which are also subject to regulation, adding complexity to the control of Ras GTPase activity.

The cellular localization of the Ras GTPases is mediated by posttranslational modification. Many Ras superfamily proteins contain a C-terminal CAAX motif (C denotes cysteine, A any aliphatic amino acid and X any amino acid). The cysteine residue serves as a substrate for prenylation (farnesyl-transferases covalently attach farnesyl-isoprenoid lipids; and geranygeranyl-transferases, geranylgeranyl-isoprenoid lipids), which mediates their subcellular localizations by the insertion into the lipid bilayers targeting the protein to cell membranes (Zhang and Casey, 1996, Okada et al.,

1996, Porfiri et al., 1994, Gotoh et al., 2001). Lipidated GTPases localization can be regulated by the action of guanine nucleotide dissociation inhibitors (GDIs), which bind to prenylated GTPases, masking and blocking the translocation to the membranes of GTPases of the Rab and Rho families (Seabra and Wasmeier, 2004).

The mechanisms of nucleotide exchange by GEFs involve multiple steps (Figure I2). First, the interaction between GEFs and the GTPases results in the formation of a low-affinity trimeric complex with the GDP-bound GTPase. This complex undergoes rapid nucleotide release and transition into a high-affinity and stable nucleotide-free GTPase-GEF complex. Despite the high structural diversity between different types of GEFs, they share equivalent mechanisms to distort the nucleotide-binding site of the GTPases (Bos et al., 2007, Cherfils and Zeghouf, 2013). Most GEFs make contacts with the switch II of the GTPase and frequently GEFs distort the switch I too, some GEFs insert acidic or hydrophobic residues near the phosphate-binding or $Mg^{2+}$ sites to cause electrostatic repulsions (Thomas et al., 2007). The interaction of the GEF with the GTPase and the release of the nucleotide are reversible and can be inverted by the re-binding of the nucleotide. The most abundant nucleotide in the cytosol is GTP. Binding of GTP to the GTPase-GEF complex displaces the GEF effectively, switching the GTPase to its active GTP-bound state. This process reaches equilibrium between the GDP-bound and GTP-bound forms of the GTPase (Lenzen et al., 1998, Klebe et al., 1995).

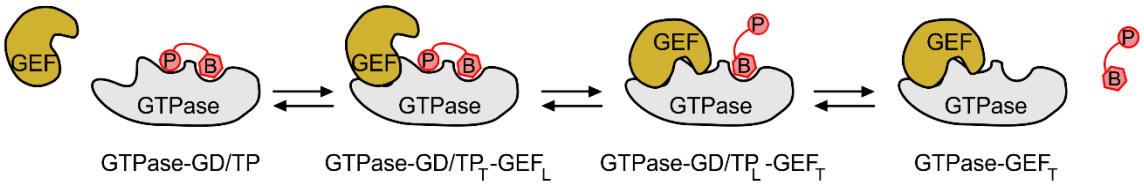



**Figure I2. Mechanism of GEF-induced GTPase nucleotide exchange.** Nucleotide exchange reaction occurs in three reversible steps. In red is the nucleotide that includes its base (B) and its phosphate moieties (P). Ternary complexes include the existence of tight (T) and loose (L) interactions of the GTPase with the GEF and the nucleotide. Adapted from (Bos et al., 2007).

Similarly to GEFs, there is a large variety of GAPs. Yet, all GAPs share common general mechanisms of action to activate the hydrolase activity of the GTPases. GAPs induce the activation of a nucleophilic water molecule through the arrangement of a pair of residues (reviewed by (Cherfils and Zeghouf, 2013)). The most common mechanism of GAPs involves an arginine residue of the GAP, known as the arginine

finger, which positions a glutamine in switch II to activate a water molecule for nucleophilic attack. Some GAPs differ from this mechanism, for example Rap1GAP that uses an asparagine residue that occupies the position of the glutamine in other GTPases (Scrima et al., 2008).

GDIs are negative regulators of small GTPases. They were originally identified as proteins that bind to the GDP-bound state, preventing the dissociation of the nucleotide and keeping the GTPase in the inactive form (Fukumoto et al., 1990). GDIs also interact with the prenylated C-terminus of some GTPases (Hoffman et al., 2000). Therefore, GDIs extract the GTPases from the membrane and sequester them in an inactive from in the cytoplasm.

**Structures and mechanisms of Ras family GEFs**

GEFs are divided in five groups attending their family specificity, their structural features, and regulatory mechanism: GEFs that activate Ras family GTPases, GEFs of the Rho family, GEFs of the Arf family, GEFs of the Ran family, and GEFs of the Rab family. The GEFs for the Ras family GTPases, Ras, Ral, and Rap, include 27 human proteins that contain a domain homologous to the Cdc25 gene of *Saccharomyces cerevisiae* (Cdc25H). These Cdc25H domains are evolutionarily conserved from yeast to human (Lowy and Willumsen, 1993); they are responsible for the binding to the GTPase and the nucleotide exchange activity. GEFs are multidomain proteins and the Cdc25H domains appear most frequently associated with a N-terminal Ras Exchanger Motif or REM domain (Boguski and McCormick, 1993).

The structures of Cdc25H domains of several mammalian Ras GEFs have been characterized in detail, such as SOS (Boriack-Sjodin et al., 1998), Epac (Rehmann et al., 2006), CalDAG-GEF II (RasGRP1) (Iwig et al., 2013), RasGRP2, RasGRP4, (Vercoulen et al., 2017), RasGRF1 (Freedman et al., 2006), PDZ-GEF, PLC$_\varepsilon$, RalGPS (Peng et al., 2011), and Rlf (Popovic et al., 2016). The Cdc25H domains consist of 11 α-helices (αA-αK) that adopt a pack bowl arrangement. Helices αH-αI extend out of the core of the domain in an antiparallel arrangement named helical hairpin (Boriack-Sjodin et al., 1998) (Figure I3).

The structure of the SOS/H-Ras complex was the first example of a Cdc25H domain bound to its cognate GTPase (Boriack-Sjodin et al., 1998). The structures of Epac2 bound to Rap1b (Rehmann et al., 2006, Rehmann et al., 2008), CalDAG-GEF I (RasGRP2) bound to Rap1, RasGRP4 bound to H-Ras (Vercoulen et al., 2017), and

Rgl II (Rlf) bound to Ral (Popovic et al., 2016) revealed common structural features and exchange mechanisms in Cdc25HD GEFs.

The GTPases bind at the concave region of the Cdc25H, wrapping around the helical hairpin. The orientation of the helical hairpin with respect to the Cdc25H core plays an important role in the regulation of the GEF activity of some Cdc25H domains (Freedman et al., 2006). Binding to the Cdc25H induces structural changes in the two switch regions of the GTPase. The helical hairpin inserts into nucleotide binding site of the GTPase and displaces the switch I by steric clash, disrupting the nucleotide-binding site and allowing the open nucleotide-binding site and dissociation. The chemical environment around the switch II and the binding site for the phosphates are also altered; the GEF displaces the associated $Mg^{2+}$ ion by repulsion forces and occludes sterically the magnesium-binding site, unfavoring binding (Cherfils and Chardin, 1999, Vetter and Wittinghofer, 2001). The switch II is necessary for tight binding to the Cdc25H but is not the critical driving force for GDP displacement (Hall et al., 2001). SOS does not block the binding sites for the base and ribose of GTP or GDP. Therefore, the nucleotide-free Cdc25H/Ras complex allows not only the release of the nucleotide, but also the re-binding; this favors the entry of the more abundant GTP and the reversal of the process (Boriack-Sjodin et al., 1998, Klebe et al., 1995, Lenzen et al., 1998).

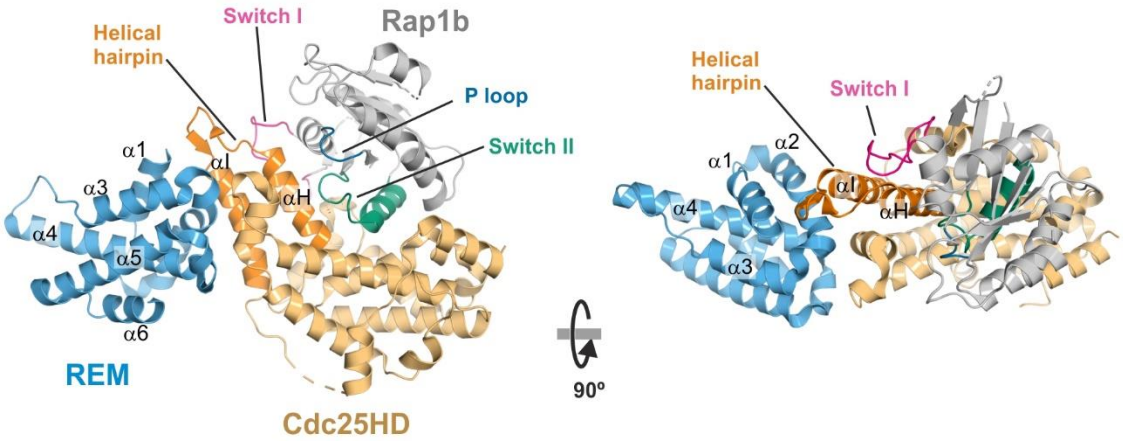



**Figure I3. Structure of Epac2 bound to Rap1b.** Representation of the crystal structure of Epac2 bound to Rap1 (PDB: 3CF6) (Rehmann et al., 2008). For clarity, the cNBD-B and RA domains of Epac2, which are present in the structure, are not shown for clarity. The main structural elements involved in the interaction are indicated.

The Cdc25H domain makes extensive interactions with the N-terminal REM domain. The REM and Cdc25H domains are frequently known as the GEF catalytic region. The

REM domain is composed of six α-helices (α1-α6). Helices α1 and α2 are in close contact with the catalytic domain, packing against the helical hairpin. However, REM domain does not contact directly the GTPase bound to the Cdc25H domain. The REM domain of SOS controls the GEF activity of the companion Cdc25H domain by modulating the orientation of the helical hairpin, and a similar mechanism has been proposed for other Cdc25H GEFs (Margarit et al., 2003, Boykevisch et al., 2006, Freedman et al., 2009).

In addition to these catalytic domains, GEFs often contain a wide variety of additional domains with diverse functions (Figure I4). These domains play crucial roles in regulating the activity and interactions of the GEFs, through intramolecular or intermolecular mechanisms (Cherfils and Zeghouf, 2013).

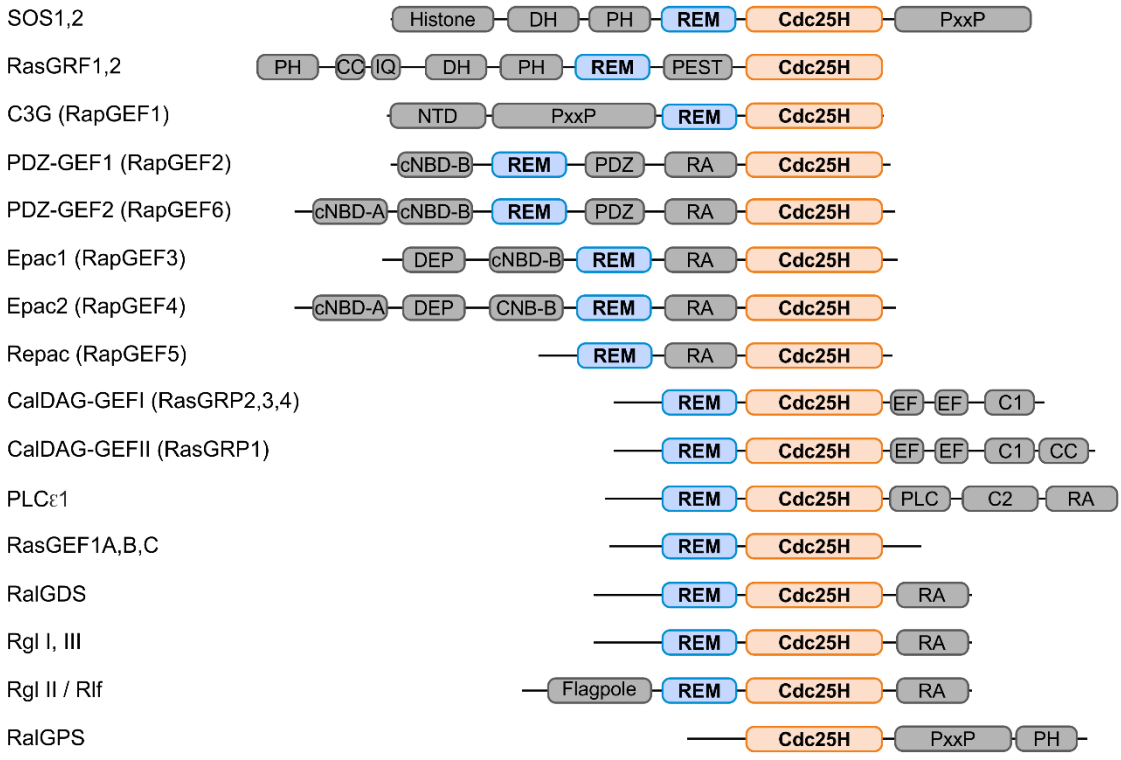



**Figure I4. Domain structure of GEFs of the Ras family of GTPases.** Commonly, the REM precedes the catalytic Cdc25HD. The rest of the domains constitute the non-catalytic regions and are mostly involved in regulatory mechanisms: C1, homologous to the DAG binding domain of protein kinase C1 (PKC1); C2, second conserved domain of PKC2, frequently involved in $Ca^{2+}$-dependent membrane binding; CC, coiled-coil; cNBD-A, cyclic nucleotide binding domain A; cNBD-B, cyclic nucleotide binding domain B; DEP, disheveled, Egl-10 and Pleckstrin domain; DH, Dbl homology domain; EF, Calcium binding domain; Flagpole, β-sheet enriched domain with binding site for SH3 domains; Histone, histone-fold domain; IQ, Calmodulin binding domain; NTD, N-terminal domain; PDZ, domain shared by PSD95, DlgA and ZO-1 proteins; PEST, domain enriched in P, E, S and T amino acids; PH, pleckstrin homology domain; PLC, phospholipase C domain; PxxP, proline-rich motif; RA, Ras-association domain. Adapted from (Vigil et al., 2010).

### Mechanisms of regulation of GEFs

The nucleotide exchange activity of many GEFs is tightly controlled. Regulation of GEFs is typically mediated by two mechanisms (i) control of the subcellular localization, and (ii) intramolecular interactions between the catalytic and non-catalytic regions that maintain a low GEF activity by blocking the binding of the GTPase (Pufall and Graves, 2002). Despite this common regulation, the mechanical characteristics, conformations, and interactions of the different types of domains involved in the regulation of GEFs can vary among different GEFs.

Two regulatory mechanisms are described below to understand the diversity of the mechanisms: SOS as the canonical Cdc25HD-containing GEF, Epac as Cdc25HD-containing GEF of Rap proteins.

### SOS

SOS proteins are GEFs for the Ras family proteins (Wang et al., 1995). Humans contain two isoforms, SOS-1, and SOS-2, with identical modular structure domain organization. The REM and Cdc25H domains are localized in the central region of the protein. In the N-terminal region, SOS contains three structural domains: a Histone-fold, a Dbl homology (DH), and a pleckstrin homology (PH) domain; in addition, there is a helical linker. The Histone and PH domains bind to phosphoinositol phosphates (PIP2 and PIP3) (Zheng et al., 1997, Rameh et al., 1997, Koshiba et al., 1997). The DH domain is homologous to the GEF catalytic domain of GEFs of Rho GTPases and mediates the activation of Rac (Cerione and Zheng, 1996, Rossman and Sondek, 2005, Nimnual et al., 1998). Downstream the Cdc25HD, there is a Proline-rich region that mediates binding to SH3-containing adaptor proteins such as Grb2, Nck and Crk (Hall et al., 2002, Lowenstein et al., 1992) (Figure I5).

SOS binds two molecules of Ras simultaneously: a first molecule of Ras binds to the catalytic site in the Cdc25HD while a second GTP-bound Ras molecule binds to a distal allosteric site. The interaction of Ras with the catalytic binding site is similar to other Cdc25HD-containing GEFs, this interaction causes the release of the nucleotide from the GTPase. The second allosteric Ras-GTP-binding site spans the REM and Cdc25H domains. Binding of the Ras-GTP to this site triggers a conformational change in the REM domain that is transmitted to the helical hairpin of the CDC25HD increasing the nucleotide exchange activity of SOS. The specificity of the two binding sites for the inactive and active forms of Ras creates a functional link between the GEF and its

effector. This leads to the activation of SOS through a positive feedback loop (Sondermann et al., 2004, Margarit et al., 2003).

The GEF activity of SOS is repressed by autoinhibitory mechanisms. The tandem of DH-PH domains interacts with the REM domain and regulates the accessibility to the distal site for the Ras allosteric molecule (Sondermann et al., 2004). The Histone domain binds to the flexible linker connecting the DH-PH tandem to the REM domain and stabilizes the autoinhibitory conformation (Corbalan-Garcia et al., 1998, Gureasko et al., 2010, Chen et al., 1997). In addition, the interaction of the Histone domain and the linker prevents the interaction of the PH domain with membrane lipids and the Rac binding to the DH, favoring the conformational autoinhibitory state. In the autoinhibitory conformation, the catalytic Ras-binding site is exposed, but the arrangement of the REM and helical hairpin is kept in a catalytic incompetent form (Nakamura et al., 2016).

It has proposed that the Proline-rich domain regulates the overall catalytic activity of SOS in cells by three mechanisms. First, the Proline-rich region participates in the recruitment of SOS to the plasma membrane in a SH3-dependent manner by the Grb2 adaptor protein (Lowenstein et al., 1992, Simon and Schreiber, 1995, Egan et al., 1993). Secondly, the Proline-rich motif does not interact with the N-terminal region but limits the SH3-independent recruitment of SOS to the membrane through the inhibition of the binding of Ras-GTP to the allosteric site (Lee et al., 2017). Thirdly, the Proline-rich region inhibits the Rac GEF activity of the DH domain and it is released upon binding of the protein E3b1/Abi-1 (Scita et al., 1999, Innocenti et al., 2003, Innocenti et al., 2002).

A multi-step model of activation of SOS has been proposed, which would be initiated with the production of Ras-GTP by other GEFs, such as RasGRP1 (Roose et al., 2007). Subsequently, the binding pf Ras-GTP to the allosteric site produces inter-domain conformational changes that cause (i) the release of Proline-rich region and Histone domain autoinhibitory interactions, (ii) the molecular configuration of the DH and PH domains favoring the interaction of Histone and PH domains with phospholipids, (iii) the rotation of the REM domain and the helical hairpin at the catalytic site adopting an active conformation, and (iv) the recruitment to the plasma membrane and the reinforce of an optimal orientation with respect to the membrane. In addition, in the final active orientation of SOS in the membrane, the DH domain is accessible to act on Rac-GDP (Lenzen et al., 1998, Gureasko et al., 2010, Freedman et al., 2006).

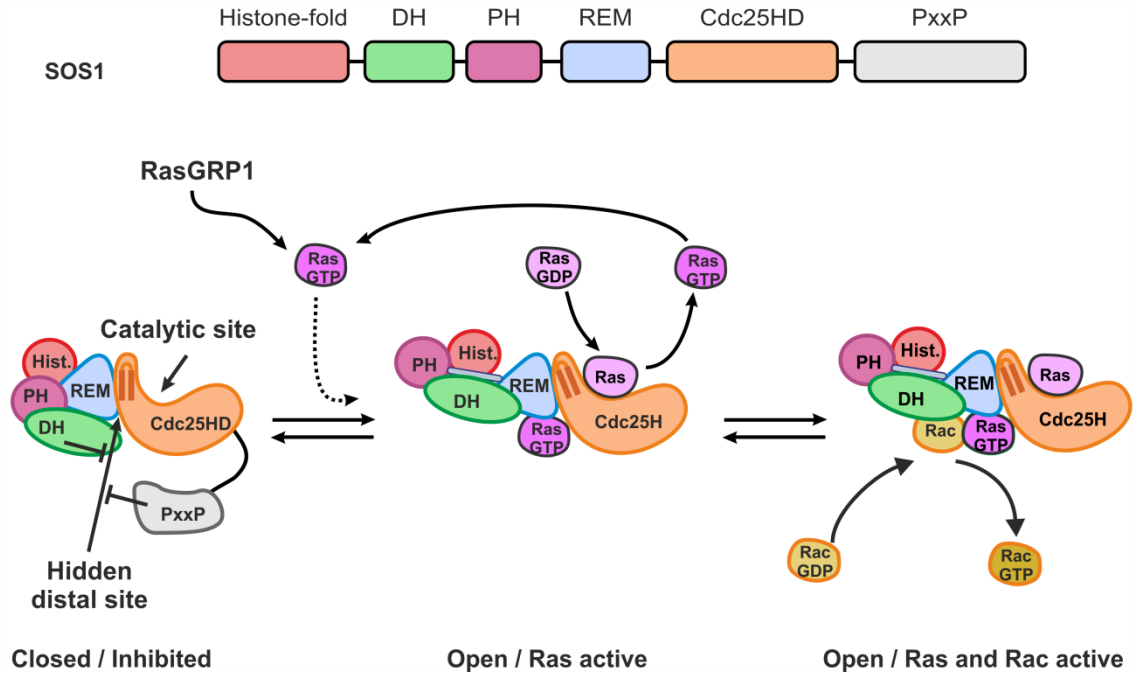



**Figure I5. Regulatory mechanism of SOS1: autoinhibition and activation.** Domain organization of SOS1 (top). In the autoinhibited conformation, both the distal binding site for Ras-GTP within the REM-Cdc25H domains and the Rac binding site in the DH domain are hidden. Upon stimulation, the distal binding site for Ras-GTP becomes accessible, adopting an activated conformation. In this state, both the Cdc25H and DH domains become catalytically competent.

## Epac

Epac proteins are GEFs for the Rap family proteins, Rap1 and Rap2, and their activation is dependent on the presence of cAMP, a second messenger molecule (de Rooij et al., 1998, Kawasaki et al., 1998). There are two Epac proteins, Epac1 and Epac2, which share a similar structural organization that includes a DEP (Disheveled, Egl-10 and Pleckstrin) domain, a cNBD-B (cyclic nucleotide binding domain) domain, REM domain, a RA (Ras association) domain and a Cdc25H domain (Figure I6). Epac2 has an additional regulatory N-terminal cNBD-A domain. Both cNBD domains bind cAMP, but cNBD-B exhibits a higher affinity for the nucleotide compared to cNBD-A. The DEP domain interacts with lipids and mediates membrane localization, while the RA domain acts as an effector for Ras-GTP and Rap-GTP.

The structure of Epac2 reveals a closely packed arrangement in which both cNBD-A and cNBD-B domains physically block access to the Rap1-binding site (Rehmann et al., 2006). The DEP domain secures this inhibitory conformation by an interaction with the Cdc25H, in a region out of the GTPase binding site, and the cNBD-B domains (Rehmann et al., 2003). In this close conformation, the helical hairpin of the Cdc25H

domain adopts an arrangement similar to active Cdc25H domains. This suggests that the autoinhibitory mechanism of Epac2 is primarily mediated by a steric barrier (Rehmann et al., 2006).

Activation of Epac proteins occurs when cAMP binds to the cNBD-B domain (de Rooij et al., 1998). This binding induces a conformational change in the protein, causing the stabilization of the cAMP-bound-cNBD-B and REM domains (Rehmann et al., 2006). As a result of this stabilization, the regulatory domains swing away from the GEF active site, exposing it to GTPases. cAMP also causes the translocation of Epac proteins from the cytoplasm to the membrane. The cAMP-induced conformational change makes de DEP domain accessible DEP for binding to acidic lipids in the membrane (Ponsioen et al., 2009, Consonni et al., 2012). Additionally, the RA domain also participates in the recruitment to the membrane through the binding to Ras-GTP (Li et al., 2006).

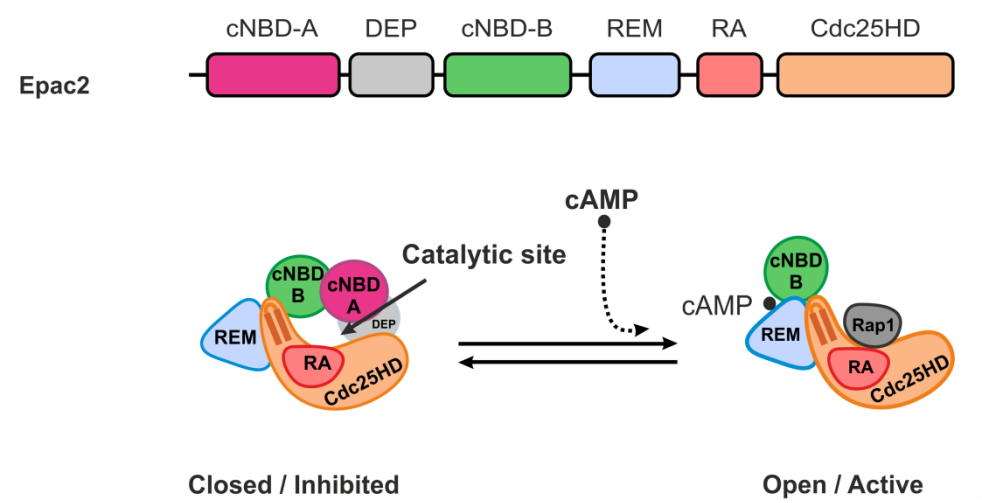



**Figure I6. Mechanism of autoinhibition and activation of Epac2.** Domain organization of Epac2 (top). In the autoinhibited structure, the GTPase binding site of the Cdc25HD is blocked by a tightly packed arrangement of the cNBD-B, DEP, and cNBD-A. Binding of cAMP to the cNBD-B induces a conformational change causing the exposure of the GTPase binding site in the Cdc25HD.

### 2. C3G

C3G (Crk Src homology-3 domain-binding guanine nucleotide exchange factor), also known as RapGEF1 or GRF2, was the first ubiquitously protein recognized as a Crk and Grb2 SH3-binding partner (Tanaka et al., 1994). C3G is a guanine exchange factor (GEF) for the GTPases belonging to the Ras family, including Rap1, R-RAS, TC-21 (Gotoh et al., 1995, Gotoh et al., 1997, Ohba et al., 2000), and Rap2, the latter with lower efficiency than other GEFs such as Epac (Popovic et al., 2013, van den Berghe et al., 1997). In addition, C3G acts as GEF of the Rho family GTPase member TC10

(Chiang et al., 2006, Chiang et al., 2001). Nevertheless, the specificity of the C3G GEF activity *in vitro* only has been demonstrated on Rap1 (van den Berghe et al., 1997, Popovic et al., 2013).

**Structure and binding partners**

The sequence of the main variant of human C3G, the isoform a (NCBI: NM_005312) encodes a modular protein of 1077 residues (~120 kDa) with three structurally and functionally distinct regions (Figure I7A).

The N-terminal domain (4-245 residues; hereafter NTD) of C3G does not have sequence similarity with other proteins. A previous work from our laboratory has shown that the initial segment of the NTD (residues 4-64) is a region rich in basic residues (pI = 9.9) that interacts with anionic lipids (Gómez-Hernández, 2014), a characteristic also described in other GEFs (Cherfils and Zeghouf, 2013, Karandur et al., 2017). The rest of the NTD (residues 90-245) is rich in α-helical content as assayed by circular dichroism (Gómez-Hernández, 2014). This segment of the NTD is predicted to form a four α-helix bundle. The segment 144-230 of the NTD interacts with E-cadherin (Hogan et al., 2004, Asuri et al., 2008). The NTD also binds to the REM domain in an intramolecular (Carabias et al., 2020, Gómez-Hernández, 2014).

The central region of C3G or SH3-binding domain (residues 246-670, hereafter SH3b) is involved in protein-protein interactions. The region is predicted to be mostly flexible and disordered. The SH3b contains five proline-rich motifs (hereafter PRMs) named P0 (265-276), P1 (282-291), P2 (452-462), P3 (537-549) and P4 (607-614). The PRMs contain the consensus sequence PPXLPXK and interact with SH3-domain containing proteins. The adaptor proteins Crk were the first identified binding partners of the SH3b (Matsuda et al., 1992); the N-terminal SH3 domain (SH3N) of Crk binds to the sites P1, P2, P3 and P4 with high affinity (Knudsen et al., 1994). Other SH3 domain-containing proteins such as Grb2 (Tanaka et al., 1994), tyrosine kinases Hck (Shivakrupa et al., 2003) and c-Abl (Radha et al., 2007), and the fusion oncoprotein Bcr-Abl (Gutierrez-Berzal et al., 2006), bind to the PRMs P1-P4. On the other hand, the SH3 domain of p130Cas binds directly to the P0 motif of C3G (Kirsch et al., 1998). The central region of C3G also facilitates the interaction with proteins that lack of SH3 domains, such as β-catenin (Dayma et al., 2012), actin (Martin-Encabo et al., 2007) and the specific T-cell phosphatases TC-PTP (Mitra et al., 2011). The C-terminal segment of the SH3b region (residues 537-646) binds to the Cdc25HD (Carabias et al., 2020). Additionally,

the SH3b region contains the residue Y504 whose phosphorylation has identified as an important event for the activation of C3G (Ichiba et al., 1997).

The C-terminal region of C3G contains a REM and a Cdc25H domain (Cdc25HD) that form the catalytic region. The REM domain binds to the NTD (Gómez-Hernández, 2014). As in other GEFs of GTPases of the Ras family, the Cdc25HD is responsible for the binding to the GTPase and the exchange of the bound nucleotide.

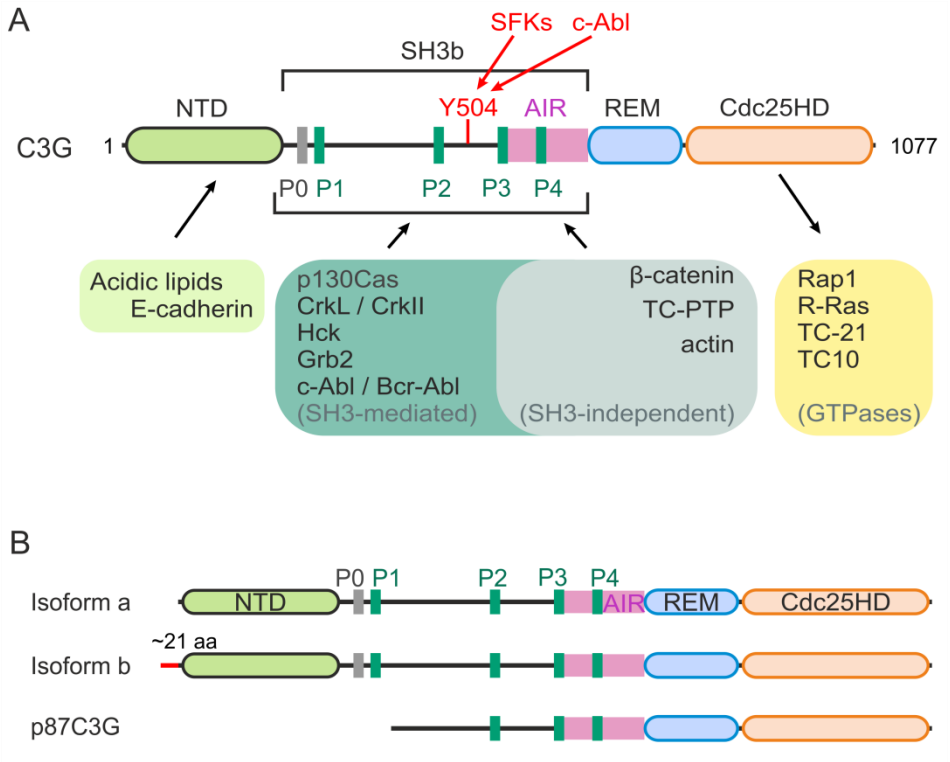



**Figure I7. Modular structure of C3G and its isoforms. (A)** Primary structure of C3G with its different domains and their binding partners. **(B)** Schematic representation of the human C3G isoforms.

In human, three isoforms of C3G are ubiquitously expressed with some tissue-specific variations in their expression levels (Figure I7B). Two isoforms result from alternative splicing of a single RNA transcript originating from a single locus, which correspond to *RAPGEF1* gene (Radha et al., 2011). Isoforms C3G-a and C3G-b differ in the N-terminal region, C3G-b isoform substitutes three residues of C3G-a by a 21-residue segment. The third isoform of 87 kDa (p87C3G) is encoded by a 4.4 kb transcript and expresses a truncated protein that lacks approximately the first 305 amino acids, corresponding to the NTD, including the E-cadherin binding site, and the two first proline-rich motifs. This isoform is overexpressed in chronic myeloid leukemia (CML) cell lines and in primary bone marrow cells from patients (Gutierrez-Berzal et al., 2006).

p87C3G is phosphorylated by Bcr-Abl kinase, which suggests that this isoform of C3G might play a role in the development of CML (Maia et al., 2009, Maia et al., 2013).

### Regulation of C3G

The molecular mechanism of regulation of C3G has revealed recently from our group. The activity of C3G is self-regulated by two intramolecular interactions (Carabias et al., 2020) (Figure I8). First, binding of the NTD to the REM domain contributes positively to the GEF activity of C3G; this suggests that similarly as observed in SOS, the REM domain of C3G may transmit conformational cues to the Cdc25HD, which modulate the GEF activity. Secondly, the interactions between the C-terminal fragment of the SH3b and the Cdc25HD is the main autoinhibitory mechanism in C3G. We refer to this region of the SH3b (residues 545-646) as the autoinhibitory region (hereafter AIR). The first part of the AIR (residues 545-569) is essential and sufficient for binding directly to the Cdc25HD; therefore, this part is named the AIR Cdc25H-binding region, or AIR-CBR. The AIR-CBR does not block completely the catalytic activity of C3G. Complete inhibition requires the second segment of the AIR (residues 570-646); therefore, this is named the Cdc25H-inhibitory tail or AIR-IT. Yet, in the absence of the AIR-CBR, the AIR-IT bind to nor block the activity of the Cdc25HD. Residues M551, Y554, and M555 of the AIR-CBR are essential for binding to the Cdc25HD; mutations of these residues disrupt the intramolecular mechanism of autoinhibition of C3G. Two of these residues Y554 and M555 are targeted by somatic missense mutations detected in non-Hodgkin´s lymphomas: Y554H had been described in a follicular lymphoma (Green et al., 2013) and M555R in a diffuse large B cell lymphoma (Morin et al., 2011). These mutations interfere with the autoinhibitory binding AIR-Cdc25HD and cause a constitutive activation of C3G-Rap1 *in vitro* and in cells (Carabias et al., 2020).

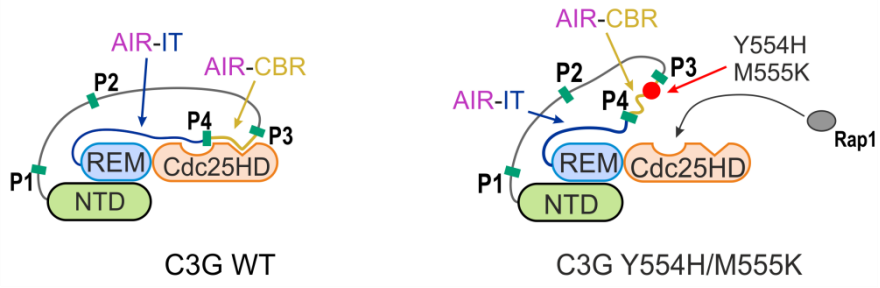



**Figure I8. Molecular mechanisms of regulation of C3G.** Model of the self-regulation of C3G by two intramolecular interactions. The interaction between the NTD and REM domains favors the GEF activity. The interaction between the AIR and Cdc25HD blocks the GEF activity. These two interactions occur in

unstimulated wild type C3G (left). Two cancer-related missense mutations disrupt the intramolecular autoinhibitory interaction inducing a constitutively active state (right).

## Functions of C3G

C3G is a ubiquitously expressed Cdc25H GEF that participates in diverse cellular processes including proliferation, migration, adhesion, differentiation, apoptosis, cell survival, suppression of malignant transformation, filopodia formation and actin remodeling (reviewed in (Radha et al., 2011)).

### Regulation of cell-cell and cell-environment contacts

C3G promotes cell adhesion via Rap1 (Arai et al., 1999), a key physiological activator of integrins through the induction of conformational changes that increase the affinity of integrins for their ligands ("inside-out" signaling) (Lagarrigue et al., 2016).

C3G is essential for mouse embryonic development, evidenced by the fact that constitutive C3G knockout (C3G-KO) mice die before embryonic day 7.5 (Ohba et al., 2001), suggesting that the functions of C3G cannot be compensated by other Rap1-GEFs during the early stages of development. The lethal phenotype of C3G-KO mice is attributed to defects in cell-cell adhesion due to alterations in signaling by Rap GTPases, essential regulators of adhesion processes (Bos, 2018). This is supported by the observation that embryonic fibroblasts (MEFs) derived form C3G-KO mice exhibit impaired cell adhesion and accelerated migration. In addition, C3G is essential for cortical neuron migration and blood vessel maturation (Voss et al., 2006, Voss et al., 2003), and for migration of liver progenitor cells (oval cells) during liver repair in chronic liver diseases (Palao et al., 2022).

C3G participates in the formation of cell-cell contacts; specifically, it contributes to the early stages of adherens junction formation through the activation of Rap1. This activation is mediated by the interaction of C3G with E-cadherin molecules, which are crucial for the assembly of adherens junctions in epithelial cells (Kooistra et al., 2007). C3G competes with β-catenin for binding to E-cadherin (Hogan et al., 2004) and strength the functions of E-cadherin, a process known as "outside-in" cell signaling. Additionally, C3G binds to β-catenin through the central SH3b region, in a SH3-independent manner, and promotes the degradation of β-catenin. Conversely, activation of β-catenin inhibits the expression of C3G (Dayma et al., 2012) .

C3G participates in the remodeling of the actin cytoskeleton. C3G and phospho-C3G colocalize with components of the cytoskeleton, such as F-actin, and alter actin dynamics (Luber et al., 2000). Phosphorylation of C3G enhances its association with cytoskeletal structures and, in association with c-Abl, it participates in filopodia formation in epithelial cell lines (Mitra and Radha, 2010, Radha et al., 2007). Through regulation of actin dynamics, C3G also regulates vesicle traffic in muscle and adipocyte tissues (Chiang et al., 2001, Chiang et al., 2006).

Role of C3G in haematopoietic cells

Rap1 plays a crucial role in T-cell function by participating in the activation of integrins in response to the TCR and chemokine receptor stimulation (Shimonaka et al., 2003, Katagiri et al., 2000). The activation of intracellular signaling pathways in response to TCR ligation regulates the cell surface integrins by clustering and activating them (known as “inside-out” signaling). Activation of the integrin LFA1 (αLβ2, CD11a/CD18) in response to TCR stimulation relies on C3G-Rap1 signaling. This process involves the WAVE2 complex, which promotes the recruitment of the CrkL-C3G complex to the membrane. The kinase Abl is involved in the process, facilitating the phosphorylation of C3G and, consequently, the activation of Rap1 (Nolz et al., 2008).

In B-cells, C3G contributes to B-cell antigen receptor-mediated signaling transduction and plays a regulatory role in the proliferation and differentiation of the B cells (Smit et al., 1996).

In mature megakaryocytes and platelets, Rap1b is the most abundant Ras family GTPase and play a crucial role in platelet functions (Stork and Dillon, 2005). Rap1b regulates platelet adhesion and aggregation through the activation of integrin αIIbβ3 (Li et al., 2010, Shattil et al., 1998) . In addition, Rap1 participate in the secretion of α- and δ-granules, thus amplifying platelet activation (Zhang et al., 2011). Platelets express several Rap GEFs (Schultess et al., 2005), being the functions of CalDAG-GEFI (RasGRP2) and C3G the best characterized (Stefanini and Bergmeier, 2016, Gutierrez-Herrero et al., 2012). While CalDAG-GEFI, which has $Ca^{2+}$ and a diacyl glycerol (DAG) binding sites, induces platelet activation through $Ca^{2+}$-dependent pathways (Stefanini and Bergmeier, 2016), C3G mediates platelet activation and aggregation in a $Ca^{2+}$-independent, PKC-dependent manner (Gutierrez-Herrero et al., 2020, Gutierrez-Herrero et al., 2012). In addition, C3G regulates megakaryocytic differentiation and proplatelet formation (Ortiz-Rivero et al., 2018), and platelet-induced

angiogenesis and metastasis by regulating the secretion of α-granule content (Martin-Granado et al., 2017, Hernandez-Cano et al., 2022).

C3G in cancer

Analysis of the role of Rap1 in tumorigenesis in various model systems revealed antagonistic functions (Looi et al., 2020). Rap1 was discovered as a suppressor of Ras transformation acting as antagonist of Ras activity (Kitayama et al., 1989, Bos, 1998). Yet, Rap1A and Rap1B promote invasion and migration of tumor cells in prostate (Bailey et al., 2009), ovarian (Lin et al., 2015), esophageal cancers (Zhang et al., 2012) and neck squamous cell carcinoma (Goto et al., 2010). Upregulation of Rap1 is associated with aggressive malignancy phenotypes (Alemayehu et al., 2013, Yang et al., 2015).

C3G can function as a tumor promoter or suppressor depending on the cellular context, tumor type and stage. Initially, C3G was found to act as tumor suppressor by down-regulation of the malignant transformation induced by overexpression of *ras*, *sis* and *v-raf* oncogenes (Guerrero et al., 1998). This role does not require the Cdc25H domain (Guerrero et al., 2004). Reduced expression levels of C3G have also been associated with invasion in hepatocellular carcinoma (Sequera et al., 2020), glioblastoma (Manzano et al., 2021a, Manzano et al., 2021b) and breast cancer (Dayma and Radha, 2011).

C3G also plays a role as tumor promoter. C3G is overexpressed in non-small cell lung cancer (Okino et al., 2006). C3G participates in the malignant cell transformation mediated v-Crk oncogene (Tanaka et al., 1997) and the RET-PTC oncogenic protein fusion in thyroid tumors (De Falco et al., 2007). In cells from chronic myeloid leukemia (CML) patients, the alternative spliced isoform p87C3G is expressed together with the Bcr-Abl fusion protein (Philadelphia chromosomal translocation). C3G binds to Bcr-Abl and is phosphorylated by its kinase activity, suggesting its involvement in the development of the CML disease (Gutierrez-Berzal et al., 2006, Maia et al., 2013). In addition, downregulation of C3G expression has been detected in chronic lymphocytic leukemia (Fernandez et al., 2008). Also, platelet C3G promotes homing and adhesion of metastatic melanoma cells by promoting tumor cell induced platelet aggregation process (Hernandez-Cano et al., 2022).

## 3. Crk adaptor proteins

Crk was initially identified as a novel transforming gene, v-Crk or Gag-Crk, within the genome of an avian sarcoma retrovirus called CT10 (chicken tumor Virus 10) (Mayer et al., 1988b, Mayer et al., 1988a). This gene lacking intrinsic tyrosine activity, contributed strongly to increase the cellular tyrosine phosphorylation level. The human Crk (CT10 regulator of a tyrosine kinase) family is formed by three members: CrkI and CrkII, which were identified as homologs of v-Crk from a unique *locus* (*CRK*); and CrkL, a paralog from an independent *locus* (*CRKL*) (Matsuda et al., 1992, ten Hoeve et al., 1993).

Crk and CrkL are co-expressed in somatic cells and it has been suggested that can compensate to each other in cellular signaling (Park and Curran, 2008, Huang et al., 2004, Matsuki et al., 2008). In contrast to this compensatory mechanism in somatic cells, CrkL and CrkII have independent and non-overlapping roles during embryonic development (Isakov, 2008, Guris et al., 2006, Guris et al., 2001).

CrkII and CrkL share a similar structural composition with three globular domains (Figure I9A) (Reichman et al., 1992). They have an N-terminal Src homology-2 (SH2) domain that binds short tyrosine phosphorylated sequences in the context of pYXXP (where pY denotes phosphorylated tyrosine, X any amino acid and P is proline) (Birge et al., 1993). Next, they have a first or N-terminal SH3 domain (hereafter SH3N) that binds to proteins with proline-rich motifs (PRMs) in the context of PXXPXR/K (Knudsen et al., 1994). Finally an atypical C-terminal SH3 domain (SH3C) does not bind to PRMs (Muralidharan et al., 2006).

CrkII and CrkI are the result of alternative splicing of the single *CRK* gene RNA (Matsuda et al., 1992). CrkI ends prematurely and lacks 50 amino acids of the SH3N-SH3C linker and the C-terminal SH3 domain. CrkII has a 17-residue proline-rich flexible loop inserted internally within the SH2 domain that is not present in the SH2 domains of other proteins. On the other hand, CrkL is encoded by a different gene *CRKL* sharing 56% of overall protein sequence identity with CrkII. Notably, CrkL and CrkII have higher structural similarity, (~72% sequence identity) in their globular domains (Kobashigawa and Inagaki, 2012).

Crk proteins bridge phospho-tyrosine-containing proteins to various binding partners involved in intracellular signaling pathways (Figure I9B). Specifically, Crk adaptor proteins exhibit a high-affinity binding to C3G via their SH3 domains and function as essential adaptor protein in many processes mediated by C3G (Radha et al., 2011).

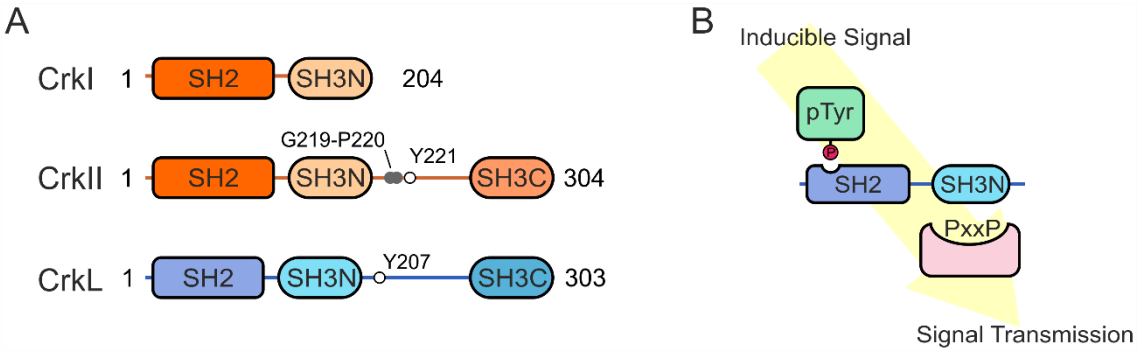



**Figure I9. Structure and role of the Crk proteins in signaling pathways. (A)** Domain organization of Crk proteins. The phosphorylalable-tyrosine residues in CrkII and CrkL and the isomerization site G219-P220 in CrkII are highlighted. **(B)** Canonical mechanism of Crk proteins in the integration of signals through their modular SH2 and SH3N domains. Adapted from (Birge et al., 2009).

Structurally, the individual SH2 and SH3 domains exhibit similarity across all three Crk members, as indicated by sequence conservation. However, distinctions are evident in their three-dimensional interdomain arrangement (Kobashigawa and Inagaki, 2012). The SH2 and SH3N of CrkL adopt a compact structure. The SH3C does not interact with the other domains of CrkL. This SH2-SH3N disposition partially occludes the pY-binding site of the SH2 domain specifically the SH3N masks the binding sub-sites for residues downstream the pY such as the P in position +3. In contrast, the binding site for PRMs within the SH3N domain remains accessible (Jankowski et al., 2012). Conversely, in CrkII, the SH2-SH3N domains maintain a compact structure, involving the C-terminal region in the conformational stabilization. This domain architecture results in the blocking of the SH3N binding site while the SH2 is accessible (Kobashigawa and Inagaki, 2012). On the other hand, CrkI do not form a compact SH2-SH3N structure, and both SH2 and SH3 binding sites are accessible (Figure I10).

Both CrkII and CrkL undergo regulation through phosphorylation of tyrosine residues in the linker that connects the SH3N and SH3C domains; Y221 in CrkII and Y207 in CrkL, which are phosphorylated by the Abl kinase (de Jong et al., 1997, Kobashigawa et al., 2007, Peterson and Long, 2008). Phosphorylation of CrkII-Y221 and CrkL-Y207 causes an intramolecular interaction between the linker and the SH2 domains, leading to the adoption of a closed pY-SH2 conformation. This closed conformation results in the inhibition of the adaptor function of both proteins. Yet, these interactions have distinct effects depending on whether it is CrkII or CrkL. Phosphorylation of Y221 in CrkII and binding to the SH2 results in blocking of the SH3N binding site by the SH3N-SH3C linker. In contrast, phosphorylation of Y207 in CrkL does not block the SH3N binding site, which remains accessible. In both cases, Crk phosphorylation by Abl

kinase compromises the ability of the SH2 domains to bind to pY motifs of other proteins (Jankowski et al., 2012).

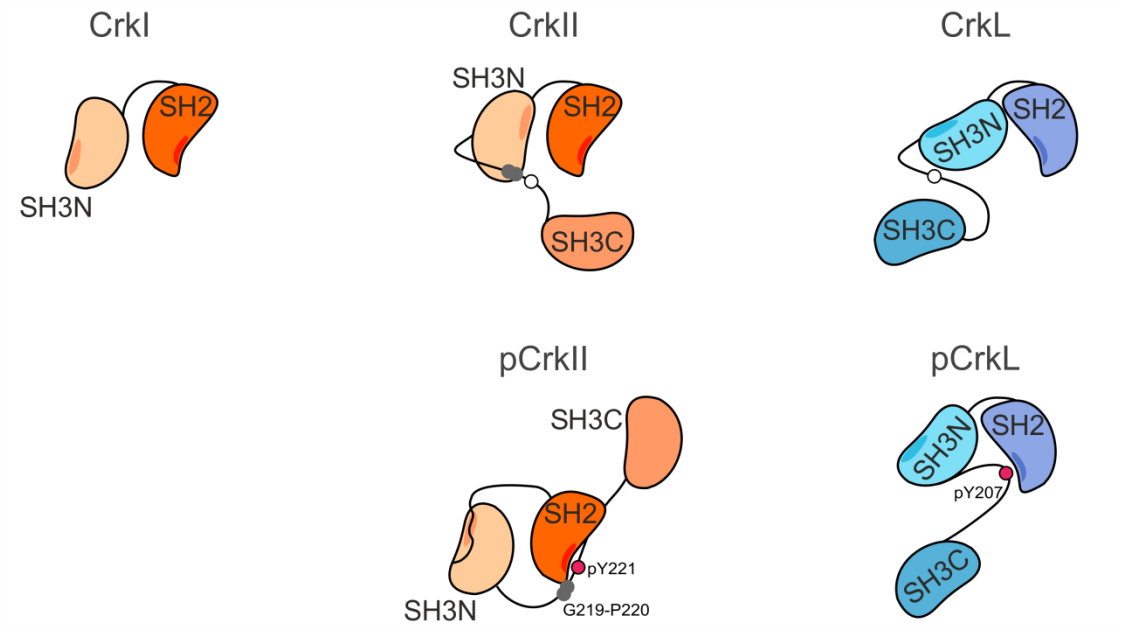



**Figure I10. Regulatory mechanism of Crk proteins.** Domain configurations of the CrkI, CrkII and CrkL, as well as the phosphorylated status of CrkII (pCrkII) and CrkL (pCrkL). The relative spatial orientations of the domains are inferred form the NMR structures of CrkI (PDB: 2EYY), CrkII (PDB: 2EYZ), CrkL (PDB: 2LQN), pCrkII (PDB: 2DVJ) and pCrkL (PDB: 2LQW) (Jankowski et al., 2012, Kobashigawa et al., 2008).

CrkII is also regulated by the *cis-trans* isomerization of a proline switch (G237-P238 in chicken CrkII, and G219-P220 in human) located adjacent to Y221, within the linker region connecting the SH3C and SH3N domains. This isomerization process regulates distinct conformations in the protein (Isakov, 2008, Sarkar et al., 2007). In the *cis*-conformation, an intramolecular inhibitory interaction is stabilized, blocking the binding site of the SH3N domain by SH3C. In contrast, the *trans* conformation adopts an open state that promotes an uninhibited arrangement where the SH3N domain is available for binding to PRMs. The equilibrium between the *cis*- and *trans*-isomers tends to favor the closed conformation and the transition between these states occurs slowly (Sarkar et al., 2011). Peptidyl-prolyl cis-trans isomerases (PPIases), such as the FK506 immunophilins and cyclophilin A (CypA), catalyze the reversibility of these conformational changes (Schmidpeter and Schmid, 2014). In human CrkII, where G219-P220 and Y221 are in close proximity, it is proposed that PPIases bound to CrkII could affect the regulation mediated by Abl-phosphorylation of Y221 (Saleh et al., 2016). CrkL does not share the same structural characteristic, as it lacks a proline near the regulatory Y207 residue and does not exhibit cis-trans isomerization.

## 4. Src kinases

c-Src (hereafter Src) is a non-receptor tyrosine kinase that was identified as the cellular homolog of the Rous Sarcoma Virus transforming gene v-Src (Hunter and Sefton, 1980, Hunter, 2015, Lipsick, 2021). Src is the prototypic member of the Src family of kinases (SFKs) composed of 11 structurally similar non-receptor protein tyrosine kinases: Src, Lyn, Fyn, Hck, Lck, Srm, Brk, Blk, Fgr, Yes and Yrk (Manning et al., 2002). SFKs share a common organization with distinctive functional domains. The N-terminal region contains a Src homology-4 (SH4) domain and a unique domain; the SH4 harbors sites for attachment of myristoyl groups responsible of the attachment to the membranes (Brown and Cooper, 1996). Downstream the unique domain there is a SH3 and a SH2 domains which allow inter- and intra-molecular interactions with PRMs (Feng et al., 1994, Musacchio et al., 1994, Yu et al., 1994) and pY containing sequences (Waksman et al., 1993), respectively. Next, Src contains the kinase catalytic domain with two distinctive lobes, N-lobe and C-lobe separated by a disordered loop. Finally, Src presents a C-terminal tail that contains regulatory sites, such as Y530.

The kinase domain of Src shows some preferences for certain features in the substrates that phosphorylates. Src has preference for substrates with isoleucine, leucine or valine preceding (position -1) the phosphorylated-Y, an acidic (glutamate or aspartate) in positions -2 and/or -3, glycine or glutamate in position +1, and phenylalanine, isoleucine, or leucine in position +3 (Deng et al., 2014, Shah et al., 2018, Songyang et al., 1995).

## 5. Activation of C3G

At the cellular level, the C3G-Rap1 pathway can be triggered by a wide range of stimuli that include ligation of B- and T-cell receptors (Medeiros et al., 2005, Nolz et al., 2008), hepatocyte growth factor (Sakkab et al., 2000), growth hormone (Ling et al., 2003) , platelet-derived growth factor (Yokote et al., 1998), epidermal growth factor (Okada and Pessin, 1997) and neuronal growth factor (York et al., 1998), as well as signals from insulin (Okada and Pessin, 1997), interferon-γ (Alsayed et al., 2000), erythropoietin (Nosaka et al., 1999) and Interleukin-3 (Arai et al., 2001). C3G-Rap1 is also activated in response to mechanical force (Tamada et al., 2004). Additionally, adhesion signals can also activate this pathway through the activation of integrins (Buensuceso and O'Toole, 2000).

C3G interacts with CrkII and CrkL in the cytosol of unstimulated cells (Okada and Pessin, 1997, Smit et al., 1996). Upon stimulation, Crk proteins recruit C3G at the plasma membrane through its adaptor function (Ichiba et al., 1997, Feller, 2001). Stretching of cells or B-cell receptor stimulation recruits the C3G-CrkL complex through the interaction with p130Cas (Ingham et al., 1996, Sawada et al., 2001, Tamada et al., 2004), while stimulation of integrins results in the recruitment of C3G-Crk complexes through both p130Cas and Cbl proteins (Vuori et al., 1996, Uemura and Griffin, 1999). At the membrane, C3G interacts with Rap1. In addition, Crk activates directly the GEF activity of C3G (Carabias et al., 2020, Popovich, 2013). C3G is phosphorylated by several tyrosine kinases such as Src family kinases (SFK): Src (Radha et al., 2004), Fyn (Nolz et al., 2008) and Hck (Shivakrupa et al., 2003), as well as c-Abl (Mitra and Radha, 2010) and its oncogenic version Bcr-cAbl (Gutierrez-Berzal et al., 2006). Phosphorylation of C3G *in vitro* by Src results in a moderate activation; Src-mediated phosphorylation and the binding of Crk proteins are independent and additive activation stimuli of the GEF activity (Carabias et al., 2020, Popovich, 2013) (Figure I11). Phosphorylation of Y504 residue is required for the CrkL-dependent activation of C3G in cells (Ichiba et al., 1999).

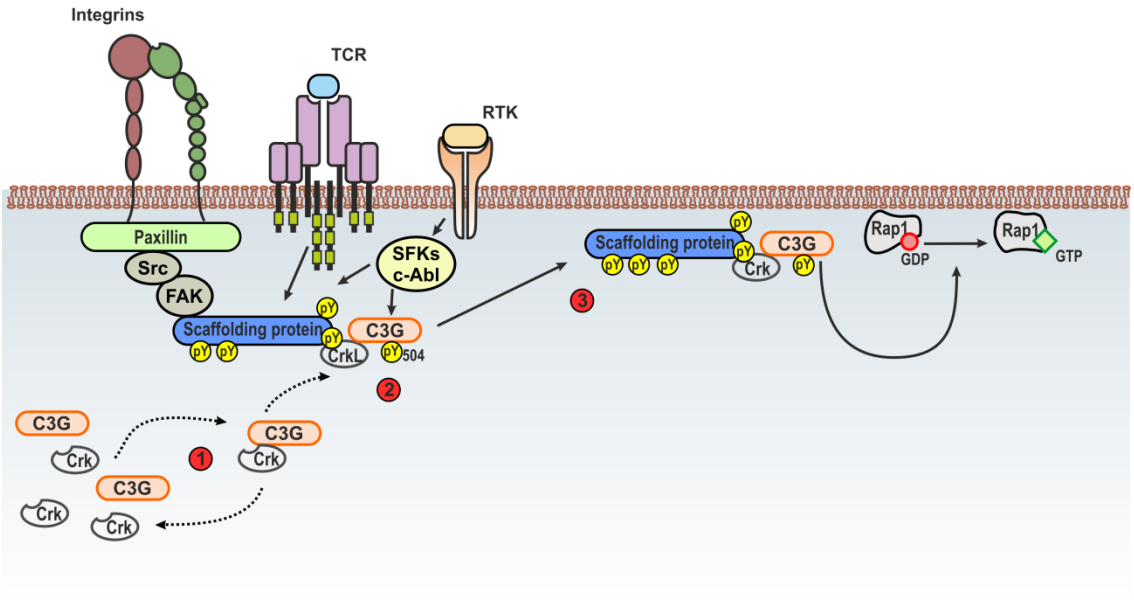



**Figure I11. The current general paradigm of C3G activation in cells.** (1) C3G is autoinhibited and forms complexes with Crk in the cytosol. (2) Upon stimulation of integrin activation, TCR ligation or receptor tyrosin kinase (RTK) activation, intramolecular kinases (e.g., SFKs, c-Abl or focal adhesion kinases (FAK)) creates binding sites for CrkL-SH2 domain on scaffolding proteins (p130Cas or Cbl) that recruits the Crk/C3G complex to the membrane. (3) Then, same SFKs or c-Abl proteins phosphorylate C3G in Y504 residue and activate it, promoting the Rap1 activation on the membrane.

# OBJECTIVES

The overall scope of this study is to gain a comprehensive understanding of the structural and mechanistic basis of the autoinhibition and physiological activation of GEF activity of C3G within the C3G-Rap1 pathway. To accomplish this, we have addressed the following specific objectives:

1. To analyze the interaction of Crk proteins with C3G at multiple proline-rich motifs and the site-specific impact on the activation of C3G.

2. To analyze the contribution of the structural domains of Crk proteins to the activation of C3G.

3. To identify the Src-phosphorylation sites in C3G and characterize their role in the activation of C3G by Crk proteins.

4. To characterize the structural bases of the autoinhibitory mechanism of C3G and use them to identify new cancer-related missense mutations that cause deregulation of C3G.

# METHODS

## 1. cDNA

A summary of the recombinant proteins employed in this study is presented in the Table M1 and cross-referenced with their corresponding cDNA and protein sequences in NCBI and Uniprot databases. The cDNA sequence of human C3G is referred to the main spliced variant, known as isoform a (1077 residues) (Tanaka et al., 1994), NCBI ID: NM_005312 with distinctions. Our sequence differs from the canonical isoform in that the first three residues (MDT) are replaced by MSGKIEKA (Guerrero et al., 1998), similar differences have also been described in other variants. We used the region 4-1077, which is common to several isoforms, as the full-length protein. Finally, our sequence includes a polymorphism in the residue S281 that is replaced by G.

**Table M1. List of the main human proteins used in this work.**

| Protein (Gene) | NCBI ID | Uniprot ID | Source |
|---|---|---|---|
| C3G (*RAPGEF1*) | NM_005312.4 | Q13905-1 | Our laboratory |
| CrkL (*CRKL*) | NM_005207.3 | P46109 | Our laboratory |
| CrkII (*CRK*) | NM_016823.3 | P43108 | Our laboratory |
| Rap1b (*RAP1B*) | NM_015646.5 | P61224 | Our laboratory |
| c-Src (*SRC*) | NM_005417.3 | P12931 | Dr. Robert Lefkowitz, Addgene #42202 |

## C3G constructs and mutants

The constructs used for the expression in bacteria of C3G and its fragments and mutants are summarized in Table M2. The construct of C3G full-length in the vector pETEV15b-His-Halo-TEV has been described (Carabias, 2019); this vector is a derivative of pETEV15b (Alonso-Garcia et al., 2009) that codes for an N-terminal octa-His-tag followed by the Halo protein and a sequence recognized by tobacco etch virus (TEV) protease. The cDNA of C3G in pETEV15b-His-Halo-TEV contains a silent point mutation that destroys an internal NdeI site around codon 340. In this work, we introduced two additional silent point mutations in C3G to generate unique internal sites for SpeI and KpnI around codons 310 and 637, respectively. A construct of the cDNA coding for the C3G amino acid region 310-637 was created in the pBluescript SK (-) vector; the region 310-637 was amplified by polymerase chain reaction (PCR) using primers that added SpeI and KpnI sites at the 5` and 3' ends (Table M3), the PCR product was digested and ligated into pBluescript SK (-) previously digested with the same enzymes. This construct was used as an intermediate vector to introduce mutations in the SH3b region of C3G that were later subcloned into the pETEV15b-His-Halo-TEV construct.

**Table M2. C3G constructs for bacterial expression**.

| Name | Limits | Description | Vector | Cloning sites | Purpose |
|---|---|---|---|---|---|
| His-Halo-C3G (WT and mutants) | 4-1077 | Full-length | pETEV15b-His-Halo-TEV | NdeI/BamHI | PD/ITC/NEK/ SEC/MS |
| GST-C3G | 4-1077 | Full-length | pGEX-4T3-NNB-TEV | NcoI/BamHI | PD/IPH |
| GST-NTD | 4-246 | NTD | pGEX-4T3-NNB-TEV | NcoI/BamHI | PD/IPH |
| GST-SH3b | 274-646 | SH3b | pGEX-4T3-NNB-TEV | NcoI/BamHI | PD/IPH |
| GST-274-500 | 274-500 | P1-P2 | pGEX-4T3-NNB-TEV | NcoI/BamHI | PD/IPH |
| GST-501-646 | 501-646 | P3-AIR-P4 | pGEX-4T3-NNB-TEV | NcoI/BamHI | PD/IPH |
| GST-537-569 | 537-569 | AIR | pGEX-4T3-NNB-TEV | NcoI/BamHI | NEK |
| GST-537-588 | 537-588 | AIR | pGEX-4T3-NNB-TEV | NcoI/BamHI | NEK |
| GST-537-597 | 537-597 | AIR | pGEX-4T3-NNB-TEV | NcoI/BamHI | NEK |
| GST-537-646 | 537-646 | AIR | pGEX-4T3-NNB-TEV | NcoI/BamHI | NEK |
| GST-REM-Cdc25HD | 670-1077 | REM-Cdc25HD | pGEX-4T3-NNB-TEV | NcoI/BamHI | PD/IPH |

PD, pull-down; ITC, isothermal titration calorimetry; NEK, nucleotide exchange kinetics; SEC, size exclusion chromatography; MS, mass-spectrometry; IPH, *in vitro* phosphorylation assay.

**Table M3. Primers used to introduce SpeI and KpnI restriction sites in the cDNA of C3G.**

| Name | Sequence (5' to 3') |
|---|---|
| C3Gh-310-Spe-For | CATGAGCCGAGCCACTAGTGGCTCCAGTTTGCCTG<br>M S R A T S G S S L P<br>310 |
| C3Gh-310-Spe-Rev | CAAACTGGAGCCACTAGTGGCTCGGCTCATGGGGG |
| C3Gh-637-KpnI-For | CTCAGCGGTCAGCGGGGTACCTGGGAAGGACAGCAGAG<br>S A V S G V P G K D S R<br>637 |
| C3Gh-637-KpnI-Rev | CTGTCCTTCCCAGGTACCCCGCTGACCGCTGAGGGATC |

The nucleotides changed in respect to the original sequence are shown in red.
Created SpeI and KpnI restriction sites are double-underlined.
The coded amino acid sequences are shown under the forward primers.

The construct coding C3G full-length in the bacterial expression vector pGEX-4T3-NNB-TEV has been described (Carabias, 2019); this vector is a derivative of pGEX-4T3 that codes for an N-terminal glutathione-S-transferase (GST) followed by a TEV site and a multi-cloning region that has NcoI, NdeI and BamHI sites as the pETEV15b vector. Constructs of the regions of C3G 274-646 (SH3b), 274-500, 501-646, 537-569, 537-588 and 537-646 in the vector pGEX-4T3-NNB-TEV, created in our group, have been described (Carabias et al., 2020, Carabias, 2019). GST-fusion constructs 4-246 (NTD) and 670-1077 (REM-Cdc25HD) in the same vector were also available in our group. The construct C3G-537-597 in the pGEX-4T3-NNB-TEV vector was created in an analogous manner. Briefly, the cDNA of interest was amplified by PCR using forward and reverse primers that added NcoI and BamHI sites respectively, the PCR product was digested and cloned using those sites.

The C3G full-length construct in the vector pEF1-mEGFP for transient expression in mammalian cells has been described (Carabias et al., 2020, Carabias, 2019). The pEF1-mEGFP vector is derivate of pEF1/V5-His (Invitrogen), in which the cDNA of monomeric enhanced green fluorescent protein (mEGFP) was inserted downstream of the multi-cloning site; therefore, C3G constructs in this vector code for fusion proteins with C-terminal mEGFP. The cDNA cloned in this vector is under the control of the elongation factor 1 promoter (EF1), allowing high levels of expression of proteins in mammalian cells. The cDNA of full-length C3G is cloned between BamHI and NotI sites of pEF1-mEGFP and lacks a stop codon.

Plasmids for the expression of C3G in mammalian cell are summarized in Table M4. The construct of C3G full-length wild type in the vector pLenti-C-mEGFP-IRES-BSD for lentiviral transduction has been described (Carabias et al., 2020, Carabias, 2019). This vector is a modified version of the commercial pLenti-C-HA-IRES-BSD lentiviral gene expression vector (Origene), which codes for a C-terminal mEGFP tag. The vector contains a chloramphenicol resistance gene for bacterial selection. It also carries the gene of the blasticidin S deaminase (BSD) within the lentiviral packaging cassette, which allows for blasticidin selection in the transfected and the transduced mammalian cells. C3G cDNA is cloned between BamHI and XhoI restriction sites following a strategy similar to that used for the cloning into pEF1-mEGFP vector.

**Table M4. C3G constructs for mammalian expression.**

| Name | Limits | Description | Vector | Sites | Purpose |
|---|---|---|---|---|---|
| C3G-mEGFP | 4-1077 | Full-length | pEF1-mEGFP | BamHI/NotI | IP/Rap1 act |
| C3G-mEGFP | 4-1077 | Full-length | pLenti-mEGFP-IRES | BamHI/NotI | IP/Rap1 act/IF |

IP, immunoprecipitation; Rap1 act, Rap activation assay; IF, immunofluorescence.

C3G mutants

The point mutants (single or multiple substitutions) of C3G constructs used in this work are summarized in the Table M5. The primers used to create these mutations are presented in Appendix I, Table A1.

**Table M5. C3G mutants and their applications.**

| Construct (Vector) | Mutations[a] | Use |
|---|---|---|
| C3G full length (pETEV15-His-Halo-TEV) | PPAA, AAPP, PPAP, PPPA, PAAA, APAA, AAPA, AAPA-Y554H, AAAP, AAAP-Y554H | Role of PRMs in the binding of CrkL and the activation of C3G |
| | Y504F, Y554F, Y554H, Y561F, Y570F, Y579F, Y590F, Y561F-Y570F-Y579F-Y590F | Role of tyrosine residues in the activation of pC3G |
| | Y329F-Y504F-Y579F-Y590F, Y504F-Y579F-Y590F (Y329WT-FL), Y329F-Y579F-Y590F (Y504WT-FL), Y329F-Y504F-Y590F (Y579WT-FL), Y329F-Y504F-Y579F (Y590WT-FL), Y329F-Y504F-Y561F-Y570F-Y579F-Y590F | Analysis of tyrosine residues phosphorylated by Src |
| | Y570N, Y579C, Y590N | COSMIC mutants |
| | D767R, N1017R | Validation of AlphaFold2 models |
| C3G-501-646 (pGEX-4T3-NNB-TEV) | Y504F-Y554F-Y561F-Y570F-Y579F-Y590F, Y554F-Y561F-Y570F-Y579F-Y590F (Y504WT), Y504F-Y561F-Y570F-Y579F-Y590F (Y554WT), Y504F-Y554F-Y570F-Y579F-Y590F (Y561WT), Y504F-Y554F-Y561F-Y579F-Y590F (Y570WT), Y504F-Y554F-Y561F-Y570F-Y590F (Y579WT), Y504F-Y554F-Y561F-Y570F-Y579F (Y590WT) | Role of tyrosine residues in the activation of pC3G |
| C3G-mEGFP full-length (pEF1-mEGFP) | PPAA, AAPP, PAAA, APAA, AAPA, AAAP | Role of PRMs in the activation of C3G |
| | Y554H[b], Y570N, Y579C, Y590N | COSMIC mutants |
| C3G-mEGFP full-length (pLenti-mEGFP-IRES-BSD) | PPAA, AAPP, PAAA, APAA, AAPA, AAAP | Role of PRMs in the recruitment and activation of C3G |

[a] Different mutants are separated by commas, while multiple substitutions within a mutant are separated by dashes. Simplified names of some mutants with multiple substitutions are indicated in parenthesis.
[b] Described in (Carabias et al., 2020, Carabias, 2019).

Point substitutions were created using a PCR-based method based on the QuikChange site-directed mutagenesis protocol (Stratagene-Agilent), using primers carrying the desired changes and the Phusion DNA polymerase (Thermo Scientific) (Xia et al., 2015). For the generation of multiple mutations, two strategies were used: (i) multiple single site-directed mutagenesis were done in a sequential manner, targeting one site at a time, or (ii) fragments were amplified from different templates carrying the desired mutations by PCR, or the desired mutations were introduced during the PCR, and were combined by overlap extension method. The latter strategy facilitated the creation of multiple mutations. For example, that method allowed the simultaneous alteration of three PRMs through a single step. It also facilitated the modification of combinations of multiple tyrosine residues (three, four, five, or six) simultaneously, using templates harboring tyrosine mutations in the fragment 274-646 (Figure M1) and full-length C3G (Figure M2).

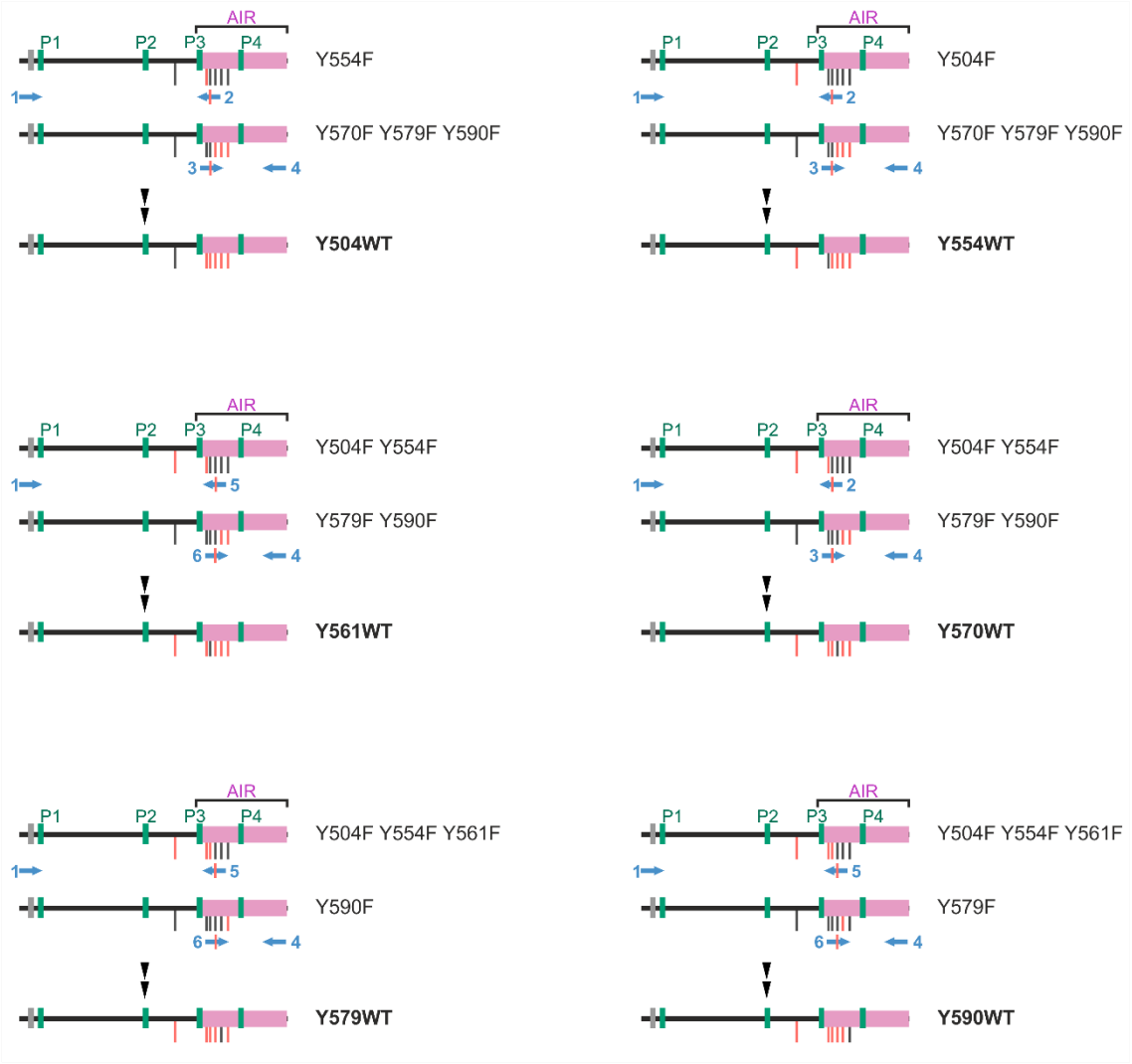



**Figure M1. Strategy to generate the single unmodified tyrosine mutants Y504WT, Y554WT, Y561WT, Y570WT, Y579WT and Y590WT by overlap extension PCR in the fragment of C3G-SH3b.** In each case, the cDNA templates and the primers used to introduce an additional mutation by overlap extension are indicated. For example, outer primers 1 and 4 were used to amplify the two overlapping fragments and, in a subsequent reaction the complete region SH3b (274-646). Primers 2 and 3 correspond to C3G-Y561F forward and reverse mutagenesis site, and primers 5 and 6 correspond to C3G-Y570F forward and reverse mutagenesis site.

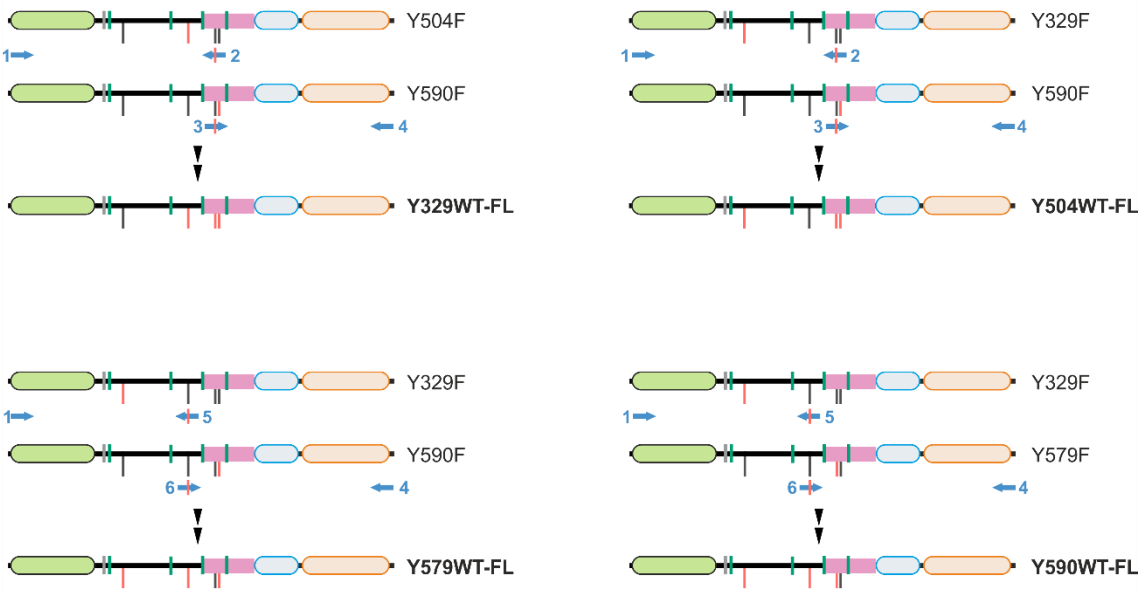



**Figure M2. Strategy to generate multiple tyrosine residue mutants in full-length C3G by overlap extension PCR.** In each case, the cDNA templates and the primers are shown. Primers 1 and 4 were used to amplify first the overlapping moieties and later on the full-length sequence of C3G. Primers 2 and 3 correspond to C3G-Y579F forward and reverse mutagenesis site, and primers 5 and 6 correspond to C3G-Y504F forward and reverse mutagenesis site.

In the case of the PRMs mutants five adjacent mutations were introduced simultaneously (i.e. using single pair of oligonucleotides); changes consisted in the substitution of five Pro, Leu and Lys residues by Ala, as they were essential for the binding of SH3 domains as described (Carabias, 2019). Mutants were named with a four-letter code, each letter corresponds to the P1 to P4 sites, to indicate whether the PRM had the wild type sequence (P) or carried the mutations to Ala (A).

The oligonucleotides used to mutate tyrosine residues of C3G are shown in Appendix I, Table A1. Initially, tyrosine mutants were introduced in the truncated construct C3G-SH3b (274-646). Subsequently, these mutations were transferred to the C3G full-length construct in the pETEV15b-His-Halo-TEV vector using the variant that contains restriction sites for SpeI and KpnI at codons 310 and 637, respectively (see above) (Figure M3). The region coding for the 310-637 segment was amplified by PCR using the C3Gh-310-SpeI-For and C3Gh-637-KpnI-Rev primers (Table M3, see above), and the desired mutant of the C3G-274-646 construct. The PCR products were digested with SpeI and KpnI and were ligated into the pETEV15b-His-Halo-TEV C3G construct with SpeI and KpnI sites previously digested with these enzymes. To facilitate the identification of clones containing the desired cassette, we used a pETEV15b-His-Halo-

TEV C3G Δ310-637 construct in which the region between the SpeI and KpnI sites was replaced by a short linker. The transfer strategy is described in the Figure M4.

Mutants carrying single tyrosine substitutions were named attending to the mutated residues; for example, C3G-Y504F. Also, multiple tyrosine substitutions were named using the mutated residues, e.g., C3G-Y561-570-579-590F. Yet, in specific C3G mutants with five mutated tyrosine residues were named by the wild type tyrosine that they contain, e.g., C3G Y554F-Y561F-Y570F-Y579F-Y590F was referred to as Y504WT.

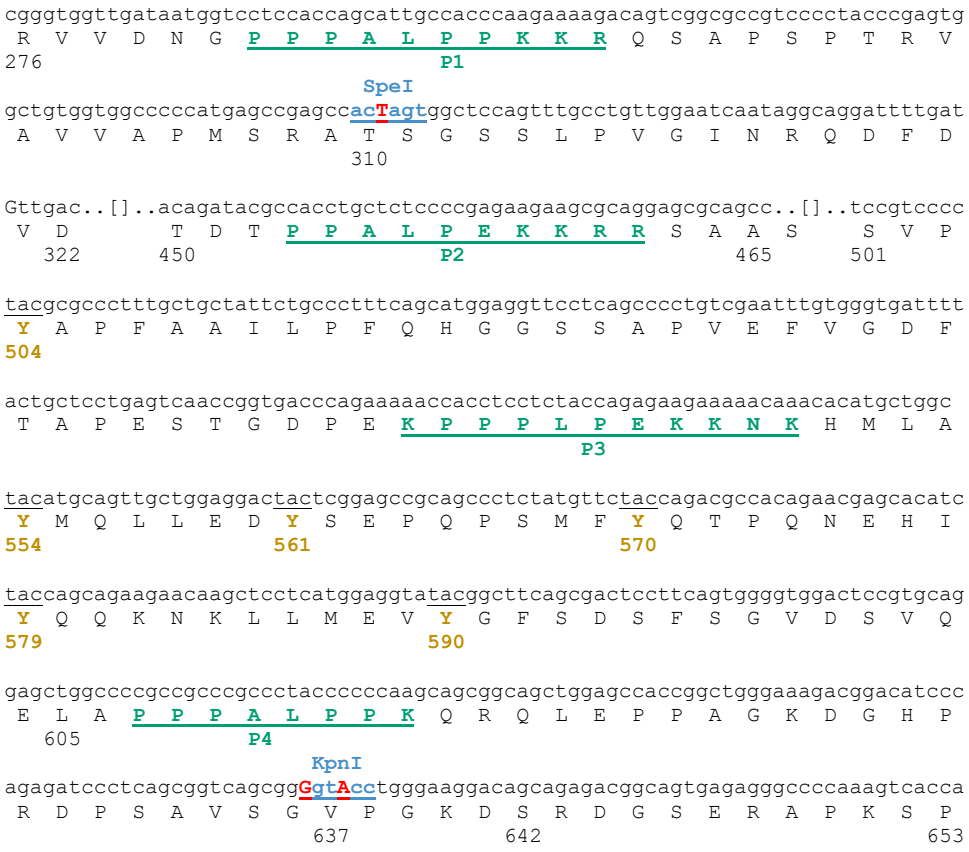



**Figure M3. Details of the sequence of C3G that include the generated box SpeI and KpnI.** Marked in blue the sequence of the SpeI and KpnI sites generated in the sequence of C3G. In red the nucleotide mutated to generate the restriction sites. Marked in green the PRMs P1 to P4. Marked in gold the tyrosine residues of interest.

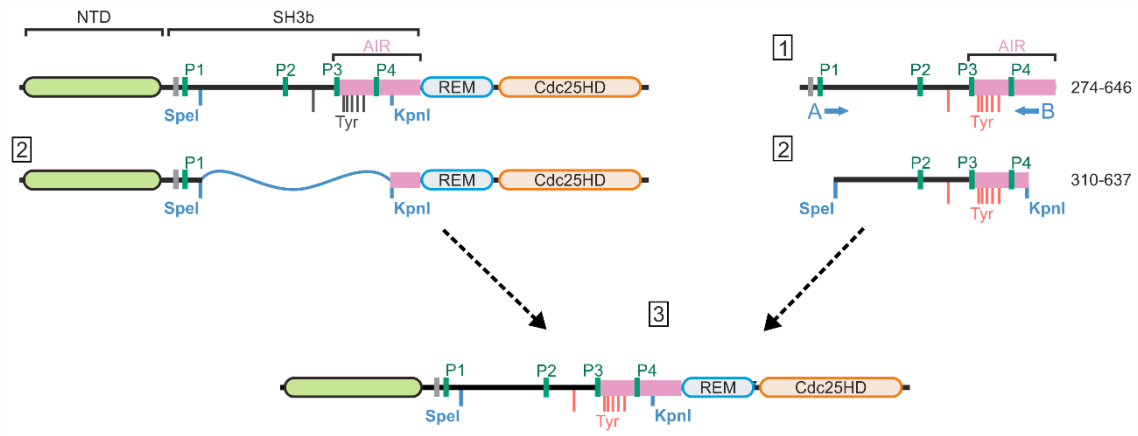



**Figure M4. Strategy used to subclone tyrosine residue mutants within the SH3b into full-length C3G. (1)** Amplification of the region containing tyrosine residues of interest using the 246-646 construct as template. Primers C3Gh-310-Spe-For (A) and C3Gh-637-KpnI-Rev (B) were used to amplify the region and introduced the restriction sites SpeI and KpnI. **(2)** Digestion of C3G full-length Δ310-637 construct and the amplified DNA fragment with SpeI and KpnI restriction enzymes. **(3)** Ligation of the two fragments generated the complete C3G sequence with the mutated SH3b box.

Mutants of C3G full-length in the pEF1-mEGFP and pLenti-C-mEGFP-IRES-BSD vectors were created by PCR amplification of the C3G cDNA in the bacterial expression vectors containing the desired mutations, using primers C3Gh-004-KKB-F (5' TGAGGTACCGCCGCCACCATG GGATCCGACTCTCAGCGTTCTCATC 3') and C3Gh-1077-noStop-R-NotI (5' GAATTCGCG GCCGCGGTCTTCTCTTCCCGGTC 3') that add restriction sites for BglII and NotI at the 5' and 3' ends, respectively (Carabias et al., 2020, Carabias, 2019). The PCR products were digested with BamHI and NotI and cloned in the pEF1-mEGFP and pLenti-C-mEGFP-IRES-BSD vectors were digested with the same enzymes.

### CrkL and CrkII constructs and mutants

Constructs for the expression in bacteria of CrkL and its fragments, as well as CrkII, are summarized in Table M6. The constructs coding for full-length CrkL in the vectors pETEV15b and pETEV15b-Avi that codes for a C-terminal Avi-Tag have been described (Carabias, 2019). Similarly, constructs of the SH2-SH3N, SH3N-SH3C, and SH3N of CrkL, and full-length CrkII in the pETEV15b vector have already been described (Carabias, 2019). Constructs coding for chimeric proteins of CrkL and CrkII were created by overlap extension PCR (Appendix I, Table A2).

**Table M6. CrkL and CrkII constructs for bacterial expression.**

| Name | Limits | Description | Vector | Sites | Purpose[c] |
|---|---|---|---|---|---|
| CrkL[a] | 1-303 | Full-length | pETEV15b | NdeI/BglII | ITC/NEK/SEC |
| CrkL-Avi[a] | 1-303 | Full-length | pETEV15b-Avi | NdeI/BglII | PD/ANIS |
| CrkL-Avi | 1-303 | Full-length | pETEV15b-Halo-Avi | NdeI/BglII | NEK/SEC |
| CrkL-SH2-SH3N[a] | 1-182 | SH2-SH3N | pETEV15b | NcoI/BamHI | NEK |
| CrkL-SH3N-SH3C[a] | 125-303 | SH3N-SH3C | pETEV15b | NdeI/BglII | NEK |
| CrkL-SH3N[a] | 125-182 | SH3N | pETEV15b | NdeI/BglII | NEK/ITC |
| GST-CrkL | 1-303 | Full-length | pGEX-4T3-NNB-TEV | NcoI/BglII | PD |
| GST-CrkL-SH3N[b] | 111-204 | SH3N | pGEX-4T3 | SalI/NotI | PD |
| CrkII[a] | 1-304 | Full-length | pETEV15b | NdeI/BamHI | NEK |
| CrkL-II-L | CrkL 1-124<br>CrkII 134-191<br>CrkL 183-303 | Chimeric protein | pETEV15b | NdeI/BamHI | NEK |
| CrkII-L-II | CrkII 1-133<br>CrkL 125-182<br>CrkII 191-304 | Chimeric protein | pETEV15b | NdeI/BamHI | NEK |

[a] Described in (Carabias, 2019).
[b] A generous gift from Carmen Guerrero.
[c] PD, pull downs; ITC, isothermal titration calorimetry; NEK, nucleotide exchange kinetics; SEC, size exclusion chromatography; ANIS, anisotropy.

Point mutations were introduced in CrkL constructs using the PCR-base method as for the C3G mutants (see above). CrkL mutants are summarized in the Table M7. Primers used to create CrkL mutants are shown in the Appendix I, Table A3. The point mutant CrkL-R39K was initially created in the pETEV15b-His-Avi vector, yet it was expressed insoluble. To improve its solubility CrkL was cloned in the pETEV15b-His-Halo-TEV vector and the mutants R39K and R39K-W160S were created in this construct.

**Table M7. CrkL mutants generated in this work.**

| Construct | Vector | Mutations |
|---|---|---|
| CrkL-Avi full-length | pETEV15b-His-Halo-TEV-Avi | R39K<br>R39K-W160S |
| CrkL-Avi full-length | pETEV15b-His-Avi | R39K,<br>W160S<br>P174R<br>Y177S |
| GST-CrkL full-length | pGEX-4T3-NNB-TEV | R39K<br>W160S<br>P174R<br>Y177S<br>R39K-W160S |

### Constructs for the expression of other proteins in bacteria

The construct of the G-domain (1-167) of human Rap1b in the pETEV15b vector, and the construct for the co-expression of the kinase domain of Src (hereafter Src-KD) (residues 251-536) and the YopH phosphatase in the pETDuet vector have been described (Carabias, 2019, Carabias et al., 2020). The cDNA encoding the nanobody abGFP4 that recognizes GFP (Kubala et al., 2010) was subcloned into a derivative of pET22b vector developed in our laboratory that lacks the peIB sequence for periplasmic localization and codes for a non-cleavable C-terminal poly-His tag followed by a cysteine; this vector is named pET22bX2-Cys. The cDNA of abGFP4 was PCR amplified using as template the plasmid pOPINE GFP nanobody (a gift from Brett Collins, Addgene plasmid #49172) and primers that added NcoI and XhoI restriction sites at the 5´ and 3´ ends, respectively (Appendix I, Table A4). The PCR product was cloned into the pET22bX2-Cys using these restriction sites. These plasmids are summarized in Table M8.

**Table M8. Constructs for the expression of other proteins in bacteria.**

| Name | Limits | Description | Vector | Sites | Purpose |
|---|---|---|---|---|---|
| Rap1b | 1-167 | G-domain | pETEV15b | NdeI/BamHI | NEK |
| Src-KD | 254-536 | Kinase domain | pETDuet-HisTEV | EcoRI/HindIII | IPH |
| GST | n.a. | GST protein | pGEX-4T3-NNB | n.a. | PD |
| abGFP4 | n.a. | anti-GFP nanobody | pET22b-X2-Cys | NcoI/XhoI | IP |

NEK, nucleotide exchange kinetics; IPH, *in vitro* phosphorylation assay; PD, pull-downs; IP, immunoprecipitation.

## 2. Bacterial protein expression

Recombinant proteins were expressed in the BL21(DE3)T1 *Escherichia coli* strain. This strain contains the gene for the T7 bacteriophage RNA polymerase under the control of the lacUV5 promotor. This RNA polymerase is inducible by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG). Bacterial cultures were grown in Terrific Broth (TB) media (Sambrook et al 1999), supplemented with 100 mg/L ampicillin. In particular cases of difficult expression of the constructs or mutants (Rap1 and C3G-full length N575C mutant), soluble expression required co-over-expression with GroEL/ES chaperones using the plasmid pBB541 (a gift from Bernd Bukau, Addgene plasmid # 27394) (de Marco et al., 2007); in these cases, the media was supplemented with 50 mg/L spectinomycin. Cultures were grown at 37 ºC with shaking until reaching an optical density at 600 nm (OD 600) ranging from 0.6 to 0.8. At this point, cultures were cooled down, and the induction of protein expression was initiated by adding 0.2 mM

IPTG. Expression was carried out at 15 ºC for 15-24 hours, depending on the specific construct. After induction, bacterial cells were harvested by centrifugation (16700 x*g* or 9000 rpm, rotor JLA 10.500) for 20 minutes at 4 ºC. The resulting cell pellets were resuspended in a volume with a proportion of 1 mL of buffer per 25mL of culture. The buffer composition consisted of 20 mM Tris-HCl (pH 7.9), 500 mM NaCl, 5 mM imidazole, supplemented with 0.1% (v/v) Triton X-100. For the C3G full-length proteins, the concentration of imidazole was increased to 20 mM. For the bacteria lysis, the cell pellet suffered one freezing-thawing cycle, followed by sonication during 6 min (cycles of 50 seconds separated by 10 seconds breaks) at 27% amplitude using a VCX500 sonicator equipped with a microtip (Sonics). The soluble portion of the bacteria lysate was separated from cell debris by centrifugation (39000 x*g* or 19000 rpm, rotor JA25.50) for 30 minutes at 4 ºC. The resulting supernatant was sonicated (2 min) to reduce its viscosity.

**Purification of His-tagged proteins**

The proteins CrkL full-length (wild type and point mutants), CrkL-SH2-SH3N, CrkL-SH3N-SH3C, CrkL-SH3N, CrkII (full-length), CrkL/CrkII chimeras, the kinase domain of Src (Src-KD), the abGFP4 nanobody and Rap1b, all of which carry a His-tag, were purified by immobilized-metal affinity chromatography (IMAC). Supernatants of the bacterial lysates were loaded on a 5 mL column of $Ni^{2+}$ chelating agarose (ABT). The proteins were eluted using an imidazole gradient ranging from 5 to 500 mM in 20 mM Tris-HCl (pH 7.9) and 500 mM NaCl. Fractions collected from the chromatography were analyzed by SDS-PAGE and Coomassie staining. Fractions containing the protein of interest were pooled and dialyzed against 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. Depending on the intended use of the preparation, the His-tag was cleaved using rTEV-His protease at room temperature (RT) for 2-4 hours while sample was dialyzed against the same buffer. After that, digestion was prolongated overnight at 4 ºC. After digestion, the cleaved protein was separated from the undigested sample and the rTEV-His protease by a second IMAC. A final purification step of size exclusion chromatography (SEC) was performed using a HiPrep Sephacryl S-200 or S-300 HR column (Cytiva) equilibrated in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. Pure proteins were concentrated by ultra-filtration in Amicon cells (Millipore) using YM3 or YM10 membranes (Millipore) or centrifugal filters with 3 or 10 kDa cut-off (Merck Millipore). Following concentration, proteins were centrifuged to eliminate any precipitants, their concentrations were measured spectrophotometrically using the extinction coefficients

calculated from the amino acid composition. Finally, samples were flash-frozen in liquid nitrogen and stored at -80 ºC for long-term preservation.

Some proteins needed minor modifications as follows. For Rap1b purification all buffers were supplemented with 10 mM $MgCl_2$. For the purification of the kinase domain of Src (Src-KD) the buffer for the first IMAC was 50 mM Tris-HCl (pH 8.0), 25 mM imidazole and 5% (v/v) glycerol, and the imidazole gradient extended up to 250 mM. Importantly, the Src-KD protein is not digested, and the His-tag is retained. After the IMAC, the Src-KD was dialyzed against 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 5% glycerol and 1 mM DTT. Next, an anion exchange chromatography was carried out using a 5 mL HiTrap Q HP column (Cytiva) equilibrated in 20 mM Tris-HCl (pH 8.0), 5% glycerol and 1 mM DTT. The Src-KD was eluted with a 0 to 35% gradient of 1 M NaCl in the previous buffer. Finally, the Src-KD was dialyzed against 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 5% glycerol and 1 mM DTT buffer and was stored as above.

**Purification of proteins with His-Halo-tag**

C3G full-length, wild type and mutants, were expressed as fusion proteins with an N-terminal tag consisting of a poly-His followed by the Halo proteins and a site recognized by the TEV protease. The Halo protein improves the expression levels of full-length C3G. Purification of His-Halo-tagged proteins was similar to that of His-tagged proteins with minor differences. In the first IMAC, the soluble bacterial fraction was loaded sequentially onto two $Ni^{2+}$ chelating columns. In the first column, the majority of free His-Halo protein was retained, while a fraction of His-Halo-C3G was not retained. The second column was used to capture the remaining His-Halo-C3G protein. Proteins from both columns were eluted with an imidazole gradient as above. Fractions containing the His-Halo-C3G protein were pooled and dialyzed against 20 mM Tris-HCl (pH 7.5), 300 mM NaCl, and digestion with rTEV-His was done as described. Similarly, for the second reverse IMAC step, sequential loading onto two columns was required to retain completely the free His-Halo protein. Next, the sample was purified by SEC in a Superdex 200 column (10/300 mm, packed in house) equilibrated in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl. Finally, C3G samples were concentrated by ultrafiltration and stored as described above.

### Purification of GST-tagged proteins

GST-C3G and GST-CrkL wild type and mutant constructs, and GST-RalGDS were purified by affinity chromatography as follows. The supernatants from bacterial lysates containing GST-fusion proteins were loaded onto 5 mL glutathione agarose columns (ABT). The proteins bound to the column were washed in two steps; first, by passing five column volumes of 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1% Triton X-100 followed by six column volumes of the previous buffer but without Triton X-100. The proteins of interest were eluted using buffer containing 10 mM reduced glutathione. The eluted fractions were checked using Bradford assay and SDS-PAGE. Protein-containing fractions were extensively dialyzed against 20 mM Tris-HCl (pH 7.5), 150 mM NaCl buffer. Some proteins eluted containing a high amount of free-GST, prompting an additional step of size exclusion chromatography (SEC) using a Superdex 200 (10/300) column (Cytiva) equilibrated in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl. Finally, proteins were concentrated by ultrafiltration, frozen and stored as above.

### Measurement of protein concentrations

Protein concentration was assessed through spectrophotometric UV absorption at 280 nm ($A_{280}$). Absorbance was measured directly (i.e., without additional dilution) using a NanoDrop spectrophotometer. Alternatively, samples were diluted in 20 mM Tris-HCl, 150 mM NaCl, 0.1% SDS and the absorbance was measured in U-2001 UV-Visible spectrophotometer (Hitachi) using quartz cuvettes with a light path of 10 mm (Hellma GmbH). For concentration determination, the mass ($E_{280}$) and the molar ($\mathcal{E}_{280}$) extinction coefficients were estimated from the theoretical protein sequences using the ProtParam server (https://web.expasy.org/protparam/). The concentration of nucleotide-loaded Rap1b protein was estimated by Bradford assay using bovine serum albumin (BSA) as a standard.

## 3. Protein phosphorylation with Src

C3G samples were phosphorylated *in vitro* with recombinant kinase domain of Src (Src-KD, residues 354-536).

Typically, 1 to 3 mg of purified full-length C3G proteins were phosphorylated with 1 µM of Src-KD in a final volume of 500 µL of a phosphorylation buffer consisting of 50 mM Tris-HCl (pH 7.5), 300 mM NaCl, 10 mM $MgCl_2$, 1mM DTT and 1 mM ATP. The

phosphorylation was carried out at 30 ºC for 30 minutes and was stopped by placing the samples on ice. Phosphorylated C3G samples were then purified to remove Src-KD and free ATP by SEC with a Superdex 200 (10/300 mm) column equilibrated in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl. The levels of phosphorylation were assessed by immunoblotting using a pan anti-phospho-tyrosine antibody (anti-pY).

To analyze phosphorylation time courses, 6 µg of C3G full-length or fragments in 20 µL volume (0.3 mg/mL) were incubated with 1 µM of Src-KD in phosphorylation buffer at 30 ºC. The reactions were stopped at various times by adding an equal volume of Laemmli sample buffer 2x (125 mM Tris-HCl (pH 6.8), 4% SDS, 55% glycerol, 2% β-mercaptoethanol, 0.02% bromophenol blue), followed by boiling at 100 ºC for 7 minutes. The levels of phosphorylation were assessed using immunoblotting with the anti-pY antibody.

**4. Nucleotide exchange experiments**

To determine the nucleotide dissociation rates from Rap1b mediated by C3G, fluorescence spectroscopy techniques were employed using the fluorescentGDP analogue mant-dGDP (2´-Deoxy-3´-O-(N-methylanthraniloyl) guanosine-5´-O-diphosphate sodium salt) (Margarit et al., 2003, Rehmann, 2006). The fluorescence of mant-dGDP is sensitive to the chemical environment. The fluorescence intensity when bound to the GTPase (hydrophobic environment) is approximately twice that of the free nucleotide (hydrophilic environment). Consequently, the dissociation of mant-dGDP from the GTPase can be monitored as a reduction in the fluorescence intensity over time in the presence of an excess of non-labeled GDP.

**Rap1b loading with mant-dGDP**

Recombinant GTPase Rap1b (G-domain; 1-167) at 200 µM was loaded with a 10-fold molar excess (2 mM) of mant-dGDP by incubation in a buffer 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, 4 mM ethylenediaminetetraacetic acid (EDTA), 1 mM DTT for 1 hour and 30 min at 4 ºC; typically, 500µL reactions were done. Afterward, to stop the reaction, 10 mM $MgCl_2$ was added, and the reaction was additionally incubated 30 minutes at 4 ºC. Precipitation was observed during mant-dGDP loading of Rap1b; therefore, the sample was clarified by centrifugation at 16000 x*g* for 20 minutes at 4 ºC. The excess of free nucleotide was removed by SEC using a preparative Superdex 200 (10/300) column equilibrated in 20 mM Tris-HCl (pH 7.5), 50 mM NaCl, and 10 mM $MgCl_2$ buffer. SEC

fractions were analyzed by SDS-PAGE, and those containing Rap1b were pooled and concentrated by ultrafiltration up to ~200 µM. The concentration of mant-dGDP-loaded Rap1b was determined by Bradford assay using BSA as standard.

**Fluorescence measurements and analysis of dissociation kinetics**

Fluorescence measurements were conducted employing a Fluoromax-3 Spectrofluorometer (Horiba-Jobin Yvon) with peltier temperature control (Newport) and using a cuvette with a light path of 3 by 3 mm (Hellma GmbH). The measurements were performed by exciting the mant fluorophore at 370 nm (1 nm bandwidth) and detecting emission at 430 nm (10 nm bandwidth), using an integration time of 0.1 seconds. In experiments longer than 60 minutes, photobleaching was prevented by applying excitation light only during the detection period. Nucleotide exchange experiments were carried out at a controlled temperature of 25 ºC in a buffer containing 50 mM Tris-HCl (pH 7.5), 150 mM NaCl, 5 mM $MgCl_2$. Typically, the reaction volume was 70 µL and Rap1b/mant-dGDP was used at 200 nM. To initiate the nucleotide reaction, free 40 µM GDP (200-fold molar excess compared to Rap1b) was manually added and mixed, resulting in a delay of 8-15 seconds before the start of the measurements, which was taken into consideration. The C3G construct and mutants were mostly employed at 1 µM, while phosphorylated C3G was used at 0.2 µM to facilitate the measurement. In some cases, additional proteins such as CrkL, CrkII, or specific C3G constructs were included as specified. Data acquisition continued until the fluorescence signal reached a plateau. Under these conditions, the nucleotide dissociation of Rap1b followed a pseudo-first-order kinetic model, which can be described as a single mono-exponential decay function that was fitted to the experimental data:

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

Where $I_t$ is the fluorescence intensity at the given time, $B$ is the fluorescence intensity at the infinite time when all the mant-dGDP is dissociated, $A_0$ is the amplitude of the change, $k_{obs}$ is the apparent nucleotide exchange rate and $t$ is the time. A python script was used to fit the single exponential decay and to estimate the parameters. In general, experimental data fitted well to the equation ($R^2 > 0.9$). For representation, data were normalized to 1 (initial intensity) and 0 (final intensity complete dissociation). Unless otherwise specified, experiments were performed in triplicate.

**Analysis of the concentration proportionality of C3G in the nucleotide exchange activity**

The relationship between C3G concentration and its nucleotide exchange activity was analyzed by measuring nucleotide exchange experiments with a Src-phosphorylated C3G (pC3G) at concentrations ranging from 0.1 to 1 µM, either alone or in the presence of CrkL (5 µM). Interestingly, within this concentration range, the activity of pC3G is directly proportional to the concentration of pC3G, following a linear-dependent model. This allowed for the presentation of the nucleotide exchange activities ($k_{obs}$) of pC3G, which were performed at 0.2 µM, as the specific activity per 1 µM pC3G using the conversion equation of $k_{obs}$/[pC3G].

**Analysis of dose response experiments**

The dose-response experiments were analyzed by measuring nucleotide dissociation experiments at a fixed concentration of C3G and multiple concentrations of the binding partner proteins. Activation or inhibition curves were used depending on these additional proteins. Independently of the type of curves used, changes in the nucleotide exchange rate constants of C3G were analyzed by fitting the following sigmoidal equation:

$$k_{[bp]} = k_{free} + \frac{\left(k_{max} - k_{free}\right)}{1 + \left(\frac{EC50}{[bp]}\right)}$$

Where, $k_{[bp]}$ is the $k_{obs}$ determined at a given concentration of the binding partner; $k_{max}$ is the maximum value of the rate constant, which corresponds to saturated conditions; $k_{free}$ is the minimum value of $k_{obs}$, which is the activity in the absence of the binding partners; $EC_{50}$ is the concentration of the binding protein that produces the half maximal activation ($AC_{50}$) or inhibition ($IC_{50}$); and [*bp*] is the concentration of the binding partner. Analysis was done with Sigmaplot (Systat Software Inc.).

**5. Size exclusion chromatography**

Size exclusion chromatography (SEC) is a technique that allows the separation of molecules based on their size and, in some cases, their molecular weight, by passing through commercially available or home-made packed columns.

We distinguish between two types of SEC runs. The first type involves using SEC as the last step of purification protocol. Typically, C3G proteins were run through a Superdex 200 (10/300) column (GE Healthcare) pre-equilibrated with a buffer containing 20 mM Tris-HCl (pH 7.5) and 300 mM NaCl. Conversely, CrkL proteins were processed through a HiPrep Sephacryl S-200 or S-300 HR column (Cytiva), depending on the specific construct. The isolated SH3N domain (7.5 kDa) was run through the S-200 column, whereas the rest of the constructs (full-length CrkL, chimeras or fragments with two domains) were run using the S-300 column. Independently of the column utilized, the buffer consisted of 20 mM Tris-HCl (pH 7.5) and 150 mM NaCl.

The second type of SEC was used to analyze the interaction between full-length CrkL and C3G proteins. We employed an analytical Superdex 200 HR (10/300) column (GE Healthcare) pre-equilibrated with a buffer containing 20 mM Tris-HCl (pH 7.5) and 300 mM NaCl. The flow rate was 0.5 mL/min, and an injection volume of 100 µL. Isolated C3G and CrkL proteins were injected at a specific concentration (20 µM and 80 µM, respectively). Mixtures of both proteins were performed using the same concentrations to compare the elution profiles. Molecular weights (MW) of the proteins were calculated from the light scattering and the refractive index at the center of the peak. Additionally, SEC fractions corresponding to the peaks of interest were precipitated using the trichloroacetic (TCA) protocol and checked by SDS-PAGE gels and Coomassie staining.

### 6. Fluorescence anisotropy-based binding assay

Fluorescence anisotropy is the polarization produced by a sample that emits light after being excited by polarized light. Fluorescence depolarization occurs due to the rotational diffusion of the fluorophores, which is related to the size and shape of the species. Thus, anisotropy is sensitive to protein associations.

Peptides corresponding to the P1 (residues 280-293), P2 (450-463), P3 (537-550) and P4 (605-618) were custom synthesized labeled with N-terminal fluorescein (Thermo Scientific).

Labeled C3G-peptides at 0.2 µM in 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1 mg/mL BSA, were titrated with CrkL wild type and mutants. The fluorescence anisotropy was measured using an Ultra Evolution plate reader (Tecan) with excitation at 485 nm and emission at 535 nm filters at 25 $^{o}$C and 384-well black microplates (Greiner Bio-One).

The apparent dissociation constant ($k_d$) of the free ($r_F$) and fully bound ($r_B$) states, along with their standard errors, were determined by nonlinear square fitting of the following 1:1 binding model using Sigmaplot (Systat Software Inc.) as described (Manso et al., 2019):

$$r = r_F + \left[ (r_B - r_F) \left( \frac{\gamma - \sqrt{\gamma^2 - 4[\text{CrkL}]_T[\text{C3G}]_T}}{2[\text{CrkL}]_T} \right) \right]$$

$$\gamma = Kd + [\text{CrkL}]_T + [\text{C3G}]_T$$

Where, $[\text{C3G}]_T$ is the total concentration of labeled-C3G-peptides, $[\text{CrkL}]_T$ is the total concentration of added CrkL, $r$ and is the observed fluorescence anisotropy of the probe.

Typically, the highest useful concentration of CrkL attained in binding assays was below 50 µM and 250 µM for wild type and mutants, respectively. In titrations with CrkL wild type, it was possible to measure the anisotropy of the free state and to obtain a reliable estimation of that of the fully bound state. However, titrations with CrkL mutants (low-affinity interaction) only reached a partial saturation fraction; for that reason, in those cases the $r_B$ value was fixed during the fitting using the value obtained for the wild type CrkL.

### 7. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a useful non-label technique used to determine the thermodynamic parameters of binding reactions. It measures the heat evolved during the association of a ligand with its binding partner. ITC experiments were performed using a MicroCal VP-ITC calorimeter (Malvern Pananalytical). Protein samples were dialyzed against a common working buffer composed of 20 mM Tris-HCl (pH 7.5), 300 mM NaCl. Both the buffer and proteins were degassed using a ThermoVac machine (Malvern) for 7 minutes at 23 ºC just before the experiment. Solutions containing C3G proteins, ranging from 5 to 50 µM, were titrated with CrkL constructs and mutants ranging from 50 to 650 µM. Titrations were done at 25 ºC and typically consisted of 28 injections of 10 µL for 20 seconds, interspersed with 240 seconds between injections. In most of the cases, ITC experiments were performed in duplicate or triplicate. Buffer-protein experiments were also performed. In certain instances, to improve the heat exchange signal, volume of the injection was modified to 15 µL.

Data from individual experiments were initially analyzed with the program MicroCal Origin ITC module (Malvern, version 7.0) using an independent-site association model, which fits the association constant ($k_a$ = 1/$k_d$) and the relative stoichiometry (N). Alternatively, heat signals were automatically integrated with the program NITPIC (Keller et al., 2012). ITC experiments were further analyzed globally using the SEDPHAT software (Zhao et al., 2015). Two different hetero-association models were employed, depending on the specific interaction to analyze. The binding of CrkL to phosphorylated C3G was analyzed using a model with three symmetrical binding sites. Interactions between CrkL and C3G mutants displaying a single or two PRMs were analyzed using a 1:1 binding model.

SEDPHAT provided estimates for the association constant ($k_a$ = 1/$k_d$) and the enthalpy of the reaction (ΔH). SEDPHAT does not fit non-integral binding sites; instead, the incompetent fraction was estimated, and the relative stoichiometry (N) was calculated as the complementary value. These three parameters ($k_d$, N and ΔH) were estimated as the value and the asymmetric errors within a 0.95 confidence interval.

## 8. Protein electrophoresis, Coomassie staining, Western blot and antibodies

### SDS polyacrylamide gel electrophoresis

Protein samples were separated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using gels with polyacrylamide concentrations ranging from 9 to 15%. Samples were previously denatured by addition of Laemmli sample buffer 2x used to purified proteins, or the buffer with a 4x concentration to cell lysis samples; in both cases, samples were heated to 95 ºC for 5-10 minutes. Identification of protein sizes were facilitated using the Page Ruler Plus Prestained Protein Ladder (Thermo Fisher).

### Coomassie staining

To detect general proteins, polyacrylamide gels were stained using a solution containing 0.1% Coomassie (R250 reagent) with 10% of acetic acid and 40% of methanol for a duration of 10 to 60 minutes.

### Protein transfer to PVDF membranes and immunodetection

For immunoanalysis, proteins were transferred to Immobilon-P membranes (Millipore) by wet transfer system using transfer buffer (66 mM Tris-HCl (pH 8.3), 386 mM glycine, 0.1% SDS, 20% methanol) at constant intensity (0.3 A) for 2 hours at 4 ºC.

Membranes were blocked with 5% non-fat dried milk in TBS-T (10 mM Tris-HCl (pH 7.3), 150 mM NaCl, 0.5% Tween-20) for 20 to 60 minutes at RT. After blocking, membranes were incubated with primary specific antibodies (Table M9) diluted in TBS-T with 2% (w/v) BSA for 2 hours at RT or overnight at 4 ºC on rotation. Membranes were washed with TBS-T and incubated with secondary antibodies (Table M10) for 1 hour at RT in 5% non-fat dried milk in TBS-T in darkness. Afterwards, membranes were washed with TBS-T before detection.

**Table M9. Primary antibodies used for Western blot detection.**

| Antibody | Host | Supplier | Reference | Dilution |
|---|---|---|---|---|
| Anti-C3G (G9) | Mouse | Santa Cruz Biotechnology | sc-393836 | 1:1000 |
| Anti-His | Mouse | SIGMA | H1029 | 1:1000 |
| Anti-pY (4G10) | Mouse | Merck Millipore | 05-321 | 1:1000 |
| Anti-GFP (B-2) | Mouse | Santa Cruz Biotechnology | sc-9996 | 1:1000 |
| Anti-CrkL (C-20) | Rabbit | Santa Cruz Biotechnology | sc-319 | 1:1000 |
| Anti-Rap1 | Rabbit | Santa Cruz Biotechnology | sc-65 | 1:800 |
| Anti-β-actin (AC-15) | Mouse | Sigma-Aldrich | A5441 | 1:1000 |

Antibody signals were detected by infra-red fluorescence using Odyssey Infrared Imaging System (LI-COR) or by chemiluminescence using Fuji X-ray films (FujiFilm Corporation) and ECL reagents (Clarity Western ECL blotting substrates kit, Bio-Rad) following the standard protocols.

**Table M10. Secondary antibodies used for Western blot detection.**

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

HRP, horseradish peroxidase; Odyssey, Odyssey Infrared Imaging System; ECL, Chemiluminescent detection

### 9. Mass spectrometry analysis of the phosphorylation *in vitro* of C3G by Src

Mass spectrometry was employed to identify the specific tyrosine residues of C3G phosphorylated *in vitro* by Src-KD. The phospho-sample were prepared according to the standard phosphorylation protocol outlined in Method point 3. The levels of phosphorylation and the quality of the samples were evaluated through Coomassie staining and immunoblotting using the anti-pY antibody.

Subsequently, for the mass spectrometry analysis, samples were sent on dry ice to the proteomic facility at The Netherlands Cancer Institute (NKI) in Amsterdam, where facility members conducted the analysis.

### 10. Analysis of protein-protein interactions *in vitro* with pull-down assays

Pull-down (PD) assays were conducted using purified proteins or bacterial lysates containing GST-C3G or GST-CrkL fusion proteins. Typically, in the case of purified proteins, between 10-30 µg of GST alone or GST-fusion proteins were immobilized with 20 µL of glutathione agarose resin (ABT) for 10-15 minutes at 4 $^{o}$C. In the case of bacterial lysate, ~120 µL of the lysate (~20-40 µg of GST-fusion protein) were used after previously checking the levels of protein expression by Coomassie staining to normalize the loading of the different samples. GST-proteins were mixed with His-tagged C3G, His-Avi-tagged CrkL, or untagged C3G full-length in a buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1% Triton X-100, supplemented with 1 mg/mL of BSA to block unspecific binding. The samples were incubated for 1 hour at RT or overnight at 4 $^{o}$C. Subsequently, the resin was washed 3-5 times with 500 µL of the previous buffer, through a centrifugation with a fast pulse at 16000 x*g* for 15 seconds at 4 $^{o}$C. Proteins were extracted by adding Laemmli sample buffer 2x and heating to 100 $^{o}$C for 5-10 minutes. The presence of GST and His-proteins in the PD or the lysates were detected either by Coomassie staining (for the GST-proteins) or Western blot (for the His-tagged, His-Avi-tagged proteins or full-length C3G). Typically, around 3 µg of GST-protein, corresponding to 10% (v/v) of the PD sample, was loaded onto SDS-PAGE for detection, and approximately ~2-8% (v/v) for Western blot analysis.

## 11. Mammalian cell line culture

Two cell lines have been used in this work: HEK293T (American Type Culture Collection (ATCC), reference CRL-3216) is a highly transfectable (containing the SV40-T antigen) derivate cell line of HEK293 from a human embryonic kidney tissue, which exhibits epithelial morphology and grows as adherent monolayers. HEK293T cells were cultured in Dulbecco´s Modified Eagle Medium (DMEM; Gibco) supplemented with 10% fetal bovine serum (FBS; Gibco), 2 mM glutamine, 100 U/mL penicillin, and 100 µg/µL streptomycin. For maintenance, cells were grown to 80-90% confluence, in 10 cm plates, and were detached by trypsinization with 0.025% Trypsin-EDTA (Gibco).

Jurkat (ATCC reference, TIB-152) is a T-lymphoblastic lymphoma cell line and is the most commonly used as T cell line. Jurkat cells were cultured in suspension in RPMI-1640 medium (Sigma-Aldrich or Gibco), supplemented with 10% FBS, 2 mM glutamine, 100U/mL streptomycin. EGFP-positive clones were grown in the same medium supplemented with 10 µg/µL of blasticidin antibiotic. Maintenance was performed using 75-$cm^2$ culture flasks (40 mL). Both cell lines were grown at 37 ºC with a 5% $CO_2$ humid atmosphere conditions.

## 12. Transfection of mammalian cells

In this work two different transfection techniques have been used.

### Standard transfection

HEK293T cells at 60-70% confluence in a 10 cm diameter plate were transfected using between 5-10 µg of DNA. DNAs were preincubated with polyethylenimine (PEI) at a ratio of 2.5 µL per 1 µg DNA, in 150 mM NaCl for 30 minutes at RT to facilitate the transfection process. Expression of constructs was assessed by fluorescence microscopy techniques after 24 hours of transfection. Cells were processed for subsequent analyzes 48-72 hours of transfection.

### Production of lentiviral particles and lentiviral transduction

Lentiviral transduction was used for the stable expression of C3G-mEGFP in Jurkat cells. Lentiviral particles were generated in HEK293T cells. For this purpose, cells were transfected with three viral plasmids: the lentiviral construct (pLenti-mEGFP or pLenti-

mEGFP-C3G lentiviral plasmids); the lentiviral packing vector (pCMV-deltaR8.91 plasmid; Lifescience Market) expressing the Gag, Pol, Rev and Tat genes; and the second-generation VSV-G envelope-expressing plasmid (pMD2.G; provided by D. Trono Addgene no. 12259) containing the G glycoprotein (VSV-G), which facilitates the formation of the lentiviral particles.

HEK293T cells were seeded in 10 cm culture dishes (at a density of 5 x $10^5$ cells per plate) 15-24 hours before transfection. The mix of the plasmids was composed of 9 µg of pLenti-mEGFP plasmid, 3 µg of pMD2.G plasmid, 5 µg of pCMV-deltaR8.91 plasmid, and 60 µL of PEI in a final volume of 500 µL of 150 mM NaCl. This mixture was incubated for 30 minutes at RT and added to the cells. 24 hours post-transfection, the transfection levels were visualized by fluorescence microscopy, and subsequently, the culture medium was replaced with the complete RPMI Jurkat medium.

To obtain the lentiviral particles, two strategies were employed; i) HEK293T culture media collected at 48- and 72-hours post-transfection were centrifuged (2 hours and 30000 x*g*) and resuspended in a 200 µL medium; ii) culture media collected at 48-, 60- and 72-hours post-transfection were centrifuged for 5 minutes at 300 x*g* to eliminate cell contaminants. Then, in both cases, the media were filtered with 0.45 µm pore filter and supplemented with 4 µg/µL polybrene before addition to Jurkat cells.

The day before infection, a total of $10^6$ Jurkat cells were seeded. The following day, Jurkat cells were harvested by centrifugation and then resuspended in the medium containing the lentiviral particles previously filtered and supplemented with polybrene. Infected cells were selected with RPMI medium supplemented with 10 µg/µL of blasticidin after 24 hours of infection. The process of selection and expansion of the infected cells occurred over a period of three to four weeks. One-month post-infection the specific selection of mEGFP-positives cells was carried out using Fluorescence-Activated Cell Sorting (FACS). mEGFP-positives clones were always grown supplemented with 10 µg/µL of blasticidin antibiotic.

### 13. Mammalian anti-GFP co-immunoprecipitation assays

To detect the interaction in cells between CrkL and C3G-mEGFP, wild type and PRMs mutants, we performed co-immunoprecipitation assays using the recombinant anti-GFP nanobody abGFP4 immobilized on agarose beads.

### Coupling nanobody

The recombinant abGFP4 nanobody against GFP, with a C-terminal cysteine, was covalently coupled to SulfoLink resin (Thermo Scientific), a cross-linked agarose bead with an iodoacetyl group (Figure M5). The concentration ratio of the nanobody to resin was 4 µg of nanobody per 1 µL volume of resin. First, the purified nanobody was incubated with 5 mM Tris(2-carboxyethyl) phosphine (TCEP), a reducing reagent, to assure reduction of the C-terminal cysteine. Incubation was prolonged for 1 hour at RT in a coupling buffer composed of 50 mM Tris (pH 8.5), 5 mM EDTA. Next, the nanobody was mixed with the resin equilibrated in the coupling buffer and incubated for 1 hour at RT in the dark with gentle rocking. Unbound sample was eliminated by centrifugation, and the resin was washed four times with coupling buffer. To block any unreacted group, the coupled resin was incubated with 50 mM L-Cysteine in coupling buffer for 1 hour at RT. Afterward, the resin was washed three times with 1 M NaCl and finally equilibrated in the storage buffer (20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 50% glycerol). This affinity resin was stored at -20 ºC until used.

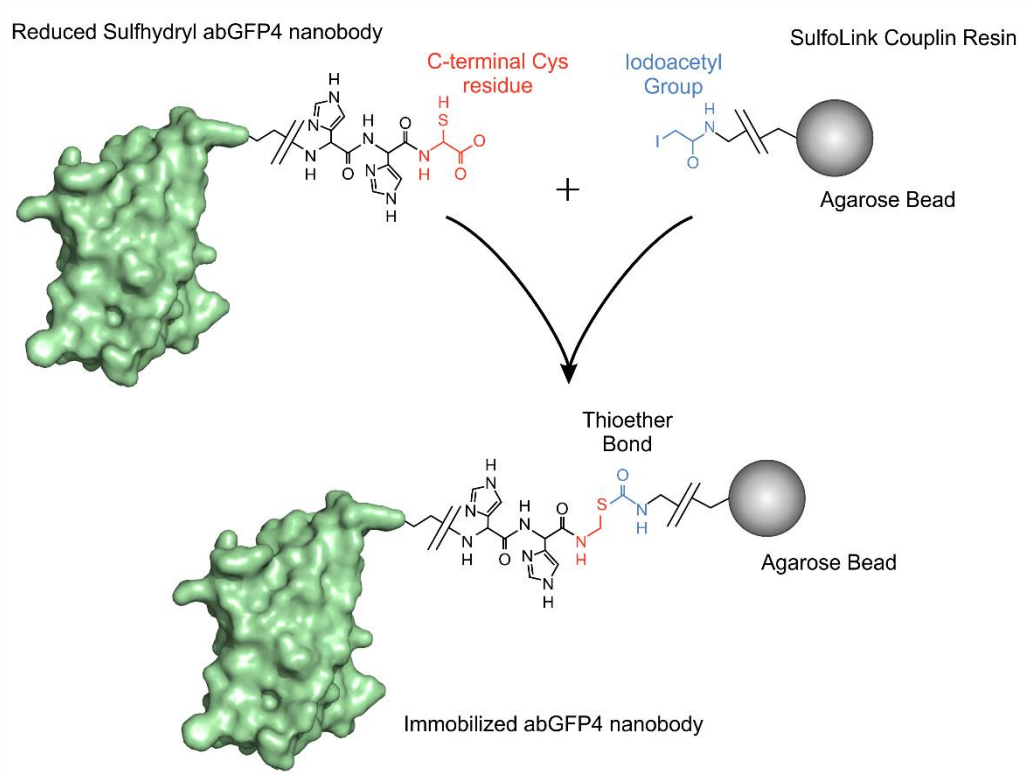



**Figure M5. AbGFP4 nanobody immobilization chemistry for SulfoLink Coupling resin.**

**Co-immunoprecipitation in Jurkat and HEK293T cell lines**

Typically, 30 x $10^6$ Jurkat cells or a confluent 10 cm plate HEK293T expressing the proteins of interest were washed with phosphate-buffered saline (PBS) and lysed in 0.3-1 mL of lysis buffer consisting of 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100, 1 mM sodium orthovanadate, 25 mM NaF, 1 mM phenylmethylsulfonyl fluoride (PMSF), and 1x protease inhibitor cocktail (cOmplete, Roche). Cells were lysed by incubation for 25 minutes on ice and vortex. Cell lysates were cleared by centrifugation at 16000 x*g* for 20 minutes at 4 ºC, and the protein concentration in the supernatant was determined by Bradford assay. Lysates containing 1 mg of total protein were then incubated with 30 µL of 33% slurry of abGFP4 nanobody resin against GFP, and the volume was completed to 500 µL with lysis buffer. Samples of the lysates before mixing with the resin were used to verify total protein levels by Western blotting. The immunoprecipitations were incubated overnight at 4 ºC with rotation. Afterward, the resins were washed three times with lysis buffer. Bound proteins were eluted with Laemmli sample buffer and boiled. The total clarified cell lysates and the co-IP samples were analyzed by Western blotting using a monoclonal anti-GFP antibody and a polyclonal anti-CrkL antibody.

**14. Confocal fluorescence microscopy**

Recruitment of C3G to the membrane in Jurkat cells upon stimulation of CD3 was analyzed by colocalization experiments using confocal microscopy. Coverslips were pre-treated with poly-L-lysine for 2 hours at RT. Jurkat cells (5 x $10^5$ cells per coverslip) were adhered to the coverslips for 30 minutes at 37 ºC. Cells were stimulated with 5 µg/mL of antibody OKT-3 (eBiosciences), which recognizes CD3, for 30 minutes at 37 ºC. After stimulation, cells were fixed using a final concentration of 4% paraformaldehyde (PFA) in PBS for 15 minutes at RT. Excess PFA was eliminated by three washes with PBS. Then, the cells were permeabilized with 0.2% Triton X-100 in PBS for 10 minutes at RT. After three washes, Jurkat cells were blocked for unspecific interactions with 1% BSA in PBS for 1 hour at RT, and then stained with 1:1500 phalloidin-iFluor-647 (ab176759, Abcam) to visualize actin filaments and delineate the plasma membrane of the cells. Nuclei were stained with 4´,6-diamidino-2-phenylindole (DAPI, Sigma) at 5 µg/mL. Coverslips were mounted onto glass slides with ProLong Diamond Antifade Mountant (Thermo Scientific).

Cells were visualized using a Leica SP5 confocal microscope with a 63x/1.4 Oil Plan-ApoChromat Ph1 immersion objective. The acquired images were processed with

Leica LAS X Software using Lightning process. Subsequently, the images were further processed and analyzed using ImageJ/Fiji Software (Schindelin et al., 2012). Protein colocalization was analyzed by using the Coloc2 plugin of Fiji distribution of ImageJ and calculating the Manders´ overlap coefficient (MOC) (Dunn et al., 2011). Additionally, the distributions of C3G-mEGFP and phalloidin-labeled proteins were analyzed using the Plot Profile Tool in representative cells for each condition.

### 15. Rap1b activation assays in cells

Activation of Rap1b in cells was analyzed by detecting active Rap1-GTP using a pull-down assay with the effector RalGDS-RBD protein fused to GST, which selectively binds to the active conformation of Rap1 (Vetter et al., 1999). Jurkat cells (~5 x $10^6$ per condition) in RPMI were stimulated with 5 µg/mL of the OKT-3 antibody against CD3 (eBiosciences) for 30 minutes at 37 ºC. After stimulation, cells were harvested by centrifugation and lysed in 200 µL of magnesium lysis buffer (MLB) composed of 25 mM HEPES (pH 7.5), 150 mM NaCl, 1% Igepal CA-630, 10 mM $MgCl_2$, 1 mM EDTA, 2% (v/v) glycerol, 1 mM $Na_3VO_4$, 1 mM PMSF, 1x protease inhibitor cocktail (cOmplete, Roche), and 1 mM DTT, in presence of 30 µg of purified GST-Ral-GDS protein. Supernatants were collected from the lysis, and approximately 15% of the samples was taken to analyze total protein amount. The remaining volume was added to 20 µL of glutathione-agarose resin and incubated for 20 minutes at 4 ºC. The resin was washed three times with 500 µL of MLB. Proteins bound to the resin were extracted by the addition of Laemmli sample buffer 2x, supplemented with 10 mM DTT, and incubated during 12 hours at RT. Samples were analyzed using 12% polyacrylamide gels followed by blotting onto Immobilon-P membranes. Total expression levels of Rap1 and mEGFP-tagged proteins were detected in the samples of total lysates, while Rap1-GTP and GST-RalGDS-RBD proteins were detected in the pull-down samples. GST-RalGDS-RBD was detected by staining the membranes with Ponceau S. Other proteins were detected by Western blot with specific antibodies (see above, Table M9).

### 16. Prediction of protein structures with AlphaFold2

The structure of the region of C3G 501-1077 that includes the AIR (537-646), the REM domain (646-810) and the catalytic domain Cdc25H (825-1077), was predicted using the artificial Intelligence-based method AlphaFold2 (Jumper et al., 2021) through the ColabFold notebook (v1.5.2) (https://github.com/sokrypton/ColabFold) using the

MMSeq2-based homology search (Mirdita et al., 2022). Five models were generated without using any existing structures as template. These models were then structurally aligned using PyMOL Molecular Graphics System, Version 2.6.0 (Schrödinger), which was also used to generate figures of molecular structures.

### 17. Statistical analysis

Data have been represented as the mean ± SD (Standard Deviation) or the mean ± SEM (Standard Error of the Mean), depending on the method, as indicated in each figure. Number of independent experiments from each experiment is indicated as well in the figures. The normal distribution of the data in the groups was first analyzed using the Shapiro-Wilk test. When the data followed a normal distribution and the variances were equivalent, as analyzed with the Brown-Forsythe homogeneity of variance test, the comparisons of multiple groups were done using one-way analysis of variance (ANOVA) with Tukey´s multiple comparisons test. For data not following a normal distribution, statistical comparisons were done using unpaired non-parametric Kruskal-Wallis test and Dunn´s multiple comparisons test. Statistical analyses were performed with the program GraphPad Prism 8 or Sigma Plot v14.0 software. Differences between groups were considered significant at $P > 0.05$. Significant levels were indicated as * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

# RESULTS

## 1. Role of Crk proteins and Src-phosphorylation in the activation of C3G

### 1.1. CrkL does not interact equally with the four binding sites in full-length C3G

Previous work from our group characterized the interaction between CrkL and C3G full-length (Carabias, 2019). CrkL binds to C3G at multiple sites with low micromolar affinity, and the interaction is highly dynamic. Four PRMs (P1 to P4) in the SH3b domain of C3G are binding sites for the SH3N domain of CrkL. Yet, on average only ~3 molecules of CrkL bind to each C3G molecule, and the P3 motif is mostly non-accessible for CrkL binding in wild type C3G (Carabias, 2019). This prompted us to analyze in further detail the interaction of CrkL with each individual PRM and the relationship between these bindings and the activation of C3G.

Initially, we analyzed the binding of CrkL full-length to the isolated PRMs by fluorescence anisotropy assays, using fluorescein-labeled peptides of the P1, P2, P3 and P4 (Figure R1A). CrkL bound to the four PRMs with very similar affinity, the estimated dissociation constants ($k_d$) ranged between 1.4 and 3 μM (Figure R1B); which are in agreement with the $k_d$ reported for the binding of the SH3N domain of c-Crk to a peptide of the P1 ($k_d$ 1.9 μM) (Wu et al., 1995).

Previous results suggested that in the context of the resting C3G full-length, which is autoinhibited, the accessibility of CrkL to the PRMs is not equal. To characterize the potential differences in accessibility, we analyzed the binding of CrkL to each of the four individual PRMs in full-length C3G. To disrupt the binding of CrkL to the PRMs, five residues that contact the SH3N domain and are essential for the binding (Wu et al., 1995) were mutated to Ala in each PRM (Figure R1A) (Carabias, 2019). We created four C3G mutants, each one conserved a single unmodified PRM and carried mutations in the other three. From this point on, these mutants are named PAAA, APAA, AAPA and AAAP, to define whether the P1, P2, P3 or P4 maintained the wild type sequence (P) or carried the mutations to Ala (A) (Figure R1C).

The interaction was analyzed using pull-down assays with the isolated SH3N domain of CrkL fused to GST (GST-CrkL-SH3N). C3G wild type bound specifically and efficiently to GST-CrkL-SH3N but not to isolated GST. C3G mutants displaying only the P1 (C3G-PAAA), P2 (C3G-APAA) or P4 (C3G-AAAP) bound to CrkL-SH3N, despite C3G-AAAP was pulled down slightly less than mutants PAAA and APAA. Finally, C3G-AAPA showed a very weak interaction with CrkL-SH3N (Figure R1D). These results suggest that P3 and P4 motifs are less accessible for CrkL binding than P1 and P2 in resting autoinhibited C3G.

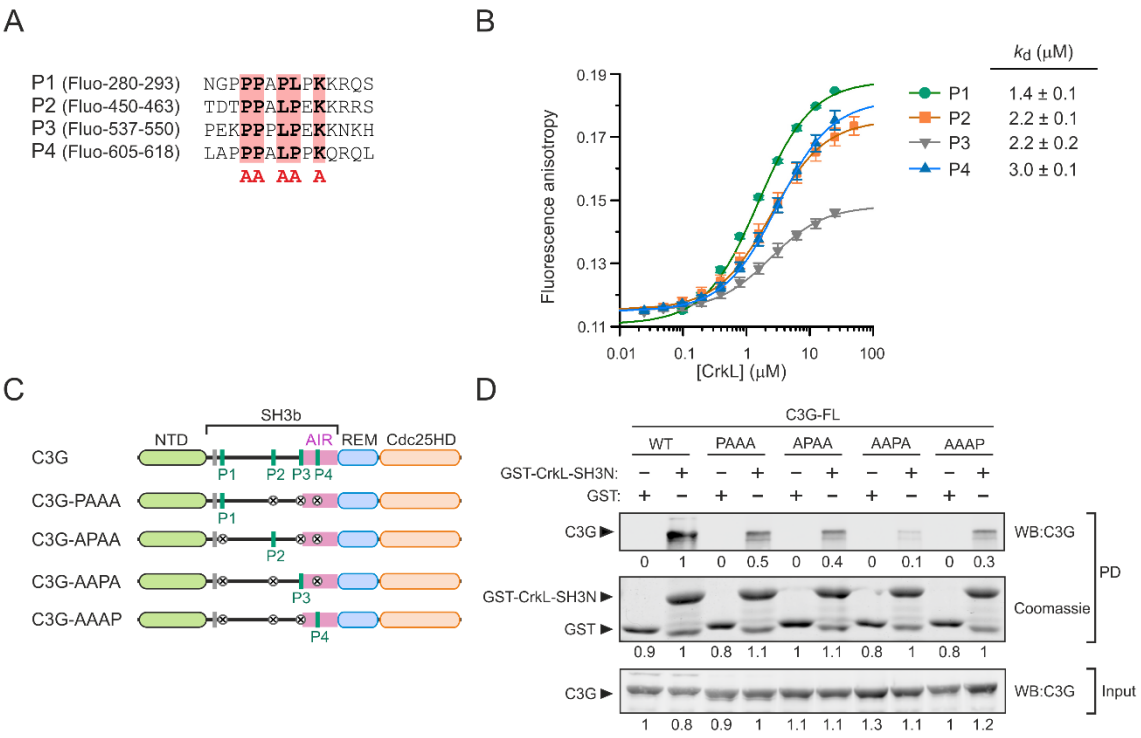



**Figure R1. Binding of CrkL to individual PRM sites of C3G. (A)** Sequences of four synthetic peptides that correspond to the P1 to P4 PRMs of C3G, which were used to determine their affinity for CrkL. The peptides were labeled with fluorescein at the amino-terminus. The five key residues in each motif that were later on mutated to Alanine (named with red A, in the figure) in C3G constructs to disrupt CrkL binding are marked by red boxes. **(B)** Fluorescence anisotropy titrations of the fluorescein-labeled peptides in A (0.2 μM) with CrkL. Lines represent the 1:1 binding model fitted to the data, which yielded the apparent dissociation constants ($k_d$) ± the asymptotic standard errors. Differences in the amplitude of the anisotropy changes induced by binding of CrkL are likely to reflect variations of the microenvironment of the fluorescein probe and are not related to the affinity. **(C)** Schematic representation of C3G full-length and the mutants, each containing a single wild type PRM. **(D)** Analysis by pull-down (PD) of the interaction of GST-CrkL-SH3N with C3G full-length (FL) wild type (WT) and mutants. C3G was detected in the PD and input samples by Western blot (WB). GST-CrkL-SH3N and GST (used as control) were visualized by Coomassie staining. Values are the relative quantification within each panel.

In resting C3G, the GEF activity is blocked by an intramolecular interaction through the AIR and Cdc25HD. The P3 is juxtaposed to the AIR and the P4 is within the AIR. Therefore, the accessibility to these sites might change between the autoinhibited and active states of C3G. To test this hypothesis, we analyzed the binding of CrkL to the P3 and P4 in the presence of the Y554H substitution that disrupts the AIR/Cdc25HD interaction and causes a constitutively activated state of C3G (Carabias et al., 2020). The active mutant C3G-AAPA-Y554H was pulled down more efficiently by GST-CrkL-SH3N than C3G-AAPA Y554 wild type. In contrast, C3G-AAAP-Y554H did not increase

binding to GST-CrkL-SH3N in comparison with C3G-AAAP wild type for Y554 (Figure R2A).

Phosphorylation of C3G by Src relaxes the autoinhibitory state and causes partial activation of C3G without disrupting the AIR/Cdc25HD interaction (Carabias et al., 2020). C3G full-length AAPA and AAAP mutants were phosphorylated *in vitro* with the kinase domain of Src (Src-KD) during 30 min at 30 ºC, and phosphorylated samples were purified by gel filtration. Phosphorylated C3G-AAPA was pulled down slightly more than the unphosphorylated protein, but less than the C3G-AAPA-Y554H mutant. In contrast, phosphorylation of the C3G-AAAP mutant did not increase the interaction with GST-CrkL-SH3N.

We used ITC as an orthogonal method to analyze quantitatively the effect of the Y554H substitution and the effect of tyrosine phosphorylation by Src on the binding of CrkL-SH3N to the P3 and P4 motifs. In these experiments, C3G full-length mutants, ranging from 7.5 to 21.8 μM, were placed in the reaction cell of the calorimeter and were titrated by injection of the SH3N of CrkL at concentrations ranging from 53 to 145 μM in the syringe. As a control, CrkL-SH3N, at the maximal concentration used to titrate C3G, was injected into buffer, resulting in minor heat exchanges (data not shown).

In all titrations, the binding of CrkL-SH3N to C3G mutants produced exothermic thermograms. The SH3N domain of CrkL bound to the P3 (C3G-AAPA) with similar stoichiometry (N ~0.2) and affinity ($k_d$ 3.5 μM) as observed with full-length CrkL (Carabias, 2019). Introduction of the activating mutation Y554H in C3G-AAPA slightly increased the affinity of CrkL-SH3N binding ($k_d$ 0.8 μM). However, the main effect of Y554H was observed in the competent fraction of the binding that increased to ~0.5 molecules of SH3N per molecule of C3G-AAPA-Y554H. In contrast, phosphorylation of C3G-AAPA with Src did not increase the fraction of active binding sites (N ~0.1) nor the affinity ($k_d$ 2.1 μM) of CrkL-SH3N with respect to the unphosphorylated protein (Figure R2B and Table R1).

CrkL-SH3N bound to the P4 site in both the constitutively active mutant C3G-AAPA-Y554H and the phosphorylated C3G-AAAP with similar stoichiometry and affinity (N 0.6-0.7 and $k_d$ 2.4-2.7 μM) as we had observed for the binding to unphosphorylated (i.e., inactive) C3G-AAAP (Figure R2C and Table R1). The interaction was also similar to the binding observed with the full-length CrkL protein to C3G-AAAP (Carabias, 2019).

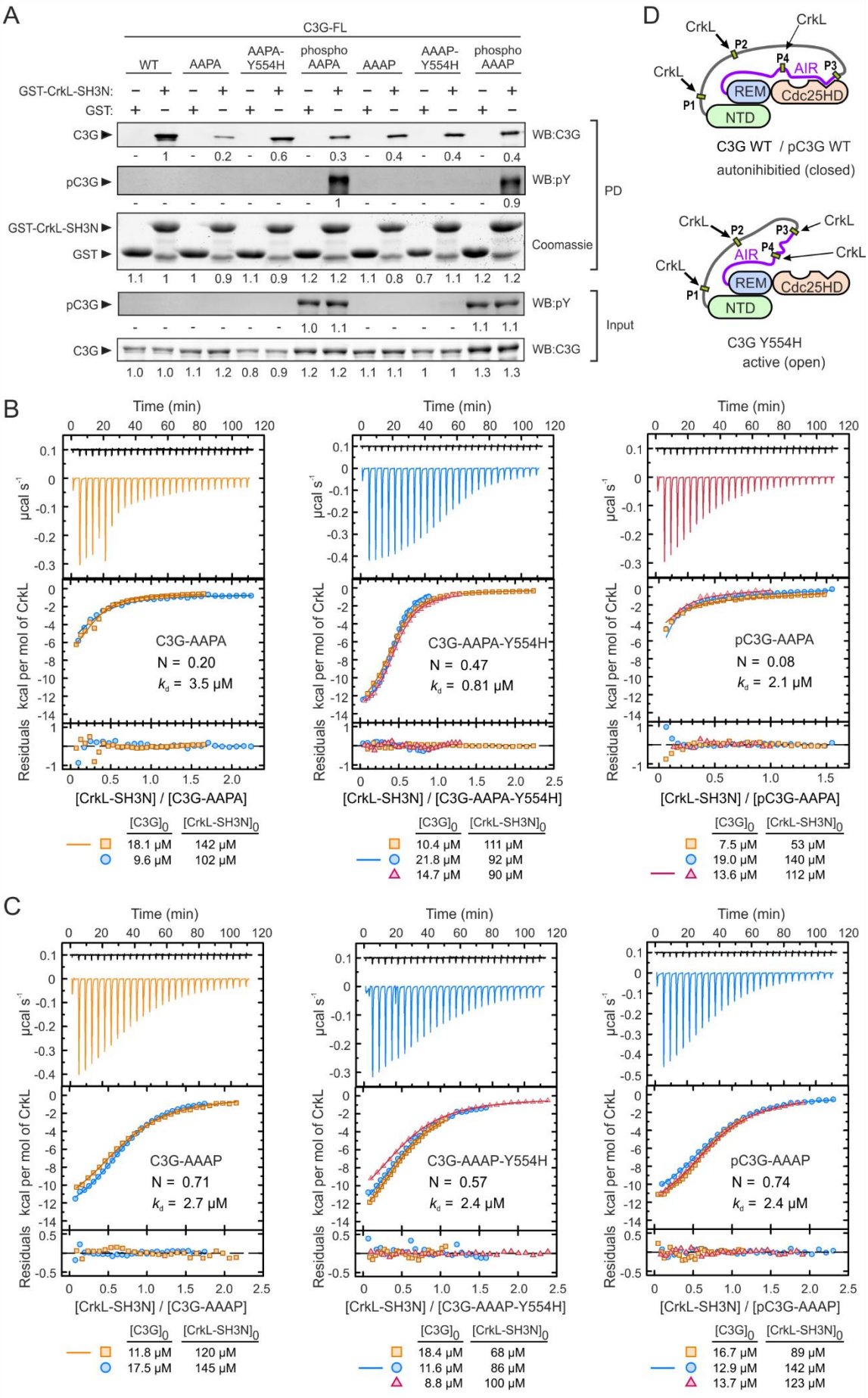

A
C3G-FL
WT
AAPA
AAPA-Y554H
phospho AAPA
AAAP
AAAP-Y554H
phospho AAAP
GST-CrkL-SH3N:
GST:
C3G
pC3G
GST-CrkL-SH3N
GST
WB:C3G
WB:pY
Coomassie
PD
Input
D
CrkL
P1
P2
P3
P4
AIR
REM
Cdc25HD
NTD
C3G WT / pC3G WT
autonihibitied (closed)
C3G Y554H
active (open)
B
Time (min)
µcal s-1
kcal per mol of CrkL
Residuals
C3G-AAPA
N = 0.20
kd = 3.5 µM
[CrkL-SH3N] / [C3G-AAPA]
[C3G]0 [CrkL-SH3N]0
18.1 µM 142 µM
9.6 µM 102 µM
C3G-AAPA-Y554H
N = 0.47
kd = 0.81 µM
[CrkL-SH3N] / [C3G-AAPA-Y554H]
10.4 µM 111 µM
21.8 µM 92 µM
14.7 µM 90 µM
pC3G-AAPA
N = 0.08
kd = 2.1 µM
[CrkL-SH3N] / [pC3G-AAPA]
7.5 µM 53 µM
19.0 µM 140 µM
13.6 µM 112 µM
C
C3G-AAAP
N = 0.71
kd = 2.7 µM
[CrkL-SH3N] / [C3G-AAAP]
11.8 µM 120 µM
17.5 µM 145 µM
C3G-AAAP-Y554H
N = 0.57
kd = 2.4 µM
[CrkL-SH3N] / [C3G-AAAP-Y554H]
18.4 µM 68 µM
11.6 µM 86 µM
8.8 µM 100 µM
pC3G-AAAP
N = 0.74
kd = 2.4 µM
[CrkL-SH3N] / [pC3G-AAAP]
16.7 µM 89 µM
12.9 µM 142 µM
13.7 µM 123 µM

**Figure R2. Accessibility of CrkL to individual P3 and P4 sites in different states of C3G. (A)** PD analysis of the binding of GST-CrkL-SH3N to C3G wild type and single-PRM mutants, in combination with the activating mutation Y554H or after phosphorylation by Src. C3G and phosphorylation of C3G was detected in the PD and input samples by Western blot (WB) using specific anti-C3G and anti-phospho-tyrosine (pY) mABs. GST-CrkL-SH3N and GST (used as control) were visualized by Coomassie staining. Values are the relative quantification within each panel. **(B)** ITC experiments of the binding of CrkL-SH3N domain to C3G-AAPA, C3G-AAPA-Y554H, and Src-phosphorylated pC3G-AAPA. ITC measurements were performed at 25 ºC. In each panel, the upper graph is a representative thermogram (colors correspond to independent titrations). The black line represents the heats of dilution of CrkL-SH3N domain in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl. The middle graphs represent the binding isotherms. A one binding site model was fitted to the data; fitted models are shown as a dashed line. The lower graph shows the residuals from the fitted model. **(C)** ITC experiments of the binding of CrkL-SH3N domain to C3G-AAAP, C3G-AAAP-Y554H, and Src-phosphorylated pC3G-AAAP. Each panel contains the same distribution of the contents as in B. **(D)** Schematic models of the structure of autoinhibited C3G WT and the constitutive active mutant C3G-Y554H and the accessibility of CrkL to the PRMs. The P3 is exposed in the active conformation.

**Table R1. Thermodynamic parameters for the binding of SH3N domain of CrkL to C3G mutants.**

| C3G mutant | Num exp[a] | N[b] | $k_d$ (μM) | ΔH (kcal/mol)[d] |
|---|---|---|---|---|
| C3G-AAPA | 2 | **0.20** [0.17-0.24][c] | **3.5** [2.7-4.6][c] | -14.0[e] |
| C3G-AAPA-Y554H | 3 | **0.47** [0.46-0.48][c] | **0.8** [0.7-0.9][c] | -14.1 |
| pC3G-AAPA | 3 | **0.08** [0.06-0.11][c] | **2.1** [1.3-3.3][c] | -14.0[e] |
| C3G-AAAP | 2 | **0.71** [0.70-0.73][c] | **2.7** [2.4-3.1][c] | -14.1 |
| C3G-AAAP-Y554H | 3 | **0.57** [0.55-0.58][c] | **2.4** [2.1-2.7][c] | -14.7 |
| pC3G-AAAP | 3 | **0.74** [0.73-0.74][c] | **2.4** [2.2-2.7][c] | -13.2 |

[a] Number of independent experiments used for the thermodynamic analysis.
[b] N refers to stoichiometry.
[c] Data in brackets represent the asymmetric errors within a 0.95 confidence interval.
[d] ΔH is the binding enthalpy.
[e] Values were fixed for analysis.

In conclusion, these results suggest that the P1 and P2 sites are constitutively accessible to CrkL in autoinhibited C3G. The P4 site is partially exposed, and its accessibility to CrkL is similar in the autoinhibited, Src-phosphorylated (partially active) and active states. Finally, the SH3N domain of CrkL accessibility to the P3 site depends on the activation state of C3G; in autoinhibited and phosphorylated (partially active) states the P3 site is mostly occluded, and only becomes exposed upon disruption of the AIR/Cdc25HD autoinhibitory interaction (Figure R2D).

### 1.2. Role of the P3 and P4 sites in the activation of C3G by CrkL

To assess the role of the binding of CrkL to the different PRMs on the regulation of the GEF activity of full-length C3G over Rap1b GTPase, we analyzed mutants containing only one wild type PRM or various combinations of unmodified PRMs. The GEF activity of C3G was measured *in vitro* using a fluorescence assay based on the release of mant-dGDP, which is a fluorescent analogue of GDP, bound to Rap1b.

These quantitative assays allow the determination of the nucleotide dissociation rate constant ($k_{obs}$), which is a measurement of the GEF activity. The GEF activity of C3G (1 μM), wild type and mutants, was measured alone and in the presence of an excess of CrkL (40 μM) (Figure R3A-C). The basal activity level of the mutants was similar to that of the autoinhibited wild type C3G, suggesting that the PRMs are not involved in the autoinhibition. CrkL did not activate C3G-PPAA above the basal level; in contrast, CrkL activated C3G-AAPP to a similar level as wild type C3G. C3G-AAPA and C3G-PPPA were also activated by CrkL but to levels significantly lower than wild type and AAPP mutant C3G. CrkL did not activate C3G-AAAP. Finally, disruption of the P3 motif alone, C3G-PPAP, was sufficient to prevent the activation by CrkL.

Next, to gain a better description of the activation, we performed CrkL dose-response experiments to analyze the dependence of the activity of C3G wild type and the mutants AAPP and AAPA (Figure R3D and Table R2). These experiments allowed the estimation of the parameters $k_{free}$, $k_{max}$ and $AC_{50}$. The $k_{max}$ measures the maximum activity that CrkL can induce, and the $AC_{50}$ indicates at which concentration of CrkL occurs the midpoint activation. CrkL activated similarly C3G wild type and C3G-AAPP mutant, with $AC_{50}$ ~23 μM and $k_{max}$ ~13.5 x $10^{-3}$ $s^{-1}$, corresponding to an increase between 5.7 and 7.3 times the basal activity, as measured by the $k_{max}/k_{free}$ ratio (Table R2). CrkL also activated C3G-AAPA with a similar $AC_{50}$ of ~28 μM, although it induced a lower $k_{max}$ of ~9 x $10^{-3}$ $s^{-1}$, corresponding to a ~4.3-fold activation. Together, these results showed that the P1 and P2 do not participate in the direct activation of C3G by CrkL; binding of CrkL to the P3 is sufficient and necessary for activation, but efficient stimulation also requires the interaction of CrkL with the P4.

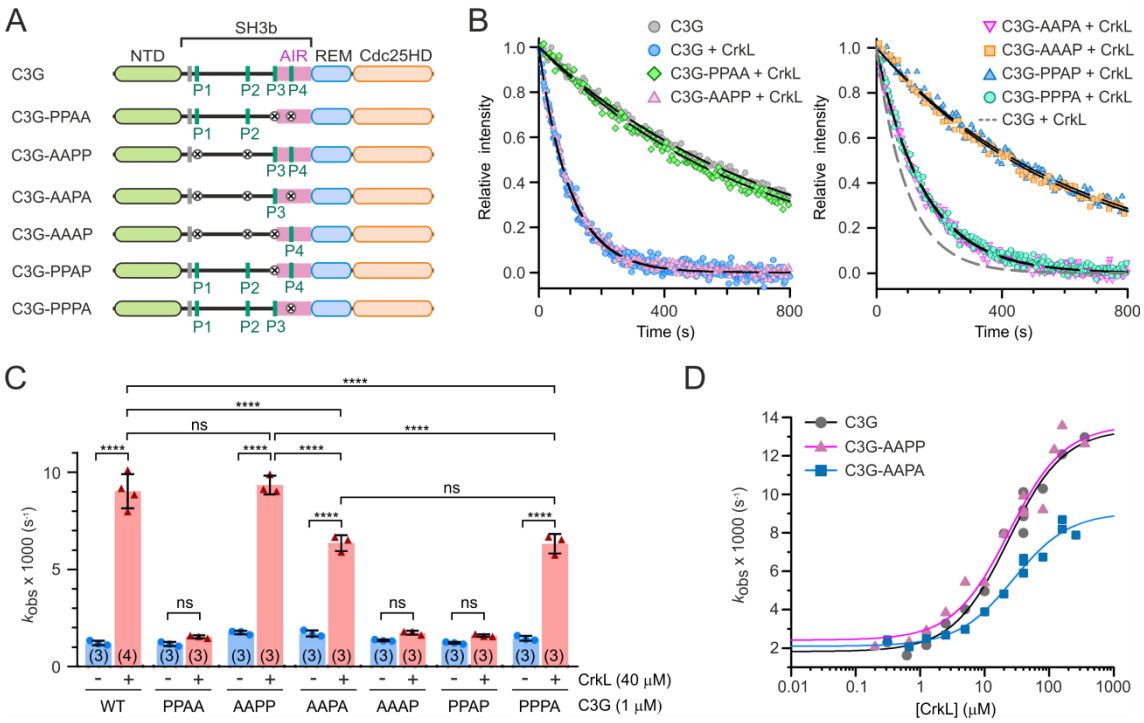



**Figure R3. Contribution of individual PRMs to the activation of C3G by CrkL. (A)** Schematic representation of the PRM mutants used to study the activation of C3G by CrkL. **(B)** Representative exchange reactions of Rap1:mant-dGDP (200 nM) catalyzed by C3G wild type and mutants (1 µM) in the presence of CrkL (40 µM). Lines are the single exponential decay models fitted to obtain the $k_{obs}$. **(C)** Nucleotide exchange rates of C3G wild type and mutants (1 µM) alone and in the presence of 40 µM CrkL. Data are shown as scatter plots with bars, means ± standard deviation. The number of independent measurements is indicated in parentheses. Statistical comparison was analyzed using ANOVA followed by Tukey´s multiple comparisons test; *** $P < 0.001$, **** $P < 0.0001$, ns $P > 0.05$. **(D)** Dose-dependent effect of CrkL on the GEF activity of C3G (1 µM) wild type and C3G-AAPP and C3G-AAPA mutants. Lines are fitted sigmoidal models.

**Table R2. Parameters of dose-response activation of C3G PRM mutants by CrkL.**

| C3G sample | $AC_{50}$ (µM)[a] | $k_{free}$ x1000 ($s^{-1}$) | $k_{max}$ x1000 ($s^{-1}$) | $k_{max}/k_{free}$ |
|---|---|---|---|---|
| C3G | 22.4 ± 4.5 | 1.7 ± 0.4 | 13.4 ± 0.6 | 8.1 ± 2.1 |
| C3G-AAPP | 23.7 ± 6.6 | 2.4 ± 0.5 | 13.6 ± 0.8 | 5.7 ± 1.2 |
| C3G-AAPA | 28.1 ± 5.7 | 2.1 ± 0.2 | 9.1 ± 0.4 | 4.3 ± 0.5 |

[a] Data are presented as the fitted value ± the associated standard errors of the nonlinear regression fit.

During activation, C3G is phosphorylated in tyrosine residues by Src kinase, and tyrosine phosphorylation of C3G and the binding with CrkL are additive stimuli in the activation of C3G. Therefore, we analyzed the role of the binding of CrkL to the different PRMs in the activation of phosphorylated C3G using wild type and four PRM mutants (Figure R4A).

Similarly to the previous analysis of unphosphorylated C3G, the GEF activity of phospho-C3G (hereafter pC3G) was measured alone and in the presence of an excess of CrkL (5 μM). These experiments were done at 0.2 μM pC3G to work below the $AC_{50}$ of CrkL (see below). In this sense, the activities of pC3G in the range of concentration between 0.1 to 1 μM alone and in the presence of an excess of CrkL were directly proportional to the concentration of C3G (Figure R4B). For comparison with the activity of the unphosphorylated C3G mutants, data of pC3G are presented as the specific activity ($k_{obs}$) per 1 μM pC3G.

Phosphorylation induced a slight ~2-fold increase in the GEF activity of C3G. The addition of CrkL caused an additional ~9-fold activation (Figure R4C). CrkL stimulated pC3G-AAPP to a similar level as wild type pC3G. The activity of the C3G mutant that conserved only the wild type P3, pC3G-AAPA, was also stimulated by CrkL, but to a significantly lower level than that observed for pC3G wild type and pC3G-AAPP. Finally, CrkL did not increase the basal activity of the pC3G-PPAA and pC3G-AAAP mutants.

We also studied the dependence of the GEF activity of pC3G, wild type and AAPP and AAPA mutants, with the concentration of CrkL (Figure R4D-E and Table R3). Wild type pC3G was stimulated at a significantly lower concentration of CrkL than the unphosphorylated protein, as shown by a ~10-fold reduction of the $AC_{50}$ upon phosphorylation. CrkL also induced a higher maximal activity ($k_{max}$) on pC3G, which was ~2.6 times higher than the $k_{max}$ induced on unphosphorylated C3G. The pC3G-AAPP mutant was activated with similar $AC_{50}$ and $k_{max}$ values as the wild type pC3G. CrkL also activated pC3G-AAPA with a similar $AC_{50}$; yet the maximum activity induced by CrkL ($k_{max}$ 27 x $10^{-3}$ $s^{-1}$) was ~25% lower than for the wild type and the AAPP mutant. Finally, the mutant pC3G-AAAP was marginally stimulated by CrkL.

**Table R3. Parameters of dose-response activation of pC3G PRM mutants by CrkL.**

| C3G mutant | $AC_{50}$ (μM)[a] | $k_{free}$ x1000 ($s^{-1}$) | $k_{max}$ x1000 ($s^{-1}$) | $k_{max}/k_{free}$ |
|---|---|---|---|---|
| pC3G | 1.7 ± 0.2 | 2.3 ± 0.6 | 34.2 ± 1.0 | 15.0 ± 4.0 |
| pC3G-AAPP | 2.2 ± 0.5 | 3.3 ± 1.0 | 35.7 ± 2.0 | 10.9 ± 3.3 |
| pC3G-AAPA | 1.3 ± 0.3 | 6.2 ± 0.9 | 26.9 ± 0.8 | 4.3 ± 0.6 |
| pC3G-AAAP | 0.6 ± 0.6 | 2.9 ± 0.4 | 4.5 ± 0.2 | 1.5 ± 0.2 |

[a] Data are presented as the fitted value ± the associated standard errors of the nonlinear regression fit.

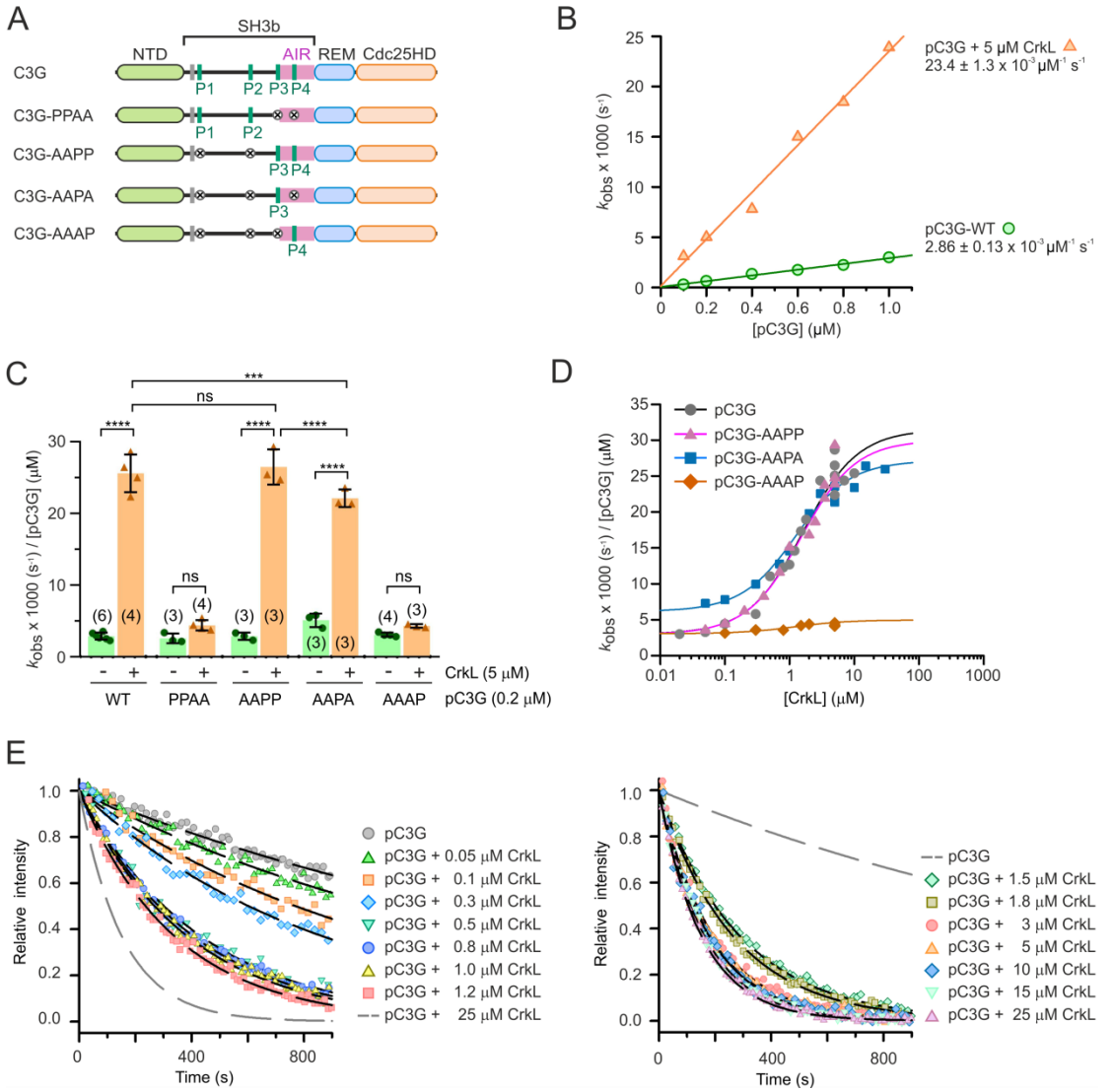



**Figure R4. Contribution of individual PRMs to the activation of Src-phosphorylated C3G by CrkL.** **(A)** Schematic representation of the PRM mutants used to study the activation of pC3G by CrkL. **(B)** Dependence of the GEF activity of pC3G on its concentration in absence and presence of CrkL (5 µM). Lines represent simple linear regression fitted to obtain the indicated values of the specific activities. **(C)** Nucleotide exchange rates of pC3G wild type and mutants (0.2 µM; data are referred to 1 µM C3G) alone and in the presence of 5 µM CrkL. Data are shown as scatter plots with bars, means ± standard deviation. The number of independent measurements is indicated in parentheses. Statistical comparison was analyzed using ANOVA followed by Tukey´s multiple comparisons test; *** $P < 0.001$, **** $P < 0.0001$, ns $P > 0.05$. **(D)** Dose-dependent effect of CrkL on the GEF activity of pC3G (0.2 µM) wild type and mutants. Data were referred as in C. **(E)** Representative exchange reactions of Rap1:mant-dGDP (200 nM) catalyzed by pC3G wild type (0.2 µM) in the presence of different concentrations of CrkL. Lines are the single exponential decay models fitted to obtain the $k_{obs}$.

Collectively, all these data revealed that only the P3 and P4 motifs are required and sufficient for activation of pC3G by CrkL, with P3 being the main activation site.

Noteworthy, phosphorylation of C3G by Src increases both the sensitivity to CrkL-mediated direct activation and the maximal exchange activity induced by CrkL.

### 1.3. The SH2 and SH3N domain of CrkL contribute to the activation of C3G

In addition to the SH3N, CrkL has two other domains, an N-terminal SH2 domain and a second C-terminal SH3 domain (SH3C) that does not bind to PRMs. The interaction of CrkL-SH3N with the PRMs P1 to P4 of C3G is well characterized. However, the role of the SH2 and SH3C domains in the activation was not well understood. Previous data from our group indicated that different CrkL fragments activated unphosphorylated C3G in a different manner (Carabias, 2019). This prompted us to analyze in detail the contribution of regions of CrkL outside the SH3N to the activation of C3G. To that end, the GEF activity of C3G was titrated with several constructs of CrkL (Figure R5A). Initially, we measured the activation of C3G by the isolated SH3N domain, which induced a maximum activity ($k_{max}$ 6.3 x $10^{-3}$ $s^{-1}$) that was half of that caused by full-length CrkL (Figure R5B and Table R4). This suggested that the SH2 and SH3C domains also contribute to the activation of C3G.

Next, we created two constructs of CrkL that lacked either the SH2 or the SH3C domain. The maximum GEF activity induced by the fragment SH2-SH3N ($k_{max}$ 8.4 x $10^{-3}$ $s^{-1}$) was slightly higher than that of the SH3N domain, but it was lower than the activation with full-length CrkL. A fragment containing the two SH3 domains, SH3N-SH3C, activated C3G to the same activity level as the full-length protein. Finally, the four constructs of CrkL analyzed showed comparable $AC_{50}$ values (16-27 μM) for the activation of the GEF activity. In summary, regions outside the SH3N domain, mainly the SH3C and the linker connecting the two SH3 domains, contribute to the efficacy of the activation of unphosphorylated C3G.

We also analyzed the role of the different domains of CrkL in the activation of phosphorylated C3G (Figure R5C and Table R4). The isolated SH3N domain activated pC3G up to a maximum activity ($k_{max}$ 13 x $10^{-3}$ $s^{-1}$ $\mu M^{-1}$) that was ~40% of the maximum activity induced by full-length CrkL. The fragment SH3N-SH3C also activated pC3G to similar maximum activity as the isolated SH3N domain. But the $AC_{50}$ of the fragment SH3N-SH3C (~4.5 μM) was lower than the SH3N alone $AC_{50}$ (~13 μM). In contrast, the fragment SH2-SH3N activated pC3G to a similar maximum activity as wild type C3G ($k_{max}$ 37.7 x $10^{-3}$ $s^{-1}$ $\mu M^{-1}$). Nonetheless, the $AC_{50}$ of the SH2-SH3N (4.9 μM) was higher

than full-length CrkL ($AC_{50}$ of 1.7 µM), suggesting that the SH3C domain of CrkL also participates in the efficient activation of phospho-C3G.

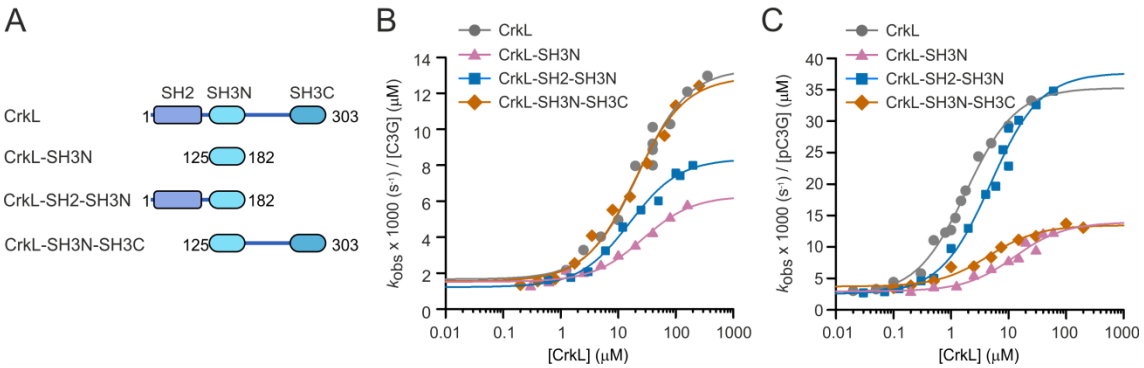



**Figure R5. Contribution of CrkL domains to the activation of C3G. (A)** Schematic representation of the domain structure of CrkL and the deletion mutants analyzed. **(B)** Analysis of the dose-dependent activation of C3G (1 µM) by CrkL full-length and the indicated fragments. **(C)** Dose-dependent activation of pC3G (0.2 µM, data are referred to 1 µM pC3G) by CrkL and its deletion mutants. Lines in B and C represent the fitted sigmoidal activation models.

**Table R4. Parameters of dose-response activation of C3G by CrkL proteins.**

| C3G | CrkL protein | $AC_{50}$ (µM)[a] | $k_{free}$ x1000 ($s^{-1}$) | $k_{max}$ x1000 ($s^{-1}$) | $k_{max}/k_{free}$ |
|---|---|---|---|---|---|
| C3G[b] | CrkL[b] | 22.4 ± 4.5 | 1.7 ± 0.4 | 13.4 ± 0.6 | 8.1 ± 2.1 |
| C3G | CrkL-SH3N | 26.8 ± 6.5 | 1.5 ± 0.1 | 6.3 ± 0.4 | 4.2 ± 0.4 |
| C3G | CrkL-SH2-SH3N | 15.8 ± 2.6 | 1.2 ± 0.2 | 8.4 ± 0.2 | 6.9 ± 1.3 |
| C3G | CrkL-SH3N-SH3C | 19.8 ± 3.9 | 1.6 ± 0.3 | 12.9 ± 0.6 | 8.3 ± 1.7 |
| pC3G[c] | CrkL[c] | 1.7 ± 0.2 | 2.3 ± 0.6 | 34.2 ± 1.0 | 15.0 ± 4.0 |
| pC3G | CrkL-SH3N | 12.8 ± 4.6 | 2.9 ± 0.5 | 13.9 ± 1.2 | 4.8 ± 0.9 |
| pC3G | CrkL-SH2-SH3N | 4.9 ± 0.8 | 2.5 ± 0.8 | 37.7 ± 1.5 | 15.0 ± 4.6 |
| pC3G | CrkL-SH3N-SH3C | 4.5 ± 1.0 | 3.7 ± 0.4 | 13.4 ± 0.4 | 3.6 ± 0.4 |

[a] Data are presented as the fitted value ± the associated standard errors of the nonlinear regression fit.
[b] Data are taken from the previous Table R3.
[c] Data are taken from the previous Table R4.

The SH2 and SH3C domains of CrkL might contribute to the activation of C3G through additional interactions with C3G. The CrkL-SH3C domain does not bind to PRMs, while the CrkL-SH2 domain binds to phosphorylated tyrosine residues in the context of pYXXP motifs (Birge et al., 1993, Sakai et al., 1994). This prompted us to analyze if the SH2 domain could bind to Src-phosphorylated C3G.

To inactivate the binding site of the SH3N domain of CrkL we created three point mutants based on the structure of the complex between SH3N domain of CrkII and the sequence peptide of the P1 of C3G (Figure R6A; PDB: 1CKA; (Wu et al., 1995)). Crk residues W169, P183 and Y186 are part of the binding site, and mutation of W169 and

Y186 disrupts the binding of the SH3N domain to PRMs (Tanaka et al., 1993, Yuwen et al., 2016). These residues of Crk correspond to W160, P174 and Y177 in CrkL; thus, we created the point mutants W160S, P174R and Y177S of CrkL full-length.

Next, we analyzed the effect of these three mutations on the binding of CrkL to the four peptides of the PRMs P1, P2, P3 and P4 of C3G using the aforementioned fluorescence anisotropy assay (Figure R6B-D). The three mutations W160S, P174R and Y177S drastically reduced the affinity for the PRMs of C3G. In addition, the effect of the mutations was analyzed using pull-down assays with the full-length C3G fused to GST (GST-C3G). GST-C3G efficiently pulled down wild type CrkL. In contrast, CrkL-W160S and CrkL-P174R were not pulled down by C3G, yet a small fraction of CrkL-Y177S bound to GST-C3G (Figure R6E). Collectively, point substitution W160S and P174R in CrkL disrupted the interaction through the SH3N domain with the PRMs of C3G (Figure R6F). To simplify the study, hereafter only the mutant CrkL-W160S was used to analyze the impact of the inactivation of the SH3N domain.

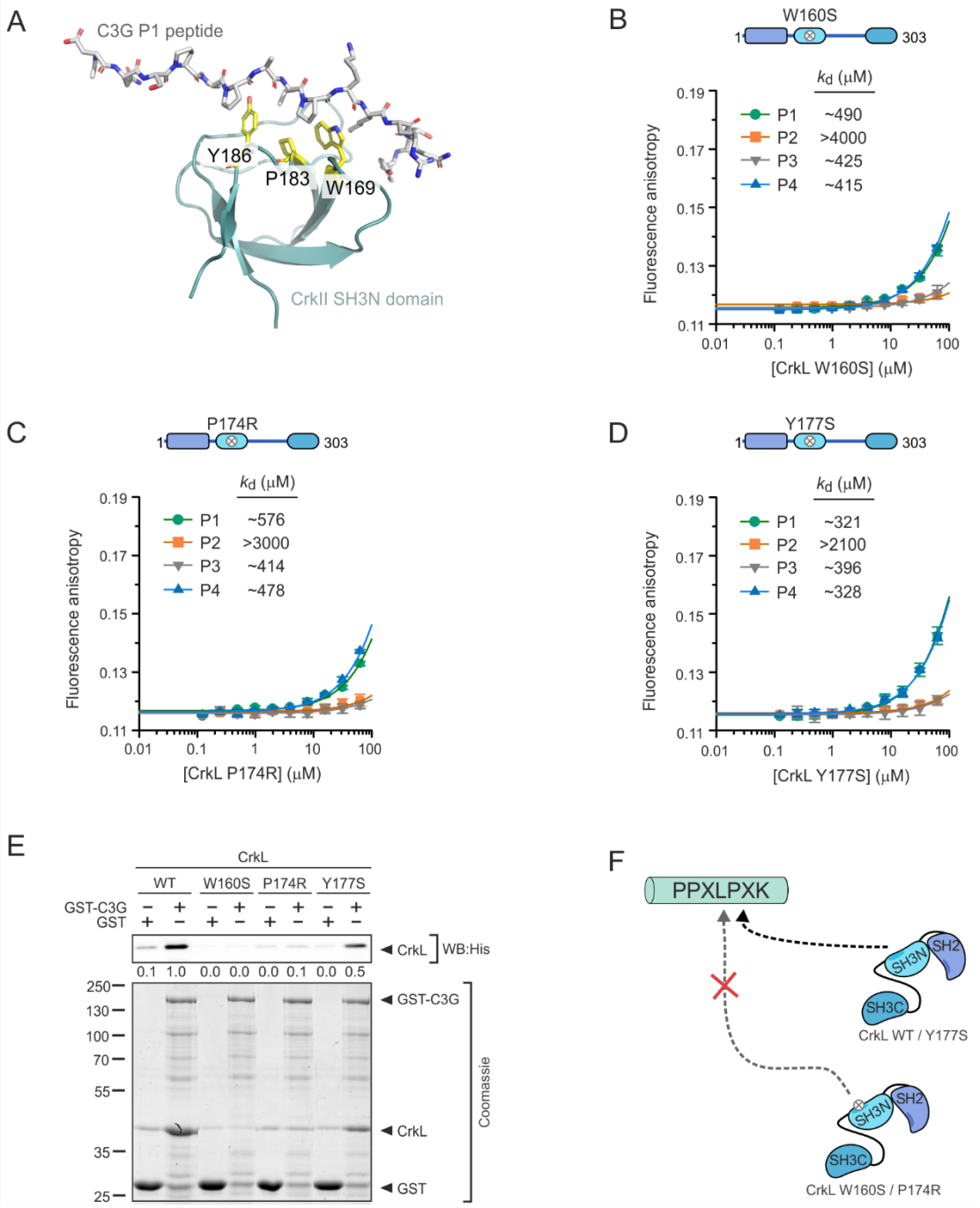



**Figure R6. Design and characterization of CrkL point mutations in the SH3N domain that disrupt binding to the PRMs of C3G. (A)** Structure of the SH3N domain of Crk (shown as ribbons) bound to a peptide of the P1 of C3G (shown as sticks) (PDB: 1CKA). The side chains of Crk residues W169, P183 and Y186 are shown as sticks. These residues form pockets that accommodate P residues of the PRM. **(B-D)** Fluorescence anisotropy titrations of the fluorescein-labeled PRMs peptides of C3G (0.2 μM) with full-length CrkL mutants W160S (B), P174R (C), and Y177S (D). Lines represent the 1:1 binding model fitted to the data, which yielded the apparent dissociation constants ($k_d$). **(E)** Analysis by pull-down (PD) of the interaction of GST-C3G full-length with CrkL wild type (WT) and the SH3N domain mutants. CrkL was detected in the PD by Western blot (WB) using an anti-His tag antibody. GST-C3G and GST (used as control) were visualized by Coomassie staining. Numbers are the relative quantification within each panel.

**(F)** Representation of the binding effect of CrkL-SH3N domain mutants and WT in the binding to the PRMs of C3G.

To design a mutant that disrupts the binding site in the SH2 domain of CrkL, we analyzed the structure of the complex between SH2 domain of CrkII and a peptide of the linker between SH3N-SH3C of CrkII that included the phosphorylated Y221 responsible of an intramolecular interaction (Figure R7A) ((PDB: 1JU5; (Donaldson et al., 2002)). Residue R38 of CrkII, which corresponds to R39 in CrkL, makes a salt bridge with the phosphate moiety and mutation of R38 in CrkII disrupts the phospho-Y binding site, and the mutation Crk-I-R38V has been used to abolish binding to phospho-Y-containing proteins (Tanaka et al., 1993). Therefore, to disrupt the binding site of the SH2 domain of CrkL we created the R39K point mutant.

Next, we used the CrkL SH2 and SH3N mutants, to analyze the contribution of each domain to the interactions and activation of full-length C3G. We used pull-down assays with full-length CrkL fused to GST (Figure R7B-C) and the different mutants to analyze the interaction with C3G full-length. Wild type CrkL pulled down C3G and pC3G with similar efficiency. GST-CrkL-W160S did not bind to unphosphorylated C3G but pulled down a small fraction of pC3G. The double mutant carrying R39K in combination with W160S did not interact with either unphosphorylated C3G or pC3G. When CrkL carried only the mutation R39K, the interaction with C3G and pC3G was remarkably similar to that of the wild type GST-CrkL. Collectively, these data support the notion that the SH2 domain of CrkL binds to pC3G.

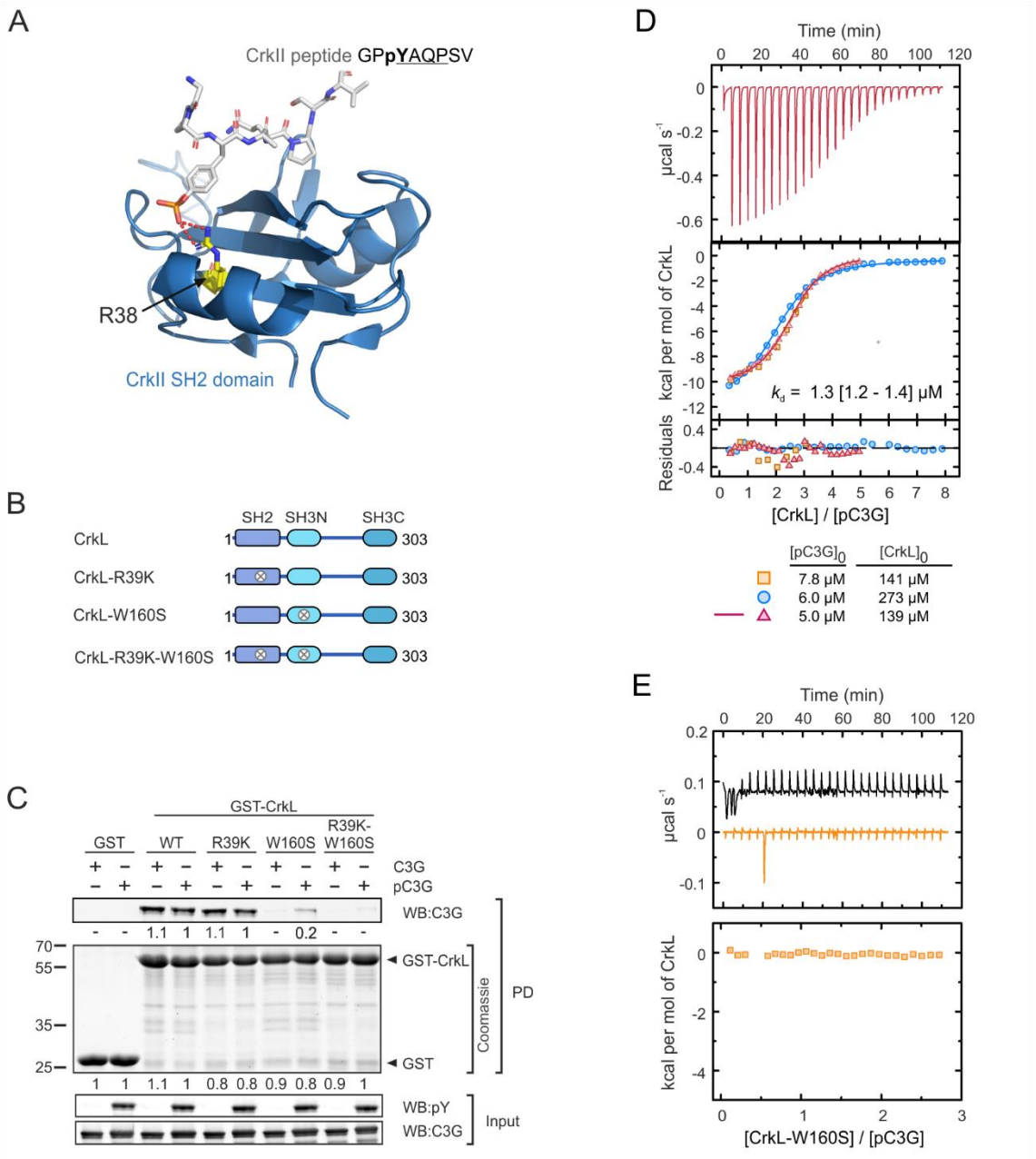



**Figure R7. Identification of a low-affinity interaction between the SH2 domain of CrkL and pC3G. (A)** Structural representation of the binding interface between the phosphorylated Y221 peptide sequence, represented as sticks, and the SH2 domain of CrkII, visualized as cartoon model. CrkII R38 residue is highlighted in yellow and as stick (modified of PDB: 1JU5). **(B)** Schematic representation of CrkL and the point mutants used to analyze the binding of CrkL to C3G and pC3G. **(C)** Pull-down (PD) analysis of the binding of GST-CrkL, wild type and mutants, to C3G and pC3G. C3G was detected in the PD and in the input samples by Western blot (WB). pC3G was detected with an antibody that recognizes phospho-Y (pY). Values are the relative quantification of the bands. **(D)** ITC analysis of CrkL binding to pC3G, the upper panel shows a representative thermogram of CrkL binding to pC3G, corresponding with the data of the same color in the middle panel. The middle panel show the binding isotherms of three independent titrations analyzed simultaneously. Residuals from the fitted model are displayed in the lower panel. **(E)** ITC analysis of the titration of pC3G with the SH3N-inactive mutant CrkL-W160S. The black line (upper trace) is heat exchanges of the dilution of CrkL-W160S in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl. The

orange line (lower trace) is the thermogram of the titration of pC3G with CrkL-W160S. No signal associated with binding was observed. The lower panel shows the integrated isotherms of the titration.

Next, we used ITC to quantitatively analyze the interaction of CrkL with pC3G by an independent method (Figure R7D and Table R5). CrkL bound to pC3G with similar affinity ($k_d$ 1.3 μM) as observed in the interaction with unphosphorylated C3G (Carabias, 2019). We also analyzed the binding of the SH3N-inactive mutant CrkL-W160S to pC3G. We used similar conditions to measure the binding of wild type CrkL to C3G and pC3G (Figure R7E and Table R5); yet no heat exchange signal was detected. These data suggested that the CrkL-SH2 domain interacts with low affinity with pC3G.

**Table R5. Thermodynamic parameters of the pC3G/CrkL or C3G/CrkL-W160S binding.**

| CrkL | Num exp | [pC3G] (μM)[a] | [CrkL] (μM)[b] | $k_d$ (μM) | ΔH (kcal/mol)[d] |
|---|---|---|---|---|---|
| CrkL-FL | 1 | 7.8 | 141.2 | **1.3** [1.2-1.4][c] | -11.5 |
| | 2 | 6.0 | 273.4 | | |
| | 3 | 5.0 | 187.2 | | |
| CrkL-W160S | 1 | 10.8 | 139.3 | - | - |

[a,b] Initial concentrations of C3G and CrkL in the reaction cell and in the syringe, respectively.
[c] Data in brackets represent the asymmetric errors within a 0.95 confidence interval.
[d] ΔH is the binding enthalpy.

In summary, our data revealed that in addition to the SH3N, the SH2 and SH3C domains of CrkL contribute to the stimulation of the GEF activity of C3G, and that the SH2 domain plays a key role in the activation of tyrosine-phosphorylated C3G through a weak interaction with pC3G.

### 1.4. Different activation of C3G by CrkII and CrkL

So far, we analyzed the activation of C3G by CrkL. CrkII, the other member of the Crk family of adaptor proteins, also stimulates the GEF activity of C3G *in vitro* (Popovich, 2013) and *in vivo* (Ichiba et al., 1999). This prompted us to analyze in a comparative manner the activation of C3G by CrkII and CrkL. First, we titrated the GEF activity of unphosphorylated C3G with increasing concentrations of CrkII (Figure R8 and Table R6). Compared to CrkL, CrkII was a weaker activator of C3G. The maximum GEF activity of C3G induced by CrkII ($k_{max}$ 7.2 x $10^{-3}$ $s^{-1}$) was about half of that induced by CrkL. In addition, the half-maximum activity of C3G was reached at three-fold higher concentration of CrkII ($AC_{50}$ 66 μM) than CrkL. Previous results from our group,

showed that CrkII binds to C3G with similar affinity (microscopic $k_d$ 2.7 µM, measured by ITC) as CrkL (Carabias, 2019).

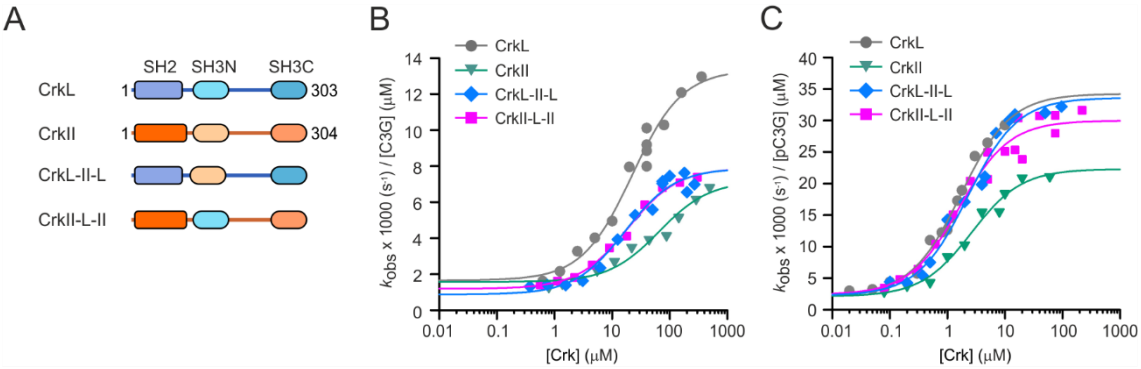



**Figure R8. Activation of C3G by CrkL, CrkII, and their constituent domains**. **(A)** Schematic representation of the domain structure of CrkII and the CrkL-CrkII chimeric proteins analyzed. **(B)** Dose-dependent analysis of the activation of C3G (1 µM) by CrkII and the CrkII-CrkL chimeric proteins. **(C)** Dose-dependent analysis of the activation of pC3G (0.2 µM, data are referred to 1 µM C3G) by CrkII and the CrkII-CrkL chimeric proteins. Rate constants in B and C are shown as the specific activities of 1 µM C3G for comparison. Lines in B and C are the fitted sigmoidal activation models. Activation curves by CrkL are the ones shown in Figure R5 and are included here for comparison.

We have shown that the SH2 and SH3C domains of CrkL contribute to the activation of C3G. This prompted us to analyze if the differences in the activation of C3G by CrkL and CrkII could be related to the contribution of regions outside the SH3N domain. To that end, we created a chimeric protein in which the SH3N domain of CrkL was replaced by the SH3N of CrkII, and the equivalent CrkII chimeric mutant with the SH3N domain of CrkL. To denote if the first (SH2), second (SH3N), and third (SH3C) domains corresponds to CrkII or CrkL, we named these chimeric proteins as CrkL-II-L and CrkII-L-II (Figure R8A).

**Table R6. Parameters of dose-response activation of C3G by CrkII, CrkL and chimeras.**

| C3G | CrkL protein | $AC_{50}$ (µM)[a] | $k_{free}$ x1000 ($s^{-1}$) | $k_{max}$ x1000 ($s^{-1}$) | $k_{max}/k_{free}$ |
|---|---|---|---|---|---|
| C3G[b] | CrkL[b] | 22.4 ± 4.5 | 1.7 ± 0.4 | 13.4 ± 0.6 | 8.1 ± 2.1 |
| C3G | CrkII | 65.7 ± 20.4 | 1.6 ± 0.2 | 7.2 ± 0.5 | 4.6 ± 0.7 |
| C3G | CrkL-II-L | 16.7 ± 4.5 | 0.9 ± 0.3 | 7.9 ± 0.3 | 9.1 ± 3.4 |
| C3G | CrkII-L-II | 18.7 ± 2.4 | 1.2 ± 0.1 | 7.9 ± 0.2 | 6.6 ± 0.8 |
| pC3G[c] | CrkL[c] | 1.7 ± 0.2 | 2.3 ± 0.6 | 34.2 ± 1.0 | 15.0 ± 4.0 |
| pC3G | CrkII | 2.6 ± 0.6 | 2.1 ± 0.8 | 22.3 ± 0.9 | 10.6 ± 4.2 |
| pC3G | CrkL-II-L | 2.3 ± 0.5 | 2.2 ± 1.3 | 33.6 ± 1.2 | 15.0 ± 8.5 |
| pC3G | CrkII-L-II | 1.6 ± 0.4 | 2.4 ± 1.6 | 30.0 ± 0.9 | 12.7 ± 8.5 |

[a] Data are presented as the fitted value ± the associated standard errors of the nonlinear regression fit.
[b] Data are taken from the previous Table R3.
[c] Data are taken from the previous Table R4.

We analyzed the dose-response activation of unphosphorylated C3G by these constructs, which revealed that the two chimeric proteins stimulated C3G in a remarkably similar manner (Figure R8B). Both chimeric proteins showed similar $AC_{50}$ values (~17 µM) as CrkL (~22 µM). In addition, the chimeric proteins induced a maximum GEF activity ($k_{max}$ 7.9 x $10^{-3}$ $s^{-1}$) half of CrkL, being similar to the activation caused by CrkII, and the SH2-SH3N fragment of CrkL.

Next, we studied the activation of Src-phosphorylated C3G by CrkII and the chimeric proteins (Figure R8C). CrkII, CrkL-II-L and CrkII-L-II stimulated pC3G with similar $AC_{50}$ values (1.6 to 2.6 µM). The maximum GEF activities ($k_{max}$ 30 - 34 x $10^{-3}$ $s^{-1}$ $\mu M^{-1}$) induced by both chimeric proteins were similar to that observed with CrkL. In contrast, CrkII was less efficient and stimulated pC3G to ~60% the maximum activity of CrkL.

In conclusion, our results suggest that the reduced capacity of CrkII to activate phosphorylated C3G is not due to differences in its individual SH2, SH3N and SH3C domains, because combinations of these domains with CrkL complementary regions support efficient activation.

### 1.5. CrkL-SH2 domain is necessary for the complete activation of phosphorylated C3G

We have shown that the SH2 domain of CrkL binds, albeit with apparently low affinity, to Src-phosphorylated C3G and deletion of the SH2 domain of CrkL reduces its capacity to stimulate phospho-C3G (see above). To understand the role of each domain in the context of the full-length CrkL, we used point substitutions in residues of the SH2 and SH3N domains that disrupt their binding sites (see above) to study their implication in the activation of C3G.

Initially, we used nucleotide exchange reactions of Rap1b to measure the GEF activity ($k_{obs}$) of unphosphorylated C3G (1 µM) alone and in the presence of an excess of CrkL full-length (5 µM) wild type or point mutants (Figure R9A-B). CrkL wild type activated C3G 3.3 times the basal level; on the other hand, CrkL-W160S mutant, which cannot bind to PRMs of C3G, did not activate C3G above its basal level.

Next, we analyzed the activation of the phosphorylated state of C3G by the CrkL mutants (Figure R9C-D). CrkL wild type increased 10.6 times the basal activity of pC3G. The SH3-inactive CrkL-W160S mutant did not activate pC3G above the basal activity, as observed with unphosphorylated C3G. Finally, the SH2-inactive CrkL-R39K mutant induced a small activation of pC3G that was significantly lower than that of wild

type CrkL. Yet, activity of pC3G induced by CrkL-R39K is higher than the activation of unphosphorylated C3G with CrkL wild type, indicating that phosphorylation also promotes the C3G activation in a CrkL SH2-independent manner.

To gain a better understanding of the role of the SH2 domain of CrkL in the activation of pC3G, we performed dose-response experiments with mutant CrkL-R39K (Figure R9E; Table R7). CrkL-R39K activated pC3G with $AC_{50}$ of ~6 µM and $k_{max}$ of ~20 x $10^{-3}$ $s^{-1}$ $\mu M^{-1}$ (~9-fold increase over the basal activity). This was weaker than the activation of pC3G by wild type CrkL. Yet, CrkL-R39K induced a stronger activation of pC3G than the activation of unphosphorylated C3G by wild type CrkL.

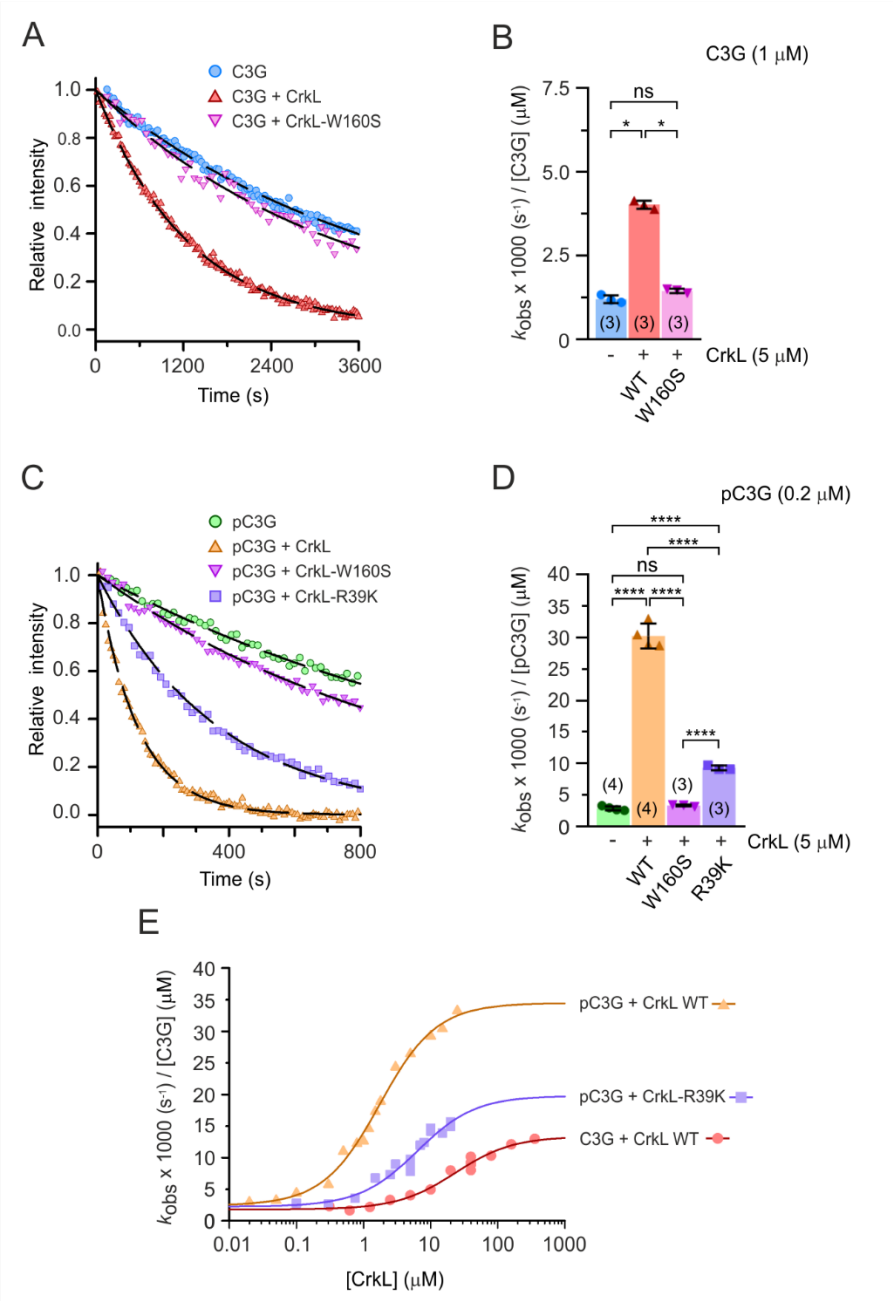


**Figure R9. Activation of C3G by CrkL wild type and SH2- or SH3N-domain mutants. (A)** Representative nucleotide exchange reactions of Rap1:mant-dGDP (200 nM) catalyzed by C3G (1 µM) in the absence and presence of CrkL wild type (WT) and CrkL-W160S mutant (5 µM). Lines are the single exponential decay models fitted to obtain the $k_{obs}$. **(B)** Nucleotide exchange rates of C3G wild type (1 µM) alone and in the presence of 5 µM CrkL wild type and CrkL-W160S mutant. Data are shown as scatter plots with bars, means ± standard deviation. The number of independent measurements is indicated in parentheses. Statistical comparison was analyzed using ANOVA followed by Tukey´s multiple comparisons test; * $P < 0.05$, ns $P > 0.05$. **(C)** Representative exchange reactions of Rap1:mant-dGDP (200 nM) catalyzed by 0.2 µM of Src-phosphorylated C3G (pC3G) in the absence and presence of CrkL wild type and CrkL-R39K or CrkL-W160S mutants (5 µM). Lines are the single exponential decay models fitted to obtain the $k_{obs}$. **(D)** Nucleotide exchange rates of pC3G (0.2 µM; data referred to 1 µM C3G) alone and in the presence of 5 µM CrkL wild type or indicated mutants. Data are shown and the statistical analysis was done as in B. **(E)** Analysis of the dose-dependent activation of pC3G (0.2 µM) by the CrkL-SH2 domain mutant (CrkL-R39K). Lines represent the fitted sigmoidal activation models. Data of the activation of C3G and pC3G by wild type CrkL are the same as presented in Figure R5 and are shown for comparison.

**Table R7. Parameters of dose-response activation of C3G and pC3G by CrkL WT and CrkL-R39K.**

| **C3G** | **CrkL protein** | $AC_{50}$ (µM)[a] | $k_{free}$ x1000 ($s^{-1}$) | $k_{max}$ x1000 ($s^{-1}$) | $k_{max}/k_{free}$ |
|---|---|---|---|---|---|
| C3G[b] | CrkL WT[b] | 22.4 ± 4.5 | 1.7 ± 0.4 | 13.4 ± 0.6 | 8.1 ± 2.1 |
| pC3G | CrkL-R39K | 6.3 ± 1.7 | 2.3 ± 0.7 | 19.8 ± 1.6 | 8.8 ± 2.3 |
| pC3Gc[c] | CrkL WT[c] | 1.7 ± 0.2 | 2.3 ± 0.6 | 34.2 ± 1.0 | 15.0 ± 4.0 |

[a] Data are presented as the fitted value ± the associated standard errors of the nonlinear regression fit.
[b] Data are taken from the previous Table R3.
[c] Data are taken from the previous Table R4.

Collectively, these results support the notion that in addition to the essential interaction of the SH3N domain of CrkL with C3G, the pY binding site of the SH2 domain is also required for the efficient activation of pC3G. This suggests that the CrkL-SH2 domain contributes to the activation through a direct interaction with pC3G. Notwithstanding, the tyrosine-phosphorylation of C3G also plays a minor but noticeable contribution to the activation independently of the functional SH2 domain of CrkL.

## 2. Characterization of Src-phosphorylation in tyrosine residues of C3G

### 2.1. Mapping tyrosine residues in C3G phosphorylated by Src

Phosphorylation of C3G in tyrosine residues by Src plays a significant role during the stimulation of the GEF activity of C3G. While phosphorylation is insufficient for activation, it is necessary for optimal activation of C3G by Crk proteins. Only Y504 of C3G had been identified as a phosphorylation site by Src and other Src-family kinases (Radha et al., 2004); phosphorylation of Y504 had been related to the activation of C3G (Ichiba et al., 1999). The incomplete understanding of the phosphorylation sites in C3G motivated us to carry out a comprehensive analysis of the phosphorylation of C3G by Src.

C3G contains 28 tyrosines (Figure10A, Table R8) that are distributed as follows: (i) 4 tyrosines in the NTD, (ii) 12 tyrosines in the SH3b region, of these 5 tyrosines are located in the AIR, (iii) 4 tyrosines in the REM domain, and (iv) 8 tyrosines in the Cdc25H domain.

**Table R8. Distribution of tyrosine residues of C3G.**

| C3G region | Limits | Tyrosine residues |
|---|---|---|
| NTD | 4-246 | Y61, Y106, Y155, Y159 |
| SH3b | 274-646 | Y329, Y341, Y386, Y390, Y478, Y485<br>Y504, Y554, Y561, Y570, Y579, Y590 |
| REM | 646-810 | Y715, Y732, Y738, Y740 |
| Cdc25HD | 825-1077 | Y857, Y891, Y929, Y957, Y970, Y984, Y1002, Y1030 |

Initially, we analyzed the phosphorylation *in vitro* of C3G by the kinase domain of Src (Src-KD). Four constructs of C3G were generated as GST fusion proteins, C3G full-length (4-1077), NTD (4-245), SH3b (274-646) and REM-Cdc25HD (602-1077); in addition, GST was included as a control (Figure 10A). Proteins were incubated with Src-KD (1 μM) at 30 ºC in phosphorylation buffer. Phosphorylation was detected at several times by Western blot using a pan-anti-pY antibody (Figure 10B). Strong pY signal was detected for C3G full-length and the SH3b region, which seemed to reach maximum levels within 10 min. Weak pY signal was observed for the REM-Cdc25HD. Finally, no phosphorylation was detected either for the NTD or GST alone.

In summary, these results suggest that Src mainly phosphorylates tyrosine residues in the central SH3b region of C3G.

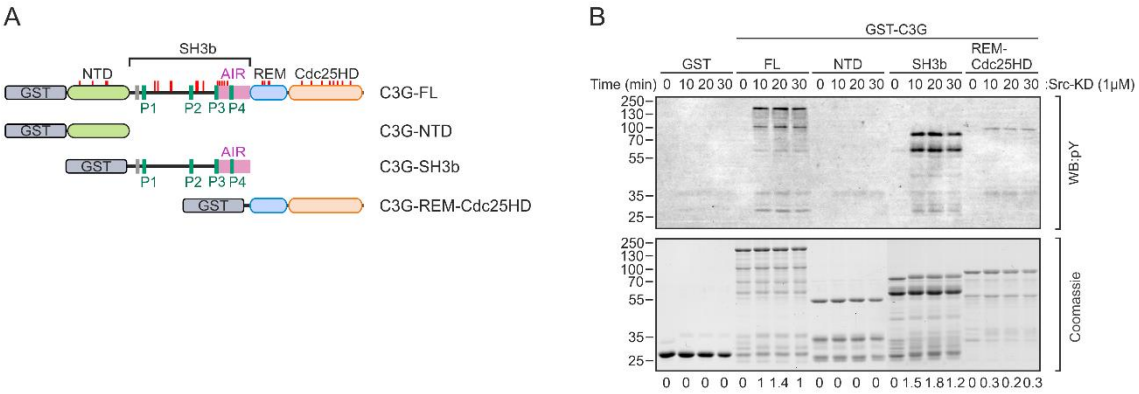



**Figure R10. Analysis of the *in vitro* Src-phosphorylation of C3G by Western blot. (A)** Schematic representation of the GST-fusion constructs used to analyze the phosphorylation of C3G. The position of tyrosine residues (red ticks) is shown in the full-length construct **(B)** Time course of the phosphorylation of the GST-C3G full-length and fragments by Src-KD. Proteins were analyzed and visualized by Coomassie staining; 2 µg of the samples were loaded, for total protein detection. Phosphorylation was detected by western blot with an anti-pY antibody (pY), 0.2 µg of the samples were loaded. Numbers are the relative quantification of the phosphorylated signal bands.

Next, to identify the specific tyrosine residues phosphorylated by Src-KD *in vitro*, we employed mass spectrometry (MS). C3G was phosphorylated by the Src-KD and was purified by size exclusion chromatography (SEC); samples of unphosphorylated C3G were also subject to SEC. Both types of samples were sent frozen to the Proteomics Facility of the Netherlands Cancer Institute (NKI, Amsterdam) for proteomic analysis.

Initially, C3G and pC3G samples were digested with trypsin and the resulting peptide maps were analyzed by mass spectrometry. Trypsin digestion resulted in coverage of ~67% of the C3G sequence (Figure R11 and Table R9). Two phospho-tyrosine pY329 and pY590 were robustly identified as Src targets (Table R10). The tryptic peptide containing Y329 was detected in the unphosphorylated and phosphorylated analysis with sufficient peptide spectrum matches (PSMs) to confirm the phosphorylation of this site. In contrast, the peptide containing Y590 was detected in the phosphorylated sample but not in the unphosphorylated. Despite not detecting the unphosphorylated peptide, the PSMs of the phosphorylated samples was sufficiently strong to confirm the Y590 as a Src target. Trypsin coverage missed out part of the sequence of C3G that included twelve tyrosines; in particular tyrosine residues in the AIR were part of a large tryptic peptide that was not detected in the MS analysis. To expand the coverage and to obtain information about remaining tyrosine residues in C3G, we performed chymotrypsin digestion of C3G and pC3G samples.

| Table R9. Trypsin and chymotrypsin digestion coverage of C3G sequence. | | | |
|---|---|---|---|
| **Protease** | **Coverage (%)** | **C3G fragments identified** | **Tyrosine residues in the identified fragments** |
| Trypsin | 66.95 | 8-25, 29-83, 87-376, 461-480, 583-649, 652-764, 789-799, 803-858, 895-943, 949-968, 1012-1034 | Y61, Y106, Y155, Y159, Y329, Y341, Y478, Y590, Y715, Y732, Y738, Y740, Y857, Y929, Y957, Y1030 |
| Chymotrypsin | 70.01 | 14-56, 75-132, 139-155, 175-244, 325-386, 393-418, 435-507, 512-554, 558-585, 597-618, 657-705, 724-738, 751-769, 788-797, 802-848, 858-865, 867-886, 893-909, 912-929, 934-984, 995-1009, 1017-1049 | Y106, Y155, Y329, Y341, Y386, Y478, Y485, Y504, Y554, Y561, Y570, Y579, Y738, Y929, Y957, Y970, Y984, Y1002, Y1030 |

Chymotrypsin peptide mapping resulted in a covered of ~70% of the sequence of C3G, which complemented the coverage of the tryptic digestion (Figure R11 and Table R9). This analysis resulted in the identification of two additional tyrosines Y504 and Y579 as Src phosphorylation sites, as well as the confirmation of the tyrosine Y329 (Table R11). The trypsin and chymotrypsin analyses yielded information about all the tyrosine residues of C3G except of Y390 and Y891. In summary, Src-KD phosphorylates C3G *in vitro* at four tyrosines located in the SH3b domain: Y329, Y504, Y579, and Y590.

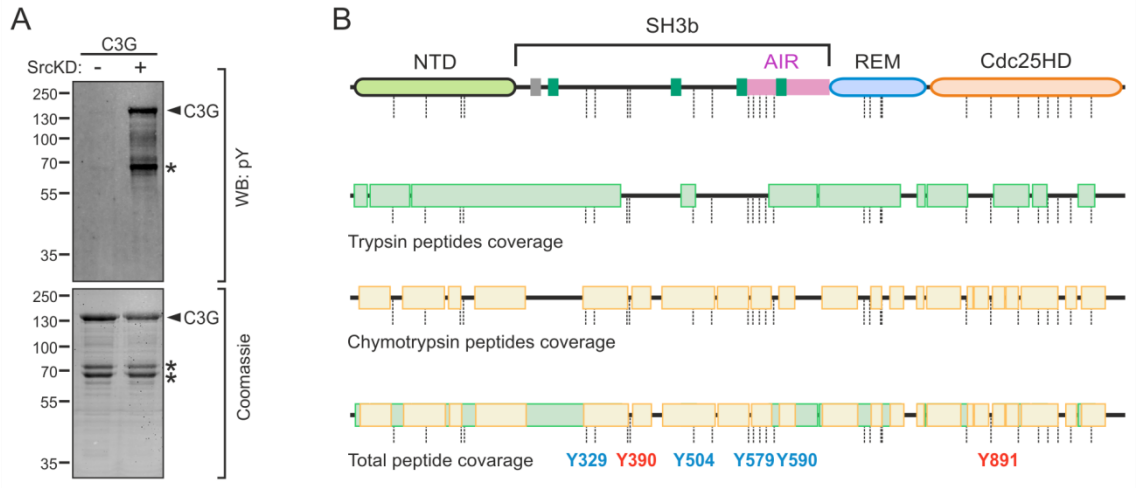



**Figure R11. Analysis of the *in vitro* Src-phosphorylation of C3G by MS. (A)** Phosphorylation of full-length C3G by Src-KD. Total protein was analyzed by Coomassie staining of non-phosphorylated and phosphorylated samples (2.5 µg). Phosphorylation was analyzed by loading 0.5 µg of the samples and pY was detected by WB using an anti-pY monoclonal antibody. **(B)** Schematic representation of the tryptic (green boxes) and chymotryptic (yellow boxes) peptides coverage. Blue numbers, corresponds to the phospho-modified tyrosines detected; red numbers correspond to the non-coverage tyrosines.

**Table R10. C3G tryptic phospho-peptide identified by mass spectrometry.**

| C3G Annotated Peptide sequence | Peptide limits | PSMs[a] | Modifications |
|---|---|---|---|
| [R].QDFDVDCYAQR.[R] | 322-332 | 81 | 1xPhospho [Y329] |
| [R].QDFDVDCYAQRR.[L] | 322-333 | 56 | 1xPhospho [Y329] |
| [R].ATSGSSLPVGINRQDFDVDCYAQR.[R] | 309-332 | 41 | 1xPhospho [Y329] |
| [R].ATSGSSLPVGINRQDFDVDCYAQRR.[L] | 309-333 | 33 | 1xPhospho [Y329] |
| [K].LLMEVYGFSDSFSGVDSVQELAPPPALPPK.[Q] | 585-614 | 12 | 1xPhospho [Y590] |
| [K].NKLLMEVYGFSDSFSGVDSVQELAPPPALPPKQR.[Q] | 583-616 | 11 | 1xPhospho [Y590] |
| [K].LLMEVYGFSDSFSGVDSVQELAPPPALPPKQR.[Q] | 585-616 | 11 | 1xPhospho [Y590] |
| [K].LLMEVYGFSDSFSGVDSVQELAPPPALPPKQR.[Q] | 585-616 | 4 | 1xPhospho [Y590] |

[a] Number of peptide spectrum matches.

**Table R11. C3G chymotryptic phospho-peptide identified by mass spectrometry.**

| C3G Annotated Peptide sequence | Peptide limits | PSMs[a] | Modifications |
|---|---|---|---|
| [R].QDFDVDCYAQR.[R] | 325-334 | 1 | 1xPhospho [Y329] |
| [R].QDFDVDCYAQRR.[L] | 486-507 | 11 | 1xPhospho [Y504] |
| [K].LLMEVYGFSDSFSGVDSVQELAPPPALPPKQR.[Q] | 570-585 | 2 | 1xPhospho [Y579] |
| [K].LLMEVYGFSDSFSGVDSVQELAPPPALPPKQR.[Q] | 570-585 | 2 | 1xPhospho [Y579] |
| [K].NKLLMEVYGFSDSFSGVDSVQELAPPPALPPKQR.[Q] | 571-585 | 2 | 1xPhospho [Y579] |

[a] Number of peptide spectrum matches.

Based on the mapping of phospho-sites by MS, we proceeded to generate a mutant in which the four tyrosine residues phosphorylated by Src-KD were replaced by phenylalanine (Y329-504-579-590F) and studied its phosphorylation *in vitro*. Among the tyrosines targeted by Src, Y579 and Y590 are situated within the AIR, which also contains Y561 and Y570 (Figure R12A). Due to the potential role of Y561 and Y570 in the autoinhibition and activation, we also created an additional mutant Y329-504-561-570-579-590F, in which these two tyrosine residues were also replaced by phenylalanine.

C3G wild type, C3G-Y329-504-579-590F and C3G-Y329-504-561-570-579-590F mutants were phosphorylated *in vitro* with Src-KD and their phosphorylation was analyzed by Western blot with anti-pY. C3G-Y329-504-579-590F showed a substantial reduction of the anti-pY signal compared to C3G wild type (Figure R12B). The mutant C3G-Y329-504-561-570-579-590F showed similarly low phosphorylation levels as the ones of C3G-Y329-504-579-590F, suggesting that Y561 and Y570 residues are not targeted by Src-KD in a detectable manner. These results support the notion that Y329, Y504, Y579 and Y590 are the main phosphorylation sites by Src-KD *in vitro*.

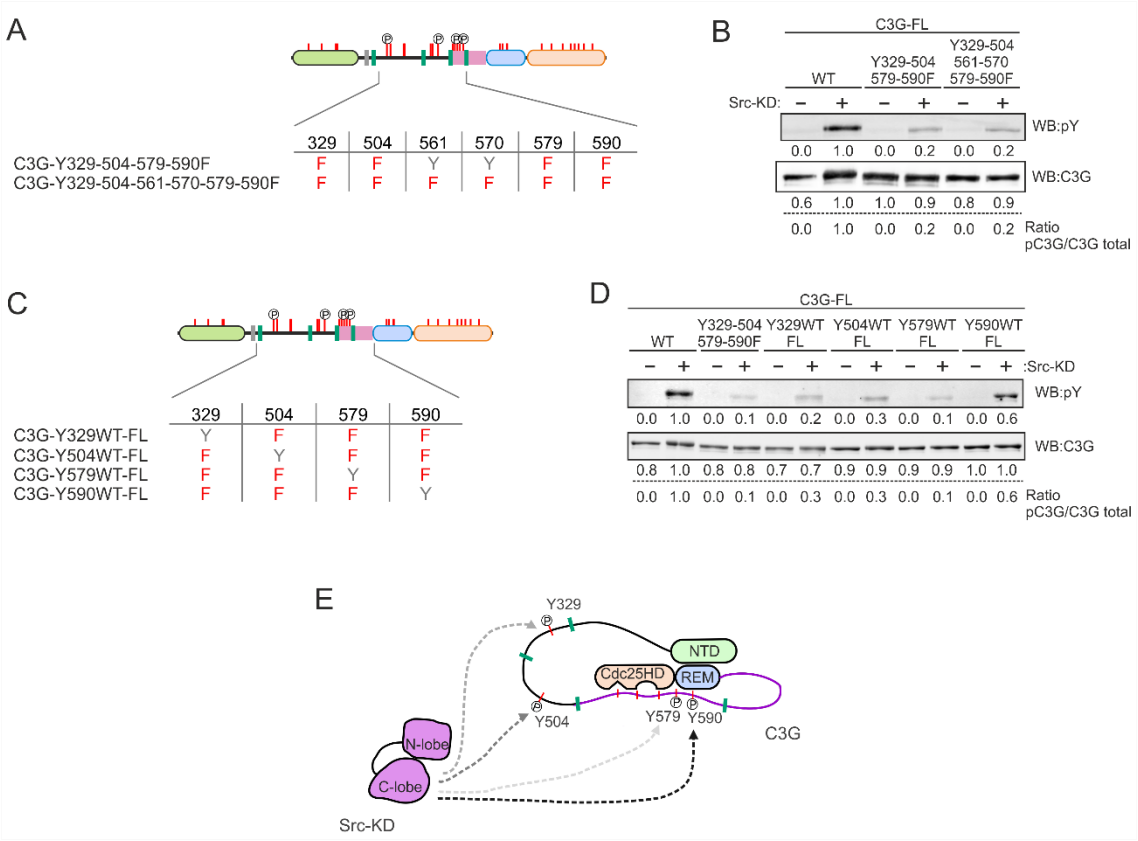



**Figure R12. Contribution of the Src-KD-targeted tyrosine residues to the phosphorylation of C3G.** **(A)** Schematic representation of the Y-to-F C3G mutants used to analyze the Src-phosphorylation sites identified by MS. **(B)** Phosphorylation of C3G tyrosine mutants by Src-KD. Tyrosine-phosphorylated and total proteins (0.5 µg per lane) were analyzed by WB using specific anti-C3G and anti-pY monoclonal antibodies. Values are the relative quantification of the bands and the relative ratios of pY signal versus C3G. **(C)** Schematic representation of additional C3G mutants used to analyze the contribution of each of the four tyrosine residues to the phosphorylation of C3G by Src-KD. **(D)** Phosphorylation of tyrosine C3G mutants by Src-KD. Samples and phosphorylation were analyzed by WB and quantified as in B. **(E)** Schematic representation of the phosphorylation of C3G by Src-KD. Arrow color intensity represents the level of phosphorylation of each tyrosine residue by Src-KD (the darker represents higher a contribution to the anti-pY signal).

Next, we were interested in assessing the individual contribution of each phosphorylatable tyrosine residue to the global phosphorylation of C3G. To this end, we designed four mutants, each of them conserving only one of the four residues unmodified, while the other three tyrosines were replaced by phenylalanine (Figure R12C). For simplicity, we named these mutants by the non-mutated phosphorylatable residue; for example, the mutant C3G-Y504-579-590F is named C3G-Y329WT-FL. Notably, we observed distinct levels of phosphorylation among these mutants (Figure R12D). C3G-Y590WT-FL displayed the highest phosphorylation levels, albeit it was slightly lower than the anti-pY signal of C3G wild type. C3G-Y504WT-FL and Y329WT-

FL exhibited intermediate phosphorylation signals, and C3G-Y579WT-FL showed a similar phospho-signal as the quadruple mutant C3G-Y329-504-579-590F. Collectively, these data suggest that the global phosphorylation levels of C3G is a cumulative contribution of the individual phosphorylation targeting Y329, Y504, Y579 and Y590, and that quantitatively Y590 is the main phosphorylation site (Figure R12E).

### 2.2. Mapping the binding site for the SH2 domain of CrkL in phosphorylated C3G

We have shown that phosphorylation of C3G by Src and a functional SH2 domain in CrkL are required for the complete activation of C3G by CrkL; in addition, the SH2 domain of CrkL binds weakly to pC3G. Taken together, these findings suggest that the SH2 domain of CrkL binds to one or more specific phospho-tyrosine residues of C3G, possibly located inside the AIR, and that such interaction contributes to the activation of C3G.

Therefore, we addressed the identification of the tyrosine residue of C3G that acts as a docking site for the CrkL-SH2 domain. To that end, we analyzed the interaction by GST pull-down assays using C3G fragments and point mutants (Figure R13).

Initially, we used GST fusions of C3G full-length (FL, 4-1077), the NTD (4-245), the central SH3b region (274-646), and the catalytic region REM-Cdc25HD (602-1077) (Figure R13A). These constructs were used unmodified or phosphorylated by Src-KD and their binding to CrkL-W160S (full-length with inactive SH3N domain) was analyzed. The phosphorylated forms of C3G full-length and the SH3b fragment bound to CrkL-W160S; but either no binding or only a marginal signal was detected with the unphosphorylated proteins (Figure R13B). In contrast, neither NTD nor GST alone bound to CrkL-W160S, in consonance with their lack of phosphorylation by Src. The phosphorylated REM-Cdc25HD region showed a marginal interaction with CrkL-W160S. In summary, the CrkL-SH2 domain binds to the SH3b of pC3G, in agreement with this region harboring the main Src phosphorylation sites.

The SH3b contains 12 tyrosine residues. To narrow down the tyrosines responsible for the binding of the CrkL-SH2 domain, we created two constructs corresponding to two halves of the SH3b, 274-500 and 501-646, each containing six tyrosine residues. We analyzed the interaction by pull-down assays as above (Figure R13C). The phosphorylated fragment 501-646 pulled down CrkL-W160S similarly to the complete SH3b region. In contrast, CrkL-W160S did not bind significantly to phosphorylated fragment 274-500, and it did not bind to the unphosphorylated proteins either.

Collectively, both pull down assays strongly suggest that CrkL-SH2 domain binds to the second half of the SH3b, 501-646 (Figure R13D). This region includes the AIR and contains two of the Src phosphorylation sites, Y579 and Y590, as well as other four tyrosines (Y504, Y554, Y561, and Y570).

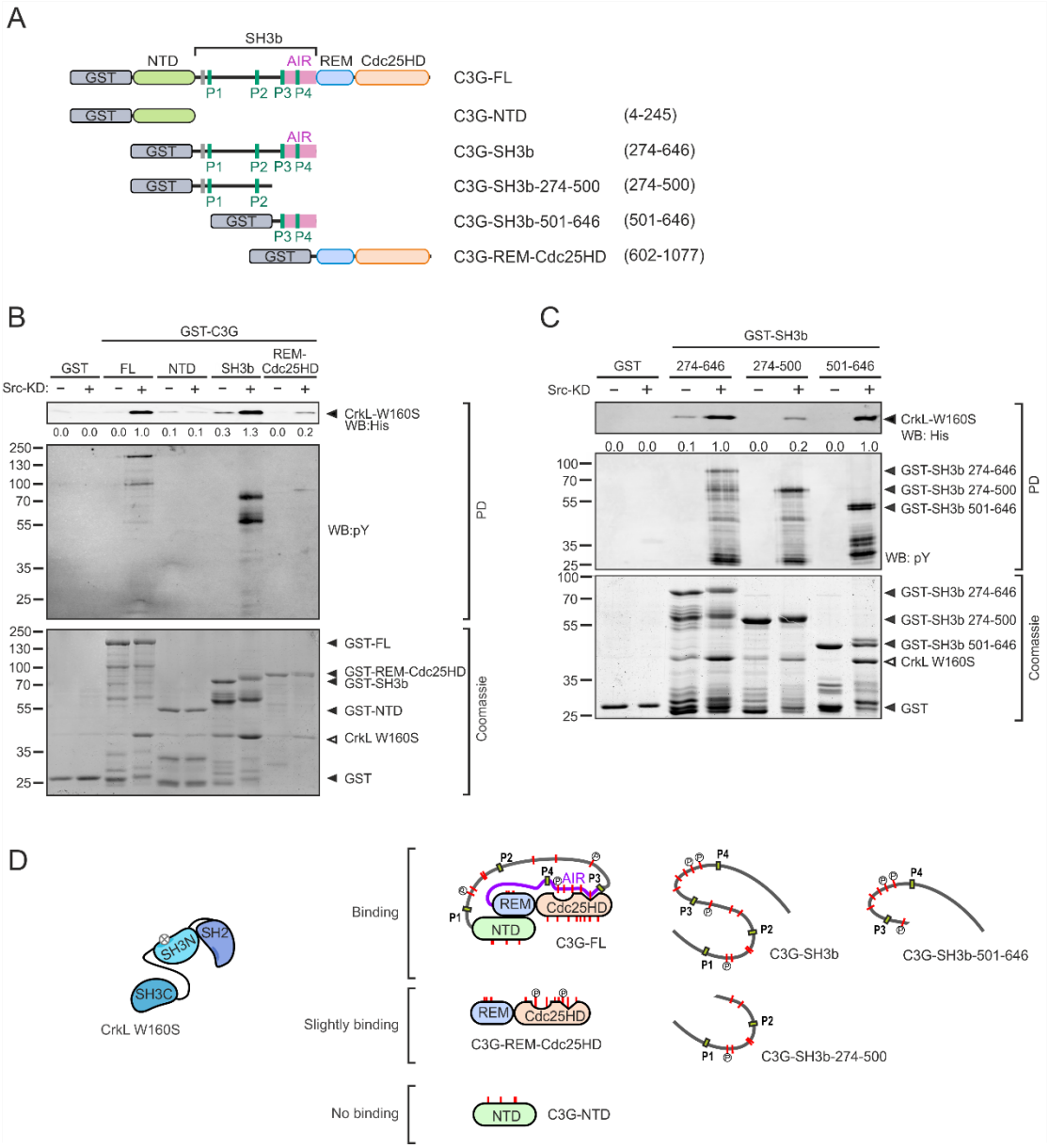



**Figure R13. Mapping the region of C3G that mediates the interaction with the SH2 domain of CrkL.** **(A)** Schematic representation of the C3G fragments used to analyze the binding of CrkL-W160S. **(B)** Pull-down (PD) analysis of the binding of CrkL-W160S to the C3G fragments in A untreated and after phosphorylation by Src. CrkL-W160S in which the His-tag was not removed and phosphorylation of C3G constructs were detected by Western blot (WB) using anti-His tag and anti-pY monoclonal antibodies, respectively. Gels were loaded with 5% (v/v) of the PD sample. Total protein of GST-C3G constructs and GST (used as control) were visualized by Coomassie staining using 2 µg of protein per lane. Values are the relative quantification of the anti-His tag signal. **(C)** Pull-down (PD) analysis of the binding of CrkL-W160S to GST-C3G SH3b region and two halves constructs, GST-274-500 and GST-501-646, before and

after phosphorylation of the C3G constructs with Src-KD. Detection of CrkL-W160S, phosphorylation and total proteins were performed as in B. Numbers are the relative quantification of the anti-His tag signal as in B. **(D)** Schematic representation of the interaction between the CrkL-W160S mutant and phosphorylated fragments of C3G.

To identify the specific tyrosine recognized by the SH2 domain of CrkL, we employed site-directed mutagenesis to replace tyrosines within the 501-646 region of C3G by phenylalanine. When we started this analysis, we had not mapped yet the phosphorylation sites, therefore we analyzed all six tyrosine residues in an unbiased manner. We generated six mutants in which five of the tyrosines were replaced by phenylalanine and only a single tyrosine remained unmodified (hereafter referred to by the non-mutated tyrosine number); we also created a mutant in which all the six tyrosines replaced (AllY-F). Initially, mutants were created in the construct 501-646 fused to GST; proteins were used unmodified and phosphorylated by Src-KD, and their binding to CrkL-W160S was analyzed with pull-down assays (Figure R14A). Out of the six mutants, only C3G-501-646-Y590WT was notably phosphorylated above the levels observed for C3G-501-646-AllF; yet the anti-pY signal of C3G-501-646-Y590WT was much lower than that of the wild type fragment. In consonance with the phosphorylation levels, pC3G-501-646-Y590WT bound to CrkL-W160S, albeit to a less extent compared to the wild type C3G-501-646 fragment. The remaining mutants, Y504WT, Y554WT, Y561WT, Y570WT, Y579WT and AllY-F, did not display any phosphorylation and they did not interact with CrkL-W160S except for a marginal interaction with pC3G-501-646-Y579WT.

In a complementary approach, we performed pull-down assays using GST-CrkL-W160S and mutants of full-length C3G in which only one tyrosine of the AIR was mutated to phenylalanine (Figure R14B). Prior to the pull-down, the full-length C3G mutants were phosphorylated by Src-KD and purified by SEC. Binding to GST-CrkL-W160S was lost in the pC3G-Y590F mutant and partially reduced with the pC3G-Y579F mutant. Yet, pC3G mutants Y504F, Y554F, Y561F and Y570F bound to GST-CrkL-W160S similarly as wild type pC3G.

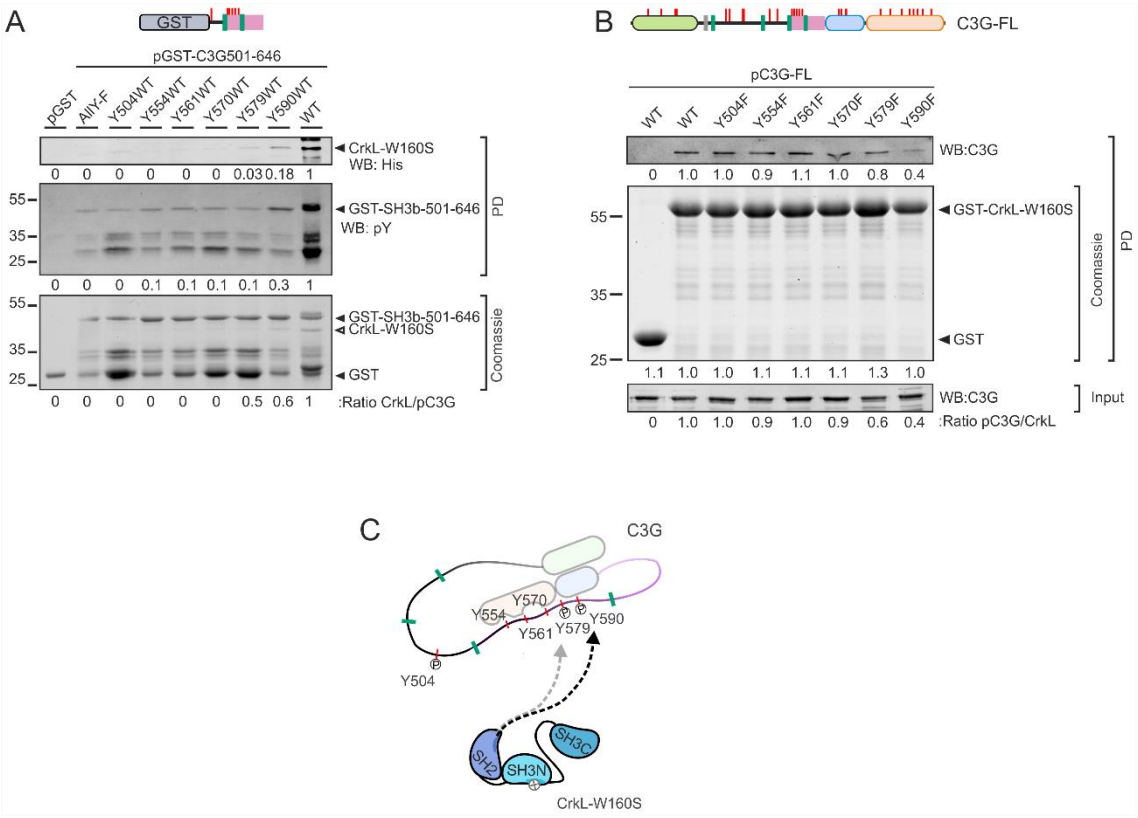



**Figure R14. Identification of tyrosine residues in the AIR essential for the binding to the CrkL-SH2 domain. (A)** Pull-down (PD) analysis of the binding of CrkL-W160S to the Src-phosphorylated GST-C3G-501-646 fragment, wild type, and tyrosine mutants. Phosphorylation of C3G proteins and presence of CrkL-W160S in the PD were detected by Western blot (WB) using anti-pY and anti-His tag monoclonal antibodies, respectively. Levels of GST-C3G constructs and GST (used as control) in the input sample before PD were visualized by Coomassie staining loading 2 µg of total protein. Values are the relative quantification within each panel. **(B)** Analysis by PD of the interaction of GST-CrkL-W160S mutant with phosphorylated full-length C3G wild type (WT) and single tyrosine mutants. C3G was detected in the PD samples by WB. C3G in the input samples before PD (0.5 µg of total protein) was analyzed by WB, and GST-CrkL-W160S and GST were visualized by Coomassie staining (2 µg of GST protein loaded). Numbers are the relative quantification within each panel. **(C)** Representation of the binding between the CrkL-SH2 domain and the AIR tyrosine residues of C3G. Arrow color intensity represents the intensity of the binding between the C3G tyrosine residues and SH2 domain of CrkL.

In summary, our results support the notion that the interaction between the CrkL-SH2 domain and pC3G is mediated by the AIR of C3G, and that Y590 plays a crucial role in this interaction (Figure R14C).

### 2.3. Role of tyrosine residues in the activation of phosphorylated C3G

Our previous results indicate that tyrosine phosphorylation of C3G by Src is important for the stimulation of the GEF activity of C3G. As a key step, phosphorylation creates a novel secondary binding site between CrkL and pC3G. This interaction between CrkL and pC3G seems to involve Y590 within the AIR of C3G. This prompted us to analyze the role of specific tyrosine residues for the activation of Src-phosphorylated C3G.

We focused on the region 501-646 of C3G, which contains a total of six tyrosine residues. Tyrosines Y554, Y561, Y570, Y579 and Y590 are situated within the AIR (537-646). Additionally, this region includes the well-characterized Y504, whose phosphorylation has been linked to the activation of C3G in cells (Radha et al., 2004, Radha et al., 2008).

To analyze the role of tyrosine residues of C3G in its activation we measured the GEF activity of mutants of full-length C3G, using the wild type protein as a reference, in four conditions: unphosphorylated C3G in the absence and in the presence of CrkL, and phosphorylated C3G in the absence and in the presence of CrkL. The activity of the unphosphorylated state informs if the autoinhibitory mechanism was affected by the mutations. Similar information is obtained by comparing the activity of the unphosphorylated state without and with CrkL, because weakening of the autoinhibition favors the activation of non-phosphorylated C3G by CrkL. Comparing the activity of unphosphorylated and phosphorylated states in the absence of CrkL quantifies the weakening of the autoinhibition by the phosphorylation alone. While differences in the activity between the unphosphorylated and phosphorylated states in the presence of CrkL give an idea of the contribution of the phosphorylation to the activation, this effect combines the contribution of the phosphorylation itself and the interaction of CrkL-SH2 domain with pC3G.

Initially, we analyzed the GEF activity of the quadruple mutant C3G-Y561-570-579-590F, as well as the one of Src-phosphorylated C3G in the absence and presence of CrkL (5 µM) (Figure R15A). In this mutant we kept unmodified Y504, as it is not part of the autoinhibitory region, and Y554, which plays an essential role to maintain the AIR/Cdc25HD autoinhibitory interaction (Carabias et al., 2020). C3G-Y561-570-579-590F showed an elevated basal GEF activity ($k_{obs}$ 9.9 x $10^{-3}$ $s^{-1}$), that was 8 times higher than the wild type C3G. The basal GEF activity of the mutant was higher than the activity of unphosphorylated C3G wild type in the presence of CrkL or the activity of phosphorylated C3G wild type in the absence of CrkL. Phosphorylation of C3G-Y561-570-579-590F did not increase the GEF activity relative to non-phosphorylated state. In

contrast, despite the high basal GEF activity observed in the mutant, CrkL activated it to a similar extent as the wild type phosphorylated C3G. The same activity was observed when CrkL was added to pC3G mutant. These results suggest that the residues Y561, Y570, Y579 and Y590 of C3G play a role in the autoinhibitory mechanism of C3G. Because this mutant is not further activated by phosphorylation (neither in absence nor presence of CrkL) these four residues, or some within them, are key sites for the stimulation of C3G triggered by Src-phosphorylation and the interaction with CrkL.

To assess the role of each tyrosine within the 501-646 fragment of C3G, we analyzed the activity of single tyrosine mutants (only one residue substituted at a time) in both non-phosphorylated C3G and pC3G full-length, in absence and presence of CrkL (5 μM) (Figure R15B-C). Under these conditions, C3G-Y504F had the same activity and was stimulated similarly to the wild type C3G, suggesting that Y504 is not involved in the direct activation of C3G.

C3G-Y554F mutant displayed a higher basal GEF activity relative to wild type C3G (a 3.8-fold increase). Both CrkL binding and Src-phosphorylation enhanced the GEF activity to similar levels; while these levels of activity were higher than those of C3G wild type, the C3G-Y554F responded to both stimuli. The GEF activity of pC3G-Y554F in the presence of CrkL was similar to that of wild type pC3G with CrkL. These results suggest that while Y554 is required for maintaining the autoinhibited state, it is not involved in the phosphorylation-dependent activation of C3G by CrkL, which is consistent with Y554 not being a phosphorylation site.

Mutants C3G-Y561F and C3G-Y590F showed a small, 2-fold, increase in basal GEF activity with respect to the wild type protein. CrkL activated both unphosphorylated mutants to higher levels than to the wild type C3G. In contrast, upon phosphorylation with Src the activity of these mutants, in the absence or presence of CrkL, was similar to that of the equivalent non-phosphorylation state.

Finally, mutants C3G-Y570F and Y579F showed a substantially higher basal GEF activity in comparison to wild type C3G (a 2.8-fold and 4.6-fold increase, respectively). Yet, these mutants responded to the activation by Src-phosphorylation, CrkL interaction, and the combined of the two stimuli; the latter resulted in activity levels comparable to those of phosphorylated wild type C3G with CrkL. In summary, C3G-Y570F and C3G-Y579F showed altered autoinhibition but seemed to retain the activation mechanisms.

Collectively these results suggest that: (i) Y504 does not participate directly in the stimulation of C3G mediated by Src-phosphorylation and CrkL interaction, (ii) Y554 has a role in maintaining the autoinhibited and closed conformation of C3G; yet, it is not required for the stimulation by Src-phosphorylation and CrkL, (iii) Y561 and Y590 are involved in the stimulation of the GEF activity of pC3G and CrkL, possibly through interaction with the CrkL-SH2 domain, (iv) Y570 and Y579 contribute to the autoinhibitory mechanism but they do not participate in the stimulation of pC3G by CrkL.

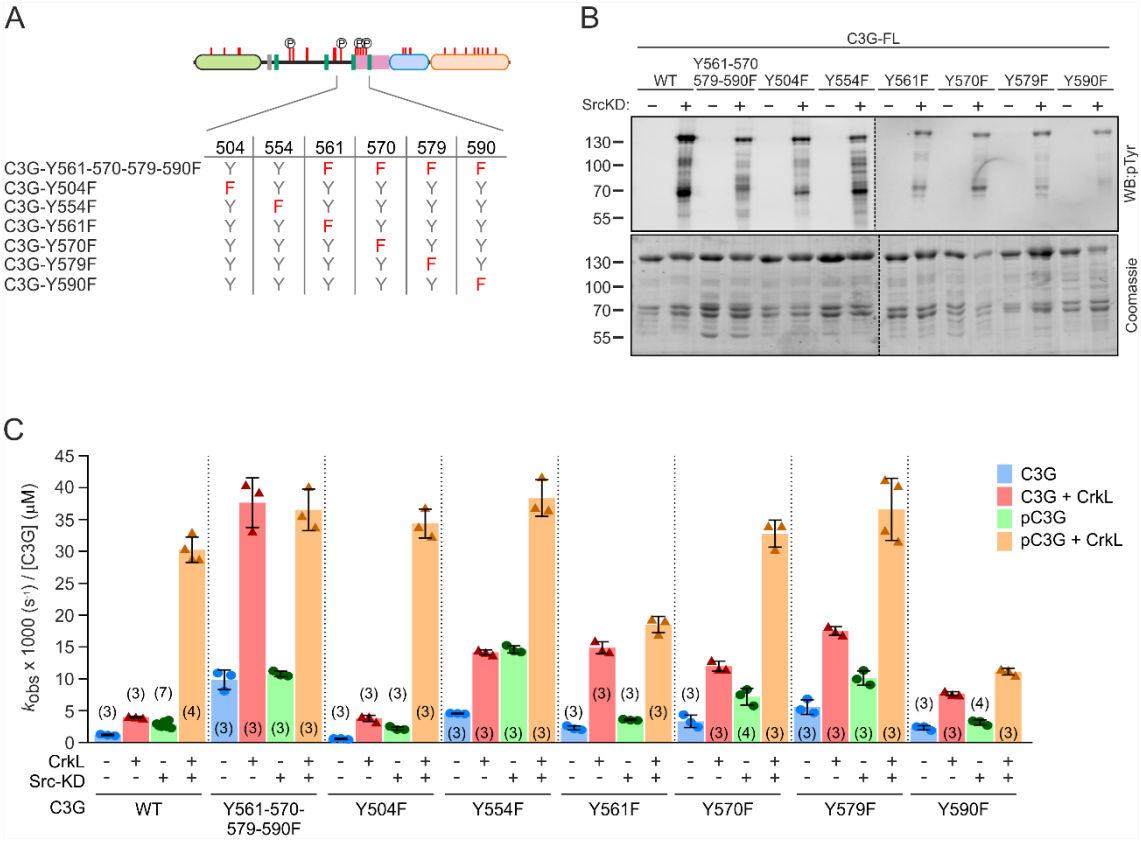



**Figure R15. Role of tyrosine residues in the SH3b in the direct activation of C3G by Src-phosphorylation and CrkL. (A)** Schematic representation of the tyrosine C3G mutants used to study the activation of C3G by Src-phosphorylation and CrkL. **(B)** Phosphorylation of tyrosine C3G mutants by Src-KD. Total protein was analyzed by Coomassie staining of non-phosphorylated and phosphorylated samples (2.5 µg). Phosphorylation was analyzed by loading 0.5 µg of the samples and pY was detected by WB using an anti-pY monoclonal antibody. **(C)** Nucleotide exchange rates ($k_{obs}$) catalyzed by full-length C3G WT and tyrosine mutants (0.2 µM, data referred to 1 µM C3G). Unphosphorylated and Src-phosphorylated C3G WT and mutants were analyzed alone and in the presence of CrkL (5 µM). Data are shown as scatter plots with bars, means ± standard deviation. The number of independent measurements is indicated in parentheses.

Subsequently, we analyzed specifically the role of the four tyrosines that we have identified as phosphorylation sites of Src (Y329, Y504, Y579 and Y590). We used the triple mutants of full-length C3G in which three of these phosphorylatable residues were replaced by phenylalanine and only one of them was kept unmodified (C3G-Y329WT-FL, C3G-Y504WT-FL, C3G-Y579WT-FL, and C3G-Y590WT-FL). We also analyzed the quadruple mutant (C3G-Y329-504-579-590F) with all four tyrosines replaced. As above, we measured the GEF activity of these mutants, and wild type C3G as a reference, in four conditions: unphosphorylated or phosphorylated, in the absence and presence of CrkL. Attending to Src-phosphorylation effect in the activation of C3G wild type and tyrosine mutants, phosphorylated C3G wild type displayed an increase in GEF activity (Figure R16A-C). The GEF activity of the mutant lacking the four Src-targeted tyrosine residues (C3G-Y329-504-579-590F), as well as the mutants C3G-Y329WT-FL, C3G-Y504WT-FL and C3G-Y579WT-FL, did not increase in response to Src-phosphorylation with respect to the unphosphorylated state (in the absence of CrkL). In contrast, the activity of C3G-Y590WT-FL increased twice upon phosphorylation. This moderate increase was similar to the effect of phosphorylation in the GEF activity of wild type C3G. CrkL activated non-phosphorylated C3G wild type and mutants in a similar manner (4-fold increase in GEF activity). Notable, Src-phosphorylated C3G-Y590WT-FL was activated by CrkL similarly to phosphorylated C3G wild type. However, the rest of mutants, which all harbor the Y590F substitution (C3G-Y329-504-579-590F, C3G-Y329WT-FL, C3G-Y504WT-FL, and C3G-Y579WT-FL), did not respond to CrkL activation after Src-phosphorylation with respect to the unphosphorylated form (Figure R16D).

Under similar conditions of concentrations, CrkL-R39K (harboring a mutation in the SH2 domain) did not increase the GEF activity of the phosphorylated C3G-Y590WT-FL mutant (Figure R16E). We also studied the dependence of the GEF activity of the Src-phosphorylated C3G-Y590WT-FL protein, on the concentration of CrkL wild type and CrkL-R39K (Figure R16F and Table R12). pC3G-Y590WT-FL was stimulated at slightly lower concentration of CrkL wild type than the SH2-inactive mutant CrkL-R39K ($AC_{50}$ 1.4 μM and 2.6 μM, respectively). The main difference was observed in the maximal activity induced by CrkL wild type and CrkL-R39K; the mutant activated pC3G-Y590WT-FL to a maximal activity ($k_{max}$ 33.7 x $10^{-3}$ $s^{-1}$) that was ~25% lower than that induced by wild type CrkL ($k_{max}$ 45.3 x $10^{-3}$ $s^{-1}$).

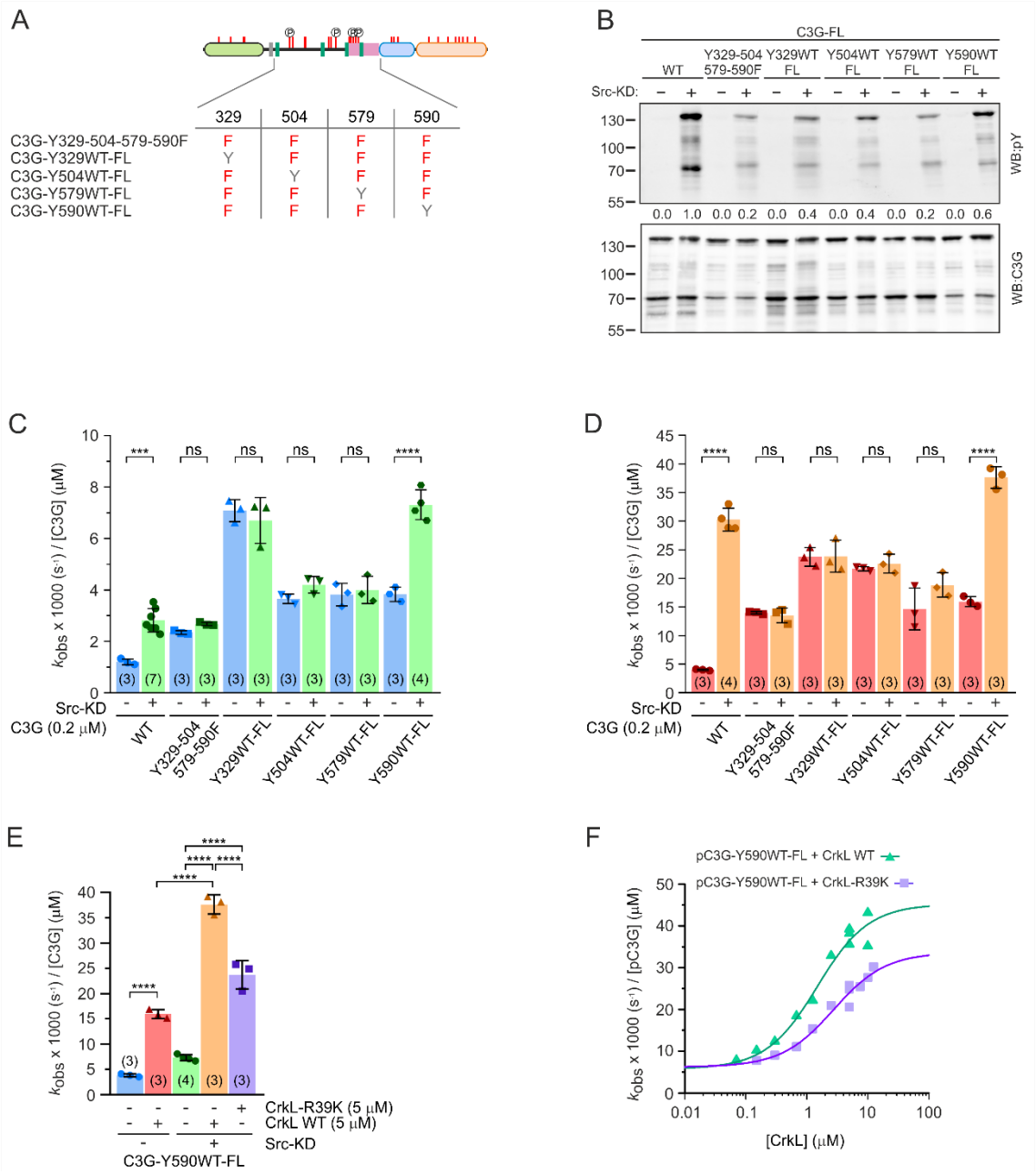



**Figure R16. Analysis of the Src-phosphorylation and CrkL effect in the GEF activity of Src-KD-target tyrosine mutants of C3G. (A)** Schematic representation of the tyrosine C3G mutants used to study the activation of C3G by Src-phosphorylation and CrkL. **(B)** Phosphorylation of C3G tyrosine mutants by Src-KD. C3G (0.5 μg) and its phosphorylation were analyzed and detected by WB using anti-C3G and anti-pY monoclonal antibodies. **(C)** Nucleotide exchange rates ($k_{obs}$) catalyzed by unphosphorylated and Src-phosphorylated full-length C3G wild type and tyrosine mutants (0.2 μM). **(D)** Nucleotide exchange rates ($k_{obs}$) catalyzed by full-length C3G WT and tyrosine mutants (0.2 μM) unphosphorylated and Src-phosphorylated analyzed in the presence of CrkL (5 μM). **(E)** Nucleotide exchange rates ($k_{obs}$) catalyzed by full-length C3G-Y590WT-FL mutant (0.2 μM), unphosphorylated and Src-phosphorylated analyzed alone and in the presence of CrkL WT and R39K mutant (5 μM). Data in C, D and E are shown as scatter plots with bars, means ± standard deviation. In C, D and E, the number of independent measurements is indicated in parenthesis. **(F)** Dose-dependent effect of CrkL WT and CrkL-R39K mutant on the GEF activity of pC3G-Y590WT-FL (0.2 μM). Lines are the fitted sigmoidal activation models. In C, D, E and F

nucleotide exchange rates are referred to 1 µM C3G. In C, D and E, statistical comparison was analyzed using ANOVA followed by Tukey´s multiple comparisons test; *** $P < 0.001$, **** $P < 0.0001$, ns $P > 0.05$.

**Table R12. Parameters of dose-response activation of pC3G-Y590WT-FL by CrkL WT and R39K.**

| C3G | CrkL protein | $AC_{50}$ (µM)[a] | $k_{free}$ x1000 ($s^{-1}$) | $k_{max}$ x1000 ($s^{-1}$) | $k_{max}/k_{free}$ |
|---|---|---|---|---|---|
| pC3G-Y590WT-FL | CrkL WT | 1.4 ± 0.5 | 5.7 ± 0.5 | 45.3 ± 0.6 | 7.9 ± 1.2 |
| pC3G-Y590WT-FL | CrkL-R39K | 2.6 ± 1.0 | 6.2 ± 0.3 | 33.7 ± 0.5 | 5.4 ± 1.6 |

[a] Data are presented as the fitted value ± the associated standard errors of the nonlinear regression fit.

Collectively, our results support the notion that while the binding of CrkL via the SH3N to the P3 and P4 sites is the main event for the activation of C3G, the complete activation of Src-phosphorylated C3G relies on Y590 and requires a functional SH2 domain of CrkL. This suggests that the CrkL-SH2 domain contributes to the activation of C3G through a secondary interaction with pY590.

### 2.4. The interaction of the CrkL-SH2 domain with pC3G contributes to the stability of the complex

Our previous results suggest that the SH2 domain of CrkL contributes to the activation of C3G through an interaction, apparently with pY590 in the AIR. Despite being a weak interaction, this secondary contact seems to make a significant contribution to the activation of C3G. To better understand the role of the SH2 domain of CrkL in the stimulation of C3G, we pursued a further analysis of this interaction.

Initially, we used ITC to analyze quantitatively the contribution of the SH2 domain of CrkL to the binding of CrkL with pC3G. In these studies, we used the C3G-AAPP mutant, in which the P1 and P2 had been inactivated. CrkL activates phosphorylated C3G-AAPP similarly to C3G wild type (see above). Thus, by using C3G-AAPP we could focus on the activating interactions of CrkL. We measured the binding of CrkL (full-length) and the isolated SH3N domain to C3G-AAPP in the unphosphorylated and Src-phosphorylated forms. In these experiments, C3G-AAPP (7.5 to 13.5 µM) was titrated with CrkL-SH3N (between 79.6 and 133.8 µM), or CrkL full-length (62 to 160 µM) in the syringe. CrkL-SH3N and CrkL full-length produced similar exothermic signals during their binding to C3G-AAPP or pC3G-AAPP mutant (Figure R17A-B). As a control, we injected the maximum concentration of CrkL-SH3N and CrkL full-length (at the highest concentrations used for binding) into buffer, resulting in minimal heat exchanges.

The isolated SH3N domain of CrkL bound to the Src-phosphorylated pC3G-AAPP with similar stoichiometry (N 0.92-0.99) and affinity ($k_d$ 2.8-4.3 µM) as observed for the binding to unphosphorylated C3G-AAPP (N 0.90-0.98 and $k_d$ 2.5-3.6 µM). The enthalpy of the interaction showed only marginal difference, with a slight reduction from -15.6 kcal/mol for unphosphorylated C3G-AAPP to -13.3 kcal/mol for Src-phosphorylated pC3G-AAPP mutant (Table R13). Binding of full-length CrkL to C3G-AAPP showed similar stoichiometry (N 0.83-0.92), affinity ($k_d$ 3.9-6.9 µM) and enthalpy (ΔH -14.2 kcal/mol) as observed for the binding of the SH3N domain. Finally, CrkL full-length bound to pC3G-AAPP with remarkably similar affinity ($k_d$ 2.3-3.5 µM), stoichiometry (N 0.92-1.08) and enthalpic change; as it bound to unphosphorylated C3G-AAPA, and as CrkL-SH3N bound to pC3G-AAPP. In summary, the subtle differences in the binding parameters did not reveal any noticeable contribution of the CrkL-SH2 domain to the interaction with pC3G.

Because we were not able to detect any contribution of the SH2 to the SH3N-driven binding of CrkL to pC3G, we attempted to measure the direct interaction of the SH2. To that end, we used a high concentration of pC3G-AAPP in the cell (46.7 µM), which was titrated with injections of 15 µL of the CrkL-W160S SH3N-inactive mutant (646.4 µM). Injection of CrkL-W160S produced endothermic signals during the binding to pC3G-AAPP (Figure R17C). To validate that the heat signal observed is produced by the protein-protein interaction, we measured the signal diluting CrkL-W160S (649 µM) into the buffer, which resulted in minimal heat exchanges. Analysis of these data was compromised by the weak heat exchange signals. Yet, an estimation of the affinity was calculated by fixing the enthalpy of the reaction to 1.0 kcal/mol, this value is based on previous ITC data of the interaction of the Src-SH2 domain with a phosphorylated tyrosine peptide (Bradshaw et al., 1999). Also, stoichiometry was assumed to be one-to-one. This yielded an estimated $k_d$ for the interaction of CrkL-W160S with pC3G-AAPP between 38 and 57 µM.

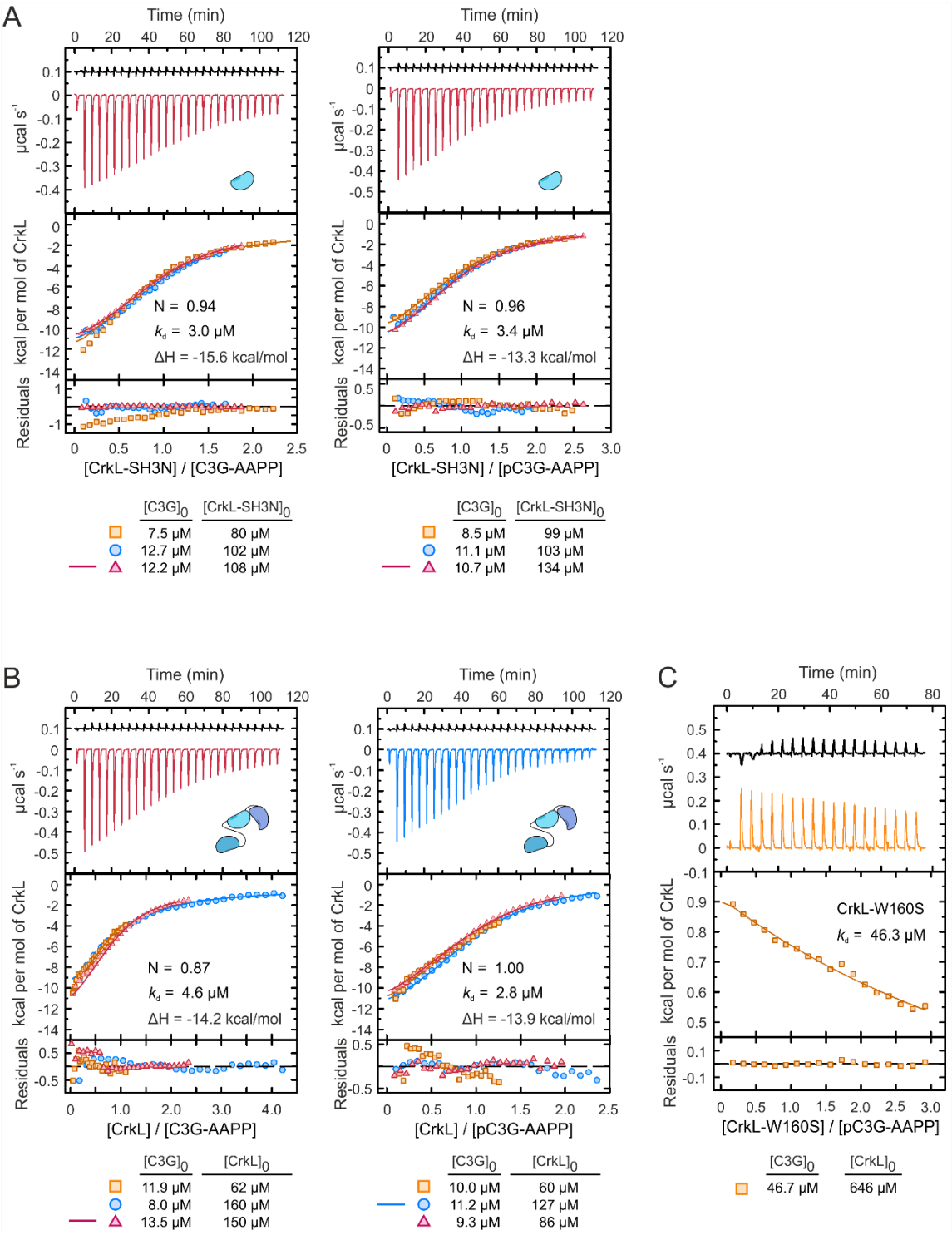



**Figure R17. ITC analysis of the contribution of the CrkL-SH2 domain to the binding to C3G. (A-B)** ITC experiments of the binding of CrkL-SH3N domain (A) and CrkL full-length (B) to C3G-AAPP and Src-phosphorylated pC3G-AAPP. ITC measurements were performed at 25 ºC. In each panel, the upper graph shows, the heats of dilution of CrkL-SH3N domain in 20 mM Tris-HCl (pH 7.5), 300 mM NaCl (black thermogram), and a representative protein-protein interaction thermogram (colors correspond to independent titrations). The middle graphs are the binding isotherms. One binding site model was fitted to the data; fitted models are shown as a dashed line. The lower graph shows the residuals from the fitted models. **(C)** ITC experiments of the binding of full-length CrkL-W160S to Src-phosphorylated pC3G-AAPP. Data are presented as in A and B.

**Table R13. Thermodynamic parameters for the binding of SH3N domain of CrkL or full-length CrkL WT or CrkL-W160S mutant to C3G-AAPP.**

| C3G mutant | CrkL construct | Num exp[a] | N[b] | $k_d$ (μM) | ΔH (kcal/mol)[d] |
|---|---|---|---|---|---|
| C3G-AAPP | CrkL-SH3N | 3 | **0.94** [0.90-0.98][c] | **3.0** [2.5-3.6][c] | -15.6 [-14.4; -17.2][c] |
| C3G-AAPP | CrkL-FL | 3 | **0.87** [0.83-0.92][c] | **4.6** [3.9-6.9][c] | -14.2 [-12.8; -16.3][c] |
| pC3G-AAPP | CrkL-SH3N | 3 | **0.96** [0.92-0.99][c] | **3.4** [2.8-4.3][c] | -13.3 [-12.3; -14.9][c] |
| pC3G-AAPP | CrkL-FL | 3 | **1.00** [0.92-1.08][c] | **2.8** [2.3-3.5][c] | -13.9 [-12.9; -15.4][c] |
| pC3G-AAPP | CrkL-W160S | 1 | **1.0**[e] | **46.3** [38.3-56.5][c] | **1.0**[e] |

[a] Number of independent experiments used for the thermodynamic analysis.

[b] N refers to stoichiometry.

[c] Data in brackets represent the asymmetric errors within a 0.95 confidence interval.

[d] ΔH is the binding enthalpy.

[e] Values were fixed for analysis.

Next, we used SEC to analyze the interaction between full-length CrkL and C3G, and the contribution of the SH2 domain of CrkL to this interaction. For these experiments, we used the C3G-AAPA mutant, which only had the P3 activation site available, while the P1, P2 and P4 sites had been disrupted. In this way, we focused on analyzing the function of the SH2 domain in the interaction of the CrkL molecule bound through the SH3N domain to the P3 site, involved in exerting the effect on pC3G activity.

First, when C3G-AAPA (full-length) was injected alone, it eluted as a major peak centered at 12.25 mL. Similarly, CrkL wild type alone eluted as a single peak at 16.72 mL. Next, we analyzed a mixture of C3G-AAPA and CrkL at 20 μM and 80 μM, respectively (corresponding to a molar ratio of 4:1 of CrkL excess over C3G). The elution profile of the mixture showed two major peaks: each one eluting in the same positions as the isolated C3G-AAPA and CrkL, respectively. We analyzed the fractions of the elution peaks by SDS-PAGE and Coomassie staining. CrkL was barely detected in the fractions corresponding to the first peak in the mixture (Figure R18A and D; red line). These results suggested the limited formation of the C3G-AAPA/CrkL complex; this is consistent with the low stoichiometry of ~0.2 molecules of CrkL per molecule of C3G-AAPA observed by ITC, and with the rapid exchange of the interaction between CrkL and C3G (Carabias, 2019).

Next, we analyzed the formation of complexes between CrkL and phosphorylated pC3G-AAPA. Phosphorylated pC3G-AAPA eluted as a major single peak (12.16 mL) slightly shifted to lower elution volume compared to unphosphorylated C3G-AAPA, possibly due to a more open or relaxed conformation of the phosphorylated state. The elution profile of the mixture containing pC3G-AAPA and CrkL showed a displacement of the first peak to a lower retention volume (11.95 mL). In addition, the area of the second peak corresponding to CrkL was reduced compared to the run of CrkL alone. Analysis of the fractions of the first peak by SDS-PAGE revealed the presence of CrkL

in that first peak. Therefore, the displacement of the elution volume is apparently caused by the formation of the complex (Figure R18B and D; orange line).

Finally, we analyzed if the complex formation could be mediated by the interaction through the CrkL-SH2 domain and the phospho-Y residues of pC3G. For that purpose, we employed the CrkL-R39K mutant. When loaded alone, CrkL-R39K eluted in a single peak (16.18 mL). The mixture of pC3G-AAPA and CrkL-R39K (20 μM and 80 μM, respectively) resulted in a subtle shift in the peak of pC3G to lower retention volume (12.10 mL). In addition, the second peak, which corresponds to CrkL-R39K, was less reduced compared to observe during the complex formation with CrkL wild type. Analysis of the fractions of the first elution peak, revealed less presence of CrkL-R39K in the first peak compared to CrkL wild type in the presence of pC3G-AAPA (Figure R18C and D; grey line). Collectively, these results support the notion that the SH2 domain stabilizes the interaction of CrkL with the P3 in pC3G, possibly by slowing down the interaction.

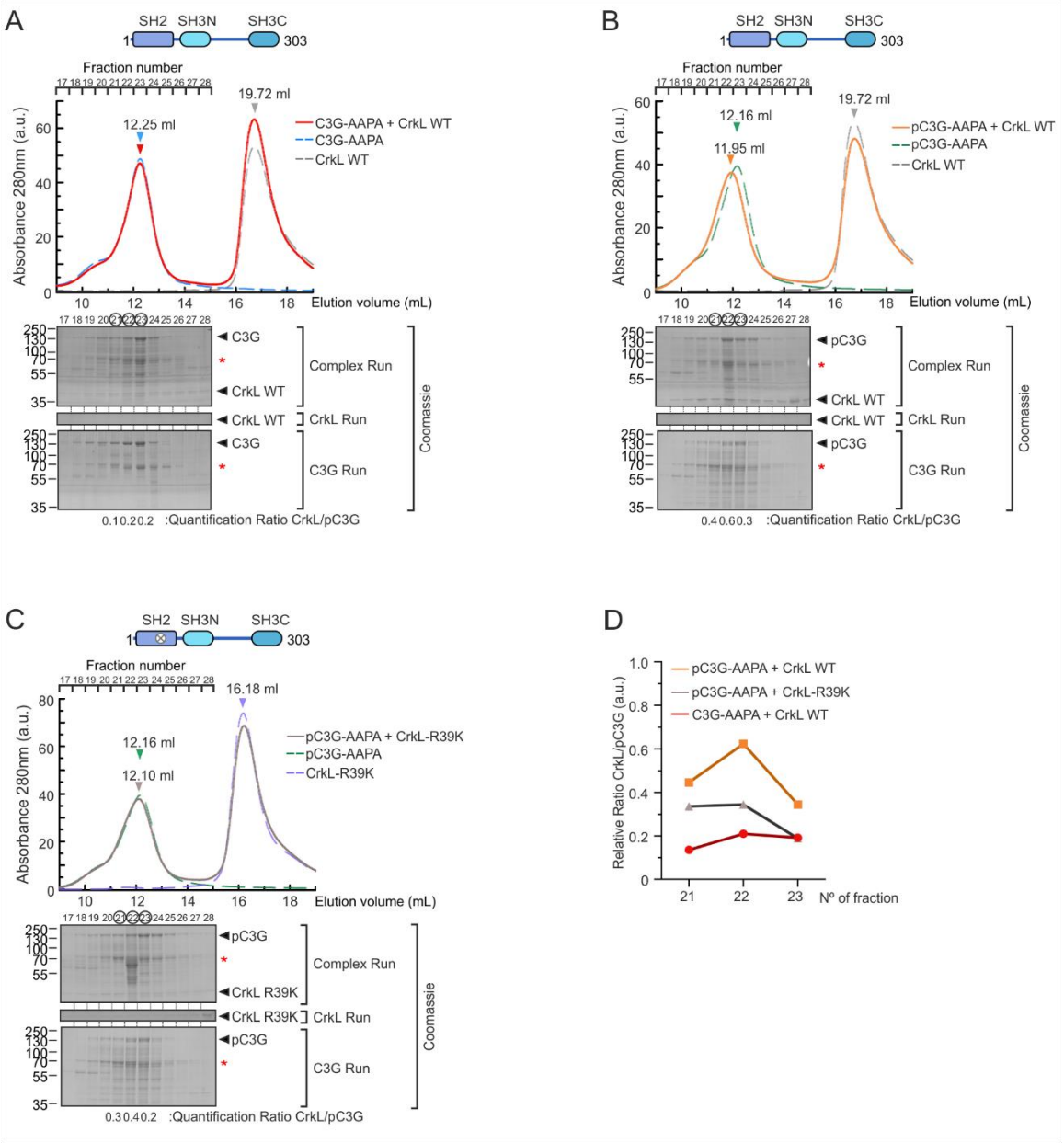



**Figure R18. Analysis by size exclusion chromatography of the contribution of phosphorylation to the interaction between CrkL and C3G. (A)** Size exclusion chromatography profiles of isolated CrkL wild type (80 µM, dashed grey line) and C3G-AAPA (20 µM, dashed blue line), and mixtures with a 4:1 molar ratio of CrkL and C3G-AAPA (red line). Superdex 200 HR 10/300 column (GE Healthcare) equilibrated in 20 mM Tris (pH 7.5), 300 mM NaCl was used. Proteins were loaded (100 µL) at the indicated concentrations. The elution volumes of the peaks are indicated. Proteins in the fractions of interest were concentrated by precipitation and were analyzed by SDS-PAGE and Coomassie staining, shown aligned below the chromatograms. **(B)** Elution profiles of isolated CrkL wild type (dashed grey line) and Src-phosphorylated pC3G-AAPA (dashed green line), and mixtures with a 4:1 molar ratio (orange line) are shown as in A. **(C)** Analysis of CrkL-R39K (dashed violet line), pC3G-AAPA (dashed green line), and 4:1 molar mixture (grey line) as in A. Red asterisks in A, B and C mark degradative products. **(D)** Representation of the C3G/CrkL signal ratio in the fractions 21, 22 and 23, (first peak) of the runs of the mixtures.

### 3. Structural basis of the autoinhibitory interaction of C3G: prediction and experimental validation

The main autoregulatory mechanism of C3G is the intramolecular interaction between the final segment of the SH3b, named the AIR, and the catalytic Cdc25H domain. This interaction blocks the GEF activity of C3G in resting conditions. The AIR has been previously mapped to the region 545-646, and two functionally distinct segments have been identified within this region (Carabias et al., 2020, Carabias, 2019). The first part (residues 545-569) drives the interaction with the Cdc25HD; hence it is called the Cdc25HD-binding region or AIR-CBR but is insufficient to block the GEF activity. A second segment of the AIR (residues 570-646), named the inhibitory tail or AIR-IT, is required to inhibit the Cdc25HD; yet, by itself does not support binding to the Cdc25HD. In order to better understand the autoinhibition of C3G and the mechanism leading to its activation, we pursued the characterization of the structural basis of C3G autoinhibition.

Multiple attempts to crystallize autoinhibited forms of C3G were unsuccessful. Therefore, we applied novel structural prediction methods based on artificial intelligence (AI) such as AlphaFold2 (Jumper et al., 2021), which appeared during this project.

Because the prediction of the structure of full-length C3G exceeds the available computing capabilities of Colab Fold, we focused on the second half of C3G (residues 501-1077) that contains the AIR, the P3 and P4 motifs, the REM domain and the catalytic Cdc25HD. We had previously shown that a similar fragment, 530-1077, is autoinhibited and it is activated by Src-phosphorylation and CrkL binding (Carabias, 2019, Carabias et al., 2020). We generated five independent models of the 501-1077 segment of C3G with AlphaFold2 (Figure19A). The models were predicted with high global confidence, having a mean per-residue local distance difference test (pLDDT) over 75 in the five models (Figure19B); pLDDT is a confidence value in a scale range from 0 to 100. The recognizable globular domains (REM and Cdc25HD) were predicted with pLDDT scores over 90. However, the region 501-665 was predicted with lower confidence, with an average pLDDT score of ~40. Yet, within this region, the segment 550-592, showed higher pLDDT scores nearing 79.

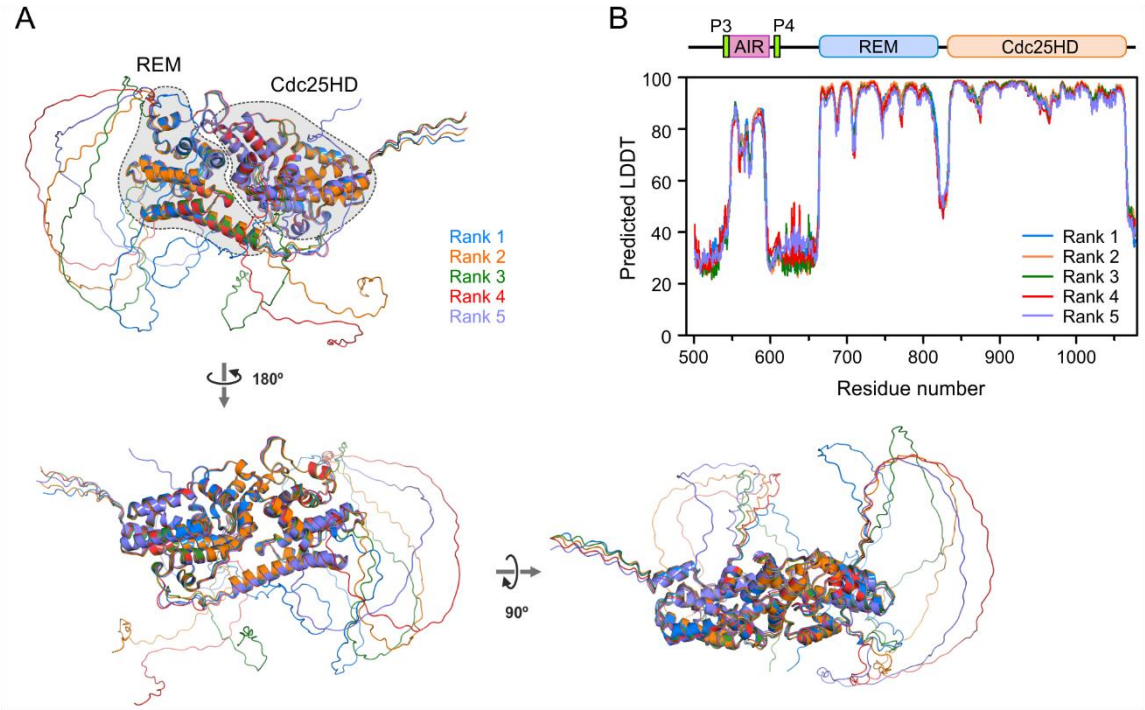



**Figure R19. Structural predictions of the C3G 501-1077 region by AlphaFold2. (A)** Three orthogonal views of a cartoon representation for the five predicted models of the C3G 501-1077 region. The five models were superposed and shown in distinct colors. **(B)** Graphical representation of the predicted LDDT scores per residue for each model. The colors of the lines match the colors of the models in A.

Superposition of the five generated models showed high structural similarity of the predictions in the REM and Cdc25H domains and the segment between the P3 and P4. After superimposition, the root-mean-square difference (rmsd) for the Cα atoms of the REM and Cdc25H domains (670-1063) was 1.09 Å, and for the segment between the P3 and P4 (549-595) it was 0.65 Å. The region 549-595 was predicted to adopt an extended structure that contains three α-helixes (α1, α2 and α3) and makes contacts with surfaces of the REM and Cdc25HD domains. Helix α1 (551-557) is positioned in a crevice of the Cdc25H domain that is predicted to interact with the Switch II of Rap1, as observed in the structures of SOS1/Ras and Epac2/Rap1 complexes (Boriack-Sjodin et al., 1998, Rehmann et al., 2008). We also observed a similar interaction between Rap1b and C3G in models obtained with AlphaFold2. Notable, Y554 is predicted to dock in a pocket of the Cdc25HD that would accommodate F64 of Rap1b; this interaction of Y554 is in agreement with its crucial role to sustain the binding to the Cdc25HD and the autoinhibition of C3G (Carabias et al., 2020). The second α-helix α2 (566-570) is predicted to cover part of the α-helical hairpin of the Cdc25HD, which is a key region for the binding of the GTPase. The third helix, α3 (576-590), was predicted to contact both the Cdc25HD and the REM. Finally, the regions 501-548 and 596-665,

which contain the P3 and P4 PRMs, respectively, were predicted to be flexible and disordered.

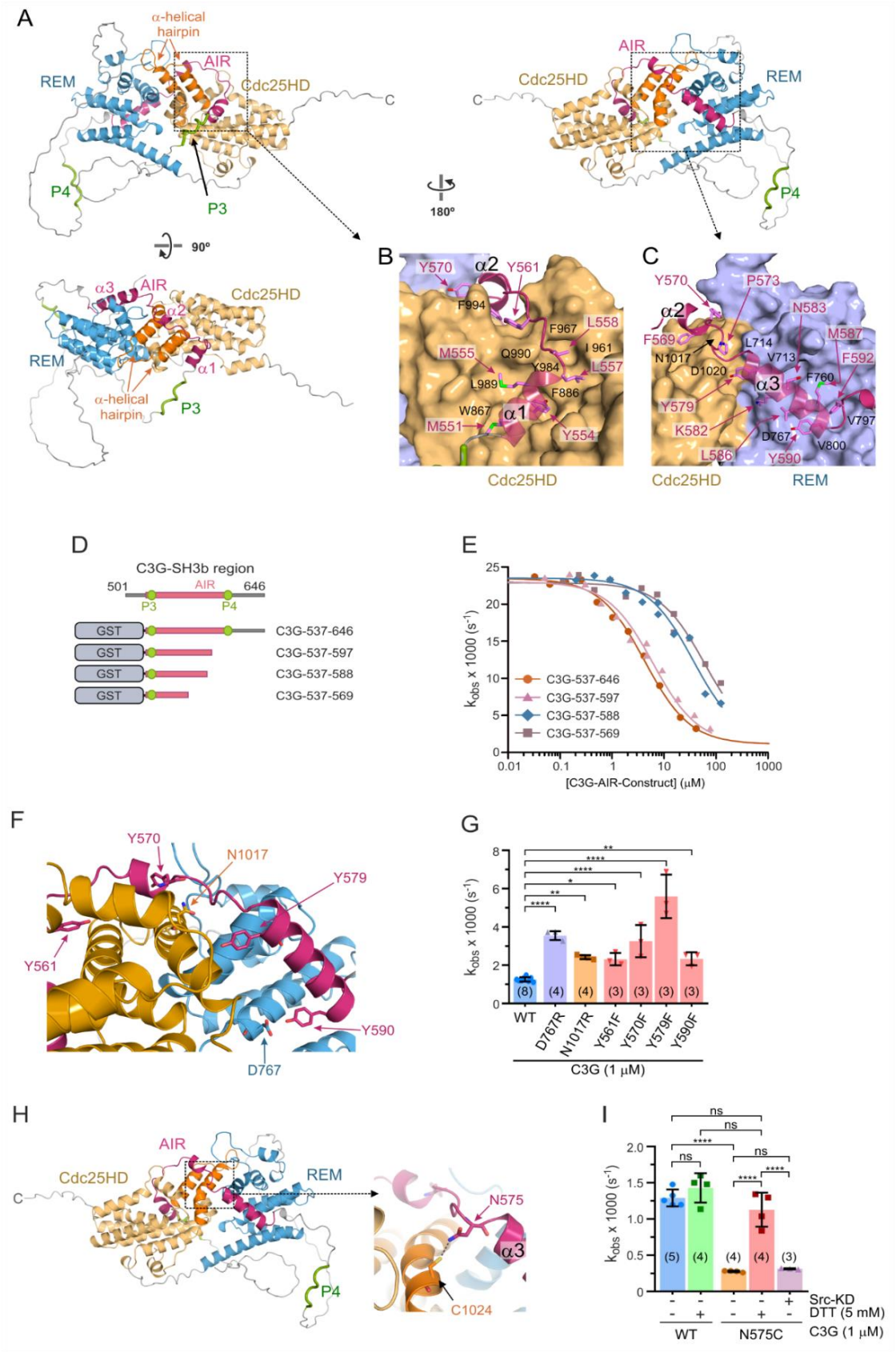


**Figure R20. Predicted structure of the autoinhibitory interaction in C3G and experimental validation of the model. (A)** Three orthogonal views of a cartoon representation of one predicted model of the C3G 501-1077 region. **(B)** Close-up view of the predicted interactions between the initial segment of the AIR (cartoon representation) and the GTPase-binding pocket of the Cdc25HD (surface). The side chains of the main interacting residues of the AIR are shown as sticks. **(C)** View in detail of the predicted interaction between the second half segment of the AIR (cartoon) and the REM and Cdc25H domains (surface). The main interacting residues of the AIR are shown as sticks. **(D)** Schematic representation of the C-terminal AIR deletion constructs used to inhibit the GEF activity of the C3G-Y554H mutant. **(E)** Dose-dependent inhibition of the GEF activity of C3G-Y554H mutant (1 μM) with the four constructs of the AIR in D. Lines represent the fitted sigmoidal models. **(F)** Close up showing the localization of Y561, Y570, Y579, Y590, D767 and N1017 in the autoinhibited model of C3G. **(G)** Nucleotide exchange rates of C3G wild type and mutants (1 μM). Data are shown as scatter plots with bars, means ± standard deviation. **(H)** View of the predicted localization of N575 and C1024 in the autoinhibited model of C3G. **(I)** Nucleotide exchange rates of C3G wild type and C3G-N575C mutant (1 μM) in the absence and presence of 5 mM DTT and Src-phosphorylation. Data are shown as scatter plots with bars, means ± standard deviation. The number of independent measurements in G and I are indicated in parentheses. Statistical comparison in G and I was analyzed using ANOVA followed by Tukey´s multiple comparisons test; * $P < 0.05$, ** $P < 0.01$, **** $P < 0.0001$, ns $P > 0.05$.

**Table R14. Parameters of dose-response inhibition of C3G-Y554H by GST-AIR fragments.**

| AIR fragments | $IC_{50}$ (μM)[a] | $k_{free}$ x1000 ($s^{-1}$) | $k_{min}$ x1000 ($s^{-1}$) |
|---|---|---|---|
| GST-C3G-537-646 | 4.2 ± 0.5 | 23.5 ± 0.3 | 1.1 ± 0.7 |
| GST-C3G-537-597 | 5.9 ± 0.7 | 23.0 ± 0.6 | 1.1[b] |
| GST-C3G-537-588 | 35.3 ± 1.2 | 23.3 ± 0.6 | 1.3 ± 2.7 |
| GST-C3G-537-569 | 58.3 ± 7.3 | 22.9 ± 0.4 | 1.1[b] |

[a] Data are presented as the fitted value ± the associated standard errors of the nonlinear regression fit.
[b] Values were fixed for analysis.

Our group had mapped the AIR to the segment 537-646; this region was identified as the minimal and sufficient segment that binds to and inhibits the GEF activity of the isolated Cdc25HD of C3G (Carabias et al., 2020). The AlphaFold2 prediction suggests that the region extending between the P3 and P4 PRMs, approximately residues 547-600, is sufficient for inhibition. In addition, the model also predicts that contacts between the last segment of helix α3 of the AIR and the REM contribute to the autoinhibition (Figure R20A-C).

To further validate the AlphaFold2 model, first we analyzed the limits of the AIR. We created four constructs of the AIR in which the AIR-IT was progressively trimmed from its C-terminus (537-646, 537-597, 537-588 and 537-569), these fragments were generated as GST fusion proteins (Figure 20D). Next, we carried out dose-dependent assays to analyze the ability of these fragments to block the GEF activity of the constitutive active mutant Y554H of full-length C3G. Nucleotide exchange reactions of

Rap1b:mant-dGDP (200 nM) were performed in the presence of C3G-Y554H (1 µM) at various concentrations of GST-C3G-AIR constructs. The GEF activity of C3G-Y554H showed a $k_{obs}$ 23 x $10^{-3}$ $s^{-1}$, which was 20 times faster than the basal GEF activity of C3G wild type, this value agrees with previous measurements (Carabias et al., 2020).

The largest construct, GST-C3G-537-646, inhibited C3G-Y554H with a half-maximal inhibitory concentration ($IC_{50}$) of ~4.2 µM (Figure R20E and Table R14). Similarly, GST-C3G-537-597 showed comparable inhibition of C3G-Y554H, with an $IC_{50}$ ~5.9 µM. In contrast, fragments 537-569 and 537-588 inhibited the GEF activity of C3G-Y554H at higher concentrations and showed $IC_{50}$ values around 10-fold higher (35 µM and 58 µM for 537-588 and 537-569, respectively), when compared with the longer fragments. Previous results from our group showed that fragments 537-569 and 537-588 only cause partial inhibition of the activity of the isolated Cdc25HD (Carabias et al., 2020); these apparent discrepancies can be explained by two experimental factors: (i) inhibition of the isolated Cdc25HD may be less efficient because the AIR cannot stablish the predicted contacts with the REM domain, and (ii) in that work inhibition was tested at 40 µM which is similar to the $IC_{50}$ of these fragments over C3G-Y554H, therefore, they could only cause partial inhibition.

Collectively, these results suggest that the fragment 537-597 is the minimal region capable of inhibiting the GEF activity of full-length C3G. This fragment includes the P3 motif and the segment between P3 and P4. We have shown that mutations in the P3, or any of the other PRMs (P1, P2 and P4), do not affect the autoinhibition of C3G. Therefore, we can infer that the AIR corresponds approximately to the region 548-597, which is the segment between the P3 and P4 PRMs. In conclusion, these results agree with the predicted structural models of the autoinhibitory mechanism of C3G.

Complementarily, we analyzed the contribution of specific residues to the autoinhibition of C3G. We assessed the effect on the GEF activity of point substitution of residues that are predicted to participate directly in the interaction between the AIR and the REM or Cdc25H domains. The residues of interest were mutated in C3G full-length and the basal GEF activity of these point mutants (1 µM) was compared with that of C3G wild type (i.e., autoinhibited). The models predict that the side chains of Y561, Y570, Y579 and Y590 dock in pockets in the catalytic region. We have shown that the mutants C3G-Y561F, C3G-Y570F, C3G-Y579F and C3G-Y590F have significantly higher activity than the wild type protein (Figure R18, data also presented in Figure R23F-G), and the quadruple mutant C3G-Y561-570-579-590F showed an even higher basal GEF activity (Figure R18C). These data suggest that the tyrosine residues in the AIR directly

contribute to the stability of the autoinhibited conformation of C3G and support the structural prediction. The cumulative weakening of the autoinhibition by the combination of multiple tyrosine-to-phenylalanine substitutions suggests that these contacts behave like a Velcro.

In addition, we created two-point mutants that target residues in the REM domain (D767R) and the Cdc25H domain (N1017R). Both, D767 and N1017, are predicted to contact residues of the AIR. The basal activity of C3G-D767R and C3G-N1017R was also significantly higher than C3G wild type (Figure R20F-G), supporting the notion that these two residues contribute to the interaction with the AIR as predicted in the AlphaFold2 models.

Finally, we designed a point substitution aimed to lock the interaction of the AIR with the Cdc25HD. Based on the AlphaFold2 model of C3G, the C1024 of the Cdc25HD is part of the AIR-binding interface, and C1024 would be near to the N575 of the AIR (Figure R20H). We introduced the mutation N575C, which according to the AlphaFold2 prediction would facilitate the formation of a disulphide bridge with C1024. This disulphide bridge would lock the closed conformation and would reduce the basal GEF activity of C3G. Indeed, C3G-N575C (1 µM) showed a basal activity ($k_{obs}$ 0.3 x $10^{-3}$ $s^{-1}$) that was 4.5 times lower than that of C3G wild type (Figure R20I). In presence of the reducing agent DTT (5 mM), the GEF activity of C3G-N575C increased significantly, reaching a similar activity as C3G wild type. Noteworthy, the activity of C3G wild type did not change in the presence of 5 mM DTT. Interestingly, Src-phosphorylation of the C3G-N575C did not affect its basal activity. These data suggest that N575C stabilizes the autoinhibited conformation of C3G in a redox reversible manner; this is compatible with the formation of a disulphide bond, possibly with C1024, as observed in the predicted structure.

Collectively, our results provide strong experimental validation for the AlphaFold2 prediction of the structure of autoinhibited C3G. The AIR has been mapped with higher precision than before. This model revealed a role of the REM domain in the autoinhibitory interaction and the direct contribution of tyrosine residues in the AIR to the contacts with the REM-Cdc25HD catalytic region. The model allows a structural interpretation of the functional data of the autoinhibition and activation mechanisms.

## 4. Identification of novel cancer somatic mutations that target the autoinhibition of C3G

Previous work demonstrated that two missense mutations found in non-Hodgkin´s lymphoma patients in residues of the AIR (Y554H and M555K) caused a constitutive activation of C3G and Rap1 *in vitro* and in cells (Carabias et al., 2020). These mutants are located in the AIR-CBR, and they were identified to disrupt the AIR/Cdc25HD interaction. Based on the AlphaFold2 models, they are predicted to locate inside the GTPase binding pocket in the helix α1. With the current information about the structural basis of the autoinhibition of C3G, we investigated if other missense mutations affecting residues in the AIR-IT could also affect the autoinhibition.

The COSMIC database contains around 306 missense somatic single nucleotide variants (SNVs) in the *RAPGEF1* gene described in samples of patients with cancer. 26 of them are in the region 548-597 (Table R15). Six residues mutated in cancer patients are predicted in contact with the REM and Cdc25H domains (L557, L558, Y570, Y579, M587 and Y590). We focused on the mutations of tyrosine residues implicated in the Src-phosphorylation of C3G. Y579C and Y590N mutations were found in two patients with thyroid carcinoma and follicular lymphoma, respectively (Pasqualucci et al., 2014). A second mutation in the tyrosine residue 590, Y590F, was found in a patient with cutaneous squamous cell carcinoma (Pickering et al., 2014). Additionally, the Y570N mutation was also evaluated because of two reasons, first it is a residue predicted in the helix α2 (different of the other mutants, Y554H and Y579C or Y590N that are located in the helix α1 and α3, respectively), and secondly Y570N mutant was found in the same patient with the M555K mutant that developed non-Hodgkin´s lymphoma.

**Table R15**. **RapGEF1 SNVs in the AIR described in COSMIC database[a]**.

| Mutation (AA)[b] | CDS mutation (CDS) | Primary tissue | Cancer type | Sample ID[c] |
|---|---|---|---|---|
| N548S | c.1694A>G | Large intestine | Adenocarcinoma | TCGA-AA-A00N-01 |
| **Y554H** | c.1711T>C | Haematopoietic and lymphoid | Follicular lymphoma | LPJ021 |
| **M555K** | c.1715T>A | Haematopoietic and lymphoid | Diffuse large B cell lymphoma | DLBCL-PatientM |
| L557S | c.1721T>C | Skin | Superficial spreading | PD42123a |
| L558P | c.1724T>C | Haematopoietic and lymphoid | Acute lymphoblastic B leukemia | SJBALL021894 |
| P564L | c.1742C>T | Large intestine | Adenocarcinoma | T352983 |
| **Y570N** | c.1759T>A | Haematopoietic and lymphoid | Diffuse large B cell lymphoma | DLBCL-PatientM |
| **Y570F** | c.1760A>T | Lung | Squamous cell carcinoma | TCGA-63-7020-01 |
| Q571R | c.1763A>G | Lung | Adenocarcinoma | TCGA-55-8302-01 |
| Q571H | c.1764G>T | Lung | Small cell carcinoma | S02139_1 |
| T572M | c.1766C>T | Small intestine | Adenoma | 66 |
| T572M | c.1766C>T | Lung | Adenocarcinoma | LUAD-RT-S01702 |
| E576K | c.1777G>A | Large intestine | Adenocarcinoma | 587332 |
| H577L | c.1781A>T | Lung | Small cell carcinoma | SCLC20_P_LU |
| H577L | c.1781A>T | Lung | Small cell carcinoma | SCLC20_R_LN |
| I578M | c.1785C>G | Stomach | Carcinoma | GCTK_508_T |
| **Y579C** | c.1787A>G | Thyroid | Carcinoma | 77C |

| Mutation (AA)[b] | CDS mutation (CDS) | Primary tissue | Cancer type | Sample ID[c] |
|---|---|---|---|---|
| **Y579C** | c.1787A>G | Thyroid | Neoplasm | PTC-77C |
| Q580K | c.1789C>A | Skin | Cutaneous melanoma | TCGA-D3-A2JL-06 |
| K584N | c.1803G>T | Large intestine | Adenocarcinoma | TCGA-AA-3966-01 |
| K584N | c.1803G>T | Stomach | Adenocarcinoma | TCGA-BR-6454-01 |
| L585R | c.1805T>G | Pancreas | Carcinoma | PCSI_0611_Pa_P_526 |
| L585R | c.1805T>G | Pancreas | Carcinoma | PCSI_0611_Pa_X_526 |
| M587T | c.1811T>C | Endometrium | Endometrioid carcinoma | TCGA-BS-A0UV-01 |
| **Y590N** | c.1822T>A | Haematopoietic and lymphoid | Diffuse large B cell lymphoma | 12_tFL |
| **Y590F** | c.1823A>T | Skin | Squamous cell carcinoma | CSCC-38-T |

[a] Version 98 accessed in September 2023.
[b] Residue numbers correspond to C3G isoform a. Numbering in the COSMIC web-portal corresponds to a longer variant (Uniprot Q13905-3).
[c] Sample ID: code of the patient in which the SNVs were found.

Initially, we analyzed the effect of the point mutations Y570N, Y579C and Y590N on the autoinhibition of C3G *in vitro* by measurement of the GEF activity (Figure R21A-B). We analyzed the Y570N, Y579C and Y590N mutants, and wild type C3G and C3G-Y554H mutant, as references. The GEF activity of the three mutants was between 7 to 20 times higher compared to wild type C3G. C3G-Y579C had the highest GEF activity that was similar to the C3G-Y554H mutant (Y579C, $k_{obs}$ 20.4 x $10^{-3}$ $s^{-1}$; Y554H, $k_{obs}$ 21.8 x $10^{-3}$ $s^{-1}$). In the meantime, mutations Y570N and Y590N showed an intermediate activation (Y570N, $k_{obs}$ 7.6 x $10^{-3}$ $s^{-1}$; Y590N, $k_{obs}$ 15.3 x $10^{-3}$ $s^{-1}$).

Complementary, the GEF activity of these mutants was analyzed in HEK293T cells by Rap1 activation assay (Figure R21C). HEK293T cells were transiently transfected with either mEGFP alone as a negative control, or constructs expressing wild type and the mutants of C3G-mEGFP. In addition, the mutant Y554H was included as positive control. Cells expressing C3G WT and Y570N showed similar Rap1-GTP levels, C3G-Y579C and Y590N increased slightly the levels of Rap1-GTP and C3G-Y554H caused significant high level of the active form of Rap1, according to previous results (Carabias et al., 2020).

In summary, these results reveal that residues outside the AIR-CBR are also important to maintain the autoinhibition of C3G and expand the number of cancer-related point mutations that cause dysregulation of C3G GEF activity.

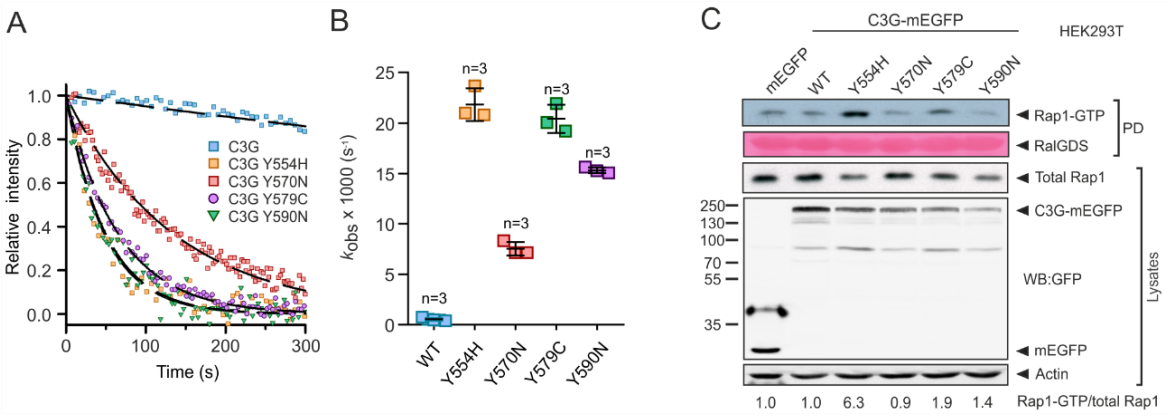



**Figure R21. Additional somatic missense mutations disrupt the autoinhibitory mechanism of C3G *in vitro* and in cell cultures. (A)** Representative exchange reactions of Rap1:mant-dGDP (200 nM) catalyzed by the presence of 1 µM C3G wild type, Y554H, Y570N, Y579C and Y590N. Lines are the single exponential decay models fitted to obtain the $k_{obs}$. **(B)** Scatter plot of estimated nucleotide exchange rates of C3G wild type and mutants (1 µM). Data are shown as scatter plots with bars, means ± standard deviation. **(C)** Analysis of Rap1 activation in HEK293T cells transiently expressing C3G-mEGFP wild type and mutants, or isolated mEGFP as control. Levels of expression of endogenous actin, total Rap1, exogenous C3G-mEGFP and isolated mEGFP were detected in cell lysates using Western blot (WB) with specific antibodies. Active Rap1 (Rap1-GTP) was pulled down by the effector protein RalGDS tagged with GST. The amount of GST-RalGDS used in pull-down (PD) was analyzed by staining the membrane with Ponceau S stain. Rap1-GTP was detected with specific anti-Rap1 antibody in the PD samples.

## 5. Role of the interaction of CrkL with the P1 and P2 sites

The PRMs P1 and P2 of C3G are accessible for the binding of CrkL *in vitro*, in the resting or autoinhibited state of C3G. In addition, P1 and P2 are dispensable for the direct stimulation of the GEF activity of C3G and pC3G by CrkL (see above). This prompted us to analyze the possible role of P1 and P2 in the Crk-mediated recruitment of C3G during activation in cells. Initially, we studied the role of the PRMs in the interaction of C3G with CrkL in Jurkat cells, where CrkL associates with C3G (Reedquist et al., 1996). We used lentiviral transduction to create Jurkat cells with a stable expression of C3G (wild type and PRM mutants) fused to a C-terminal mEGFP tag, and control cells expressing mEGFP alone. We analyzed the interaction of endogenous CrkL with C3G-mEGFP by co-immunoprecipitation (co-IP) using anti-GFP affinity resin (Figure R22A). CrkL co-immunoprecipitated with wild type C3G-mEGFP and with the mutant that contains the wild type P1 and P2 motifs (C3G-PPAA-mEGFP); yet CrkL was not detected in co-IPs of cells expressing C3G-AAPP-mEGFP. The single PRM mutants C3G-PAAA-mEGFP and C3G-APAA-mEGFP also immunoprecipitated CrkL, indicating that CrkL interacts with both the P1 and P2 motifs in Jurkat cells.

We also analyzed the interaction between CrkL and the C3G-mEGFP PRM mutants in HEK293T cells (Figure R22B). In this case, HEK cells were transiently transfected with constructs expressing wild type and the same PRMs mutants of C3G-mEGFP, or mEGFP alone as a control, and the interaction was analyzed by anti-GFP co-IP. Similarly, as observed in Jurkat cells, CrkL interacts strongly with C3G wild type and the mutant C3G-PPAA-mEGFP. CrkL detection was lower in co-IPs with C3G-AAPP and the single-PRM mutants C3G-PAAA and C3G-APAA.

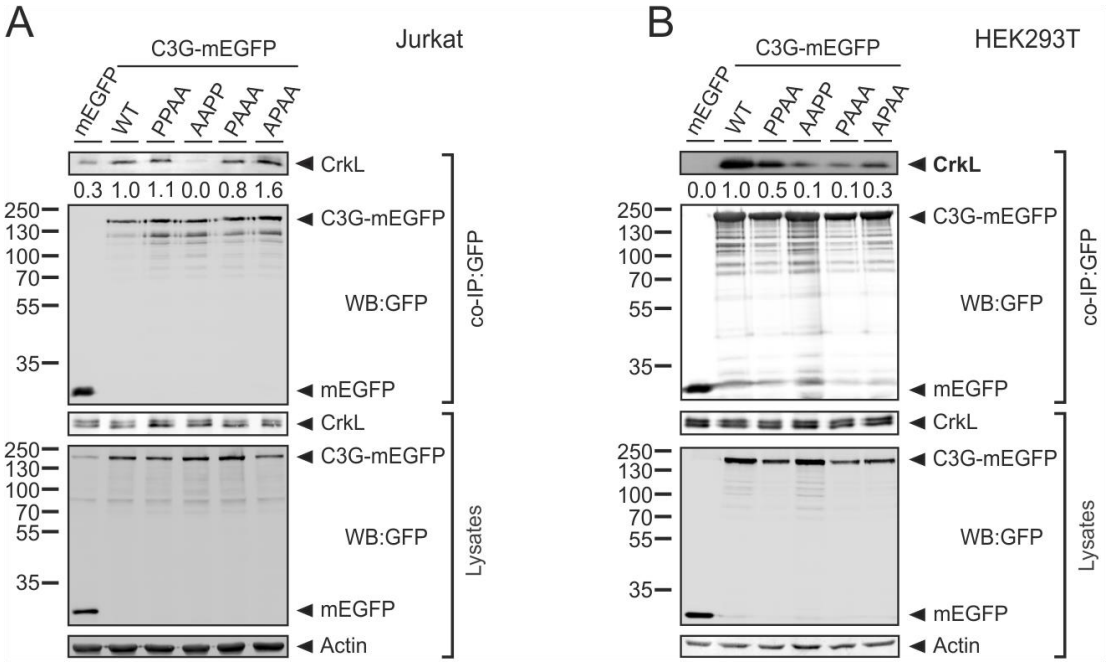



**Figure R22. CrkL interacts with the P1 and P2 sites of C3G in cells. (A)** Analysis by co-immunoprecipitation (co-IP) in Jurkat cells of the interaction between stably expressed exogenous C3G-mEGFP variants and endogenous CrkL. Proteins were immunoprecipitated with affinity resin against GFP. CrkL and mEGFP-tagged proteins were detected in cell lysates and in the co-IP by Western blot (WB). **(B)** Analysis of the interaction between C3G-mEGFP and CrkL in HEK293T cells. mEGFP-tagged C3G WT and mutants were transiently expressed and the interaction with endogenous CrkL was analyzed by co-IP as in A.

Crk proteins interact with the PRMs P1 and P2 in cells, and P3 and P4 participated in the activation of C3G *in vitro*. With this information, we were interested in understanding the contribution of the interaction of Crk protein with the two distinct pairs of PRMs of C3G during the activation of Rap1 in cells. In Jurkat cells, Crk and C3G are required for the activation of Rap1 upon ligation of the T-cell receptor (TCR), in a process that includes the recruitment of the complex CrkL-C3G to the plasma membrane (Nolz et al., 2008).

We analyzed the translocation of C3G to the cell periphery in Jurkat cells. We used confocal microscopy to analyze the localization of C3G-mEGFP wild type, C3G-PPAA-mEGFP, and C3G-AAPP-mEGFP, before and 30 min after ligation with CD3 (Figure R23). We used phalloidin staining of cortical actin as a marker for the plasma membrane at the resolution of conventional confocal microscopy (Manso et al., 2019).

Wild type C3G-mEGFP and C3G-PPAA-mEGFP co-localization with cortical actin increased significantly upon CD3 ligation, suggesting that they were recruited to the plasma membrane. In contrast, C3G-AAPP-mEGFP mutant did not show significant differences in the co-localization with cortical actin before and after CD3 ligation.

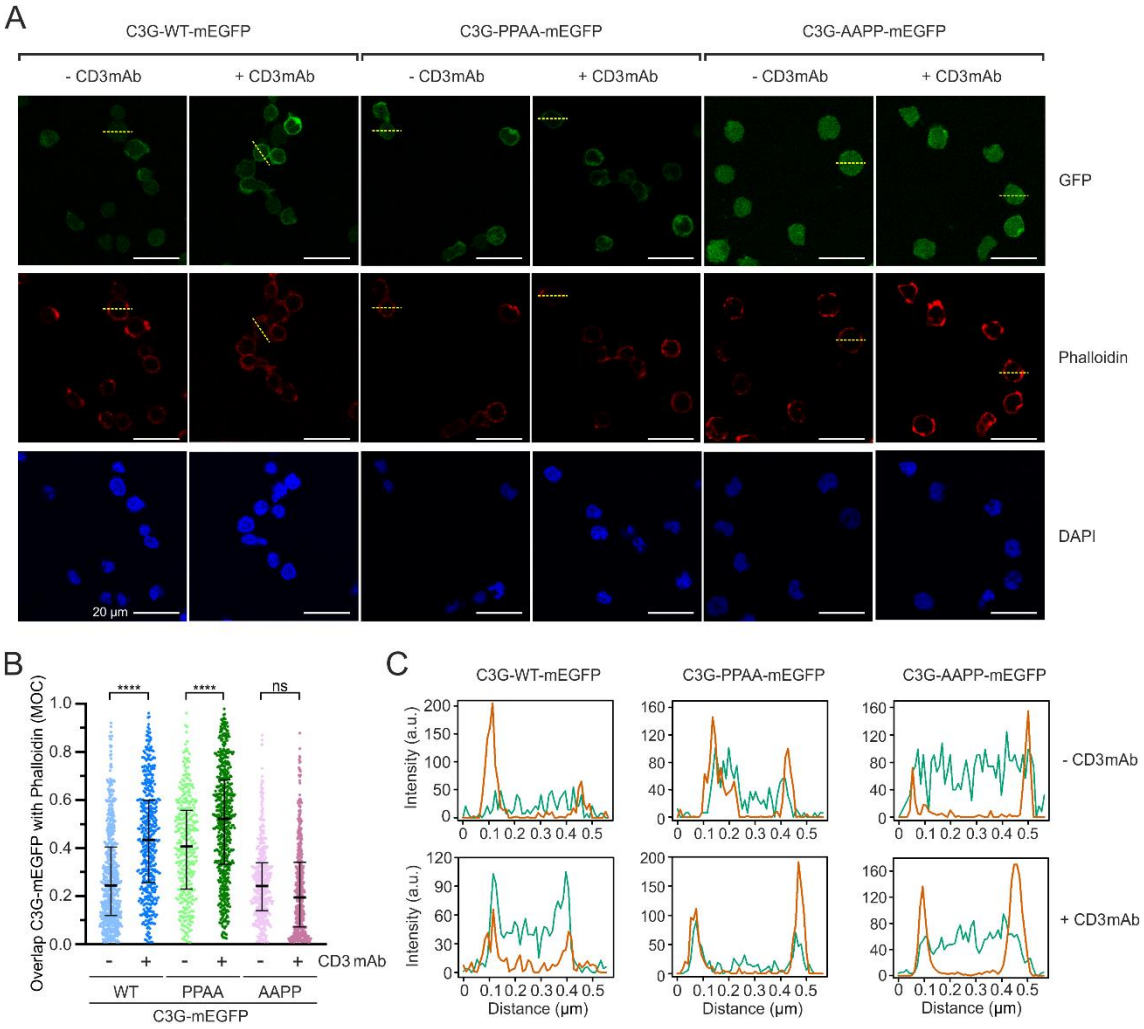



**Figure R23. Role of PRMs in the recruitment of C3G to the plasma membrane. (A)** Imaging of C3G-mEGFP (upper panels), and cortical actin (stained with phalloidin-iFluor 647, middle panels) in Jurkat cells expressing C3G wild type or the indicated mutants. Nuclei were stained with DAPI (lower panels). Cells were plated on coverslips treated with poly-L-lysine and were unstimulated (- CD3mAb) or stimulated with OKT-3 antibody against CD3 (+CD3mAb). Representative fields are shown, in which the median of the Manders´ overlap coefficient (MOC) for the cells in each field is similar to the value observed for the total of cells analyzed. Scale bars, 20 µm. **(B)** Quantification of the co-localization of C3G-mEGFP with phalloidin-

stained actin in confocal microscopy images of Jurkat cells as shown in A. MOC values are shown as scatter plots (from left to right, n = 637, 466, 346, 565, 423 and 459 cells). Middle bars mark the median and whiskers are the 25th and 75th percentiles. Statistical analysis was done with non-parametric Kruskal-Wallis and Dunn´s multiple comparisons tests (**** $P < 0.0001$, ns $P > 0.05$). **(C)** Fluorescence intensity of C3G-mEGFP (green lines) and phalloidin staining (orange lines) along the yellow dashed lines in B, which are selected to run across representative cells of the experimental conditions.

Next, to understand the contribution of the different PRMs to the activation of C3G in cells, we analyzed the activation of Rap1 upon ligation of TCR in Jurkat cells expressing the previous constructs used in the co-localization assay (wild type C3G-mEGFP, C3G-PPAA-mEGFP and C3G-AAPP-mEGFP). These cells also expressed endogenous C3G, which can participate in the activation of Rap1. The contribution of endogenous C3G to Rap1 activation was observed in control cells that expressed mEGFP alone, in which Rap1-GTP increased upon ligation of CD3 (Figure R24). Cells expressing wild type C3G-mEGFP showed a stronger activation of Rap1, which corresponds to the contribution of the exogenous C3G. Rap1-GTP was not noticeably increased upon CD3 ligation in cells expressing mutant C3G-PPAA-mEGFP. This can be interpreted as a dominant negative effect of C3G-PPAA-mEGFP that binds to CrkL and reduces the accessibility of CrkL to the endogenous C3G, but this mutant cannot be activated. Finally, in Jurkat cells expressing C3G-AAPP-mEGFP the increase in the levels of active Rap1-GTP upon stimulation was similar to that observed in the cells expressing mEGFP alone, in agreement with the inability of this mutant to bind to CrkL and be recruited to signaling sites by CrkL; in this case the activation of Rap1 would be produced by endogenous C3G.

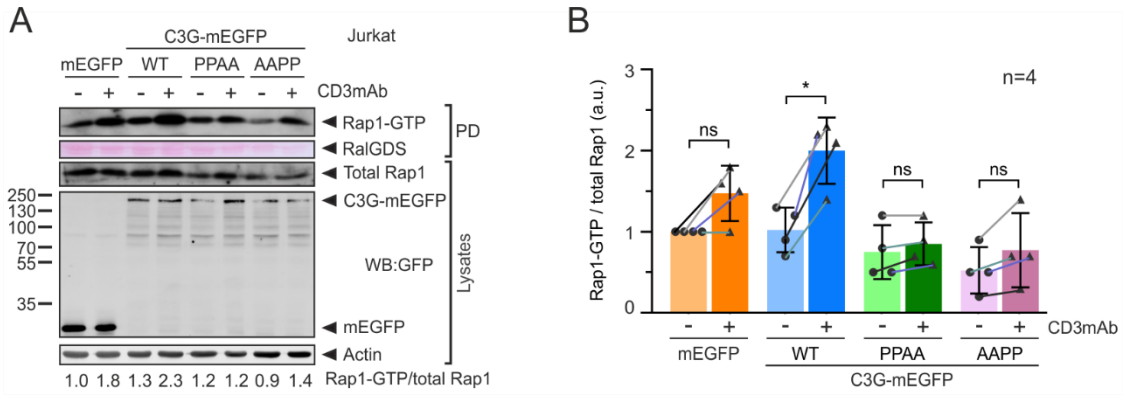



**Figure R24. Role of PRMs in the recruitment of C3G to the plasma membrane. (A)** Representative assay of Rap1 activation in Jurkat cells stably expressing C3G-mEGFP WT and mutants, or isolated mEGFP as control, before and after stimulation with antibody against CD3. Levels of expression of endogenous actin, total Rap1, exogenous C3G-mEGFP WT, mutants and isolated mEGFP were detected

in cell lysates using Western blot (WB) technique and specific antibodies for each protein. Active Rap1 (Rap1-GTP) was pulled down by the effector protein RalGDS tagged with GST. The amount of GST-RalGDS used in pull-down (PD) was analyzed by staining the membrane with Ponceau S stain. Rap1-GTP was detected with specific anti-Rap1 antibody in the PD samples. **(B)** Quantification of the ratio Rap1-GTP/total Rap1 from four independent experiments (a.u. arbitrary units). The ratio in non-stimulated cells expressing mEGFP was set to 1 for normalization. Lines represent the data of each experiment. Statistical comparison was analyzed using ANOVA followed by Tukey´s multiple comparisons tests, * $P < 0.01$, ns $P > 0.05$.

In summary, the four PRMs of C3G have distinct and specialized roles during the C3G-mediated activation of Rap1 in Jurkat cells. Sites P1 and P2 are necessary for the recruitment of C3G to the plasma membrane, while sites P3 and P4 are essential for the direct stimulation of the GEF activity; both processes are mediated by the interaction with Crk adaptor proteins.

# DISCUSSION

To gain a comprehensive and thorough understanding of the spatiotemporal regulation of small GTPases, it is crucial to explore deeply the specific mechanisms controlling the autoregulation and activation of various GEFs. This complexity is exemplified by the Rap1 proteins, which respond to multiple stimuli in the cells by activating distinct GEFs. For example, in mice platelets, low thrombin stimulation triggers the activation of Rap1b through CalDAG-GEFI (RasGRP2), while C3G assumes the primary role as the Rap1GEF in response to high thrombin stimulation (Franke et al., 2000, Gutierrez-Herrero et al., 2020, Gutierrez-Herrero et al., 2012). We have elucidated the mechanisms of C3G activation and thereby expanded our understanding of Rap1 signaling.

**"Velcro" mechanism of autoinhibition of C3G**

We have obtained an experimentally validated structural model of the autoinhibitory interaction between the AIR and the REM-Cdc25H region of C3G. The AIR adopts an extended structure, consisting of three α-helices that block directly the GTPase binding site in the Cdc25HD. This model is compatible with the bi-partite mechanisms of inhibition by the AIR and provides a structural framework for that model (Carabias et al., 2020). Helix α1 (551-557) contains three residues essential for the binding of the AIR to the Cdc25HD: M551, Y554, and M555. Therefore, helix α1 is likely to correspond to the AIR-CBR, which is necessary and sufficient for binding to the Cdc25HD. Helix α1 occupies a pocket in the Cdc25HD that binds to the switch II region of the GTPase. This is in apparent contrast with the previous lack of inhibition by the AIR-CBR alone; we have now shown that segment of the AIR 537-569 does inhibit, although it requires higher concentrations. The second part of the AIR encompasses helices α2 (566-570) and α3 (576-590), which are predicted to establish contacts with both the Cdc25H and REM domains. These two helices are required for efficient inhibition of the GEF activity; therefore, they correspond to the inhibitory tail of the AIR (AIR-IT).

Noteworthy, the five tyrosine residues of the AIR engage in direct contacts with the catalytic region. The hydroxyl groups of these residues contribute to stabilize the autoinhibitory interaction, because substitution by phenylalanine has a cumulative activating effect (i.e., destabilization of the closed conformation). Two of these tyrosine residues, Y579 and Y590, are phosphorylated by Src *in vitro*. Therefore, the predicted model provides the structural basis of the partial activation caused by Src-

phosphorylation of C3G. Collectively, we have shown that the tyrosine residues of the AIR actively contribute to stabilize the autoinhibited conformation of C3G, and these interactions can be likened to the functioning of a Velcro-like mechanism.

The interaction between the AIR and the Cdc25HD of C3G shows similarities with autoinhibitory mechanisms of other Cdc25H GEFs. In RasGRP1 (CalDAG-GEFII) the short linker that connects the Cdc25HD to the EF hand blocks the GTPase binding site in the Cdc25HD (Iwig et al., 2013). The linker binds to a surface of the Cdc25HD equivalent to that used by the α1 helix of C3G, and RasGRP1 W454 of the linker docks in a pocket equivalent to the binding site of C3G-Y554. Despite these similarities, the inhibitory linker of RasGRP1 is stabilized by the interaction of the EF and C1 domains with an area of the Cdc25HD not involves in the binding of the GTPase. A similar autoinhibitory mechanism has been recently identified in the *Candida albicans* CaCdc25H GEF (Manso et al., 2023), where a short sequence upstream of the REM domain blocks the GEF activity. The inhibitory region of CaCdc25H is predicted to bind to the switch II binding site of the Cdc25H, with F850 of the inhibitory region occupying a position similar to that of C3G-Y554. Collectively, the switch II binding region of Cdc25H domains emerges as a potential regulatory site exploited by several GEFs as an autoinhibitory site. Neither the inhibitory segments of RasGRP1 nor CaCdc25 stablish contacts similar to those of the AIR-IT of C3G (helices α2 and α3), suggesting that this second part of the inhibitory interaction is unique to C3G.

Autoinhibition by direct blocking of the GTPases binding site has also been described in other non-Cdc25H GEFs. For example, in the Arf-GEF Grp1, the catalytic Sec7 domain is blocked by a linker, but this alone is not sufficient to completely inhibit the GEF activity, requiring an additional interaction through a C-terminal helix to fully secure the inhibitory mechanism (DiNitto et al., 2007, Sztul et al., 2019). Similarly, Rac-GEFs Vav1 proteins are inhibited by a helical linker that directly blocks the catalytic binding site, yet several interactions involving both the catalytic and non-catalytic domains are necessary to maintain autoinhibition (Yu et al., 2010) . Notably, much like C3G, the transition between the inactive and active conformations of Vav1 proteins are also tightly regulated by the phosphorylation status of three key tyrosine residues responsible for initiating their activation (Yu et al., 2010, Miletic et al., 2006, Amarasinghe and Rosen, 2005).

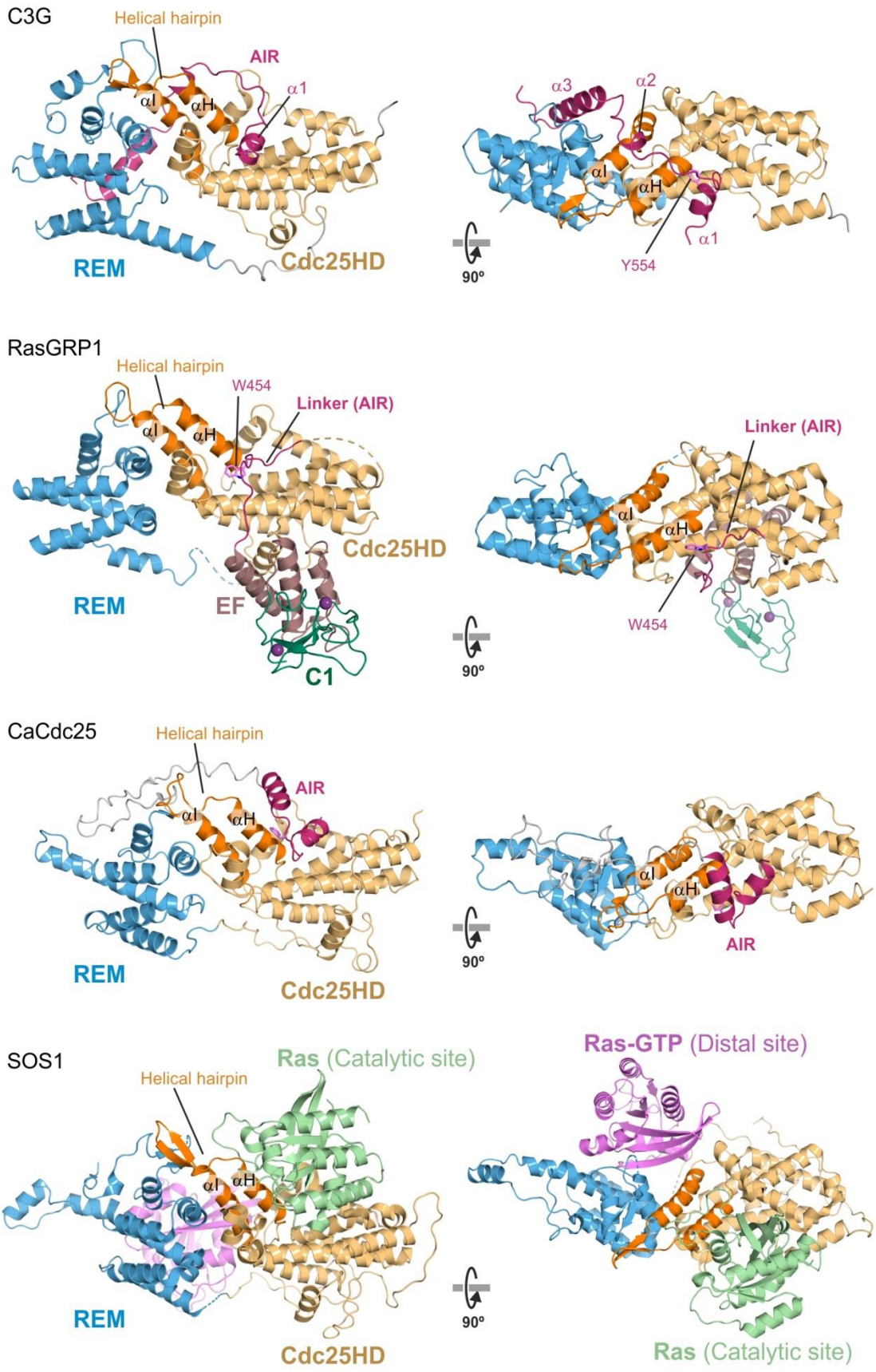



**Figure D1. Similarities of autoinhibition mechanisms in C3G and other Cdc25H GEFs.** Representation of the predicted structure of autoinhibited C3G (top). For clarity, only the AIR, REM and

Cdc25HD are shown. Crystal structure of autoinhibited RasGRP1 (PDB: 4L9M) (Iwig et al., 2013), AlphaFold2 model of the autoinhibition of CaCdc25H GEF (Manso et al., 2023), and crystal structure of the catalytic region of SOS1 with a Ras (nucleotide-free) molecule bound to the GEF catalytic site and a Ras-GTP molecule bound to the distal or allosteric site (PDB: 1NVU) (Margarit et al., 2003). All structures are shown in similar orientations. Equivalent domains and structural elements are colored in an analogous manner in all structures.

### Potential allosteric regulatory site in C3G

Comparison of the structure of autoinhibited C3G with structures of the catalytic region of SOS (REM and Cdc25H domains) in complex with two molecules of Ras revealed that the α3 helix of the AIR-IT of C3G occupies a similar position as the Ras-GTP molecule bound to the allosteric or distal site in SOS. Binding of the Ras-GTP to the distal site of SOS stimulates the GEF activity of the Cdc25HD through conformational changes transmitted by the REM domain. This stimulation of the activity of SOS by Ras-GTP is referred to as a positive feedback loop (Margarit et al., 2003). Until now no positive feedback mechanisms have been described in C3G, and no inter-molecular allosteric regulators of C3G are known. Yet, our group has shown that the NTD of C3G binds to the REM domain and favors the GEF activity, possibly through the stabilization of a catalytic-competent conformation of the Cdc25HD (Carabias et al., 2020). Mutagenesis analysis revealed that the NTD-binding area in the REM domain is different from the surface that interacts with the AIR-IT α3 helix. Should the surface of the REM that binds to α3 helix be involved in allosteric regulation of C3G, the autoinhibitory mechanisms would block simultaneously the catalytic site in the Cdc25HD and a putative regulatory site. Conversely, release of the autoinhibition would also expose such hypothetical regulatory site.

This observation suggests a potential feedback loop between the Ras and Rap1 signaling pathways. This loop may involve the binding of Ras-GTP molecules to the AIR-IT-binding site in REM domain of C3G. This would promote the stabilization of the active conformation of C3G, and consequently, the activation of the Rap1-GTP signaling pathway. While this interaction could lead to a competitive inhibition of SOS1 and the activation of Ras.

It is worth noting that Rap1 has previously been associated with antagonizing Ras signaling by trapping Ras effectors (Kitayama et al., 1989, Cook et al., 1993), particularly Raf1, in T lymphocytes (Boussiotis et al., 1997). However, two independent studies have demonstrated that the conversion of Ras-GTP to its active state is

effectively suppressed, suggesting that the inhibition of the GEF activity of Ras occurs upstream of Ras itself. This finding aligns with the hypothesis that C3G may obstruct the activation of Ras by trapping Ras-GTP, preventing the interaction and activation of its GEF, SOS, through the allosteric site mechanism.

**Crk proteins favor an active state of C3G**

Based on the present study, we propose that the GEF activity of C3G is modulated through the control of a conformational equilibrium. C3G would exist in equilibrium between an inactive or closed state in which AIR binds and blocks the Cdc25HD, and an active or open conformation in which AIR is dissociated from the Cdc25HD (Figure D2). In resting conditions, the equilibrium would be mostly shifted towards the closed conformation. However, the existence of a small proportion of active C3G before any stimulation is supported by the residual GEF activity observed in the unphosphorylated wild type C3G ($k_{obs}$ ~1 x $10^{-3}$ $s^{-1}$ $\mu M^{-1}$), and by the lower activity of the C3G-N575C point mutant ($k_{obs}$ ~0.3 x $10^{-3}$ $s^{-1}$ $\mu M^{-1}$) that apparently creates an intramolecular disulphide bridge with C1024, which traps the inactive conformation and breaks the conformational equilibrium. A comparable equilibrium between inactive and active states has been proposed for other Cdc25H GEFs such as Epac proteins, which are mediated by cAMP nucleotide binding (Rehmann et al., 2003).

The conformational balance of C3G is altered by the interaction of the Crk adaptor proteins and the phosphorylation in tyrosine residues (Figure D2). The PRM P3 is the main site for the activation of C3G by Crk proteins. The accessibility of CrkL to the P3 is linked to the activation state of C3G, being the P3 occluded in the closed conformation and exposed, at least partially, in the active state. The P3 (539-549) is situated adjacent to the α1 helix (549-555) the AIR, which corresponds to the region essential for binding to the Cdc25HD, known as AIR-CBR. Therefore, binding of Crk proteins to the P3 site in the closed conformation is likely hampered by steric hindrance. Mutations in P3 did not affect the autoinhibition of C3G, further supporting the structural prediction in which the P3 does not contribute directly to the interaction with the Cdc25HD. The exposure of P3, albeit not completely, occurs through the activating mutation Y554H, which disrupts the binding of the AIR-CBR to the Cdc25HD (Figure R2). When produced as independent proteins, the binding of CrkL to P3 displaces the interaction between AIR and Cdc25HD (Carabias et al., 2020). CrkL would only bind to P3 in the small fraction of C3G that adopts an open conformation, as indicated by the low stoichiometry observed in the CrkL/P3 interaction ((Carabias,

2019) and this work). In summary, the binding of Crk proteins to P3 appears to stabilize the active state, shifting the conformational equilibrium of C3G.

Binding of CrkL to the P4 is also necessary for optimal activation. Yet, the CrkL/P4 interaction alone cannot displace the AIR/Cdc25HD autoinhibitory interaction (Carabias et al., 2020) and does not stimulate C3G (Figure R3 and R4). It is reasonable to assume that CrkL binding to P4 contributes to the stabilization of the open state only when it is initially induced through the binding of CrkL to P3.

At high concentrations, that is, at high fractional levels of binding saturation, CrkL and CrkII activate unphosphorylated C3G. Yet, they only induce a maximum GEF activity that is 30-40% of the activity they stimulate in phosphorylated C3G. This suggests that Crk proteins alone have a limited capacity to stabilize the active state of unphosphorylated C3G; either they induce a partial activation state or the activation of a fraction of the C3G molecules.

In conclusion, the binding of Crk proteins to both P3 and P4 is a common and crucial event for the activation of C3G; therefore, we refer to the P3 and P4 as activation sites. However, the efficiency activation requires the tyrosine phosphorylation of C3G and depends on the specific Crk protein that binds to C3G.

**Phosphorylation of C3G by Src contributes to the activation in two ways**

Tyrosine phosphorylation of C3G by Src leads to multiple effects that collectively favor the activation of C3G by Crk proteins. Phosphorylation of C3G does not disrupt the interaction between the AIR-CBR and the Cdc25HD (Carabias et al., 2020), nor does it expose the P3 and P4 sites for CrkL binding (Figure R2). Out of the four Src-phosphorylation sites in C3G that we have identified, Y579 and Y590 are located in the α3 helix of the AIR, which is part of the AIR-IT region, and are predicted to contact with the Cdc25H and REM domains respectively (Figure R19-20). The slight increase of the GEF activity of pC3G compared to unphosphorylated C3G (Carabias et al., 2020) can be attributed to a partial disruption of interactions between the AIR-IT and the REM and Cdc25H domains. That is, Src-phosphorylation of C3G would cause a moderate displacement of the conformational equilibrium of C3G towards the open-active state, independently of the binding of Crk proteins. This also reflects in a higher activation of pC3G than unphosphorylated C3G by the isolated CrkL-SH3N domain. This effect is independent of the contributions of the SH2 and SH3C domains.

In addition, phosphorylation of C3G favors the activation by CrkL through a second mechanism that depends on the pY-binding site of the SH2 domain of CrkL.

CrkL enhances the GEF activity of pC3G to levels similar to those induced by point mutations in the AIR-CBR that disrupt the AIR/Cdc25HD autoinhibitory interaction, such as Y554R and Y554H. This suggests that akin to activating mutations, the ultimate effect of CrkL binding to pC3G is to displace the AIR/Cdc25HD autoinhibitory interaction and stabilize an open and active state.

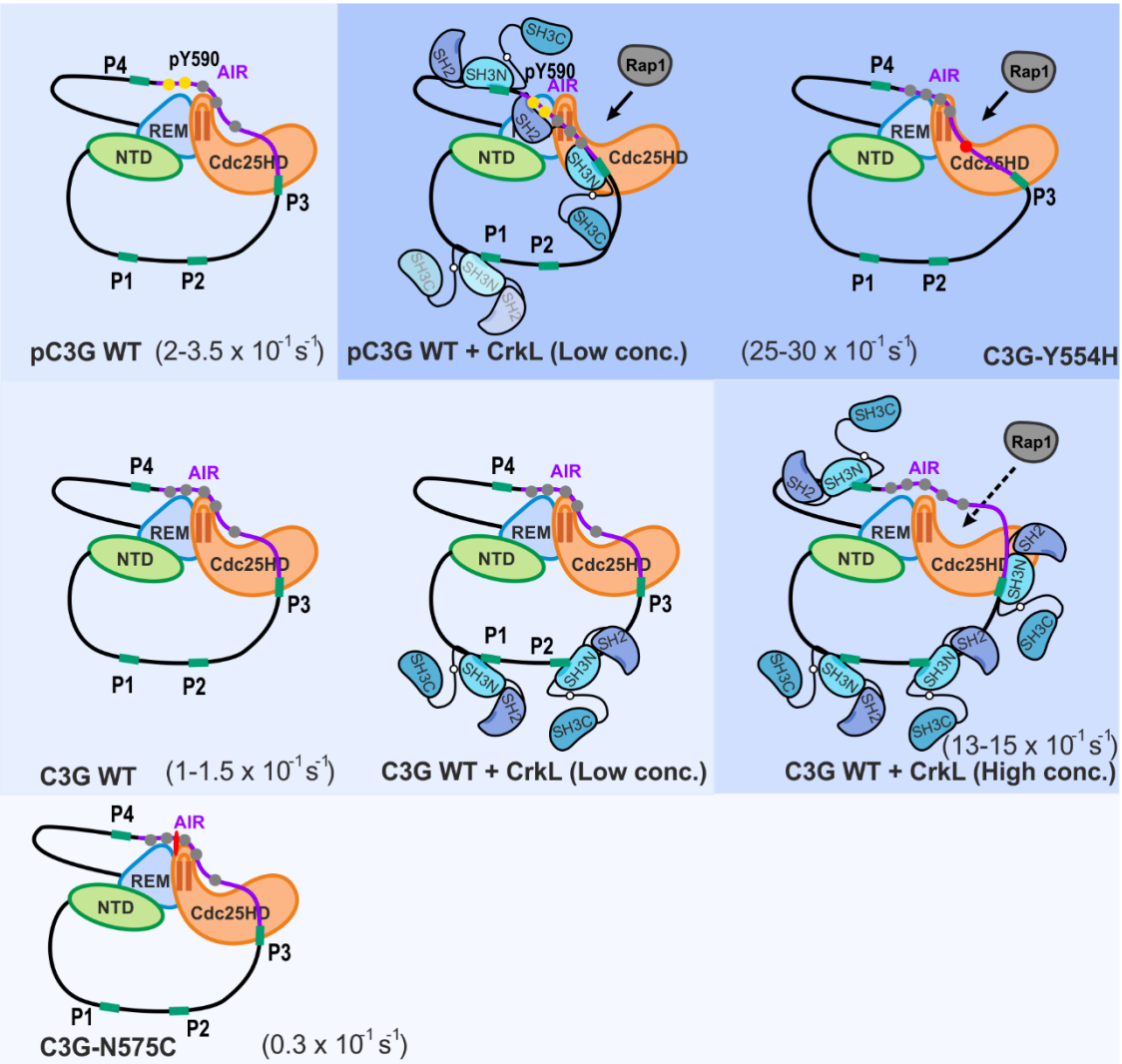



**Figure D2. Schematic representation of the conformational landscape of C3G.** Model of the different activation states of C3G in resting conditions and induced by phosphorylation, binding of CrkL, or engineered disulphide bride (N575C). Closed conformation is regulated by the AIR-Cdc25HD and AIR-REM interactions. In resting conditions (C3G WT) the equilibrium is mostly closed favoring to the stable AIR-Cdc25HD interaction. C3G-N575C traps the inactive conformation. CrkL-binding and phosphorylation in tyrosine residues alters the conformation balance of C3G favoring the release of the inhibitory interaction.

## Binding of CrkL-SH2 to a non-canonical pY motif of C3G is required for activation

We have unveiled a novel interaction between CrkL and C3G, which requires the CrkL-SH2 domain and phosphorylation of Y590 of C3G, suggesting a direct binding between these two elements. This interaction is crucial for the efficient activation of pC3G (Figure R9). The SH2 domain of CrkL binds preferably to phosphor-sites with the consensus sequence pYXXP; CrkL full-length CrkL and its isolated SH2 domain bind to a canonical phospho-peptide with $k_d$ values of 23 µM and 7 µM, respectively (Jankowski et al., 2012). However, our data reveals that the CrkL-SH2 domain binds to pC3G with considerably lower affinity ($k_d$ ~46 µM), which is consistent with the sequence context of Y590 ($^{590}$YGSF$^{593}$) not matching the preferred canonical site (Figure R17). In contrast, the CrkL-SH3N domain binds to the PRMs with high affinity ($k_d$ ranged between 1.4 and 3 µM). These results support the idea that the interacting SH2 domain belongs to a Crk molecule simultaneously engaging C3G through its SH3N domain. The initial binding of the SH3N to a PRM of C3G would facilitate the subsequent interaction of the companion SH2 domain, compared to the binding of a CrkL molecule relying solely on its SH2 domain. This notion of a bivalent interaction between Crk proteins with pC3G gains further support from our observations.

Collectively, Src-phosphorylation of C3G has two major contributions to the activation by Crk proteins. Firstly, phosphorylation enhances the sensitivity of C3G, reducing the concentration of Crk protein required for activation. This is evident from a ten-fold reduction of the CrkL $AC_{50}$ between unphosphorylated and Src-phosphorylated C3G (Figure R4 Table R3). Phosphorylation essentially couples CrkL binding to C3G activation; that is, the activation curve of pC3G by CrkL now matches the binding isotherm, the $AC_{50}$ (1.7 µM) is similar to the $k_d$ (1.4 and 3 µM). Secondly, phosphorylation increases 3 times the maximal activity induced by Crk proteins with respect to unphosphorylated C3G, as demonstrated by a $k_{max}$ value of 34.2 x $10^{-3}$ $s^{-1}$.

## Crk proteins activates C3G in an adaptor-independent mechanism

Crk proteins serve as adaptor molecules involved in the recruitment of proteins from the cytoplasm to the plasma membrane upon different stimuli. In an adaptor-dependent mechanism, Crk proteins use their SH2 domain to bind to phosphorylated tyrosine sites on scaffolding proteins, thereby facilitating the recruitment of effector proteins to the plasma membrane through SH3 domain interactions with PRMs.

Our results have demonstrated an additional role for Crk proteins in directly activating C3G. This activation process involves a dual interaction, with both the SH3N and SH2 domains of Crk proteins binding to pC3G, enhancing the activation process. The Crk molecule or molecules that engage in the dual interaction via the SH3N and SH2 domains do not function as a bridge between C3G and other tyrosine-phosphorylated proteins. As a result, the direct activation of pC3G by CrkL and CrkII represents, to the best of our knowledge, the first example of Crk proteins signaling that is independent of their typical adaptor function.

**Model for C3G activation**

Our results give information to present the following comprehensive multi-step mechanism for the regulation and physiological activation of C3G (Figure D3). Before stimulation, C3G predominantly resides in the cytoplasm, bound to Crk proteins (Okada et al., 1998, Buensuceso and O'Toole, 2000) through the P1 and P2 sites, which are constitutively available for binding (Carabias, 2019). Because the activation of unphosphorylated C3G occurs at concentrations of Crk protein ($AC_{50}$ is ~22.4 µM) that are over 7 times higher than those requirement for binding ($k_d$ ~3 µM) (Carabias, 2019) (Figure R3D, Table R2), there is an uncoupling between Crk-binding and Crk-mediated C3G activation. That is, there is significant binding of Crk proteins without significant activation of C3G. This, combined with the presumable relatively low abundance of Crk proteins in the cytoplasm, provides a mechanism for the presence of inactive Crk/C3G complexes. The preformed inactive complexes likely play a pivotal role in expediting rapid responses during activation.

The activation of C3G is triggered by signals that induce tyrosine kinase activity at the plasma membrane. At that moment, Crk molecules, tethered primarily through the SH3N domain to the accessible P1 and P2 sites of C3G (Figure R22), bind via their SH2 domain to tyrosine-phosphorylated proteins such as p130Cas, CasL, or paxillin (Maia et al., 2013, Vuori et al., 1996, Ohashi et al., 1998). These interactions result in the translocation of the Crk/C3G complexes to the signaling sites (Figure R23). Therefore, P1 and P2 are recruiting sites. Until this point, Crk proteins function as conventional adaptors, connecting C3G to phospho-proteins. In a second step, C3G is phosphorylated in tyrosine residues at the plasma membrane by Src-family kinases or c-Abl. The key role of phosphorylation is to facilitate the subsequent activation by Crk proteins in a final third step. Phosphorylation, mainly at Y590, enhances the sensitivity of C3G, reducing the Crk protein concentrations required for activation and increases

the maximal activity induced by CrkL and CrkII (Figure R4 and table R3). Binding of Crk proteins to C3G at signaling sites is likely facilitated by the recruitment of additional CrkL or CrkII molecules to the phospho-tyrosine sites that increase their local concentration. This facilitates the binding of Crk proteins to the activation sites P3 and P4. The interaction is further stabilized by the bivalent interaction through the SH3N and SH2 domains, which results in a release of the autoinhibition, the stabilization of C3G in an open and active state, and the local activation of Rap1 molecules.

Finally, the GEF activity of C3G is attenuated through the action of tyrosine phosphatases. It is known that the T-cell protein tyrosine phosphatase TC-PTP dephosphorylated C3G (Mitra et al., 2011), and the tyrosine phosphatase Shp2 inhibits the GEF activity of C3G in platelets (Gutierrez-Herrero et al., 2020). Dephosphorylation of C3G would lead to the rapid dissociation of CrkL, as observed in the dynamic nature of the CrkL/C3G interaction (Carabias, 2019), and results in the dispersion of inactive C3G throughout the cytoplasm.

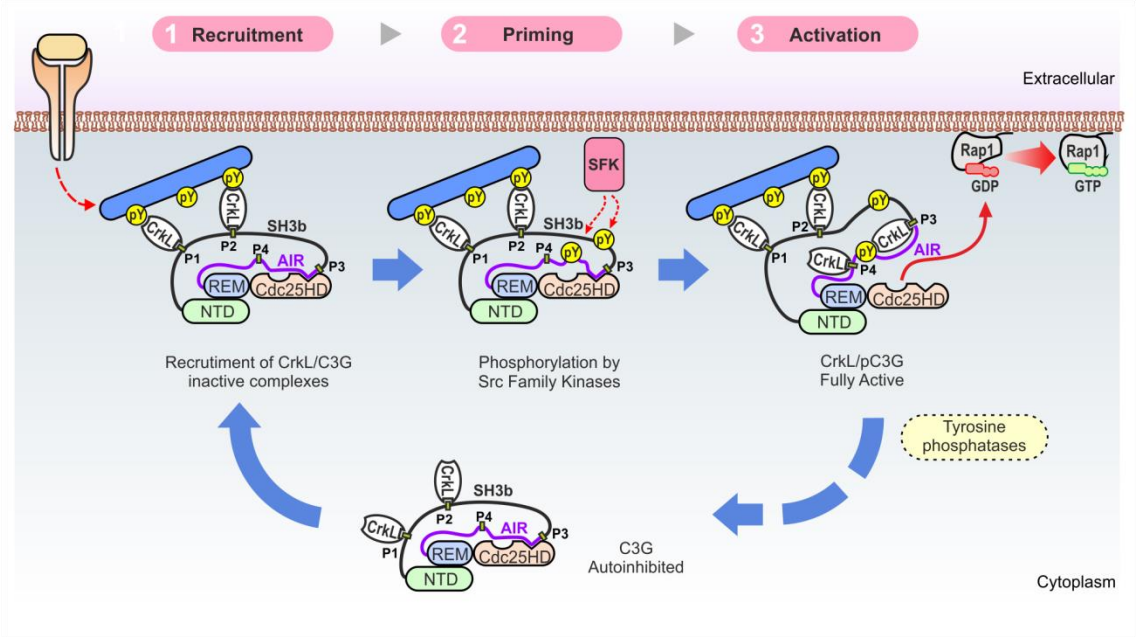



**Figure D3. Model for the activation of C3G.** Activation of C3G involves a three-step process. Initially, C3G is recruited by the interaction with CrkL through the P1 and P2 sites. Translocation to the plasma membrane is independent of the activation of C3G. The second phase is the priming of C3G that involves the phosphorylation of C3G at tyrosine residues by Src-family kinases (SFKs) or other kinases, such as Abl. Phosphorylation alone is not sufficient for activation. The third step encompasses the actual activation process, which occurs when CrkL engages with the phosphorylated C3G at the P3 and P4 sites. CrkL plays a key role in stabilizing an open state of pC3G, preventing the autoinhibitory interaction between the AIR and Cdc25HD and thus unleashing the GEF activity of the Cdc25HD. Activation is likely to be reversed through various mechanisms, one of which involves the action of tyrosine phosphatases. This

action results in the dissociation of CrkL from the P3 activation site and the dispersion of inactive CrkL/C3G complexes into the cytoplasm.

Efficient activation of C3G requires the convergence of tyrosine phosphorylation of C3G and the binding of Crk proteins. This dual-factor mechanism ensures precise spatiotemporal control of C3G signaling, allowing activation only when and where both conditions are concurred.

Significantly, the same Crk proteins participate in the two crucial steps of the activation of C3G: recruiting C3G to signaling sites and unlocking of its GEF activity. The four Crk-binding sites in C3G are chemically equivalent, meaning that CrkL exhibits similar affinity for all four PRMs. However, these binding sites are functionally specialized; they are not equally exposed in full-length C3G, and binding of Crk triggers different responses. This underscores the need to thoroughly study interactions involving short linear motifs, such as PRMs, in the context of the full-length proteins. In addition, this work highlights the relevance of weak or low affinity interactions, such as that between CrkL-SH2 and pC3G, in the regulation of biological processes.

**Activation of C3G by the different members of the Crk adaptor protein family**

We have shown that CrkL stimulates pC3G to higher activity levels than CrkII. Yet, both Crk proteins activate pC3G with similar $AC_{50}$ values, which is a parameter that reflects the affinity of the activator (Rehmann et al., 2006). We have also shown that CrkII and CrkL bind, through their SH3N domains, with similarly affinity to C3G. Therefore, differences in the activation of C3G by CrkII and CrkL are apparently caused by regions outside the SH3N domains. Using chimeric CrkL-CrkII proteins, we showed that the SH3N domain alone or the SH2 and SH3C domains of CrkII support efficient activation when combined with the complementary domains of CrkL. This suggests that the moderate activation of C3G by CrkII is related to its inter-domain architecture rather than to the contribution of individual domains. CrkII adopts in solution a compact structure stabilized by contacts between the linker region that connects the two SH3 domains and the SH2, SH3N, and SH3C domains (Kobashigawa et al., 2007). In contrast, the SH3C of CrkL is mobile and does not make stable interactions with the SH2-SH3N region (Jankowski et al., 2012). Given the prominent role of the SH2 domain in the activation of pC3G, the lower activation by CrkII may also be due to the relative arrangement of its SH2 and SH3N domains, which is different from the SH2-SH3N inter-domain organization of CrkL. Notably, CrkI, which only has the SH2 and

SH3N domains, has high inter-domain flexibility (Kobashigawa et al., 2007), suggesting that the SH2-SH3N of CrkII may be locked by the C-terminal region in a suboptimal conformation for activating C3G. We expect that the inter-domain arrangements of CrkL and CrkII would not be preserved in the chimeric proteins. Therefore, the chimeric proteins are likely to adopt relaxed architectures that would favor the simultaneous binding of the SH2 and SH3N domain with pC3G, as suggested by their efficient stimulation of the GEF activity.

**C3G activation mechanism impacts in physiological signaling and diseases**

Differences in the activation of C3G by CrkL and CrkII *in vitro* m may contribute to divergent Rap1 signaling outcomes by these two Crk proteins *in vivo*. For example, during processes of trans-endothelial migration or diapedesis, chemokines-induced Rap1 activation plays a key role (Shimonaka et al., 2003). Notable, the knockout of CrkL results in more pronounced defects in T cell diapedesis than the depletion of CrkII (Huang et al., 2015). This suggests that the activation of Rap1 in response to chemokines relies more on CrkL than on CrkII, aligning with our *in vitro* observations of higher efficiency of CrkL in activating C3G-Rap1 compared to CrkII. However, it might consider that other factors, such as expression levels and regulatory mechanisms, may also contribute to the differential signaling by CrkII and CrkL.

In various human cancers, the levels of Crk proteins are frequently increased and are associated with poor prognosis (Park, 2021). Ectopic overexpression of either CrkL or CrkII stimulates C3G-dependent Rap1 activation (Carabias et al., 2020, Ichiba et al., 1997), implying that high Crk protein levels in cancer cells may lead to abnormal C3G-Rap1 activation. According to the proposed model of C3G-Rap1 activation, in situations with elevated Crk protein concentrations, the activation of C3G may not necessarily require, or may depend less on, tyrosine phosphorylation of C3G. In addition, elevated levels of Crk protein might induce the atypical targeting of C3G to sites characterized by low levels of tyrosine phosphorylation, which would be insufficient to recruit C3G under normal physiological conditions. This notion is supported by the observed increase in C3G phosphorylation induced by Crk overexpression (Radha et al., 2004).

In summary, our detailed mechanistic description of C3G activation provides insights into the understanding of physiological CrkL/C3G/Rap1 signaling and opens the way for investigating alterations in this pathway associated with various diseases.

# CONCLUSIONS / CONCLUSIONES

1. CrkL binds differentially to the proline-rich motifs (PRMs) P1, P2, P3 and P4 in full-length C3G. In resting conditions, sites P1 and P2 are constitutively fully accessible. Exposure of the P3 site is linked to the activation state of C3G. The P4 site is partially accessible independently of the activation state of C3G.

2. Binding of CrkL to the P3 and P4 is required and sufficient for the direct activation of C3G by CrkL. Binding of CrkL to the P3 is the main activation cue. Sites P1 and P2 sites do not participate in this activation.

3. The three globular domains of CrkL (SH2, SH3N and SH3C) contribute to the stimulation of the GEF activity of C3G. The SH2 domain plays a key role in the activation of tyrosine-phosphorylated C3G.

4. CrkII activates C3G with lower efficiency than CrkL. These differences cannot be accounted solely by differences in individual SH2, SH3N and SH3C domains, and are apparently due to different inter-domain arrangements of CrkII and CrkL.

5. *In vitro* phosphorylation of C3G by Src occurs mainly at Y329, Y504, Y579 and Y590.

6. CrkL binds to phospho-C3G through a non-canonical low affinity interaction via the SH2 domain of CrkL and the phosphorylated Y590 of C3G. This interaction is essential to achieve efficient activation of C3G, possibly through the stabilization of the interaction of CrkL with the activation sites in C3G.

7. An experimentally validated structural prediction of the autoinhibited state of C3G revealed that the AIR corresponds to the segment 537-597. The AIR blocks the catalytic site in the Cdc25HD and makes additional contacts with the REM domain.

8. Two missense mutations caused by SNVs, Y570N and Y590N, found in non-Hodgkin lymphoma patients, and a third one, Y579C, found in a thyroid carcinoma patient, affect the final part of the AIR, known as the inhibitory tail, and disrupt the autoinhibition of C3G *in vitro*, and in HEK293T cells.

9. Crk proteins bind constitutively to sites P1 and P2 in HEK293T and Jurkat cell lines and are required for the recruitment of C3G to the plasma membrane upon TCR stimulation of Jurkat cells.

1. CrkL se une de distinta manera a los motivos ricos en prolina (PRMs) P1, P2, P3 y P4 en C3G completo. En condiciones de reposo, los sitios P1 y P2 están completamente accesibles de manera constitutiva. La exposición del sitio P3 está vinculada al estado de activación de C3G. El sitio P4 está parcialmente accesible independientemente del estado de activación de C3G.

2. La unión de CrkL a los sitios P3 y P4 es necesaria y suficiente para la activación directa de C3G por CrkL. La unión de CrkL al sitio P3 es la principal señal de activación. Los sitios P1 y P2 no participan en esta activación.

3. Los tres dominios globulares de CrkL (SH2, SH3N y SH3C) contribuyen a estimular la actividad de GEF de C3G. El dominio SH2 desempeña un papel clave en la activación de C3G fosforilada en tirosina.

4. CrkII activa C3G con menor eficiencia que CrkL. Estas diferencias no pueden explicarse únicamente por diferencias en los dominios individuales SH2, SH3N y SH3C, y aparentemente se deben a disposiciones diferentes entre los dominios de CrkII y CrkL.

5. La fosforilación in vitro de C3G por Src ocurre principalmente en las tirosinas Y329, Y504, Y579 e Y590.

6. CrkL se une a la fosfo-C3G a través de una interacción no canónica de baja afinidad mediante el dominio SH2 de CrkL y la tirosina fosforilada Y590 de C3G. Esta interacción es esencial para lograr una activación eficiente de C3G, posiblemente a través de la estabilización de la interacción de CrkL con los sitios de activación en C3G.

7. Una predicción estructural validada experimentalmente del estado autoinhibido de C3G reveló que el AIR corresponde al segmento 537-597. El AIR bloquea el sitio catalítico en el Cdc25HD y establece contactos adicionales con el dominio REM.

8. Dos mutaciones de sentido erróneo causadas por variantes de un solo nucleótido (SNVs), Y570N e Y590N, encontradas en pacientes con linfoma no Hodgkin, y una tercera, Y579C, encontrada en un paciente con carcinoma de tiroides, afectan a la parte final del AIR, conocida como la cola inhibitoria, y perturban la autoinhibición de C3G in vitro y en células HEK293T.

9. Las proteínas Crk se unen de manera constitutiva a los sitios P1 y P2 en las líneas celulares HEK293T y Jurkat, y son necesarias para el reclutamiento de C3G a la membrana plasmática cuando se estimulan las células Jurkat mediante el receptor de células T (TCR).

# BIBLIOGRAPHY

ALAN, J. K. & LUNDQUIST, E. A. 2013. Mutationally activated Rho GTPases in cancer. *Small GTPases,* 4**,** 159-63.

ALEMAYEHU, M., DRAGAN, M., PAPE, C., SIDDIQUI, I., SACKS, D. B., DI GUGLIELMO, G. M., BABWAH, A. V. & BHATTACHARYA, M. 2013. beta-Arrestin2 regulates lysophosphatidic acid-induced human breast tumor cell migration and invasion via Rap1 and IQGAP1. *PLoS One,* 8**,** e56174.

ALONSO-GARCIA, N., INGLES-PRIETO, A., SONNENBERG, A. & DE PEREDA, J. M. 2009. Structure of the Calx-beta domain of the integrin beta4 subunit: insights into function and cation-independent stability. *Acta Crystallogr D Biol Crystallogr,* 65**,** 858-71.

ALSAYED, Y., UDDIN, S., AHMAD, S., MAJCHRZAK, B., DRUKER, B. J., FISH, E. N. & PLATANIAS, L. C. 2000. IFN-gamma activates the C3G/Rap1 signaling pathway. *J Immunol,* 164**,** 1800-6.

AMARASINGHE, G. K. & ROSEN, M. K. 2005. Acidic region tyrosines provide access points for allosteric activation of the autoinhibited Vav1 Dbl homology domain. *Biochemistry,* 44**,** 15257-68.

ARAI, A., NOSAKA, Y., KANDA, E., YAMAMOTO, K., MIYASAKA, N. & MIURA, O. 2001. Rap1 is activated by erythropoietin or interleukin-3 and is involved in regulation of beta1 integrin-mediated hematopoietic cell adhesion. *J Biol Chem,* 276**,** 10453-62.

ARAI, A., NOSAKA, Y., KOHSAKA, H., MIYASAKA, N. & MIURA, O. 1999. CrkL activates integrin-mediated hematopoietic cell adhesion through the guanine nucleotide exchange factor C3G. *Blood,* 93**,** 3713-22.

ASURI, S., YAN, J., PARANAVITANA, N. C. & QUILLIAM, L. A. 2008. E-cadherin dis-engagement activates the Rap1 GTPase. *J Cell Biochem,* 105**,** 1027-37.

BAILEY, C. L., KELLY, P. & CASEY, P. J. 2009. Activation of Rap1 promotes prostate cancer metastasis. *Cancer Res,* 69**,** 4962-8.

BIRGE, R. B., FAJARDO, J. E., REICHMAN, C., SHOELSON, S. E., SONGYANG, Z., CANTLEY, L. C. & HANAFUSA, H. 1993. Identification and characterization of a high-affinity interaction between v-Crk and tyrosine-phosphorylated paxillin in CT10-transformed fibroblasts. *Mol Cell Biol,* 13**,** 4648-56.

BIRGE, R. B., KALODIMOS, C., INAGAKI, F. & TANAKA, S. 2009. Crk and CrkL adaptor proteins: networks for physiological and pathological signaling. *Cell Commun Signal,* 7**,** 13.

BOGUSKI, M. S. & MCCORMICK, F. 1993. Proteins regulating Ras and its relatives. *Nature,* 366**,** 643-54.

BORIACK-SJODIN, P. A., MARGARIT, S. M., BAR-SAGI, D. & KURIYAN, J. 1998. The structural basis of the activation of Ras by Sos. *Nature,* 394**,** 337-43.

BOS, J. L. 1997. Ras-like GTPases. *Biochim Biophys Acta,* 1333**,** M19-31.

BOS, J. L. 1998. All in the family? New insights and questions regarding interconnectivity of Ras, Rap1 and Ral. *EMBO J,* 17**,** 6776-82.

BOS, J. L. 2018. From Ras to Rap and Back, a Journey of 35 Years. *Cold Spring Harb Perspect Med,* 8.

BOS, J. L., REHMANN, H. & WITTINGHOFER, A. 2007. GEFs and GAPs: critical elements in the control of small G proteins. *Cell,* 129**,** 865-77.

BOURNE, H. R., SANDERS, D. A. & MCCORMICK, F. 1991. The GTPase superfamily: conserved structure and molecular mechanism. *Nature,* 349**,** 117-27.

BOUSSIOTIS, V. A., FREEMAN, G. J., BEREZOVSKAYA, A., BARBER, D. L. & NADLER, L. M. 1997. Maintenance of human T cell anergy: blocking of IL-2 gene transcription by activated Rap1. *Science,* 278**,** 124-8.

BOYKEVISCH, S., ZHAO, C., SONDERMANN, H., PHILIPPIDOU, P., HALEGOUA, S., KURIYAN, J. & BAR-SAGI, D. 2006. Regulation of ras signaling dynamics by Sos-mediated positive feedback. *Curr Biol,* 16**,** 2173-9.

BRADSHAW, J. M., MITAXOV, V. & WAKSMAN, G. 1999. Investigation of phosphotyrosine recognition by the SH2 domain of the Src kinase. *Journal of molecular biology,* 293**,** 971-985.

BROWN, M. T. & COOPER, J. A. 1996. Regulation, substrates and functions of src. *Biochim Biophys Acta,* 1287**,** 121-49.

BUENSUCESO, C. S. & O'TOOLE, T. E. 2000. The association of CRKII with C3G can be regulated by integrins and defines a novel means to regulate the mitogen-activated protein kinases. *J Biol Chem,* 275**,** 13118-25.

CARABIAS, A. 2019. *Mechanisms of auto-regulation and activation of the guanine nucleotide exchange factor C3G (Thesis).*

CARABIAS, A., GOMEZ-HERNANDEZ, M., DE CIMA, S., RODRIGUEZ-BLAZQUEZ, A., MORAN-VAQUERO, A., GONZALEZ-SAENZ, P., GUERRERO, C. & DE PEREDA, J. M. 2020. Mechanisms of autoregulation of C3G, activator of the GTPase Rap1, and its catalytic deregulation in lymphomas. *Sci Signal,* 13.

CERIONE, R. A. & ZHENG, Y. 1996. The Dbl family of oncogenes. *Curr Opin Cell Biol,* 8**,** 216-22.

COLICELLI, J. 2004. Human RAS superfamily proteins and related GTPases. *Sci STKE,* 2004**,** RE13.

CONSONNI, S. V., GLOERICH, M., SPANJAARD, E. & BOS, J. L. 2012. cAMP regulates DEP domain-mediated binding of the guanine nucleotide exchange factor Epac1 to phosphatidic acid at the plasma membrane. *Proc Natl Acad Sci U S A,* 109**,** 3814-9.

COOK, S. J., RUBINFELD, B., ALBERT, I. & MCCORMICK, F. 1993. RapV12 antagonizes Ras-dependent activation of ERK1 and ERK2 by LPA and EGF in Rat-1 fibroblasts. *EMBO J,* 12**,** 3475-85.

CORBALAN-GARCIA, S., MARGARIT, S. M., GALRON, D., YANG, S. S. & BAR-SAGI, D. 1998. Regulation of Sos activity by intramolecular interactions. *Mol Cell Biol,* 18**,** 880-6.

CHEN, R. H., CORBALAN-GARCIA, S. & BAR-SAGI, D. 1997. The role of the PH domain in the signal-dependent membrane targeting of Sos. *EMBO J,* 16**,** 1351-9.

CHERFILS, J. & CHARDIN, P. 1999. GEFs: structural basis for their activation of small GTP-binding proteins. *Trends Biochem Sci,* 24**,** 306-11.

CHERFILS, J. & ZEGHOUF, M. 2013. Regulation of small GTPases by GEFs, GAPs, and GDIs. *Physiol Rev,* 93**,** 269-309.

CHIANG, S. H., BAUMANN, C. A., KANZAKI, M., THURMOND, D. C., WATSON, R. T., NEUDAUER, C. L., MACARA, I. G., PESSIN, J. E. & SALTIEL, A. R. 2001. Insulin-stimulated GLUT4 translocation requires the CAP-dependent activation of TC10. *Nature,* 410**,** 944-8.

CHIANG, S. H., CHANG, L. & SALTIEL, A. R. 2006. TC10 and insulin-stimulated glucose transport. *Methods Enzymol,* 406**,** 701-14.

DAYMA, K. & RADHA, V. 2011. Cytoskeletal remodeling by C3G to induce neurite-like extensions and inhibit motility in highly invasive breast carcinoma cells. *Biochim Biophys Acta,* 1813**,** 456-65.

DAYMA, K., RAMADHAS, A., SASIKUMAR, K. & RADHA, V. 2012. Reciprocal Negative Regulation between the Guanine Nucleotide Exchange Factor C3G and beta-Catenin. *Genes Cancer,* 3**,** 564-77.

DE FALCO, V., CASTELLONE, M. D., DE VITA, G., CIRAFICI, A. M., HERSHMAN, J. M., GUERRERO, C., FUSCO, A., MELILLO, R. M. & SANTORO, M. 2007. RET/papillary thyroid carcinoma oncogenic signaling through the Rap1 small GTPase. *Cancer Res,* 67**,** 381-90.

DE JONG, R., TEN HOEVE, J., HEISTERKAMP, N. & GROFFEN, J. 1997. Tyrosine 207 in CRKL is the BCR/ABL phosphorylation site. *Oncogene,* 14**,** 507-13.

DE MARCO, A., DEUERLING, E., MOGK, A., TOMOYASU, T. & BUKAU, B. 2007. Chaperone-based procedure to increase yields of soluble recombinant proteins produced in E. coli. *BMC Biotechnol,* 7**,** 32.

DE ROOIJ, J., ZWARTKRUIS, F. J., VERHEIJEN, M. H., COOL, R. H., NIJMAN, S. M., WITTINGHOFER, A. & BOS, J. L. 1998. Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. *Nature,* 396**,** 474-7.

DEGEER, J. & LAMARCHE-VANE, N. 2013. Rho GTPases in neurodegeneration diseases. *Exp Cell Res,* 319**,** 2384-94.

DENG, Y., ALICEA-VELAZQUEZ, N. L., BANNWARTH, L., LEHTONEN, S. I., BOGGON, T. J., CHENG, H. C., HYTONEN, V. P. & TURK, B. E. 2014. Global analysis of human nonreceptor tyrosine kinase specificity using high-density peptide microarrays. *J Proteome Res,* 13**,** 4339-46.

DIEKMANN, Y., SEIXAS, E., GOUW, M., TAVARES-CADETE, F., SEABRA, M. C. & PEREIRA-LEAL, J. B. 2011. Thousands of rab GTPases for the cell biologist. *PLoS Comput Biol,* 7**,** e1002217.

DINITTO, J. P., DELPRATO, A., GABE LEE, M. T., CRONIN, T. C., HUANG, S., GUILHERME, A., CZECH, M. P. & LAMBRIGHT, D. G. 2007. Structural basis and mechanism of autoregulation in 3-phosphoinositide-dependent Grp1 family Arf GTPase exchange factors. *Mol Cell,* 28**,** 569-83.

DONALDSON, L. W., GISH, G., PAWSON, T., KAY, L. E. & FORMAN-KAY, J. D. 2002. Structure of a regulatory complex involving the Abl SH3 domain, the Crk SH2 domain, and a Crk-derived phosphopeptide. *Proc Natl Acad Sci U S A,* 99**,** 14053-8.

DUNN, K. W., KAMOCKA, M. M. & MCDONALD, J. H. 2011. A practical guide to evaluating colocalization in biological microscopy. *Am J Physiol Cell Physiol,* 300**,** C723-42.

EGAN, S. E., GIDDINGS, B. W., BROOKS, M. W., BUDAY, L., SIZELAND, A. M. & WEINBERG, R. A. 1993. Association of Sos Ras exchange protein with Grb2 is implicated in tyrosine kinase signal transduction and transformation. *Nature,* 363**,** 45-51.

FELLER, S. M. 2001. Crk family adaptors-signalling complex formation and biological roles. *Oncogene,* 20**,** 6348-71.

FENG, S., CHEN, J. K., YU, H., SIMON, J. A. & SCHREIBER, S. L. 1994. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science,* 266**,** 1241-7.

FERNANDEZ, V., JARES, P., SALAVERRIA, I., GINE, E., BEA, S., AYMERICH, M., COLOMER, D., VILLAMOR, N., BOSCH, F., MONTSERRAT, E. & CAMPO, E. 2008. Gene expression profile and genomic changes in disease progression of early-stage chronic lymphocytic leukemia. *Haematologica,* 93**,** 132-6.

FRANKE, B., VAN TRIEST, M., DE BRUIJN, K. M., VAN WILLIGEN, G., NIEUWENHUIS, H. K., NEGRIER, C., AKKERMAN, J. W. & BOS, J. L. 2000. Sequential regulation of the small GTPase Rap1 in human platelets. *Mol Cell Biol,* 20**,** 779-85.

FREEDMAN, T. S., SONDERMANN, H., FRIEDLAND, G. D., KORTEMME, T., BAR-SAGI, D., MARQUSEE, S. & KURIYAN, J. 2006. A Ras-induced conformational switch in the Ras activator Son of sevenless. *Proc Natl Acad Sci U S A,* 103**,** 16692-7.

FREEDMAN, T. S., SONDERMANN, H., KUCHMENT, O., FRIEDLAND, G. D., KORTEMME, T. & KURIYAN, J. 2009. Differences in flexibility underlie functional differences in the Ras activators son of sevenless and Ras guanine nucleotide releasing factor 1. *Structure,* 17**,** 41-53.

FUKUMOTO, Y., KAIBUCHI, K., HORI, Y., FUJIOKA, H., ARAKI, S., UEDA, T., KIKUCHI, A. & TAKAI, Y. 1990. Molecular cloning and characterization of a novel type of regulatory protein (GDI) for the rho proteins, ras p21-like small GTP-binding proteins. *Oncogene,* 5**,** 1321-8.

GÓMEZ-HERNÁNDEZ, M. 2014. *Caracterización de la organización estructural del factor de intercambio de nucleótido de guanina C3G (Thesis).*

GOTO, M., MITRA, R. S., LIU, M., LEE, J., HENSON, B. S., CAREY, T., BRADFORD, C., PRINCE, M., WANG, C. Y., FEARON, E. R. & D'SILVA, N. J. 2010. Rap1 stabilizes beta-catenin and enhances beta-catenin-dependent transcription and invasion in squamous cell carcinoma of the head and neck. *Clin Cancer Res,* 16**,** 65-76.

GOTOH, T., HATTORI, S., NAKAMURA, S., KITAYAMA, H., NODA, M., TAKAI, Y., KAIBUCHI, K., MATSUI, H., HATASE, O., TAKAHASHI, H. & ET AL. 1995. Identification of Rap1 as a target for the Crk SH3 domain-binding guanine nucleotide-releasing factor C3G. *Mol Cell Biol,* 15**,** 6746-53.

GOTOH, T., NIINO, Y., TOKUDA, M., HATASE, O., NAKAMURA, S., MATSUDA, M. & HATTORI, S. 1997. Activation of R-Ras by Ras-guanine nucleotide-releasing factor. *J Biol Chem,* 272**,** 18602-7.

GOTOH, T., TIAN, X. & FEIG, L. A. 2001. Prenylation of target GTPases contributes to signaling specificity of Ras-guanine nucleotide exchange factors. *J Biol Chem,* 276**,** 38029-35.

GREEN, M. R., GENTLES, A. J., NAIR, R. V., IRISH, J. M., KIHIRA, S., LIU, C. L., KELA, I., HOPMANS, E. S., MYKLEBUST, J. H., JI, H., PLEVRITIS, S. K., LEVY, R. & ALIZADEH, A. A. 2013. Hierarchy in somatic mutations arising during genomic evolution and progression of follicular lymphoma. *Blood,* 121**,** 1604-11.

GUERRERO, C., FERNANDEZ-MEDARDE, A., ROJAS, J. M., FONT DE MORA, J., ESTEBAN, L. M. & SANTOS, E. 1998. Transformation suppressor activity of C3G is independent of its CDC25-homology domain. *Oncogene,* 16**,** 613-24.

GUERRERO, C., MARTIN-ENCABO, S., FERNANDEZ-MEDARDE, A. & SANTOS, E. 2004. C3G-mediated suppression of oncogene-induced focus formation in fibroblasts involves inhibition of ERK activation, cyclin A expression and alterations of anchorage-independent growth. *Oncogene,* 23**,** 4885-93.

GUREASKO, J., KUCHMENT, O., MAKINO, D. L., SONDERMANN, H., BAR-SAGI, D. & KURIYAN, J. 2010. Role of the histone domain in the autoinhibition and activation of the Ras activator Son of Sevenless. *Proc Natl Acad Sci U S A,* 107**,** 3430-5.

GURIS, D. L., DUESTER, G., PAPAIOANNOU, V. E. & IMAMOTO, A. 2006. Dose-dependent interaction of Tbx1 and Crkl and locally aberrant RA signaling in a model of del22q11 syndrome. *Dev Cell,* 10**,** 81-92.

GURIS, D. L., FANTES, J., TARA, D., DRUKER, B. J. & IMAMOTO, A. 2001. Mice lacking the homologue of the human 22q11.2 gene CRKL phenocopy neurocristopathies of DiGeorge syndrome. *Nat Genet,* 27**,** 293-8.

GUTIERREZ-BERZAL, J., CASTELLANO, E., MARTIN-ENCABO, S., GUTIERREZ-CIANCA, N., HERNANDEZ, J. M., SANTOS, E. & GUERRERO, C. 2006. Characterization of p87C3G, a novel, truncated C3G isoform that is overexpressed in chronic myeloid leukemia and interacts with Bcr-Abl. *Exp Cell Res,* 312**,** 938-48.

GUTIERREZ-HERRERO, S., FERNANDEZ-INFANTE, C., HERNANDEZ-CANO, L., ORTIZ-RIVERO, S., GUIJAS, C., MARTIN-GRANADO, V., GONZALEZ-PORRAS, J. R., BALSINDE, J., PORRAS, A. & GUERRERO, C. 2020. C3G contributes to platelet activation and aggregation by regulating major signaling pathways. *Signal Transduct Target Ther,* 5**,** 29.

GUTIERREZ-HERRERO, S., MAIA, V., GUTIERREZ-BERZAL, J., CALZADA, N., SANZ, M., GONZALEZ-MANCHON, C., PERICACHO, M., ORTIZ-RIVERO, S., GONZALEZ-PORRAS, J. R., ARECHEDERRA, M., PORRAS, A. & 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**,** 1366-77.

HALL, B. E., YANG, S. S. & BAR-SAGI, D. 2002. Autoinhibition of Sos by intramolecular interactions. *Front Biosci,* 7**,** d288-94.

HALL, B. E., YANG, S. S., BORIACK-SJODIN, P. A., KURIYAN, J. & BAR-SAGI, D. 2001. Structure-based mutagenesis reveals distinct functions for Ras switch 1 and switch 2 in Sos-catalyzed guanine nucleotide exchange. *J Biol Chem,* 276**,** 27629-37.

HEASMAN, S. J. & RIDLEY, A. J. 2008. Mammalian Rho GTPases: new insights into their functions from in vivo studies. *Nat Rev Mol Cell Biol,* 9**,** 690-701.

HERNANDEZ-CANO, L., FERNANDEZ-INFANTE, C., HERRANZ, O., BERROCAL, P., LOZANO, F. S., SANCHEZ-MARTIN, M. A., PORRAS, A. & GUERRERO, C. 2022. New functions of C3G in platelet biology: Contribution to ischemia-induced angiogenesis, tumor metastasis and TPO clearance. *Front Cell Dev Biol,* 10**,** 1026287.

HOFFMAN, G. R., NASSAR, N. & CERIONE, R. A. 2000. Structure of the Rho family GTP-binding protein Cdc42 in complex with the multifunctional regulator RhoGDI. *Cell,* 100**,** 345-56.

HOGAN, C., SERPENTE, N., COGRAM, P., HOSKING, C. R., BIALUCHA, C. U., FELLER, S. M., BRAGA, V. M., BIRCHMEIER, W. & FUJITA, Y. 2004. Rap1 regulates the formation of E-cadherin-based cell-cell contacts. *Mol Cell Biol,* 24**,** 6690-700.

HUANG, Y., CLARKE, F., KARIMI, M., ROY, N. H., WILLIAMSON, E. K., OKUMURA, M., MOCHIZUKI, K., CHEN, E. J., PARK, T. J., DEBES, G. F., ZHANG, Y., CURRAN, T., KAMBAYASHI, T. & BURKHARDT, J. K. 2015. CRK proteins selectively regulate T cell migration into inflamed tissues. *J Clin Invest,* 125**,** 1019-32.

HUANG, Y., MAGDALENO, S., HOPKINS, R., SLAUGHTER, C., CURRAN, T. & KESHVARA, L. 2004. Tyrosine phosphorylated Disabled 1 recruits Crk family adapter proteins. *Biochem Biophys Res Commun,* 318**,** 204-12.

HUNTER, T. 2015. Discovering the first tyrosine kinase. *Proc Natl Acad Sci U S A,* 112**,** 7877-82.

HUNTER, T. & SEFTON, B. M. 1980. Transforming gene product of Rous sarcoma virus phosphorylates tyrosine. *Proc Natl Acad Sci U S A,* 77**,** 1311-5.

ICHIBA, T., HASHIMOTO, Y., NAKAYA, M., KURAISHI, Y., TANAKA, S., KURATA, T., MOCHIZUKI, N. & MATSUDA, M. 1999. Activation of C3G guanine nucleotide exchange factor for Rap1 by phosphorylation of tyrosine 504. *J Biol Chem,* 274**,** 14376-81.

ICHIBA, T., KURAISHI, Y., SAKAI, O., NAGATA, S., GROFFEN, J., KURATA, T., HATTORI, S. & MATSUDA, M. 1997. Enhancement of guanine-nucleotide exchange activity of C3G for Rap1 by the expression of Crk, CrkL, and Grb2. *J Biol Chem,* 272**,** 22215-20.

INGHAM, R. J., KREBS, D. L., BARBAZUK, S. M., TURCK, C. W., HIRAI, H., MATSUDA, M. & GOLD, M. R. 1996. B cell antigen receptor signaling induces the formation of complexes containing the Crk adapter proteins. *J Biol Chem,* 271**,** 32306-14.

INNOCENTI, M., FRITTOLI, E., PONZANELLI, I., FALCK, J. R., BRACHMANN, S. M., DI FIORE, P. P. & SCITA, G. 2003. Phosphoinositide 3-kinase activates Rac by entering in a complex with Eps8, Abi1, and Sos-1. *J Cell Biol,* 160**,** 17-23.

INNOCENTI, M., TENCA, P., FRITTOLI, E., FARETTA, M., TOCCHETTI, A., DI FIORE, P. P. & SCITA, G. 2002. Mechanisms through which Sos-1 coordinates the activation of Ras and Rac. *J Cell Biol,* 156**,** 125-36.

ISAKOV, N. 2008. A new twist to adaptor proteins contributes to regulation of lymphocyte cell signaling. *Trends Immunol,* 29**,** 388-96.

IWIG, J. S., VERCOULEN, Y., DAS, R., BARROS, T., LIMNANDER, A., CHE, Y., PELTON, J. G., WEMMER, D. E., ROOSE, J. P. & 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. & KALODIMOS, C. G. 2012. Domain organization differences explain Bcr-Abl's preference for CrkL over CrkII. *Nat Chem Biol,* 8**,** 590-6.

JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ZIDEK, A., POTAPENKO, A., BRIDGLAND, A., MEYER, C., KOHL, S. A. A., BALLARD, A. J., COWIE, A., ROMERA-PAREDES, B., NIKOLOV, S., JAIN, R., ADLER, J., BACK, T., PETERSEN, S., REIMAN, D., CLANCY, E., ZIELINSKI, M., STEINEGGER, M., PACHOLSKA, M., BERGHAMMER, T., BODENSTEIN, S., SILVER, D., VINYALS, O., SENIOR, A. W., KAVUKCUOGLU, K., KOHLI, P. & HASSABIS, D. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature,* 596**,** 583-589.

KARANDUR, D., NAWROTEK, A., KURIYAN, J. & 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**,** 11416-11421.

KARNOUB, A. E. & WEINBERG, R. A. 2008. Ras oncogenes: split personalities. *Nat Rev Mol Cell Biol,* 9**,** 517-31.

KASHATUS, D. F. 2013. Ral GTPases in tumorigenesis: emerging from the shadows. *Exp Cell Res,* 319**,** 2337-42.

KATAGIRI, K., HATTORI, M., MINATO, N., IRIE, S., TAKATSU, K. & KINASHI, T. 2000. Rap1 is a potent activation signal for leukocyte function-associated antigen 1 distinct from protein kinase C and phosphatidylinositol-3-OH kinase. *Mol Cell Biol,* 20**,** 1956-69.

KAWASAKI, H., SPRINGETT, G. M., MOCHIZUKI, N., TOKI, S., NAKAYA, M., MATSUDA, M., HOUSMAN, D. E. & GRAYBIEL, A. M. 1998. A family of cAMP-binding proteins that directly activate Rap1. *Science,* 282**,** 2275-9.

KELLER, S., VARGAS, C., ZHAO, H., PISZCZEK, G., BRAUTIGAM, C. A. & SCHUCK, P. 2012. High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal Chem,* 84**,** 5066-73.

KIRSCH, K. H., GEORGESCU, M. M. & HANAFUSA, H. 1998. Direct binding of p130(Cas) to the guanine nucleotide exchange factor C3G. *J Biol Chem,* 273**,** 25673-9.

KITAYAMA, H., SUGIMOTO, Y., MATSUZAKI, T., IKAWA, Y. & NODA, M. 1989. A ras-related gene with transformation suppressor activity. *Cell,* 56**,** 77-84.

KLEBE, C., PRINZ, H., WITTINGHOFER, A. & GOODY, R. S. 1995. The kinetic mechanism of Ran--nucleotide exchange catalyzed by RCC1. *Biochemistry,* 34**,** 12543-52.

KNUDSEN, B. S., FELLER, S. M. & HANAFUSA, H. 1994. Four proline-rich sequences of the guanine-nucleotide exchange factor C3G bind with unique specificity to the first Src homology 3 domain of Crk. *J Biol Chem,* 269**,** 32781-7.

KOBASHIGAWA, Y. & INAGAKI, F. 2012. Structural biology: CrkL is not Crk-like. *Nat Chem Biol,* 8**,** 504-5.

KOBASHIGAWA, Y., SAKAI, M., NAITO, M., YOKOCHI, M., KUMETA, H., MAKINO, Y., OGURA, K., TANAKA, S. & INAGAKI, F. 2007. Structural basis for the transforming activity of human cancer-related signaling adaptor protein CRK. *Nat Struct Mol Biol,* 14**,** 503-10.

KOBASHIGAWA, Y., TANAKA, S. & INAGAKI, F. 2008. [Structural basis for the transforming activity of human cancer-related signaling adaptor protein Crk]. *Tanpakushitsu Kakusan Koso,* 53**,** 148-56.

KOOISTRA, M. R., DUBE, N. & BOS, J. L. 2007. Rap1: a key regulator in cell-cell junction formation. *J Cell Sci,* 120**,** 17-22.

KOSHIBA, S., KIGAWA, T., KIM, J. H., SHIROUZU, M., BOWTELL, D. & YOKOYAMA, S. 1997. The solution structure of the pleckstrin homology domain of mouse Son-of-sevenless 1 (mSos1). *J Mol Biol,* 269**,** 579-91.

KUBALA, M. H., KOVTUN, O., ALEXANDROV, K. & COLLINS, B. M. 2010. Structural and thermodynamic analysis of the GFP:GFP-nanobody complex. *Protein Sci,* 19**,** 2389-401.

LAGARRIGUE, F., KIM, C. & GINSBERG, M. H. 2016. The Rap1-RIAM-talin axis of integrin activation and blood cell function. *Blood,* 128**,** 479-87.

LEE, Y. K., LOW-NAM, S. T., CHUNG, J. K., HANSEN, S. D., LAM, H. Y. M., ALVAREZ, S. & 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. & 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**,** 7420-30.

LI, Y., ASURI, S., REBHUN, J. F., CASTRO, A. F., PARANAVITANA, N. C. & QUILLIAM, L. A. 2006. The RAP1 guanine nucleotide exchange factor Epac2 couples cyclic AMP and Ras signals at the plasma membrane. *J Biol Chem,* 281**,** 2506-14.

LI, Z., DELANEY, M. K., O'BRIEN, K. A. & DU, X. 2010. Signaling during platelet adhesion and activation. *Arterioscler Thromb Vasc Biol,* 30**,** 2341-9.

LIN, K. T., YEH, Y. M., CHUANG, C. M., YANG, S. Y., CHANG, J. W., SUN, S. P., WANG, Y. S., CHAO, K. C. & WANG, L. H. 2015. Glucocorticoids mediate induction of microRNA-708 to suppress ovarian cancer metastasis through targeting Rap1B. *Nat Commun,* 6**,** 5917.

LING, L., ZHU, T. & LOBIE, P. E. 2003. Src-CrkII-C3G-dependent activation of Rap1 switches growth hormone-stimulated p44/42 MAP kinase and JNK/SAPK activities. *J Biol Chem,* 278**,** 27301-11.

LIPSICK, J. 2021. A History of Cancer Research: Tumor Viruses. *Cold Spring Harb Perspect Biol,* 13.

LOIRAND, G., SAUZEAU, V. & PACAUD, P. 2013. Small G proteins in the cardiovascular system: physiological and pathological aspects. *Physiol Rev,* 93**,** 1659-720.

LOOI, C. K., HII, L. W., NGAI, S. C., LEONG, C. O. & MAI, C. W. 2020. The Role of Ras-Associated Protein 1 (Rap1) in Cancer: Bad Actor or Good Player? *Biomedicines,* 8.

LOWENSTEIN, E. J., DALY, R. J., BATZER, A. G., LI, W., MARGOLIS, B., LAMMERS, R., ULLRICH, A., SKOLNIK, E. Y., BAR-SAGI, D. & SCHLESSINGER, J. 1992. The SH2 and SH3 domain-containing protein GRB2 links receptor tyrosine kinases to ras signaling. *Cell,* 70**,** 431-42.

LOWY, D. R. & WILLUMSEN, B. M. 1993. Function and regulation of ras. *Annu Rev Biochem,* 62**,** 851-91.

LUBER, B., CANDIDUS, S., HANDSCHUH, G., MENTELE, E., HUTZLER, P., FELLER, S., VOSS, J., HOFLER, H. & BECKER, K. F. 2000. Tumor-derived mutated E-cadherin influences beta-catenin localization and increases susceptibility to actin cytoskeletal changes induced by pervanadate. *Cell Adhes Commun,* 7**,** 391-408.

MACARA, I. G., LOUNSBURY, K. M., RICHARDS, S. A., MCKIERNAN, C. & BAR-SAGI, D. 1996. The Ras superfamily of GTPases. *FASEB J,* 10**,** 625-30.

MAIA, V., ORTIZ-RIVERO, S., SANZ, M., GUTIERREZ-BERZAL, J., ALVAREZ-FERNANDEZ, I., GUTIERREZ-HERRERO, S., DE PEREDA, J. M., PORRAS, A. & 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**,** 9.

MAIA, V., SANZ, M., GUTIERREZ-BERZAL, J., DE LUIS, A., GUTIERREZ-UZQUIZA, A., PORRAS, A. & GUERRERO, C. 2009. C3G silencing enhances STI-571-induced apoptosis in CML cells through p38 MAPK activation, but it antagonizes STI-571 inhibitory effect on survival. *Cell Signal,* 21**,** 1229-35.

MANNING, G., WHYTE, D. B., MARTINEZ, R., HUNTER, T. & SUDARSANAM, S. 2002. The protein kinase complement of the human genome. *Science,* 298**,** 1912-34.

MANSO, J. A., CARABIAS, A., SARKANY, Z., DE PEREDA, J. M., PEREIRA, P. J. B. & MACEDO-RIBEIRO, S. 2023. Pathogen-specific structural features of Candida albicans Ras1 activation complex: uncovering new antifungal drug targets. *mBio,* 14**,** e0063823.

MANSO, J. A., GOMEZ-HERNANDEZ, M., CARABIAS, A., ALONSO-GARCIA, N., GARCIA-RUBIO, I., KREFT, M., SONNENBERG, A. & DE PEREDA, J. M. 2019. Integrin alpha6beta4 Recognition of a Linear Motif of Bullous Pemphigoid Antigen BP230 Controls Its Recruitment to Hemidesmosomes. *Structure,* 27**,** 952-964 e6.

MANZANO, S., GUTIERREZ-UZQUIZA, A., BRAGADO, P., CUESTA, A. M., GUERRERO, C. & PORRAS, A. 2021a. C3G Protein, a New Player in Glioblastoma. *Int J Mol Sci,* 22.

MANZANO, S., GUTIERREZ-UZQUIZA, A., BRAGADO, P., SEQUERA, C., HERRANZ, O., RODRIGO-FAUS, M., JAUREGUI, P., MORGNER, S., RUBIO, I., GUERRERO, C. & PORRAS, A. 2021b. C3G

downregulation induces the acquisition of a mesenchymal phenotype that enhances aggressiveness of glioblastoma cells. *Cell Death Dis,* 12**,** 348.

MARGARIT, S. M., SONDERMANN, H., HALL, B. E., NAGAR, B., HOELZ, A., PIRRUCCELLO, M., BAR-SAGI, D. & KURIYAN, J. 2003. Structural evidence for feedback activation by Ras.GTP of the Ras-specific nucleotide exchange factor SOS. *Cell,* 112**,** 685-95.

MARTIN-ENCABO, S., SANTOS, E. & GUERRERO, C. 2007. C3G mediated suppression of malignant transformation involves activation of PP2A phosphatases at the subcortical actin cytoskeleton. *Exp Cell Res,* 313**,** 3881-91.

MARTIN-GRANADO, V., ORTIZ-RIVERO, S., CARMONA, R., GUTIERREZ-HERRERO, S., BARRERA, M., SAN-SEGUNDO, L., SEQUERA, C., PERDIGUERO, P., LOZANO, F., MARTIN-HERRERO, F., GONZALEZ-PORRAS, J. R., MUNOZ-CHAPULI, R., PORRAS, A. & GUERRERO, C. 2017. C3G promotes a selective release of angiogenic factors from activated mouse platelets to regulate angiogenesis and tumor metastasis. *Oncotarget,* 8**,** 110994-111011.

MATSUDA, M., TANAKA, S., NAGATA, S., KOJIMA, A., KURATA, T. & SHIBUYA, M. 1992. Two species of human CRK cDNA encode proteins with distinct biological activities. *Mol Cell Biol,* 12**,** 3482-9.

MATSUKI, T., PRAMATAROVA, A. & HOWELL, B. W. 2008. Reduction of Crk and CrkL expression blocks reelin-induced dendritogenesis. *J Cell Sci,* 121**,** 1869-75.

MAYER, B. J., HAMAGUCHI, M. & HANAFUSA, H. 1988a. Characterization of p47gag-crk, a novel oncogene product with sequence similarity to a putative modulatory domain of protein-tyrosine kinases and phospholipase C. *Cold Spring Harb Symp Quant Biol,* 53 Pt 2**,** 907-14.

MAYER, B. J., HAMAGUCHI, M. & HANAFUSA, H. 1988b. A novel viral oncogene with structural similarity to phospholipase C. *Nature,* 332**,** 272-5.

MEDEIROS, R. B., DICKEY, D. M., CHUNG, H., QUALE, A. C., NAGARAJAN, L. R., BILLADEAU, D. D. & 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**,** 213-26.

MILBURN, M. V., TONG, L., DEVOS, A. M., BRUNGER, A., YAMAIZUMI, Z., NISHIMURA, S. & KIM, S. H. 1990. Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science,* 247**,** 939-45.

MILETIC, A. V., SAKATA-SOGAWA, K., HIROSHIMA, M., HAMANN, M. J., GOMEZ, T. S., OTA, N., KLOEPPEL, T., KANAGAWA, O., TOKUNAGA, M., BILLADEAU, D. D. & SWAT, W. 2006. Vav1 acidic region tyrosine 174 is required for the formation of T cell receptor-induced microclusters and is essential in T cell development and activation. *J Biol Chem,* 281**,** 38257-65.

MIRDITA, M., SCHUTZE, K., MORIWAKI, Y., HEO, L., OVCHINNIKOV, S. & STEINEGGER, M. 2022. ColabFold: making protein folding accessible to all. *Nat Methods,* 19**,** 679-682.

MITRA, A., KALAYARASAN, S., GUPTA, V. & RADHA, V. 2011. TC-PTP dephosphorylates the guanine nucleotide exchange factor C3G (RapGEF1) and negatively regulates differentiation of human neuroblastoma cells. *PLoS One,* 6**,** e23681.

MITRA, A. & RADHA, V. 2010. F-actin-binding domain of c-Abl regulates localized phosphorylation of C3G: role of C3G in c-Abl-mediated cell death. *Oncogene,* 29**,** 4528-42.

MORIN, R. D., MENDEZ-LAGO, M., MUNGALL, A. J., GOYA, R., MUNGALL, K. L., CORBETT, R. D., JOHNSON, N. A., SEVERSON, T. M., CHIU, R., FIELD, M., JACKMAN, S., KRZYWINSKI, M., SCOTT, D. W., TRINH, D. L., TAMURA-WELLS, J., LI, S., FIRME, M. R., ROGIC, S., GRIFFITH, M., CHAN, S., YAKOVENKO, O., MEYER, I. M., ZHAO, E. Y., SMAILUS, D., MOKSA, M., CHITTARANJAN, S.,

RIMSZA, L., BROOKS-WILSON, A., SPINELLI, J. J., BEN-NERIAH, S., MEISSNER, B., WOOLCOCK, B., BOYLE, M., MCDONALD, H., TAM, A., ZHAO, Y., DELANEY, A., ZENG, T., TSE, K., BUTTERFIELD, Y., BIROL, I., HOLT, R., SCHEIN, J., HORSMAN, D. E., MOORE, R., JONES, S. J., CONNORS, J. M., HIRST, M., GASCOYNE, R. D. & MARRA, M. A. 2011. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature,* 476**,** 298-303.

MURALIDHARAN, V., DUTTA, K., CHO, J., VILA-PERELLO, M., RALEIGH, D. P., COWBURN, D. & MUIR, T. W. 2006. Solution structure and folding characteristics of the C-terminal SH3 domain of c-Crk-II. *Biochemistry,* 45**,** 8874-84.

MUSACCHIO, A., SARASTE, M. & WILMANNS, M. 1994. High-resolution crystal structures of tyrosine kinase SH3 domains complexed with proline-rich peptides. *Nat Struct Biol,* 1**,** 546-51.

NAKAMURA, Y., HIBINO, K., YANAGIDA, T. & SAKO, Y. 2016. Switching of the positive feedback for RAS activation by a concerted function of SOS membrane association domains. *Biophys Physicobiol,* 13**,** 1-11.

NIMNUAL, A. S., YATSULA, B. A. & BAR-SAGI, D. 1998. Coupling of Ras and Rac guanosine triphosphatases through the Ras exchanger Sos. *Science,* 279**,** 560-3.

NOLZ, J. C., NACUSI, L. P., SEGOVIS, C. M., MEDEIROS, R. B., MITCHELL, J. S., SHIMIZU, Y. & BILLADEAU, D. D. 2008. The WAVE2 complex regulates T cell receptor signaling to integrins via Abl- and CrkL-C3G-mediated activation of Rap1. *J Cell Biol,* 182**,** 1231-44.

NOSAKA, Y., ARAI, A., MIYASAKA, N. & MIURA, O. 1999. CrkL mediates Ras-dependent activation of the Raf/ERK pathway through the guanine nucleotide exchange factor C3G in hematopoietic cells stimulated with erythropoietin or interleukin-3. *J Biol Chem,* 274**,** 30154-62.

OHASHI, Y., TACHIBANA, K., KAMIGUCHI, K., FUJITA, H. & MORIMOTO, C. 1998. T cell receptor-mediated tyrosine phosphorylation of Cas-L, a 105-kDa Crk-associated substrate-related protein, and its association of Crk and C3G. *J Biol Chem,* 273**,** 6446-51.

OHBA, Y., IKUTA, K., OGURA, A., MATSUDA, J., MOCHIZUKI, N., NAGASHIMA, K., KUROKAWA, K., MAYER, B. J., MAKI, K., MIYAZAKI, J. & MATSUDA, M. 2001. Requirement for C3G-dependent Rap1 activation for cell adhesion and embryogenesis. *EMBO J,* 20**,** 3333-41.

OHBA, Y., MOCHIZUKI, N., YAMASHITA, S., CHAN, A. M., SCHRADER, J. W., HATTORI, S., NAGASHIMA, K. & MATSUDA, M. 2000. Regulatory proteins of R-Ras, TC21/R-Ras2, and M-Ras/R-Ras3. *J Biol Chem,* 275**,** 20020-6.

OKADA, S., MATSUDA, M., ANAFI, M., PAWSON, T. & 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**,** 2554-65.

OKADA, S. & PESSIN, J. E. 1997. Insulin and epidermal growth factor stimulate a conformational change in Rap1 and dissociation of the CrkII-C3G complex. *J Biol Chem,* 272**,** 28179-82.

OKADA, T., MASUDA, T., SHINKAI, M., KARIYA, K. & KATAOKA, T. 1996. Post-translational modification of H-Ras is required for activation of, but not for association with, B-Raf. *J Biol Chem,* 271**,** 4671-8.

OKINO, K., NAGAI, H., NAKAYAMA, H., DOI, D., YONEYAMA, K., KONISHI, H. & TAKESHITA, T. 2006. Inactivation of Crk SH3 domain-binding guanine nucleotide-releasing factor (C3G) in cervical squamous cell carcinoma. *Int J Gynecol Cancer,* 16**,** 763-71.

ORTIZ-RIVERO, S., BAQUERO, C., HERNANDEZ-CANO, L., ROLDAN-ETCHEVERRY, J. J., GUTIERREZ-HERRERO, S., FERNANDEZ-INFANTE, C., MARTIN-GRANADO, V., ANGUITA, E., DE PEREDA, J. M.,

PORRAS, A. & GUERRERO, C. 2018. C3G, through its GEF activity, induces megakaryocytic differentiation and proplatelet formation. *Cell Commun Signal,* 16**,** 101.

PAI, E. F., KRENGEL, U., PETSKO, G. A., GOODY, R. S., KABSCH, W. & WITTINGHOFER, A. 1990. Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *EMBO J,* 9**,** 2351-9.

PALAO, N., SEQUERA, C., CUESTA, A. M., BAQUERO, C., BRAGADO, P., GUTIERREZ-UZQUIZA, A., SANCHEZ, A., GUERRERO, C. & PORRAS, A. 2022. C3G down-regulation enhances pro-migratory and stemness properties of oval cells by promoting an epithelial-mesenchymal-like process. *Int J Biol Sci,* 18**,** 5873-5884.

PARK, T. 2021. Crk and CrkL as Therapeutic Targets for Cancer Treatment. *Cells,* 10.

PARK, T. J. & CURRAN, T. 2008. Crk and Crk-like play essential overlapping roles downstream of disabled-1 in the Reelin pathway. *J Neurosci,* 28**,** 13551-62.

PASQUALUCCI, L., KHIABANIAN, H., FANGAZIO, M., VASISHTHA, M., MESSINA, M., HOLMES, A. B., OUILLETTE, P., TRIFONOV, V., ROSSI, D., TABBO, F., PONZONI, M., CHADBURN, A., MURTY, V. V., BHAGAT, G., GAIDANO, G., INGHIRAMI, G., MALEK, S. N., RABADAN, R. & DALLA-FAVERA, R. 2014. Genetics of follicular lymphoma transformation. *Cell Rep,* 6**,** 130-40.

PENG, W., XU, J., GUAN, X., SUN, Y., ZHANG, X. C., LI, X. & RAO, Z. 2011. Structural study of the Cdc25 domain from Ral-specific guanine-nucleotide exchange factor RalGPS1a. *Protein Cell,* 2**,** 308-19.

PETERSON, M. E. & LONG, E. O. 2008. Inhibitory receptor signaling via tyrosine phosphorylation of the adaptor Crk. *Immunity,* 29**,** 578-88.

PICKERING, C. R., ZHOU, J. H., LEE, J. J., DRUMMOND, J. A., PENG, S. A., SAADE, R. E., TSAI, K. Y., CURRY, J. L., TETZLAFF, M. T., LAI, S. Y., YU, J., MUZNY, D. M., DODDAPANENI, H., SHINBROT, E., COVINGTON, K. R., ZHANG, J., SETH, S., CAULIN, C., CLAYMAN, G. L., EL-NAGGAR, A. K., GIBBS, R. A., WEBER, R. S., MYERS, J. N., WHEELER, D. A. & FREDERICK, M. J. 2014. Mutational landscape of aggressive cutaneous squamous cell carcinoma. *Clin Cancer Res,* 20**,** 6582-92.

PONSIOEN, B., GLOERICH, M., RITSMA, L., REHMANN, H., BOS, J. L. & JALINK, K. 2009. Direct spatial control of Epac1 by cyclic AMP. *Mol Cell Biol,* 29**,** 2521-31.

POPOVIC, M., RENSEN-DE LEEUW, M. & REHMANN, H. 2013. Selectivity of CDC25 homology domain-containing guanine nucleotide exchange factors. *J Mol Biol,* 425**,** 2782-94.

POPOVIC, M., SCHOUTEN, A., RENSEN-DE LEEUW, M. & REHMANN, H. 2016. The structure of the Guanine Nucleotide Exchange Factor Rlf in complex with the small G-protein Ral identifies conformational intermediates of the exchange reaction and the basis for the selectivity. *J Struct Biol,* 193**,** 106-14.

POPOVICH, M. 2013. *Regulation and Selectivity of Exchange Factors for G-proteins of the Ras-family*

PORFIRI, E., EVANS, T., CHARDIN, P. & HANCOCK, J. F. 1994. Prenylation of Ras proteins is required for efficient hSOS1-promoted guanine nucleotide exchange. *J Biol Chem,* 269**,** 22672-7.

PUFALL, M. A. & GRAVES, B. J. 2002. Autoinhibitory domains: modular effectors of cellular regulation. *Annu Rev Cell Dev Biol,* 18**,** 421-62.

RADHA, V., MITRA, A., DAYMA, K. & SASIKUMAR, K. 2011. Signalling to actin: role of C3G, a multitasking guanine-nucleotide-exchange factor. *Biosci Rep,* 31**,** 231-44.

RADHA, V., RAJANNA, A., GUPTA, R. K., DAYMA, K. & RAMAN, T. 2008. The guanine nucleotide exchange factor, C3G regulates differentiation and survival of human neuroblastoma cells. *J Neurochem,* 107**,** 1424-35.

RADHA, V., RAJANNA, A., MITRA, A., RANGARAJ, N. & SWARUP, G. 2007. C3G is required for c-Abl-induced filopodia and its overexpression promotes filopodia formation. *Exp Cell Res,* 313**,** 2476-92.

RADHA, V., RAJANNA, A. & 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.

RAMEH, L. E., ARVIDSSON, A., CARRAWAY, K. L., 3RD, COUVILLON, A. D., RATHBUN, G., CROMPTON, A., VANRENTERGHEM, B., CZECH, M. P., RAVICHANDRAN, K. S., BURAKOFF, S. J., WANG, D. S., CHEN, C. S. & CANTLEY, L. C. 1997. A comparative analysis of the phosphoinositide binding specificity of pleckstrin homology domains. *J Biol Chem,* 272**,** 22059-66.

RAUEN, K. A. 2013. The RASopathies. *Annu Rev Genomics Hum Genet,* 14**,** 355-69.

REEDQUIST, K. A., FUKAZAWA, T., PANCHAMOORTHY, G., LANGDON, W. Y., SHOELSON, S. E., DRUKER, B. J. & BAND, H. 1996. Stimulation through the T cell receptor induces Cbl association with Crk proteins and the guanine nucleotide exchange protein C3G. *J Biol Chem,* 271**,** 8435-42.

REHMANN, H. 2006. Characterization of the activation of the Rap-specific exchange factor Epac by cyclic nucleotides. *Methods Enzymol,* 407**,** 159-73.

REHMANN, H., ARIAS-PALOMO, E., HADDERS, M. A., SCHWEDE, F., LLORCA, O. & BOS, J. L. 2008. Structure of Epac2 in complex with a cyclic AMP analogue and RAP1B. *Nature,* 455**,** 124-7.

REHMANN, H., DAS, J., KNIPSCHEER, P., WITTINGHOFER, A. & BOS, J. L. 2006. Structure of the cyclic-AMP-responsive exchange factor Epac2 in its auto-inhibited state. *Nature,* 439**,** 625-8.

REHMANN, H., RUEPPEL, A., BOS, J. L. & WITTINGHOFER, A. 2003. Communication between the regulatory and the catalytic region of the cAMP-responsive guanine nucleotide exchange factor Epac. *J Biol Chem,* 278**,** 23508-14.

REICHMAN, C. T., MAYER, B. J., KESHAV, S. & HANAFUSA, H. 1992. The product of the cellular crk gene consists primarily of SH2 and SH3 regions. *Cell Growth Differ,* 3**,** 451-60.

REUTHER, G. W. & DER, C. J. 2000. The Ras branch of small GTPases: Ras family members don't fall far from the tree. *Curr Opin Cell Biol,* 12**,** 157-65.

ROJAS, A. M., FUENTES, G., RAUSELL, A. & VALENCIA, A. 2012. The Ras protein superfamily: evolutionary tree and role of conserved amino acids. *J Cell Biol,* 196**,** 189-201.

ROJAS, J. M. & SANTOS, E. 2002. Ras genes and human cancer: different implications and different roles. *Current Genomics,* 3**,** 295-311.

ROOSE, J. P., MOLLENAUER, M., HO, M., KUROSAKI, T. & 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**,** 2732-45.

ROSSMAN, K. L. & SONDEK, J. 2005. Larger than Dbl: new structural insights into RhoA activation. *Trends Biochem Sci,* 30**,** 163-5.

SAKAI, R., IWAMATSU, A., HIRANO, N., OGAWA, S., TANAKA, T., MANO, H., YAZAKI, Y. & HIRAI, H. 1994. A novel signaling molecule, p130, forms stable complexes in vivo with v-Crk and v-Src in a tyrosine phosphorylation-dependent manner. *EMBO J,* 13**,** 3748-56.

SAKKAB, D., LEWITZKY, M., POSERN, G., SCHAEPER, U., SACHS, M., BIRCHMEIER, W. & FELLER, S. M. 2000. Signaling of hepatocyte growth factor/scatter factor (HGF) to the small GTPase Rap1 via the large docking protein Gab1 and the adapter protein CRKL. *J Biol Chem,* 275**,** 10772-8.

SALEH, T., JANKOWSKI, W., SRIRAM, G., ROSSI, P., SHAH, S., LEE, K. B., CRUZ, L. A., RODRIGUEZ, A. J., BIRGE, R. B. & KALODIMOS, C. G. 2016. Cyclophilin A promotes cell migration via the Abl-Crk signaling pathway. *Nat Chem Biol,* 12**,** 117-23.

SANTOS, E. & NEBREDA, A. R. 1989. Structural and functional properties of ras proteins. *FASEB J,* 3**,** 2151-63.

SANTOS, E., TRONICK, S. R., AARONSON, S. A., PULCIANI, S. & BARBACID, M. 1982. T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes. *Nature,* 298**,** 343-7.

SARASTE, M., SIBBALD, P. R. & WITTINGHOFER, A. 1990. The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci,* 15**,** 430-4.

SARKAR, P., REICHMAN, C., SALEH, T., BIRGE, R. B. & KALODIMOS, C. G. 2007. Proline cis-trans isomerization controls autoinhibition of a signaling protein. *Mol Cell,* 25**,** 413-26.

SARKAR, P., SALEH, T., TZENG, S. R., BIRGE, R. B. & KALODIMOS, C. G. 2011. Structural basis for regulation of the Crk signaling protein by a proline switch. *Nat Chem Biol,* 7**,** 51-7.

SAWADA, Y., NAKAMURA, K., DOI, K., TAKEDA, K., TOBIUME, K., SAITOH, M., MORITA, K., KOMURO, I., DE VOS, K., SHEETZ, M. & ICHIJO, H. 2001. Rap1 is involved in cell stretching modulation of p38 but not ERK or JNK MAP kinase. *J Cell Sci,* 114**,** 1221-7.

SCITA, G., NORDSTROM, J., CARBONE, R., TENCA, P., GIARDINA, G., GUTKIND, S., BJARNEGARD, M., BETSHOLTZ, C. & DI FIORE, P. P. 1999. EPS8 and E3B1 transduce signals from Ras to Rac. *Nature,* 401**,** 290-3.

SCRIMA, A., THOMAS, C., DEACONESCU, D. & WITTINGHOFER, A. 2008. The Rap-RapGAP complex: GTP hydrolysis without catalytic glutamine and arginine residues. *EMBO J,* 27**,** 1145-53.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods,* 9**,** 676-82.

SCHMIDPETER, P. A. & SCHMID, F. X. 2014. Molecular determinants of a regulatory prolyl isomerization in the signal adapter protein c-CrkII. *ACS Chem Biol,* 9**,** 1145-52.

SCHULTESS, J., DANIELEWSKI, O. & SMOLENSKI, A. P. 2005. Rap1GAP2 is a new GTPase-activating protein of Rap1 expressed in human platelets. *Blood,* 105**,** 3185-92.

SEABRA, M. C. & WASMEIER, C. 2004. Controlling the location and activation of Rab GTPases. *Curr Opin Cell Biol,* 16**,** 451-7.

SEQUERA, C., BRAGADO, P., MANZANO, S., ARECHEDERRA, M., RICHELME, S., GUTIERREZ-UZQUIZA, A., SANCHEZ, A., MAINA, F., GUERRERO, C. & PORRAS, A. 2020. C3G Is Upregulated in Hepatocarcinoma, Contributing to Tumor Growth and Progression and to HGF/MET Pathway Activation. *Cancers (Basel),* 12.

SHAH, N. H., LOBEL, M., WEISS, A. & KURIYAN, J. 2018. Fine-tuning of substrate preferences of the Src-family kinase Lck revealed through a high-throughput specificity screen. *Elife,* 7.

SHATTIL, S. J., KASHIWAGI, H. & PAMPORI, N. 1998. Integrin signaling: the platelet paradigm. *Blood,* 91**,** 2645-57.

SHIMONAKA, M., KATAGIRI, K., NAKAYAMA, T., FUJITA, N., TSURUO, T., YOSHIE, O. & KINASHI, T. 2003. Rap1 translates chemokine signals to integrin activation, cell polarization, and motility across vascular endothelium under flow. *J Cell Biol,* 161**,** 417-27.

SHIVAKRUPA, R., RADHA, V., SUDHAKAR, C. & SWARUP, G. 2003. Physical and functional interaction between Hck tyrosine kinase and guanine nucleotide exchange factor C3G results in apoptosis, which is independent of C3G catalytic domain. *J Biol Chem,* 278**,** 52188-94.

SIMON, J. A. & SCHREIBER, S. L. 1995. Grb2 SH3 binding to peptides from Sos: evaluation of a general model for SH3-ligand interactions. *Chem Biol,* 2**,** 53-60.

SMIT, L., VAN DER HORST, G. & 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**,** 8564-9.

SONDERMANN, H., SOISSON, S. M., BOYKEVISCH, S., YANG, S. S., BAR-SAGI, D. & KURIYAN, J. 2004. Structural analysis of autoinhibition in the Ras activator Son of sevenless. *Cell,* 119**,** 393-405.

SONGYANG, Z., CARRAWAY, K. L., 3RD, ECK, M. J., HARRISON, S. C., FELDMAN, R. A., MOHAMMADI, M., SCHLESSINGER, J., HUBBARD, S. R., SMITH, D. P., ENG, C. & ET AL. 1995. Catalytic specificity of protein-tyrosine kinases is critical for selective signalling. *Nature,* 373**,** 536-9.

STEFANINI, L. & BERGMEIER, W. 2016. RAP1-GTPase signaling and platelet function. *J Mol Med (Berl),* 94**,** 13-9.

STORK, P. J. & DILLON, T. J. 2005. Multiple roles of Rap1 in hematopoietic cells: complementary versus antagonistic functions. *Blood,* 106**,** 2952-61.

SZTUL, E., CHEN, P. W., CASANOVA, J. E., CHERFILS, J., DACKS, J. B., LAMBRIGHT, D. G., LEE, F. S., RANDAZZO, P. A., SANTY, L. C., SCHURMANN, A., WILHELMI, I., YOHE, M. E. & KAHN, R. A. 2019. ARF GTPases and their GEFs and GAPs: concepts and challenges. *Mol Biol Cell,* 30**,** 1249-1271.

TAMADA, M., SHEETZ, M. P. & SAWADA, Y. 2004. Activation of a signaling cascade by cytoskeleton stretch. *Dev Cell,* 7**,** 709-18.

TANAKA, S., HATTORI, S., KURATA, T., NAGASHIMA, K., FUKUI, Y., NAKAMURA, S. & MATSUDA, M. 1993. Both the SH2 and SH3 domains of human CRK protein are required for neuronal differentiation of PC12 cells. *Mol Cell Biol,* 13**,** 4409-15.

TANAKA, S., MORISHITA, T., HASHIMOTO, Y., HATTORI, S., NAKAMURA, S., SHIBUYA, M., MATUOKA, K., TAKENAWA, T., KURATA, T., NAGASHIMA, K. & ET AL. 1994. C3G, a guanine nucleotide-releasing protein expressed ubiquitously, binds to the Src homology 3 domains of CRK and GRB2/ASH proteins. *Proc Natl Acad Sci U S A,* 91**,** 3443-7.

TANAKA, S., OUCHI, T. & HANAFUSA, H. 1997. Downstream of Crk adaptor signaling pathway: activation of Jun kinase by v-Crk through the guanine nucleotide exchange protein C3G. *Proc Natl Acad Sci U S A,* 94**,** 2356-61.

TEN HOEVE, J., MORRIS, C., HEISTERKAMP, N. & GROFFEN, J. 1993. Isolation and chromosomal localization of CRKL, a human crk-like gene. *Oncogene,* 8**,** 2469-74.

THOMAS, C., FRICKE, I., SCRIMA, A., BERKEN, A. & WITTINGHOFER, A. 2007. Structural evidence for a common intermediate in small G protein-GEF reactions. *Mol Cell,* 25**,** 141-9.

THUMKEO, D., WATANABE, S. & NARUMIYA, S. 2013. Physiological roles of Rho and Rho effectors in mammals. *Eur J Cell Biol,* 92**,** 303-15.

UEMURA, N. & GRIFFIN, J. D. 1999. The adapter protein Crkl links Cbl to C3G after integrin ligation and enhances cell migration. *J Biol Chem,* 274**,** 37525-32.

VAN DEN BERGHE, N., COOL, R. H., HORN, G. & WITTINGHOFER, A. 1997. Biochemical characterization of C3G: an exchange factor that discriminates between Rap1 and Rap2 and is not inhibited by Rap1A(S17N). *Oncogene,* 15**,** 845-50.

VERCOULEN, Y., KONDO, Y., IWIG, J. S., JANSSEN, A. B., WHITE, K. A., AMINI, M., BARBER, D. L., KURIYAN, J. & ROOSE, J. P. 2017. A Histidine pH sensor regulates activation of the Ras-specific guanine nucleotide exchange factor RasGRP1. *Elife,* 6.

VETTER, I. R., LINNEMANN, T., WOHLGEMUTH, S., GEYER, M., KALBITZER, H. R., HERRMANN, C. & WITTINGHOFER, A. 1999. Structural and biochemical analysis of Ras-effector signaling via RalGDS. *FEBS Lett,* 451**,** 175-80.

VETTER, I. R. & WITTINGHOFER, A. 2001. The guanine nucleotide-binding switch in three dimensions. *Science,* 294**,** 1299-304.

VIGIL, D., CHERFILS, J., ROSSMAN, K. L. & DER, C. J. 2010. Ras superfamily GEFs and GAPs: validated and tractable targets for cancer therapy? *Nat Rev Cancer,* 10**,** 842-57.

VOSS, A. K., GRUSS, P. & THOMAS, T. 2003. The guanine nucleotide exchange factor C3G is necessary for the formation of focal adhesions and vascular maturation. *Development,* 130**,** 355-67.

VOSS, A. K., KREBS, D. L. & THOMAS, T. 2006. C3G regulates the size of the cerebral cortex neural precursor population. *EMBO J,* 25**,** 3652-63.

VUORI, K., HIRAI, H., AIZAWA, S. & RUOSLAHTI, E. 1996. Introduction of p130cas signaling complex formation upon integrin-mediated cell adhesion: a role for Src family kinases. *Mol Cell Biol,* 16**,** 2606-13.

WAKSMAN, G., SHOELSON, S. E., PANT, N., COWBURN, D. & KURIYAN, J. 1993. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell,* 72**,** 779-90.

WANG, W., FISHER, E. M., JIA, Q., DUNN, J. M., PORFIRI, E., DOWNWARD, J. & EGAN, S. E. 1995. The Grb2 binding domain of mSos1 is not required for downstream signal transduction. *Nat Genet,* 10**,** 294-300.

WU, X., KNUDSEN, B., FELLER, S. M., ZHENG, J., SALI, A., COWBURN, D., HANAFUSA, H. & 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**,** 215-26.

XIA, Y., CHU, W., QI, Q. & XUN, L. 2015. New insights into the QuikChange process guide the use of Phusion DNA polymerase for site-directed mutagenesis. *Nucleic Acids Res,* 43**,** e12.

YANG, Y., LI, M., YAN, Y., ZHANG, J., SUN, K., QU, J. K., WANG, J. S. & DUAN, X. Y. 2015. Expression of RAP1B is associated with poor prognosis and promotes an aggressive phenotype in gastric cancer. *Oncol Rep,* 34**,** 2385-94.

YOKOTE, K., HELLMAN, U., EKMAN, S., SAITO, Y., RONNSTRAND, L., SAITO, Y., HELDIN, C. H. & MORI, S. 1998. Identification of Tyr-762 in the platelet-derived growth factor alpha-receptor as the binding site for Crk proteins. *Oncogene,* 16**,** 1229-39.

YORK, R. D., YAO, H., DILLON, T., ELLIG, C. L., ECKERT, S. P., MCCLESKEY, E. W. & STORK, P. J. 1998. Rap1 mediates sustained MAP kinase activation induced by nerve growth factor. *Nature,* 392**,** 622-6.

YU, B., MARTINS, I. R., LI, P., AMARASINGHE, G. K., UMETANI, J., FERNANDEZ-ZAPICO, M. E., BILLADEAU, D. D., MACHIUS, M., TOMCHICK, D. R. & ROSEN, M. K. 2010. Structural and energetic mechanisms of cooperative autoinhibition and activation of Vav1. *Cell,* 140**,** 246-56.

YU, H., CHEN, J. K., FENG, S., DALGARNO, D. C., BRAUER, A. W. & SCHREIBER, S. L. 1994. Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell,* 76**,** 933-45.

YUWEN, T., XUE, Y. & SKRYNNIKOV, N. R. 2016. Role of Electrostatic Interactions in Binding of Peptides and Intrinsically Disordered Proteins to Their Folded Targets: 2. The Model of Encounter Complex Involving the Double Mutant of the c-Crk N-SH3 Domain and Peptide Sos. *Biochemistry,* 55**,** 1784-800.

ZHANG, F. L. & CASEY, P. J. 1996. Protein prenylation: molecular mechanisms and functional consequences. *Annu Rev Biochem,* 65**,** 241-69.

ZHANG, G., XIANG, B., YE, S., CHRZANOWSKA-WODNICKA, M., MORRIS, A. J., GARTNER, T. K., WHITEHEART, S. W., WHITE, G. C., 2ND, SMYTH, S. S. & LI, Z. 2011. Distinct roles for Rap1b protein in platelet secretion and integrin alphaIIbbeta3 outside-in signaling. *J Biol Chem,* 286**,** 39466-77.

ZHANG, M., ZHOU, S., ZHANG, L., ZHANG, J., CAI, H., ZHU, J., HUANG, C. & WANG, J. 2012. miR-518b is down-regulated, and involved in cell proliferation and invasion by targeting Rap1b in esophageal squamous cell carcinoma. *FEBS Lett,* 586**,** 3508-21.

ZHAO, H., PISZCZEK, G. & SCHUCK, P. 2015. SEDPHAT--a platform for global ITC analysis and global multi-method analysis of molecular interactions. *Methods,* 76**,** 137-148.

ZHENG, J., CHEN, R. H., CORBLAN-GARCIA, S., CAHILL, S. M., BAR-SAGI, D. & COWBURN, D. 1997. The solution structure of the pleckstrin homology domain of human SOS1. A possible structural role for the sequential association of diffuse B cell lymphoma and pleckstrin homology domains. *J Biol Chem,* 272**,** 30340-4.

# APPENDIX

## APPENDIX I

Table A1 includes the list of the primers used for mutagenesis of C3G sequence. Table A2 contains the list of primers used to generate the chimeric proteins of CrkL and CrkII. Table A3 includes the list of the primers used for mutagenesis of Crk sequences. Table A4 includes the list of primers used to subclone the abGFP4 nanobody.

**Table A1. Mutagenesis primers of C3G.**

| Name | Sequence (5´ to 3´) |
|---|---|
| C3Gh-NdeIX-For | GGAGGCAGCCACTCCTATGGTGGAGAGTCGC |
| C3Gh-NdeIX-Rev | GCGACTCTCCACCATAGGAGTGGCTGCCTCC |
| C3Gh-P1A-For | GATAATGGTCCTGCAGCAGCAGCGGCACCCGCGAAAAGACAGTCGGCGC |
| C3Gh-P1A-Rev | GCGCCGACTGTCTTTTCGCGGGTGCCGCTGCTGCTGCAGGACCATTATC |
| C3Gh-P2A-For | GCAGACAGATACGGCAGCTGCTGCCGCCGAGGCGAAGCGCAGGAG |
| C3Gh-P2A-Rev | CTCCTGCGCTTCGCCTCGGCGGCAGCAGCTGCCGTATCTGTCTGC |
| C3Gh-P3A-For | GACCCAGAAAAAGCAGCTCCTGCAGCAGAGGCGAAAAACAAACACATGCTGGCC |
| C3Gh-P3A-Rev | GGCCAGCATGTGTTTGTTTTTCGCCTCTGCTGCAGGAGCTGCTTTTTCTGGGTC |
| C3Gh-P4A-For | GGCCCCGGCAGCCGCCGCAGCCCCCGCGCAGCGGCAGCTG |
| C3Gh-P4A-Rev | CCGCTGCGCGGGGGCTGCGGCGGCTGCCGGGGCCAGCTCC |
| C3Gh-310-Spe-For | CATGAGCCGAGCCACTAGTGGCTCCAGTTTGCCTG |
| C3Gh-310-Spe-Rev | CAAACTGGAGCCACTAGTGGCTCGGCTCATGGGGG |
| C3Gh-637-KpnI-For | CTCAGCGGTCAGCGGGGTACCTGGGAAGGACAGCAGAG |
| C3Gh-637-KpnI-Rev | CTGTCCTTCCCAGGTACCCCGCTGACCGCTGAGGGATC |
| C3Gh-Y329F-For | GATGTTGACTGTTTCGCACAGAGGCGACTGTCAGG |
| C3Gh-Y329F-Rev | CCTCCTGACAGTCGCCTCTGTGCGAAACAGTCAAC |
| C3Gh-Y504F-For | CCCATCCGTCCCCTTCGCGCCCTTTGCTGC |
| C3Gh-Y504F-Rev | GCAGCAAAGGGCGCGAAGGGGACGGATGGG |
| C3Gh-Y554F-For | CAAACACATGCTGGCCTTCATGCAGTTGCTGGAG |
| C3Gh-Y554F-Rev | CTCCAGCAACTGCATGAAGGCCAGCATGTGTTTG |
| C3Gh-Y561F-For | GTTGCTGGAGGACTTCTCGGAGCCGCAGCCC |
| C3Gh-Y561F-Rev | GGGCTGCGGCTCCGAGAAGTCCTCCAGCAAC |
| C3Gh-Y570F-For | CCGCAGCCCTCTATGTTCTTCCAGACGCCACAGAAC |
| C3Gh-Y570F-Rev | GTTCTGTGGCGTCTGGAAGAACATAGAGGGCTGCGG |
| C3Gh-Y579F-For | CAGAACGAGCACATCTTCCAGCAGAAGAACAAG |
| C3Gh-Y579F-Rev | CTTGTTCTTCTGCTGGAAGATGTGCTCGTTCTG |
| C3Gh-Y590F-For | CTCCTCATGGAGGTATTCGGCTTCAGCGACTCC |
| C3Gh-Y590F-Rev | GGAGTCGCTGAAGCCGAATACCTCCATGAGGAG |
| C3Gh-X579Y-For | CAGAACGAGCACATCTACCAGCAGAAGAACAAG |
| C3Gh-X579Y-Rev | GAGCTTGTTCTTCTGCTGGTAGATGTGCTCG |
| C3Gh-X590Y-For | CTCCTCATGGAGGTATACGGCTTCAGCGACTCC |
| C3Gh-X590Y-Rev | GAAGGAGTCGCTGAAGCCGTATACCTCCATGAG |
| C3Gh-Y554H-For | CAAACACATGCTGGCCCACATGCAGTTGCTGGAG |
| C3Gh-Y554H-Rev | CTCCAGCAACTGCATGTGGGCCAGCATGTGTTTG |
| C3Gh-Y570N-For | CCGCAGCCCTCTATGTTCAACCAGACGCCACAGAAC |
| C3Gh-Y570N-Rev | GTTCTGTGGCGTCTGGTTGAACATAGAGGGCTGCGG |
| C3Gh-Y579C-For | CAGAACGAGCACATCTTCCAGCAGAAGAACAAG |
| C3Gh-Y579C-Rev | CTTGTTCTTCTGCTGGAAGATGTGCTCGTTCTG |
| C3Gh-Y590N-For | CTCCTCATGGAGGTAAACGGCTTCAGCGACTCC |
| C3Gh-Y590N-Rev | GGAGTCGCTGAAGCCGTTTACCTCCATGAGGAG |
| C3Gh-N575C-For | GACGCCACAGTGCGAGCACATCTACCAGCAGAAG |
| C3Gh-N575C-Rev | TAGATGTGCTCGCACTGTGGCGTCTGGTAGAAC |
| C3Gh-D767R-For | GTGGTGCGTGAGCTCTGCCTGGTGGAGTTGACAAAG |
| C3Gh-D767R-Rev | GTTCTTTGTCAACTCCACCAGGCAGAGCTCACGCACC |
| C3Gh-N1017R-For | GCAGTTCCGCATCCTCGACAGCATGCGCTGCTTCC |
| C3Gh-N1017R-Rev | GCTGGAAGCAGCGCATGCTGTCGAGGATGCGGAAC |

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

**Table A2. Primers used to amplify chimeric proteins of CrkL and CrkII.**

| Name | Sequence (5´ to 3´) |
|---|---|
| CrkL-001-For-NdeI | TGACCATGGCATATGTCCTCCGCCAGGTTC |
| CrkII-001-For-NdeI | TGACCATGGCATATGGCGGGCAACTTCGACTC |
| CrkII133-CrkL125-For | GATTCTCAGGCAGGAGGAGCTGGAATATGTACGGACTCTG |
| CrkII133-CrkL125-Rev | CAGAGTCCGTACATATTCCAGCTCCTCCTGCCTGAGAATC |
| CrkL124-CrkII134-For | CCTGCCTACAGCAGAAGATAACGCGGAGTATGTGCGAGCC |
| CrkL124-CrkII134-Rev | GGCTCGCACATACTCCGCGTTATCTTCTGCTGTAGGCAGG |
| CrkL182-CrkII192-For | GTCCCTTATGTCGAAAAGCTTGTGCCTGCCTCCGCCTCAG |
| CrkL182-CrkII192-Rev | CTGAGGCGGAGGCAGGCACAAGCTTTTCGACATAAGGGAC |
| CrkII191-CrkL183-For | TTACGTCGAGAAGTATAGAAGATCCTCACCACACGGAAAG |
| CrkII191-CrkL183-Rev | CTTTCCGTGTGGTGAGGATCTTCTATACTTCTCGACGTAA |
| CrkL-303-Stop-Rev-BglII | GCCGTCGACAGATCTACTCGTTTTCATCTGGGTTTTGAG |
| CrkII-304-Stop-Rev-BamHI | GCCGTCGACGGATCCTCAGCTGAAGTCCTCATCGGG |

Nucleotides of CrkL sequence are highlighted in blue, and the nucleotides of CrkII sequence in red.

**Table A3. Mutagenesis primers of CrkL.**

| Name | Sequence (5´ to 3´) |
|---|---|
| CrkL-BamHI-X-For | GGAATTCCAACAGTTATGGCATCCCAGAACCTGCTCATG |
| CrkL-BamHI-X-Rev | CATGAGCAGGTTCTGGGATGCCATAACTGTTGGAATTCC |
| CrkL-R39K-For | GTATGTTCCTCGTCAAGGATTCTTCCACCTGCCCTGGGG |
| CrkL-R39K-Rev | GCAGGTGGAAGAATCCTTGACGAGGAACATACCGTGGCG |
| CRKL-W160S-For | GAAGCCTGAAGAACAGTCGTGGAGTGCCCGGAAC |
| CRKL-W160S-Rev | GTTCCGGGCACTCCACGACTGTTCTTCAGGCTTC |
| CRKL-P174R-For | CGGGTTGGGATGATTCGTGTCCCTTATGTCGAAAAGC |
| CRKL-P174R-Rev | GCTTTTCGACATAAGGGACACGAATCATCCCAACCCG |
| CRKL-Y177S-For | GGGATGATTCCTGTCCCTTCTGTCGAAAAGCTTGTGAG |
| CRKL-Y177S-Rev | CTCACAAGCTTTTCGACAGAAGGGACAGGAATCATCCC |

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

**Table A4. Primers to subclone the abGFP4 nanobody in a modified pET22b vector.**

| Name | Sequence (5´ to 3´) |
|---|---|
| NbGFP-001-For-Nco | GAATTCCCATGGCCCAGGTTCAACTGGTGGAAAGCGGC |
| NbGFP-116-Rev-Xho | GCCGAATTCCTCGAGAGAGCTCACCGTCACCTGAGTCC |

# LIST OF FIGURES

# LIST OF TABLES

**APPENDIX**