

University of Groningen

Studying dimerization of Roco proteins in living cells

Nederveen-Schippers, Laura

DOI:
[10.33612/diss.207599709](https://doi.org/10.33612/diss.207599709)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2022

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Nederveen-Schippers, L. (2022). *Studying dimerization of Roco proteins in living cells: a tool to better understand Parkinson's Disease*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen. <https://doi.org/10.33612/diss.207599709>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

STUDYING DIMERIZATION OF ROCO PROTEINS IN LIVING CELLS

A tool to better understand Parkinson's Disease

Laura Nederveen-Schippers

About the cover: This is a schematic representation of a focused light beam with in the middle the confocal volume, illuminating a sample containing monomeric and dimeric LRRK2 proteins. As LRRK2 is linked to Green Fluorescent Protein (not drawn), protein molecules entering the blue light beam emit green light and have therefore been given a green color in this cartoon. Protein structures of LRRK2 have been published by Mayasnikov et al. (2021) Cell 184: 3519. For this thesis these were retrieved from the RCSB Protein Data Bank (PDB 7LHW and 7LHT) and formatted in PyMOL to create high resolution images viewed from different angles. The seize of the molecules on the cover is not to scale.



university of
 groningen

faculty of science
 and engineering



The research described in this thesis was carried out at the Department of Cell Biochemistry of the Groningen Biomolecular Sciences and Biotechnology institute (GBB) of the University of Groningen, and was funded by a Vidi grant from NWO, awarded to Prof. Arjan Kortholt.

Cover design: Ilse Modder | www.ilsemodder.nl

Layout design: Ilse Modder | www.ilsemodder.nl

Printed by: Gildeprint Enschede | www.gildeprint.nl



© 2022 Laura M. Nederveen-Schippers. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission in writing from the proprietor.



rijksuniversiteit
groningen

Studying dimerization of Roco proteins in living cells

A tool to better understand Parkinson's Disease

Proefschrift

ter verkrijging van de graad van doctor aan de
Rijksuniversiteit Groningen
op gezag van de
rector magnificus prof. dr. C. Wijmenga
en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

dinsdag 5 april 2022 om 16.15 uur

door

Laura Myrna Schippers

geboren op 17 december 1988
te Leeuwarden

Promotores

Prof. dr. P.J.M. van Haastert

Prof. dr. A. Kortholt

Beoordelingscommissie

Prof. dr. G. van den Bogaart

Prof. dr. A.J.M. Driessen

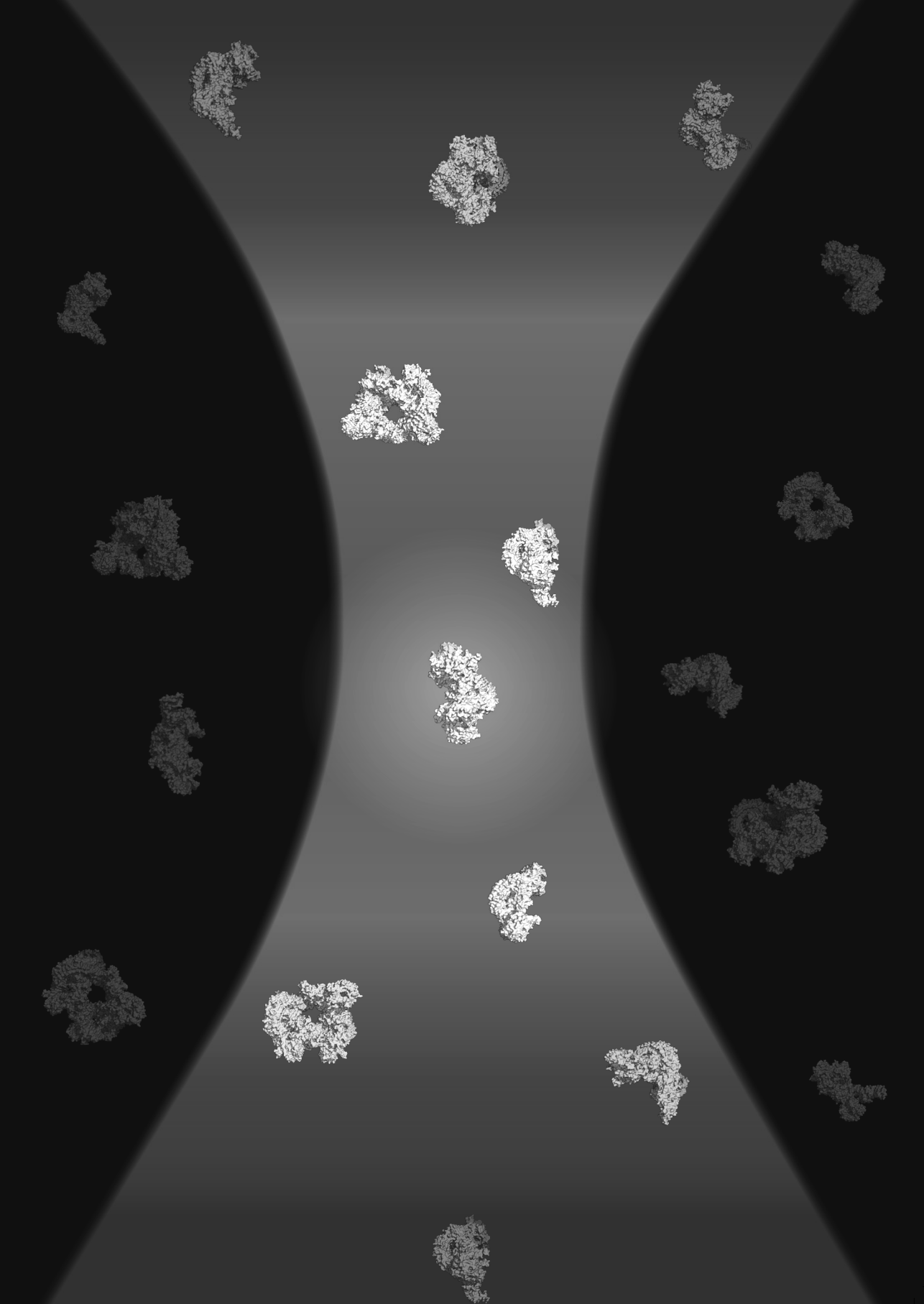
Prof. dr. C. Beta

Table of Contents

Abbreviations		9
Chapter 1	Introduction: The unconventional G-protein cycle of LRRK2 and Roco proteins Aim of the thesis	11
Chapter 2	A homologue of the Parkinson's disease-associated protein LRRK2 undergoes a monomer-dimer transition during GTP turnover	25
Chapter 3	Complex analysis of fluorescence intensity fluctuations of molecular compounds	55
Chapter 4	Combined FCS and PCH analysis to quantify protein dimerization in living cells	71
Chapter 5	Using Brightness and Diffusion Global Analysis to study the dimerization of Parkinson's Disease-related Roco proteins in living Dictyostelium and HEK293 cells	111
Chapter 6	Summary and Discussion	147
Addendum	Nederlandse samenvatting	158
	Acknowledgements	164
	List of publications	168
	Curriculum vitae	169

Abbreviations

PD	Parkinson's disease
LRRK2	Leucine rich repeat kinase 2
Roco4	Roc-of-complex proteins 4
FKBP12	FK506 binding protein 12
GFP	green fluorescent protein
diGFP	tandem-dimer GFP
Roc	Ras of complex proteins
COR	C-terminal of Roc
GAD	G-proteins activated by nucleotide dependent dimerization
GEF	G-nucleotide exchange factor
GAP	GTPase activating protein
GDP	Guanosindiphosphate
GTP	Guanosinetriphosphate
GppNHp	5'-Guanylyl imidodiphosphate
<i>Ct</i>	<i>Chlorobium tepidum</i>
<i>Mb</i>	<i>Methanosarcina barkeri</i>
<i>Hs</i>	<i>Homo sapiens</i>
<i>Dictyostelium</i>	<i>Dictyostelium discoideum</i>
HEK cells	Human embryonic kidney cells
SEC	size exclusion chromatography
SAXS	small angle X-ray scattering
MALS	multi-angle light scattering
MS	native mass spectrometry
AUC	analytical ultracentrifugation
EM	electron microscopy
FFS	fluorescence fluctuation spectroscopy
FCS	fluorescence correlation spectroscopy
FIDA	fluorescence intensity distribution analysis
PCH	photon counting histogram
ACF	autocorrelation function
PCD	photon counting distribution
PCA	principle component analysis
BDGA	brightness and diffusion global analysis
SD	standard deviation
CI	confidence interval



1

The unconventional G-protein cycle of LRRK2 and Roco proteins

Susanne Terheyden¹, Laura M. Nederveen-Schippers¹ and Arjan Kortholt^{1,*}

¹Department of Cell Biochemistry, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands.

*Corresponding author: a.kortholt@rug.nl

Author contributions: ST, LNS and AK wrote the manuscript.

Abstract

Mutations in the human Leucine- Rich Repeat Kinase 2 (LRRK2) are the most frequent cause of hereditary Parkinson's Disease (PD). LRRK2 belongs to the Roco family of proteins, which are characterized by the presence of a Ras-like G-domain (Roc), a C-terminal of Roc domain (COR), and a kinase domain. Despite intensive research much remains unknown about activity and the effect of PD associated mutations. Recent biochemical and structural studies suggest that LRRK2 and Roco proteins are non-canonical G-proteins that do not depend on Guanine nucleotide exchange factors (GEFs) or GTPase activating proteins (GAPs) for activation. In this review we will discuss the unusual G-protein cycle of LRRK2 in the context of the complex intramolecular LRRK2 activation mechanism.

Keywords

Leucine- Rich Repeat Kinase 2 (LRRK2), Roco proteins, G-proteins activated by nucleotide dependent dimerization (GAD), Parkinson's disease

Introduction

One of the most frequently mutated genes in hereditary PD is LRRK2 [1]. LRRK2 is a large, ~280 kD, multi-domain protein and belongs to the Roco family of complex proteins [2]. The catalytic core region consists of a Ras-of-complex proteins (Roc) G-domain, a C-terminal-of-Roc (COR) dimerization domain and a kinase domain. This region is flanked by protein-interaction domains: armadillo repeats (ARM), ankyrin repeats (ANK), a leucine rich repeats (LRR) at the N-terminus, and a WD40 repeat domain at the C-terminus [3]. One of the major challenges is that despite a vast amount of research on LRRK2 was conducted in the past years, the cellular and pathological functions are largely unknown. Several auto-phosphorylation sites within LRRK2 have been identified [4,5], however for a long time no other kinase substrate could be validated. Importantly, Steger *et al.* [6] recently identified a subgroup of Rab proteins as the first bona fide kinase substrate. Another caveat to study LRRK2 on a molecular level was the lack of reasonable amounts of high quality purified protein. Initially, our work with related Roco proteins from lower organisms provided the first important progress in the structural understanding of LRRK2. In the meanwhile Guaitoli *et al.* [7] are now able to purify the full-length protein, so more insights are to be expected.

PD-linked mutations in LRRK2 are found in nearly every domain, but are primarily located in the catalytic core of the protein (RocCOR-Kinase) [8]. Several of the PD-mutations have been linked to a decrease in GTPase and/or an increase in kinase activity [9–19]. However, it is not well understood how LRRK2 activity is regulated and how mutations in nearly every domain of the protein can alter the protein activity and function. Several lines of evidence suggest that the nucleotide binding state (GDP/GTP) of the Roc domain influences the kinase activity [9,20,21], and auto-phosphorylation by the kinase domain influences the GTPase activity of the Roc domain [22,23]. In this review we focus on the recent progress in the biochemical and structural characterization of the G-domain of LRRK2 and related Roco proteins and discuss the implications for the LRRK2 activation mechanism.

The Roc domain cycle is non canonical

The Roc domains of Roco proteins are homologous to proteins of the Ras superfamily and always appear in tandem with the COR domain. Like other G-Proteins, LRRK2 Roc is able to cycle between a GDP-bound state and a GTP-bound state [24]. Classical G-proteins depend on guanine nucleotide exchange factors (GEFs) for their GDP/GTP exchange because of their high affinity to the nucleotide and slow dissociation (**Figure 1A**). For example, the nucleotide dissociation of H-Ras ($2 * 10^{-5} \text{ s}^{-1}$) is accelerated by

10⁵ fold by its GEF Cdc25 (3.9 s⁻¹) [25]. A number of studies have found that Roco proteins have a relatively low affinity for nucleotides (in the μM range), as well as a fast dissociation of nucleotides [16,20,26–28]. Recently Rudi *et al.* [11] performed GTP binding assays with purified human LRRK2 Roc-COR-kinase fragments and measured a K_m of 343 μM, indicating that also LRRK2 has a low nucleotide affinity. The Roc domain alone has a nucleotide affinity of 553 μM [16]. These findings suggest that Roco proteins including LRRK2 do not need GEFs for their nucleotide exchange. The thus far only reported GEF for LRRK2, ARHGEF7 [29], has only a minor effect on nucleotide exchange and does not directly bind to the Roc domain, suggesting that it does not act as a classical GEF for LRRK2.

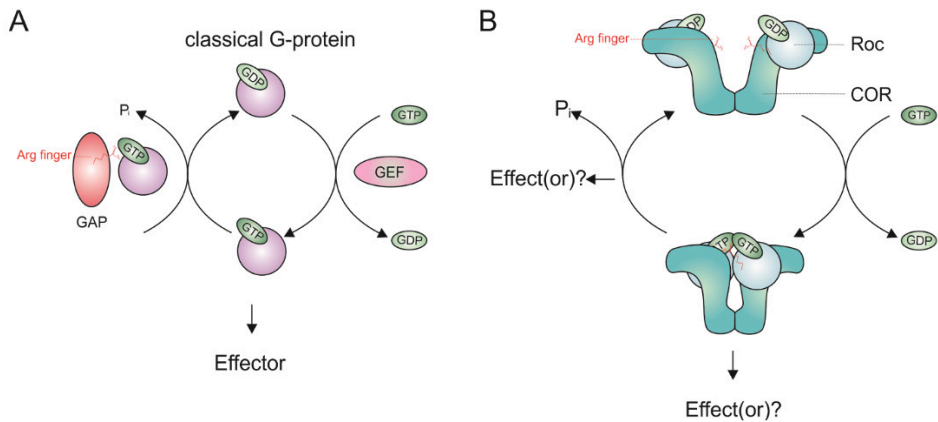


Figure 1: Classical (A) and proposed (LRRK2) RocCOR (B) G-protein cycle. The Arginine finger is highlighted in red. (GAP: GTPase activation protein; GEF: Guanine-nucleotide exchange factor; GTP: Guanosine-triphosphate; GDP: Guanosine-diphosphate; P_i: free inorganic phosphate; Roc: Ras of complex proteins; COR: C-terminal of Roc.)

The hydrolysis of GTP by conventional G-Proteins is stimulated by GTPase activating proteins (GAPs). The intrinsic GTPase activity is usually limited by the lack of catalytic residues or a non-optimal conformation. Often, the missing residue is an arginine which is then provided by the GAP (Arg-finger, **Figure 1B**) [24]. However, some G-proteins possess all required catalytic residues but need the GAPs to stabilize the catalytic machinery and the transition state of the hydrolysis reaction. Ran and RanGAP1 are such an example: Ran provides the catalytic residues, but requires RanGAP1 to position the catalytic glutamine correctly [30].

Roco proteins and LRRK2 can form dimers via the C-terminal subdomain of the COR domain. In 2008, the dimeric structure of the RocCOR tandem from *Chlorobium tepidum*

(*Ct*) was solved by Gotthardt *et al.* [26]. Moreover our recently published structural model of full-length LRRK2 [7] and pull-down experiments with various truncated LRRK2 constructs [28] confirmed that the C-terminus of COR is most likely the most important dimerization interface (also reported in [31]). However, in addition several other LRRK2 domain-domain interactions have been identified that might contribute to the stability and activity of the LRRK2 dimer [7,32]. Dimeric LRRK2 has a hydrolysis rate that is significantly faster than that of monomeric Roc domain (RocCOR-Kinase: $k_{\text{cat}} = 48 \text{ s}^{-1}$, [11]; Roc: $k_{\text{cat}} = 1.2 \text{ s}^{-1}$ [16]). Similarly, Roco proteins from lower organisms, such as *Mechanosarcina barkeri* (*Mb*) and *Ct*, also depend on dimerization (reviewed by Gilsbach *et al.* [33], [26]) and have a rather high intrinsic hydrolysis when dimerized. A truncated monomeric form of *Mb* Roco hydrolyses GTP significantly slower than the dimeric RocCOR construct [28]. Our structural studies suggested that the active site of one protomer is complemented by a residue from the other protomer [26,28]. Consistently, mutating the corresponding arginine in both prokaryotic Roco proteins impairs GTP hydrolysis [26,28].

Together this suggests that Roco proteins including LRRK2 belongs to the GAD (G-proteins activated by nucleotide dependent dimerization) class of molecules. The COR domain allows formation of a stable dimer, while Roc dimerization enables complementation of the active site and the subsequent hydrolysis of GTP to GDP and inorganic phosphate. Binding of effector or GTPase co-regulators, including ARHGEF7 [29], ArfGAP1 and RGS2 [21,34,35] might modulate the hydrolysis reaction.

The role of the Roc domain in LRRK2 activity and function

So far it remains to be determined if the kinase, the Roc domain or both regulate the output signal of LRRK2 [13,17,36]. Nevertheless the Roc, COR and kinase domains and their associated activities appear to be the key players in regulating LRRK2 function [37]. There is clear evidence that the LRRK2 R1441H/G/C mutations have reduced GTPase activity [16–19] and in addition some studies suggest that these mutations also result in increased kinase activity *in vitro* [9,17]. Furthermore, it has been consistently and independently shown by several different groups that the kinase activity of LRRK2 is dependent on the guanine nucleotide binding capacity of the Roc domain. Disruption of nucleotide binding results in reduced kinase activity of LRRK2 [20,38–40]. However, it is not entirely clear if the nucleotide state of the Roc domain influences kinase activity. For canonical G-proteins, the GDP-bound state is the inactive state and the GTP-bound state is active. In the GDP-state the switch regions are flexible and cannot be recognized by effectors [24], while in the active GTP-state the switch regions bind

1

to the γ -phosphate of the nucleotide and thereby adapt distinct conformations that are recognized by effectors. Liao *et al.* [16] suggested that also the GTP bound form of Roc domain is the active conformation that can stimulate LRRK2 kinase activity, either through direct binding via the switch regions or indirectly via conformational change in the protein [16]. In contrast, several other studies showed that LRRK2 kinase activity did not change upon addition of GDP, GTP, or non-hydrolysable GTP analogues [40,41], while others suggested that an intermediate state during hydrolysis presents the active state of LRRK2 [11,21]. In the latter case this would imply that, similar to the GAD protein MnmE [42], the G-protein needs to cycle between the GTP bound and GDP bound state in order to be able to fully activate the kinase. In MnmE, dissociation of the G-domain induces major conformational changes that drive the function of the protein. Rudi *et al.* [11] could show that within the dimer the flexibility of the Roc domain differs depending on the nucleotide bound, but that there are no major conformational changes. This might suggest that stabilization of a specific conformation, rather than major structural rearrangements, stimulates the activity of Roco proteins [11]. Recent data suggest that the cross-talk between the G-domain and kinase domain might be mediated by the N- and/or C-terminal domains of LRRK2 [7]. The first structural model for full-length LRRK2 implies that the protein has a very compact folding in which both the N-terminus and C-terminus are in close proximity to the kinase domain.

Recently it was shown that the Roc domain not only regulates the kinase domain, but that vice-versa the kinase domain also regulates the LRRK2 G-protein cycle. Mutants with increased kinase activity have normal GTP binding affinity [9,10,13,14], but reduced GTPase activity [41], while mutants with reduced kinase activity have increased GTPase activity [43]. In contrast to that, Taymans *et al.* [40] found that LRRK2 auto-phosphorylation of residues within the Roc domain increases the GTP-binding and thereby the activity of the Roc domain, suggesting that Roc functions downstream of the kinase domain [40]. Also, Webber *et al.* [22] found that auto-phosphorylation in the Roc domain stabilizes the GTP bound state and speculated that the two catalytic domains stimulate each other via a feed-forward loop, in which the domains are both upstream and downstream of each other.

Effect of PD-mutations in the G-domain and therapeutic opportunities

Four of the common LRRK2 mutations that were shown to segregate with PD are located in the RocCOR domain: R1441H/G/C and Y1699C. The R1441H/G/C substitutions were found to prolong the active GTP-bound state of the Roc domain by reduction of the GTPase activity [11,16–19] and by a 2-fold increase in GTP binding affinity [16].

Daniëls *et al.* [44] could show that the Y1699C mutation resulted in a stronger RocCOR interaction along with a weaker COR-COR dimer interaction and reduced GTPase activity. Structural studies with prokaryotic Roco protein revealed that both mutations are located in the conserved hydrophobic cleft between the Roc and COR domain [26,28]. Rudi *et al.* [11] investigated the dynamic changes in conformations upon mutation in the *Ct* RocCOR tandem model in more detail and could show that the dynamics of the Roc domain changes. They postulated that these changes in dynamics influence the RocCOR interface and thereby the GTP hydrolysis reaction [11].

Since PD-mutations in the Roc and COR domain have been linked to reduced GTPase activity, the G-domain might represent a good therapeutic target. Developing GTPase inhibitors may be of special interest, since most of the identified LRRK2 specific kinase inhibitors seem to have dramatic side effects in primates and rodents [45–47]. Targeting the G-domain could be done by using small compounds that increase the GTPase reaction or bind to the nucleotide binding site to block the G-protein cycle. It has been a major challenge to identify such compounds for the canonical G-protein Ras. However, since LRRK2 has a much lower nucleotide affinity and the GTPase activity most likely is regulated by dimerization, the non-canonical LRRK2 G-domain may provide a better therapeutic target. In this perspective it is very promising that Li and colleagues [48] demonstrated a beneficial effect of inhibitors of LRRK2 GTP-binding, reducing degeneration in PD cell- and mouse-models. These inhibitors did not alter the activity of the kinase domain directly, but do target the activity of the full-length protein.

Conclusions and Outlook

Based on the biochemical and structural data discussed in this review, we postulate a model for the activation mechanism of LRRK2 (**Figure 2**): LRRK2 can exist as a monomer in the cytosol and as an active kinase dimer/oligomer at the membrane [49]. Dimerization is mainly mediated via the COR domain. In this model, the Roc domain of one protomer can complete the catalytic site of the other Roc domain presumably via an arginine finger, thereby stimulating the GTPase activity. Thus only the dimeric protein at the membrane is able to complete the processive G-protein cycle, which might be required for optimal LRRK2 kinase activity. Because of the low nucleotide affinity of LRRK2, the nucleotide off rate is fast and there is no need for a GEF. Since the GTP concentration is approximately 10-fold higher than the GDP concentration [50] this also would imply that the monomeric LRRK2 protein in the cytosol is mainly GTP-bound. Additionally, auto-phosphorylation by the kinase domain may regulate the Roc domain activity [22]. Important questions that need to be answered include: how is oligomerization of LRRK2 regulated? Does the monomeric and presumable GTP-loaded

LRRK2 form has an additional cellular function? How does the Roc domain regulate kinase activity and ,vice versa, the kinase domain the GTPase activity? How does each PD-mutation interfere with these properties? And how can we therapeutically target LRRK2? To answer these questions and further understand the complex intramolecular LRRK2 activation mechanism, detailed biochemical data and high resolution structures of full-length protein and domain structures will be instrumental.

Funding

This work is supported by the Michael J. Fox Foundation and a VIDI [723.012.108] grant of the Netherlands Organisation for Scientific Research (NWO).

Competing Interests

The Authors declare that there are no competing interests associated with the manuscript.

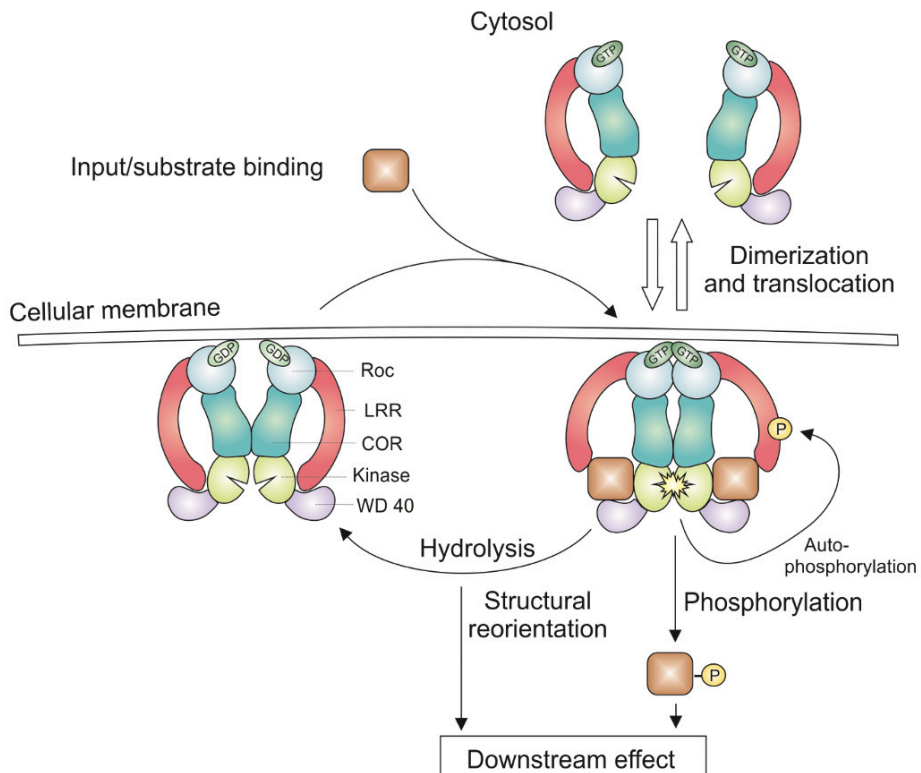


Figure 2: Schematic drawing of LRRK2 activation. When active as a dimer, the Roc domain does not need a GEF for nucleotide exchange and is able to hydrolyze GTP in the dimer. The kinase is also activated and can phosphorylate LRRK2 or its substrate(s). The kinase downstream targets are a subgroup of Rab proteins (possibly among others). Auto-phosphorylation might serve as a feed-forward/-back loop to increase LRRK2 activity.

Aim of the thesis

Aim of this thesis is to develop and use techniques to study the dimerization of LRRK2 in living cells. In **Chapter 2** we demonstrate that dimerization is an integral part of the activation cycle of Roco proteins *in vitro*. While this work was done with purified proteins and cell lysate, measurements in living cells would shed light on the regulation of dimerization in a cellular context. However, measuring monomer-dimer equilibria in living cells is a challenging task, often involving low signal-to-noise ratios or two different fluorophores that are difficult to align. In chapter 3-5 of this thesis, we have exploited a technique called Fluorescence Correlation Spectroscopy (FCS), in order to distinguish single-color fluorescently labelled monomers from dimers at the particle level. In **Chapter 3**, a method was explored to analyze these FCS data without fitting the data to a theoretical model, but using a Principle Component Analysis (PCA) instead. In **Chapter 4**, a still more sophisticated and accurate methodology was developed, which we named Brightness and Diffusion Global Analysis (BDGA). BDGA enabled us to quantify the proportion of monomer and dimer, as was demonstrated by comparing monomeric GFP and dimeric diGFP in the amoeba *Dictyostelium discoideum*. In **Chapter 5**, BDGA was further adapted for the use of complex proteins, using Roco4 in *Dictyostelium* as a model Roco protein. A two-component analysis enabled us to distinguish freely diffusing monomeric particles from slowly diffusing oligomeric particles, which were likely vesicle/membrane bound. Finally, the developed methodology was applied to Parkinson's Disease (PD)-related LRRK2, expressed in HEK293 cells. The obtained results are in agreement with the model that LRRK2 is monomeric in the cytosol and dimeric at membranes. As discussed in **Chapter 6**, the BDGA methodology developed in this thesis is now ready to be deployed to study the LRRK2 activation cycle in more detail, by exploring the influence of relevant mutations and upstream regulatory proteins, as well as dimerization modifying compounds, on the monomer-dimer equilibrium of LRRK2 in living cells.

References

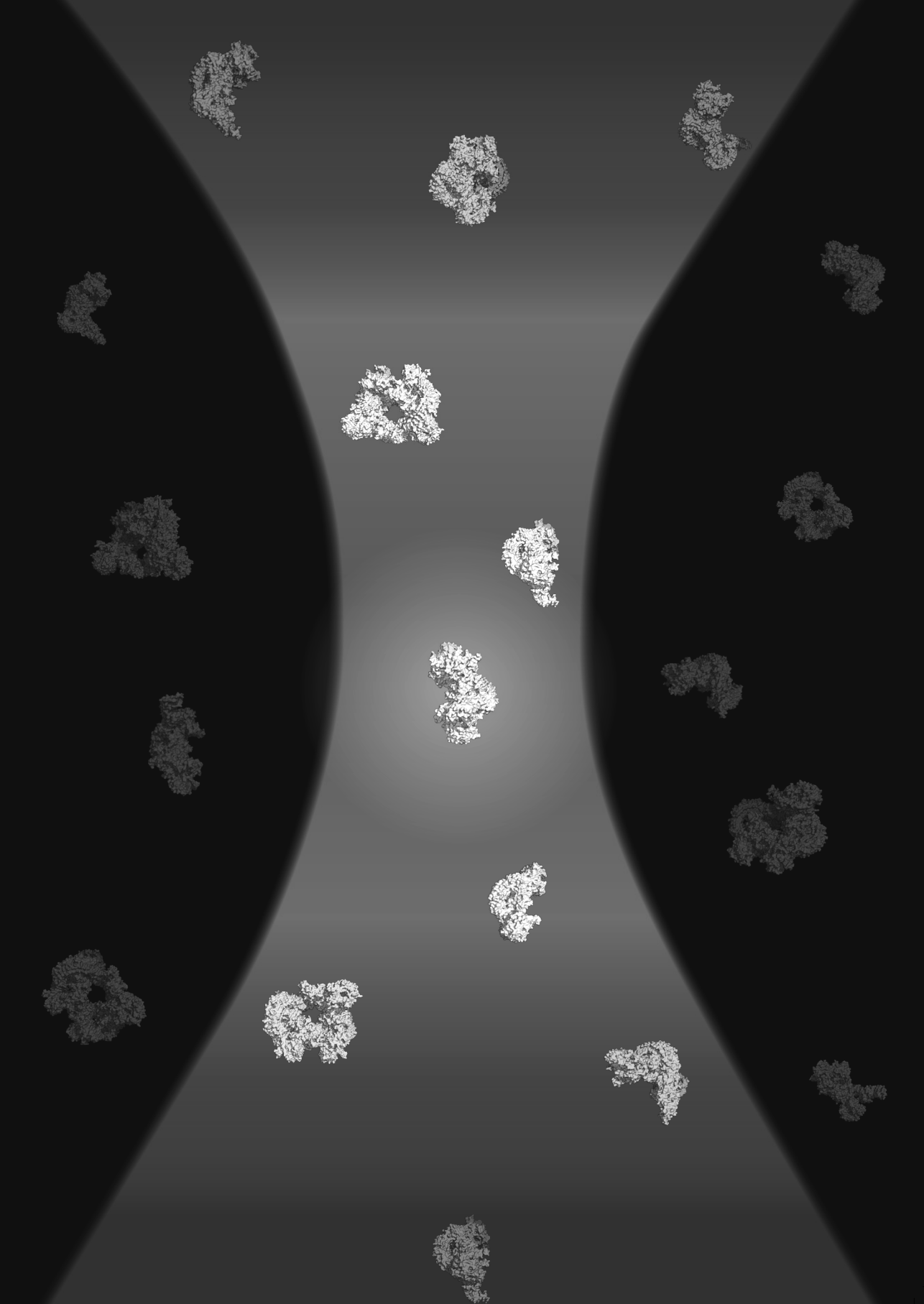
1. Sharma, M.; Ioannidis, J.P.A.; Aasly, J.O.; Annesi, G.; Brice, A.; Van Broeckhoven, C.; Bertram, L.; Bozi, M.; Crosiers, D.; Clarke, C.; et al. Large-scale replication and heterogeneity in Parkinson disease genetic loci. *Neurology* **2012**, *79*, 659–67, doi:10.1212/WNL.0b013e318264e353.
2. Marín, I. Ancient origin of the parkinson disease gene LRRK2. *J. Mol. Evol.* **2008**, *67*, 41–50, doi:10.1007/s00239-008-9122-4.
3. Mills, R.D.; Mulhern, T.D.; Liu, F.; Culvenor, J.G.; Cheng, H.-C. Prediction of the Repeat Domain Structures and Impact of Parkinsonism-Associated Variations on Structure and Function of all Functional Domains of Leucine-rich Repeat Kinase 2 (LRRK2). *Hum. Mutat.* **2014**, *2*, 1–74, doi:10.1002/humu.22515.
4. Gloeckner, C.J.; Boldt, K.; von Zweydford, F.; Helm, S.; Wiesent, L.; Sarioglu, H.; Ueffing, M. Phosphopeptide analysis reveals two discrete clusters of phosphorylation in the N-terminus and the Roc domain of the Parkinson-disease associated protein kinase LRRK2. *J. Proteome Res.* **2010**, *9*, 1738–1745, doi:10.1021/pr9008578.
5. Sheng, Z.; Zhang, S.; Bustos, D.; Kleinheinz, T.; Le Pichon, C.E.; Dominguez, S.L.; Solanoy, H.O.; Drummond, J.; Zhang, X.; Ding, X.; et al. Ser1292 autophosphorylation is an indicator of LRRK2 kinase activity and contributes to the cellular effects of PD mutations. *Sci. Transl. Med.* **2012**, *4*, 164ra161, doi:10.1126/scitranslmed.3004485.
6. Steger, M.; Tonelli, F.; Ito, G.; Davies, P.; Trost, M.; Vetter, M.; Wachter, S.; Lorentzen, E.; Duddy, G.; Wilson, S.; et al. Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *Elife* **2016**, *5*, 1–28, doi:10.7554/eLife.12813.001.
7. Guaitoli, G.; Raimondi, F.; Gilsbach, B.K.; Gómez-Llorente, Y.; Deyaert, E.; Renzi, F.; Li, X.; Schaffner, A.; Jagtap, P.K.A.; Boldt, K.; et al. Structural model of the dimeric Parkinson's protein LRRK2 reveals a compact architecture involving distant interdomain contacts. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, doi:10.1073/pnas.1523708113.
8. Cookson, M.R. The role of leucine-rich repeat kinase 2 (LRRK2) in Parkinson's disease. *Nat. Rev. Neurosci.* **2010**, *11*, 791–797, doi:10.1038/nrn2935.
9. West, A.B.; Moore, D.J.; Biskup, S.; Bugayenko, A.; Smith, W.W.; Ross, C.A.; Dawson, V.L.; Dawson, T.M. Parkinson's disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 16842–7, doi:10.1073/pnas.0507360102.
10. Anand, V.S.; Reichling, L.J.; Lipinski, K.; Stochaj, W.; Duan, W.; Kelleher, K.; Pungaliya, P.; Brown, E.L.; Reinhart, P.H.; Somberg, R.; et al. Investigation of leucine-rich repeat kinase 2 : enzymological properties and novel assays. *FEBS J.* **2009**, *276*, 466–78, doi:10.1111/j.1742-4658.2008.06789.x.
11. Rudi, K.; Ho, F.Y.; Gilsbach, B.K.; Pots, H.; Wittinghofer, A.; Kortholt, A.; Klare, J.P. Conformational heterogeneity of the Roc domains in *C. tepidum* Roc-COR and implications for human LRRK2 Parkinson mutations. *Biosci. Rep.* **2015**, *35*, e00254–e00254, doi:10.1042/BSR20150128.
12. Luzón-Toro, B.; Rubio de la Torre, E.; Delgado, A.; Pérez-Tur, J.; Hilfiker, S. Mechanistic insight into the dominant mode of the Parkinson's disease-associated G2019S LRRK2 mutation. *Hum. Mol. Genet.* **2007**, *16*, 2031–9, doi:10.1093/hmg/ddm151.
13. Greggio, E.; Jain, S.; Kingsbury, A.; Bandopadhyay, R.; Lewis, P.; Kaganovich, A.; van der Brug, M.P.; Beilina, A.; Blackinton, J.; Thomas, K.J.; et al. Kinase activity is required for the toxic effects of mutant LRRK2/dardarin. *Neurobiol. Dis.* **2006**, *23*, 329–41, doi:10.1016/j.nbd.2006.04.001.
14. Jaleel, M.; Nichols, R.J.; Deak, M.; Campbell, D.G.; Gillardon, F.; Knebel, A.; Alessi, D.R. LRRK2 phosphorylates moesin at threonine-558: characterization of how Parkinson's disease mutants affect kinase activity. *Biochem. J.* **2007**, *405*, 307–17, doi:10.1042/BJ20070209.
15. Ho, D.H.; Jang, J.; Joe, E.; Son, I.; Seo, H.; Seol, W. G2385R and I2020T Mutations Increase LRRK2 GTPase Activity. **2016**, *2016*, doi:10.1155/2016/7917128.
16. Liao, J.; Wu, C.-X.; Burlak, C.; Zhang, S.; Sahm, H.; Wang, M.; Zhang, Z.-Y.; Vogel, K.W.; Federici,

- M.; Riddle, S.M.; et al. Parkinson disease-associated mutation R1441H in LRRK2 prolongs the “active state” of its GTPase domain. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4055–60, doi:10.1073/pnas.1323285111.
17. Guo, L.; Gandhi, P.N.; Wang, W.; Petersen, R.B.; Wilson-Delfosse, A.L.; Chen, S.G. The Parkinson's disease-associated protein, leucine-rich repeat kinase 2 (LRRK2), is an authentic GTPase that stimulates kinase activity. *Exp. Cell Res.* **2007**, *313*, 3658–70, doi:10.1016/j.yexcr.2007.07.007.
 18. Li, X.; Tan, Y.-C.; Poulou, S.; Olanow, C.W.; Huang, X.-Y.; Yue, Z. Leucine-rich repeat kinase 2 (LRRK2)/PARK8 possesses GTPase activity that is altered in familial Parkinson's disease R1441C/G mutants. *J. Neurochem.* **2007**, *103*, 238–47, doi:10.1111/j.1471-4159.2007.04743.x.
 19. Lewis, P.A.; Greggio, E.; Beilina, A.; Jain, S.; Baker, A.; Cookson, M.R. The R1441C mutation of LRRK2 disrupts GTP hydrolysis. *Biochem. Biophys. Res. Commun.* **2007**, *357*, 668–71, doi:10.1016/j.bbrc.2007.04.006.
 20. Ito, G.; Okai, T.; Fujino, G.; Takeda, K.; Ichijo, H.; Katada, T.; Iwatsubo, T. GTP binding is essential to the protein kinase activity of LRRK2, a causative gene product for familial Parkinson's disease. *Biochemistry* **2007**, *46*, 1380–8, doi:10.1021/bi061960m.
 21. Biosa, A.; Trancikova, A.; Civiero, L.; Glauser, L.; Bubacco, L.; Greggio, E.; Moore, D.J. GTPase activity regulates kinase activity and cellular phenotypes of Parkinson's disease-associated LRRK2. *Hum. Mol. Genet.* **2013**, *22*, 1140–56, doi:10.1093/hmg/ddt522.
 22. Webber, P.J.; Smith, A.D.; Sen, S.; Renfrow, M.B.; Mobley, J.A.; West, A.B. Autophosphorylation in the leucine-rich repeat kinase 2 (LRRK2) GTPase domain modifies kinase and GTP-binding activities. *J. Mol. Biol.* **2011**, *412*, 94–110, doi:10.1016/j.jmb.2011.07.033.
 23. Liu, Z.; Mobley, J.A.; DeLucas, L.J.; Kahn, R.A.; West, A.B. LRRK2 autophosphorylation enhances its GTPase activity. *FASEB J.* **2016**, *30*, 336–47, doi:10.1096/fj.15-277095.
 24. Vetter, I.R.; Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* **2001**, *294*, 1299–304, doi:10.1126/science.1062023.
 25. Lenzen, C.; Cool, R.H.; Prinz, H.; Kuhlmann, J.; Wittinghofer, A. Kinetic analysis by fluorescence of the interaction between Ras and the catalytic domain of the guanine nucleotide exchange factor Cdc25Mm. *Biochemistry* **1998**, *37*, 7420–30, doi:10.1021/bi972621j.
 26. Gotthardt, K.; Weyand, M.; Kortholt, A.; Van Haastert, P.J.M.; Wittinghofer, A. Structure of the Roc-COR domain tandem of *C. tepidum*, a prokaryotic homologue of the human LRRK2 Parkinson kinase. *EMBO J.* **2008**, *27*, 2239–49, doi:10.1038/emboj.2008.150.
 27. Civiero, L.; Vancaenenbroeck, R.; Belluzzi, E.; Beilina, A.; Lobbestael, E.; Reyniers, L.; Gao, F.; Micetic, I.; de Maeyer, M.; Bubacco, L.; et al. Biochemical Characterization of Highly Purified Leucine-Rich Repeat Kinases 1 and 2 Demonstrates Formation of Homodimers. *PLoS One* **2012**, *7*, e43472, doi:10.1371/journal.pone.0043472.
 28. Terheyden, S.; Ho, F.Y.; Gilsbach, B.K.; Wittinghofer, A.; Kortholt, A. Revisiting the Roco G protein cycle. *Biochem. J.* **2014**, doi:10.1042/BJ20141095.
 29. Haebig, K.; Gloeckner, C.J.; Miralles, M.G.; Gillardon, F.; Schulte, C.; Riess, O.; Ueffing, M.; Biskup, S.; Bonin, M. ARHGEF7 (Beta-PIX) acts as guanine nucleotide exchange factor for leucine-rich repeat kinase 2. *PLoS One* **2010**, *5*, e13762, doi:10.1371/journal.pone.0013762.
 30. Seewald, M.J.; Körner, C.; Wittinghofer, A.; Vetter, I.R. RanGAP mediates GTP hydrolysis without an arginine finger. *Nature* **2002**, *415*, 662–666, doi:10.1038/415662a.
 31. Tsika, E.; Moore, D.J. Contribution of GTPase activity to LRRK2-associated Parkinson disease. *Small GTPases* **2013**, *4*, 164–170.
 32. Deng, J.; Lewis, P. a; Greggio, E.; Sluch, E.; Beilina, A.; Cookson, M.R. Structure of the ROC domain from the Parkinson's disease-associated leucine-rich repeat kinase 2 reveals a dimeric GTPase. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 1499–1504.
 33. Gilsbach, B.K.; Kortholt, A. Structural biology of the LRRK2 GTPase and kinase domains: implications

for regulation. *Front. Mol. Neurosci.* **2014**, *7*, 32, doi:10.3389/fnmol.2014.00032.

34. Xiong, Y.; Yuan, C.; Chen, R.; Dawson, T.M.; Dawson, V.L. ArfGAP1 is a GTPase activating protein for LRRK2: reciprocal regulation of ArfGAP1 by LRRK2. *J. Neurosci.* **2012**, *32*, 3877–86, doi:10.1523/JNEUROSCI.4566-11.2012.
35. Dusonchet, J.; Li, H.; Guillily, M.; Liu, M.; Stafa, K.; Derada, C.; Liu-yesucevitz, L.; Schneider, B.L.; Aebischer, P.; Derada Troletti, C.; et al. A Parkinson's disease gene regulatory network identifies the signaling protein RGS2 as a modulator of LRRK2 activity and neuronal toxicity. *Hum. Mol. Genet.* **2014**, *23*, 1–19, doi:10.1093/hmg/ddu202.
36. Gandhi, P.N.; Wang, X.; Zhu, X.; Chen, S.G.; Wilson-Delfosse, A.L. The Roc domain of leucine-rich repeat kinase 2 is sufficient for interaction with microtubules. *J. Neurosci. Res.* **2008**, *86*, 1711–20, doi:10.1002/jnr.21622.
37. Zimprich, A.; Biskup, S.; Leitner, P.; Lichtner, P.; Farrer, M.; Lincoln, S.; Kachergus, J.; Hulihan, M.; Uitti, R.J.; Calne, D.B.; et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **2004**, *44*, 601–7, doi:10.1016/j.neuron.2004.11.005.
38. Smith, W.W.; Pei, Z.; Jiang, H.; Dawson, V.L.; Dawson, T.M.; Ross, C.A. Kinase activity of mutant LRRK2 mediates neuronal toxicity. *Nat. Neurosci.* **2006**, *9*, 1231–3, doi:10.1038/nn1776.
39. West, A.B.; Moore, D.J.; Choi, C.; Andrabi, S.A.; Li, X.; Dikeman, D.; Biskup, S.; Zhang, Z.; Lim, K.-L.; Dawson, V.L.; et al. Parkinson's disease-associated mutations in LRRK2 link enhanced GTP-binding and kinase activities to neuronal toxicity. *Hum. Mol. Genet.* **2007**, *16*, 223–32, doi:10.1093/hmg/ddl471.
40. Taymans, J.-M.; Vancraenenbroeck, R.; Ollikainen, P.; Beilina, A.; Lobbestael, E.; De Maeyer, M.; Baekelandt, V.; Cookson, M.R. LRRK2 kinase activity is dependent on LRRK2 GTP binding capacity but independent of LRRK2 GTP binding. *PLoS One* **2011**, *6*, e23207, doi:10.1371/journal.pone.0023207.
41. Liu, M.; Kang, S.; Ray, S.; Jackson, J.; Zaitsev, A.D.; Gerber, S.A.; Cuny, G.D.; Glicksman, M.A. Kinetic, mechanistic, and structural modeling studies of truncated wild-type leucine-rich repeat kinase 2 and the G2019S mutant. *Biochemistry* **2011**, *50*, 9399–408, doi:10.1021/bi201173d.
42. Prado, S.; Villarroya, M.; Medina, M.; Armengod, M.-E.E. The tRNA-modifying function of MnmE is controlled by post-hydrolysis steps of its GTPase cycle. *Nucleic Acids Res.* **2013**, *41*, 6190–6208, doi:10.1093/nar/gkt320.
43. Greggio, E.; Taymans, J.-M.M.; Zhen, E.Y.; Ryder, J.; Vancraenenbroeck, R.; Beilina, A.; Sun, P.; Deng, J.; Jaffe, H.; Baekelandt, V.; et al. The Parkinson's disease kinase LRRK2 autophosphorylates its GTPase domain at multiple sites. *Biochem. Biophys. Res. Commun.* **2009**, *389*, 449–454, doi:10.1016/j.bbrc.2009.08.163.
44. Daniëls, V.; Vancraenenbroeck, R.; Law, B.M.H.; Greggio, E.; Lobbestael, E.; Gao, F.; De Maeyer, M.; Cookson, M.R.; Harvey, K.; Baekelandt, V.; et al. Insight into the mode of action of the LRRK2 Y1699C pathogenic mutant. *J. Neurochem.* **2011**, *116*, 304–315, doi:10.1111/j.1471-4159.2010.07105.x.
45. Fuji, R.N.; Flagella, M.; Baca, M.; S. Baptista, M.A.; Brodbeck, J.; Chan, B.K.; Fiske, B.K.; Honigberg, L.; Jubb, A.M.; Katavolos, P.; et al. Effect of selective LRRK2 kinase inhibition on nonhuman primate lung. *Sci. Transl. Med.* **2015**, *7*, 273ra15–273ra15, doi:10.1126/scitranslmed.aaa3634.
46. Baptista, M.A.S.; Dave, K.D.; Frasier, M.A.; Sherer, T.B.; Greeley, M.; Beck, M.J.; Varsho, J.S.; Parker, G.A.; Moore, C.; Churchill, M.J.; et al. Loss of Leucine-Rich Repeat Kinase 2 (LRRK2) in Rats Leads to Progressive Abnormal Phenotypes in Peripheral Organs. *PLoS One* **2013**, *8*, e80705, doi:10.1371/journal.pone.0080705.
47. Herzig, M.C.; Kolly, C.; Persohn, E.; Theil, D.; Schweizer, T.; Hafner, T.; Stemmelen, C.; Troxler, T.J.; Schmid, P.; Danner, S.; et al. LRRK2 protein levels are determined by kinase function and are crucial for kidney and lung homeostasis in mice. *Hum. Mol. Genet.* **2011**, *20*, 4209–23, doi:10.1093/hmg/ddr348.
48. Li, T.; Yang, D.; Zhong, S.; Thomas, J.M.; Xue, F.; Liu, J.; Kong, L.; Voulalas, P.; Hassan, H.E.; Park, J.-S.; et al. Novel LRRK2 GTP-binding inhibitors reduced degeneration in Parkinson's disease cell and mouse models. *Hum. Mol. Genet.* **2014**, doi:10.1093/hmg/ddu341.
49. James, N.G.; Digman, M. a.; Gratton, E.; Barylko, B.; Ding, X.; Albanesi, J.P.; Goldberg, M.S.; Jameson,

- D.M. Number and brightness analysis of LRRK2 oligomerization in live cells. *Biophys. J.* **2012**, *102*, L41-3, doi:10.1016/j.bpj.2012.04.046.
50. Traut, T.W. Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem.* **1994**, *140*, 1-22, doi:10.1007/BF00928361.



2

A homologue of the Parkinson's disease-associated protein LRRK2 undergoes a monomer-dimer transition during GTP turnover

Egon Deyaert^{1,2,\$}, Lina Wauters^{1,2,3,\$}, Giambattista Guaitoli^{4,5}, Albert Konijnenberg⁶, Margaux Leemans^{1,2}, Susanne Terheyden^{3,7}, Arsen Petrovic⁸, Rodrigo Gallardo^{9,10}, **Laura M. Nederveen-Schippers**³, Panagiotis S. Athanasopoulos³, Henderikus Pots³, Peter J.M. Van Haastert³, Frank Sobott^{6,11,12}, Christian Johannes Gloeckner^{4,5}, Rouslan Efremov^{1,2}, Arjan Kortholt³, Wim Versées^{1,2,*}

¹VIB-VUB Center for Structural Biology, Pleinlaan 2, 1050 Brussels, Belgium. ²Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. ³Department of Cell Biochemistry, University of Groningen, Groningen 9747 AG, The Netherlands. ⁴German Center for Neurodegenerative Diseases (DZNE), 72076 Tübingen, Germany. ⁵Eberhard Karls University, Institute for Ophthalmic Research, Center for Ophthalmology, 72076 Tübingen, Germany. ⁶Department of Chemistry, Biomolecular & Analytical Mass Spectrometry group, University of Antwerp, 2020 Antwerp, Belgium. ⁷Structural Biology Group, Max-Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. ⁸Department of Mechanistic Cell Biology, Max-Planck Institute of Molecular Physiology, 44227 Dortmund, Germany. ⁹VIB Switch Laboratory, VIB, 3000 Leuven, Belgium. ¹⁰Switch Laboratory, Department of Cellular and Molecular Medicine, KU Leuven, Herestraat 49, PB 802, 3000 Leuven, Belgium. ¹¹Astbury Centre for Structural Molecular Biology, University of Leeds, LS2 9JT Leeds, United Kingdom. ¹²School of molecular and Cellular Biology, University of Leeds, LS2 9JT Leeds, United Kingdom. ^{\$}These authors contributed equally to this work. ^{*}Corresponding author: wim.versees@vib-vub.be

Author Contributions: ED, LW, PV, AP, **LNS**, RE, CJG, FS, ArK and WV designed the experiments. ED, GG, ML, HP, PA, and ST purified the proteins. LW and RE performed EM analysis. ED and RG performed the SEC-MALS analysis. ED and WV performed the SEC-SAXS analysis. ST, AP and ED performed the AUC analysis. AIK and ED performed the native MS analysis. ED performed the kinetic and FRET experiments. **LNS** performed and analyzed the FCS experiments. ED and WV wrote the manuscript, with contributions from **LNS** regarding FCS. All the authors read and edited the manuscript.

Part of this chapter has been published in: Nature Communications. 2017; Nature Communications 2017 Oct 18;8(1):1008. Doi: 10.1038/s41467-017-01103-4

Abstract

Mutations in LRRK2 are a common cause of genetic Parkinson's disease (PD). LRRK2 is a multi-domain Roco protein, harbouring kinase and GTPase activity. In analogy with a bacterial homologue, LRRK2 was proposed to act as a GTPase activated by dimerization (GAD), while recent reports suggest LRRK2 to exist under a monomeric and dimeric form *in vivo*. It is however unknown how LRRK2 oligomerization is regulated. Here, we show that oligomerization of a homologous bacterial Roco protein depends on the nucleotide load. The protein is mainly dimeric in the nucleotide-free and GDP-bound states, while it forms monomers upon GTP binding, leading to a monomer-dimer cycle during GTP hydrolysis. An analogue of a PD-associated mutation stabilizes the dimer and decreases the GTPase activity. Importantly, FCS data from HEK cells expressing GFP-RocCOR suggest that oligomerization of LRRK2 is also nucleotide-dependent. This work thus provides insights into the conformational cycle of Roco proteins and suggests a link between oligomerization and disease-associated mutations in LRRK2.

Introduction

Mutations in the gene encoding leucine-rich-repeat kinase 2 (LRRK2) are the most common genetic cause of Parkinson's disease (PD) [1,2]. LRRK2 mutations account for 5-6 % of familial PD cases, and are identified as a risk factor for sporadic forms of the disease [3]. LRRK2 is a large (2527 amino acids) multi-domain protein, belonging to the Roco protein family. This protein family is characterized by the presence of a GTPase domain, called Roc (Ras of complex proteins), fused to a COR domain (C-terminal of Roc) [4,5]. In many cases this Roc-COR module is preceded by a leucine-rich-repeat (LRR) domain and followed by a protein kinase domain. In LRRK2 a number of additional protein-protein interaction domains are present [6,7]. The pathogenic LRRK2 mutations mainly cluster in the catalytic Roc-COR and kinase domains, and the most prevalent mutations result in decreased GTPase activity and/or enhanced kinase activity [8–13]. This coupled phenotype may point towards an intramolecular regulatory mechanism between the Roc and kinase domains, thus underscoring the central role of the Roc GTPase cycle in PD pathology.

Significant progress in our understanding of the structure and mechanism of the Roc-COR module of LRRK2 comes from studies with related Roco proteins from prokaryotes and lower eukaryotes [7,14–16]. Most importantly, the model that LRRK2 functions as a GAD (G protein activated by nucleotide-dependent dimerization) is heavily based on the crystal structure of the dimeric Roc-COR module of the Roco protein from *Chlorobium tepidum* [14,17,18]. This model implies that the COR domain acts as a permanent dimerization device and that the stimulation of GTPase activity depends on reciprocal complementation of two Roc active sites [7,14,15,17,19]. Similar to the prokaryotic Roco proteins, various studies report that LRRK2 can also form dimers through its Roc-COR domain *in vitro* [15,20–22], although some other studies suggest that the protein is mainly monomeric [23]. A number of recent results indicate that *in vivo* functional LRRK2 cycles between a predominantly monomeric kinase-inactive form in the cytosol and a dimeric kinase-active form at the plasma membrane [24,25]. However, so far the mechanisms regulating these changes in LRRK2 translocation and oligomerization are poorly understood.

Here, we show that Roco proteins cycle between a monomeric and dimeric form during GTP turnover. Using small angle X-ray scattering (SAXS), multi-angle light scattering (MALS), native mass spectrometry (MS), analytical ultracentrifugation (AUC) and electron microscopy (EM), we demonstrate that the *C. tepidum* Roco protein is mainly dimeric in the nucleotide-free and GDP-bound states, while it is mainly monomeric when bound to GTP. Moreover, using time resolved FRET and EM we show that the GTP-induced monomerization occurs on a catalytically relevant time scale and that the monomer-

dimer cycle occurs concomitant with GTP turnover. A mutation linked to PD decreases the GTPase activity by interfering with the monomer-dimer equilibrium. Furthermore, FCS analysis on cell lysates from HEK cells expressing human GFP-RocCOR suggest that this nucleotide dimerization mechanism is conserved in LRRK2. Together these results shed new light on a long-standing discussion regarding the oligomeric state of Roco/LRRK2 proteins, and propose a model for their GTP hydrolysis mechanism.

Results

GppNHp binding induces monomerization of the Roc-COR module

Like most prokaryotic Roco proteins, the Roco protein from *Chlorobium tepidum* (CtRoco) consists of an N-terminal LRR domain (a.a. 1 – 411), a central Roc-COR module (a.a. 412 – 946) and a C-terminal region of unknown structure and function (a.a. 947 - 1102) (Supplementary Fig. 1) [4,14,15]. A crystal structure of the Roc-COR domain construct of CtRoco (CtRoc-COR) was solved, with the Roc GTPase domain in a nucleotide-free state [14]. This structure shows the protein as a homodimer, for which most of the contacts between the subunits are mediated via COR domain residues. To investigate the influence of nucleotide binding on the conformation of CtRoc-COR we set out to perform small angle X-ray scattering coupled to size-exclusion chromatography (SEC-SAXS) experiments with the protein in either the nucleotide-free state or saturated with GDP or the GTP mimic 5'-guanylyl imidodiphosphate (GppNHp) (Supplementary Fig. 2). The scattering profile of the nucleotide-free protein is in excellent agreement with the symmetrical CtRoc-COR dimer of the crystal structure (pdb 3DPU, Gotthardt *et al.* 2008), and a comparison of the experimental and theoretical scattering profiles yields a χ^2 -value of 0.9 (Fig. 1a). Moreover, a good agreement is found between the *ab initio* molecular envelope based on the SAXS profile and the CtRoc-COR crystal structure (Fig. 1b). However, when CtRoc-COR is saturated with either GDP or GppNHp conformational changes are taking place, as the experimental scattering curves yield a worse fit with the theoretical scattering curve based on the nucleotide-free crystal structure (translated in a $\chi^2 = 1.6$ and $\chi^2 = 6.0$ for GDP- and GppNHp-bound protein respectively). To obtain further insights in the nature of these conformational changes we calculated the estimated molecular masses based on the Porod volumes. For both the nucleotide-free and the GDP-bound form, a molecular mass around 120 kDa is found, which is in good agreement with the expected molecular mass of the dimer (theoretical $MM_{\text{dimer}} = 130$ kDa) (Fig. 1c). In the GppNHp-bound state an average molecular mass of 90 kDa is found in between the values expected for a monomer and a dimer. This indicates a monomer/dimer equilibrium with overlapping peaks in the chromatogram at the concentration used in SEC-SAXS (50 μ l of an 8 mg/ml protein solution injected on the column). Such a shift in oligomerization is also translated in the

pair-distance distribution functions, where an overlay shows that the curves overlap for the nucleotide-free and GDP loaded protein, while a shift toward on average smaller distances is found for the GppNHp-bound protein (Supplementary Fig. 2d).

Next, we performed multi-angle light scattering coupled to size-exclusion chromatography (SEC-MALS) and sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments to determine the molecular mass of CtRoc-COR in different nucleotide states (Fig. 1c,d). For the SEC-MALS experiments, 10 μ l of 8 mg/ml protein solutions were injected on the SEC column. For the nucleotide-free and GDP-bound forms a molecular mass of 120 kDa and 125 kDa is obtained respectively, again corresponding to a dimer. For the GppNHp-loaded protein SEC-MALS yielded a molecular mass of 78 kDa, just slightly above the expected value of a monomer. The SV-AUC experiments were performed at an even lower protein concentration (0.3 mg/ml, 4.6 μ M) (Fig. 1c and Supplementary Fig. 3). Again, for nucleotide-free and GDP-loaded CtRoc-COR molecular masses of 121 and 115 kDa are obtained (sedimentation coefficients of 4.7S and 4.8S, respectively). In the presence of GppNHp the monomer-dimer equilibrium is almost completely shifted to the monomer, clearly seen as a shift in sedimentation coefficient from around 4.8S (dimer) to 3.6S (monomer), corresponding to a molecular mass of 69 kDa.

Finally, to resolve monomeric and dimeric species we performed native-MS experiments (Fig. 1e). The mass spectrum of nucleotide-free CtRoc-COR shows exclusive dimeric species ($MM_{\text{exp}} = 130$ kDa). Also with GDP dimeric species dominate, while a very small amount of monomer appears ($MM_{\text{exp}} = 65$ kDa). However, in agreement with our previous data, we observe a clear shift toward the monomeric species in the presence of GppNHp, resulting in an approximately 50:50 monomer/dimer ratio under the conditions used in this experiment.

In conclusion, all these experiments show that while the CtRoc-COR is mainly a dimer in the nucleotide-free and GDP-bound form, the GTP mimic GppNHp induces monomerization in a concentration dependent manner.

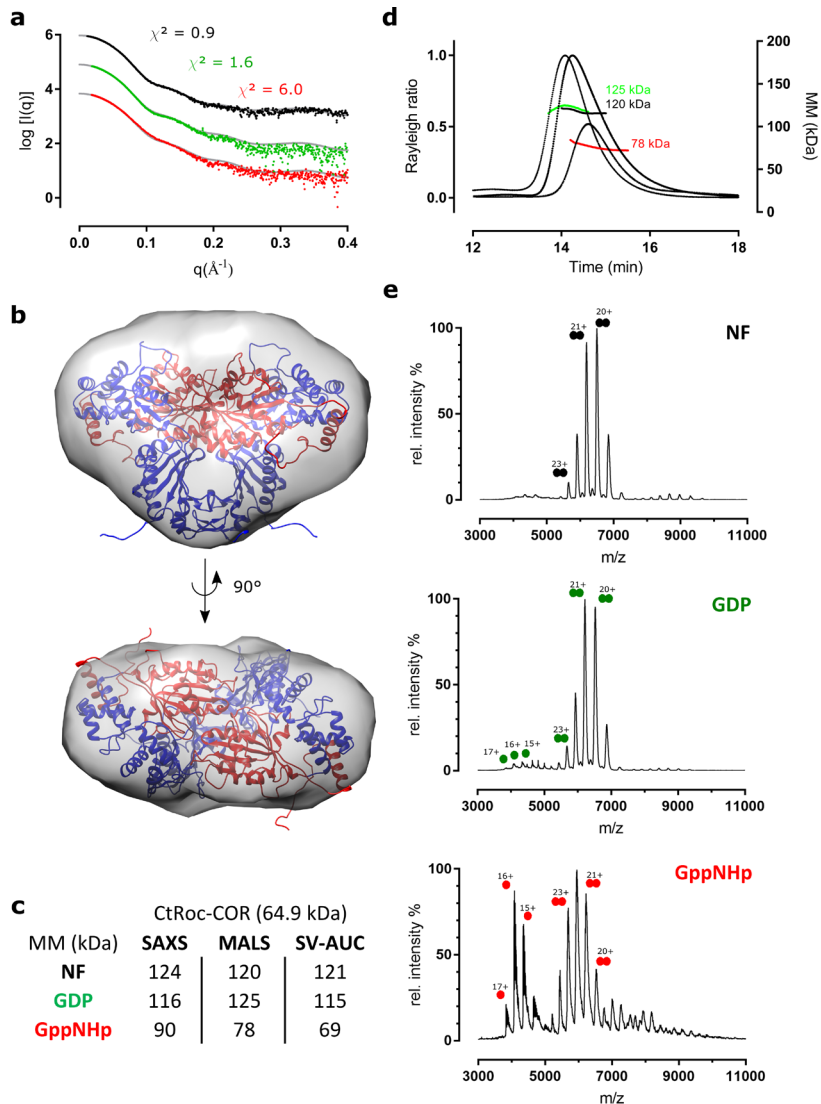


Figure 1: The *Chlorobium tepidum* Roc-COR domain monomerizes upon binding of a non-hydrolysable GTP analogue (GppNHp). (a) Pairwise-comparison using CRYSOLE of the theoretical scattering curve derived from the crystallographic dimer model of CtRoc-COR (PDB 3DPU, grey line) with the experimental scattering curves of CtRoc-COR in the absence of nucleotides (black dots) or in the presence of GDP (green dots) or GppNHp (red dots). (b) Superposition of the crystallographic CtRoc-COR dimer model (Roc in red and COR in blue) on the *ab initio* SAXS envelope, constructed starting from the scattering curve of CtRoc-COR in the absence of nucleotides. (c) Overview of the molecular masses obtained via SAXS (based on Porod volume/1.7) SEC-MALS and SV-AUC for CtRoc-COR in different nucleotide-bound states. The theoretical molecular mass of the monomer is given in between brackets. (NF = nucleotide-free) (d) SEC-MALS data for CtRoc-COR in the absence (black) or presence of nucleotides GDP (green) or GppNHp (red). (e) Native mass spectra of CtRoc-COR for the three different nucleotide states: nucleotide-free (NF), GDP-bound and GppNHp-bound. Peaks corresponding to dimeric and monomeric species are labeled with two circles and one circle, respectively.

GppNHp binding shifts the equilibrium toward the monomeric state in full length CtRoco

Considering the GppNHp-induced monomerization observed in the CtRoc-COR construct, we next collected SAXS data for the full length CtRoco protein, either in the absence of nucleotides or bound to GDP or GppNHp (Supplementary Fig. 4). As there is, thus far, no crystal structure available of the full length CtRoco protein (nor of any other Roco protein), no fitting of the SAXS data on a theoretical curve could be performed. However, visual inspection of the scattering curves clearly shows that conformational changes occur in both the CtRoco GppNHp- and GDP-bound states compared to the nucleotide-free state (Fig. 2a and Supplementary Fig. 4e). Docking of the CtRoc-COR crystal structure into the *ab initio* envelope generated from the SAXS data of the nucleotide-free CtRoco protein show additional features at the N- and C-terminus of CtRoc-COR presumably corresponding to the LRR and C-terminal domains (Fig. 2b).

Molecular mass calculation from the Porod analysis of the SAXS curves, from SEC-MALS and from SV-AUC (Fig. 2c,d and Supplementary Fig 5), yields values between 240 and 290 kDa for the nucleotide-free and GDP-bound forms, close to the theoretical value expected for a dimer (theoretical $MM_{\text{dimer}} = 254.2$ kDa). In the presence of GppNHp, SEC-SAXS (50 μl of an 8 mg/ml protein solution injected on a column) gives a molecular mass of 209 kDa and SEC-MALS (10 μl of an 8 mg/ml protein solution injected on a column) a molecular mass of 173 kDa, while SV-AUC (protein at 0.6 mg/ml, 4.5 μM) gives a broader distribution with a molecular mass at the peak of 158 kDa (sedimentation coefficient of 4.9S) (Supplementary Fig. 5). As for CtRoc-COR this indicates that GppNHp induces monomerization, with the protein existing in a concentration dependent monomer/dimer equilibrium. Finally, native-MS spectra of CtRoco show mainly dimeric species ($MM_{\text{exp}} = 256$ kDa) with a very small amount of monomers ($MM_{\text{exp}} = 128$ kDa) for nucleotide-free and GDP-loaded protein and a 50:50 monomer-dimer ratio for the GppNHp loaded protein (Fig. 2e).

It might seem remarkable at first sight that despite the difference in molecular mass the nucleotide-free, GppNHp- and GDP-bound CtRoco proteins elute at approximately the same volume in SEC (Fig. 2d). Correspondingly, Guinier analysis of the SAXS data shows that the radius of gyration (R_g) even increases upon going from the nucleotide-free state, over the GDP-bound state to the GppNHp-bound state (Supplementary Fig. 4). This could be explained by a conformational change occurring in the CtRoco subunits upon monomerization or upon nucleotide binding. Indeed, as is the case for the CtRoc-COR construct, the SAXS pair-distance distribution function of CtRoco shows a shift toward shorter average inter-atom distances in the presence of GppNHp compared to GDP or nucleotide-free states, indicative of monomerization (Supplementary Fig. 4d).

However, the curve of the GppNHp-bound CtRoco protein also shows a tailing toward longer maximal distances. This strongly suggests the occurrence of a GppNHp-bound monomer that is more elongated than the corresponding subunits in the nucleotide-free dimer. Also in the GDP-bound state tailing of the pair-distance distribution function is observed indicating that already some conformational changes occur in the dimeric nucleotide-bound protein. These conformational changes also explain why the observed monomer/dimer equilibrium would probably be missed when assessed only with size-exclusion chromatography.

Since we observe a monomer/dimer equilibrium in nucleotide-bound CtRoco, we next determined the dissociation constant of this equilibrium under the different nucleotide conditions, using sedimentation equilibrium AUC (SE-AUC) experiments (Supplementary Fig. 6). Global analysis of the data using a single species model indicated predominantly dimeric species for the nucleotide-free (239 kDa) and the GDP-bound CtRoco (225 kDa), whereas a clear shift to the monomeric state for GppNHp-bound CtRoco (177 kDa) is observed. Fitting the data for GDP- and GppNHp-bound CtRoco using a monomer/dimer model results in an approximate K_D of 2 μM for GDP-bound CtRoco and 30 μM for GppNHp-bound CtRoco (exclusively dimers are found for nucleotide-free CtRoco) (Supplementary Fig. 6). Since protein concentrations used for the SEC-SAXS, SEC-MALS, SV-AUC and native MS experiments range between approximately 4.5 and 16 μM , these dissociation constants generally agree with the predominantly dimeric form observed for CtRoco bound to GDP and the shift to the monomeric form upon GppNHp binding.

Finally, to further characterize the observed nucleotide-induced changes in oligomerization we turned to negative stain electron microscopy (EM). Thereto, CtRoco was applied on grids at a concentration of 0.01 mg/ml (0.08 μM) either in a nucleotide-free state or pre-incubated with 1mM GDP or GppNHp (Fig. 2f and Supplementary Fig. 7). In the nucleotide-free state a rather uniform population of “X-shaped” particles is observed, displaying seemingly a two-fold symmetry. We interpret these particles as being the dimeric form of CtRoco. In contrast, in the presence of GppNHp a uniform population of more elongated “worm-like” particles is observed lacking the two-fold symmetry. We interpret these particles to correspond to the monomeric form of the CtRoco, in agreement with the SAXS, SEC-MALS, native MS and AUC experiments. For the GDP-bound state a mixture of particles corresponding to dimers and monomers is observed.

In conclusion, we show that nucleotide binding also induces monomerization of CtRoco in a concentration dependent manner, with the GTP mimic GppNHp having a much stronger effect than GDP.

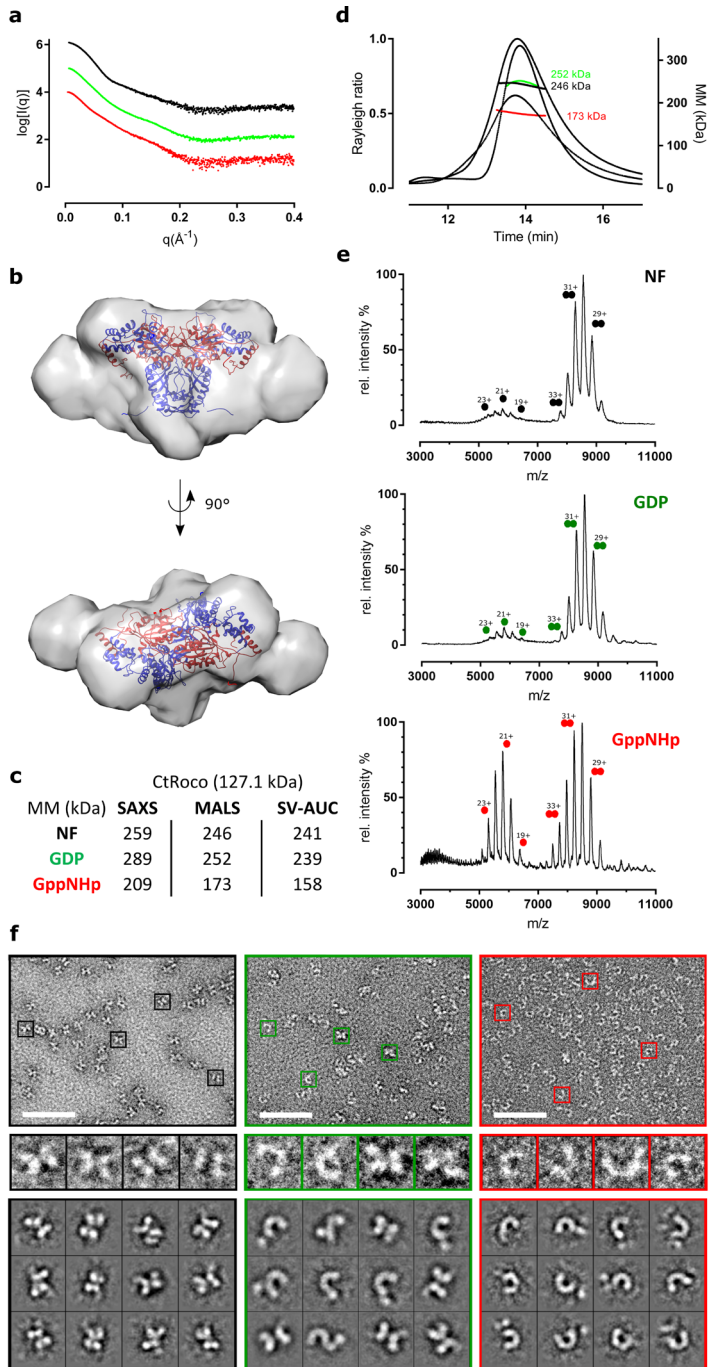


Figure 2: Binding of a non-hydrolysable GTP analogue (GppNHp) to the *Chlorobium tepidum* Roco protein shifts the equilibrium toward the monomeric form. (a) Scattering curves of CtRoco in the absence (black dots) or presence of nucleotides GDP (green dots) or GppNHp (red dots). (b) Superposition of the crystallographic CtRoc-COR dimer model (Roc in red and COR in blue) onto the *ab initio* SAXS envelope, constructed starting from the scattering curve of CtRoco in the absence

of nucleotides. The envelope shows clear additional features corresponding to the N-terminal LRR domain and the C-terminal domain, which are not present in the CtRoc-COR crystal structure. (c) Overview of the molecular masses obtained via SAXS (based on Porod volume/1.7), SEC-MALS and SV-AUC for CtRoco in different nucleotide-bound states. The theoretical molecular mass of the monomer is given in between brackets. (NF = nucleotide-free) (d) SEC-MALS data for CtRoco in the absence (black) or presence of nucleotides GDP (green) or GppNHp (red). (e) Native mass spectra of CtRoco for the three different nucleotide states: nucleotide-free (NF), GDP-bound and GppNHp-bound. Peaks corresponding to dimeric and monomeric species are labeled with two circles and one circle, respectively. (f) Top panels: negative-stain EM images (scale bar: 50 nm) of CtRoco in the nucleotide-free state (black border), bound to GDP (green border) or bound to GppNHp (red border). Middle panels: 4x enlargements of boxed particles in top panels (from left to right). Bottom panels: representative class averages (box size: 18.2 x 18.2 nm) for CtRoco in nucleotide-free state (black border), or bound to GDP (green border) or GppNHp (red border). See also Supplementary Fig. 7 for an overview of all class averages.

Monomerization occurs on a time scale relevant for GTP turnover

In order to be relevant for the CtRoco-catalysed GTP hydrolysis reaction, the observed monomerization should occur within a time frame that is consistent with the time of a full GTP turnover cycle. The k_{cat} value for GTP hydrolysis of CtRoco and CtRoc-COR is 0.1 min^{-1} (Supplementary Fig. 8), meaning that it takes a protein molecule, on average, approximately 10 minutes to travel through an entire GTPase cycle under conditions of substrate saturation. For the CtRoc-COR construct the rate of monomerization could be measured using time-resolved FRET experiments making use of a site-specific single cysteine mutant of this protein construct [19]. A CtRoc-COR variant containing a single cysteine residue at position 928 in the COR domain (S928C) was randomly labelled with a Cy3/Cy5 FRET pair using maleimide-chemistry (Fig. 3a). Statistical incorporation should result in 50% incorporation with the donor/acceptor pair. Subsequently, this nucleotide-free protein was rapidly mixed in a stopped flow apparatus with either buffer or $50 \mu\text{M}$ GDP, GppNHp or GTP, and the decrease in FRET signal linked to monomerization was followed (Fig. 3b). Compared to the buffer control, a very small and slow decrease in FRET signal is observed with GDP. In contrast, a fast decrease in FRET signal is observed upon mixing the protein with an excess of either GppNHp or GTP. The decrease in signal and thus the monomerization is significantly faster upon mixing with GTP compared to the GTP mimic GppNHp. While monomerization reaches a steady state within less than 25 seconds with GTP, equilibrium is only reached after more than 100 seconds with GppNHp. Such differences between physiologically relevant nucleotides and nucleotide analogues were described before [26]. In any case, the observed time for monomerization is significantly faster than the GTP turnover time of about 10 minutes meaning that monomerization happens on a catalytically relevant time scale.

Since cysteine-free full length CtRoco could not be obtained, we turned to time-resolved EM measurements to estimate the time frame of monomerization. Hereto, CtRoco was mixed with an excess of GppNHp or GTP and samples were taken

every two minutes and immediately spotted on grids. In agreement with the FRET experiments nearly full conversion of CtRoco from the “X-shaped” dimers to the “worm-shaped” monomers is observed at the 2 minutes time point (Fig. 3c), while a significant amount of monomerization has even taken place within the dead time (2 minutes) of the experiment. These experiments thus confirm that GTP-induced monomerization of CtRoco is fast and occurs within the time frame of a catalytic cycle of GTP turnover.

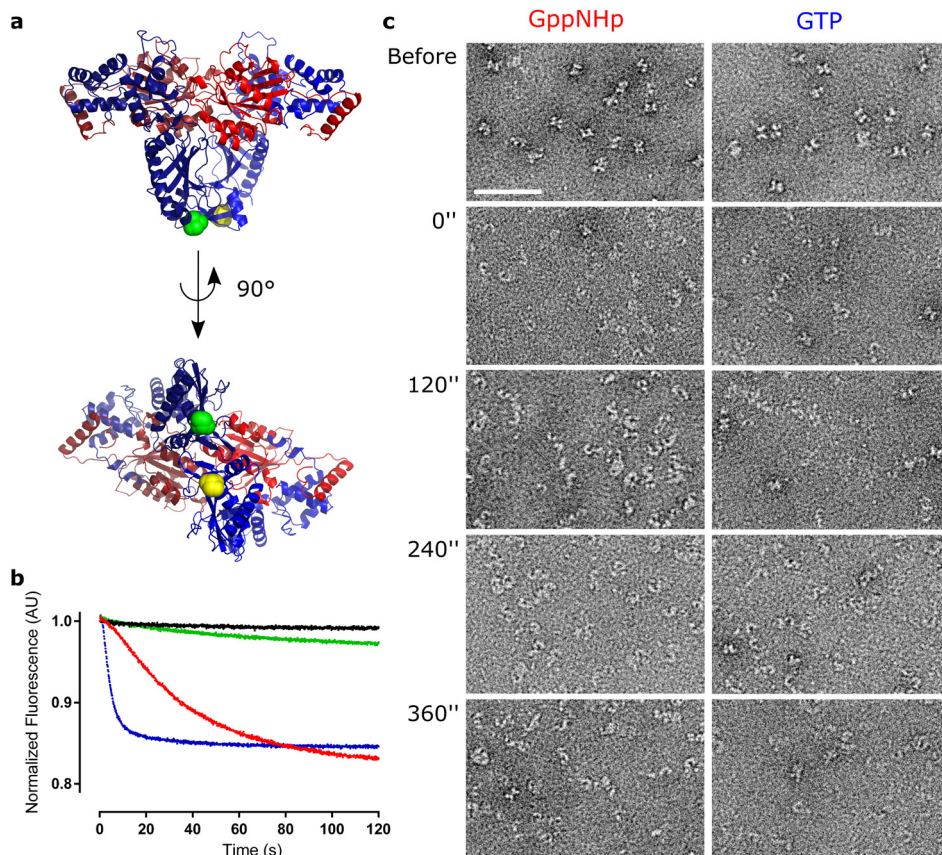


Figure 3: GppNHp- and GTP-induced monomerization of the *Chlorobium tepidum* Roco protein occurs in a time scale relevant for GTP turnover. (a) Two different views of the dimeric CtRoc-COR crystal structure with amino acid 928 colored green and yellow for protomer A and B, respectively. Within the dimeric protein these residues are separated by a distance of 20 Å allowing them to be used for Cy3/Cy5 FRET experiments. The Roc and COR domains are colored red and blue (dark for protomer A and light for protomer B), respectively. (b) Stopped-flow FRET traces of labelled CtRoc-COR upon mixing with different nucleotides to follow monomerization in time. The FRET signal of Cy3/Cy5-labeled CtRoc-COR (S928C) over time is shown after rapid mixing with buffer (black), or 50 μ M GDP (green), GppNHp (red) or GTP (blue). (c) Monomerization of CtRoco followed via time-resolved negative stain EM. CtRoco was mixed with 1 mM of GppNHp or GTP and samples were taken every 2 minutes. Representative EM images of the samples just before adding nucleotide and for each time point are shown. (scale bar: 50 nm)

CtRoco completes a monomer-dimer cycle during GTP turnover

Considering that GTP-induced formation of CtRoco / CtRoc-COR monomers is fast compared to a complete GTP turnover, we subsequently assessed whether the protein undergoes a monomer-dimer cycle coupled to GTP turnover. We therefore resorted to single turnover kinetic measurements (1 μM protein + 1 μM GTP), such that a complete single GTPase cycle can be monitored.

Reverse-phase HPLC measurements show that, under the experimental (non-saturating) conditions, it takes about 6000 seconds for 1 μM CtRoc-COR to completely convert 1 μM GTP to GDP (Fig. 4a). In turn, stopped-flow measurements where 1 μM of fluorescent 2'-(or-3')-O-(N-methylanthraniloyl)-GTP (mant-GTP) is mixed with 1 μM of CtRoc-COR show a fast increase in fluorescence, coupled to mant-GTP binding and the associated conformational changes, that occurs in a time frame of about 250 seconds (Fig. 4a). This fast binding phase is followed by a slow decrease in fluorescence coupled to mant-GTP hydrolysis and return to the initial conformation. In agreement with the data obtained from the reverse-phase HPLC measurements, the total GTPase cycle is finished within about 6000 seconds. Subsequently we followed the monomerization of CtRoc-COR during single GTP-turnover using our time-resolved FRET approach. 1 μM Cy3/Cy5-labelled CtRoc-COR (S928C) was rapidly mixed with 1 μM GTP and the FRET signal was followed over time (Fig. 4b). These traces show a fast decrease in the FRET signal associated with monomerization, followed by a slow increase in signal associated with dimerization. This shows that the GTP-induced monomerization is reversible and re-dimerization occurs upon GTP hydrolysis. The FRET signal reaches a minimum at about 250 seconds. This time point of maximal monomerization corresponds to the time point of maximal GTP binding as determined from the fluorescence stopped-flow experiments with mant-GTP. After this point the FRET signal increases in a time frame that corresponds to GTP hydrolysis and probably also GDP release (note that for CtRocCOR $K_D(\text{GDP}) = 30.2 \mu\text{M}$ and $K_D(\text{GppNHp}) = 0.33 \mu\text{M}$, while a concentration of 1 μM is used here [14]).

Finally, we also assessed the oligomeric state of full length CtRoco during single turnover GTP hydrolysis using time-resolved negative stain EM measurements. Hereto, 1 μM of CtRoco was mixed with 1 μM GTP and samples were taken at different time points (0, 120, 240, 360, 1500, 3000 and 4500 seconds) (Fig. 4c and Supplementary Fig. 9). Under these single turnover conditions a large fraction of CtRoco molecules are converted from a dimeric into a monomeric form after 240 seconds. This is expected for a single turnover experiment where a mixture of unbound, GTP-bound and post-hydrolysis GDP-bound CtRoco molecules co-exist. After that time point the fraction of dimeric protein is again increasing due to GTP hydrolysis, and after 4500 seconds, when all GTP is hydrolysed, a dimeric population reforms.

Thus, together these experiments clearly show that during the GTPase cycle, CtRoco undergoes monomerization concomitant with GTP binding and subsequent dimerization coupled to GTP hydrolysis and GDP release.

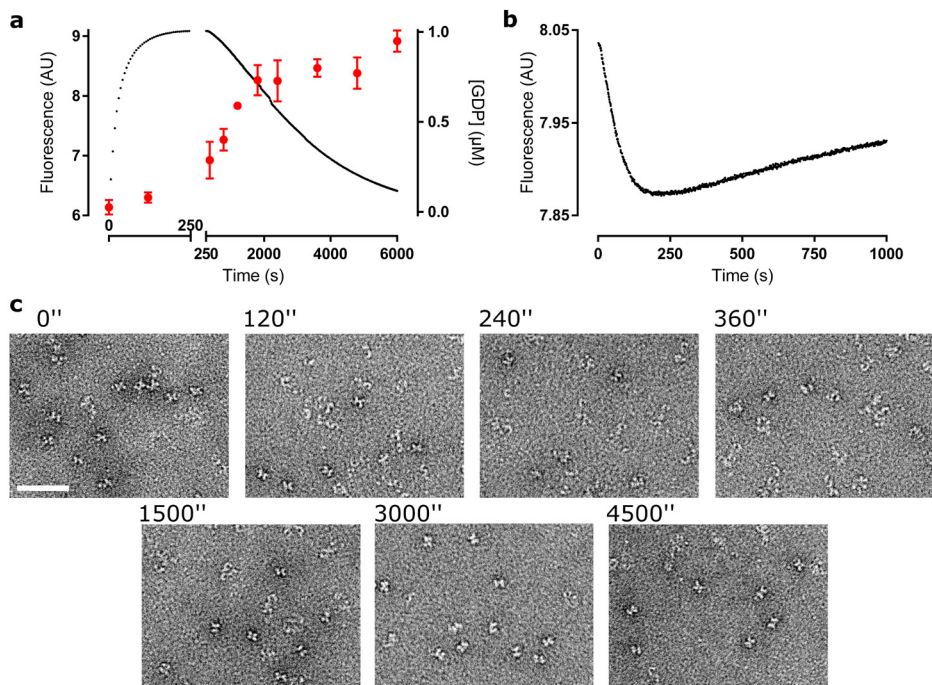


Figure 4: The *Chlorobium tepidum* Roco protein undergoes a monomer-dimer cycle during GTP turnover. (a) Single turnover (mant-)GTP hydrolysis by CtRoc-COR (S928C) followed by stopped flow fluorescence (black curve) and reverse-phase HPLC (red data points; each data point is the average (\pm s.d.) of 3 independent measurements). Rapid mixing of 1 μ M mant-GTP and 1 μ M unlabeled CtRoc-COR (S928C) in a stopped-flow apparatus yields a rapid increase in fluorescence (time frame 1- 250 seconds), followed by a slow decrease in a time frame of 6000 seconds. Following production of GDP from GTP in time via reverse-phase HPLC shows that the increase in fluorescence occurs prior to GDP production while all GTP is converted in the time scale of 6000 seconds. A split time axis is used to highlight the two phases: fast fluorescence increase associated with GTP binding and slow fluorescence decrease concomitant with GTP hydrolysis. (b) Stopped-flow FRET signal obtained by mixing 1 μ M of Cy3/Cy5- labeled CtRoc-COR (S928C) with 1 μ M GTP. The traces show relatively fast monomerization (first phase from 1 – 250 seconds) followed by a slow return to the dimeric state after GTP hydrolysis. (c) Single GTP turnover of CtRoco followed by time resolved EM. 1 μ M CtRoco was mixed with 1 μ M GTP and samples were taken at the indicated time points. Representative images for each time point are shown (see also Supplementary Fig. 9). (scale bar: 50 nm)

An analogue of a PD-associated mutation stabilizes the CtRoco dimer

PD-associated mutations in LRRK2 are mainly located in the kinase and Roc-COR domains. Mapping of the Roc-COR mutations (LRRK2 I1371V/R1441C/Y1699C) onto

2

the *C. tepidum* Roc-COR crystal structure showed that the bacterial analogues of these mutations (L487A/Y558A/Y804C) are all located in the conserved interface between the Roc and COR sub-domains [14]. Since, in our hands, the Y558A and Y804C mutants show significant aggregation when analysed on size exclusion chromatography, we focused on the L487A mutant and analysed its effect on the oligomeric state of the CtRoco and CtRoc-COR proteins using SEC-MALS. Residue L487 corresponds to I1371 in LRRK2 and is located in the Roc domain, in the C-terminus of Switch 1, at the interface of the Roc and COR domains. In their nucleotide-free and GDP-bound states, both the CtRoco and CtRoc-COR proteins that contain the L487A mutation are entirely dimeric, as was also observed in the wild-type counterparts (Fig. 5a,b). However, while in the GppNHp-bound form the monomer-dimer equilibrium of the wild-type CtRoc-COR and CtRoco is clearly shifted toward the monomeric form, the corresponding proteins harboring the L487A mutation remain nearly completely dimeric (Fig. 5a,b). Thus we find that the PD-associated L487A mutation stabilizes the dimeric form of the CtRoco protein. The same behavior is observed when we compare the SAXS curves of wild-type and L487A CtRoco in different nucleotide states (Supplementary Fig. 10,11). While the scattering curves, the normalized Kratky plots [27] and the pair distance distribution function of the wild-type and L487A proteins nearly overlap in the nucleotide-free and GDP states, the curves of both protein variants clearly differ in the GppNHp-bound forms. This again indicates that, while the L487A mutation does not cause any large-scale conformational changes in the nucleotide-free and GDP-bound dimers, it does stabilize the dimeric form of the protein in the GppNHp-bound state.

Subsequently, we measured the GTPase activity of the wild-type and L487A CtRoc-COR proteins under single-turnover conditions and find that the L487A mutation causes a 4-fold decrease in the single-turnover rate constant (Fig. 5c). This is in agreement with previous reports that show that the PD mutations in the Roc-COR domain of LRRK2, as well as the counterparts in CtRoco, decrease the GTPase activity [10,12,28–30]. We thus speculate that monomerization of the Roco protein, which we showed to be an integral part of the GTPase cycle, is hindered by this PD-analogous mutation, thereby causing the pathological decrease in GTPase activity. Hence, our data present a link between the Roco monomer-dimer transition, the rate of GTP turnover and disease.

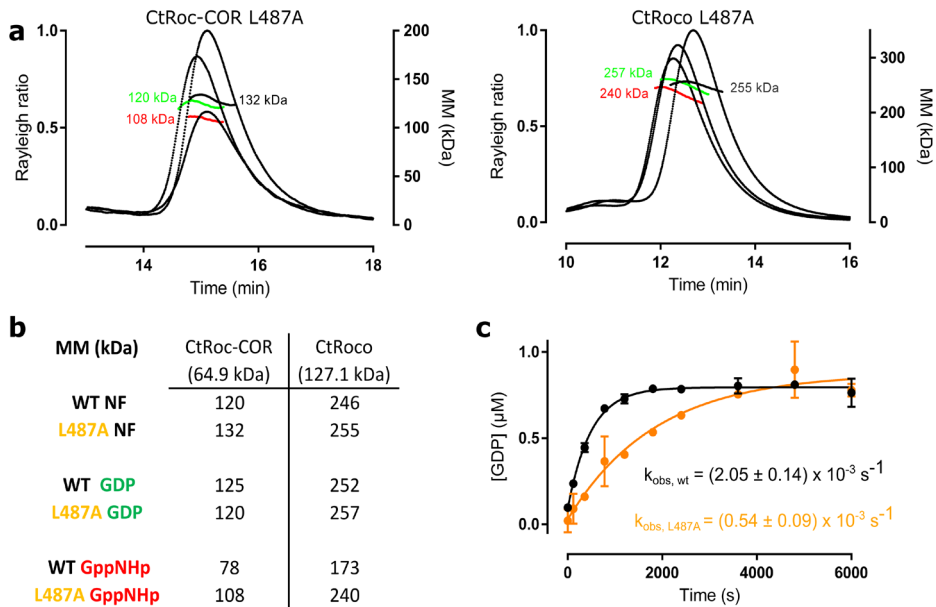


Figure 5: Effect of a PD-analogous mutation on the CtRoco and CtRoc-COR monomer-dimer equilibrium. (a) SEC-MALS data for the CtRoc-COR L487A mutant (left panel) and the CtRoco L487A mutant (right panel) in the absence (black) or presence of nucleotides GDP (green) or GppNHp (red). (b) Overview of molecular masses determined by SEC-MALS for CtRoc-COR wt, CtRoc-COR L487A, CtRoco wt and CtRoco L487A in absence and presence of GDP or GppNHp. The theoretical molecular masses of the monomers are given in between brackets. (NF = nucleotide-free). (c) Single turnover GTP hydrolysis of CtRoc-COR wt (black) and CtRoc-COR L487A (orange) using 1 μM protein and 1 μM GTP. GDP formation is followed using reverse-phase HPLC. The observed rate constants (k_{obs}) ± s.e. for wild-type and mutant CtRoc-COR are indicated. Each data point is the average (± s.d.) of 3 independent measurements.

Nucleotide binding also influences the monomer-dimer equilibrium of hSLRRK2 Roc-COR

Next we analysed if also human LRRK2 has a nucleotide dependent monomer-dimer cycle. For this we turned to fluorescence correlation spectroscopy (FCS). Cell lysates of HEK293 cells transfected with GFP-RocCOR (of hSLRRK2) were supplemented with excess GDP, GppNHp (a non-hydrolysable GTP analogue), or EDTA (resulting in nucleotide-free protein), and the average brightness per particle was determined using FCS (Figure 6). The brightness of GppNHp bound RocCOR was similar to monomeric GFP, while the brightness of nucleotide free and GDP bound RocCOR was significantly higher, indicating a shift to the dimeric state. These results confirm that HsLRRK2, like CtRoco, monomerizes in the presence of GppNHp. More elaborate brightness analysis methods would be required to quantify the kinetics of dimerization.

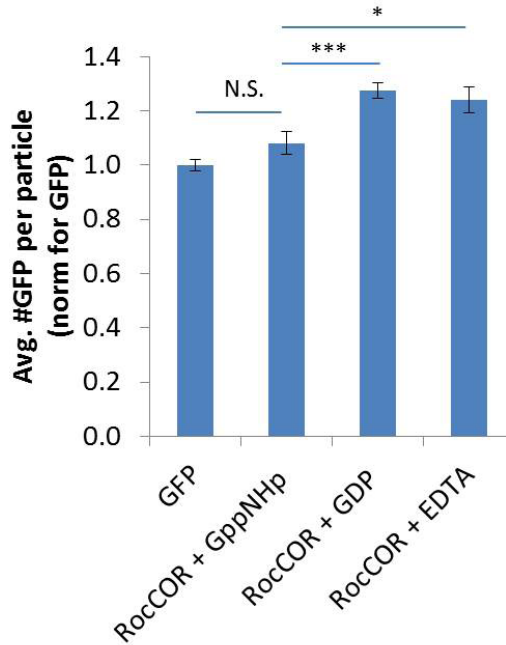


Figure 6: FCS Brightness analysis of HsRoc-COR-GFP oligomerization in HEK cell lysates. HEK293 cells transiently expressing GFP or Roc-COR-GFP from hLRRK2 were lysed in the presence of 10 mM EDTA, 0.5 mM GDP or 0.5 mM GppNHp. The total fluorescence signal (photons/sec) was divided by the average number of particles in the confocal volume and normalized for the brightness of GFP, yielding the average number of GFP molecules per particle/complex. Data are mean \pm SEM, based on two biological replicates. GFP, $n = 29$; RocCOR +EDTA, $n = 11$; RocCOR +GDP, $n = 12$, RocCOR +GppNHp, $n = 14$. *, $p < 0.05$; ***, $p < 0.001$; N.S., non-significant.

Discussion

In the current study we show that the oligomerization of the *Chlorobium tepidum* Roco protein depends on its nucleotide state. In contrast to previous suggestions, we find that the protein is mainly monomeric when bound to GTP and dimeric in the nucleotide-free state, while an intermediate situation is observed in the GDP state. Moreover, we show that the CtRoco protein cycles through the dimeric and monomeric conformations during a round of GTP hydrolysis, meaning that these changes in oligomerization are an integral feature of the GTPase catalytic cycle. Consequently, although many details still need to be resolved, a mechanism as outlined in Figure 7 can be proposed based on our results. The reaction cycle starts with the rapid binding of a GTP molecule to the Roc domains, followed by monomerization of the protein as observed in SEC-MALS, SEC-SAXS, native MS, AUC and EM (Fig. 1,2&3). Inter-domain and inter-subunit conformational changes leading to monomerization are likely triggered via changes

in conformation of the switch regions. The crystal structure of CtRoc-COR shows that especially switch II is ideally positioned to transfer GTP-induced conformational changes from the Roc domain to the COR domain within the same protomer as well as to the adjacent COR domain [14]. The monomerization step is followed by slow GTP hydrolysis concomitant with protein dimerization (Fig. 4). Considering that single turnover GDP production, as monitored via reverse-phase HPLC, occurs on the same time scale as dimerization (Fig. 4), we assume GTP hydrolysis to be the rate-limiting step in this process. Finally, GDP is released, which may occur either after or concomitant with protein dimerization since we find CtRoco to be in equilibrium between a monomeric and dimeric state in the presence of GDP.

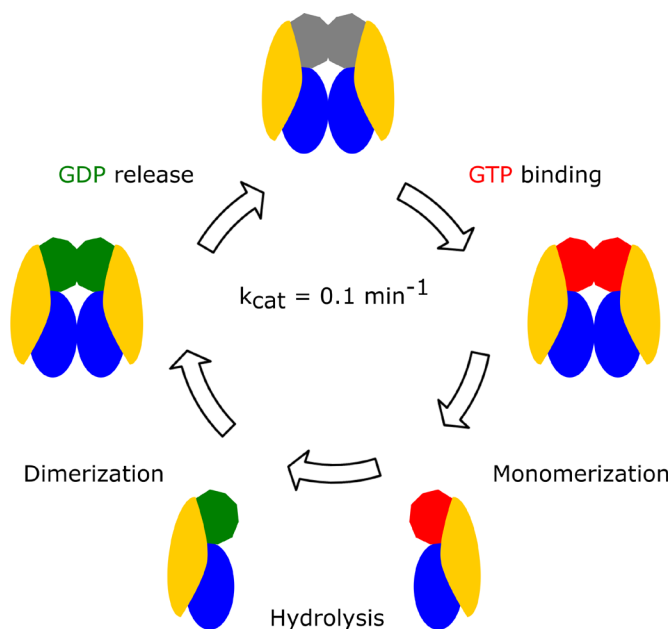


Figure 7: Proposed monomer-dimer transitions during the GTP hydrolysis cycle of CtRoco. In its nucleotide-free state the protein is dimeric. GTP binding induces conformational changes that lead to fast monomerization. In the monomeric state the protein hydrolyses GTP to GDP. Depending on the protein concentration, protein dimerization could take place either after or before release of GDP. After GDP is released, the cycle can restart. In the cartoon of CtRoco, the LRR domain is colored yellow, the COR domain blue, the Roc domain dependent on the nucleotide state grey (nucleotide-free), green (GDP-bound) or red (GTP-bound).

Our observation that GTP hydrolysis occurs after monomerization contrasts with the current model that assumes that Roco proteins, including LRRK2, function as GADs [17,18,31]. This latter model is based largely on the crystal structure of the nucleotide-free Roc-COR module of the *C. tepidum* Roco protein, being the only available structure

2

of a full Roc-COR tandem domain [14]. This structure shows a Roc-COR dimer where the dimer contacts are formed by the COR domain and with the GTP binding sites of the Roc domains juxtaposed. This observation has led to the hypothesis that GTP binding induces dimerization of the Roc domains, with the reciprocal complementation of the active sites finally leading to catalysis of GTP hydrolysis. We show here, using the exact same protein, that GTP binding induces monomerization and that GTP hydrolysis occurs after the monomerization step. So far, it remains unclear whether Roco monomerization is a strict requirement for GTP hydrolysis to occur or whether GTP-induced monomerization and GTP hydrolysis are two uncoupled phenomena. However, both EM and SAXS indicate the occurrence of secondary inter-domain conformational changes in the CtRoco monomers compared to the dimers, with the protomers in the monomeric arrangement being more elongated compared to the dimers (Fig. 2 and Supplementary Fig. 4d). High resolution structural information of the monomeric states could shed further light on the mechanism of GTP hydrolysis. A second question that is triggered by our findings concerns the role of dimerization within the reaction cycle. A potential scenario would be that dimerization of the Roco subunits leads to a decrease in the affinity for GDP, thus enhancing the rate of GDP release.

Since the presence of the Roc-COR module is highly conserved throughout all Roco proteins, our current findings also throw new light on LRRK2 functioning. However, apart from the Roc-COR and LRR domain, LRRK2 also contains a kinase domain and a number of domains typically involved in protein-protein interactions, which potentially can have an additional influence on oligomerization [21,32]. Detailed and quantitative biophysical studies with LRRK2 to determine its oligomeric state remain highly challenging since purification yields are low, and purified LRRK2 is very heterogeneous and prone to degradation. However, it has been shown that LRRK2 can form dimers via its Roc-COR domain [15,20–22], while other studies indicate that LRRK2 can also exist as a monomer [18,20,23]. Both observations can be reconciled with our model. Consistently, our FCS experiments on the cytosolic fraction of cell lysates expressing human LRRK2 GFP-RocCOR confirm that oligomerization of the Roc-COR domain of LRRK2 is taking place in the GDP and nucleotide free states, while the GTP-bound state is mainly monomeric.

Interestingly, recent data suggest that LRRK2 can exist as either a monomer or a dimer in the cell, where the monomeric form prevails in the cytosol and the dimeric form is more prominently present at the membrane [21,24,25,33]. A higher prevalence of a monomeric GTP-bound form under multiple GTP turnover conditions such as present in the cytosol (GTP concentrations in the cell are in a range of 1-1.6 mM [34,35]) corresponds to our *in vitro* observation in EM (Fig. 3). Following the same argument, LRRK2 would exist in a more GDP-bound form at the membrane. Apart from these considerations it is likely that additional regulatory factors, such as phosphorylation or interaction with other proteins,

influence the monomer-dimer equilibrium of LRRK2 *in vivo* [36].

A number of point mutations in the Roc (R1441C/G/H) and COR (Y1699C) domains of LRRK2 are directly linked to PD [1,37,38]. These mutations have been related to a reduced GTPase activity [10,12,28–30], but so far no widely-accepted consensus on the effect of PD mutations on homodimerization exists [22,29,39] [33]. According to our model, whereby the CtRoco protein needs to cycle through a monomeric and dimeric state, stabilization of either could lead to a reduced GTPase activity. We show that the L487A mutation in the Roc-COR interface (analogous to I1371V in LRRK2) decreases the GTPase activity by stabilizing the dimer interface (Fig. 5). Our results thus provide a direct link between oligomerization, GTPase activity and disease phenotype.

Because the kinase activity of LRRK2 is directly responsible for neuronal toxicity [9,40], most efforts for therapeutic intervention are focussed on the development of kinase-inhibitors [41,42]. However, long-term inhibition of LRRK2 by many of these inhibitors increases the risk for morphological changes in lungs, similar to what has been observed in LRRK2 knock-out models [43–45]. Our findings now show that either increased dimerization or increased monomerization could lead to changes in GTPase activity and might form an appealing target for future drug development.

Methods

Protein expression and purification

The DNA fragments coding for CtRoco (a.a. 1-1102) and CtRoc-COR (a.a. 412-946) were cloned in the pProEX plasmid with N-terminal His-tags [14]. The plasmids were transformed in *E. coli* BL21 (DE3) cells. Cells were grown at 37 °C in Terrific Broth medium with 100 µgml⁻¹ ampicillin. Once an optical density at 600 nm (OD600) of 0.7 was reached, 0.1mM β-D-1-thiogalactopyranoside (IPTG) was added and the temperature was dropped to 28 °C. After overnight incubation, the cells were harvested by centrifugation and resuspended in buffer A (20mM Hepes pH7.5, 150mM NaCl, 5mM MgCl₂, 5% glycerol, 1mM DTT) with 50 µgml⁻¹ DNase, 0.1 mg ml⁻¹ AEBSF, 1 µgml⁻¹ leupeptin, and 200 µM GDP. Afterwards the cells were disrupted using a cell disruptor (Constant systems) and the soluble fraction was separated from the cell debris via centrifugation. As a first purification step, an immobilized metal affinity chromatography (IMAC) step was performed. The supernatant containing the soluble protein was loaded on Ni²⁺-NTA sepharose resin. The resin was extensively washed with buffer A and subsequently the protein was eluted using buffer A including 300mM imidazole. The obtained protein was subsequently dialysed against 20mM Hepes pH 7.5, 150mM NaCl, 5% glycerol, 1mM DTT (buffer B). Before the final size exclusion chromatography purification step, 1mM

EDTA was added to the sample and the mixture was incubated for 1 h at 4 °C. Finally the sample was loaded on a Superdex S200 26/60 column (GE Healthcare) equilibrated with buffer B. Fractions containing the protein were supplemented with 5 mM MgCl₂ and the nucleotide-load was determined using reversed-phase chromatography coupled to HPLC (see further).

Size Exclusion Chromatography Small-angle X-ray scattering (SEC-SAXS) and Multi-angle light scattering (SEC-MALS)

SAXS experiments were performed on the BM29 beamline at the ESRF (Grenoble, France) and the SWING beamline at SOLEIL (Paris, France), always with an inline HPLC set-up. 50 µl of an 8 mg/ml protein sample was injected on a Bio SEC-3 HPLC column (Agilent, 3 µm 300 Å) equilibrated with 20 mM Hepes pH 7.5, 150 mM NaCl, 5 mM MgCl₂, 5% glycerol and 1 mM DTT (assay buffer). For data collection in the presence of nucleotides, the nucleotide-free proteins were pre-incubated with 1mM of the nucleotide, and 200µM (= molar excess) of the nucleotide was added to the assay buffer. Initial processing of the data was done with DATASW [46]. PRIMUS was used for determination of the radius of gyration (Rg) using the Guinier approximation [47], and GNOM for the calculation of the pair distance distribution function P(r) [48]. Modelling of missing loops in the crystallographic dimer model of the CtRoc-COR was done using ModLoop [49] and the missing N- and C-terminus were added with CORAL [50]. The final dimer model was compared with the experimental data using CRY SOL [51]. *Ab initio* envelopes were calculated using DAMMIN (average of 19 runs for CtRocCOR and 20 runs for CtRoco) [52] followed by DAMAVER [53]. The final model was generated using one round of DAMMIN [52], starting from the damstart model generated by DAMAVER. Docking into the envelope was performed with Supcomb [54]. For docking into the CtRoco envelope, the position of the CtRoc-COR dimer structure was manually adjusted. The molmap command in chimera was used to convert the *ab initio* bead models into 20 Å density maps [55]. In overview of the experimental and modelling parameters is provided in Supplementary Tables 1, 2 and 3.

For the SEC-MALS experiments, the samples were prepared and run in the same way as for the SEC-SAXS experiments. 10 µl of an 8 mg/ml protein sample was injected on a Bio SEC-3 HPLC column (Agilent, 3 µm 300 Å). A Dawn Heleos detector (using 9 angles) and Optilab T-rEX detector (Wyatt technology) were attached to a HPLC (Shimadzu). The molar masses were calculated with the ASTRA 5.3.4.20 software.

Native mass spectrometry

Protein samples were exchanged to a buffer containing 150 mM Ammonium Acetate pH 7.5 using micro Bio-spin columns (Bio-gel P6, Bio-rad). For measurements in the presence of nucleotide (GDP or GppNHp), samples were pre-incubated with 500 µM

nucleotide and 100 μM nucleotide was added to the Ammonium Acetate buffer. Final protein concentrations ranged between 12.5 μM and 16 μM . Samples were introduced into the vacuum of the mass spectrometer using nanoelectrospray ionization with in-house-prepared, gold-coated borosilicate glass capillaries with a spray voltage of +1.4 kV. Spectra were recorded on a quadrupole TOF instrument (Q-TOF2, Waters) modified for transmission of native, high-m/z protein assemblies, as described elsewhere [56]. Critical voltages and pressures throughout the instrument were 120 V and 25 V for the sampling cone and collision voltage respectively, with pressures of 10 and $2\text{E}-2$ mbar for the source and collision cell.

Analytical ultracentrifugation

Sedimentation velocity (SV-AUC) experiments on CtRoco and CtRoc-COR were carried out at 20°C in 20 mM HEPES pH 7.5, 150 mM NaCl, 5 mM MgCl_2 , 5% Glycerol and 100 μM of the respective nucleotides on a Beckman Coulter ProteomLab™ XL-I analytical ultracentrifuge using the absorbance at 280 nm. Samples were prepared at a concentration of 0.57 mg/ml (4.5 μM) for CtRoco and 0.3 mg/ml (4.6 μM) for CtRoc-COR. Standard double sector centrepieces were used. The cells were scanned every minute and in total 200 scans were collected. The data was analysed using SEDFIT 15.01b [57] with the continuous $c(s)$ distribution model. Solution density ρ , viscosity η and partial specific volumes \bar{v} were calculated using SEDNTERP [58] ($\rho = 1.02061$ g/l; $\eta = 0.01197$ kg/(s*m), \bar{v} (CtRoco) = 0.7446, \bar{v} (CtRoc-COR) = 0.7398). The $c(s)$ analysis was carried out with an s range of 0 to 15 with a resolution of 200 and a confidence level of 0.68. In all cases, fits were good, with root mean square deviation (rmsd) values ranging from 0.005 to 0.012. Results were prepared for publication using GUSI 1.2.1 [59].

Sedimentation equilibrium experiments (SE-AUC) on CtRoco were conducted on the same instrument and under the same buffer conditions as SV-AUC experiments using an Epon six-channel centrepiece and measurement of the absorbance at 280 nm. Each sample was used at three different concentrations (0.9 mg/ml, 7.2 μM / 0.58 mg/ml, 4.5 μM and 0.35 mg/ml, 2.7 μM). Samples were centrifuged at 1164, 3407 and 9757 x g until sedimentation equilibrium was reached. Three scans were taken at 280nm at each point. The data was analysed with a SEDPHAT [60] software using either monomer-dimer or single-species model. Figures were generated using GUSI 1.2.1. [59].

Förster Resonance Energy Transfer (FRET) and fluorescence stopped flow analysis

The S928C mutant was generated in a cysteine-free variant of CtRoc-COR using quick-change mutagenesis, as described earlier [19]. 10 mg of completely reduced pure CtRoc-COR S928C protein was loaded on a S200 16/60 column equilibrated with a degassed buffer composed of 20 mM Hepes pH 7, 150 mM NaCl, 5 mM MgCl_2

and 5% glycerol. Peak fractions were used to perform the labelling reaction with 8 μM protein, 36 μM maleimide-Cy3 (Lumiprobe) and 36 μM maleimide-Cy5 (Lumiprobe) in a total volume of 5 ml. After incubation at room temperature for 2 hours, the unreacted fluorophores were separated from the labelled protein using size-exclusion chromatography on a S200 16/60 column equilibrated with 20 mM Hepes pH 7.5, 150 mM NaCl, 5 mM MgCl_2 , 5% glycerol and 1 mM DTT. Labelling stoichiometry was spectrophotometrically determined to be approximately 30% for each fluorophore. To determine the monomerization rate, 0.2 μM double labelled protein was mixed with 100 μM nucleotide in a stopped flow apparatus (Applied Photophysics). To follow monomerization during single GTP-turnover, 2 μM Cy3/Cy5-labelled CtRoc-COR was rapidly mixed with 2 μM GTP. The Cy3 fluorophore was excited at 540 nm and change in Cy5 emission was monitored using a cut-off filter of 645 nm.

mant-GTP binding and hydrolysis was followed in a stopped-flow apparatus (Applied Photophysics) by rapidly mixing 2 μM of protein with 2 μM of mant-GTP. The mant-fluorophore was excited at 360 nm and emission was followed through a cut-off filter of 405 nm. All experiments were performed at 25°C and at least 3 time traces were averaged.

Negative stain Electron Microscopy

For negative stain electron microscopy, 2 μL of a 0.01 mg/ml protein sample was applied on a glow discharged carbon-coated copper grid. After three short wash and blot steps with MilliQ water, the grids were stained with a 1% uranyl formate solution. Grids were visualized with a JEOL JEM-1400 electron microscope operating at 120 kV and equipped with a LaB_6 cathode. Images were recorded on a CMOS TemCam-F416 camera (TVIPS, Germany) at a nominal magnification of 80 000, a defocus of approximately 2 μm and a corresponding pixel size of 1.42 Å.

For imaging the different nucleotide-bound states, 0.01 mg/ml CtRoco protein solutions were pre-incubated with 1 mM GDP or GppNHp. In the multiple turnover experiment, 1 mM GTP/GppNHp was added to 0.01 mg/ml CtRoco protein. For the single turnover experiment, where turnover of 1 μM GTP by 1 μM CtRoco was followed, samples were diluted to 0.01 mg/ml right before spotting. In both experiments samples were taken at the following time points: 0, 2, 4, 6, 25, 50 and 75 minutes. The time-resolved EM experiments were performed at 25°C. For the calculation of class averages, 11,571 particles for the nucleotide-free state, 9620 particles for the GDP-bound state and 11,164 for the GppNHp-bound state were selected using e2boxer [61]. Further classification was done with SPARX [62].

GTP hydrolysis assays

GTP hydrolysis experiments were performed at 25°C in 20 mM Hepes pH 7.5, 150 mM NaCl, 5 mM MgCl₂, 5% glycerol and 1mM DTT. The GTP-GDP mixture was separated using a C18-reversed phase column (Phenomenex, Jupiter 5µm C18 300 Å) coupled to a HPLC system (Waters), using 100 mM KH₂PO₄ pH 6.4, 10 mM Tetra butyl-Ammonium-Bromide and 7.5% acetonitrile as a mobile phase. Nucleotide elution was followed using absorbance at 254 nm. The area of GDP was converted to concentration using a standard curve.

Time points for the single turnover experiments with 1 µM protein and 1 µM GTP were 0, 2, 6, 13, 20, 30, 40, 60, 80 and 100 minutes. For the multiple turnover steady state experiments 0.2 µM protein was incubated with different concentrations of GTP (25, 50, 100, 175, 250µM) and samples were taken after 0, 30, 60, 90 and 120 minutes. Kinetic parameters were determined by fitting the data to a single exponential (single turnover) or the Michaelis-Menten equation (multiple turnover) using GraphPad Prism 6. All experiments were done in triplicate.

FCS analysis

HEK293 cells were maintained at 37 °C in tissue culture treated Greiner flasks in Dulbecco's Modified Eagle medium (DMEM) supplemented with high glucose, 4mM L-Glutamine, 10% fetal bovine serum, 1% antibiotics (Penicillin and Streptomycin), and 5% CO₂. Approximately 0.5 x 10⁶ cells were seeded per well in 6-well plates for 24 hours and transfected for 48 hours with 2.5 µg plasmid per well (pEGFP LRRK2_RocCOR or pCDNA 3.0 GFP, the latter diluted 50x with empty pCDNA 3.0) using Lipofectamine LTX™ reagent (Invitrogen, Waltham, Massachusetts, USA) and incubated for 48 hours. The cells were washed in the well in 50 mM Tris buffer (pH 7.5, 100 mM NaCl, 5 mM MgCl₂ and 5% Glycerol), resuspended, pelleted and lysed for 30 min on ice with lysis buffer (50 mM Tris, pH 7.5, 100 mM NaCl, 5 mM MgCl₂, 5% Glycerol, 1% Triton X-100, protease inhibitor P2714 Prln (Sigma-Aldrich, St. Louis, Missouri, United States)), supplemented with 10 mM EDTA, 0.5 mM GDP or 0.5 mM GppNHp. Cell debris was cleared from the lysate by centrifugation (10 min, 20,800× *g*, 4 °C). Lysates were further diluted in lysis buffer supplemented with 0.5 mM GDP/GppNHp to a concentration of 3–5 molecules in the confocal spot (*N*) (accordingly to the estimates reported by the ZEN acquisition software (Zeiss, Jena, Germany)). FCS measurements of the lysates were performed on a LSM710 ConfoCor 3 microscope (Zeiss, Jena, Germany) and analyzed globally using the FFS data processor 2.6 software (SSTC, Department of Systems Analysis and Computer Modelling, Belarussian State University, Minsk, Belarus, www.sstcenter.com), as described in more detail in Chapter 4 of this thesis (FCS global analysis without PCH). The total measured fluorescence signal (photons/sec) was divided by the average number of particles in the confocal volume (calculated by FCS) and normalized for the

brightness of GFP, yielding the average number of GFP molecules per particle/complex.

Data availability

All data supporting the findings of this study are available from the corresponding author upon reasonable request. All SAXS data and derived models were deposited in the SASDB (CtRocCOR NF: SASDCB2, CtRocCOR GDP: SASDCC2, CtRocCOR GppNHp: SASDCD2, CtRoco NF: SASDC82, CtRoco GDP: SASDCA2, CtRoco GppNHp: SASDC92, CtRoco L487A NF: SASDCG2, CtRoco L487A GDP: SASDCE2, CtRoco L487A GppNHp: SASDCF2)

Acknowledgements

We would like to thank the staff at the beamlines SWING of Soleil (France) and BM29 of ESRF (France) for assistance during data collection. This work was supported by the Fonds voor Wetenschappelijk Onderzoek (ML, RE, WV), BioStruct-X by the European Community's Seventh Framework Programme (WV), a Strategic Research Program Financing of the VUB (WV), VUB/RUG collaboration agreement (OZR2544; LW), the Hercules foundation (WV), the Michael J. Fox Foundation for Parkinson's Research (PV, ArK, GG, CJG, WV) and a NWO-VIDI grant (ArK). We thank Annelore Stroobants for assistance with electron microscopy, Pragya Pathak for help with data analysis and Roise Mc Govern for carefully reading the manuscript.

Competing interests

The authors declare no competing financial interests.

Supplementary information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-017-01103-4>



Supplementary Information

References

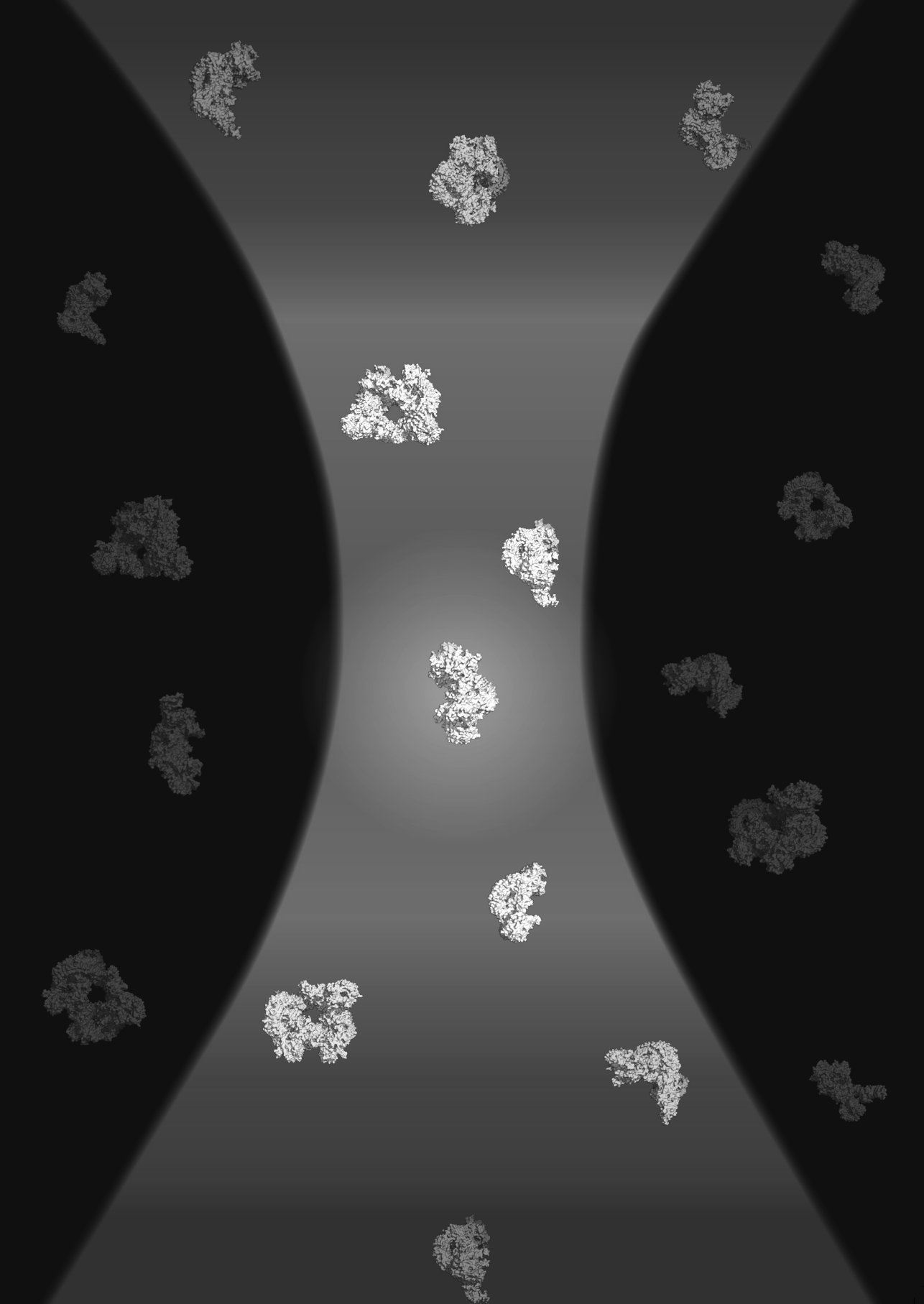
1. Zimprich, A.; Biskup, S.; Leitner, P.; Lichtner, P.; Farrer, M.; Lincoln, S.; Kachergus, J.; Hulihan, M.; Uitti, R.J.; Calne, D.B.; et al. Mutations in LRRK2 Cause Autosomal-Dominant Parkinsonism with Pleomorphic Pathology. *Neuron* **2004**, *44*, 601–607, doi:10.1016/j.neuron.2004.11.005.
2. Paisán-Ruiz, C.; Jain, S.; Evans, E.W.; Gilks, W.P.; Simón, J.; van der Brug, M.; López de Munain, A.; Aparicio, S.; Gil, A.M.; Khan, N.; et al. Cloning of the Gene Containing Mutations That Cause PARK8-Linked Parkinson's Disease. *Neuron* **2004**, *44*, 595–600, doi:10.1016/j.neuron.2004.10.023.
3. Gilks, W.P.; Abou-Sleiman, P.M.; Gandhi, S.; Jain, S.; Singleton, A.; Lees, A.J.; Shaw, K.; Bhatia, K.P.; Bonifati, V.; Quinn, N.P.; et al. A Common LRRK2 Mutation in Idiopathic Parkinson's Disease. *Lancet* **2004**, *365*, 415–416, doi:10.1016/S0140-6736(05)17830-1.
4. Bosgraaf, L.; Van Haastert, P.J.M. Roc, a Ras/GTPase Domain in Complex Proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **2003**, *1643*, 5–10, doi:10.1016/j.bbamcr.2003.08.008.
5. Marín, I.; van Egmond, W.N.; van Haastert, P.J.M. The Roco Protein Family: A Functional Perspective. *FASEB J.* **2008**, *22*, 3103–3110, doi:10.1096/fj.08-111310.
6. Cookson, M.R. The Role of Leucine-Rich Repeat Kinase 2 (LRRK2) in Parkinson's Disease. *Nat. Rev. Neurosci.* **2010**, *11*, 791–797, doi:10.1038/nrn2935.
7. Gilsbach, B.K.; Kortholt, A. Structural Biology of the LRRK2 GTPase and Kinase Domains: Implications for Regulation. *Front. Mol. Neurosci.* **2014**, *7*, 32, doi:10.3389/fnmol.2014.00032.
8. West, A.B.; Moore, D.J.; Biskup, S.; Bugayenko, A.; Smith, W.W.; Ross, C.A.; Dawson, V.L.; Dawson, T.M. Parkinson's Disease-Associated Mutations in Leucine-Rich Repeat Kinase 2 Augment Kinase Activity. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 16842–16847, doi:10.1073/pnas.0507360102.
9. Greggio, E.; Jain, S.; Kingsbury, A.; Bandopadhyay, R.; Lewis, P.; Kaganovich, A.; van der Brug, M.P.; Beilina, A.; Blackinton, J.; Thomas, K.J.; et al. Kinase Activity Is Required for the Toxic Effects of Mutant LRRK2/Dardarin. *Neurobiol. Dis.* **2006**, *23*, 329–341, doi:10.1016/j.nbd.2006.04.001.
10. Lewis, P.A.; Greggio, E.; Beilina, A.; Jain, S.; Baker, A.; Cookson, M.R. The R1441C Mutation of LRRK2 Disrupts GTP Hydrolysis. *Biochem. Biophys. Res. Commun.* **2007**, *357*, 668–671, doi:10.1016/j.bbrc.2007.04.006.
11. Steger, M.; Tonelli, F.; Ito, G.; Davies, P.; Trost, M.; Vetter, M.; Wachter, S.; Lorentzen, E.; Duddy, G.; Wilson, S.; et al. Phosphoproteomics Reveals That Parkinson's Disease Kinase LRRK2 Regulates a Subset of Rab GTPases. *Elife* **2016**, *5*, 1–28, doi:10.7554/eLife.12813.001.
12. Guo, L.; Gandhi, P.N.; Wang, W.; Petersen, R.B.; Wilson-Delfosse, A.L.; Chen, S.G. The Parkinson's Disease-Associated Protein, Leucine-Rich Repeat Kinase 2 (LRRK2), Is an Authentic GTPase That Stimulates Kinase Activity. *Exp. Cell Res.* **2007**, *313*, 3658–3670, doi:10.1016/j.yexcr.2007.07.007.
13. Gloeckner, C.J.; Kinkl, N.; Schumacher, A.; Braun, R.J.; O'Neill, E.; Meitinger, T.; Kolch, W.; Prokisch, H.; Ueffing, M. The Parkinson Disease Causing LRRK2 Mutation I2020T Is Associated with Increased Kinase Activity. *Hum. Mol. Genet.* **2006**, *15*, 223–232, doi:10.1093/hmg/ddi439.
14. Gotthardt, K.; Weyand, M.; Kortholt, A.; Van Haastert, P.J.M.; Wittinghofer, A. Structure of the Roc-COR Domain Tandem of C. Tepidum, a Prokaryotic Homologue of the Human LRRK2 Parkinson Kinase. *EMBO J.* **2008**, *27*, 2239–2249, doi:10.1038/emboj.2008.150.
15. Terheyden, S.; Ho, F.Y.; Gilsbach, B.K.; Wittinghofer, A.; Kortholt, A. Revisiting the Roco G Protein Cycle. *Biochem. J.* **2014**, doi:10.1042/BJ20141095.
16. Gilsbach, B.K.; Ho, F.Y.; Vetter, I.R.; Van Haastert, P.J.M.; Wittinghofer, A.; Kortholt, A. Roco Kinase Structures Give Insights into the Mechanism of Parkinson Disease-Related Leucine-Rich-Repeat Kinase 2 Mutations. *Proc. Natl. Acad. Sci.* **2012**, *109*, 10322–10327, doi:10.1073/pnas.1203223109.
17. Gasper, R.; Meyer, S.; Gotthardt, K.; Sirajuddin, M.; Wittinghofer, A. It Takes Two to Tango: Regulation of G Proteins by Dimerization. *Nat. Rev. Mol. Cell Biol.* **2009**, *10*, 423–429.

18. Nixon-Abell, J.; Berwick, D.C.; Harvey, K. L'RRK de Triomphe: A Solution for LRRK2 GTPase Activity? *Biochem. Soc. Trans.* **2016**, *44*.
19. Rudi, K.; Ho, F.Y.; Gilsbach, B.K.; Pots, H.; Wittinghofer, A.; Kortholt, A.; Klare, J.P. Conformational Heterogeneity of the Roc Domains in *C. Tepidum* Roc-COR and Implications for Human LRRK2 Parkinson Mutations. *Biosci. Rep.* **2015**, *35*, e00254–e00254, doi:10.1042/BSR20150128.
20. Nixon-Abell, J.; Berwick, D.C.; Grannó, S.; Spain, V.A.; Blackstone, C.; Harvey, K. Protective LRRK2 R1398H Variant Enhances GTPase and Wnt Signaling Activity. *Front. Mol. Neurosci.* **2016**, *9*, doi:10.3389/fnmol.2016.00018.
21. Greggio, E.; Zambrano, I.; Kaganovich, A.; Beilina, A.; Taymans, J.-M.; Daniëls, V.; Lewis, P.; Jain, S.; Ding, J.; Syed, A.; et al. The Parkinson Disease-Associated Leucine-Rich Repeat Kinase 2 (LRRK2) Is a Dimer That Undergoes Intramolecular Autophosphorylation. *J. Biol. Chem.* **2008**, *283*, 16906–16914, doi:10.1074/jbc.M708718200.
22. Klein, C.L.; Rovelli, G.; Springer, W.; Schall, C.; Gasser, T.; Kahle, P.J. Homo- and Heterodimerization of ROCO Kinases: LRRK2 Kinase Inhibition by the LRRK2 ROCO Fragment. *J. Neurochem.* **2009**, *111*, 703–715, doi:10.1111/j.1471-4159.2009.06358.x.
23. Ito, G.; Iwatsubo, T. Re-Examination of the Dimerization State of Leucine-Rich Repeat Kinase 2: Predominance of the Monomeric Form. *Biochem. J.* **2012**, *441*, 987–994, doi:10.1042/BJ20111215.
24. Berger, Z.; Smith, K.A.; Lavoie, M.J. Membrane Localization of LRRK2 Is Associated with Increased Formation of the Highly Active LRRK2 Dimer and Changes in Its Phosphorylation. *Biochemistry* **2010**, *49*, 5511–5523, doi:10.1021/bi100157u.
25. James, N.G.; Digman, M. a.; Gratton, E.; Barylko, B.; Ding, X.; Albanesi, J.P.; Goldberg, M.S.; Jameson, D.M. Number and Brightness Analysis of LRRK2 Oligomerization in Live Cells. *Biophys. J.* **2012**, *102*, L41-3, doi:10.1016/j.bpj.2012.04.046.
26. Partridge, J.R.; Lavery, L. a.; Elnatan, D.; Naber, N.; Cooke, R.; Agard, D. a A Novel N-Terminal Extension in Mitochondrial TRAP1 Serves as a Thermal Regulator of Chaperone Activity. *Elife* **2014**, *3*, 1–21, doi:10.7554/eLife.03487.
27. Durand, D.; Vivès, C.; Cannella, D.; Pérez, J.; Pebay-Peyroula, E.; Vachette, P.; Fieschi, F. NADPH Oxidase Activator P67phox Behaves in Solution as a Multidomain Protein with Semi-Flexible Linkers. *J. Struct. Biol.* **2010**, *169*, 45–53, doi:10.1016/j.jsb.2009.08.009.
28. Liao, J.; Wu, C.-X.; Burlak, C.; Zhang, S.; Sahm, H.; Wang, M.; Zhang, Z.-Y.; Vogel, K.W.; Federici, M.; Riddle, S.M.; et al. Parkinson Disease-Associated Mutation R1441H in LRRK2 Prolongs the “Active State” of Its GTPase Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4055–4060, doi:10.1073/pnas.1323285111.
29. Daniëls, V.; Vancraenenbroeck, R.; Law, B.M.H.; Greggio, E.; Lobbestael, E.; Gao, F.; De Maeyer, M.; Cookson, M.R.; Harvey, K.; Baekelandt, V.; et al. Insight into the Mode of Action of the LRRK2 Y1699C Pathogenic Mutant. *J. Neurochem.* **2011**, *116*, 304–315, doi:10.1111/j.1471-4159.2010.07105.x.
30. Xiong, Y.; Coombes, C.E.; Kilaru, A.; Li, X.; Gitler, A.D.; Bowers, W.J.; Dawson, V.L.; Dawson, T.M.; Moore, D.J. GTPase Activity Plays a Key Role in the Pathobiology of LRRK2. *PLoS Genet.* **2010**, *6*, 19–22, doi:10.1371/journal.pgen.1000902.
31. Terheyden, S.; Nederveen-Schippers, L.M.; Kortholt, A. The Unconventional G-Protein Cycle of LRRK2 and Roco Proteins. *Biochem. Soc. Trans.* **2016**, *44*, doi:10.1042/BST20160224.
32. Jorgensen, N.D.; Peng, Y.; Ho, C.C.-Y.; Rideout, H.J.; Petrey, D.; Liu, P.; Dauer, W.T. The WD40 Domain Is Required for LRRK2 Neurotoxicity. *PLoS One* **2009**, *4*, e8463, doi:10.1371/journal.pone.0008463.
33. Sen, S.; Webber, P.J.; West, A.B. Dependence of Leucine-Rich Repeat Kinase 2 (LRRK2) Kinase Activity on Dimerization. *J. Biol. Chem.* **2009**, *284*, 36346–36356, doi:10.1074/jbc.M109.025437.
34. Bochner, B.R.; Ames, B.N. Complete Analysis of Cellular Nucleotides by Two-Dimensional Thin Layer Chromatography. *J. Biol. Chem.* **1982**, *257*, 9759–9769.
35. Buckstein, M.H.; He, J.; Rubin, H. Characterization of Nucleotide Pools as a Function of Physiological

State in Escherichia Coli. *J. Bacteriol.* **2008**, *190*, 718–726, doi:10.1128/JB.01020-07.

36. Liu, Z.; Mobley, J.A.; DeLucas, L.J.; Kahn, R.A.; West, A.B. LRRK2 Autophosphorylation Enhances Its GTPase Activity. *FASEB J.* **2016**, *30*, 336–347, doi:10.1096/fj.15-277095.
37. Mata, I.F.; Taylor, J.P.; Kachergus, J.; Hulihan, M.; Huerta, C.; Lahoz, C.; Blazquez, M.; Guisasola, L.M.; Salvador, C.; Ribacoba, R.; et al. LRRK2 R1441G in Spanish Patients with Parkinson's Disease. *Neurosci. Lett.* **2005**, *382*, 309–311, doi:10.1016/j.neulet.2005.03.033.
38. Mata, I.F.; Kachergus, J.M.; Taylor, J.P.; Lincoln, S.; Aasly, J.; Lynch, T.; Hulihan, M.M.; Cobb, S.A.; Wu, R.-M.; Lu, C.-S.; et al. Lrrk2 Pathogenic Substitutions in Parkinson's Disease. *Neurogenetics* **2005**, *6*, 171–177, doi:10.1007/s10048-005-0005-1.
39. Deng, J.; Lewis, P. a; Greggio, E.; Sluch, E.; Beilina, A.; Cookson, M.R. Structure of the ROC Domain from the Parkinson's Disease-Associated Leucine-Rich Repeat Kinase 2 Reveals a Dimeric GTPase. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 1499–1504.
40. Smith, W.W.; Pei, Z.; Jiang, H.; Dawson, V.L.; Dawson, T.M.; Ross, C.A. Kinase Activity of Mutant LRRK2 Mediates Neuronal Toxicity. *Nat. Neurosci.* **2006**, *9*, 1231–1233, doi:10.1038/nn1776.
41. Estrada, A.A.; Liu, X.; Baker-Glenn, C.; Beresford, A.; Burdick, D.J.; Chambers, M.; Chan, B.K.; Chen, H.; Ding, X.; Dipasquale, A.G.; et al. Discovery of Highly Potent, Selective, and Brain-Penetrable Leucine-Rich Repeat Kinase 2 (LRRK2) Small Molecule Inhibitors. *J. Med. Chem.* **2012**, *55*, 9416–9433, doi:10.1021/jm301020q.
42. Fell, M.J.; Mirescu, C.; Basu, K.; Cheewatrakoolpong, B.; DeMong, D.E.; Ellis, J.M.; Hyde, L.A.; Lin, Y.; Markgraf, C.G.; Mei, H.; et al. MLI-2, a Potent, Selective, and Centrally Active Compound for Exploring the Therapeutic Potential and Safety of LRRK2 Kinase Inhibition. *J. Pharmacol. Exp. Ther.* **2015**, *355*, 397–409, doi:10.1124/jpet.115.227587.
43. Fuji, R.N.; Flagella, M.; Baca, M.; S. Baptista, M.A.; Brodbeck, J.; Chan, B.K.; Fiske, B.K.; Honigberg, L.; Jubb, A.M.; Katavolos, P.; et al. Effect of Selective LRRK2 Kinase Inhibition on Nonhuman Primate Lung. *Sci. Transl. Med.* **2015**, *7*, 273ra15-273ra15, doi:10.1126/scitranslmed.aaa3634.
44. Baptista, M.A.S.; Dave, K.D.; Frasier, M.A.; Sherer, T.B.; Greeley, M.; Beck, M.J.; Varsho, J.S.; Parker, G.A.; Moore, C.; Churchill, M.J.; et al. Loss of Leucine-Rich Repeat Kinase 2 (LRRK2) in Rats Leads to Progressive Abnormal Phenotypes in Peripheral Organs. *PLoS One* **2013**, *8*, e80705, doi:10.1371/journal.pone.0080705.
45. Herzig, M.C.; Kolly, C.; Persohn, E.; Theil, D.; Schweizer, T.; Hafner, T.; Stemmelen, C.; Troxler, T.J.; Schmid, P.; Danner, S.; et al. LRRK2 Protein Levels Are Determined by Kinase Function and Are Crucial for Kidney and Lung Homeostasis in Mice. *Hum. Mol. Genet.* **2011**, *20*, 4209–4223, doi:10.1093/hmg/ddr348.
46. Shkumatov, A. V.; Strelkov, S. V. DATASW, a Tool for HPLC-SAXS Data Analysis. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2015**, *71*, 1347–1350, doi:10.1107/S1399004715007154.
47. Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M.H.J.; Svergun, D.I. PRIMUS: A Windows PC-Based System for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2003**, *36*, 1277–1282, doi:10.1107/S0021889803012779.
48. Svergun, D.I. Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria. *J. Appl. Crystallogr.* **1992**, *25*, 495–503, doi:10.1107/S0021889892001663.
49. Fiser, A.; Sali, A. ModLoop: Automated Modeling of Loops in Protein Structures. *Bioinformatics* **2003**, *19*, 2500–2501, doi:10.1093/bioinformatics/btg362.
50. Petoukhov, M. V.; Franke, D.; Shkumatov, A. V.; Tria, G.; Kikhney, A.G.; Gajda, M.; Gorba, C.; Mertens, H.D.T.; Konarev, P. V.; Svergun, D.I. New Developments in the ATSAS Program Package for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2012**, *45*, 342–350, doi:10.1107/S0021889812007662.
51. Svergun, D.; Barberato, C.; Koch, M.H. CRY SOL – a Program to Evaluate X-Ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **1995**, *28*, 768–773, doi:10.1107/S0021889895007047.

- 2
52. Svergun, D.I. Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophys.J.* **1999**, *76*, 2879–2886, doi:10.1016/S0006-3495(99)77443-6.
 53. Volkov, V. V.; Svergun, D.I. Uniqueness of *Ab Initio* Shape Determination in Small-Angle Scattering. *J. Appl. Crystallogr.* **2003**, *36*, 860–864, doi:10.1107/S0021889803000268.
 54. Kozin, M.B.; Svergun, D.I. Automated Matching of High- and Low-Resolution Structural Models. *J. Appl. Crystallogr.* **2001**, *34*, 33–41, doi:10.1107/S0021889800014126.
 55. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612, doi:10.1002/JCC.20084.
 56. Sobott, F.; Hernández, H.; McCammon, M.G.; Tito, M.A.; Robinson, C. V. A Tandem Mass Spectrometer for Improved Transmission and Analysis of Large Macromolecular Assemblies. *Anal. Chem.* **2002**, *74*, 1402–1407, doi:10.1021/AC0110552.
 57. Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophys. J.* **2000**, *78*, 1606–1619, doi:10.1016/S0006-3495(00)76713-0.
 58. Laue, T.M.; Shah, B.D.; Ridgeway, T.M.; Pelletier, S.L. Analytical Ultracentrifugation in Biochemistry and Polymer Science. In *Royal Society of Chemistry*; 1992; pp. 90–125.
 59. Brautigam, C.A. Chapter Five – Calculations and Publication-Quality Illustrations for Analytical Ultracentrifugation Data. In *Methods in Enzymology*; 2015; Vol. 562, pp. 109–133 ISBN 9780128029084.
 60. Schuck, P. On the Analysis of Protein Self-Association by Sedimentation Velocity Analytical Ultracentrifugation. *Anal. Biochem.* **2003**, *320*, 104–124.
 61. Tang, G.; Peng, L.; Baldwin, P.R.; Mann, D.S.; Jiang, W.; Rees, I.; Ludtke, S.J. EMAN2: An Extensible Image Processing Suite for Electron Microscopy. *J. Struct. Biol.* **2007**, *157*, 38–46, doi:10.1016/j.jsb.2006.05.009.
 62. Hohn, M.; Tang, G.; Goodyear, G.; Baldwin, P.R.; Huang, Z.; Penczek, P.A.; Yang, C.; Glaeser, R.M.; Adams, P.D.; Ludtke, S.J. SPARX, a New Environment for Cryo-EM Image Processing. *J. Struct. Biol.* **2007**, *157*, 47–55, doi:10.1016/j.jsb.2006.07.003.



3

Complex analysis of fluorescence intensity fluctuations of molecular compounds

M.M. Yatskou^{1,*}, V.V. Skakun¹, **L.M. Nederveen-Schippers**²,
A. Kortholt² and V.V. Apanasovich³

¹ Department of Systems Analysis and Computer Modelling, Belarusian State University, Minsk, 220030, Belarus. ² University of Groningen, 9747AG Groningen, The Netherlands. ³ Institute of IT & Business Administration, Minsk, 220004, Belarus. * Corresponding author: yatskou@bsu.by

Author Contributions: MY, VS, **LNS**, AK, and VA designed the experiments, MY and VS performed and analyzed the simulation experiments, **LNS** performed the biological experiments and contributed to the analysis. MY, VS and VA wrote the manuscript.

This chapter has been published in: Journal of Applied Spectroscopy. 2020; J Appl Spectrosc. 2020 Sept 11;87(4):685–692. doi:10.1007/s10812-020-01055-6 (translated from Zhurnal Prikladnoi Spektroskopii 2020 July-August; 87(4):628–636)

Abstract

A method is proposed for the complex analysis of fluctuations in the fluorescence intensity of molecular compounds, which allows determining the structural composition of protein oligomers. The idea of the method is to analyze the photon counting histograms of experimental measurements using principal component analysis to assess the presence of oligomeric compounds, and to perform hierarchical cluster analysis, to determine the data classes corresponding to various molecular compounds, followed by selecting cluster medoids to determine the oligomeric composition of protein complexes. The efficiency of the analysis algorithms developed within the framework of the proposed method was confirmed on simulated and experimental photon counting histograms of the measured fluorescence intensity fluctuations of monomeric and dimeric forms of green-fluorescent protein (GFP).

Keywords

Fluorescence intensity fluctuation, photon counting histogram, molecular compounds, protein oligomers, data mining, principal component analysis, hierarchical cluster analysis, green-fluorescent protein (GFP).

Introduction

Fluorescence fluctuation spectroscopy is widely used to study the diffusion of proteins and their interactions in living cells [1–3]. In the course of the experiment, the fluorescence of molecules bound or freely moving in a solution or a cell is recorded in a certain small volume (up to 10^{-18} m³) formed by an extremely focused laser beam. Fluctuations in fluorescence intensity are primarily due to changes in the number and location of molecules in the recorded volume, as well as their interaction and the properties of the medium. The oligomeric composition of a protein compound can be determined by analyzing the amplitude of fluctuations in fluorescence intensity over time (methods for analyzing the distribution of fluorescence intensity — PCH (photon counting histogram) [4] and FIDA (fluorescence intensity distribution analysis) [5]). In the PCH and FIDA methods, a histogram of the number of photocounts (PC) is plotted at a given recording time interval to determine the concentration of a protein freely emitting or labeled with a luminescent dye. The recorded fluorescence intensity of the sample is directly proportional to the number of fluorescent molecules that form the studied molecular complex, which makes it possible to estimate the number of molecules inside the protein complex and the size of the complex [6, 7].

To analyze the distribution of the number of photocounts, various mathematical models [4–7] and optimization methods are usually used, among which the least squares method with Levenberg–Marquardt optimization [8] is used most often, which makes it possible to obtain information on the diffusion and structural properties of the studied protein compounds in the first approximation. However, the classical iterative algorithms for data analysis have a number of significant limitations. They do not allow one to accurately determine the number and type of molecular oligomers, perform a local rather than global search for model parameters, and require significant computational costs for data analysis. An alternative approach to solving this problem is the use of mining algorithms and large multidimensional data, the essence of which is the simultaneous global analysis of the entire data set as a whole [9–12].

In the present work, we propose a method for the complex analysis of fluorescence intensity fluctuations and the PCHs based on them using intelligent analysis algorithms in order to determine the oligomeric composition of molecular compounds.

Methodology

The developed method is based on the hypothesis of the separability of a set of multidimensional experimental data in a certain information space into several populations representing various molecular oligomeric compounds [10]. A small measurement volume is considered, in which molecular compounds of the same type prevail in a series of short time intervals. A normal distribution of the measured attributes is assumed for molecular compounds of the same type in the allocated space. For example, protein monomers can form a cloud or spherical Gaussian cluster of data in a multidimensional space based on measurable attributes. If, however, protein oligomers are added to the monomeric forms of the protein, then the cloud is extended or divided into two parts along a certain line connecting the centers of the two populations. In the extreme case, two clouds or clusters of these monomers and oligomers are expected. Thus, if groups of data are divided into clusters in a multidimensional space of attributes, this confirms the presence of several forms of protein compounds. Tasks of this kind are solved using data mining algorithms such as data dimensionality reduction and cluster analysis [10, 13, 14]. Dimensionality reduction algorithms allow switching to a low-dimensional space without losing the essence of information [15, 16]. Cluster analysis algorithms make it possible to determine clusters of data specified in varying degrees of similarity, the number of which may be associated with aggregates of molecular compounds. Thus, applying principal components analysis (PCA) will make it possible to carry out such a rotation, as a result of which the axis of the first principal component coincides with the diagonal of the data cloud in multidimensional space [17]. Therefore, the relative fraction of the scatter attributable to the first principal component for two types of molecular compounds (an elongated ellipsoid or two spherical data clouds in a multidimensional space of attributes is expected) should differ significantly from that for a monomer solution (one spherical cloud). It should be noted that the scatter diagram of the first two principal components is informative in the sense of defining the data structure in two-dimensional space.

The idea behind the method of complex analysis is to calculate the PCH based on the recorded fluorescence intensities (it is possible to use other attributes, for example, the autocorrelation function or factorial cumulants of the distribution of the number of photocounts [18]), the use of the PCA to assess the presence of oligomeric compounds and hierarchical cluster analysis to determine groups of data, corresponding to various molecular compounds, followed by the isolation of cluster medoids (PCHs having the smallest average distances to the remaining objects of the corresponding clusters) to assess the parameters of the oligomeric composition of protein complexes. Comprehensive analysis requires the availability of experimental data for the reference (monomers) and tested (oligomeric forms) samples. The block diagram of the developed method is shown in Fig. 1 consider the main stages of the method.

Calculation of the PCH

We calculate N of the PCH based on the registered sets of fluorescence intensities S_i , $i = 1, 2, \dots, N$, and form objects n_1, n_2, \dots, n_N , characterized by attributes X_1, X_2, \dots, X_K , – histogram channels representing the frequencies of occurrence f_j of the number of photons $l = (j - 1), j = 1, 2, \dots, K$, during a certain (short) time interval Δt . As a standard or reference sample, we use the experimental data of the monomer solution, and as a test sample – data for the oligomeric forms of the protein.

Data Dimensionality Reduction

The PCA method is applied to datasets of reference and test samples. In the PCA, such a linear transformation is defined, as a result of which the initial data X_1, X_2, \dots, X_K are expressed by a set of principal components Z_1, Z_2, \dots, Z_K , where the first M principal components ($M \ll K$) provide the required fraction γ of the variance of groups of attributes. In expanded form, the principal component Z_j is expressed through the attribute vectors X_1, X_2, \dots, X_K :

$$Z_j = a_{1j} X_1 + a_{2j} X_2 + \dots + a_{Kj} X_K, \quad (1)$$

where α_{ij} are the loading parameters of the principal components. The relative proportion of the scatter (%) attributable to the principal component Z_j is:

$$\alpha_j = 100 \frac{D(Z_j)}{D(Z_1) + D(Z_2) + \dots + D(Z_K)}, \quad (2)$$

where $D(Z_j)$ is the variance of the component Z_j . If the relative proportions of the scatter in the reference and the tested samples, which fall on the first principal component Z_1 , are the same, then to assume that there are no oligomers means to stop the algorithm. Otherwise, permit the presence of oligomers and continue the algorithm.

Hierarchical Cluster Analysis of the Reference Sample (HCARS)

A hierarchical cluster analysis of the histograms of the reference sample $n_1^R, n_2^R, \dots, n_N^R$ is performed in the space of initial attributes. In this case, it is necessary to specify a method for comparing objects to each other (or a measure of similarity, for example, Euclidean, Minkowski, correlation distance). In the developed method to eliminate inter-experimental inhomogeneities associated with separate measurements of the reference and test samples, we propose to use the standardized Euclidean distance (invariant to inhomogeneity in the data) [10]:

$$d_e(n_i, n_j) = \sqrt{\sum_{l=1}^K \frac{(x_{il} - x_{jl})^2}{\sigma_l^2}}, \quad (3)$$

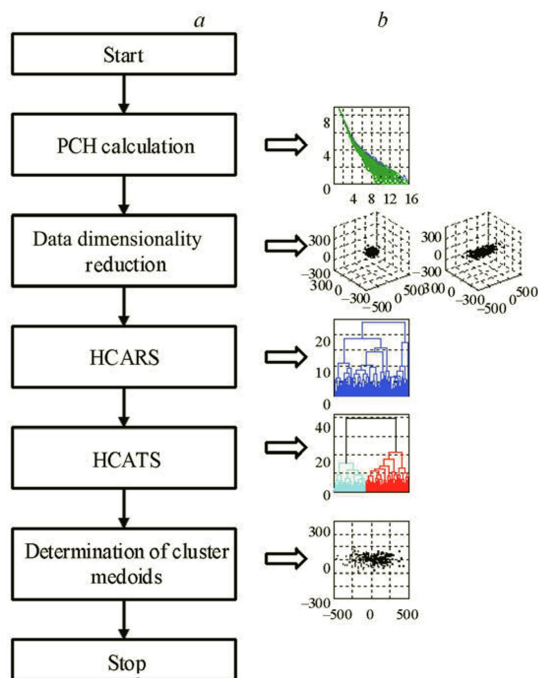


Figure 1: Block diagram of the method (a) and diagram of the results of its main stages (b) for studying fluctuations of the fluorescence intensity of molecular compounds using data mining algorithms.

where χ_{ii} and χ_{ji} are coordinates of objects n_i and n_j ; σ_l^2 is the variance of the attribute X_l . We determine the maximum connection distance (or threshold) d_1 on the dendrogram, at which the data are combined into one cluster. The maximum connection distance d_1 is used as a threshold for finding the number of oligomer clusters on the dendrogram for the test data.

Hierarchical Cluster Analysis of the Test Sample (HCATS)

A hierarchical cluster analysis of the histograms of the tested sample $n_1^T, n_2^T, \dots, n_N^T$ is performed in the space of initial attributes. Using the threshold d_1 found in the previous step of the algorithm, we select data clusters on the dendrogram. Assume that one cluster belongs to monomers, and the other(s) – to oligomeric forms.

Determination of Cluster Medoids

Clusters of monomers and oligomers are displayed on the scatter diagram of the first two principal components. Datasets are formed by calculating medoids in each cluster to accurately determine the parameters of molecular compounds using PCH and FIDA methods.

Materials and Methods

Consider simulated and experimental data. The simulated data make it possible to qualitatively assess the performance of the method and explore the limits of application. The experimental data are used to confirm the fundamental possibility of applying the developed approach to solving real problems of experimental research.

A simulation model of the photocount flow with a given distribution of the number of photocounts is presented in [19]. The number of photons emitted by the molecule during the observation time T is approximated by the Poisson distribution with the intensity

$$\lambda_f = \langle q \rangle TB(r), \quad (4)$$

where $\langle q \rangle$ is the brightness, or the average number of photons emitted by one molecule per unit of time; $B(r)$ is the exposure profile function; $r(x, y, z)$ is the radius vector of the molecule. A three-dimensional Gaussian distribution is used as a function of the exposure profile $B(r)$. The number of molecules in solution in a certain volume obeys the Poisson distribution with the parameter

$$\lambda_m = \langle N_m \rangle V_0, \quad (5)$$

where $\langle N_m \rangle$ is the average number of molecules of the test sample per unit volume; V_0 is the exposure volume. For each molecule, the coordinates of the location in the volume V_0 (according to the uniform distribution law) and the number of emitted photons (according to the Poisson distribution with the intensity λ_f) are generated. If a mixture of molecules of different types is simulated, then it is necessary to perform photon generation cycles for each type of molecule. The generation cycle is repeated iteratively until the accumulation of the number of photons, at which a PCH with a given signal-to-noise ratio is formed. To take into account the effect of scattering of data or “blurring” of PCH clusters caused by the influence of various distortions, such as the presence of unremovable impurities that quench or stimulate fluorescence of molecules, high background noise, flare and degradation of dyes, we use modeling of model parameters that have a normal distribution with a given mathematical expectation and standard deviation σ . Variation of σ makes it possible to control the scatter of data or the blur of clusters of PCH curves in a multidimensional space of time samples.

The simulated data is an example of an idealized system of two types of molecules: a monomer (M) and a dimer (D) of a certain protein (for example, GFP in solution), separately generated PCHs of which are characterized by the average

number of molecules in the recording volume and their average brightness $\langle N^M \rangle = 2$, $\langle q^M \rangle = 5 \cdot 10^4$ and $\langle N^D \rangle = 1$, $\langle q^D \rangle = 10^5$. Observation interval is $T = 5 \cdot 10^{-5}$ s. Modeling was carried out with $\sigma = 0.02$ and 0.2 of the absolute values of the parameters $\langle N^M \rangle$, $\langle q^M \rangle$, $\langle N^D \rangle$, and $\langle q^D \rangle$.

Experimental data — well-known monomeric and dimeric forms of the green fluorescent protein GFP S65T [20] — were provided by the Cell Biochemistry Laboratory of the University of Groningen (Netherlands). Reference samples: GFP protein in buffered lysis solution (50 mM Tris, 50 mM NaCl, 5 mM DTT, 5 mM MgCl₂, 1% PI mix, 1% Triton X-100); separate measurements of the monomer (mGFP) and the stable dimer (diGFP, synthesized by liganding the pDM313 vector into pDM334 at the Spel/XBAl binding sites) of GFP protein in lysates of *Dictyostelium* cells. A test sample is a mixture of equal proportions of low concentrations ($\langle N_m \rangle < 1$) of mGFP and diGFP proteins in *Dictyostelium* cell lysate. The measurements of the first sample were performed using a Leika TCS fluorescence confocal inverted microscope equipped with a lens immersed in oil (100×, 1.4NA) and a PicoHarp 300 (PicoQuant) photocount counting and recording system. The second and third samples were examined using a scanning inverted confocal microscope LSM 710 (Carl Zeiss) equipped with a lens immersed in water (100×, 1.2NA) and a Confocor3 measurement system (Carl Zeiss). The fluorescence of the samples was excited at $\lambda = 488$ nm and recorded in the $\lambda = 505$ – 610 nm range.

The simulated data make it possible to investigate the applicability of the developed method in the case of different separability of data clusters (varied by the parameter σ) corresponding to protein compounds. The data representing the GFP protein in the buffer solution and the cell lysate are experimentally confirmed and make it possible to check the efficiency of the method using examples of real model data. A mixture of monomeric and dimeric forms of the GFP protein is an example of a dataset specifically containing various forms of protein aggregation. Assuming that molecules of the same type were predominantly found in the observation volume, the PCHs of the experimental samples were constructed over a time interval of $5 \cdot 10^{-2}$ s or less in one measurement of fluorescence intensity fluctuations with a duration of 120 s.

The algorithms were implemented in the Matlab mathematical programming environment using the `pdist`, `linkage`, `cluster`, and `eig` functions, which integrate algorithms for hierarchical cluster analysis and PCA [21]. The hierarchical method of cluster analysis was used, and the most common method for calculating the distance (standardized Euclidean) and the measure of cluster similarity (Ward) were investigated [13]. The data centering procedure is applied in the PCA. To assess the error ϵ of restoring the PCHs of various types of molecules, the ratio of incorrectly determined PCHs to the total number of PCHs (in %) was considered.

Results and Discussion

The results of the analysis of the simulated datasets using the algorithms of the integrated approach are shown in Fig. 2 and in Table 1. The analysis of the simulated data was carried out separately for monomers and dimers (Fig. 2a and 2b). The relative proportion of the scatter α_1 for the first principal component is 54.6 and 58.8% for monomers and dimers, and the data clouds in the space of the principal components have a spherical Gaussian shape. The threshold value of the similarity measure, at which molecules form a single cluster $d_1 = 15$, is a criterion for determining clusters of different molecular shapes. The connection distance of the resulting clusters into one is <2 , which indicates a significant similarity of the combined clusters.

The application of the algorithms of the developed method to the analysis of the combined set of simulated data makes it possible to accurately determine the samples of monomeric and dimeric forms of proteins (error $\epsilon = 0$), which is confirmed by the high relative fraction of the scatter falling on the first principal component, $\alpha_1 > 98\%$ (for monomers 54.6%), clear separability of data into two clusters in the space of the principal components Z_1 and Z_2 (Fig. 2c), long connection distances of the resulting clusters into one (>50), which confirms the importance of the difference between clusters. It should be noted that the method successfully works under the conditions of the considered example of blurring and partial overlapping of data clusters ($\sigma = 0.2$, $\epsilon = 1.5\%$; Fig. 2d), which is typical for molecular systems such as a mixture of GFP monomers and dimers in a cell lysate. Samples of monomeric and dimeric forms of proteins were determined: the relative proportion of scatter $\alpha_1 = 99\%$, the data form two clusters in the space of the principal components Z_1 and Z_2 (Fig. 2d), the line length of the unification of the resulting clusters into one is >30 .

In the course of the study, together with the standardized Euclidean distance, three additional measures for calculating the similarity between objects, invariant to data heterogeneity, such as Mahalanobis, correlation and Spearman were considered [9, 13, 14]. The best results were obtained for the distances of the standardized Euclidean distance and Mahalanobis. However, the Mahalanobis measure requires the computation of the covariance matrix of the input data, which can be costly in the case of analyzing large datasets ($N \rightarrow \infty$, $K \rightarrow \infty$).

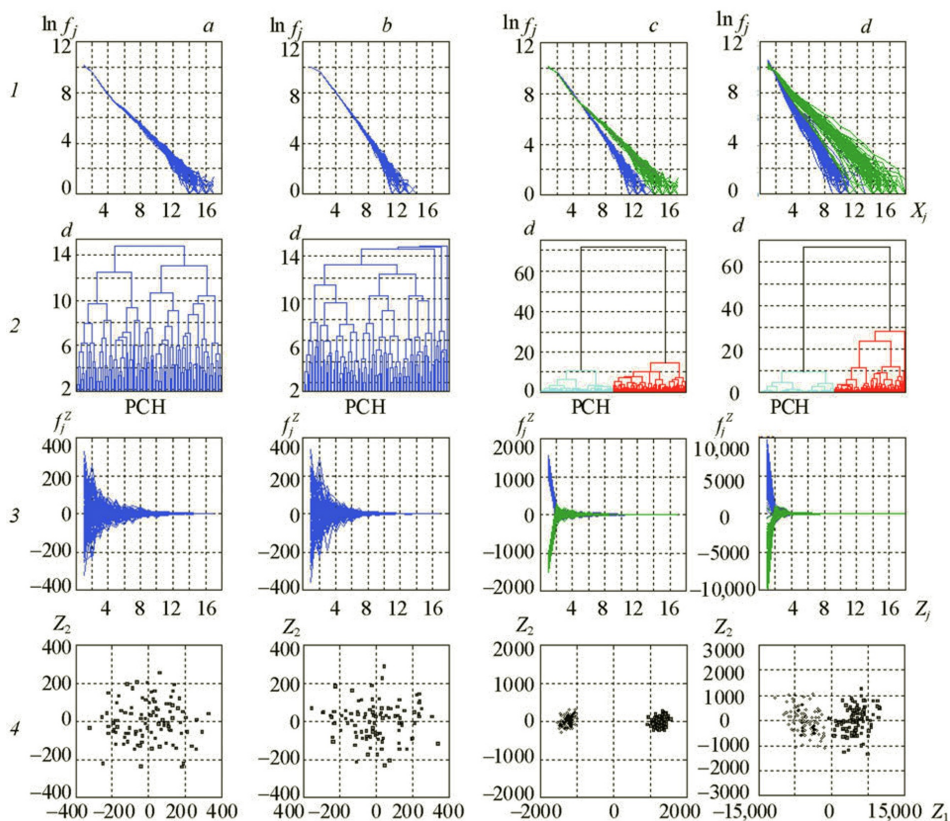


Figure 2: The results of the analysis of the simulated data using the developed method, based on the algorithms of the principal component method (data centering is performed) and hierarchical cluster analysis (the standardized Euclidean measure of similarity of objects and the Ward connection distance for combining clusters are implemented); modeling parameters: $\langle N^M \rangle = 2$, $\langle q^M \rangle = 5 \cdot 10^4$ and $\langle N^D \rangle = 1$, $\langle q^D \rangle = 10^5$; a) monomers, $\sigma = 0.02$; b) dimers, $\sigma = 0.02$; c and d) combined sets of monomers and dimers with $\sigma = 0.02$ and 0.2 ; 1) photon counting histograms on a logarithmic scale in the space of the initial attributes X_1, X_2, \dots, X_K ; 2) dendrograms of photon counting histograms PCHs, d is the measure of cluster similarity; 3) photon counting histograms in the space of principal components $Z_1, Z_2, \dots, Z_K, f_j^Z$ — linearly transformed frequencies of occurrence of the number of photons in the coordinates of principal components; 4) histograms of photon counts in space of the first two principal components; the dimensionality of the axes of the principal components is represented by the linearly transformed frequencies of occurrence of the number of photons in the coordinates of components 1 and 2; shades of gray indicate monomeric and dimeric forms of proteins.

The results of the analysis of experimental datasets using the algorithms of the integrated approach are shown in Fig. 3 and in Table 1. Study of the data for the GFP protein in a buffer solution allows one to determine the threshold value of the similarity measure ($d_1 = 23$), at which the monomers form a single cluster, for use in the subsequent analysis of protein compounds (Fig. 3a). The connection distance of the resulting clusters into one (<5), the spherical shape of the data cloud in the space of the first two principal

components (Fig. 3a) and a low relative proportion of the scatter $\alpha_1 = 50.5\%$ (Table 1), which falls on the first principal component, qualitatively confirm the fundamental principle of the working hypothesis proposed in the implemented method. As a result of the analysis of the combined experimental data of mGFP and diGFP proteins in cell lysates, the presence of two forms of proteins corresponding to monomeric and dimeric forms (Fig. 3b) was confirmed: $\alpha_1 = 99.9\%$, the data form two clusters in the space of the principal components, the connection distance of the resulting clusters into one >40 . Analysis of the experimental data of a mixture of mGFP and diGFP proteins in the cell lysate revealed the presence of two forms of protein oligomers. The relative proportion of the scatter α_1 , which falls on the first principal component of the tested data, at 93.6% significantly exceeds the value of 50.5% obtained for monomeric forms of the GFP protein in a buffer solution. The connection distance at which the final cluster is formed is 40 (Fig. 3c), the data form two clusters in the space of the principal components, at 18 the connection distance of the resulting clusters into one significantly exceeds the value of 5 for GFP monomers. The value ≥ 23 should be taken as the threshold value for determining the number of nonmonomeric form clusters. At a connection distance of 23, two clusters formed by the majority of mGFP or diGFP molecules can be distinguished on the dendrogram of the tested data (Fig. 3c). Further evaluation of the parameters of protein complexes can be carried out in the course of analysis of medoids of the obtained PCH clusters using classical algorithms for analyzing fluorescence spectroscopy data [5, 6]. Note that the monomers of the GFP protein form a spherical cluster of data in the space of the first two principal components (Fig. 3a), while an elongated ellipsoidal cloud is observed for a mixture of mGFP or diGFP, formed by clusters of monomers and dimers of compounds (Fig. 3c).

Table 1: Relative Proportion of Scatter (in %) for the First 10 Principal Components Obtained During Analysis of Simulated (SD) and Experimental Datasets Using Principal Component Analysis

Components	1	2	3	4	5	6	7	8	9	10
SD, monomers	54.564	29.263	8.134	3.079	2.318	1.120	0.701	0.471	0.166	0.094
SD, dimers	58.775	25.822	9.317	3.410	1.611	0.704	0.195	0.100	0.045	0.014
SD 1*	98.768	0.823	0.206	0.104	0.048	0.027	0.012	0.007	0.003	0.001
SD 2**	98.998	0.812	0.160	0.017	0.008	0.003	0.002	0.001	0.000	0.002
GFP	50.502	16.554	12.656	9.545	6.331	2.674	1.137	0.343	0.150	0.077
mGFP/diGFP	99.869	0.041	0.025	0.018	0.014	0.012	0.007	0.005	0.004	0.003
mGFP/diGFP mixture	93.592	4.161	1.360	0.470	0.175	0.104	0.055	0.028	0.023	0.011

* Monomers/dimers, $\sigma = 0.02$;

** Monomers/dimers, $\sigma = 0.2$.

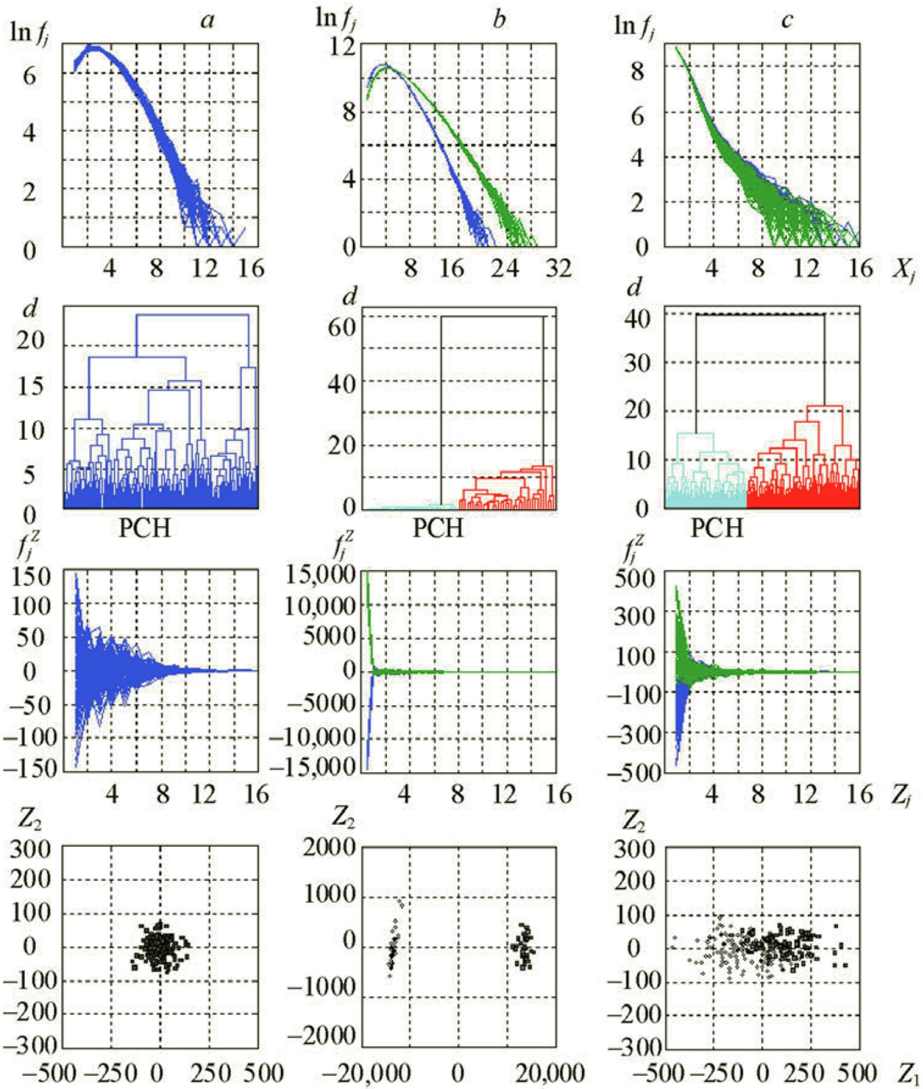


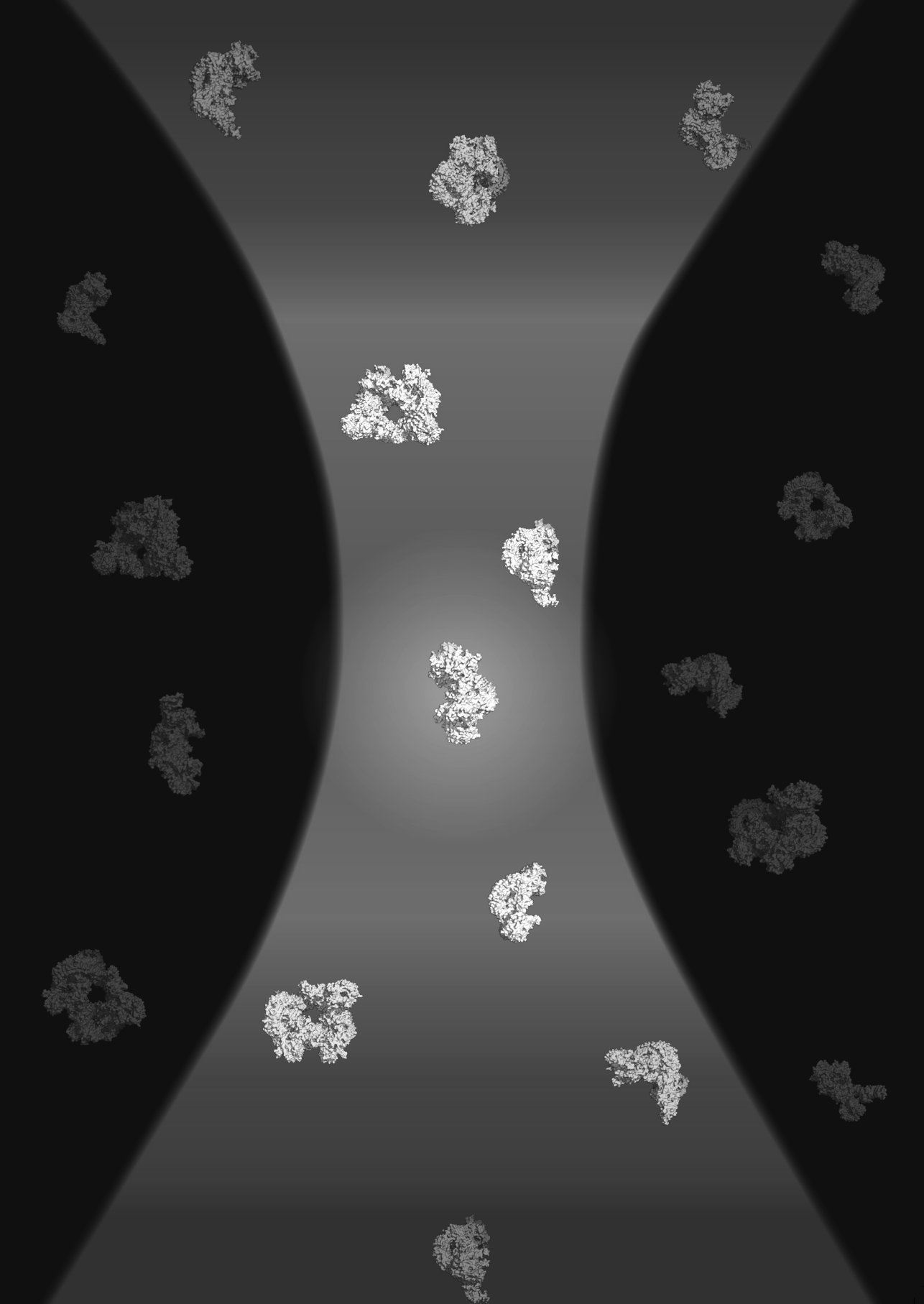
Figure 3: Results of the analysis of experimental data sets using the developed method, based on the algorithms of the principal component analysis (data centering was performed) and hierarchical cluster analysis (the Euclidean measure of similarity of objects and the Ward connection distance for combining clusters were implemented): a) GFP protein in a buffer solution, b) mGFP and diGFP proteins in cell lysates, c) mixture of mGFP and diGFP proteins in cell lysates; designations as in Fig. 2.

Conclusions

A method for the complex analysis of fluctuations of the fluorescence intensity of molecular compounds is proposed, which makes it possible to determine the structural composition of protein oligomers and complements the classical methods of PCH and FIDA analysis. The efficiency of the algorithms developed within the framework of the proposed method was confirmed during the analysis of simulated and experimental data representing the fluorescence of monomeric and dimeric forms of the GFP protein. The developed method has the following advantages over the classical method for analyzing data from fluorescence fluctuation spectroscopy: it improves the accuracy of data analysis, since it uses the entire data set, rather than individual histograms; provides computational performance due to the high speed of execution of procedures of the method of principal components and cluster analysis in comparison with a separate analysis of the full set of histograms; provides the ability to visualize data in the space of the first two principal components, which is much more informative than a diagram of a complete set of initial histograms.

References

1. E. L. Elson and D. Magde, *Biopolymers*, **13**, No. 1, 1–27 (1974).
2. A. Kitamura and M. Kinjo, *Int. J. Mol. Sci.*, **19**, No. 4, pii: E964, 1–18 (2018).
3. S. Veerapathiran and T. Wohland, *J. Biosci.*, **43**, No. 3, 541–553 (2018).
4. Y. Chen, J. D. Müller, P. T. So, and E. Gratton, *Biophys. J.*, **77**, 553–567 (1999).
5. P. Kask, K. Palo, D. Ullmann, and K. Gall, *Proc. Natl. Acad. Sci. USA*, **96**, No. 24, 13756–13761 (1999).
6. Y. Chen, L. N. Wei, and J. D. Müller, *Proc. Natl. Acad. Sci. USA*, **100**, No. 26, 15492–15497 (2003).
7. V. V. Skakun and V. V. Apanasovich, *Vestnik BSU. Ser. 1, Physics. Mathematics. Informatics*, **1**, 52–59 (2016).
8. D. Marquardt, *SIAM J. Appl. Math.*, **11**, No. 2, 431–441 (1963).
9. N. N. Yatskov, *Data Mining* [in Russian], BSU, Minsk (2014).
10. N. N. Yatskov, V. V. Skakun, and V. V. Apanasovich, *Applied Problems of Optics, Informatics, Radiophysics and Condensed Matter Physics* [in Russian], NII PFP BSU, Minsk (2019), pp. 122–124.
11. M. Bramer, *Principles of Data Mining*, Springer, London (2013).
12. C. C. Aggarwal, *Data Mining: The Textbook*, eBook, Springer (2015).
13. I. D. Mandel', *Cluster Analysis* [in Russian], Finansy and Statistika, Moscow (1988).
14. M. B. Lagutin, *Visual Mathematical Statistics* [in Russian], BINOM, Laboratoriya Znaniy, Moscow (2007).
15. P. V. Nazarov, A. K. Wienecke-Baldacchino, A. Zinovyev, U. Czerwińska, A. Muller, D. Nashan, G. Dittmar, F. Azuaje, and S. Kreis, *BMC Med. Genom.*, **12**, No. 1, 132(1–17) (2019).
16. N. Sompairac, P. V. Nazarov, U. Czerwinska, L. Cantini, A. Biton, A. Molkenov, Z. Zhumadilov, E. Barillot, F. Radvanyi, A. Gorban, U. Kairov, and A. Zinovyev, *Int. J. Mol. Sci.*, **20**, No. 18, E4414 (1–27) (2019).
17. I. T. Jolliffe, *Principal Component Analysis*, Springer, New York (2002).
18. V. V. Skakun, E. G. Novikov, T. V. Apanasovich, and V. V. Apanasovich, *Methods Appl. Fluores.*, **3**, No. 4, 1–12 (2015).
19. I. P. Shingaryov, V. V. Skakun, and V. V. Apanasovich, *Methods Mol. Biol.*, **1076**, 743–755 (2014).
20. A. Kortholt, J. S. King, I. Keizer-Gunnink, A. J. Harwood, and P. J. M. Van Haastert, *Mol. Biol. Cell*, **18**, No. 12, 4772–4779 (2007).
21. N. N. Yatskov and E. V. Lisitsa, *Data Mining: Guidelines for Laboratory Work* [in Russian], BSU, Minsk (2019).



4

Combined FCS and PCH Analysis to Quantify Protein Dimerization in Living Cells

Laura M. Nederveen-Schippers¹, Pragya Pathak¹, Ineke Keizer-Gunnink¹,
Adrie H. Westphal², Peter J. M. van Haastert¹, Jan Willem Borst², Arjan Kortholt^{1,*}
and Victor Skakun^{3,*}

¹Department of Cell Biochemistry, University of Groningen, 9747 AG Groningen, The Netherlands.

²Laboratory of Biochemistry, Wageningen University & Research, 6708 WE Wageningen, The Netherlands.

³Department of Systems Analysis and Computer Simulation, Belarusian State University, 220030 Minsk, Belarus. *Corresponding authors: a.kortholt@rug.nl (AK), skakun@bsu.by (VS)

Author Contributions: PvH, AK, JWB, VS, and **LNS** designed the experiments; **LNS**, PP, and IKG performed the experiments with the help of AW and JWB; VS developed and adapted the data analysis algorithms and software, and **LNS**, PP and IKG analyzed the data. All authors contributed to writing of the manuscript and approved it prior to submission. All authors have read and agreed to the published version of the manuscript.

Abstract

Protein dimerization plays a crucial role in the regulation of numerous biological processes. However, detecting protein dimers in a cellular environment is still a challenge. Here we present a methodology to measure the extent of dimerization of GFP-tagged proteins in living cells, using a combination of fluorescence correlation spectroscopy (FCS) and photon counting histogram (PCH) analysis of single-color fluorescence fluctuation data. We named this analysis method brightness and diffusion global analysis (BDGA) and adapted it for biological purposes. Using cell lysates containing different ratios of GFP and tandem-dimer GFP (diGFP), we show that the average brightness per particle is proportional to the fraction of dimer present. We further adapted this methodology for its application in living cells, and we were able to distinguish GFP, diGFP, as well as ligand-induced dimerization of FKBP12-GFP. While other analysis methods have only sporadically been used to study dimerization in living cells and may be prone to errors, this paper provides a robust approach for the investigation of any cytosolic protein using single-color fluorescence fluctuation spectroscopy.

Keywords

Brightness and diffusion global analysis; *Dictyostelium discoideum*; dimeric protein; GFP; FK506 binding protein 12; fluorescence correlation spectroscopy; fluorescence fluctuation spectroscopy; photon counting histogram

Introduction

Oligomerization and complex formation of proteins in organisms is one of the key mechanisms for the orchestration of protein function and activities. Classic examples are the dimerization and activation of protein kinases like extracellular signal-regulated kinase 2 (ERK2) and rapidly accelerated fibrosarcoma (RAF) [1–3]. Most nuclear receptors are monomeric in the cytosol and become dimeric and active upon the binding of ligand molecules [4,5], although, for the human androgen receptor, this may be the other way around [6]. Unraveling the dimerization mechanism of the androgen receptor is important because it has been implicated in hormone-related diseases [7]. In fact, dimerization and oligomerization play a role in many diseases, especially in neurodegenerative diseases, e.g., leucine-rich repeat kinase 2 and alpha-synuclein in Parkinson's disease and tau protein in Alzheimer's disease.

To study the dynamic process of dimerization it is logical to use in cellulo techniques such as 'live cell imaging'. Over the past decades, numerous approaches have been developed, all based on the non-invasive visualization of fluorescently-labeled proteins. For the detection of dimers it is possible to use a combination of two differently colored labels, as used in techniques like Förster resonance energy transfer (FRET) and fluorescence cross-correlation spectroscopy (FCCS) [8,9]. However, the creation of protein constructs with the right proximity and spectral properties of fluorophores is often a daunting task. Therefore, several techniques have been developed using a single fluorophore, most of which are based on fluorescence fluctuation spectroscopy (FFS).

FFS exploits the fluctuation of fluorescence signals, caused by either diffusion of fluorescent particles, through a tiny, illuminated volume or reversible change of its spectroscopic properties, e.g., singlet-state–triplet-state transitions, recorded from a stack of images or from a time-observed single spot. This fluctuation pattern is influenced by the diffusion speed and brightness of the particles. The diffusion and brightness per particle are different for monomer and dimer because of their different size or mass and number of fluorophores. In this way co-diffusion of multiple proteins in a dimer/oligomer form can be observed without any requirements for proximity of fluorophores within the complex, and monomer can be distinguished from the dimer with the same fluorophore.

To calculate the diffusion and brightness per particle from FFS data, various analysis techniques have been developed [10–12]. Some of these techniques have been applied for analysis of the dimerization of GFP-tagged proteins in living cells, i.e., number and brightness analysis (N&B [13]), raster image correlation spectroscopy (RICS [14]), photon counting histogram (PCH [15]) fluorescence intensity distribution analysis (FIDA

[16]) and time integrated fluorescence cumulant analysis (TIFCA [12]).

N&B and RICS are based on the analysis of images, yielding valuable information on the location of oligomeric complexes in the cell, especially on membranes. However, the vast number of pixels analyzed simultaneously is at the expense of precision. For an accurate determination of molecular brightness and diffusion parameters in the intracellular environment, single-point FFS as in FCS, PC(M)H, and FI(M)DA would be the method of choice [17–23].

In single-point FFS, the fluorescence intensity fluctuations are caused by the diffusion of proteins through a very small volume, which is generated by a tightly focused laser beam of a confocal microscope. The fluorescence intensity at a specific time point and the next time point are correlated, and the amount of correlation of a specific time-step is dependent on the diffusion speed of the molecules crossing the observation volume. The correlation shows a time-dependent decay and can be fitted to a chosen autocorrelation function (ACF) describing the physical model of diffusion.

In fluorescence correlation spectroscopy (FCS), these ACF curves are used to estimate the diffusion time and number of fluorescent particles. By simply comparing the latter to the total photon count rate, the molecular brightness can be calculated, which is defined as the average number of photons detected per molecule per second (cpms). FCS has been widely used to study protein concentration, diffusion, kinetics, and aggregation in living cells, as well as dimerization [24,25]. However, to study dimerization in living cells, a more sophisticated and accurate method is required to calculate the molecular brightness such as PCH or FIDA analysis.

PCH and FIDA are two related techniques, both based on the analysis of the amplitudes in the fluctuation of fluorescence intensity. The distributions of these amplitudes are plotted as photon counting distributions (PCD) and are derived from the same FFS data as the ACF curves. In the case of analysis of multiple distributions, they are called PCMH and FIMDA [18,22,23]. PCMH and FIMDA allow to extract information not only about brightness, but also about diffusion parameters. However, to get the same information content about the diffusion as available in the analysis of a single ACF in FCS, the number of analyzed PCDs should be equal to the number of points in the ACF, which is around one hundred and fifty. The collecting and analysis of a high number of PCDs, calculated at different binning times, is time consuming, whereas the same information, specifically on diffusion, is obtained by fitting just one ACF in FCS. For the determination of brightness, however, PCH and FIDA, and their extensions PCMH and FIMDA, are suitable methods. PCH was first applied to dimerization in cells by Chen et al. using tandem-dimer EGFP as well as concentration-dependent dimerization of

several nuclear receptors [15]. For samples with low signal-to-noise ratios, which is often the case in cellular samples, additional analysis tools are required.

Here we describe the combination of FCS and PCH in the analysis of molecular brightness and diffusion speed and its application in the analysis of dimerization of cytosolic proteins in living cells. A specific analysis technique has been developed which allows the simultaneous analysis of FCS and PCH by linking common parameters [26]. The advantage of this combined FCS and PCH global analysis method is that it enhances the advantages of PCH in estimation of brightness by adding the power of FCS in estimation of diffusion to the analysis [26]. Compared to individual FCS and PCH analyses (as well as other techniques), the combined global FCS and PCH analysis is superior at lower signal-to-noise ratios, making the analysis more robust and accurate. Moreover, this technique can be performed on a standard FCS microscope using single-color fluorescence.

Previously, the combined FCS and PCH global analysis method has been applied to purified dimeric GFP (diGFP) *in vitro* [26], as well as monomeric GFP in cells [27]. However, the detection and quantification of dimeric species in living cells is a challenging task, and the accuracy of this global analysis, as for all FFS methods, depends on the method of application. We have optimized the combined FCS and PCH global analysis method to detect the state of dimerization of fluorescent molecules in living cells. This combined analysis will be referred here as brightness and diffusion global analysis (BDGA). We use *Dictyostelium* cells as a model system, expressing either monomeric GFP, diGFP, or FKBP12-GFP, a small protein that dimerizes upon ligand binding [28,29], and demonstrate that this BDGA methodology is a suitable approach to study the *in vivo* dimerization of GFP-tagged proteins.

Results and Discussion

Optimization of the BDGA Method in Order to Measure Monomer-Dimer Equilibria

In cells, dimeric proteins often exist in an equilibrium between monomers and dimers; however, resolving monomer-dimer mixtures using FFS methods is still a challenging task. Both species differ in mass and brightness by just a factor of two, and even when we applied the standard BDGA method, see details in the Methods section, it was not possible to clearly resolve a monomer-dimer mixture at the low signal-to-noise ratios typical for FFS measurements in cells. Therefore, we aimed to optimize BDGA for a multi-species application in cells.

To be able to quantify protein dimerization in cells, the method was first validated on data obtained from mixes of cell lysates expressing monomeric GFP and tandem-dimer GFP

(diGFP), see Materials section for details. In this way, various stable monomer-dimer equilibria were mimicked in a biological context. The measurements of these samples were performed at typical settings for standard FCS measurements, as described in the Methods section. The total averaged fluorescence intensity ranged from 40 kHz to 250 kHz, depending on the (fluorescent) protein concentration.

Initial Steps in the Analysis of FFS Data

For the data analysis, raw data sets of photon arrival times were converted into autocorrelation curves (ACF) and photon counting distributions (PCD) (**Figure 1**, top panels). In order to do so, a set of data was imported into the measurement database of the FFS data processor (FFS DP) and each measurement was split into traces of 5 s. A 5-s trace-length provides enough data points for validated samplings, while yielding a high enough number of traces in the global analysis. Per trace one ACF and three PCDs were generated. The binning time step of the ACF curves was set at 2×10^{-7} s, roughly 500 times shorter than the residence time of GFP and diGFP in the confocal volume, which was in the order of 100 μ s. The binning times of the PCD curves were set such that the histograms contained a minimum of five data points and the histogram peak value did not pass the 10-photon counts position. The first constraint is required because the PCH model has five parameters for an analysis with two components and the latter to maintain the super-Poissonian shape of the distribution (i.e., to avoid effect of averaging).

Next, all generated ACF and PCD curves of one measurement were imported in the analysis platform of FFS DP to perform the global analysis. The FCS 3D free diffusion model with triplet state term and brightness correction (Equation (1)) was used to fit ACFs, and the PCH model with out of focus (first order), diffusion and triplet correction (Equations (2)–(9)) was used to fit PCDs. A detector dead-time correction was included for the analysis of measurements obtained at high intensities. These equations can be used for one, two, or more component analyses. We started with a two-component analysis because of our goal to analyze samples that contain a mixture of two species.

Two-Component BDGA of Lysate Mixtures

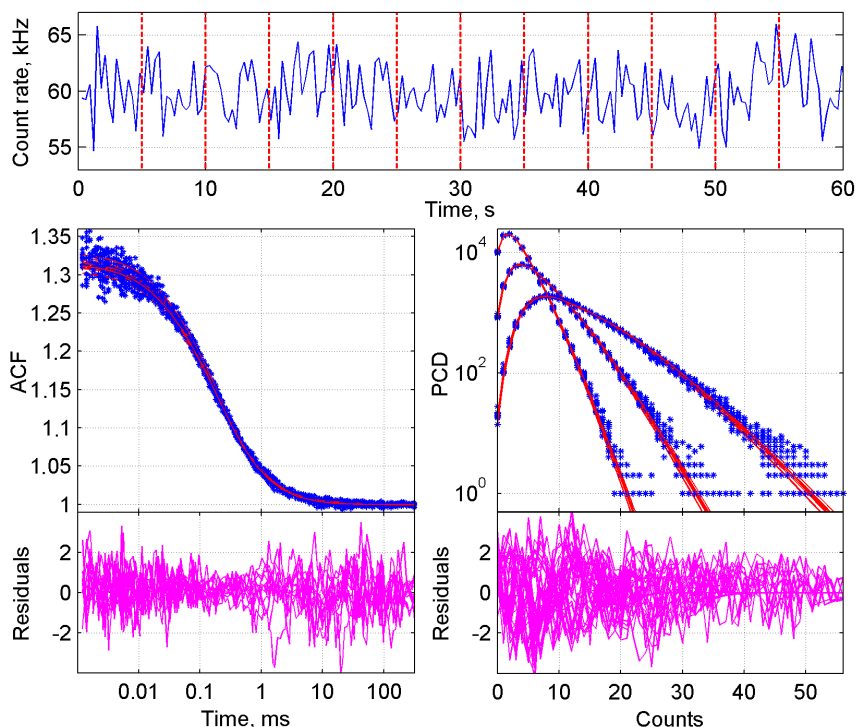


Figure 1. BDGA analysis of the ACF and PCD curves from a monomer-dimer ‘equilibrium’ in cell lysate. The sample consisted of 50% GFP and 50% diGFP, mimicking a stable equilibrium between two species. Top panel: raw FFS data showing photon counts over time, from which the ACF and PCD curves were calculated. The measurement was divided into twelve 5-s traces, as indicated by vertical lines. Bottom left: 1-component fit of all ACF curves, with residuals below. Bottom right: 1-component fit of all PCD curves, with residuals below. PCD curves were generated with three different time steps of 5×10^{-5} , 1×10^{-4} and 2×10^{-4} s, respectively.

In the simplified system of artificial monomer-dimer equilibria in cell lysates, we expected that the parameters of the two species could be resolved via a two-component analysis. Initially, the diffusion parameter $\tau_{diff\ 1}$, see description of model parameters in section 3.1 and **Table S2**, was grouped, as well as the triplet state parameters τ_{trip} and F_{trip} ($\tau_{diff\ 2}$ were not grouped to allow some freedom in their estimation). Reliable values for these parameters were retrieved from the separate FCS analyses and used to fix $\tau_{diff\ 1}$, τ_{trip} , and F_{trip} in the combined BDGA analysis. As a result, parameters N_1 and N_2 should indicate the amount of monomeric and dimeric particles in the confocal volume, and the brightness parameters q_1 and q_2 should differ by a factor of two. However, the values for N_1 and N_2 varied a lot between different traces within one measurement, while q_1 and q_2 yielded similar values, in-between the brightness of monomeric and dimeric controls (**Figure S3**), indicating that the two species were not resolved properly. In order to obtain good fits, we applied different approaches in the global analysis,

mostly including variation in the different parameters to be fixed or free. In the example of a lysate mix with 50% GFP and 50% diGFP, best results were obtained when diffusion parameters $\tau_{diff 1}$ and $\tau_{diff 2}$ were fixed to the values found for GFP ($\tau_{diff 1}$) and diGFP ($\tau_{diff 2}$), while q_1 was fixed to the brightness of GFP (q_{GFP}) and the q_2/q_1 ratio to $r = 1.8$ (Table 1, Figure S4). Alternative approaches, like fixing q_2 to the brightness found for diGFP, resulted in N_1 - and N_2 -values that deviated more from the expected percentages for GFP and diGFP (Table S3). However, the optimal value for r varied for measurements with other proportions of diGFP (Table S4), thus a generally applicable fixed value of r , which would yield the expected N -values under different conditions, could not be determined. Therefore, the two-component analysis is considered to not be a reliable method for analysis of our measurements of a mix of two components that vary only two-fold in mass and brightness. Moreover, this method is not suitable for unknown samples, since τ_{diff} and q were fixed to known values from the pure samples, and in biological situations it is difficult to obtain proteins that are completely monomer or completely dimer. We concluded that most probably the signal-to-noise ratio needed to resolve such mixtures into monomer and dimer was not reached.

Table 1. Analyses of representative lysate samples from one measurement day.

Each measurement consists of 12 traces of 5 s. Deviations of all parameters except N are presented as confidence intervals, calculated as asymptotic standard errors (ASE), as reported by the software. F_{trip} : triplet state fraction; τ_{trip} : triplet state time; τ_{diff} : diffusion parameter as reported by the software; N : number of particles in the confocal volume, with the standard deviation (SD) between traces indicated; q_{true} : true brightness; χ^2 : value of the global fit criterion. ASEs represent uncertainties of estimated parameters obtained in analyses performed per one trace. For comparison we additionally calculated standard deviations of brightness between traces (SD of q_{true}). Fixed values are indicated in italics (have been determined in the rows above). cpms: counts per molecule per second; $r = q_2/q_1$.

Sample	Analysis method	F_{trip} ($\times 10^{-2}$)	τ_{trip} (μ s)	$\tau_{diff 1}$ (μ s)	N (\pm SD)	q_{true} ($\times 10^4$ cpms)	SD of q_{true}	χ^2
R110	1-component	9.10 \pm 0.78	6.9 \pm 1.2	35 \pm 1	4.50 (\pm 0.06)	4.04 \pm 0.023	0.055	1.147
GFP	1-component	12.2 \pm 0.90	35.1 \pm 2.0	151 \pm 4	3.88 (\pm 0.09)	3.20 \pm 0.020	0.096	1.05
diGFP	1-component	7.08 \pm 0.43	31.0 \pm 2.0	221 \pm 3	3.63 (\pm 0.06)	5.12 \pm 0.023	0.090	1.233
50% GFP	1-component	8.97 \pm 0.47	22.2 \pm 2.5	191 \pm 2	3.47 (\pm 0.07)	4.88 \pm 0.025	0.107	1.193
+ 50% diGFP	2-component ($r = 1.8$)	8.07 \pm 0.46	19.1 \pm 2.4	<i>151 ($\tau_{diff 1}$);</i> <i>221 ($\tau_{diff 2}$)</i>	2.00 (\pm 0.18) (N_1); 1.76 (\pm 0.12) (N_2)	<i>3.20 (q_1);</i> <i>5.76 (q_2)</i>	-	1.068

Although the two-component model could not resolve two species which differ by only a factor of two in brightness and mass, in more complex samples the addition of a second component may very well be required. In our experience, adherence to cellular structures may cause part of the particles to diffuse very slowly (in the seconds time-scale), requiring the two-component model for a good fit. But for the lysate mixes, which do not have such a slow-diffusing component, we changed to a simpler variant of the BDGA method, namely the one-component analysis.

One-Component BDGA of Lysate Mixes

While in a two-component analysis the number of molecules (N_1 and N_2) reflects the proportion of each species, in a one-component analysis the parameter of interest is the average brightness q . This apparent brightness per particle increases with a higher dimer to monomer ratio. The BDGA analysis was performed as described above, except that a one-component model was applied. F_{trip} , τ_{trip} , and $\tau_{diff 1}$ (now called τ_{diff}) were first calculated in the FCS global analysis, and subsequently fixed in the BDGA analysis. An example of the fit of one of the analyses is shown in **Figure 1**. The brightness parameter q was determined per trace and corrected for correction factor F_{c_1} (Equation (12)) before further calculations were made. Data from multiple measurement days were normalized and averaged as described in the Methods section. Of note: the GFP lysate was used for normalization rather than the R110 calibration dye, because it corrects for day-to-day variation between the biological samples rather than only the technical setup.

Brightness per sample was plotted as a ratio compared to the brightness of GFP (**Figure 2** and **Table 2**). Despite some deviation due to technical fluctuation per day, a linear trend is observed with higher average brightness when more dimer is present (solid line). This is slightly lower than the theoretical value when diGFP would be twice as bright as the monomer (dashed line). This deviation is in accordance with previous reports, where a ratio of 1.7 was observed, possibly due to fluorescence energy transfer [26]. A proportion of dimer as small as 5% could be distinguished ($p < 0.001$). Therefore, one-component BDGA can be used to determine whether dimeric proteins are present in the sample and to estimate the monomer to dimer ratio.

Table 2. Diffusion rate and brightness of various GFP-diGFP mixes in cell lysate. Percentages of diGFP particles are indicated, relative to the total number of GFP and diGFP particles in the sample. The presented values are mean \pm standard deviation (SD) based on 42–126 traces from 2–5 measurement days. τ_{diff} : diffusion parameter before any normalization; D : diffusion coefficient calculated from τ_{diff} , corrected for the diffusion of R110 on each measurement day; q_{true} : average true brightness without any correction for instrumental variation between days; q_{true} norm. to GFP: true brightness normalized for the true brightness of the GFP sample per measurement day.

diGFP%	τ_{diff} (μ s)	D (μ m ² s ⁻¹)	q_{true} ($\times 10^4$ cps)	q_{true} norm. to GFP
0	131 \pm 21	97.3 \pm 6.5	4.43 \pm 1.22	1.00 \pm 0.02
5	126 \pm 32	98.3 \pm 6.6	5.13 \pm 1.19	1.02 \pm 0.03
10	154 \pm n.a.	99.0 \pm n.a.	5.16 \pm 1.60	1.10 \pm 0.03
12.5	132 \pm 37	94.1 \pm 3.3	5.19 \pm 1.15	1.13 \pm 0.03
33.3	137 \pm 32	87.7 \pm 3.9	6.43 \pm 1.42	1.37 \pm 0.04
50	169 \pm 31	74.0 \pm 6.0	5.69 \pm 1.62	1.45 \pm 0.06
66.7	168 \pm 37	78.4 \pm 0.1	6.20 \pm 1.84	1.55 \pm 0.25
87.5	156 \pm 37	76.8 \pm 2.2	8.22 \pm 2.02	1.75 \pm 0.08
90	175 \pm 46	69.5 \pm 6.9	8.27 \pm 2.88	1.84 \pm 0.25
95	158 \pm 36	75.7 \pm 2.4	8.68 \pm 1.70	1.74 \pm 0.07
100	173 \pm 40	71.4 \pm 5.9	8.82 \pm 2.77	1.89 \pm 0.24

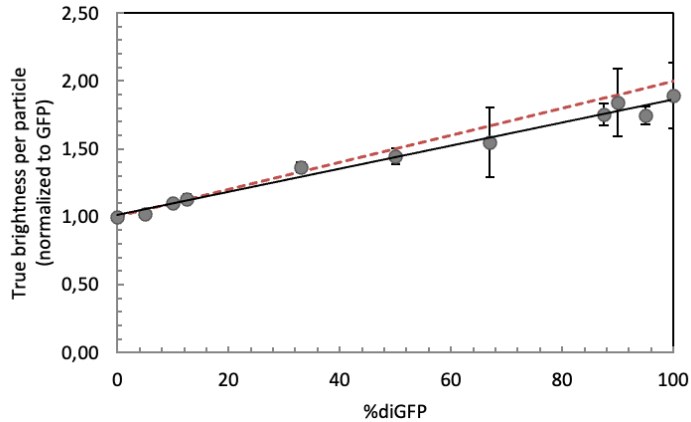


Figure 2. Average brightness of various GFP-diGFP mixes in cell lysate. Percentages of diGFP particles are relative to the total number of GFP and diGFP particles in the sample. Average true brightness is calculated as in **Table 2**, normalized to GFP per day. Error bars indicate standard deviations. The black line indicates the linear fit ($R^2 = 0.9837$). The dashed line indicates the theoretical expected values, when a dimer would be twice as bright as a monomer.

Improved BDGA Methodology for the Analysis of Cellular Data

Analysis of cellular data brings new challenges compared to data from cell lysates. Due to internal movement of non-fluorescent biomolecules or compartments in living cells, such as lysosomes or vacuoles, the fluorescence intensity in the small FCS spot will fluctuate over time (**Figure 3**, top panel). Therefore, each measurement was split into 3-s traces, and instable traces were excluded from further analysis based on their typical aberrant tail in the ACF. This approach has proven to be effective to analyze the non-stationary data [30]. Of note, by using 3- rather than 5-s traces (as for lysate data), less data had to be removed per aberrant trace, because a deviation lasted usually for about three seconds and the selection of fluctuating regions in the measurement became more precise.

Other challenges to overcome are phototoxicity and photobleaching. Focusing the laser beam on a single spot in a cell for a few minutes will cause local stress such as phototoxicity to the cells. Moreover, photobleaching of the fluorophore can play a role, because if a dimer of which one molecule is bleached it will behave as dimer in FCS and as monomer in PCH. In the small, confined volume of a cell, the bleached molecules are not diluted by diffusion as in lysates. Therefore, *Dictyostelium* cells were measured for only 45 s. All other steps in the ACF and PCD curve generation were the same as above. The resulting ACF as well as the PCD curves from a series of traces did overlay but deviated due to bleaching (**Figure 3**).

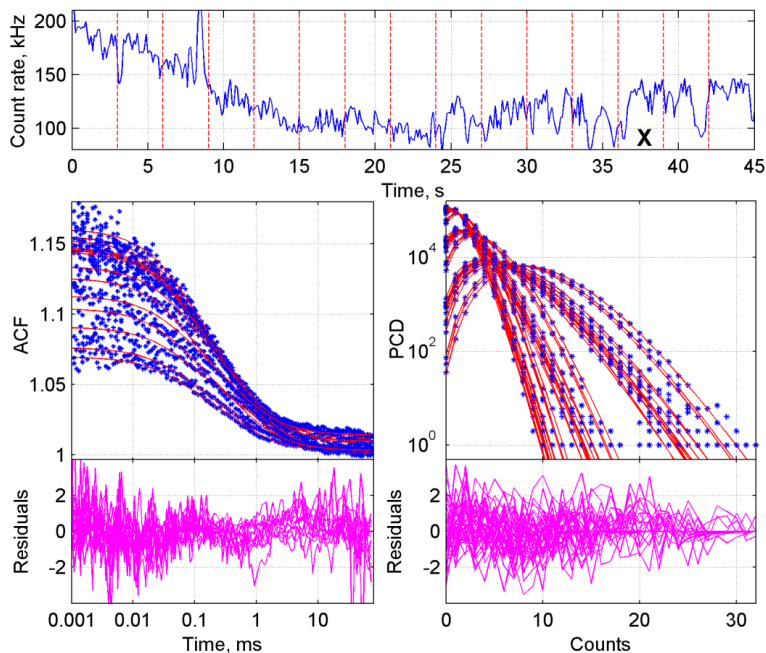


Figure 3. BDGA analysis of the ACF and PCD curves from GFP in living cells. The sample consisted of cells expressing GFP. Panels are explained in **Figure 1**. The measurement was divided into fifteen 3-s traces. Traces marked with an 'X' have been deleted based on visual inspection of the ACF curves. In the first 15 s, bleaching is taking place. The slight wave in the ACF residuals is expected because of internal movement of cellular compartments and fixation of the triplet state parameters. PCD curves were generated with three different time steps of 1×10^{-5} , 2×10^{-5} , and 5×10^{-5} s, respectively.

As with the lysate data, $\tau_{diff\ 1}$, τ_{trip} , and F_{trip} were first calculated via the global analysis of only ACF curves. Importantly, τ_{trip} and F_{trip} were the same for all samples with single GFP and GFP-fusion proteins, while diGFP had significantly different triplet state values (**Figure S5A,B**). Therefore, τ_{trip} and F_{trip} were averaged over a large number of cells, and all samples except diGFP were analyzed with these values fixed for τ_{trip} and F_{trip} in both the FCS and BDGA analyses. Fixation of triplet parameters improved the analysis, yielding lower standard deviations for the brightness parameters than when these parameters were left free (**Figure S5C**). In samples with complex proteins, the standard deviations will be higher and the reduction in standard deviation upon fixing the triplet state parameters is expected to be clearer. More details about the triplet state in FCS can be found in [31,32]. In general, fixing some parameters to the previously known values improved the estimation of any other parameters of interest.

Representative examples of the global analysis of individual cells are shown in **Table 3**, while the averages of multiple cells are shown in **Table 4** (two first rows). The fit per cell

yielded in many cases a χ^2 close to unity and the standard deviations of fit parameters were similar to what was obtained for measurements in cell lysate (**Table 1**). However, the variation between cells was much larger compared to the cell lysate measurements. Therefore, 27–35 cells from three measurement days were analyzed per experimental condition, normalized towards GFP and averaged for the parameters q and τ_{diff} (**Table 4**).

Unlike with cell lysate, in cells the additional recalculation of the apparent brightness to q_{true} was not performed. The required correction parameter F_{c_1} varied to a great extent and did not reflect the real instrumental correction factor due to ‘absorption’ of any deviations which are not accounted by the model and which are more prevalent in cells. Since all measurements were performed on the same instrument with similar settings, a similar F_{c_1} correction parameter was expected for the whole data set. Any deviations caused by different F_{c_1} should be averaged out because of the large sample size and linearity of the correction. Finally, the influence of the correction parameter F_{c_1} to the relative brightness should be canceled out due to the performed normalization. In order to validate the results, a pilot of seven cells per sample was analyzed using PCH with the polynomial model (instead of Gaussian), which does not include parameter F_{c_1} . Indeed, the differences in relative brightness were small (**Figure S6A**). Therefore, we are confident that the additional calculation of q_{true} was not required for the comparison of relative values of cellular samples as was performed on the lysate samples.

Table 3. Global analysis results of single cells. Representative cells have been chosen from the same measurement day. Each measurement of 45 s was divided into 3-s traces. Deviations of all parameters except N are presented as confidence intervals. F_{trip} , τ_{trip} , and τ_{diff} , see **Table 1**. N : number of particles in the confocal volume, lowest and highest values indicate the amount of bleaching; q : apparent brightness; # traces: number of traces from one measurement, on which the analysis was based; χ^2 : value of the global fit criterion. ASEs represent uncertainties of estimated parameters obtained in analyses performed per one trace. For comparison we additionally calculated standard deviations of brightness between traces (SD of q). +dim: 1 μ M dimerizer was added to the cells for 3 h. Trip free: triplet state parameters were free in the ACF analysis. Fixed values are indicated in italics and have been determined by averaging multiple analyses with free triplet state parameters. The two GFP results were based on the same measurement but with different analysis settings.

Sample	F_{trip}	τ_{trip} (μ s)	τ_{diff} (μ s)	N (range)	q ($\times 10^4$ cpms)	SD of q ($\times 10^4$ cpms)	# traces	χ^2
GFP (trip free)	0.241 \pm 0.008	53.8 \pm 2.8	381 \pm 8	12.1 (8.5–19.9)	4.71 \pm 0.06	0.55	14	1.197
GFP (trip fixed)	<i>0.178</i>	<i>40.0</i>	330 \pm 2	11 (7.8–18.2)	5.18 \pm 0.07	0.55	14	1.108
diGFP	<i>0.128</i>	<i>61.9</i>	841 \pm 6	6.5 (4.9–7.9)	8.32 \pm 0.10	1.04	13	1.079
FKBP12-GFP + dim	<i>0.178</i>	<i>40.0</i>	1018 \pm 9	19 (16.5–21.7)	6.73 \pm 0.18	0.79	9	1.015

The results of the analysis of GFP and diGFP in cells are shown in **Table 4** (upper rows). The diffusion of both GFP and diGFP was slower in cells and had a broader distribution compared to cell lysate (**Table 2**). This broad deviation in diffusion may be due to variation in viscosity, depending on the position in the cell, which could not be distinguished in our setup. Still, the diffusion of GFP and diGFP was significantly different in the separate samples ($p < 0.001$; **Table 4**), as determined in the initial analysis step with FCS without PCH.

Considering the significant difference in diffusion between GFP and diGFP, the question may arise why FCS should not be used. However, if a protein with unknown dimerization property is observed, the amount of dimer cannot be estimated because a reference with pure monomer of the same protein is often lacking. Moreover, the diffusion of proteins in cells is often influenced by adhesion to cellular structures like membranes and protein complexes. Nevertheless, FCS may be used to estimate the protein concentration N , from which the average brightness per particle may be calculated by dividing the total fluorescence intensity by N ($q = \langle I \rangle / N$). In complex samples which are difficult to fit, this may be a fast way to assess whether higher order oligomers are present in the cells, since this average brightness can be compared to the brightness of monomeric GFP in cells. This simplification comes at the expense of any corrections for background fluorescence, protein aggregates and other artifacts.

With the brightness and diffusion global analysis as performed here, the brightness of GFP and diGFP could clearly be distinguished with a significance of $p < 0.001$ (**Table 4** and **Figure 4**, first rows/bars). The magnitude of this difference is 1.6-fold, which is comparable to what was found before with purified protein (1.72; [26]), but smaller than the two-fold difference which is expected when two fluorophores are present per particle. The low apparent brightness of the dimer may be caused by resonance energy transfer, triplet state formation, photobleaching, or blinking of one of the subunits [33–35]. In the small confined volume of cells, the effect of photobleaching is substantial, which can be avoided by adjusting the laser power or it may be reduced by using short measurement times (<1 min) as well as more photostable fluorophores or two-photon excitation [15]. However, we strived for a simple, broadly applicable method which could be performed with any existing GFP construct on any confocal microscope with FCS module. Using diGFP as a reference for the brightness of dimeric samples, it will be possible to estimate the amount of dimer in the sample. Taken together, with the above data we have shown that in our setup the BDGA method can distinguish GFP from diGFP in *Dictyostelium* cells.

Application of the Developed Methodology to the Analysis of Induced Dimerization in Cells

To mimic real dimerization and create a monomer-dimer mix in living cells, GFP was fused to the FKBP12 dimerization domain [29,36]. Two of these domains can dimerize via a linker molecule called B/B homodimerizer (or 'dimerizer', **Figure S7A**). Measurement and analysis of *Dictyostelium* cells with FKBP12-GFP were performed as described in the Methods section. The ACF and PCD curves of FKBP12-GFP appear similar to those of GFP in cells (**Figure 5** compared to **Figure 3**).

When dimerizer was added to the cells, the apparent brightness of the FKBP12-GFP molecules was 1.33-fold higher than GFP, which was significantly higher compared to FKBP12-GFP without dimerizer or the GFP controls ($p < 0.001$), indicating that dimerization of FKBP12-GFP was taking place. Considering the linear relationship between average brightness and the monomer-dimer ratio (**Figure 2**), and the fact that 100% diGFP was 1.6-fold brighter than GFP (**Table 4**), a brightness of $1.33 \times q_{\text{GFP}}$ would correspond to approximately 55% dimer. Taken together, the amount of dimeric protein could be estimated in a cellular context using monomeric and dimeric GFP as a reference.

The difference in diffusion between FKBP12-GFP and 'FKBP12-GFP plus dimerizer' is smaller than the difference between GFP and diGFP, but due to a large sample size this difference is still significant ($p < 0.001$). Thus, the presence of FKBP12 dimer can already be distinguished by analyzing diffusion alone. However, as mentioned above, 'clean' diffusion (without adherence to other cellular components) and the availability of a 100% monomeric sample, are required to determine whether a sample would contain oligomers based on diffusion.

Taken together, we show that BDGA is an adequate analysis procedure, even in an in cellulo situation, to quantify the dimerization status of cytosolic proteins with statistical significance. To our knowledge, this is the first time that this method is applied to dimerizing proteins in living cells.

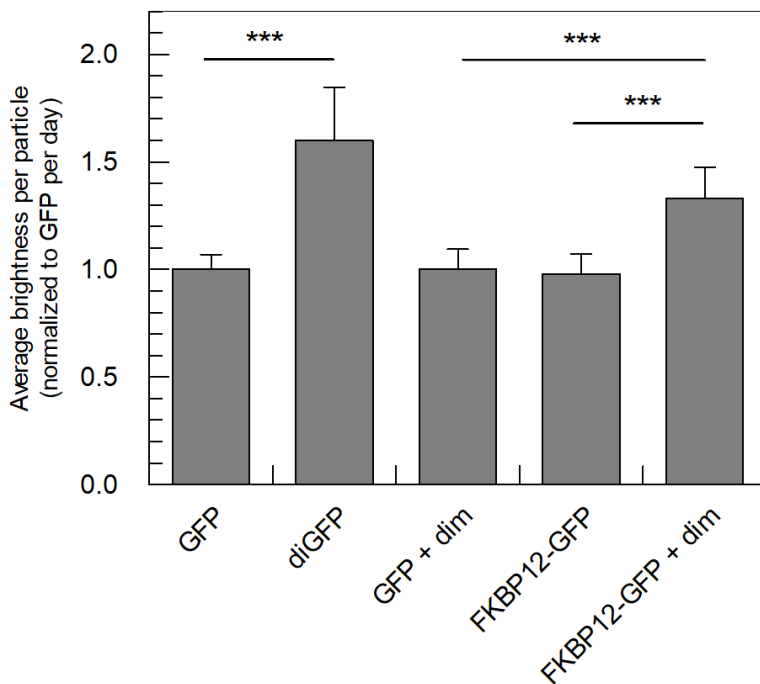


Figure 4. Average brightness calculated from analysis of multiple cells. The apparent brightness of several proteins was measured in living *Dictyostelium* cells and normalized to GFP per day. Averages were calculated from 24–35 cells per sample from 3 independent measurement days, as in **Table 4**. Error bars indicate standard deviations. FKBP12-GFP: GFP-linked dimerizing domain; +dim: cells have been incubated with 1 μM dimerizer for 3 h; *** $p < 0.001$.

Table 4. Induced dimerization of GFP in cells via the FKBP12 domain. Addition of dimerizer increased the average brightness per particle when GFP was fused to the FKBP12 dimerizing domain, but not to GFP alone. The presented values are mean \pm standard deviation (SD) based on 27–35 cells from 3 measurement days. Parameters are explained in **Table 2** and **3**. The diffusion coefficient D is calculated from τ_{diff} and the diffusion rate of R110 on each measurement day. q : average apparent brightness without any correction for variation between days; q norm. to GFP: brightness normalized for the brightness of the GFP sample per measurement day. +dim: 1 μM dimerizer was added to the cells for 3 h. * $p < 0.001$ relative to GFP; ** $p < 0.001$ relative to GFP and FKBP12-GFP.

Sample	τ_{diff} (μs)	D ($\mu\text{m}^2 \text{s}^{-1}$)	q ($\times 10^4$ cps)	q normalized to GFP	# cells
GFP	506 \pm 115	16.7 \pm 3.6	4.9 \pm 0.6	1.00 \pm 0.07	35
diGFP	871 \pm 336	10.2 \pm 2.4 *	7.8 \pm 1.4	1.60 \pm 0.25 *	27
GFP + dim	502 \pm 133	17.1 \pm 4.3	4.9 \pm 0.6	1.00 \pm 0.09	28
FKBP12-GFP	882 \pm 269	9.8 \pm 2.2	4.8 \pm 0.7	0.98 \pm 0.09	31
FKBP12-GFP + dim	1101 \pm 336	7.8 \pm 1.7 **	6.6 \pm 0.9	1.33 \pm 0.15 **	35

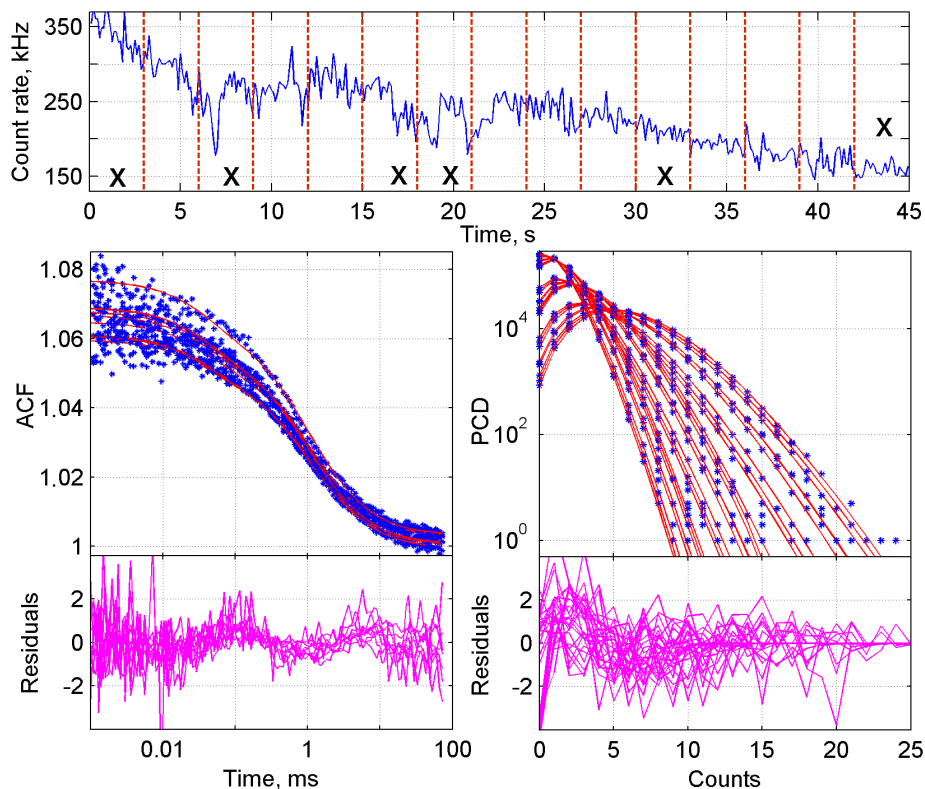


Figure 5. Results of BDGA analysis of the ACF and PCD curves from a monomer-dimer equilibrium of FKBP12 in living cells. The sample consisted of cells expressing FKBP12-GFP, supplemented with dimerizer. Panels are explained in **Figure 1**. The measurement was divided into fifteen 3-s traces. Traces marked with an 'X' have been deleted based on visual inspection of the ACF curves. PCD curves were generated with three different time steps of 5×10^{-6} , 1×10^{-5} , and 2×10^{-5} s, respectively.

Materials and Methods

Theory of the Global Analysis of ACF and PCD

The set of measured autocorrelation functions (ACF) and photon counting distributions (PCD) is analyzed globally using a combination of two models, FCS and PCH, by linking parameters having the same meaning (and name) in both models [26]. The analysis is performed using the constrained nonlinear iterative least-squares method with the Levenberg–Marquardt optimization [37]. Application of a global analysis approach increases the information content available from a single measurement that results in more accurate values of molecular diffusion coefficients and triplet-state parameters, and also in robust, time-independent estimates of molecular brightness and number of molecules.

The FCS model used for fitting the autocorrelation function describes a number of independent molecular species, which diffuse freely in a 3D Gaussian-shaped observation volume and undergo the triplet process, and is written as [38]:

$$G(t) = G_{\text{inf}} + \frac{1 - F_{\text{trip}} + F_{\text{trip}} e^{-t/\tau_{\text{trip}}}}{(1 - F_{\text{trip}}) \left(\sum_i q_{0\text{eff}i} N_{0\text{eff}i} \right)^2} \times \sum_i \frac{q_{0\text{eff}i}^2 N_{0\text{eff}i}}{\left(1 + t/\tau_{\text{diff}i}\right) \sqrt{\left(1 + t/a^2 \tau_{\text{diff}i}\right)}}, \quad (1)$$

where $G_{\text{inf}} = G_{\text{corr}}(\infty)$; F_{trip} and τ_{trip} are, respectively, the fraction and the relaxation time of molecules in the triplet state; $q_{0\text{eff}i}$ is the apparent molecular brightness of species i , expressed in counts per molecule per second (cpms); $N_{0\text{eff}i}$ is the number of molecules of species i in the effective volume V_{eff} ; $a = W_z/W_{xy}$, ω_{xy} and ω_z are, respectively, the lateral and axial radii of the confocal detection volume; and $\tau_{\text{diff}i}$ is the lateral diffusion time of species i , which is related to the diffusion coefficient D via $\tau_{\text{diff}} = \omega_{xy}^2/(4D)$. The effective volume is calculated as $V_{\text{eff}} = C_1^2/C_2$, where $\chi_k = \int_V B^k(\mathbf{r}) dV$, and $B(\mathbf{r})$ is the brightness profile function, which is the convolution of excitation intensity and detection efficiency profiles, depending on the radius r of the overlapping excitation and detection spots. The subscript 0 in $q_{0\text{eff}i}$ and $N_{0\text{eff}i}$ means that these parameters do not depend on time. Equation (1) has been written in assumption that each molecular species has the same triplet-state characteristics.

The PCH model with triplet, free 3D diffusion and out of focus emission corrections is calculated by a numerical algorithm consisting of the following steps (the full protocol is described in details in [38]):

1. Calculate the time dependent parameters
- 2.

$$q_{\text{eff}i}(T) = q_{0\text{eff}i} B_2(T), \quad N_{\text{eff}i}(T) = N_{0\text{eff}i} / B_2(T) \quad (2)$$

for each molecular species $i = 1, 2, \dots$, where T is the counting time interval (bin time), $B_2(T)$ is the binning correction factor

$$B_2(T) = \frac{2}{T^2} \int_0^T (T-t) g(t) dt \quad (3)$$

calculated over a time dependent term of the autocorrelation function in FCS

$$g(t) = \frac{1 - F_{\text{trip}} + F_{\text{trip}} e^{-t/\tau_{\text{trip}}}}{\left(1 - F_{\text{trip}}\right) \left(1 + t/\tau_{\text{diff}i}\right) \sqrt{\left(1 + t/a^2 \tau_{\text{diff}i}\right)}} \quad (4)$$

and $q_{0eff\ i}$, $N_{0eff\ i}$, a , F_{trip} , τ_{trip} , $\tau_{diff\ i}$ are fit parameters;

3. Calculate single-molecular PCD $p^{(1)}(k, q_{eff\ i})$ for each molecular species i

$$p^{(1)}(0, q_{eff\ i}) = 1 - \sum_{k=1}^K p^{(1)}(k, q_{eff\ i}),$$

$$p^{(1)}(k, q_{eff\ i}) = \frac{1 + Fc_2}{(1 + Fc_1)^2} \left[p_{3DG}^{(1)}(k, q_{eff\ i}) + \frac{1}{k! \Theta} \sum_{n=k}^{\infty} \frac{(-1)^{n-k} (q_{eff\ i} T)^n Fc_n}{(n-k)! (2n)^{3/2}} \right],$$
(5)

where $k = 1, 2, \dots, K$ is the number of photons detected in an interval T , K is the maximal number of photons, Fc_n , $n = 1, 2$ are instrumental out-of-focus correction parameters (Fc_n are also fit parameters) and

$$p_{3DG}^{(1)}(k, q_{eff\ i}) = \frac{1}{\Theta \sqrt{\pi k}} \int_0^{\infty} \gamma(k, q_{eff\ i} T e^{-x^2}) dx,$$
(6)

In Equation (6), $\gamma()$ is the incomplete gamma function and parameter Θ is varied depending on the value of the product of $q_{eff} T$ (from 1 to 20), see details in [38];

4. Calculate PCD $P(k)$ for each brightness component assuming the Poissonian distribution of a number of molecules in an open observation volume

$$P(k, q_{eff\ i}, N_{eff\ i}) = \sum_{M=0}^{\infty} p^{(M)}(k, q_{eff\ i}) Poi(M, QN_{eff\ i}),$$
(7)

where $p^{(M)}(k, q_{eff\ i}) = \underbrace{p^{(1)}(k, q_{eff\ i}) \otimes \dots \otimes p^{(1)}(k, q_{eff\ i})}_{M \text{ times}}$ is M -times convolution of the single-

molecule PCD and $Poi(k, \eta)$ denotes the Poisson distribution with the mean value η ;

5. Calculate the total PCD for a molecular system. PCD of a number of independent species is given by a convolution of PCD of each species

$$P(k) = P(k, N_{eff\ 1}, q_{eff\ 1}) \otimes \dots \otimes P(k, N_{eff\ n}, q_{eff\ n}),$$
(8)

6. The correction on dead-time is performed accordingly to the following equation [39]:

$$P_{DT\ corr}(k) = \sum_{j=0}^{\infty} P_0(k+j) P_{binomial} \left(j; k+j, \frac{(k+j)\tau_{dt}}{T + (k+j)\tau_{dt}} \right),$$
(9)

where τ_{dt} is the detector dead time (fit parameter) and $P_{binomial}(j, n, p)$ is the binomial probability distribution. We omitted the dependence on T from $q_{eff\ i}$, $N_{eff\ i}$, and therefore

from all related expressions in the Equations (5)–(9) for the sake of simplicity.

Both FCS and PCH models can be modified to fit the brightness ratio $r_i = q_{0eff i} / q_{0eff 1}$ instead of fitting of the absolute brightness values, thus, to have the following set of fitting parameters: $q_{0eff 1}, r_2, r_3, \dots$. The ratio can be then fixed to the expected value or constrained to reasonable boundaries, which increases the ability of the model to resolve multicomponent samples. The FCS model will take the form:

$$G(t) = G_{inf} + \frac{1 - F_{trip} + F_{trip} e^{-t/\tau_{trip}}}{(1 - F_{trip}) \left(\sum_i r_i N_{0eff i} \right)^2} \times \sum_i \frac{r_i^2 N_{0eff i}}{(1 + t/\tau_{diff i}) \sqrt{(1 + t/a^2 \tau_{diff i})}} \quad (10)$$

In the PCH model, one has to recalculate the time dependent parameters $q_{eff i}(T)$ accordingly:

$$q_{eff i}(T) = r_i q_{0eff 1} B_2(T) \quad (11)$$

For the sake of simplicity, q will be used instead of q_{0eff} , N instead of N_{0eff} , and r instead of r_2 throughout the text.

In the described method the FCS and PCH models are linked through the common parameter N . Values of N in FCS are corrected by values of q from PCH, and values of q and N in PCH are corrected by values of diffusion and triplet state parameters from FCS. When using the two-component model to resolve two species in the sample, it is important that the two fractions are separated in the same way for each trace in the global analysis of multiple traces, especially at low signal-to-noise ratios and/or when q and N values are close for the two species. To prevent the components to be exchanged in some individual traces, e.g., q_1 of that trace is low and q_2 is high while for the rest of the traces it is the other way around, the allowed range of each component should either be linked across all traces or even better, be tightly constrained or fixed. The swapping of components during the fit process (and therefore trapping in the local minima) is probable, even inside a global analysis of one ACF and one PCD in the case when brightness and diffusion are close to each other, which is typical for the dimerization studies. If necessary, the swapped values can be manually corrected and the analysis should then be repeated. The use of the model with brightness ratio fixed to the expected value may drastically improve the resolvability of the method in this case.

Sample Preparation and Measurement

Plasmids and Cloning

Vectors pDM317 and pDM334 [40] were used for constitutive or inducible expression of GFP-S65T (further referred to as GFP), respectively. For the construction of tandem-dimer GFP (diGFP), an additional GFP gene was added to these plasmids by digesting GFP from pDM313 with SpeI/XbaI and ligating into the SpeI site of pDM317 or pDM334. This created the following linker: 5'-SGLRSSTSS-3' between the two GFP inserts.

FKBP12 was cloned into pDM334 using BglII/SpeI restriction sites. To improve binding of the dimerizing ligand, an F36V point mutation was created by PCR-based site directed mutagenesis using Phusion (ThermoScientific) polymerase and primer: 5'-GATGGAAAGAAAGTTGATTCCTCCC resulting in FKBP12 F36V (hereafter referred to as FKBP12).

Cell Culture

The axenic *Dictyostelium discoideum* wild-type strain was used for all experiments. Cells were grown in HL5-C medium (Formedium) at 22 °C. The indicated constructs were transformed in AX2 cells by electroporation and selected with either 10 µg/mL geneticin or 50 µg/mL hygromycin B. To induce expression from the tetracycline inducible vectors, the cells were overnight incubated with 1 µg/mL doxycycline. Dimerization of FKBP12-GFP was induced by incubation with B/B homodimerizer (Clontech, also called AP20187) to a final concentration of 1 µM for 3 h.

Measurement in Cell Lysate

Vegetative *Dictyostelium* cells containing constitutive expression vectors (pDM317 hyg GFP or pDM317 diGFP) were put on LoFlo medium (Formedium) overnight, 10^8 cells were pelleted, washed in 10 mM Na-K-phosphate buffer (pH = 6.5). Cells were lysed by incubation in 1 mL lysis buffer (50 mM Tris (pH 7.5), 50 mM NaCl, 5 mM DTT, 5 mM MgCl₂, and 1% of a modified *Dictyostelium* protease inhibitor mix consisting of 2 µg/mL pepstatin (Carl Roth), 100 µg/mL N-tosyl-L-lysine chloromethyl ketone (Sigma), 80 µg/mL N-p-tosyl-L-arginine-methyl ester hcl (Sigma), 5 µg/mL leupeptin (Carl Roth), 0.1 mM PMSF (Carl Roth), 5 mM benzamidine (Sigma), 2 mM N-CBZ-Pro-ALA (Sigma) [41]) containing 1% triton for 30 min on ice.

Lysate was cleared from cell debris by centrifugation (10 min, 20,800× *g*, 4 °C). The protein concentrations were in the range of 5–30 mg/mL, as determined with the Bradford assay. The lysates were diluted 5–20 fold to a concentration corresponding to 3–5 molecules detected in the confocal spot (*N*) accordingly to the estimates reported by the ZEN acquisition software (Zeiss, Jena, Germany). To prepare solutions with the same *N*, the dilution factor was determined by measuring both the GFP and diGFP

lysates. Subsequently, the lysates were mixed according to the percentage of diGFP as indicated. Lysates were kept on ice for a maximum of two hours.

All lysate samples were measured on a LSM710 ConfoCor 3 microscope (Zeiss, Jena, Germany), supplied with a 488 nm solid state laser, a triple dichroic (488/543/633) excitation filter, NFT 635 VIS dichroic mirror, BP 505-610 IR emission filter, Zeiss C-Apochromat 40x/1.2 NA water objective with coverslip thickness correction collar, and APD detector. The pinhole was set at 1 AU, and the laser power at 0.5%, as reported by the Zen software. Samples with a minimum volume of 100 μL were measured at 20 °C, in an 8-chambered coverslip (μ -Slide 8 Well Glass Bottom, Ibidi, Gräfelfing, Germany). Measurements were performed 30 μm above the glass surface, for 120–180 s in 1–3 spots per well. Rhodamine 110 (R110, $D = 4.3 \times 10^{-10} \text{ m}^{-2} \text{ s}^{-1}$ (Invitrogen, Breda, The Netherlands)) in water was used for calibration measurements.

Measurement of Cells

Dictyostelium cells containing inducible expression vectors (pDM334hyg GFP, pDM334 diGFP, or pDM334 FKBP12 F36V) were starved for 3 h in Na-K-phosphate buffer (10 mM, pH = 6.5) at a cell density of 10^7 cells per mL. DMSO was added to a final concentration of 2% v/v to impair cell movement, after which the cells were immediately transferred to 8-chambered coverslips (μ -Slide 8 Well Glass Bottom, Ibidi, Gräfelfing, Germany) and were allowed to settle down for 3–5 min. All cellular samples were measured on a TCS SP8 X SMD system (Leica Microsystems, Wetzlar, Germany), supplied with a super continuum laser (emitting a continuous spectrum from 470 to 670 nm), bandpass-adjustable spectral filters, a 63x 1.20 NA water immersion objective with coverslip thickness correction collar, and a HyD internal hybrid detector, coupled to a PicoHarp 300 TCSPC module (PicoQuant, Berlin, Germany). The pinhole was set at 80 μm , the laser line at 488 nm with a pulsed frequency of 40 MHz, and the spectral filter at 495–545 nm. Raw intensity fluctuation data consisting of about 10^7 photons were collected at 20 °C from single measurements. Generally, data were obtained from one measurement of 45 s per cell in one spot, in 10–20 cells per sample. Spots were selected in cells with low brightness, in areas with homogeneous distribution of fluorescence rather than nuclear area or large intracellular vesicles (**Figure S1**).

Data Analysis Procedure

Fitting Software

The FCS and global FCS and PCH analyses were performed using the FFS data processor 2.6 software (SSTC, Department of Systems Analysis and Computer Modelling, Belarussian State University, Minsk, Belarus, www.sstcenter.com). This software offers the complete set of tools for the global analysis of FFS data. First the raw data, which is the sequence of photon arrival times stored in a binary form,

are loaded in the program. This is followed by calculating a set of ACFs and PCDs, performing the non-linear constrained least-squares minimization, estimating of confidential intervals of fit parameters, displaying the analysis results in a user-friendly form, and finally storing both measured and analyzed data in databases.

The software has a powerful algorithm to process raw data by splitting the measured data into parts and allowing for fully automated calculation of ACFs and PCDs at user defined binning times from each part. Weight factors for ACF are calculated as follows: the measurement (or part of the measurement) is split into a number of sub-parts, ACF is calculated from each sub-part and standard deviations of each point of resulted ACFs are calculated in a usual way. Weight factors for PCDs are calculated assuming the binomial distribution. The length of a PCD (a range of photon counts) is extended automatically by a factor of 2 each time the software counts more photons than allocated for the last channel.

The non-linear constrained least-squares analysis is performed using the Levenberg–Marquardt optimization and reduced χ^2 criterion. The fit parameters can be either constrained or fixed to the expected values. Confidential intervals are estimated using asymptotic standard errors (ASE) approach [42].

Analysis of the Obtained Data

The obtained raw data were imported into the measurement database of the FFS data processor. Each measurement was divided into traces of 3–5 s, and for each trace one ACF and three PCDs were calculated. Each trace was subdivided by 10 for the ACF weight factors calculation. The ACF was calculated with 140 points and a time step of 1×10^{-7} s (cells) or 2×10^{-7} s (lysate). The lower binning times (1×10^{-7} s) and number of points in ACF (in comparison with conventionally used) were selected to ensure enough time length for the calculation of weights factors for ACF because of splitting the raw data in relatively short time traces. PCDs were calculated with 32 points as initial value and three different binning times, usually at 2×10^{-5} , 5×10^{-5} and 1×10^{-4} s, depending on the total fluorescence intensity.

Next, the ACF curves from all traces of one measurement were analyzed globally using a free diffusion 3D Gaussian model (Equation (1)), with the parameters specified in **Table S1**. Parameters a , F_{trip} , τ_{trip} , and τ_{diff} were grouped between traces, and a was fixed to the value found for the R110 calibration dye. Confidence intervals (CI) with standard errors were calculated for F_{trip} , τ_{trip} and τ_{diff} . In case of cellular data, the outcome of F_{trip} and τ_{trip} were averaged for all GFP (and diGFP) control samples of a single measurement day. These averaged values were fixed in the subsequent analyses of the experiment.

Next, the PCD and ACF curves of all traces of one measurement were imported in a new worksheet and collectively analyzed using the BDGA method as described above (Equations (1)–(9)). For lysate data, parameter a was fixed to the value found in the control samples, while F_{trip} , τ_{trip} , and τ_{diff} were fixed to the values found in the previous analyses of the ACF curves of that particular sample. For cells, F_{trip} and τ_{trip} were also based on the control samples. The background parameter bg was fixed to 0 because, for the first-order out-of-focus correction, the parameters bg and $Fc1$ are mutually correlated (Skakun et al., 2015) and one of them must be fixed, G_{inf} was free, and Fc_1 and τ_{dt} were free but grouped for all traces together. The parameters N and q were grouped per trace, i.e., per one ACF and three PCD curves (see **Figure S2** for an illustration of the grouping). An example of the analysis output in the FFS data processor can be found in **Table S2**.

All parameter estimates were imported into Microsoft Excel for the calculation of averages, standard deviations, statistics, and performing normalizations. In the case of lysate data, the apparent brightness q obtained from global analysis was recalculated into the true brightness q_{true} by [43], using

$$q_{true} = q_{0eff} / (1 + Fc_1) \quad (12)$$

Each lysate sample was measured once or twice per experiment day and normalized towards GFP of that day. Because of the complex composition of the data (multiple experiment days, with variable multiple measurements per sample consisting of multiple traces), multiple options were explored to determine the best way to perform the statistics. Since measurements were performed during just a few days, and the number of measurements per day varied, data traces of different days were combined before averages and standard deviations were calculated.

In the case of cellular data, averages were first calculated from the traces per cell, after which the averages of all cells (normalized towards GFP per day) from several days were calculated. Therefore, standard deviations here also account for the variation between cells. The standard t-test (unequal variances) was applied each time when the question of significance of the difference between obtained values was studied.

Conclusions

Protein dimerization is abundant in nature [44] but studying the dimerization status of proteins in living cells is a challenge. In this paper we present a methodology to measure the oligomeric state of proteins using a standard confocal microscope including an FCS module. Diffusion and brightness information were combined using the BDGA method, exploiting all available information from the FCS data. We have developed a fully optimized analysis protocol by which monomeric and dimeric particles can be distinguished in cell extracts and in cells.

We applied this methodology to analyze dimerization in living cells, as well as a mixture of GFP and diGFP *in vitro*. Compared to our previous global ACF and PCH experiments [26,27], changes were made regarding preparation of the samples, data acquisition, ACF and PCD curve calculation from raw data, global analysis of ACF and PCD curves, and further data processing. The changes regarding cellular sample preparation and data acquisition entailed (1) using starved rather than vegetative *Dictyostelium* cells, (2) adding 2% DMSO to the cells before the measurement, (3) selecting dim cells, (4) applying low laser power, (5) measuring for only 45 s per cell, and (6) using a sample for only 30 min. Regarding the raw data processing, measurements were separated into 3-s traces, from which ACF and PCD curves were then calculated as before.

Regarding the global analysis of ACF and PCD curves (BDGA), the main adaptations entailed (1) discarding aberrant traces, (2) fixing parameters a , F_{trip} , τ_{trip} , and τ_{diff} (determined from only the ACF curves), rather than setting their values as initial guess, (3) grouping N and q per trace, rather than for all traces together, and (4) fixing F_{trip} and τ_{trip} for the cellular measurements with values as determined from the GFP sample. When processing the results, (1) the GFP sample rather than the R110 dye was used for normalization between measurement days, and (2) the apparent brightness rather than the true brightness (corrected for F_c) was used for comparisons between cells. We have demonstrated that the ratio of monomer and dimer can be estimated with our methodology, by mixing the content of cells containing monomeric GFP and dimeric GFP. We have explained how we developed and optimized this method, and therefore several variants of the analysis as well as considerations regarding experimental setup and analysis were discussed. Next, we applied this method to distinguish GFP and diGFP in living cells, using *Dictyostelium discoideum* as a model organism. Finally, using the inducible dimerization domain FKBP12 [28,29], we have demonstrated that our BDGA method is well applicable to the *in cellulo* dimerization of GFP-tagged proteins. In our setup, dimerization of proteins with a monomeric GFP tag could be measured on a standard single-color, single-photon FCS microscope. Therefore, we consider this method to be very suitable for examining the dimerization status of any GFP-labeled

cytosolic protein in living cells. Moreover, the presented methodology is in principle independent of the means of labeling and could therefore be applied to any fluorescent protein or synthetic dye. Besides cytosolic proteins, BDGA may also be applied to extracellular protein dimerization with no change in the used models. For non-mobile cells, regions at the cell surface may also be selected to study the oligomerization of membrane proteins, which should then be analyzed using 2D models rather than 3D (e.g., triplet-state with 2D diffusion model and the 2D Gaussian model for PSF profile). With the brightness and diffusion global analysis of FFS data, the abundant role of dimerization in nature can be further explored.

Funding

This research was supported by a NWO VIDI, grant number 723.012.108 to AK

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Khokhlatchev, A.V.; Canagarajah, B.; Wilsbacher, J.; Robinson, M.; Atkinson, M.; Goldsmith, E.; Cobb, M.H. Phosphorylation of the MAP kinase ERK2 promotes its homodimerization and nuclear translocation. *Cell* **1998**, *93*, 605–615, doi:10.1016/S0092-8674(00)81189-7.
2. Brummer, T.; McInnes, C. RAF kinase dimerization: Implications for drug discovery and clinical outcomes. *Oncogene* **2020**, *39*, 4155–4169.
3. Rajakulendran, T.; Sahmi, M.; Lefrançois, M.; Sicheri, F.; Therrien, M. A dimerization-dependent mechanism drives RAF catalytic activation. *Nature* **2009**, *461*, 542–545, doi:10.1038/nature08314.
4. Yaşar, P.; Ayaz, G.; User, S.D.; Güpür, G.; Muyan, M. Molecular mechanism of estrogen–estrogen receptor signaling. *Reprod. Med. Biol.* **2017**, *16*, 4–20.
5. Levin, E.R.; Hammes, S.R. Nuclear receptors outside the nucleus: Extranuclear signalling by steroid receptors. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 783–797.
6. Shizu, R.; Yokobori, K.; Perera, L.; Pedersen, L.; Negishi, M. Ligand induced dissociation of the AR homodimer precedes AR monomer translocation to the nucleus. *Sci. Rep.* **2019**, *9*, doi:10.1038/s41598-019-53139-9.
7. Jiménez-Panizo, A.; Pérez, P.; Rojas, A.M.; Fuentes-Prior, P.; Estébanez-Perpiñá, E. Non-canonical dimerization of the androgen receptor and other nuclear receptors: Implications for human disease. *Endocr. Relat. Cancer* **2019**, *26*, R479–R497.
8. Mahajan, N.P.; Linder, K.; Berry, G.; Gordon, G.W.; Heim, R.; Herman, B. Bcl-2 and bax interactions in mitochondria probed with green fluorescent protein and fluorescence resonance energy transfer. *Nat. Biotechnol.* **1998**, *16*, 547–552, doi:10.1038/nbt0698-547.
9. Bacia, K.; Kim, S.A.; Schwille, P. Fluorescence cross-correlation spectroscopy in living cells. *Nat. Methods* **2006**, *3*, 83–89, doi:10.1038/nmeth822.
10. Jameson, D.M.; Ross, J.A.; Albanesi, J.P. Fluorescence fluctuation spectroscopy: Ushering in a new age of enlightenment for cellular dynamics. *Biophys. Rev.* **2009**, *1*, 105–118, doi:10.1007/s12551-009-0013-8.
11. Kitamura, A.; Kinjo, M. State-of-the-art fluorescence fluctuation-based spectroscopic techniques for the study of protein aggregation. *Int. J. Mol. Sci.* **2018**, *19*, 964.
12. Wu, B.; Singer, R.H.; Mueller, J.D. Time-integrated fluorescence cumulant analysis and its application in living cells. In *Methods in Enzymology*; Academic Press Inc.: Cambridge, MA, USA, 2013; Volume 518, pp. 99–119, ISBN 9780123884220.
13. Nagy, P.; Claus, J.; Jovin, T.M.; Arndt-Jovin, D.J. Distribution of resting and ligand-bound ErbB1 and ErbB2 receptor tyrosine kinases in living cells using number and brightness analysis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 16524–16529, doi:10.1073/pnas.1002642107.
14. Sarkar-Banerjee, S.; Sayyed-Ahmad, A.; Prakash, P.; Cho, K.J.; Waxham, M.N.; Hancock, J.F.; Gorge, A.A. Spatiotemporal Analysis of K-Ras Plasma Membrane Interactions Reveals Multiple High Order Homo-oligomeric Complexes. *J. Am. Chem. Soc.* **2017**, *139*, 13466–13475, doi:10.1021/jacs.7b06292.
15. Chen, Y.; Wei, L.-N.; Müller, J.D. Probing protein oligomerization in living cells with fluorescence fluctuation spectroscopy. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15492–15497, doi:10.1073/pnas.2533045100.
16. Saffarian, S.; Li, Y.; Elson, E.L.; Pikey, L.J. Oligomerization of the EGF receptor investigated by live cell fluorescence intensity distribution analysis. *Biophys. J.* **2007**, *93*, 1021–1031, doi:10.1529/biophysj.107.105494.
17. Kask, P.; Palo, K.; Ullmann, D.; Gall, K. Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 13756–13761, doi:10.1073/pnas.96.24.13756.
18. Palo, K.; Mets, Ü.; Jäger, S.; Kask, P.; Gall, K. Fluorescence intensity multiple distributions analysis:

- Concurrent determination of diffusion times and molecular brightness. *Biophys. J.* **2000**, *79*, 2858–2866, doi:10.1016/S0006-3495(00)76523-4.
19. Politz, J.C.; Browne, E.S.; Wolf, D.E.; Pederson, T. Intranuclear diffusion and hybridization state of oligonucleotides measured by fluorescence correlation spectroscopy in living cells. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6043–6048, doi:10.1073/pnas.95.11.6043.
 20. Schwille, P.; Haupts, U.; Maiti, S.; Webb, W.W. Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophys. J.* **1999**, *77*, 2251–2265, doi:10.1016/S0006-3495(99)77065-7.
 21. Pan, P.-Y.; Li, X.; Wang, J.; Powell, J.; Wang, Q.; Zhang, Y.; Chen, Z.; Wicinski, B.; Hof, P.; Ryan, T.A.; et al. Parkinson's Disease-Associated LRRK2 Hyperactive Kinase Mutant Disrupts Synaptic Vesicle Trafficking in Ventral Midbrain Neurons. *J. Neurosci.* **2017**, *37*, 11366–11376, doi:10.1523/JNEUROSCI.0964-17.2017.
 22. Chen, Y.; Müller, J.D.; So, P.T.C.; Gratton, E. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophys. J.* **1999**, *77*, 553–567, doi:10.1016/S0006-3495(99)76912-2.
 23. Muller, J.D.; Chen, Y.; Gratton, E. Resolving heterogeneity on the single molecular level with the photon-counting histogram. *Biophys. J.* **2000**, *78*, 474–486, doi:10.1016/S0006-3495(00)76610-0.
 24. Werner, A.; Skakun, V.V.; Meyer, C.; Hahn, U. RNA dimerization monitored by fluorescence correlation spectroscopy. *Eur. Biophys. J.* **2011**, *40*, 907–921, doi:10.1007/s00249-011-0701-8.
 25. Oasa, S.; Sasaki, A.; Yamamoto, J.; Mikuni, S.; Kinjo, M. Homodimerization of glucocorticoid receptor from single cells investigated using fluorescence correlation spectroscopy and microwells. *FEBS Lett.* **2015**, *589*, 2171–2178, doi:10.1016/j.febslet.2015.07.003.
 26. Skakun, V.V.; Engel, R.; Digris, A.V.; Borst, J.W.; Visser, A.J.W.G. Global analysis of autocorrelation functions and photon counting distributions. *Front. Biosci. Elit.* **2011**, *3*, 489–505, doi:10.2741/e264.
 27. Skakun, V.V.; Engel, R.; Borst, J.W.; Apanasovich, V.V.; Visser, A.J.W.G. Simultaneous diffusion and brightness measurements and brightness profile visualization from single fluorescence fluctuation traces of GFP in living cells. *Eur. Biophys. J.* **2012**, *41*, 1055–1064, doi:10.1007/s00249-012-0864-y.
 28. Spencer, D.M.; Wandless, T.J.; Schreiber, S.L.; Crabtree, G.R. Controlling signal transduction with synthetic ligands. *Science* **1993**, *262*, 1019–1024, doi:10.1126/science.7694365.
 29. Guan, Y.; Meurer, M.; Raghavan, S.; Rebane, A.; Lindquist, J.R.; Santos, S.; Kats, I.; Davidson, M.W.; Mazitschek, R.; Hughes, T.E.; et al. Live-cell multiphoton fluorescence correlation spectroscopy with an improved large Stokes shift fluorescent protein. *Mol. Biol. Cell* **2015**, *26*, 2054–2066, doi:10.1091/mbc.E14-10-1473.
 30. Skakun, V.V.; Hink, M.A.; Digris, A.V.; Engel, R.; Novikov, E.G.; Apanasovich, V.V.; Visser, A.J.W.G. Global analysis of fluorescence fluctuation data. *Eur. Biophys. J.* **2005**, *34*, 323–334, doi:10.1007/s00249-004-0453-9.
 31. Davis, L.; Shen, G. Accounting for Triplet and Saturation Effects in FCS Measurements. *Curr. Pharm. Biotechnol.* **2006**, *7*, 287–301, doi:10.2174/13892010677950843.
 32. Widengren, J.; Mets, Ü.; Rigler, R. Fluorescence correlation spectroscopy of triplet states in solution: A theoretical and experimental study. *J. Phys. Chem.* **1995**, *99*, 13368–13379, doi:10.1021/j100036a009.
 33. Vámosi, G.; Mücke, N.; Müller, G.; Krieger, J.W.; Curth, U.; Langowski, J.; Tóth, K. EGFP oligomers as natural fluorescence and hydrodynamic standards. *Sci. Rep.* **2016**, *6*, 33022, doi:10.1038/srep33022.
 34. Youker, R.T.; Teng, H. Measuring protein dynamics in live cells: Protocols and practical considerations for fluorescence fluctuation microscopy. *J. Biomed. Opt.* **2014**, *19*, 090801, doi:10.1117/1.jbo.19.9.090801.
 35. Dickson, R.M.; Cubitt, A.B.; Tsient, R.Y.; Moerner, W.E. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature* **1997**, *388*, 355–358, doi:10.1038/41048.
 36. Pruschy, M.N.; Spencer, D.M.; Kapoor, T.M.; Miyake, H.; Crabtree, G.R.; Schreiber, S.L. Mechanistic studies of a signaling pathway activated by the organic dimerizer FK1012. *Chem. Biol.* **1994**, *1*, 163–172, doi:10.1016/1074-5521(94)90006-X.

37. Marquardt, D.W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431–441, doi:10.1137/0111030.
38. Skakun, V.V.; Digris, A.V.; Apanasovich, V.V. Global analysis of autocorrelation functions and photon counting distributions in fluorescence fluctuation spectroscopy. *Methods Mol. Biol.* **2014**, *1076*, 719–741, doi:10.1007/978-1-62703-649-8_33.
39. Palo, K.; Mets, Ü.; Loorits, V.; Kask, P. Calculation of photon-count number distributions via master equations. *Biophys. J.* **2006**, *90*, 2179–2191, doi:10.1529/biophysj.105.066084.
40. Veltman, D.M.; Keizer-Gunnink, I.; Haastert, P.J.M.V. An extrachromosomal, inducible expression system for Dictyostelium discoideum. *Plasmid* **2009**, *61*, 119–125, doi:10.1016/j.plasmid.2008.11.002.
41. Kollmar, M. Use of the myosin motor domain as large-affinity tag for the expression and purification of proteins in Dictyostelium discoideum. *Int. J. Biol. Macromol.* **2006**, *39*, 37–44, doi:10.1016/j.ijbiomac.2006.01.005.
42. Bevington, P.R.; Robinson, D.K. *Data Reduction and Error Analysis for the Physical Sciences*; 3rd ed.; McGraw-Hill Publishing Company: New York City, NY, USA, 2003; ISBN 0072472278/9780072472271.
43. Skakun, V.V.; Novikov, E.G.; Apanasovich, T.V.; Apanasovich, V.V. Fluorescence cumulants analysis with non-ideal observation Profiles. *Methods Appl. Fluoresc.* **2015**, *3*, doi:10.1088/2050-6120/3/4/045003.
44. Matthews, J.M.; Sunde, M. Dimers, oligomers, everywhere. *Adv. Exp. Med. Biol.* **2012**, *747*, 1–18, doi:10.1007/978-1-4614-3229-6_1.

Supplementary Information

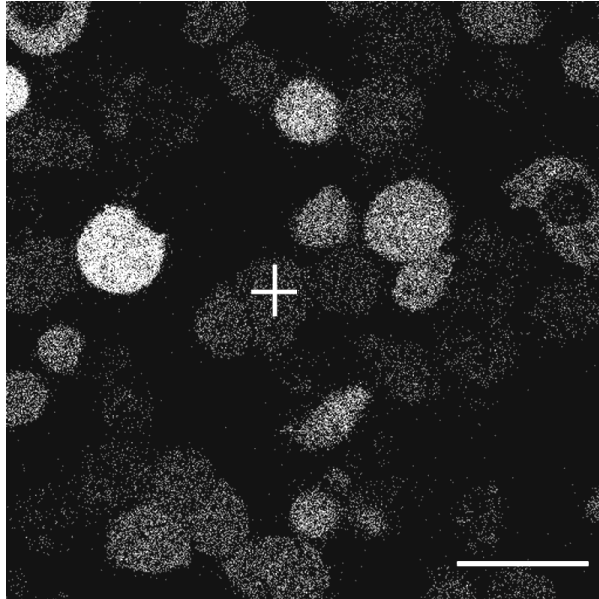


Figure S1: Selection of cells. Green fluorescent images of *Dictyostelium* cells during the FCS experiment. Shown are vegetative GFP-expressing *Dictyostelium* cells in the presence of 2% DMSO. The crosshair (+) was placed on relatively dim cells, for a better resolution. Scale bar: 20 μm .

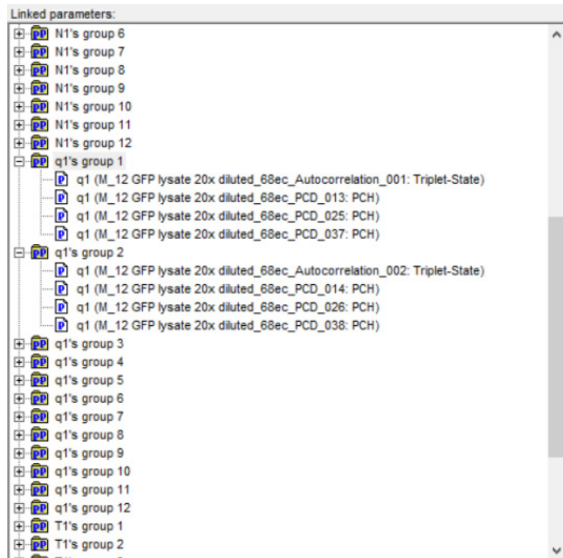


Figure S2: Grouping of parameters per trace. Illustration of linking parameter q1 per trace. Three PCD curves and one ACF curve derived from the same 5-second trace were grouped. In this example twelve groups were formed for parameter q1 and twelve groups for parameter N1. Icons do not illustrate files and folders but parameters and groups.

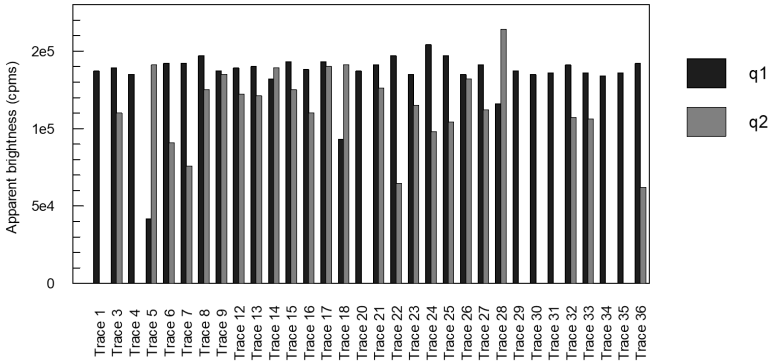


Figure S3: Brightness, estimated in 2 components fit when all parameters are free. Representative example of the initial 2 component global analysis of monomer-dimer mixtures (50%) with free parameters. F_{trip} , τ_{trip} and τ_{diff} were grouped in the global analysis of only ACFs, and their outcomes were subsequently fixed in the combined global analysis of ACFs and PCDs, while $\tau_{diff 2}$, q_1 , q_2 , N_1 and N_2 were free and only grouped per trace (parameters of one ACF and three PDC curves were grouped per trace, not all traces together). The sample shown here is a mix of 50% GFP and 50% diGFP in cell lysate. The brightness of GFP and diGFP on the same measurement day were 0.92 ± 0.015 and $1.95 \pm 0.030 \times 10^5$ cpms, respectively. Average $q_1 = 1.34 \pm 0.20 \times 10^5$ cpms; average $q_2 = 1.15 \pm 0.25 \times 10^5$ cpms, which is in-between the brightness GFP and diGFP. For some traces q_2 could not be determined.

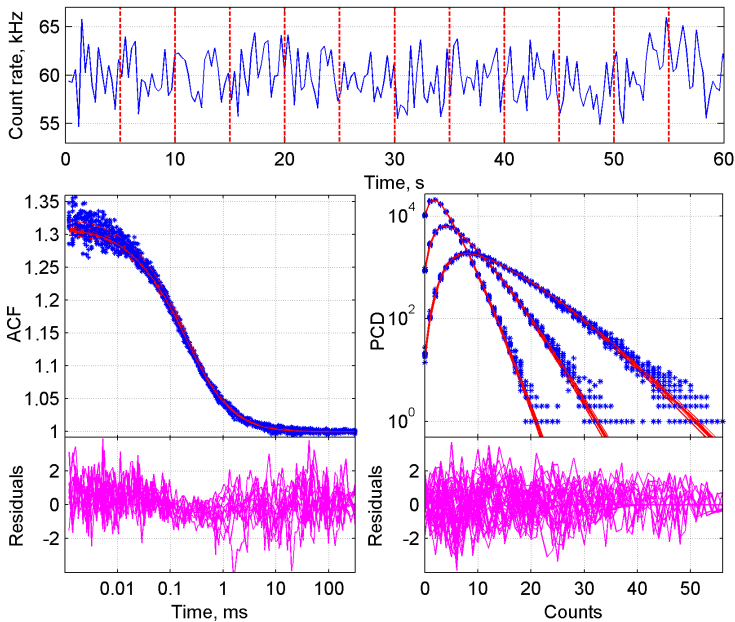


Figure S4: Results of 2-component FSC+PCH global analysis of the ACF and PCD curves from a monomer-dimer equilibrium in cell lysate. The analysis was based on the same data as **Figure 1**. Top panel: raw FFS data showing photon counts over time, from which the ACF and PCD curves were calculated. The measurement was divided into 12 5-second traces, as indicated by vertical lines. Bottom left: fit of all ACF curves, with residuals below. Bottom right: fit of all PCD curves, with residuals below. PCD curves were generated with three different time steps of 5×10^{-5} , 1×10^{-4} and 2×10^{-4} s.

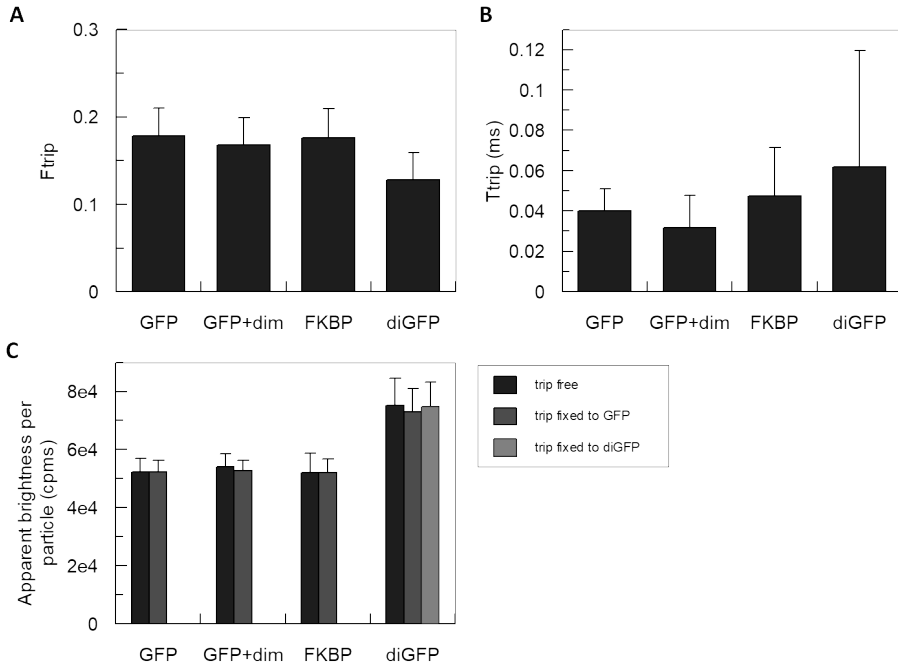


Figure S5: Triplet state parameters may be fixed for all cells. A: Average triplet state fraction (F_{trip}) as determined in the ACF analyses from one measurement day. B: Average triplet state time (τ_{trip}) as determined in the ACF analyses from the same measurement day. C: Average brightness from all cells per sample of one measurement day, with either fixed or free triplet state parameters. Free triplet state parameters were determined per cell in the ACF analysis and subsequently fixed in the combined BDGA analysis. Fixed triplet state parameters were determined by the average of all GFP (or diGFP) cells and fixed in both the ACF and the BDGA analysis. Error bars: standard deviations (SD); 'trip': triplet state parameters.

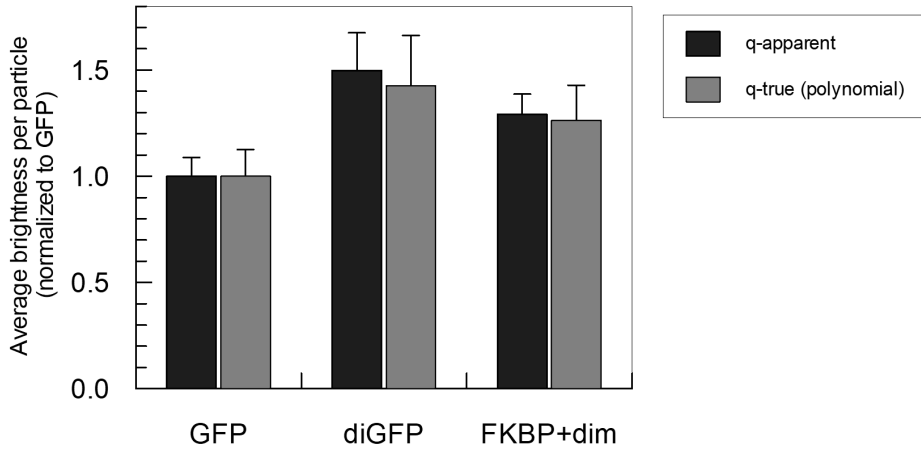


Figure S6: Similar results for the 3D Gaussian and the Polynomial model. Calculations have been repeated with the polynomial model in FFS Data Processor, for a subset of seven cells per sample from one measurement day. q -apparent indicates the apparent relative brightness calculated with the Gaussian model, without correction for Fc_1 , q -polynomial indicates the brightness as calculated with the polynomial model. For both models the same cells have been used for comparison. All brightness values have been normalized to the brightness of GFP control.

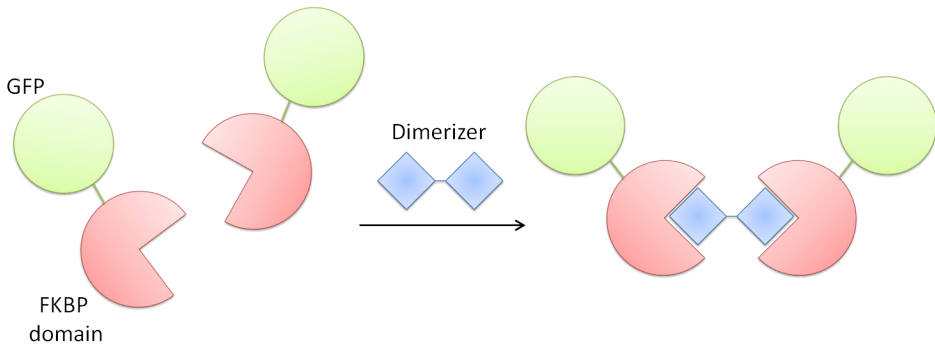


Figure S7: Ligand induced FKBP12 dimerization. Cartoon of the FKBP12 binding domain linked to GFP. When B/B homodimerizer ('Dimerizer') is added to the cells, it symmetrically binds and connects two FKBP12 domains, inducing an indirect link between two GFP fluorophores.

Table S1: List of parameters. The definition and status of the parameters are indicated. All parameters from either one or two component analysis are listed together. ‘Fixed to (the value of) GFP’ means fixed to the average of the GFP measurements of the same measurement day. Free: parameter was not grouped, nor fixed to a value. Grouped: parameter was grouped between different traces, but not fixed to a value. Grouped per trace: different ACF and PCH curves from one trace were grouped together, but not fixed to a value. Fixed: parameter was fixed to a specific value that has been determined before.

Parameter	Explanation	Status in ACF analysis	Status in BDGA
F_1	Fraction of component 1	Free	-
G_{inf}	Value of ACF curve at infinity	Free	Free
F_{trip}	Triplet state fraction	Fixed to the value of GFP	Fixed to the value of GFP
τ_{trip}	Triplet state relaxation time	Fixed to the value of GFP	Fixed to the value of GFP
$\tau_{diff} (\tau_{diff \nu})$	Residence time of component 1	Grouped	Fixed to the value from the ACF analysis
a	Structural parameter	Fixed to the value of R110	Fixed to the value of R110
N	Number of particles in the confocal volume	Free	-
N_1	Number of particles in the confocal volume (of component 1 or 2)	-	Grouped per trace
FC_1	Out-of-focus correction factor	-	Grouped
τ_{dt}	Dead-time correction	-	Grouped
Bg	Background	-	Fixed to 0
q_1	Molecular brightness	-	Grouped per trace
χ^2	Fit criterion	<i>Output</i>	<i>Output</i>

Additional parameters in case of 2-component analysis of lysate data:

F_2	Fraction of component 2 (if applicable)	Free	-
N_2	Number of particles in the confocal volume (of component 2)	-	Grouped per trace
$\tau_{diff 2}$	Residence time of component 2	Fixed to the value of diGFP (and $\tau_{diff 1}$ fixed to GFP)	Fixed to the value of diGFP (and $\tau_{diff 1}$ fixed to GFP)
q_2	Brightness parameter	-	Grouped per trace

Table S2: Example of BDGA output in FFS Data Processor. The standardized BDGA method was applied on the measurement of a cell with FKBP12-GFP and dimerizer (the same cell with FKBP12 as in **Table 3** and **Figure 4**). Some values have been fixed to the values from GFP, while others have been fixed to the values from the ACF analysis of this particular cell, as summarized in **Table S1**. T1: $\tau_{diff 1}$.

Name	Value	Minimum	Maximum	Fixed
Fc1's group	0.927	0	3	False
Ftrip's group 10	0.178	0	0.999	True
Ftrip's group 12	0.178	0	0.999	True
Ftrip's group 13	0.178	0	0.999	True
Ftrip's group 14	0.178	0	0.999	True
Ftrip's group 2	0.178	0	0.999	True
Ftrip's group 4	0.178	0	0.999	True
Ftrip's group 5	0.178	0	0.999	True
Ftrip's group 8	0.178	0	0.999	True
Ftrip's group 9	0.178	0	0.999	True
Ginf	1	0.5	1.5	False
Ginf	0.999	0.5	1.5	False
Ginf	1.001	0.5	1.5	False
Ginf	1.003	0.5	1.5	False
Ginf	1.004	0.5	1.5	False
Ginf	1.003	0.5	1.5	False
Ginf	1.001	0.5	1.5	False
Ginf	1.003	0.5	1.5	False
Ginf	1.003	0.5	1.5	False
N1's group 10	20.15	0	100	False
N1's group 12	17.459	0	100	False
N1's group 13	17.81	0	100	False
N1's group 14	16.525	0	100	False
N1's group 2	21.665	0	100	False
N1's group 4	19.628	0	100	False
N1's group 5	20.205	0	100	False
N1's group 8	19.069	0	100	False
N1's group 9	18.775	0	100	False
T1's group 10	1.018	1.00E-03	10	True
T1's group 12	1.018	1.00E-03	10	True
T1's group 13	1.018	1.00E-03	10	True
T1's group 14	1.018	1.00E-03	10	True
T1's group 2	1.018	1.00E-03	10	True
T1's group 4	1.018	1.00E-03	10	True
T1's group 5	1.018	1.00E-03	10	True
T1's group 8	1.018	1.00E-03	10	True
T1's group 9	1.018	1.00E-03	10	True
Tdt's group	6.34E-05	1.00E-06	1.00E-03	False
Ttrip's group	4.00E-02	1.00E-06	1	True
a's group	5.9	1.001	20	True
bg's group	0	0	1.00E+06	True
q1's group 10	6.30E+04	0	1.00E+07	False

Table S2: Continued.

Name	Value	Minimum	Maximum	Fixed
q1's group 12	6.15E+04	0	1.00E+07	False
q1's group 13	5.64E+04	0	1.00E+07	False
q1's group 14	5.74E+04	0	1.00E+07	False
q1's group 2	7.50E+04	0	1.00E+07	False
q1's group 4	7.55E+04	0	1.00E+07	False
q1's group 5	7.67E+04	0	1.00E+07	False
q1's group 8	6.94E+04	0	1.00E+07	False
q1's group 9	7.05E+04	0	1.00E+07	False

Table S3: Comparison of various settings of the 2-component analysis of one monomer-dimer mix. All analyses have been performed on the same measurement of one sample, which was a mix of 50% GFP and 50% diGFP, consisting of 12 traces of 5 seconds. ' q and τ free': $\tau_{diff,1}$ was grouped and calculated in the global analysis of only ACFs and fixed in BDGA, while $\tau_{diff,2}$, q_1 and q_2 were only grouped per trace not for all traces together. ' q free': $\tau_{diff,1}$ and $\tau_{diff,2}$ were fixed to the values found for GFP and diGFP (**Table 1**), while q_1 and q_2 were fixed to the values found for GFP and diGFP. ' $r = 1.8$ ' (or 2): $\tau_{diff,1}$ and q_1 were fixed to the values found for GFP and diGFP. ' q free': $\tau_{diff,1}$ and $\tau_{diff,2}$ were fixed to the value found for diGFP, while q_2 was fixed to 1.8x or 2x the value of q_1 . ' q_2 of diGFP': $\tau_{diff,1}$, $\tau_{diff,2}$, q_1 and q_2 were fixed to the values found for GFP and diGFP. All parameters except N and $\tau_{diff,2}$ are presented with confidence intervals, calculated as asymptotic standard errors (ASE), as reported by the software. F_{trip} : triplet state fraction; τ_{trip} : triplet state time; $\tau_{diff,1}$ and $\tau_{diff,2}$: diffusion parameters as reported by the software; N_1 and N_2 : number of particles in the confocal volume, SD is indicated; q_{True} and q_{2True} : true brightness, with ASEs representing uncertainties of estimated parameters obtained in analyses performed per one trace. We additionally calculated standard deviations of brightness between traces (SD of q). χ^2 : global analysis fit criterion value. Cpm/s: counts per molecule per second. Fixed values are indicated in italics and have been determined in **Table 1**. Values of N which are closest to the expected 50-50 ratio are indicated in bold.

Analysis method	F_{trip} ($\times 10^{-2}$ ms)	τ_{trip} ($\times 10^{-2}$ ms)	$\tau_{diff,1}$ (ms)	$\tau_{diff,2}$ (\pm SD)	N_1 (\pm SD)	N_2 (\pm SD)	q_{True} ($\times 10^4$ cpm/s)	SD q_{True} ($\times 10^4$ cpm/s)	q_{2True} ($\times 10^4$ cpm/s)	SD q_{2True} ($\times 10^4$ cpm/s)	χ^2
2-comp (q and τ free)	7.13 \pm 0.61	1.47 \pm 0.25	0.171 \pm 0.004	6.2 \pm 11.2	1.73 \pm 0.90	5.28 \pm 3.98	6.58 \pm 0.242	2.089	2.65 \pm 0.21	2.49	0.971
2-comp (q free)	8.07 \pm 0.46	1.91 \pm 0.24	0.151	0.221	1.89 \pm 1.37	2.23 \pm 1.40	6.08 \pm 0.222	3.101	5.42 \pm 0.18	2.24	0.832
2-comp ($r = 1.8$)	8.07 \pm 0.46	1.91 \pm 0.24	0.151	0.221	2.00 \pm 0.18	1.76 \pm 0.12	3.20		5.76		1.068
2-comp ($r = 2$)	8.07 \pm 0.46	1.91 \pm 0.24	0.151	0.221	2.65 \pm 0.14	1.23 \pm 0.08	3.20		6.4		0.924
2-comp (q_2 of diGFP)	8.07 \pm 0.46	1.91 \pm 0.24	0.151	0.221	1.19 \pm 0.27	2.49 \pm 0.17	3.20		5.12		1.294

Table S4: Comparison of various q-ratios of multiple monomer-dimer mixes. All samples have been measured on the same day. Fixed values are indicated in italics. Since this analysis was performed in an earlier stage, F_{trip} and τ_{trip} have been determined per sample (not shown), instead of fixed to the same value for all analyses, and $\tau_{diff 1}$, $\tau_{diff 2}$ and q_1 have been fixed to different values than determined for GFP and diGFP in **Table 1**. N_1 and N_2 have been calculated into percentages by the formulas $N_1\% = N_1 / (N_1 + N_2) \times 100\%$ and $N_2\% = N_2 / (N_1 + N_2) \times 100\%$. The $N_2\%$ closest to the % diGFP (input) are indicated in bold, along with the corresponding q_2/q_1 ratio. The $N_2\%$ may become even closer to the diGFP% when $\tau_{diff 2}$, $\tau_{diff 1}$ and/or q_1 are varied. As shown here, optimal settings varied a lot per sample and therefore no standardized procedure could be determined.

% diGFP	$\tau_{diff 1}$	$\tau_{diff 2}$	q_1	r q_2/q_1	N_1 %	N_2 %
10	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.8	93.2	6.8
	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.7	91.8	8.2
12.5	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.8	88.2	11.8
	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.7	85.7	14.3
33.3	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.8	71.2	28.8
	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.6	56.2	43.8
50	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.7	64.7	35.3
	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.8	50.0	50.0
67.7	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.6	14.0	86.0
	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.8	45.3	54.7
87.5	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.7	31.5	68.5
	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.8	28.2	71.8
90	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.6	7.5	92.5
	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.7	9.7	90.3
90	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.8	30.6	69.4
	<i>0.14</i>	<i>0.2</i>	<i>5.50E+04</i>	1.6	9.6	90.4

Step-by-step protocol of FCS Data analysis using the BDGA methodology

1. Data preparation

- a. For cellular data, split each measurement into traces of 3-5 seconds. Note. *In vitro* data is more stable and may consist of longer traces.
- b. For each trace, calculate one ACF (time step 1×10^{-7} s (cells) or 2×10^{-7} s (lysate); point count 140) and three PCDs (point count 32), the latter with different time steps depending on the total fluorescence intensity, as described in section 3.1.1.

2. Curve fitting with the 1-component BDGA model

2.1 ACF only

1. Fit globally all ACF traces from one measurement using the free diffusion 3D Gaussian model with the triplet-state term (Eq. 1).
 - a. Group parameters F_{trip} , τ_{trip} , and τ_{diff1} between traces, and fix parameter a to the value found for the Rhodamine 110 calibration dye.
 - b. Generate initial guesses.
 - c. Run the analysis.
 - d. Remove traces with a very aberrant fit (e.g., with a steep drop of the ACF curve on the right side of the plot, or an overall alleviation of the curve compared to the other traces; this correlated with irregularities in the original traces), and run the analysis again.
 - e. The chi-square value is a measure for how good the fit is and should be close to 1, but any value between 0.7 and 1.4 is considered acceptable.

2. Calculate averages F_{trip} and τ_{trip} from all GFP measurements of one day and then fix them for all samples of that day (except diGFP), as recommended in section 3.2.

Note. In the 2-component BDGA analysis, parameters τ_{diff2} , N_2 , and q_2 were added to the model and grouped per trace. F_{trip} , τ_{trip} , τ_{diff1} and a were fixed like in the 1-component model, but τ_{diff2} was not, to allow for differences between traces.

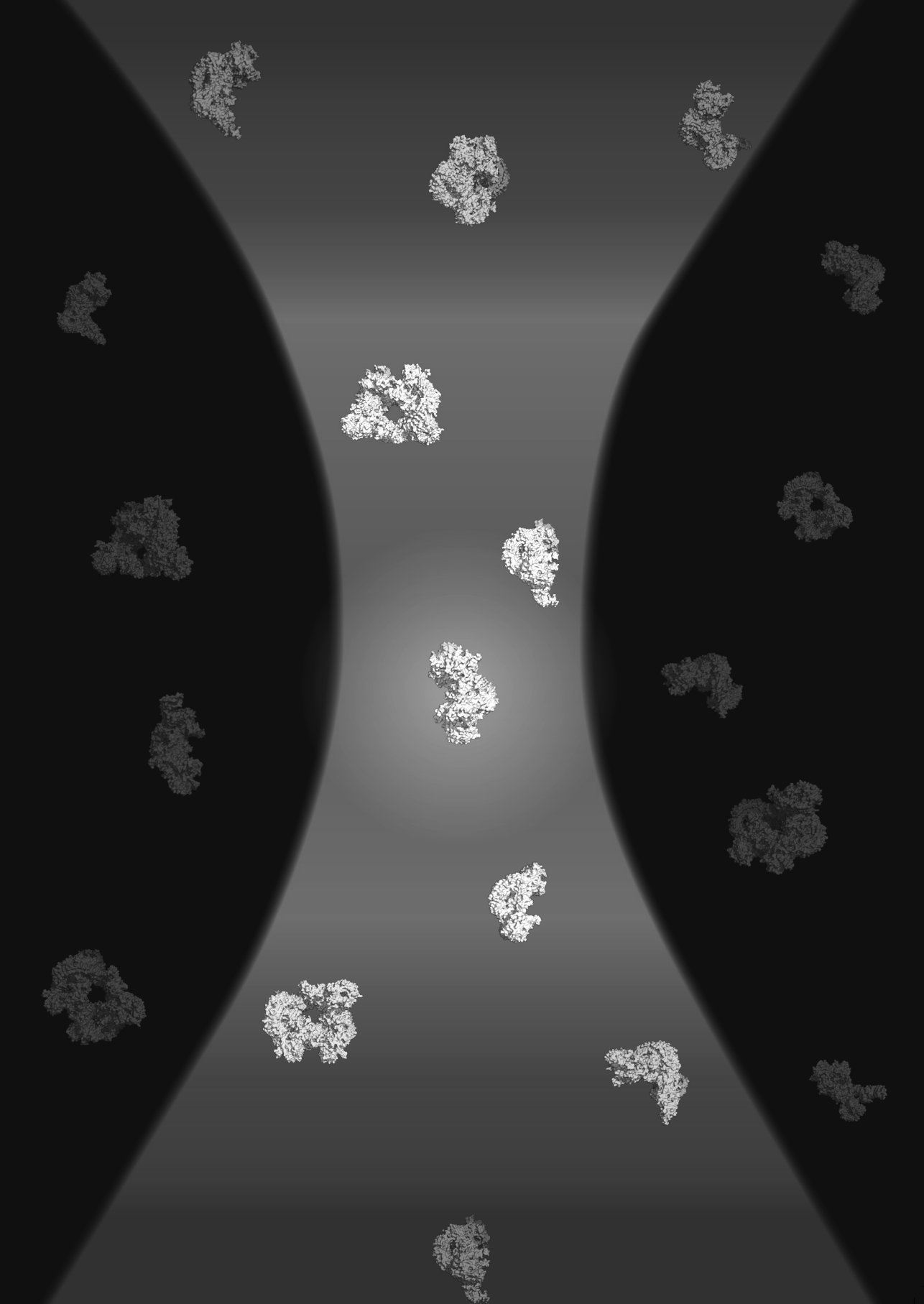
2.2 Combined ACF and PCD

1. Fit globally all ACF and PCD curves from the same measurement using the protocol described in detail in [38]. The FCS and PCH models are calculated as described in section 2.1.
 - a. Group the parameters τ_{diff1} , F_{trip} , N_1 , and q_1 per trace (one ACF and three PCDs), while grouping τ_{trip} , a and some other parameters (Fc_1 , τ_{dt} , bg) for all traces together.
 - b. Fix parameters a , F_{trip} , and τ_{trip} to the same values as fixed during the analysis of only the ACF curves, while fix τ_{diff1} to the value found during that ACF analysis.
 - c. Run the analysis immediately without generation of initial guesses.

- d. Examine the fits and residuals of the ACF and the PCD curves for aberrations, before exporting the resulting parameter values to a text file.

3. Postprocessing of fit results

- a. Import the resulting data in a spreadsheet program like Microsoft Excel.
- b. Calculate the average q_1 , q_2 , τ_{diff1} and τ_{diff2} per cell.
- c. Normalize obtained brightness values towards GFP on that day.
- d. Average both brightness and diffusion times between all cells of multiple days.
- e. Convert τ_{diff1} and τ_{diff2} into D_1 and D_2 by the formulas given in section 2.1.
- f. The significance of apparent differences between samples may be determined using the standard two-sided t-test with unequal variances.
- g. Calculate standard deviations from the variation between cells, based on the normalized data from multiple measurement days.



5

Using Brightness and Diffusion Global Analysis to study the dimerization of Parkinson's Disease-related Roco proteins in living Dictyostelium and HEK293 cells

Laura M. Nederveen-Schippers¹, Pragya Pathak¹, Ineke Keizer-Gunnink¹,
Adrie H. Westphal², Jan Willem Borst², Peter J.M. van Haastert¹, Victor Skakun³,
and Arjan Kortholt^{1,*}

¹Department of Cell Biochemistry, University of Groningen, 9747 AG Groningen, The Netherlands.

²Laboratory of Biochemistry, Wageningen University & Research, 6708 WE Wageningen, The Netherlands.

³Department of Systems Analysis and Computer Simulation, Belarusian State University, 220030 Minsk, Belarus. *Corresponding author: a.kortholt@rug.nl

Author contributions: PvH, AK, JWB, VS, and **LNS**. designed the experiments; **LNS**, PP, and IKG performed the experiments with the help of AW and JWB; VS developed and adapted the data analysis algorithms and software, and **LNS**, PP and IKG. analyzed the data. **LNS**, PP, IKG, VS, and AK contributed to writing of the manuscript.

Abstract

Dimerization is an important step in the activation mechanism of LRRK2, a Roco protein involved in Parkinson's Disease (PD). Elucidating this mechanism is essential for understanding the etiology of the disease and for developing new drugs targeting LRRK2 activation, as LRRK2 has a higher kinase activity in PD patients. However, studying LRRK2 dimerization *in vivo* has been challenging. Therefore, we adapted the Brightness and Diffusion Global Analysis (BDGA) methodology for its application to complex proteins in living cells. BDGA combines Fluorescence Correlation Spectroscopy (FCS) with Photon Counting Histogram (PCH) analysis of Fluorescence Fluctuation Spectroscopy (FFS) data, exploiting the property that dimers exhibit different brightness and diffusion patterns than monomers. To further develop BDGA for complex proteins, we used the amoeba *Dictyostelium discoideum* (*Dictyostelium*) as a model organism. First, we optimized the FFS measurement and BDGA analysis for *Dictyostelium* by studying reversible induced dimerization of FKBP12. Subsequently we used the optimized BDGA methodology to study oligomerization of *Dictyostelium* Roco4 and humanLRRK2 in HEK293 cells. Our data confirm previous models that both Roco4 and LRRK2 are mainly monomeric in the cytosol, with a small fraction of dimeric particles that are very slow diffusing and are most likely bound to membranes. Our research opens new avenues to study the *in cellulo* dimerization of complex proteins in various cell types, and study the influence of different conditions like PD-mutations, inhibitors, upstream proteins and nucleotide states on localization and dimerization of LRRK2.

Keywords

BDGA or Brightness and Diffusion Global Analysis, *Dictyostelium discoideum*, FKBP12, fluorescence correlation spectroscopy, LRRK2, reversible dimerization, Roco4

Introduction

The Roco protein family was first identified in the amoeba *Dictyostelium discoideum* and is characterized by the presence of a Roc GTPase followed directly by a C-terminal-of-Roc (COR) domain [1](**Figure 1**). In addition to the Roc-COR domains, most members of the Roco family contain multiple protein-protein interaction domains and a kinase domain, although the latter is lacking in bacteria [2]. Roco proteins have a complex activation mechanism, influenced by both enzymatic domains, protein-protein interactions, localization and dimerization. Importantly, dimerization is essential for maximum activity of Roco proteins [3].

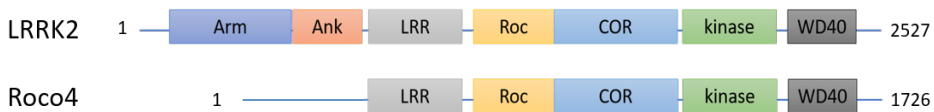


Figure 1: Domain structure of Roco proteins LRRK2 and Roco4. Catalytic core: Roc, Ras-of-complex proteins (G-nucleotide binding domain); COR, C-terminal of Roc (facilitating dimerization); and a kinase domain. Protein-protein interaction domains: Arm, armadillo repeats; Ank, ankyrin repeats; LRR, Leucine Rich Repeats, and a WD40 domain.

The most prominent member of the Roco family is the human Leucine Rich Repeat Kinase 2 (LRRK2), which is a major factor for the cause of Parkinson’s Disease (PD) [4]. PD is an increasing problem, with the number of cases growing faster than the population of elderly. There is currently no cure available [5] nor are there good biomarkers to detect the disease. Some treatments to delay the disease progression are currently being tested in clinical phase, but those are not ideal due to side effects, heaviness (e.g. brain surgery), and unknown effectiveness [6–8]. Therefore, there is an urgent need to identify the pathways underlying the disease and find novel targets targeting PD.

LRRK2 kinase activity has been reported to be increased in both sporadic and familial forms of PD [4,9,10]. Strikingly, different PD mutations lead to an increased kinase activity, even though they are located in different domains of the protein [11]. Because compounds that directly inhibit the LRRK2 kinase activity may cause problems in lung and kidney [12,13], indirect allosteric inhibition of LRRK2 activity through its activation mechanism is now being explored.

One essential step in the activation mechanism of LRRK2 is dimerization, which is therefore an ideal target for intervention [3,14–16]. Importantly, based on in vitro data it has been reported that some PD mutations most likely affect dimerization [17–19].

However, tools to study the dynamics of LRRK2 dimerization in living cells are limited. One previous study assessed LRRK2 dimerization using Photon Counting Histograms, Number and Brightness analysis and Autocorrelation analysis, showing an increase in dimerization of the G2019S mutant variant [19]. However, these were relatively simple, singular analyses and thus no direct relationship could be established between diffusion and brightness per particle. Therefore, we here aimed to adopt and optimize our previous developed Brightness and Diffusion Global Analysis (BDGA) of Fluorescence Fluctuation Spectroscopy (FFS) data methodology [20] to study the dimerization of Roco and LRRK2 proteins in their cellular environment.

FFS is based on the Brownian diffusion of fluorescent molecules through a fixed confocal observation volume, causing fluctuations in the fluorescence signal. Other processes that may influence these fluctuations are triplet state dynamics, reversible reactions, and protonation. FFS data is most commonly analyzed through autocorrelation of the fluctuations and then called Fluorescence Correlation Spectroscopy (FCS) [21–23]. In FCS, AutoCorrelation Functions (ACFs) are calculated from the raw FFS data and fitted with a model to extract the diffusion speed and number of particles in the confocal observation volume. Alternatively, FFS data can be converted into Photon Counting Distributions (PCDs), which are then analyzed by Photon Counting Histogram (PCH) analysis, yielding information on the number of particles as well as the brightness per particle [24,25]. BDGA combines the FCS and PCH analyses by fitting the ACF and PCD curves from the same FCS data with linked mathematical models [20].

To optimize the BDGA method for the analysis of complex proteins in living cells, we initially used *Dictyostelium discoideum* as a model organism, a unicellular amoeba which forms multicellular fruiting bodies upon starvation. *Dictyostelium* contains multiple Roco proteins, among which Roco4 has been most well studied and has a domain architecture similar to LRRK2 [26–29]. Moreover, Roco4 overlaps with LRRK2 regarding activation mechanism and function. Both proteins phosphorylate Rab proteins and influence mitochondrial respiration [11,30,31].

BDGA analysis in *Dictyostelium* was previously developed using FKBP12 [20], which functions as a ligand-inducible dimerizing domain when fused to a protein of interest [32,33]. FKBP12 naturally binds rapamycin and mTOR [34], and has been extensively used to induce dimerization in mammalian cell cultures and *in vivo* (e.g. [35–38]). In the first part of this chapter, we describe the optimization of the FKBP12 system in more detail, focusing on specific adaptations of the sample preparation and measurement of *Dictyostelium* cells. Moreover, we will show that this induced dimerization of FKBP12 is reversible. In the second part of this chapter, we will apply the optimized BDGA to the analysis of Roco4 in *Dictyostelium* and LRRK2 in HEK293 cells. We demonstrate that

cellular Roco4 and LRRK2 exist in a monomer-dimer equilibrium, which opens the way to study the dimerization of other complex proteins in living cells.

Results and discussion

Brightness and Diffusion Global Analysis of FCS data: five steps

Measuring dimerization in living cells is a challenging task, as monomeric and dimeric proteins differ only by a factor of two in mass and brightness, while the cellular environment results in more complex data than measurements *in vitro*. Recently, we optimized the Brightness and Diffusion Global Analysis (BDGA) method for the resolution of monomeric and dimeric species in living cells [20]. Here we describe the specific application of this method to *Dictyostelium*.

In general, the full procedure consists of five steps, i.e. (1) cell culture and sample preparation, (2) sample measurement, (3) ACF and PCD curve calculation from raw data, (4) curve fitting, and (5) parameter analysis (**Figure 2**). In this section we will describe the sample preparation and measurement in detail, followed by a brief summary of the three data analysis steps, as they have been extensively covered in [20].

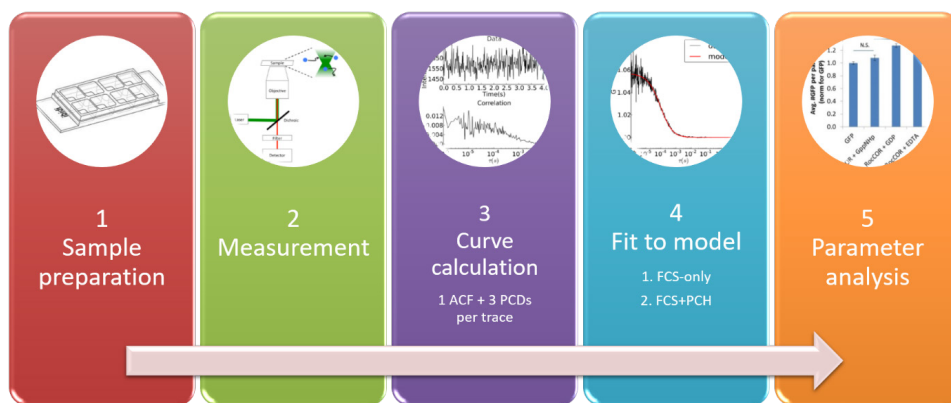


Figure 2: General workflow of the FFS experiments with BDGA analysis. Data collection starts with preparation (1) and measurement (2) of the sample. FFS measurements generate raw fluorescence fluctuation traces from which ACF and PCD curves are being calculated (3). These curves are imported in the FFS analysis interface (Figure S2B), after which first the ACF curves are analyzed with the FCS global analysis method (4.1) and subsequently the corresponding parameters are fixed in the BDGA analysis of all ACF and PCD curves (4.2). The resulting parameters are exported to a spreadsheet program like Microsoft Excel, by which averages and statistics are calculated (5).

Sample preparation of living *Dictyostelium* cells

In a simple setup, cultured *Dictyostelium* cells expressing a fluorescent protein may be transferred to a glass slide and be directly measured with an FCS microscope. Here, we show some adaptations to optimize this setup. Unlike many other techniques, FCS benefits from low protein concentrations. Because the fluorescence fluctuations are caused by molecules diffusing through the confocal volume, the changes per event are relatively greater with fewer fluorescent proteins present, resulting in higher resolution. To control the protein levels, proteins were expressed from plasmids with inducible promoters. We used 10-20 times less inducer than required for full expression [39] and applied it overnight, resulting in a substantial proportion of low-fluorescent cells.

As unicellular amoeba, *Dictyostelium* cells drink liquid medium by a process called macropinocytosis, resulting in a large number of vesicles in the cytoplasm. This could lead to a heterogeneous distribution of fluorescent protein throughout the cell and an unstable signal. We hypothesized that starvation of *Dictyostelium* cells in buffer may reduce the number of vesicles in these cells and increase signal stability. However, when comparing the fluorescence signal of vegetative and starved cells expressing GFP, the stability of the signal did not clearly improve (**Figure 3A** and **Figure S1**). In addition, a simple calculation of the brightness per particle gave similar results and did not become more accurate for starved cells (**Figure 3B**). Based on these observations (**Figure 3** and **Figure S1**), both vegetative and starved cells can be used, at least for small proteins such as GFP. Because starved cells are less sensitive to laser power and our protein of interest (Roco4) has been reported to play a role in development [28], we decided to use starved cells by default in all our experiments.

Starved *Dictyostelium* cells have a higher tendency to move away from the focal point, resulting in a sudden loss of signal. Therefore, the cells were immobilized by the addition of DMSO. Addition of 2% (v/v) DMSO to cells starved for 5 hours at room temperature resulted in round and non-motile cells (own observations). *Dictyostelium* cells starved overnight in the refrigerator required more DMSO for immobilization and were therefore not used for further experiments. DMSO is a common solvent for drugs and other compounds and the used concentrations are not toxic for *Dictyostelium* cells [40]. But of course, in general, appropriate controls to investigate the effect of DMSO on the pathway of interest should be included in the experimental setup.

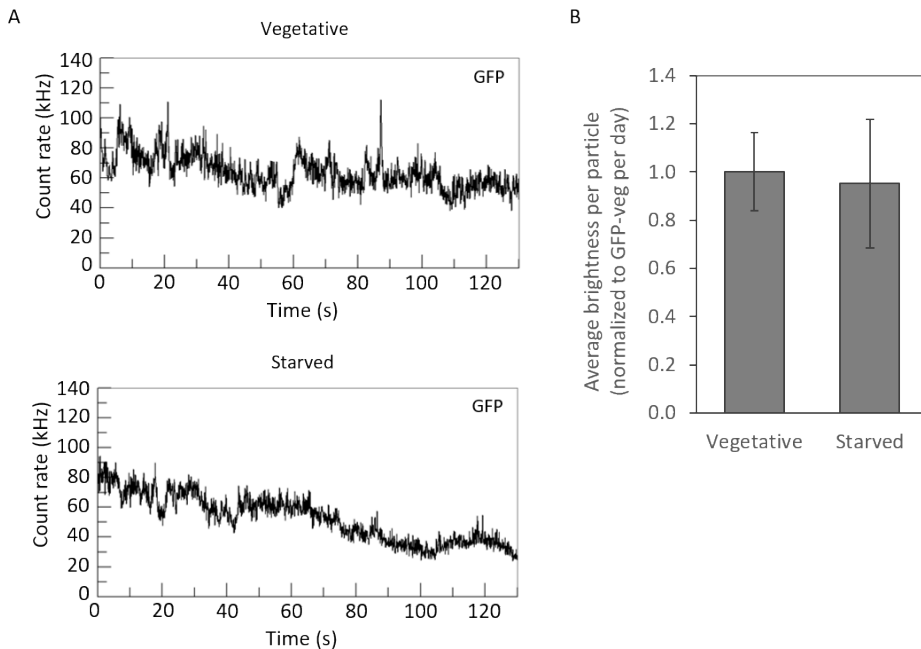


Figure 3: Vegetative and starved cells. **A:** Representative measurements of vegetative and 5 h starved cells expressing monomeric GFP, showing variation in the signal. **B:** Brightness results of GFP are similar for vegetative and 5 h starved cells. The FCS measurements have been analyzed with the simple brightness analysis. Measurements consisted of 30 traces of 5.5 seconds. Brightness values were normalized to the averaged vegetative samples (GFP-veg) per measurement day. $n= 8-10$ cells from 2-3 measurement days.

FCS measurements of Dictyostelium cells

The samples were examined with a confocal microscope with FCS module. First, the fluorescence of GFP in cells was induced and imaged in the confocal scanning mode, to find the correct focus (z-position) and to search for an area with good cells. An example of a good density and distribution of cells is given in **Figure 4A**. Then, a spot for the FCS measurement was marked by placing a cross hair in a region with homogeneous GFP expression in a dim cell (**Figure 4B**), after which the FCS measurement was started. Note that low laser intensities were used for both imaging and FCS measurement to avoid bleaching of the fluorophore. Multiple cells in one well were usually measured for 45 seconds per cell, while a fresh sample was prepared every 30 minutes.

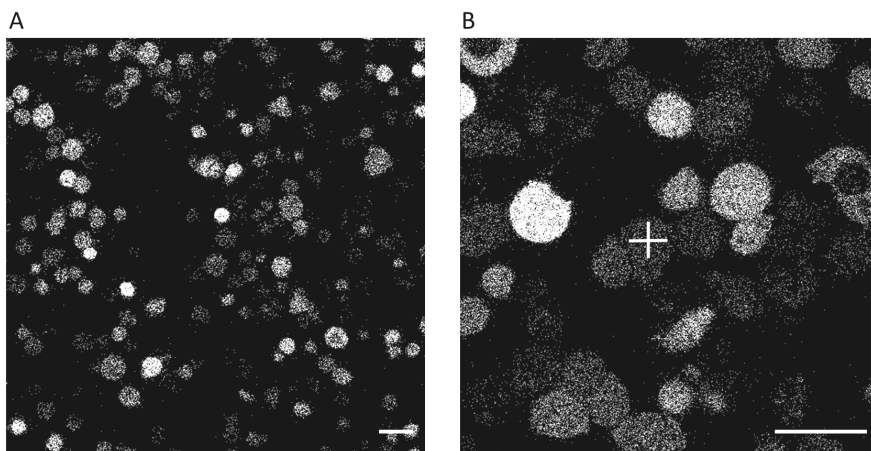


Figure 4: Selection of cells. Green fluorescent images of *Dictyostelium* cells during the FCS experiment. Shown are vegetative GFP-expressing *Dictyostelium* cells in the presence of 2% DMSO. Starved cells gave a similar picture. **A:** Representative illustration of a good cell density: a single layer of cells with some space in between. **B:** Closer view of cells as used during the experiments to select a good spot for the experiment, indicated by a crosshair (+). The crosshair was placed on relatively dim cells, for a better resolution. Scale bar: 20 μm .

FCS data analysis using the BDGA method

The analysis of the FCS data (step 3-5 in **Figure 2**) has been extensively described in [20]. An impression of the data analysis program (FFS Data Processor (FFS DP)) is given in Supplementary **Figure S2**. In short, the raw data was first imported into the FFS DP Measurement Database, loaded into the Raw Data Builder and displayed as a fluorescence intensity plot (photon frequency (kHz) over time (s) (**Figure S2A**, top panel)). Each measurement was divided into short 3-second traces to minimize artefacts caused by a high non-stationarity of the signal, and one autocorrelation function (ACF) and three photon counting distributions (PCDs) were calculated for each trace.

To determine the values of dynamic parameters such as diffusion and triplet state characteristic times in a simpler and thus more reliable model compared to the combined FCS and PCH analysis, ACF curves were first analyzed separately using the global FCS analysis (without PCH), as described in [20,41] (point 1 of step 4 in **Figure 2**). Briefly, the generated ACF curves of all traces from one measurement were loaded into the analysis platform of FFS DP (**Figure S2B**) and aberrant traces were removed (e.g. with a steep drop of the ACF curve on the right side of the plot, or an overall alleviation of the curve compared to the other traces; the number of traces removed ranged from a few to half of the traces from a single measurement in a cell). The ACF curves were fitted with the FCS global analysis method, using the ‘1-component free diffusion 3D Gaussian model’ as described [20]. Relevant parameters in this model are diffusion time τ_{diff} , structural parameter a , triplet state fraction F_{trip} , and triplet state

decay time τ_{trip} . (Figure S2C). The quality of the fit is estimated by calculation of χ^2 criterion values and visual inspection of residuals (as in Figure S2E).

Next, all ACF and PCD curves were fitted together (point 2 of step 4 in Figure 2), while fixing the relevant parameters from the previous step. The N and q parameters associated with the ACF and PCD curves were grouped by trace (Figure S2D). N is the number of particles in the confocal volume, and q is the brightness per particle. The resulting fit (Figure S2E-F) was inspected and the average brightness q was first calculated per cell (of all measured traces), after which the average of multiple cells in the sample was calculated. All measurements were normalized to the brightness of (cellular) GFP on that day, and averaged over multiple days.

One measurement of – in our case – 45-150 seconds ultimately results in one data point. Therefore, dynamic processes in the order of seconds cannot be observed in real time with this method. Cellular responses lasting only for a few minutes could potentially be measured in a different setup, for example using a flow chamber or channel slide. In our setup, however, more sustained changes were investigated, for example through the addition of dimerizing compounds, as discussed below.

Reversible homodimerization via the FKBP12 binding domain in *Dictyostelium*

When the brightness of tandem-dimer GFP (diGFP) was compared to monomeric GFP in *Dictyostelium* cells, the dimer was significantly brighter than the monomer (Figure 5 and previous data [20]). However, diGFP is expressed as one protein, resulting in a uniform population of ‘dimer’, while natural proteins often switch between a monomer-dimer equilibria in the cell.

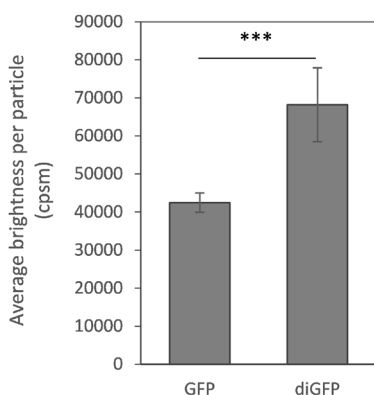


Figure 5: Brightness analysis of GFP and diGFP in cells before normalization. 3h-starved *Dictyostelium* cells expressing GFP or diGFP were measured on a single day and analyzed with the described BDGA analysis method (3 sec per trace). Error bars indicate the standard deviation (SD), $n=10$ cells per cell line. Normalized averages of GFP and diGFP on multiple measurement days can be found in Figure 6. *** $p<0.001$.

We have adopted the TaKaRa/Clontech's iDimerize inducible dimerization system (based on [32]) for *Dictyostelium* to test and optimize our BDGA methodology. In this system, dimerization occurs via the DmrB domain, a derivative of the FKBP12 protein, which is a member of the FK506 Binding Protein (FKBP) family. An F36V point mutation facilitates the binding of the dimerizing ligand (B/B Homodimerizer, hereinafter referred to as 'Dimerizer'), while endogenous FKBP12 is insensitive to the compound [42]. FKBP12-F36V is further referred to as FKBP12. FKBP12 is relatively small and functions as a binding domain when fused to a protein of interest, in our case GFP (**Figure 6A**). In this way, a monomer-dimer equilibrium of the protein of interest can be created in the cells.

To investigate the appropriate experimental setup for this induced dimerization in *Dictyostelium*, several conditions were tested using FKBP12-GFP (**Figure 6B**). The highest brightness per particle was achieved when cells were incubated for 3 hours with 1 μ M Dimerizer, indicating that dimerization is occurring. Interestingly, incubation for longer periods or at higher concentrations did not improve dimerization, but rather decreased it, possibly because excess Dimerizer would compete for and thereby saturate the binding pockets of FKBP12, leaving too few unbound proteins to form dimers, as hypothesized before [33].

Based on the Clontech iDimerize system, we expected that the dimerization of FKBP12 would be reversible. Indeed, when an excess of B/B Washout Ligand (hereinafter referred to as Splitter) was added to the samples incubated with Dimerizer, no dimerization was observed (**Figure 6C**). To confirm that this effect was not caused by the ethanol solvent, mock samples with 2% EtOH were also analyzed. The ethanol may cause a larger standard deviation, but the average brightness per particle was not affected (**Figure 6C**).

BDGA analysis of Roco4 and LRRK2

Next we applied the BDGA method to *Dictyostelium* Roco4 and human LRRK2. LRRK2 and Roco4 both consist of a catalytic core, flanked by protein-protein interaction domains (**Figure 1**). Biochemical fractionation experiments have shown that LRRK2 is largely monomeric in the cytosol and dimeric at membranes [14]. Importantly, the membrane fraction yielded higher enzyme activity than the cytosolic fraction, although the latter contained the majority of all LRRK2, indicating that dimerization and translocation are important for protein activation. We initially investigated this mechanism in *Dictyostelium*, using Roco4 as a model for LRRK2, and asked whether LRRK2 also has a monomer-dimer equilibrium.

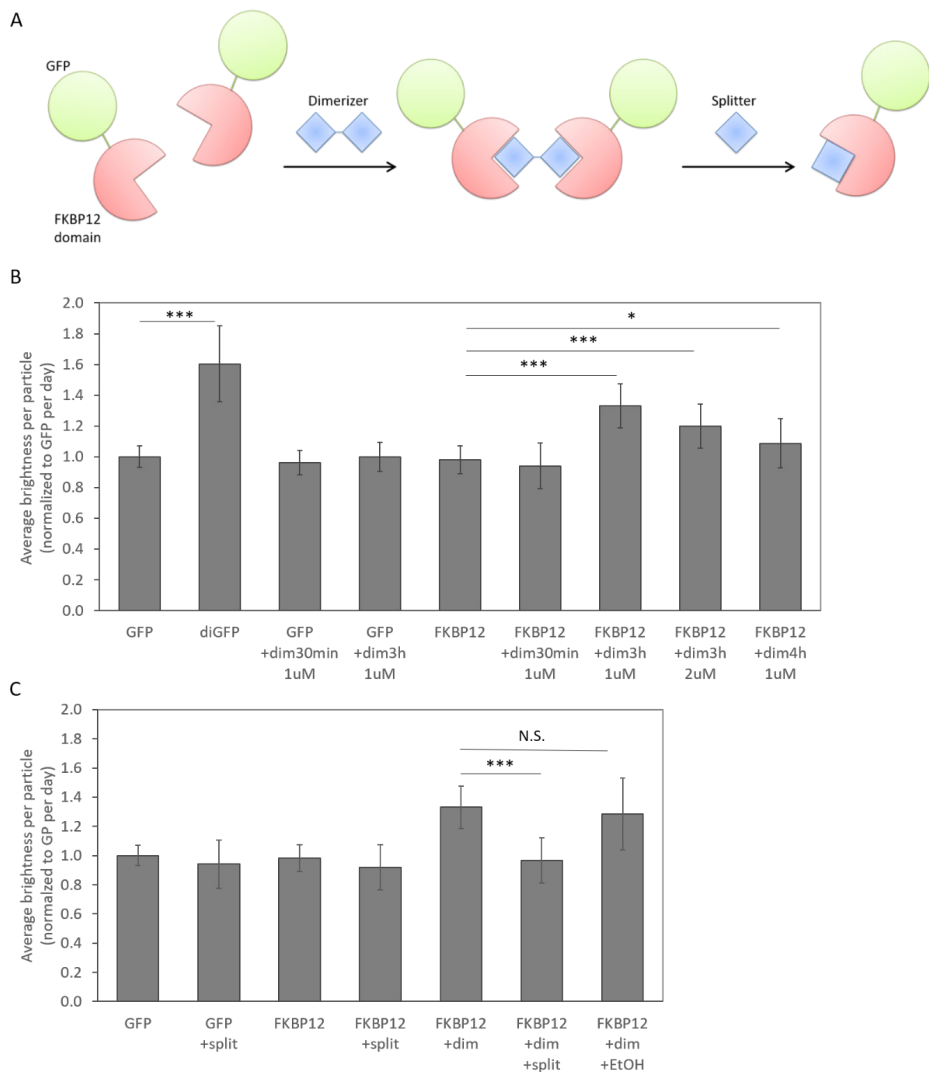


Figure 6: Reversible induced dimerization through the FKBP12 domain. **A:** Cartoon of dimerization through FKBP12. Addition of a dimerizing compound called Dimerizer induces dimerization, while subsequent addition of the washout ligand called Splitter induces monomerization. FKBP12 could also be linked to other proteins than GFP, to investigate cellular effects of dimerization. **B-C:** BDGA analysis of FKBP12 under various conditions. GFP, FKBP12, and FKBP12+dim3h_1uM are repeated in both figures. As Dimerizer was dissolved in ethanol, 0.2% EtOH was added to each mock sample during starvation (not indicated in the figure). All samples (except FKBP12+dim4h_1uM) were starved for 3 hours and each measurement was divided into 3-sec traces. **B:** Dimerization at various incubation times and concentrations of Dimerizer. An incubation time of 3 hours and a concentration of 1 μ M Dimerizer (dim) seem most effective to induce dimerization, compared to the conditions tested. **C:** Dimerization is reversible through the addition of Splitter. Dimerizer was incubated for 3 hours at a concentration of 1 μ M, while Splitter (split) was added just before the measurement at a concentration of 10 μ M. Because Splitter was dissolved in ethanol, the same volume of EtOH (2% of the total volume) was added as a control (final bar). * p <0.05, *** p <0.001, N.S.: non-significant.

Roco4 sample preparation and analysis

To address this question, the optimized BDGA methodology has been adapted for measurements with *Dictyostelium* cells expressing Roco4-GFP. Consistent with the other data, the difference in brightness per particle between vegetative and starved cells was negligible (**Figure 7**, **Figure S3**, **Figure S4A**), although the presence of fluorescence-free vesicles was observed in vegetative cells (**Figure 7A**), therefore we decided to continue our experiments with 5-6 h starved cells.

Although the measurement of Roco4-GFP-expressing *Dictyostelium* cells was the same as for GFP, the data analysis presented new challenges. When our optimized BDGA was applied to the Roco4 data, the fit yielded a clear wave in the residuals of the ACF curve, indicating that the mathematical model was incomplete (**Figure 8A**, left panel). Therefore, we increased the number of components to two. This 2-component model describes the presence of two species with different diffusion velocities, and yielded a good fit in the ACF-only analysis step (**Figure 8A**, right panel).

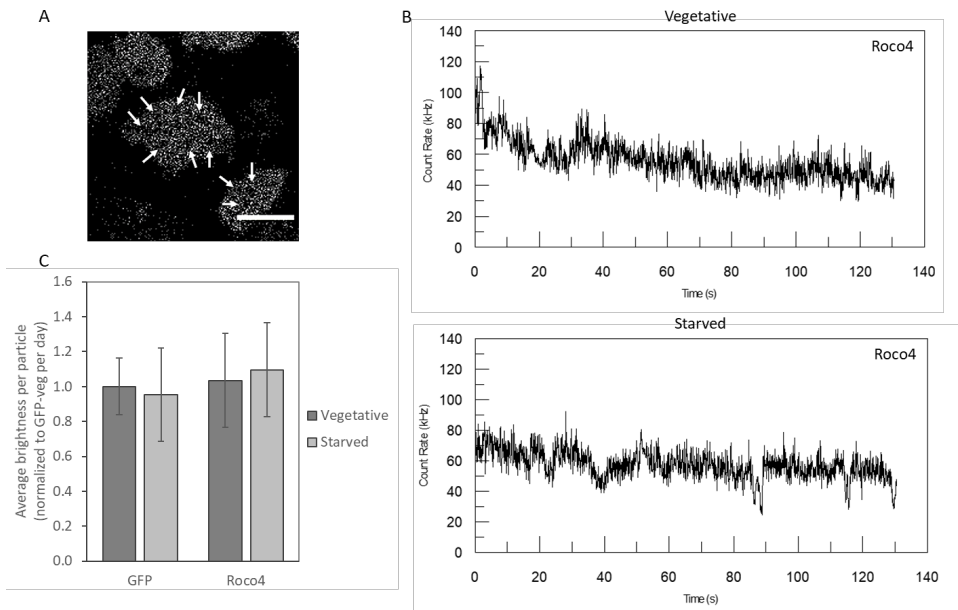


Figure 7: Vegetative and starved Roco4 samples. **A:** Confocal image of vegetative cells expressing the Roco4-GFP construct (180 kD). Arrows indicate vesicles without fluorescent protein. Scale bar: 10 μm **B:** Representative FFS fluorescence fluctuation graphs for vegetative and 5h-starved Roco4 cells. **C:** Brightness results of Roco4-GFP are similar for vegetative and 5 h starved cells. The FCS measurements have been analyzed with the simple brightness analysis. Measurements consisted of 30 traces of 5.5 seconds. Brightness values were normalized to the averaged vegetative samples (GFP-veg) per measurement day. n= 8-22 cells from 2-3 measurement days. Error bars: SD.

From this 2-component ACF fit, the diffusion speed per component was calculated, as well as the fraction of each component (**Figure 8B**). The majority of the Roco4-GFP containing particles had a diffusion coefficient of $7.5 \pm 1.7 \mu\text{m}^2/\text{s}$ ($n=14$) (mean and standard deviation, with the number of independent observations in parenthesis). This is slower than expected when compared to the size and speed of GFP, but this difference can be explained by the putative non-globular shape of Roco4 [43]. The second component consisted of very slowly diffusing particles ($0.23 \pm 0.34 \mu\text{m}^2/\text{s}$ ($n=14$)). Similar patterns were found for mutant versions of Roco4, with very slowly diffusing particles in the second component, which was the minority fraction (**Figure S4B**). This slow diffusion corresponds to a residence time of the order of seconds in the confocal volume, which is even visible as waves in the fluorescence fluctuation signal (**Figure 9A**) and cannot be explained by mere dimerization or oligomerization. Instead, the slow diffusion of this second component may be due to adhesion of Roco4 to vesicles or organelles.

To test whether the slow diffusion could possibly be caused by the adhesion of Roco4 to vesicles, the fluorescence fluctuation signal of Roco4-GFP was compared with that of a mitochondrial membrane protein (GFP-GemATM). Indeed, the seconds-scale fluctuations are also observed for GFP-GemATM (**Figure 9B**), indicating that such fluctuations may be caused by adhesion to cellular membrane structures. Thus, the presence of a slow diffusing component may be caused by its interaction with membrane structures, but does not provide information about the presence of dimers or oligomers. Therefore, diffusion analysis by itself is not suitable to study dimerization, and we continued our research with brightness analysis.

Roco4 is largely monomeric, with a minor slow-diffusing brighter fraction

Initially, the average fluorescence per particle was calculated by simply dividing the total fluorescence intensity per second by the total number of particles in the confocal volume, as calculated from the global analysis of ACF curves (without PCH), called 'simple brightness analysis'. However, the results of this experiment were not satisfactory. The average brightness per particle for Roco4 may be higher than monomeric GFP, but the large standard deviation of this simple method does not allow for conclusions (**Figure 10**). Similarly, the analysis of some Roco4 mutants revealed no significant differences in the average brightness per particle compared to GFP or Roco4 WT (**Figure S4C**). Therefore, the 2-component BDGA analysis was applied to Roco4. As shown above for the analysis of ACF curves only (**Figure 8B**), here again the first component was the fast diffusion component, and the second component represented the slowly diffusing particles. Now with this combined brightness and diffusion analysis, the brightness per component could be determined. Interestingly, the brightness of the first component of Roco4 was similar to monomeric GFP, while the average brightness of the second component was similar to diGFP (**Figure 11A**).

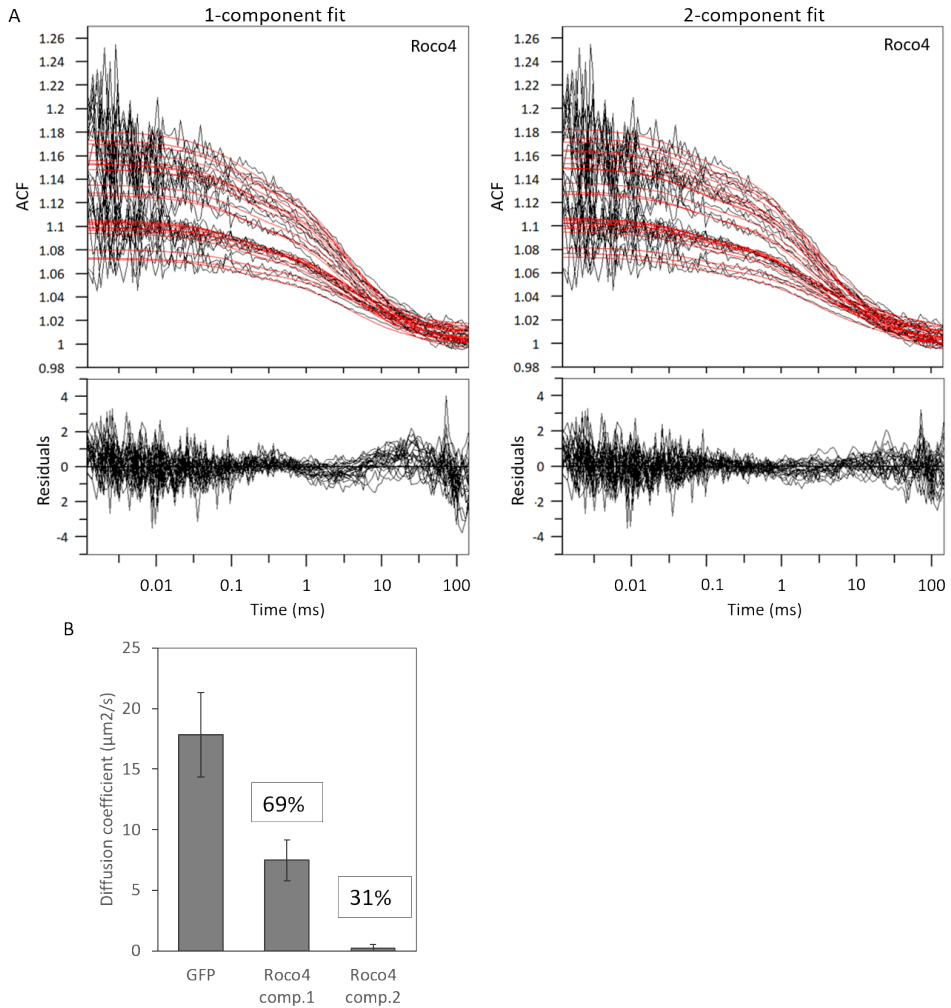


Figure 8: 2-comp FCS diffusion analysis of Roco4. **A:** Autocorrelation curves (ACF) from one representative Dictyostelium cell expressing Roco4-GFP (27 5.5-second traces). Roco4 data yields a better fit when analyzed with a two-component model, as shown by the decreased wave in the residuals. Red line: fit calculated by the free diffusion 3D Gaussian model with triplet-state term [20]. **B:** Diffusion speed of the two Roco4 components, averaged for multiple 5h starved cells. Each measurement consisted of 30 traces of 5.5 seconds. $n = 13-14$ cells from 1-2 measurement days. Error bars: SD.

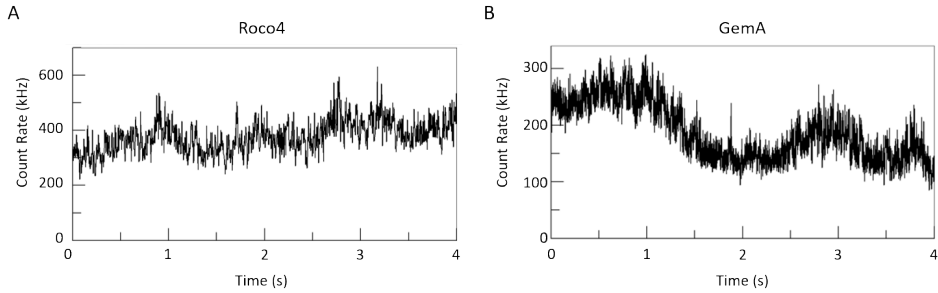


Figure 9: GemA TM fluorescence fluctuates at the seconds time scale, comparable to Roco4. Roco4-GFP and the GFP-coupled trans-membrane domain of GemA were expressed in vegetative Dictyostelium cells and measured with a Zeiss LSM 710 confocal microscope with FCS module.

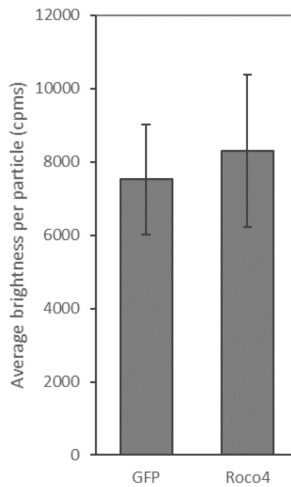
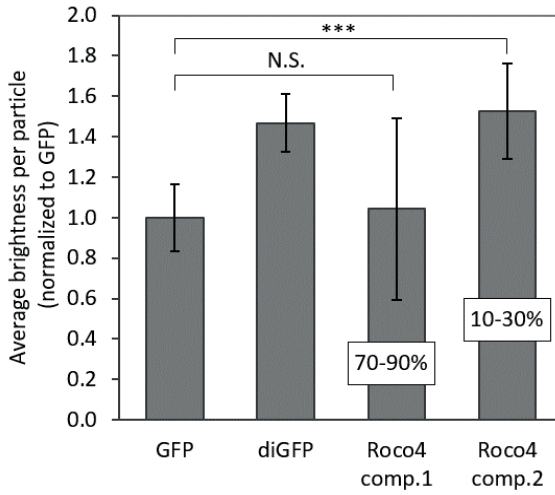
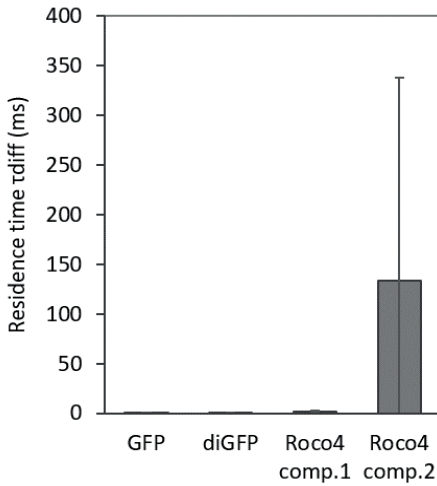


Figure 10: Simple brightness analysis of GFP and Roco4. Total photon counts per second have been divided by the number of particles in the confocal volume (N , calculated from ACF-only), yielding an average brightness per particle. Data from vegetative and 5h-starved Dictyostelium cells have been averaged. Each measurement of a cell consisted of 30 traces of 5.5 seconds. Error bars: SD. $n = 18-32$ cells from 4 measurement days.

A



B



C

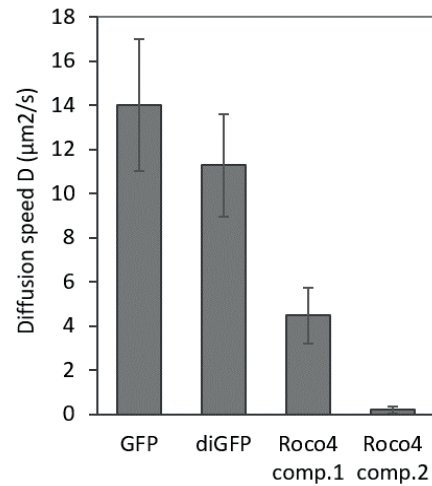


Figure 11: 2-component BDGA analysis of Roco4-GFP in Dictyostelium cells. **A:** 2-component BDGA brightness analysis, normalized to average GFP. Percentages indicate the proportion of the first and second component, varying between cells. **B-C:** BDGA diffusion results represented as the residence time in the confocal volume (**B**) and recalculated into diffusion speed D (**C**). The residence time τ_{diff1} of the first component is fixed with the values from the ACF-only analysis, while the τ_{diff2} of the second component was left free. Each measurement consisted of 26 5-second traces, measured in 5h starved cells. Error bars: SD; *** $p < 0.001$; N.S. non-significant.

The fast diffusing particles from the first component, which was the largest fraction, had a residence time of 2.2 ± 0.6 ms ($n=8$) in the confocal volume ($D_1 = 4.5 \pm 1.3$ $\mu\text{s}^2/\text{s}$ ($n= 8$))(Figure 11B-C). The residence time of the second component varied widely and could even vary between 1.9 ms and 600 ms between the 3-second traces within one cell, with an average D_2 of 0.19 ± 0.16 $\mu\text{s}^2/\text{s}$ ($n= 8$), which is much slower than expected from a freely diffusing dimeric particle and may indicate vesicle binding as suggested above (Figure 9). These diffusion parameters, calculated by the BDGA method, are in the same order of magnitude as with the FCS analysis performed on previous data (Figure 8B). Although the diffusion and brightness of Roco4 mutants did not change much when analyzed with FCS and a simple brightness calculation (Figure S4), testing these variants with the BDGA method could potentially yield more information. Taken together, the diffusion of Roco4 in *Dictyostelium* cells could be split in a large normally diffusing fraction with the brightness of a monomer, and a minor slow diffusing fraction with the brightness of a dimer.

Dimerization of LRRK2 in HEK293 cells analyzed with BDGA

To investigate whether the results for *Dictyostelium* Roco4 could be translated to human LRRK2, similar experiments were performed with GFP-LRRK2 expressed in HEK293 cells. Good results were achieved when HEK293 cells were seeded and transfected on 8-well IBIDI glass slides where they were also measured. The FCS measurements were performed with similar settings as for *Dictyostelium* cells. During the analysis, a 2-component BDGA model was required as for Roco4, as the simple brightness analysis did not yield a significant increase in average brightness of LRRK2 WT and mutants, compared to monomeric GFP (**Figure S5**).

Interestingly, the BDGA results of LRRK2 were very similar to Roco4. The fast-diffusing component 1 represented the largest fraction and had a brightness comparable to monomeric GFP, while the slowly diffusing second component represented the minor fraction (**Figure 12B**). However, the average brightness of the second component was higher than for Roco4-GFP, indicating the presence of higher order oligomers. This is in agreement with the previous observation that LRRK2 is able to form oligomeric filaments, as seen, for example, on microtubule under specific conditions [44]. We did not observe these higher order oligomers when only the RocCOR dimerizing domain was expressed (**Figure 12A-B**).

The diffusion of LRRK2 was also similar to Roco4: component 1 had a diffusion corresponding to a monomer ($\tau_{diff1} = 1.7 \pm 0.38$ ms ($n=10$); $D_1 = 5.0 \pm 1.2$ $\mu\text{s}^2/\text{s}$ ($n=10$)), while component 2 was much slower and displayed a large variation between cells and between traces within a measurement of one cell ($\tau_{diff2} = 51 \pm 50$ ms ($n=10$); $D_2 = 0.39 \pm 0.31$ $\mu\text{s}^2/\text{s}$ ($n = 10$))(**Figure 12C-D**). The first component of the RocCOR domain had a

residence time τ_{diff1} of 0.78 ± 0.13 ms ($n=8$) in the confocal volume ($D_1 = 11 \pm 1.9 \mu\text{s}^2/\text{s}$ ($n=8$)), while the second component was slower ($\tau_{diff2} = 6.8 \pm 2.9$ ms ($n=8$); $D_2 = 1.4 \pm 0.62 \mu\text{s}^2/\text{s}$ ($n=8$)), but clearly not as slow as for LRRK2. Above, we hypothesized that the very slow diffusion of Roco4 (and thus of LRRK2) may be due in part to organelle binding. Whether the moderate diffusion speed of the second component of RocCOR would also be caused by translocation is not clear from these data. To speculate, RocCOR may still bind to some cellular structures or protein complexes, but not to large organelles.

The two components of full-length LRRK2 may represent a cytosolic fraction with a diffusion and brightness corresponding to monomeric LRRK2 and an oligomeric membrane-bound fraction with slow diffusion and high brightness per particle. Under the conditions tested (over expression in HEK293 cells, without any stimulus), the vast majority of the LRRK2 full-length and RocCOR particles were in the monomeric cytosolic form. However, as with Roco4, the possibility that the oligomeric fraction and the diffusion-limited fraction represent different sets of particles cannot be excluded. Cell fractionation and N&B experiments indicate that the membrane fraction is indeed predominantly oligomeric [14,19], suggesting that the oligomeric and diffusion-limited particles likely overlap in our analysis. Strikingly, the percentages of the two components resemble the cytosolic and membrane fractions in the aforementioned biochemical fractionation experiments [14]. In addition, Sanstrum et al. found that the PD-related G2019S variant had increased dimer formation [19]. It would be very interesting to test these and other PD mutations with BDGA. Taken together, our *in cellulo* data support the model that LRRK2 is monomeric in the cytosol and dimeric at membranes.

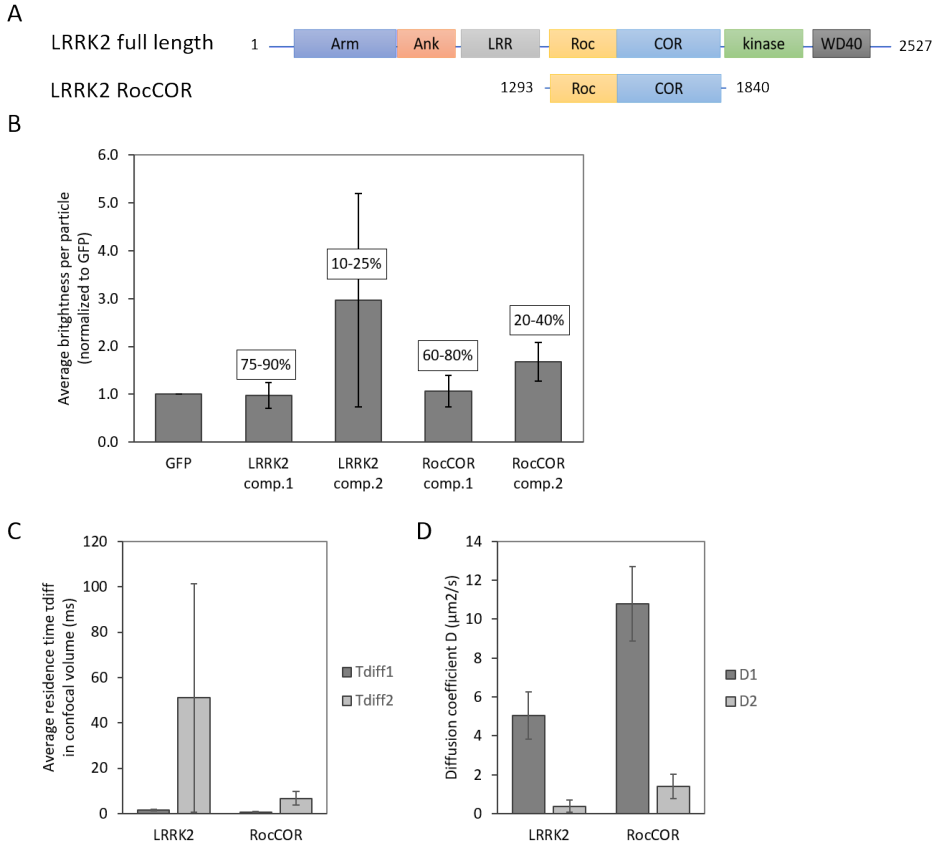


Figure 12: 2-component BDGA analysis of GFP-LRRK2 in HEK293 cells. A: LRRK2 and RocCOR domain structure. **B:** 2-component BDGA brightness analysis of LRRK2 and RocCOR in HEK293 cells, normalized to average GFP. Percentages indicate the proportion of the first and second component, varying between cells. **C-D:** BDGA diffusion results represented as the residence time in the confocal volume (**B**) and recalculated into diffusion speed D (**C**). The residence time τ_{diff1} of the first component is fixed with the values from the ACF-only analysis, while the τ_{diff2} of the second component was left free. Each measurement consisted of 25 5.5-second traces, measured in 5h starved cells. Error bars: SD.

Conclusion and outlook

Our data confirm that our previous developed Brightness and Diffusion Global Analysis of FCS data is a suitable method to study homodimerization in living *Dictyostelium* and HEK cells [20]. Furthermore, to our knowledge, we are the first to apply FKBP12-induced dimerization to *Dictyostelium*, as well as the first to quantitatively measure this FKBP12-mediated dimerization in living cells, as most FKBP12-induced dimerization assays are based on the cellular output of dimerization, rather than on the extent of the dimerization itself.

Previously, the dimerization of various proteins in *Dictyostelium* has been studied using *in vitro* techniques such as protein purification and crystallography, velocity sedimentation of cell extracts, co-immunoprecipitation, NMR, chemical crosslinking, gel filtration, small-angle X-ray scattering, as well as some *in cellulo* techniques such as fluorescence complementation and FCS [45–51], although to our knowledge the latter was only used for heterodimerization of two different molecules [48]. Studying homodimerization *in vivo* has been challenging until now. The single-color BDGA as described here can be a game changer in this regard. Known homodimers such as CAP [45], STAT [46,47], TgrC1 [49], Skp1 [50,51], or some small GTPases can now be studied in more detail, while new ones could be discovered by screening existing GFP-fusion constructs using BDGA.

Subsequently we successfully applied the BDGA methodology to study Roco4 and LRRK2 dimerization in *Dictyostelium* and HEK cells respectively. Our analysis revealed that the majority of both LRRK2 and Roco4 are monomeric, while a small fraction has a brightness comparable to dimeric species. This indicates that both proteins indeed have a monomer-dimer equilibrium. Moreover, the major fraction yielded a diffusion speed comparable to that of cytosolic proteins, while the minor fraction showed a very slow diffusion comparable to vesicle bound proteins.

This is in agreement with previous cellular fractionation experiments, as well as Number and Brightness and PCH analyses, showing LRRK2 is monomeric in the cytosol and oligomeric at membrane structures [14,19,52], and that slowly diffusing LRRK2 is likely vesicle-bound [19]. LRRK2 has been found to localize to mitochondria, lysosomes and other vesicles [30,53,54]. In the case of Roco4, techniques such as cell fractionation and colocalization are needed to investigate which organelles and vesicles Roco4 could bind to. However, since the majority of both Roco4 and LRRK2 are cytosolic, overexpression of GFP fusion constructs would yield a very high cytosolic fraction that could mask any organelle or membrane localization. Developing good antibodies for immunostaining of Roco4, as has been done with LRRK2, could provide more insight in the localization of endogenous Roco4. Alternatively, a knock-in Roco4-GFP strain could be useful in this perspective.

Using our new method, we were now able to demonstrate in living cells that Roco proteins exist in a monomeric cytosolic form and a dimeric/oligomeric membrane-bound form (**Figure 13**). Interestingly, recent proximity biotinylation experiments – which allowed to specifically purify dimeric LRRK2 from cell extracts – show that LRRK2 dimers yielded a higher kinase activity than monomeric LRRK2 [55], indicating that dimerization is an important step in the activation mechanism of Roco proteins. Since BDGA quantifies the equilibrium between cytosolic monomers and diffusion-

impaired dimers/oligomers, any condition that potentially changes this equilibrium could now be tested *in cellulo*. For example, PD mutations or kinase inhibitors may influence the amount of dimer, while LRRK2-recruiting proteins such as 14-3-3 and Rabs usually affect localization. 14-3-3 is responsible for the recruitment of LRRK2 to the cytosol, while some Rabs facilitate the localization of LRRK2 and Roco4 to the trans-Golgi network [31,56–58]. Inhibition, knock-down or mutation of any of these proteins is expected to affect localization and thus possibly the monomer-dimer equilibrium. New factors involved in regulating Roco4/LRRK2 dimerization and recruitment may be identified.

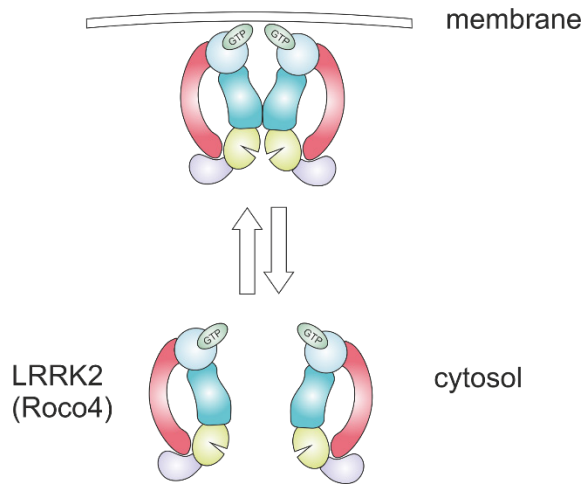


Figure 13: Model of the monomer-dimer equilibrium of LRRK2. In this model, dimerization and translocation are closely connected. LRRK2 and Roco4 are largely monomeric in the cytosol and dimeric at cellular membranes.

In addition, BDGA may be further exploited to investigate to which organelles Roco proteins can bind. LRRK2 has been found to bind to a wide variety of vesicles and organelles, e.g. mitochondria, endosomes and lysosomes (Erb & Moore, 2020; Kuwahara & Iwatsubo, 2020; Singh et al., 2019). Any significant change in the recruitment of LRRK2 to (one of) these structures is expected to influence the brightness, diffusion and/or fraction of the second component in the BDGA analysis. Inhibition of (the migration of) mitochondria, phagosomes or other organelles is expected to reduce the presence of a second component – if indeed Roco4/LRRK2 were to bind to one of these structures. Moreover, Stapled Peptides have recently been developed that inhibit LRRK2 dimerization by binding to the dimerization interface in the RocCOR domain [59]. It would be interesting to use BDGA to investigate the degree of inhibition *in cellulo*, as well as its influence on LRRK2 recruitment to cellular structures.

Taken together, we have now optimized the BDGA methodology to investigate the dimerization of complex proteins such as Roco4 and LRRK2, and our results support the hypothesis that both Roco4 and LRRK2 are monomeric in the cytosol and dimeric at cellular membranes. The cytosolic fraction was the major component, while a minor proportion was most likely bound to vesicles and other structures. Importantly, the 2-component BDGA analysis methodology developed for *Dictyostelium* required no further adaptation for the use in HEK293 cells, suggesting that BDGA is a universal method with broad applicability. For the analysis of proteins that are membrane-bound without a substantial cytosolic fraction, further optimization may be required. Proteins in the outer membrane of the cell are usually measured with Total Internal Reflection Fluorescence (TIRF) microscopy [19,60], and the resulting fluorescence fluctuation data may subsequently be analyzed using BDGA to obtain detailed information about the brightness and diffusion per particle. The application of BDGA in *Dictyostelium* and other cell types could potentially significantly increase our knowledge about dimerization.

Materials and Methods

Cell lines and plasmids

Most experiments were performed with the AX2 axenic cell line from *Dictyostelium*, containing plasmid pDM334hyg GFP-S65T (further referred to as GFP), pDM334 diGFP, pDM334 FKBP12-F36V-GFP (further referred to as FKBP12), pDM363 Roco4-GFP, or pDM488 GFP-gemATM [28]. The construction of pDM334 FKBP12-F36V and pDM334 diGFP has been described in [20]. GemATM consisted of the trans membrane domain of *Dictyostelium* GemA (a.a. 1810-1977).

Handling of *Dictyostelium* cells has been described by [61,62]. Selection of cells containing the expression vectors was performed with 10 µg/mL G418 (pDM334 diGFP and pDM334 FKBP) or 50 µg/mL Hygromycin B (pDM334hyg GFP, pDM363 Roco4, pDM488 gemA). Cells were grown on 100 mm coated Nunc plates supplemented with HL5-C medium (Formedium) at 22°C. These cultures were transferred to shake flasks containing 20 mL HL5-C medium and harvested in the exponential growth phase.

HEK293 cells were maintained in tissue culture treated T75 Greiner flasks in Dulbecco's Modified Eagle medium (DMEM) supplemented with high glucose, 4mM L-Glutamine, 10% fetal bovine serum and 1% antibiotics (Penicillin and Streptomycin). The cells were grown at 37 °C and supplemented with 5% CO₂. The cells were grown until 80% confluency in T75 flask and 30,000 cells were seeded per well in 8-well glass bottom Ibidi slides (Cat.No: 80827). Cells were transfected for 48 hours with 200 ng

of pcDNA 3.0/eGFP, pcDNA 3.1/eGFP-LRRK2 WT or pEGFP_RocCOR [63] using Lipofectamine LTX™ reagent (Invitrogen, Waltham, Massachusetts, USA) according to the protocol given by Invitrogen transfecting plasmid DNA into HEK293 cells used for 1 cm² surface area per well. In order to have low expression of 'GFP only' plasmid, cells were transfected with a mixture of 196 ng of pcDNA 3.0 and 4 ng of pcDNA3/eGFP.

FCS sample preparation and measurement

Before each experiment, 1 µg/mL doxycycline (a tetracycline) was added to the *Dictyostelium* cells and incubated overnight to induce protein expression. The next day, cell densities were determined and 1 × 10⁷ cells were washed in 10mM Na-K-Phosphate buffer (PB, pH=6.5), and resuspended in 1 mL PB. In case of starvation, the suspension was transferred to 10 mL glass jars on a shaker (150 rpm) and starved for 3-6h as indicated. For both starved and vegetative cells, the cell suspension was concentrated or diluted in PB depending on the cell density, and mixed just before the measurement in a 1:1 ratio with a stock solution of 4% DMSO (dissolved in PB). Subsequently, the cells were loaded on an 8-well IBIDI glass slide and allowed to settle down for 3-5 minutes, after which the effect of DMSO was clearly visible (round cells) and the measurements were started.

Induction of dimerization

Dimerization of FKBP12-GFP was induced by the addition of B/B Homodimerizer (Clontech, also called 'Dimerizer' and known as AP20187, closely related to AP1903) to a final concentration of 1-2 µM for the whole starvation period of 3-4 hours as indicated. In the case of 30 min starvation, 1 µM Dimerizer was added to 2.5-hour starved cells for 30 minutes (total starvation time 3 hours). Mock samples without Dimerizer were supplied with 0.2% EtOH (final concentration) before starvation, to compensate for the solvent of the Dimerizer. To some samples B/B Washout Ligand (Clontech, also called splitter) or mock EtOH was added immediately after starvation to a final concentration of 10 µM splitter and 2% EtOH.

FCS measurements in cells

All samples except GemA were measured on a TCS SP8 X SMD system (Leica Microsystems) as described in [20]. The GemA expressing *Dictyostelium* cells were measured on a Zeiss LSM 780 with the same settings as in [20]. Fluorescent images were captured in scanning mode before each FCS measurement. Spots were selected in cells with low brightness, in areas with homogeneous distribution of fluorescence rather than vesicles. Generally, data were obtained from one measurement of 45 seconds per cell in one spot, in 10-20 cells per sample, unless indicated differently. For calibration measurements, Rhodamine 110 (R110, $D = 4.3 \times 10^{-10} \text{ m}^2\text{s}^{-1}$) (Invitrogen, Breda, The Netherlands) in water was used.

FCS Data analysis using the BDGA methodology

Data preparation

The raw FCS data was converted to ACF and PCD before further processing, using the Raw Data Processing tool of the FFS Data Processor 2.6 software (SSTC, Department of Systems Analysis and Computer Modelling, Belarussian State University, Minsk, Belarus, www.sstcenter.com). Each measurement was split into traces of 3-5 seconds, and for each trace one ACF (time step 1×10^{-7} s; point count 140) and three PCDs (point count 32, with different binning times) were calculated, the latter with different time steps depending on the total fluorescence intensity, as described in section 2.1.1 of [20].

Curve fitting with the 1-component BDGA model

All ACF traces from one measurement were loaded in FFS Data Processor and fitted globally with the free diffusion 3D Gaussian model with triplet-state term [20]. Parameters F_{trip} , τ_{trip} , and τ_{diff1} were grouped between traces, and parameter a was fixed to the value found for the Rhodamine 110 calibration dye. An 'Initial Guess' was created and the analysis was run.

Traces with a very aberrant fit were removed (e.g. with a steep drop of the ACF curve on the right side of the plot, or an overall alleviation of the curve compared to the other traces; this correlated with irregularities in the original traces), and the analysis was run again. The chi-square criterion value is a measure for how good the fit is and should be close to 1, but any value between 0.7 and 1.4 is considered acceptable because of the broader shape of the χ^2 distribution at relatively low number of degrees of freedom. Average F_{trip} and τ_{trip} were calculated from all GFP measurements of one day and then fixed for all samples of that day (except diGFP), as recommended in section 2.2 of [20]. Next, all ACF and PCD curves from the same measurement were fitted globally using the protocol described in detail in [64]. The FCS and PCH models are calculated as described in [20]. The parameters τ_{diff1} , F_{trip} , N_1 , and q_1 were grouped per trace (one ACF and three PCDs), while τ_{trip} , a and some other parameters (Fc_1 , τ_{ct} , bg) were grouped for all traces together. Parameters a , F_{trip} , and τ_{trip} were fixed to the same values as fixed during the analysis of only the ACF curves, while τ_{diff1} was fixed to the value found during that ACF analysis. The analysis was run immediately without generation of initial guesses. The fits and residuals of the ACF and the PCD curves were examined for clearly visible aberrations, before the resulting parameter values were exported to a text file.

Curve fitting with the 2-component BDGA model

In the 2-component BDGA analysis, parameters τ_{diff2} , N_2 , and q_2 were added to the model and grouped per trace. F_{trip} , τ_{trip} , τ_{diff1} and a were fixed like in the 1-component

model, but τ_{diff2} was not, to allow for differences between traces.

Post processing of fit results

The resulting data was imported in Microsoft Excel. Average q_1 , q_2 , τ_{diff1} and τ_{diff2} were calculated per cell, and the obtained brightness values were normalized towards GFP on that day. Brightness values as well as diffusion times were averaged between all cells of multiple days. τ_{diff1} and τ_{diff2} were converted into D_1 and D_2 by the formulas given in section 3.1 of [20]. The significance of apparent differences between samples was determined using the standard two-sided t-test with unequal variances. Standard deviations were calculated from the variation between cells, based on the normalized data from multiple measurement days.

Simple brightness analysis

ACF curves were generated and fitted as described for the BDGA analysis. From these results, the average number of particles in the confocal volume (N) was calculated per cell. The total photon count of the measurement was divided by N and the duration of the measurement, resulting in the average counts per molecule per second (cpms).

Acknowledgements

We would like to thank Ahmed Soliman and Franz Ho for providing the LRRK2 and RocCOR plasmids, Wouter van Egmond for the Roco4 constructs, and Jason King for providing the GemA TM plasmid.

Funding

This research was supported by a NWO VIDI, grant number 723.012.108 to AK

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bosgraaf, L.; Van Haastert, P.J.M. Roc, a Ras/GTPase domain in complex proteins. *Biochim. Biophys. Acta - Mol. Cell Res.* **2003**, *1643*, 5–10, doi:10.1016/j.bbamcr.2003.08.008.
2. Wauters, L.; Terheyden, S.; Gilsbach, B.K.; Leemans, M.; Athanasopoulos, P.S.; Guaitoli, G.; Wittinghofer, A.; Gloeckner, C.J.; Versées, W.; Kortholt, A. Biochemical and kinetic properties of the complex Roco G-protein cycle. *Biol. Chem.* **2018**, *0*, doi:10.1515/hsz-2018-0227.
3. Deyaert, E.; Wauters, L.; Guaitoli, G.; Konijnenberg, A.; Leemans, M.; Terheyden, S.; Petrovic, A.; Gallardo, R.; Nederveen-Schippers, L.M.; Athanasopoulos, P.S.; et al. A homologue of the Parkinson's disease-associated protein LRRK2 undergoes a monomer-dimer transition during GTP turnover. *Nat. Commun.* **2017**, *8*, doi:10.1038/s41467-017-01103-4.
4. Zimprich, A.; Biskup, S.; Leitner, P.; Lichtner, P.; Farrer, M.; Lincoln, S.; Kachergus, J.; Hulihan, M.; Uitti, R.J.; Calne, D.B.; et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* **2004**, *44*, 601–7, doi:10.1016/j.neuron.2004.11.005.
5. Ray Dorsey, E.; Elbaz, A.; Nichols, E.; Abd-Allah, F.; Abdelalim, A.; Adsuar, J.C.; Ansha, M.G.; Brayne, C.; Choi, J.Y.J.; Collado-Mateo, D.; et al. Global, regional, and national burden of Parkinson's disease, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **2018**, *17*, 939–953, doi:10.1016/S1474-4422(18)30295-3.
6. Kambey, P.A.; Kanwore, K.; Ayanlaja, A.A.; Nadeem, I.; Du, Y.Z.; Buberwa, W.; Liu, W.Y.; Gao, D. Failure of Glial Cell-Line Derived Neurotrophic Factor (GDNF) in Clinical Trials Orchestrated By Reduced NR4A2 (NURR1) Transcription Factor in Parkinson's Disease. A Systematic Review. *Front. Aging Neurosci.* **2021**, *13*.
7. Tolosa, E.; Vila, M.; Klein, C.; Rascol, O. LRRK2 in Parkinson disease: challenges of clinical trials. *Nat. Rev. Neurol.* **2020**, *16*, 97–107.
8. Lang, A.E.; Espay, A.J. Disease Modification in Parkinson's Disease: Current Approaches, Challenges, and Future Considerations. *Mov. Disord.* **2018**, *33*, 660–677.
9. Paísán-Ruiz, C. LRRK2 gene variation and its contribution to Parkinson disease. *Hum. Mutat.* **2009**, *30*, 1153–1160.
10. Di Maio, R.; Hoffman, E.K.; Rocha, E.M.; Keeney, M.T.; Sanders, L.H.; De Miranda, B.R.; Zharikov, A.; Van Laar, A.; Stepan, A.F.; Lanz, T.A.; et al. LRRK2 activation in idiopathic Parkinson's disease. *Sci. Transl. Med.* **2018**, *10*, eaar5429, doi:10.1126/scitranslmed.aa5429.
11. Steger, M.; Tonelli, F.; Ito, G.; Davies, P.; Trost, M.; Vetter, M.; Wachter, S.; Lorentzen, E.; Duddy, G.; Wilson, S.; et al. Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *Elife* **2016**, *5*, 1–28, doi:10.7554/eLife.12813.001.
12. Fuji, R.N.; Flagella, M.; Baca, M.; S. Baptista, M.A.; Brodbeck, J.; Chan, B.K.; Fiske, B.K.; Honigberg, L.; Jubb, A.M.; Katavolos, P.; et al. Effect of selective LRRK2 kinase inhibition on nonhuman primate lung. *Sci. Transl. Med.* **2015**, *7*, 273ra15-273ra15, doi:10.1126/scitranslmed.aaa3634.
13. Herzig, M.C.; Kolly, C.; Persohn, E.; Theil, D.; Schweizer, T.; Hafner, T.; Stemmelen, C.; Troxler, T.J.; Schmid, P.; Danner, S.; et al. LRRK2 protein levels are determined by kinase function and are crucial for kidney and lung homeostasis in mice. *Hum. Mol. Genet.* **2011**, *20*, 4209–23, doi:10.1093/hmg/ddr348.
14. Berger, Z.; Smith, K.A.; Lavoie, M.J. Membrane localization of LRRK2 is associated with increased formation of the highly active LRRK2 dimer and changes in its phosphorylation. *Biochemistry* **2010**, *49*, 5511–5523, doi:10.1021/bi100157u.
15. Berwick, D.C.; Heaton, G.R.; Azeggagh, S.; Harvey, K. LRRK2 Biology from structure to dysfunction: Research progresses, but the themes remain the same. *Mol. Neurodegener.* **2019**, *14*, 1–22.
16. Terheyden, S.; Nederveen-Schippers, L.M.; Kortholt, A. The unconventional G-protein cycle of LRRK2 and Roco proteins. *Biochem. Soc. Trans.* **2016**, *44*, doi:10.1042/BST20160224.

17. Wu, C.-X.; Liao, J.; Park, Y.; Reed, X.; Engel, V.A.; Hoang, N.C.; Takagi, Y.; Johnson, S.M.; Wang, M.; Federici, M.; et al. Parkinson's disease-associated mutations in the GTPase domain of LRRK2 impair its nucleotide-dependent conformational dynamics. *J. Biol. Chem.* **2019**, jbc.RA119.007631, doi:10.1074/jbc.RA119.007631.
18. Huang, X.; Wu, C.; Park, Y.; Long, X.; Hoang, Q.Q.; Liao, J. The Parkinson's disease-associated mutation N1437H impairs conformational dynamics in the G domain of LRRK2. *FASEB J.* **2019**, *33*, 4814, doi:10.1096/FJ.201802031R.
19. Sanstrum, B.J.; Goo, B.M.S.S.; Holden, D.Z.Y.; Delgado, D.D.; Nguyen, T.P.N.; Lee, K.D.; James, N.G. Fluctuation Imaging of LRRK2 Reveals that the G2019S Mutation Alters Spatial and Membrane Dynamics. *Molecules* **2020**, *25*, doi:10.3390/molecules25112561.
20. Nederveen-Schippers, L.M.; Pathak, P.; Keizer-Gunnink, I.; Westphal, A.H.; Haastert, P.J.M. van; Borst, J.W.; Kortholt, A.; Skakun, V. Combined FCS and PCH Analysis to Quantify Protein Dimerization in Living Cells. *Int. J. Mol. Sci.* **2021**, Vol. 22, Page 7300 **2021**, *22*, 7300, doi:10.3390/IJMS22147300.
21. Magde, D.; Elson, E.L.; Webb, W.W. Fluorescence correlation spectroscopy. II. An experimental realization. *Biopolymers* **1974**, *13*, 29–61, doi:10.1002/bip.1974.360130103.
22. Schwille, P.; Haupts, U.; Maiti, S.; Webb, W.W. Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophys. J.* **1999**, *77*, 2251–2265, doi:10.1016/S0006-3495(99)77065-7.
23. Politz, J.C.; Browne, E.S.; Wolf, D.E.; Pederson, T. Intranuclear diffusion and hybridization state of oligonucleotides measured by fluorescence correlation spectroscopy in living cells. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 6043–6048, doi:10.1073/pnas.95.11.6043.
24. Muller, J.D.; Chen, Y.; Gratton, E. Resolving heterogeneity on the single molecular level with the photon-counting histogram. *Biophys. J.* **2000**, *78*, 474–486, doi:10.1016/S0006-3495(00)76610-0.
25. Chen, Y.; Müller, J.D.; So, P.T.C.; Gratton, E. The photon counting histogram in fluorescence fluctuation spectroscopy. *Biophys. J.* **1999**, *77*, 553–567, doi:10.1016/S0006-3495(99)76912-2.
26. Gilsbach, B.K.; Ho, F.Y.; Vetter, I.R.; Van Haastert, P.J.M.; Wittinghofer, A.; Kortholt, A. Roco kinase structures give insights into the mechanism of Parkinson disease-related leucine-rich-repeat kinase 2 mutations. *Proc. Natl. Acad. Sci.* **2012**, *109*, 10322–10327, doi:10.1073/pnas.1203223109.
27. Liu, Z.; Mobley, J.A.; DeLucas, L.J.; Kahn, R.A.; West, A.B. LRRK2 autophosphorylation enhances its GTPase activity. *FASEB J.* **2016**, *30*, 336–47, doi:10.1096/fj.15-277095.
28. van Egmond, W.N.; Van Haastert, P.J.M. Characterization of the Roco protein family in *Dictyostelium discoideum*. *Eukaryot. Cell* **2010**, *9*, 751–761, doi:10.1128/EC.00366-09.
29. Rosenbusch, K.E.; Oun, A.; Sanislav, O.; Lay, S.T.; Keizer-Gunnink, I.; Annesley, S.J.; Fisher, P.R.; Dolga, A.M.; Kortholt, A. A Conserved Role for LRRK2 and Roco Proteins in the Regulation of Mitochondrial Activity. *Front. Cell Dev. Biol.* **2021**, *0*, 2479, doi:10.3389/FCCELL.2021.734554.
30. Singh, A.; Zhi, L.; Zhang, H. LRRK2 and mitochondria: Recent advances and current views. *Brain Res.* **2019**, *1702*, 96–104, doi:10.1016/j.brainres.2018.06.010.
31. Rosenbusch, K.E. Investigation of Roco protein activity: Hydra's head, University of Groningen, 2021.
32. Spencer, D.M.; Wandless, T.J.; Schreiber, S.L.; Crabtree, G.R. Controlling signal transduction with synthetic ligands. *Science (80-.)* **1993**, *262*, 1019–1024, doi:10.1126/science.7694365.
33. Guan, Y.; Meurer, M.; Raghavan, S.; Rebane, A.; Lindquist, J.R.; Santos, S.; Kats, I.; Davidson, M.W.; Mazitschek, R.; Hughes, T.E.; et al. Live-cell multiphoton fluorescence correlation spectroscopy with an improved large Stokes shift fluorescent protein. *Mol. Biol. Cell* **2015**, *26*, 2054–66, doi:10.1091/mbc.E14-10-1473.
34. Tong, M.; Jiang, Y. FK506-Binding Proteins and Their Diverse Functions. *Curr. Mol. Pharmacol.* **2015**, *9*, 48–65, doi:10.2174/1874467208666150519113541.
35. Zhang, X.; Atala, A.; Godbey, W.T. Expression-targeted gene therapy for the treatment of transitional cell carcinoma. *Cancer Gene Ther.* **2008**, *15*, 543–552, doi:10.1038/cgt.2008.7.

36. Jin, H.; Wang, J.Y.J. Abl Tyrosine Kinase Promotes Dorsal Ruffles but Restrains Lamellipodia Extension during Cell Spreading on Fibronectin. *https://doi.org/10.1091/mbc.e07-01-0085* **2007**, *18*, 4143–4154, doi:10.1091/MBC.E07-01-0085.
37. Soetandyo, N.; Wang, Q.; Ye, Y.; Li, L. Role of intramembrane charged residues in the quality control of unassembled T-cell receptor α -chains at the endoplasmic reticulum. *J. Cell Sci.* **2010**, *123*, 1031, doi:10.1242/JCS.059758.
38. Abell, A.N.; Johnson, G.L. MEKK4 Is an Effector of the Embryonic TRAF4 for JNK Activation *. *J. Biol. Chem.* **2005**, *280*, 35793–35796, doi:10.1074/JBC.C500260200.
39. Veltman, D.M.; Keizer-Gunnink, I.; Haastert, P.J.M. V An extrachromosomal, inducible expression system for Dictyostelium discoideum. *Plasmid* **2009**, *61*, 119–125, doi:10.1016/j.plasmid.2008.11.002.
40. Min, J.; Sridevi, P.; Alexander, S.; Alexander, H. Sensitive cell viability assay for use in drug screens and for studying the mechanism of action of drugs in Dictyostelium discoideum. *Biotechniques* **2006**, *41*, 591–595, doi:10.2144/000112260/ASSET/IMAGES/LARGE/FIGURE3.JPEG.
41. Skakun, V. V.; Hink, M.A.; Digris, A. V.; Engel, R.; Novikov, E.G.; Apanasovich, V. V.; Visser, A.J.W.G. Global analysis of fluorescence fluctuation data. *Eur. Biophys. J.* **2005**, *34*, 323–334, doi:10.1007/s00249-004-0453-9.
42. Clackson, T.; Yang, W.; Rozamus, L.W.; Hatada, M.; Amara, J.F.; Rollins, C.T.; Stevenson, L.F.; Magari, S.R.; Wood, S.A.; Courage, N.L.; et al. Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 10437–10442, doi:10.1073/pnas.95.18.10437.
43. Myasnikov, A.; Zhu, H.; Hixson, P.; Xie, B.; Yu, K.; Pitre, A.; Peng, J.; Sun, J. Structural analysis of the full-length human LRRK2. *Cell* **2021**, *184*, 3519–3527.e10, doi:10.1016/J.CELL.2021.05.004.
44. Deniston, C.; Salogiannis, J.; Mathea, S.; Snead, D.; Lahiri, I.; Matyszewski, M.; Donosa, O.; Watanabe, R.; Böhning, J.; Shiau, A.; et al. Structure of LRRK2 in Parkinson's disease and model for microtubule interaction. *Nature* **2020**, *588*, 344, doi:10.1038/S41586-020-2673-2.
45. Dodatko, T.; Fedorov, A.A.; Grynberg, M.; Patskovsky, Y.; Rozwarski, D.A.; Jaroszewski, L.; Aronoff-Spencer, E.; Kondraskina, E.; Irving, T.; Godzik, A.; et al. Crystal structure of the actin binding domain of the cyclase-associated protein. *Biochemistry* **2004**, *43*, 10628–10641, doi:10.1021/bi049071r.
46. Soler-Lopez, M.; Petosa, C.; Fukuzawa, M.; Ravelli, R.; Williams, J.G.; Müller, C.W. Structure of an activated Dictyostelium STAT in its DNA-unbound form. *Mol. Cell* **2004**, *13*, 791–804, doi:10.1016/S1097-2765(04)00130-3.
47. Fukuzawa, M.; Araki, T.; Adrian, I.; Williams, J.G. Tyrosine phosphorylation-independent nuclear translocation of a Dictyostelium STAT in response to DIF signaling. *Mol. Cell* **2001**, *7*, 779–788, doi:10.1016/S1097-2765(01)00222-2.
48. Knol, J.C.; Engel, R.; Blaauw, M.; Visser, A.J.W.G.; van Haastert, P.J.M. The Phosducin-Like Protein PhLP1 Is Essential for G β Dimer Formation in Dictyostelium discoideum. *Mol. Cell. Biol.* **2005**, *25*, 8393–8400, doi:10.1128/mcb.25.18.8393-8400.2005.
49. Chen, G.; Xu, X.; Wu, X.; Thomson, A.; Siu, C.H. Assembly of the TgrB1-TgrC1 cell adhesion complex during Dictyostelium discoideum development. *Biochem. J.* **2014**, *459*, 241–249, doi:10.1042/BJ20131594.
50. Kim, H.W.; Eletsky, A.; Gonzalez, K.J.; Van Der Wel, H.; Strauch, E.M.; Prestegard, J.H.; West, C.M. Skp1 Dimerization Conceals Its F-Box Protein Binding Site. *Biochemistry* **2020**, *59*, 1527–1536, doi:10.1021/acs.biochem.0c00094.
51. Sheikh, M.O.; Schafer, C.M.; Powell, J.T.; Rodgers, K.K.; Mooers, B.H.M.; West, C.M. Glycosylation of Skp1 affects its conformation and promotes binding to a model F-box protein. *Biochemistry* **2014**, *53*, 1657–1669, doi:10.1021/bi401707y.
52. James, N.G.; Digman, M. a.; Gratton, E.; Barylko, B.; Ding, X.; Albanesi, J.P.; Goldberg, M.S.; Jameson, D.M. Number and brightness analysis of LRRK2 oligomerization in live cells. *Biophys. J.* **2012**, *102*, L41-3, doi:10.1016/j.bpj.2012.04.046.

53. Erb, M.L.; Moore, D.J. LRRK2 and the Endolysosomal System in Parkinson's Disease. *J. Parkinsons. Dis.* **2020**, *10*, 1271–1291.
54. Kuwahara, T.; Iwatsubo, T. The Emerging Functions of LRRK2 and Rab GTPases in the Endolysosomal System. *Front. Neurosci.* **2020**, *14*.
55. Leandrou, E.; Markidi, E.; Memou, A.; Melachroinou, K.; Greggio, E.; Rideout, H.J. Kinase activity of mutant LRRK2 manifests differently in hetero-dimeric vs. Homo-dimeric complexes. *Biochem. J.* **2019**, *476*, 559–579, doi:10.1042/BCJ20180589.
56. Purlyte, E.; Dhekne, H.S.; Sarhan, A.R.; Gomez, R.; Lis, P.; Wightman, M.; Martinez, T.N.; Tonelli, F.; Pfeffer, S.R.; Alessi, D.R. Rab29 activation of the Parkinson's disease-associated LRRK2 kinase. *EMBO J.* **2018**, *37*, 1–18, doi:10.15252/embj.201798099.
57. Liu, Z.; Bryant, N.; Kumaran, R.; Beilina, A.; Abeliovich, A.; Cookson, M.R.; West, A.B. LRRK2 phosphorylates membrane-bound Rabs and is activated by GTP-bound Rab7L1 to promote recruitment to the trans-Golgi network. *Hum. Mol. Genet.* **2018**, *27*, 385–395, doi:10.1093/hmg/ddx410.
58. Nichols, R.J.; Dzamko, N.; Morrice, N.A.; Campbell, D.G.; Deak, M.; Ordureau, A.; Macartney, T.; Tong, Y.; Shen, J.; Prescott, A.R.; et al. 14-3-3 binding to LRRK2 is disrupted by multiple Parkinson's disease-associated mutations and regulates cytoplasmic localization. *Biochem. J.* **2010**, *430*, 393–404, doi:10.1042/BJ20100483.
59. Helton, L.G.; Soliman, A.; Zweydford, F. von; Kentros, M.; Manschwetus, J.T.; Hall, S.; Gilsbach, B.; Ho, F.Y.; Athanasopoulos, P.S.; Singh, R.K.; et al. Allosteric Inhibition of Parkinson's-Linked LRRK2 by Constrained Peptides. *ACS Chem. Biol.* **2021**, doi:10.1021/ACSCHEMBIO.1C00487.
60. Axelrod, D. Cell-substrate contacts illuminated by total internal reflection fluorescence. *J. Cell Biol.* **1981**, *89*, 141–145, doi:10.1083/jcb.89.1.141.
61. Ruchira, Hink, M. a; Bosgraaf, L.; Van Haastert, P.J.M.; Visser, A.J.W.G. Pleckstrin homology domain diffusion in Dictyostelium cytoplasm studied using fluorescence correlation spectroscopy. *J Biol Chem* **2004**, *279*, 10013–10019, doi:10.1074/jbc.M310039200.
62. Fey, P.; Kowal, A.S.; Gaudet, P.; Pilcher, K.E.; Chisholm, R.L. Protocols for growth and development of Dictyostelium discoideum. *Nat. Protoc.* **2007**, *2*, 1307–1316, doi:10.1038/nprot.2007.178.
63. Terheyden, S.; Ho, F.Y.; Gilsbach, B.K.; Wittinghofer, A.; Kortholt, A. Revisiting the Roco G protein cycle. *Biochem. J.* **2014**, doi:10.1042/BJ20141095.
64. Skakun, V. V.; Digris, A. V.; Apanasovich, V. V. Global analysis of autocorrelation functions and photon counting distributions in fluorescence fluctuation spectroscopy. *Methods Mol. Biol.* **2014**, *1076*, 719–741, doi:10.1007/978-1-62703-649-8_33.
65. van Egmond, W.N. Biochemical and functional aspects of GbpC and other Roco proteins in Dictyostelium discoideum, University of Groningen, 2010.

Supplementary figures

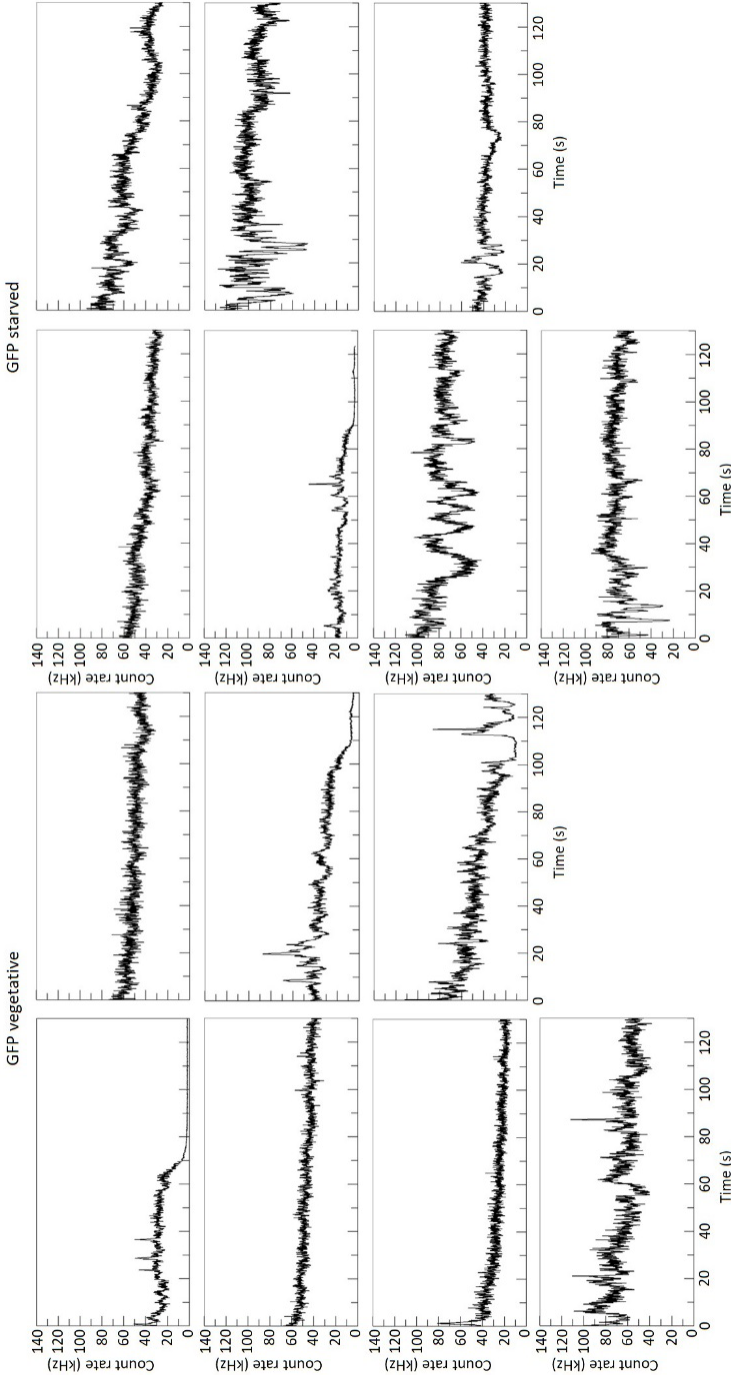
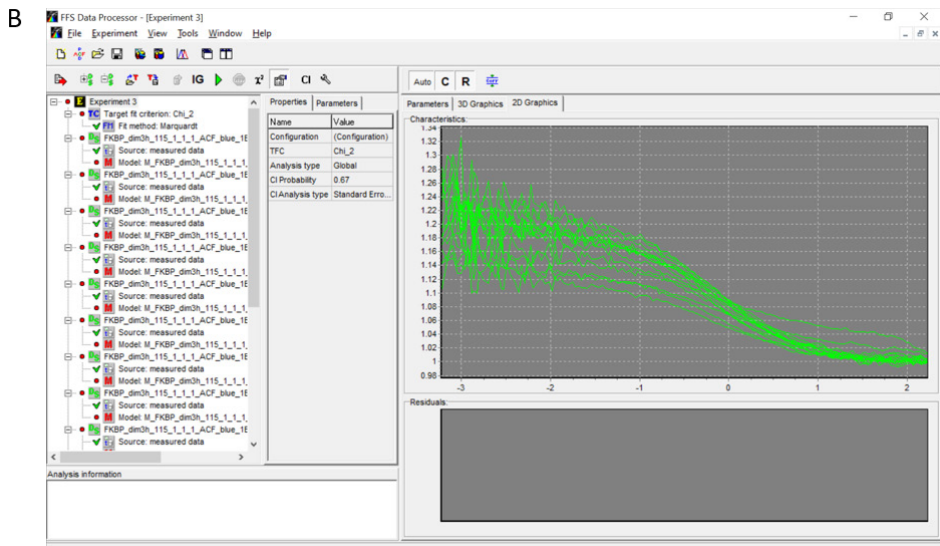
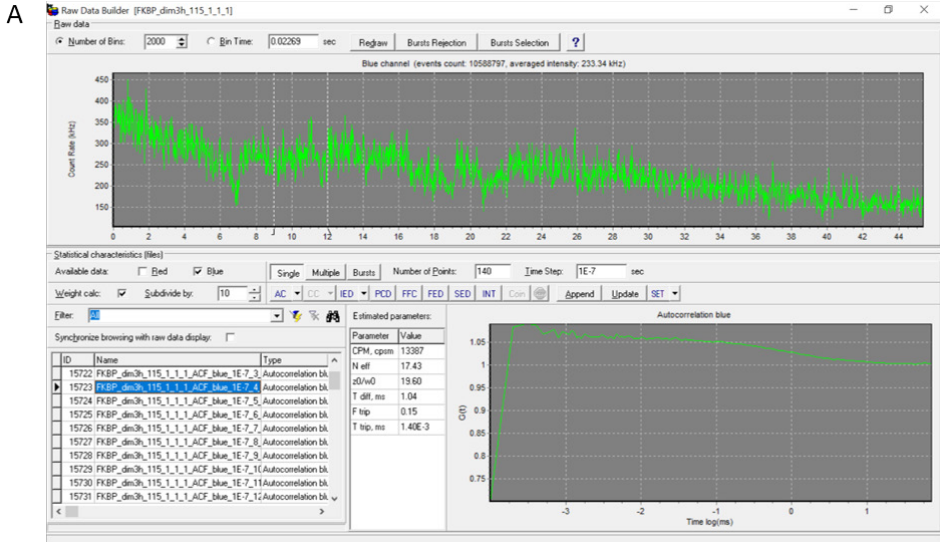


Figure S1: Traces of vegetative and starved cells expressing GFP. Fluorescent traces of vegetative and 5h starved *Dictyostellium* cells expressing monomeric GFP, all measured on the same day. Under both conditions, some cells show more irregularities in the signal than others.



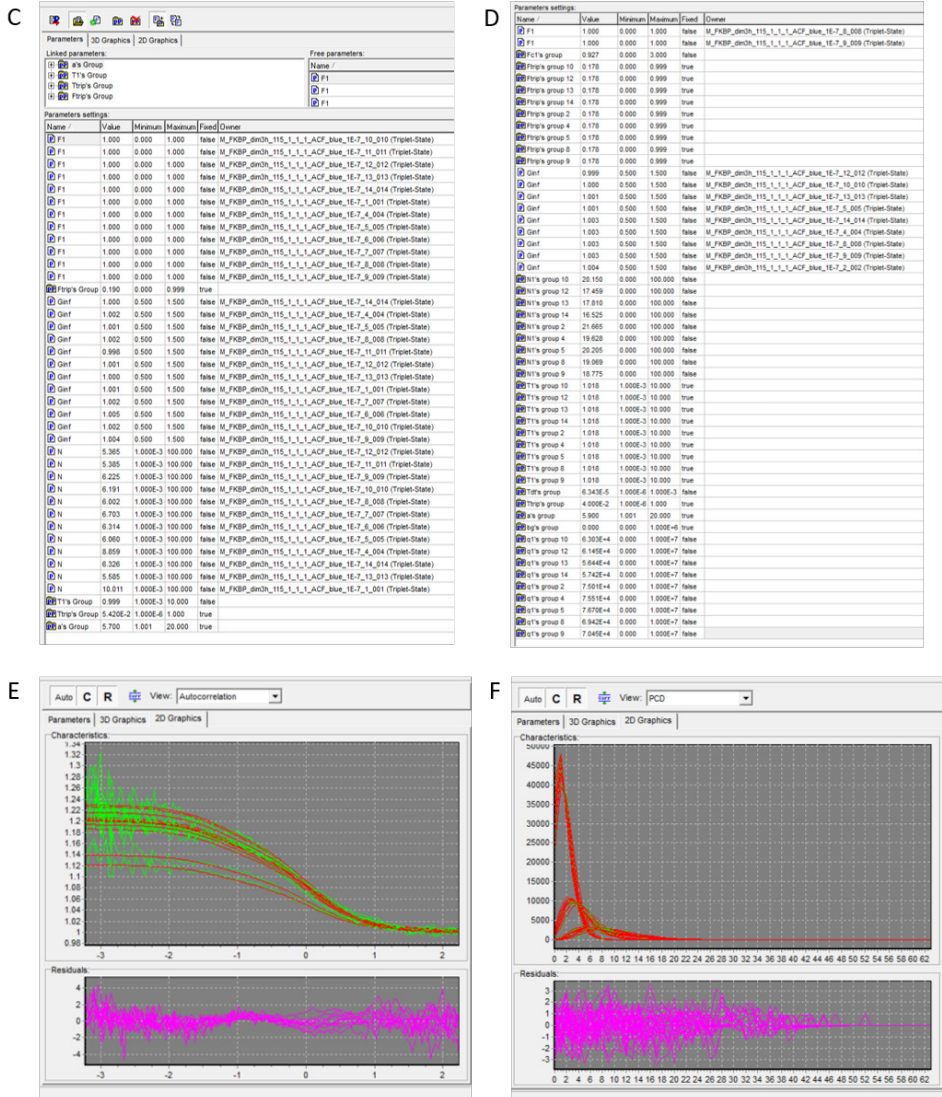


Figure S2: Impression of analysis program FFS Data Processor. Raw data is first imported into the Measurement Database, from which the Raw Data Builder (A) can be opened, where ACF and PCD curves are calculated as described. Next, all ACF curves from one measurement are imported into FFS Data Processor (B). (In this example, some aberrant traces are visible, which are usually removed at this stage.) Before running the ACF analysis, some parameters are grouped and/or fixed (C). Next, ACF and PCD curves are analyzed together in a new global analysis (D), simultaneously fitting the ACF (E) and PCD (F) curves. Green lines: measured data; red lines: theoretical curves obtained after the fit (free diffusion 3D Gaussian model with triplet-state term).

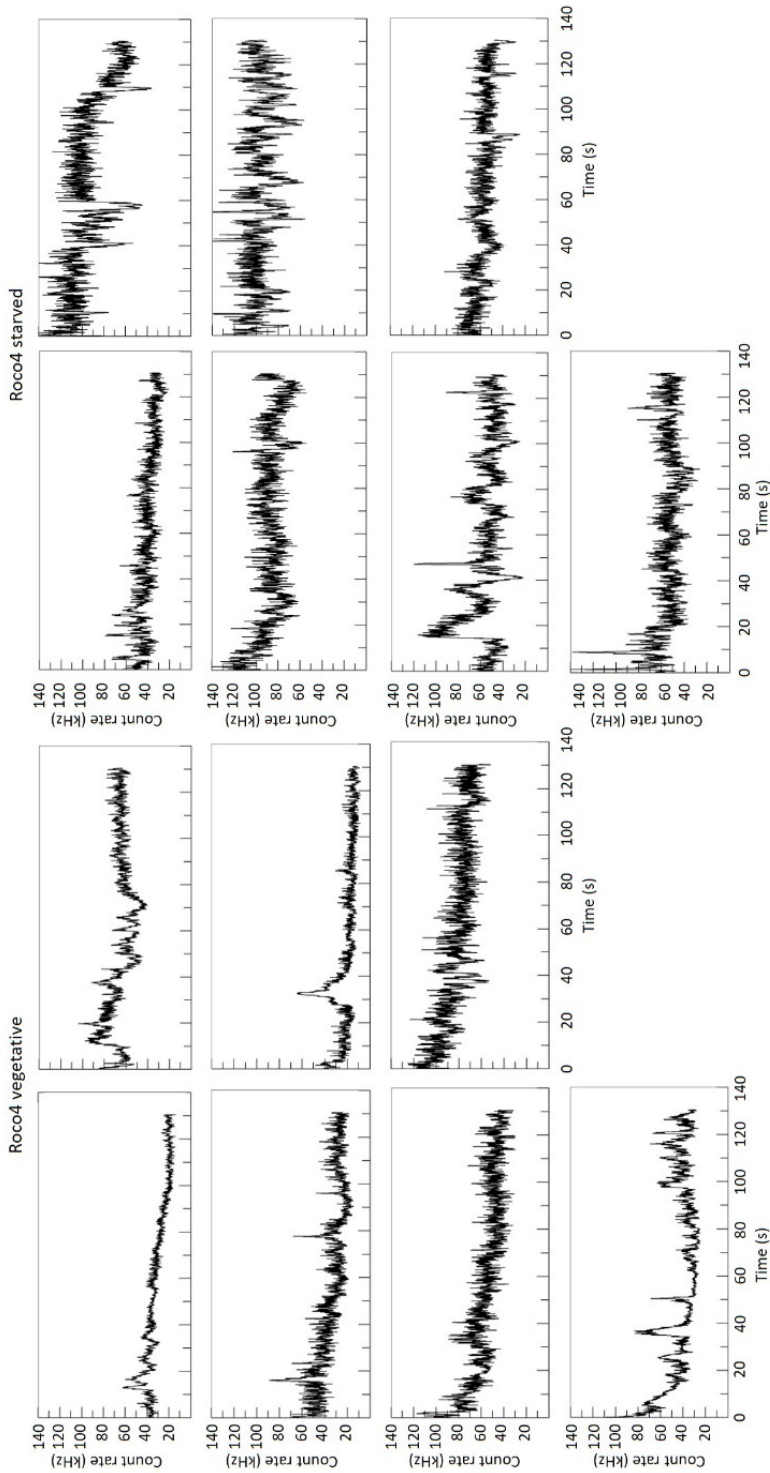


Figure S3: Traces of vegetative and starved cells expressing Roco4-GFP. Fluorescent traces of vegetative and 5h starved Dictyostelium cells expressing Roco4-GFP, all measured on the same day. Under both conditions, some cells show more irregularities in the signal than others.

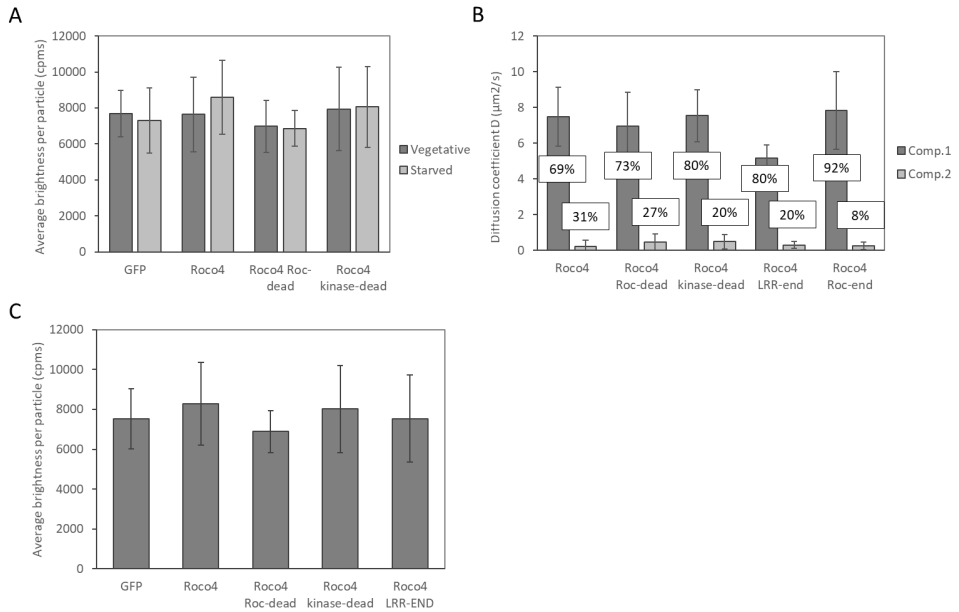


Figure S4: Brightness and diffusion of Roco4 WT and mutant samples. Roco4 WT and mutants have been expressed in Dictyostelium from plasmid pDM363, containing N-terminal GFP [65]. **A:** Simple brightness analysis of vegetative and starved cells. **B:** diffusion coefficients based on 2-comp ACF analysis of Roco4 mutants (starved cells). Percentages of the two components are indicated. For comparison, Roco4 data from Figure 8B have also been included. **C:** Simple brightness analysis of Roco4 mutants, averaged between vegetative and starved cells. For comparison, GFP and Roco4 data from Figure 10 have also been included. All panels are based on the same raw data, although the simple brightness analysis was performed on a smaller subset of samples than the 2-comp ACF analysis (diffusion in panel B). Roco4 LRR-END was only based on starved cells, and therefore omitted in panel A. Roc-dead: Roco4 K383N, unable to bind G-nucleotides; kinase-dead: K1055W; LRR-end: amino acid 244-1726; Roc-end: amino acid 341-1726. Error bars: SD.

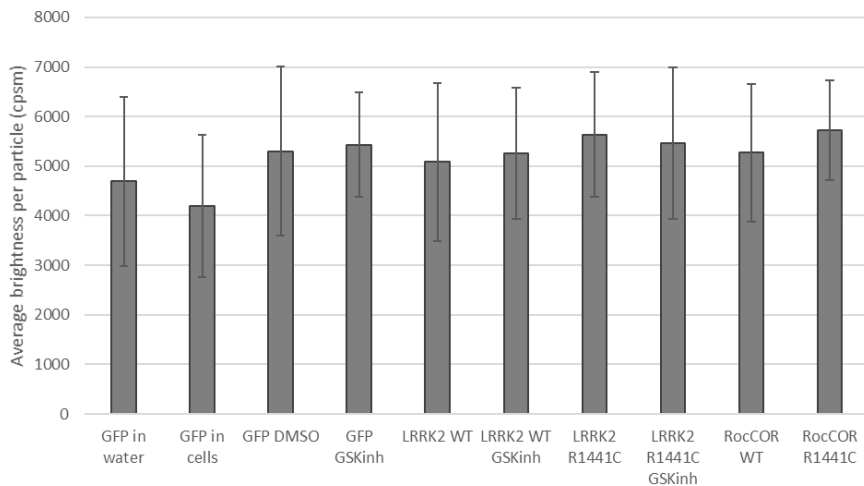
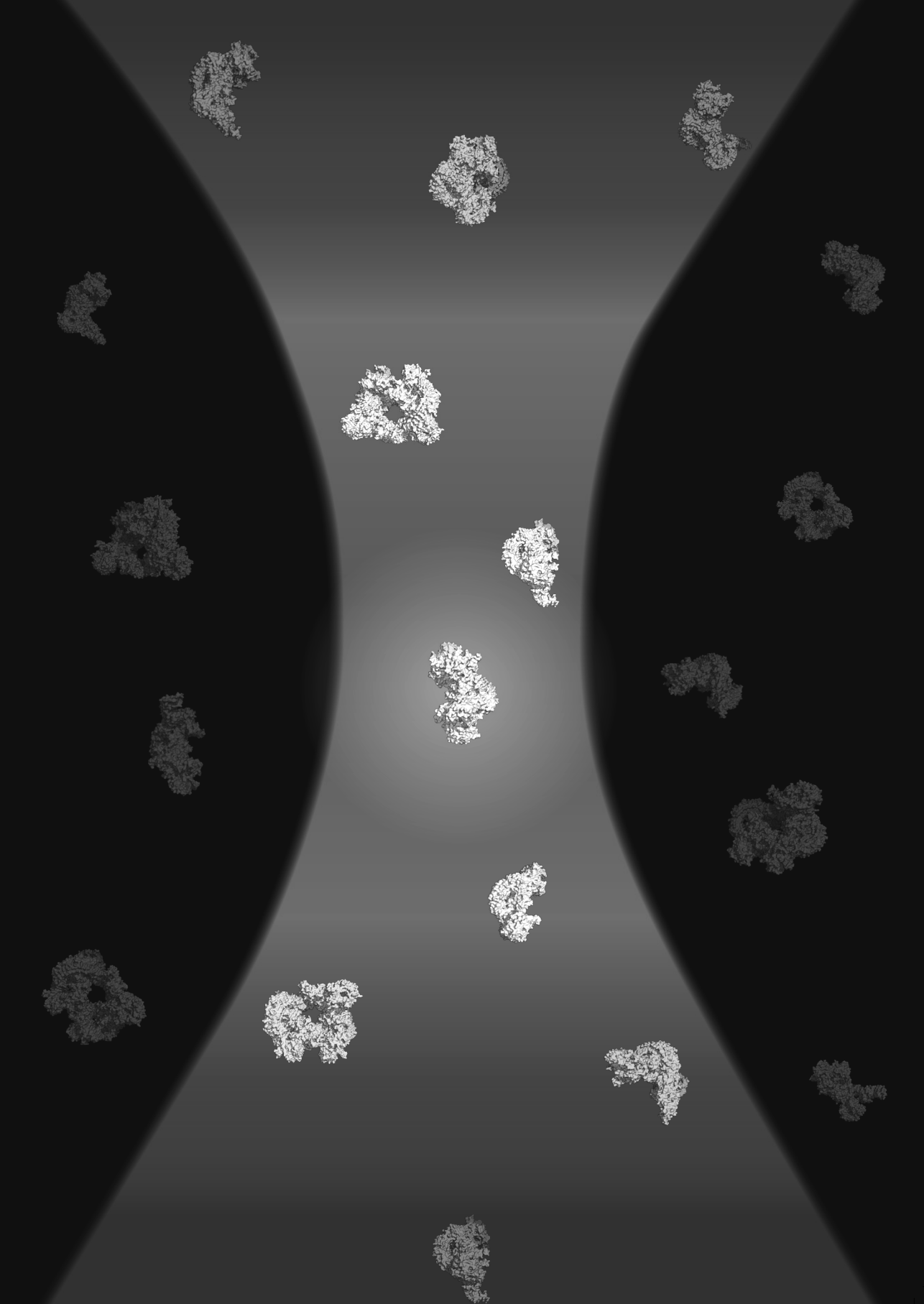


Figure S5: Brightness results for LRRK2 mutants and inhibitor. LRRK2 WT, LRRK2 R1441C, and RocCOR R1441C have been expressed from pcDNA3.1/eGFP, RocCOR WT from pEGFP, and eGFP from pcDNA3.0/eGFP plasmids in the 293T cell line. R1441C point mutations had been generated via quick-change site-directed mutagenesis. GSKinh: 1 μ M GSK2578215A kinase inhibitor (GSK, Brentford, UK) incubated for 90 min; DMSO: control sample with 0.004% DMSO, similar to the final DMSO concentration in the sample with inhibitor; RocCOR: amino acid 1293-1840 from LRRK2. The FCS measurements have been analyzed with a simple brightness analysis. Measurements consisted of 25 traces of 5 seconds. n= 17-38 cells from 2-4 measurement days.



6

Summary and discussion

Laura M. Nederveen-Schippers

Summary and discussion

Protein clustering processes such as dimerization and complex formation play an important role in the activation and function of various proteins in living cells. For example, protein kinases such as ERK2 and RAF become active upon dimerization [1–3]. Complex formation contributes to various diseases, e.g. aggregation of Tau protein in Alzheimer’s disease and alpha-synuclein in Parkinson’s disease. In addition, dimerization is important for the activation of LRRK2, a protein that also contributes to Parkinson’s disease [4,5].

Parkinson’s Disease (PD) is the fastest growing neurodegenerative disease, affecting over 6 million people worldwide (Ray Dorsey et al. 2018). Despite decades of research, no cure has been found, in part because molecular mechanisms leading to neurodegeneration are complex and still poorly understood. The vast majority of cases are idiopathic (not familial), while for a small subset of patients, pathogenic mutations have been identified in multiple genes, including SNCA/ α -synuclein, PINK1, DJ-1 GBA and LRRK2 [6–8].

Our research focuses on the dimerization of LRRK2 (Leucine Rich Repeat Kinase 2), which belongs to the Roco protein family. Roco proteins are characterized by the presence of a Ras-like G-domain called Roc (Ras of complex proteins), immediately followed by a dimerization domain called COR (C-terminal of Roc)[9]. In addition, LRRK2 and other eukaryotic Roco proteins also contain a kinase domain, thus harboring two catalytic domains within one protein. In LRRK2, this ‘catalytic core’ is flanked by multiple protein interaction domains, i.e. Armadillo, Ankyrin and Leucine Rich Repeats (LRR) at the N-terminus, and a WD40 domain at the C-terminus.

In both familial and idiopathic Parkinson’s Disease, LRRK2 has increased kinase activity [10]. However, the exact mechanism by which the activity is regulated remains to be elucidated. The complex structure of LRRK2 allows for multiple protein-protein interactions as well as intra-molecular regulation between domains through G-nucleotide-dependent kinase activity and autophosphorylation.

Canonical small G-proteins such as Ras depend on G-nucleotide Exchange Factors (GEFs) and GTPase Activating Proteins (GAPs) for their GTPase cycle. Roco proteins, on the other hand, seem to function through a different mechanism. In **Chapter 1** we argued that Roco proteins may belong to the GAD (G-proteins activated by nucleotide-dependent dimerization) class of molecules, based on the low nucleotide affinity of the Roc domain, the crystal structures of *C. tepidum* Roco implying *trans*-complementation of the active site, and the increased GTPase activity of the dimer [11–13]. In the GAD

mechanism, the active state of a protein would be a GTP-bound dimer. However, this model has been challenged by the findings in **Chapter 2**, which showed that GTP-bound Roco proteins are predominantly monomeric, while the dimeric proteins were largely GDP-bound or nucleotide free. Based on several observations, a mechanism was proposed in which the GTPase cycle involves both monomerization and dimerization steps. Thus, in this model, monomerization and dimerization are an integral part of the GTPase cycle. Importantly, other reports suggest that the GTPase domain of Roco proteins need to actively cycle to gain full kinase activity, rather than having distinct on- and off-states [12,14]. Taken together, this would imply that a dynamic alternation between dimerization and monomerization is required for kinase activity.

The mechanism described above is based on biochemical and structural data of bacterial Roco proteins. A next step would be to verify these results in the biological context of living cells (*in cellulo*). However, the difference between monomers and dimers cannot be easily distinguished by simple microscopic observation. In recent decades, more sophisticated approaches have been developed based on the visualization of fluorescently labeled proteins. These include Fluorescence Correlation Spectroscopy (FCS), Förster resonance energy transfer (FRET), fluorescence cross-correlation spectroscopy (FCCS), number and brightness analysis (N&B), raster image correlation spectroscopy (RICS), and photon counting histogram (PCH) analysis [15–20]. As discussed in Chapter 4, these techniques are either difficult to implement, or not very accurate at low signal-to-noise ratios, which are typical for living cells. However, previous analyses of FCS data yielded some promising results [21,22]. Therefore, **the aim of this thesis was to optimize the FCS analysis method for measuring dimerization of LRRK2 and other proteins in living cells.**

Measuring dimerization in living cells

Initial estimations of dimerization

FCS is a method to correlate the fluctuations in the fluorescence signal of fluorescently labeled molecules in solution, caused by their diffusion through the confocal observation volume. Based on the generated autocorrelation curves, diffusion speed and particle concentration can be calculated. When the observed total fluorescence intensity in the confocal volume is divided by the average number of particles in the same volume, the average brightness per particle is the result. When compared to the brightness monomeric fluorophores, the presence of dimerization/oligomerization can be observed.

We used this simple method to estimate the per-particle brightness of GFP-bound LRRK2 RocCOR in cell lysate under different G-nucleotide conditions (**Chapter 2**). Similar to the bacterial *in vitro* data, GTP-bound RocCOR was monomeric, while

nucleotide free and GDP-bound RocCOR was partly oligomeric. These initial findings support the idea that for LRRK2, like for bacterial Roco proteins, a similar mechanism of continuous cycling between GTP-bound monomer and GDP-bound dimer may represent the active state.

Principle Component Analysis

In search of a more robust and accurate method, we aimed to apply the combined FCS and PCH global analysis which was previously developed and successively applied to distinguish between the brightness of monomeric and dimeric GFP in vitro [21], as well as to measure the brightness of monomeric GFP in living *Dictyostelium* cells [22]. However, when this technique was initially applied to Roc4, a *Dictyostelium* Roco protein, it was problematic to fit the data. In **Chapter 3**, an extended and adapted data analysis approach was therefore developed, which takes into account the big variability of FCS and PCH curves calculated from the raw fluorescence intensity data of different samples and the relatively low signal to noise ratio typical for the measurements in living cells. Also, through a data mining technique called Principle Component Analysis (PCA), we could show in cell lysate that the PCH curves of tandem-dimer GFP (diGFP) are shaped differently from monomeric GFP, without fitting the data to a mathematical model. Moreover, in different measurements of GFP and diGFP, or even of their mixtures, two clusters were found in the principle components space, indicating that PCH curves representing monomeric and dimeric particles could be easily distinguished, as well as one mixture from another. This PCA based method may be further developed for nonparametric analysis of complex proteins in living cells.

Brightness and Diffusion Global Analysis

To still develop the combined FCS and PCH global analysis for the analysis of Roco proteins in living cells, we optimized this technique in **Chapter 4** for simplified model systems and named it Brightness and Diffusion Global Analysis (BDGA). Compared to our previous global FCS and PCH experiments [21,22], changes were made in the preparation of the samples, global analysis of FCS and PCH curves, and further data processing. Using a series of different mixtures of GFP and diGFP containing lysates, we were able to show that the average brightness per particle calculated by BDGA is proportional to the dimer/monomer ratio present in the sample. Thus, when comparing an unknown sample to the brightness of GFP and diGFP, the proportion of the monomer and dimer can be estimated quantitatively. Next, we demonstrated that living *Dictyostelium* cells expressing diGFP yielded a significantly higher brightness per particle than monomeric GFP. In addition, when GFP was linked to FKBP12, homo-dimerization could be chemically induced [23–26]. Upon addition of ligand to the cells, we showed a clear increase in average brightness per particle, in-between monoGFP and diGFP, indicating the presence of a monomer-dimer equilibrium. From this we

concluded that the developed BDGA methodology is a suitable approach to study the dimerization of GFP-tagged proteins in living cells.

Specific application of BDGA and FKBP12-mediated dimerization in Dictyostelium discoideum

In the first part of **Chapter 5**, we explained in more detail the adaptations made to utilize BDGA and FKBP12-induced dimerization in *Dictyostelium* cells. The FKBP12-induced dimerization method is derived from the iDimerize system from Takara bio, which is usually applied to the dimerization of proteins involved in signal transduction and other processes, and the effect of dimerization is usually assessed by cellular outputs like cell proliferation and kinase assays [27,28]. To our knowledge, this is the first time that the induced dimerization levels of the iDimerize system were quantified in a more direct way. We have performed some initial experiments using different ligand concentrations. This preliminary data suggested that an excess of ligand may impair the total amount of dimerization taking place. More elaborate titration experiments would shed light on this effect, and the optimal ligand concentration, which probably depends on the expression level of the FKBP12-fused protein, may be determined using BDGA. Both techniques – BDGA and FKBP12-induced dimerization – could be used independently of each other and would be of great benefit to the ‘Dicty field’.

Adaptation of BDGA to study the dimerization of Roco proteins and LRRK2

In the second part of **Chapter 5**, this BDGA method was further developed for the application to Roco proteins, after which the dimerization of Roco4 and LRRK2 could be investigated in living cells. Since BDGA had already been developed for *Dictyostelium* and this model organism was easier to culture than human cell lines, it was a logical step to first apply this method to *Dictyostelium* Roco proteins before investigating LRRK2 itself. Of the eleven *Dictyostelium* Roco proteins, Roco4 is the most well studied, has a similar domain architecture to LRRK2, and overlaps with LRRK2 in activation mechanism and function [9,29–33]. The main adjustment was the introduction of a two-component model in the BDGA analysis, which enabled us to assess differences in diffusion, abundance and brightness per particle between the two fractions. Importantly, the two-component BDGA methodology developed for Roco4 in *Dictyostelium* could be applied directly to the FCS data from LRRK2 in HEK293 cells. Thus, we have optimized the BDGA methodology such that it is now possible to study the dimerization of both *Dictyostelium* Roco and human LRRK2 in living cells.

Interestingly, Roco4 and LRRK2 yielded similar results. The first component represented the major fraction of freely diffusing monomeric particles, and the second component represented the minority of diffusion-limited, dimeric particles, which we expect to be vesicle/membrane bound. The presence of monomeric LRRK2/Roco in the cytosol

is in agreement with the model proposed in Chapter 2, where GTP – abundantly present in the cytosol – would cause monomerization. To investigate the hypothesis that the slow diffusion of the second component is caused by adhesion to organelles and vesicles, more research is needed, but some preliminary observations indicated that the diffusion pattern of the Roco proteins resembled that of mitochondria-bound protein. Other reports have shown that membrane association and dimerization are essential for LRRK2 signaling [5,34–37].

Conclusions and outlook

Taken together, we have developed a technique to study the dimerization of LRRK2 in living cells. Our cellular data supports previous *in vitro* observations that LRRK2 is monomeric in the cytosol and dimeric at membranes. The BDGA methodology developed in this thesis is now ready to be deployed to study the LRRK2 activation cycle in more detail, by exploring the influence of relevant mutations, upstream regulatory proteins, and dimerization modifying compounds on the monomer-dimer equilibrium of LRRK2 in living cells.

One of the remaining questions is whether the GTPase cycle, besides monomerization and dimerization, also involves association and dissociation with membranes. FCS-TIRF microscopy of outer cell membrane-associated proteins may shed light on this question. The resulting data can potentially be analyzed with BDGA using a 2D diffusion model, which would then indicate whether some monomer is present at the membranes. If not, then membrane dissociation would be required for the monomerization step during the GTPase cycle. Another approach would be to inhibit dimerization using compounds like the recently developed LRRK2-specific stapled peptides [4]. If monomerization is directly coupled to membrane dissociation, we would not only expect a decrease in the brightness of the slow-diffusing second component, but also in the fraction of this component, when measured with BDGA.

Besides investigating the influence of stapled peptides and other inhibitors on dimerization and translocation, our method may also assess the influence of LRRK2-recruiting proteins, such as Rab29 [34,37], as knockdown of this protein is also expected to result in a decrease in the second component. In addition, the validity of the *in vitro* results of Chapter 2 may be examined in living cells by introducing nucleotide-free and GTP- and GDP-mimetic mutations into the G-nucleotide binding pocket of LRRK2 to investigate the influence of G-nucleotide binding on dimerization. Finally, the influence of Parkinson-related mutations on the monomer-dimer equilibrium and localization of LRRK2 can now be assessed in more detail using BDGA.

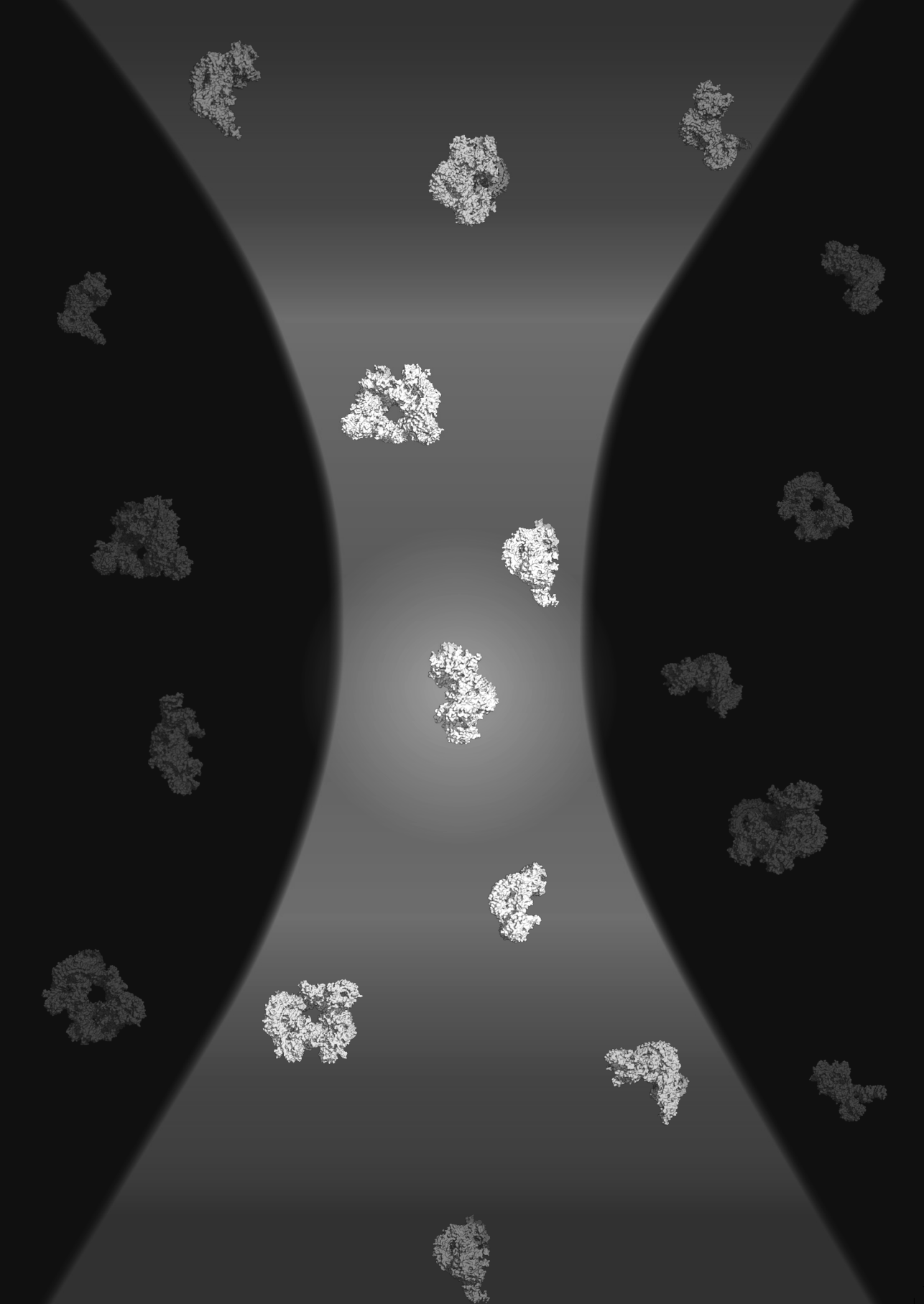
In general, our developed technique can also be applied to other proteins than LRRK2.

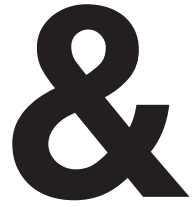
As our setup only requires a standard single-color, single-photon FCS microscope, and standard fluorescent tags like GFP, we expect it to be widely applicable to any fluorescently-labeled cytoplasmic protein with a potential monomer-dimer equilibrium in living cells. In addition, we expect that with a few adjustments, oligomerization of membrane proteins could also be studied using TIRF with BDGA. Compared to individual FCS and PCH analyses, the combined global FCS and PCH analysis of BDGA is superior at lower signal-to-noise ratios, making the analysis more robust and accurate. While two-component BDGA was optimized using Roco4 and LRRK2, we expect that this method is also suitable for other complex proteins which partly adhere to cellular structures. This makes the BDGA method widely applicable to a large variety of proteins.

References

1. Khokhlatchev, A. V.; Canagarajah, B.; Wilsbacher, J.; Robinson, M.; Atkinson, M.; Goldsmith, E.; Cobb, M.H. Phosphorylation of the MAP kinase ERK2 promotes its homodimerization and nuclear translocation. *Cell* **1998**, *93*, 605–615, doi:10.1016/S0092-8674(00)81189-7.
2. Brummer, T.; McInnes, C. RAF kinase dimerization: implications for drug discovery and clinical outcomes. *Oncogene* **2020**, *39*, 4155–4169.
3. Rajakulendran, T.; Sahmi, M.; Lefrançois, M.; Sicheri, F.; Therrien, M. A dimerization-dependent mechanism drives RAF catalytic activation. *Nature* **2009**, *461*, 542–545, doi:10.1038/nature08314.
4. Helton, L.G.; Soliman, A.; Zweydford, F. von; Kentros, M.; Manschwetus, J.T.; Hall, S.; Gilsbach, B.; Ho, F.Y.; Athanasopoulos, P.S.; Singh, R.K.; et al. Allosteric Inhibition of Parkinson's-Linked LRRK2 by Constrained Peptides. *ACS Chem. Biol.* **2021**, doi:10.1021/ACSCHEMBIO.1C00487.
5. Berger, Z.; Smith, K.A.; Lavoie, M.J. Membrane localization of LRRK2 is associated with increased formation of the highly active LRRK2 dimer and changes in its phosphorylation. *Biochemistry* **2010**, *49*, 5511–5523, doi:10.1021/bi100157u.
6. Spatola, M.; Wider, C. Genetics of Parkinson's disease: the yield. *Parkinsonism Relat. Disord.* **2014**, *20 Suppl 1*, S35-8, doi:10.1016/S1353-8020(13)70011-7.
7. Clarimón, J.; Kulisevsky, J. Parkinson's disease: from genetics to clinical practice. *Curr. Genomics* **2013**, *14*, 560–7, doi:10.2174/1389202914666131210212305.
8. Gan-Or, Z.; Liong, C.; Alcalay, R. GBA-Associated Parkinson's Disease and Other Synucleinopathies. *Curr. Neurol. Neurosci. Rep.* **2018**, *18*, doi:10.1007/S11910-018-0860-4.
9. van Egmond, W.N.; Van Haastert, P.J.M. Characterization of the Roco protein family in Dictyostelium discoideum. *Eukaryot. Cell* **2010**, *9*, 751–761, doi:10.1128/EC.00366-09.
10. Di Maio, R.; Hoffman, E.K.; Rocha, E.M.; Keeney, M.T.; Sanders, L.H.; De Miranda, B.R.; Zharikov, A.; Van Laar, A.; Stepan, A.F.; Lanz, T.A.; et al. LRRK2 activation in idiopathic Parkinson's disease. *Sci. Transl. Med.* **2018**, *10*, eaar5429, doi:10.1126/scitranslmed.aar5429.
11. Gotthardt, K.; Weyand, M.; Kortholt, A.; Van Haastert, P.J.M.; Wittinghofer, A. Structure of the Roc-COR domain tandem of *C. tepidum*, a prokaryotic homologue of the human LRRK2 Parkinson kinase. *EMBO J.* **2008**, *27*, 2239–49, doi:10.1038/emboj.2008.150.
12. Rudi, K.; Ho, F.Y.; Gilsbach, B.K.; Pots, H.; Wittinghofer, A.; Kortholt, A.; Klare, J.P. Conformational heterogeneity of the Roc domains in *C. tepidum* Roc-COR and implications for human LRRK2 Parkinson mutations. *Biosci. Rep.* **2015**, *35*, doi:10.1042/BSR20150128.
13. Terheyden, S.; Ho, F.Y.; Gilsbach, B.K.; Wittinghofer, A.; Kortholt, A. Revisiting the Roco G-protein cycle. *Biochem. J.* **2015**, *465*, 139–147, doi:10.1042/BJ20141095.
14. Biosia, A.; Trancikova, A.; Civiero, L.; Glauser, L.; Bubacco, L.; Greggio, E.; Moore, D.J. GTPase activity regulates kinase activity and cellular phenotypes of Parkinson's disease-associated LRRK2. *Hum. Mol. Genet.* **2013**, *22*, 1140–56, doi:10.1093/hmg/dds522.
15. Mahajan, N.P.; Linder, K.; Berry, G.; Gordon, G.W.; Heim, R.; Herman, B. Bcl-2 and bax interactions in mitochondria probed with green fluorescent protein and fluorescence resonance energy transfer. *Nat. Biotechnol.* **1998**, *16*, 547–552, doi:10.1038/nbt0698-547.
16. Bacia, K.; Kim, S.A.; Schwille, P. Fluorescence cross-correlation spectroscopy in living cells. *Nat. Methods* **2006**, *3*, 83–89, doi:10.1038/nmeth822.
17. Nagy, P.; Claus, J.; Jovin, T.M.; Arndt-Jovin, D.J. Distribution of resting and ligand-bound ErbB1 and ErbB2 receptor tyrosine kinases in living cells using number and brightness analysis. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16524–16529, doi:10.1073/pnas.1002642107.
18. Sarkar-Banerjee, S.; Sayyed-Ahmad, A.; Prakash, P.; Cho, K.J.; Waxham, M.N.; Hancock, J.F.; Gorfe, A.A. Spatiotemporal Analysis of K-Ras Plasma Membrane Interactions Reveals Multiple High Order Homo-oligomeric Complexes. *J. Am. Chem. Soc.* **2017**, *139*, 13466–13475, doi:10.1021/jacs.7b06292.

19. Chen, Y.; Wei, L.-N.; Müller, J.D. Probing protein oligomerization in living cells with fluorescence fluctuation spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 15492–7, doi:10.1073/pnas.2533045100.
20. Schwille, P.; Haupts, U.; Maiti, S.; Webb, W.W. Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. *Biophys. J.* **1999**, *77*, 2251–2265, doi:10.1016/S0006-3495(99)77065-7.
21. Skakun, V. V.; Engel, R.; Digris, A. V.; Borst, J.W.; Visser, A.J.W.G. Global analysis of autocorrelation functions and photon counting distributions. *Front. Biosci. - Elit.* **2011**, *3 E*, 489–505, doi:10.2741/e264.
22. Skakun, V. V.; Engel, R.; Borst, J.W.; Apanasovich, V. V.; Visser, A.J.W.G. Simultaneous diffusion and brightness measurements and brightness profile visualization from single fluorescence fluctuation traces of GFP in living cells. *Eur. Biophys. J.* **2012**, *41*, 1055–64, doi:10.1007/s00249-012-0864-y.
23. Amara, J.F.; Clackson, T.; Rivera, V.M.; Guo, T.; Keenan, T.; Natesan, S.; Pollock, R.; Yang, W.; Courage, N.L.; Holt, D.A.; et al. A versatile synthetic dimerizer for the regulation of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 10618–10623, doi:10.1073/pnas.94.20.10618.
24. Clackson, T.; Yang, W.; Rozamus, L.W.; Hatada, M.; Amara, J.F.; Rollins, C.T.; Stevenson, L.F.; Magari, S.R.; Wood, S.A.; Courage, N.L.; et al. Redesigning an FKBP-ligand interface to generate chemical dimerizers with novel specificity. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 10437–10442, doi:10.1073/pnas.95.18.10437.
25. Guan, Y.; Meurer, M.; Raghavan, S.; Rebane, A.; Lindquist, J.R.; Santos, S.; Kats, I.; Davidson, M.W.; Mazitschek, R.; Hughes, T.E.; et al. Live-cell multiphoton fluorescence correlation spectroscopy with an improved large Stokes shift fluorescent protein. *Mol. Biol. Cell* **2015**, *26*, 2054–66, doi:10.1091/mbc.E14-10-1473.
26. Spencer, D.M.; Wandless, T.J.; Schreiber, S.L.; Crabtree, G.R. Controlling signal transduction with synthetic ligands. *Science (80-.)*. **1993**, *262*, 1019–1024, doi:10.1126/science.7694365.
27. L, J.; H, A.; CA, B. Stimulating cell proliferation through the pharmacologic activation of c-kit. *Blood* **1998**, *91*, 890–897.
28. Abell, A.N.; Johnson, G.L. MEKK4 Is an Effector of the Embryonic TRAF4 for JNK Activation *. *J. Biol. Chem.* **2005**, *280*, 35793–35796, doi:10.1074/JBC.C500260200.
29. Gilsbach, B.K.; Ho, F.Y.; Vetter, I.R.; Van Haastert, P.J.M.; Wittinghofer, A.; Kortholt, A. Roco kinase structures give insights into the mechanism of Parkinson disease-related leucine-rich-repeat kinase 2 mutations. *Proc. Natl. Acad. Sci.* **2012**, *109*, 10322–10327, doi:10.1073/pnas.1203223109.
30. Liu, Z.; Mobley, J.A.; DeLucas, L.J.; Kahn, R.A.; West, A.B. LRRK2 autophosphorylation enhances its GTPase activity. *FASEB J.* **2016**, *30*, 336–47, doi:10.1096/fj.15-277095.
31. Rosenbusch, K.E. Investigation of Roco protein activity: Hydra's head, University of Groningen, 2021.
32. Singh, A.; Zhi, L.; Zhang, H. LRRK2 and mitochondria: Recent advances and current views. *Brain Res.* **2019**, *1702*, 96–104, doi:10.1016/j.brainres.2018.06.010.
33. Steger, M.; Tonelli, F.; Ito, G.; Davies, P.; Trost, M.; Vetter, M.; Wachter, S.; Lorentzen, E.; Duddy, G.; Wilson, S.; et al. Phosphoproteomics reveals that Parkinson's disease kinase LRRK2 regulates a subset of Rab GTPases. *Elife* **2016**, *5*, 1–28, doi:10.7554/eLife.12813.001.
34. Purylyte, E.; Dhekne, H.S.; Sarhan, A.R.; Gomez, R.; Lis, P.; Wightman, M.; Martinez, T.N.; Tonelli, F.; Pfeffer, S.R.; Alessi, D.R. Rab29 activation of the Parkinson's disease-associated LRRK2 kinase. *EMBO J.* **2018**, *37*, 1–18, doi:10.15252/embj.201798099.
35. Sen, S.; Webber, P.J.; West, A.B. Dependence of leucine-rich repeat kinase 2 (LRRK2) kinase activity on dimerization. *J. Biol. Chem.* **2009**, *284*, 36346–36356, doi:10.1074/jbc.M109.025437.
36. Schapansky, J.; Nardozi, J.D.; Felizia, F.; Lavoie, M.J. Membrane recruitment of endogenous LRRK2 precedes its potent regulation of autophagy. *Hum. Mol. Genet.* **2014**, doi:10.1093/hmg/ddu138.
37. Liu, Z.; Bryant, N.; Kumaran, R.; Beilina, A.; Abeliovich, A.; Cookson, M.R.; West, A.B. LRRK2 phosphorylates membrane-bound Rabs and is activated by GTP-bound Rab7L1 to promote recruitment to the trans-Golgi network. *Hum. Mol. Genet.* **2018**, *27*, 385–395, doi:10.1093/hmg/ddx410.





ADDENDUM

**Nederlandse samenvatting
Acknowledgements
List of publications
Curriculum vitae**

Nederlandse samenvatting

voor geïnteresseerd publiek

Dit proefschrift beschrijft de ontwikkeling van een methode om de zogenaamde ‘dimerisatie’ (paarvorming) van eiwitten te meten. Dit is onder andere van belang voor het eiwit LRRK2 dat betrokken is bij de ziekte van Parkinson en waar onze onderzoeksgroep uitgebreid onderzoek naar doet. De fundamentele inzichten die voortvloeien uit dit onderzoek dragen bij aan de ontwikkeling van medicijnen die de ziekte van Parkinson kunnen remmen of zelfs voorkomen.

De ziekte van Parkinson

Parkinson is een neurodegeneratieve ziekte die in Nederland ongeveer één procent van de zestigplussers treft en die bij een groeiend aantal mensen voorkomt. Bij parkinsonpatiënten sterven dopamine-producerende hersencellen langzaam af, wat op den duur leidt tot stijfheid, trillen en moeite met bewegen, alsook cognitieve problemen zoals concentratieverlies en slaperigheid. De huidige medicijnen bestrijden tijdelijk de symptomen, maar kunnen niet voorkomen dat steeds meer neuronen afsterven.

LRRK2

De oorzaak van de ziekte van Parkinson is bij de meeste patiënten onbekend, maar in sommige families zijn genetische afwijkingen gevonden, onder andere in het LRRK2-gen. LRRK2 (spreek uit: lark-two) is een relatief groot eiwit dat behoort tot de Roco familie en bestaat uit meerdere domeinen (onderdelen) met verschillende functies (Figuur 1). LRRK2 kan aan andere eiwitten binden via de Ankyrin-, LRR- en WD40-domeinen, het kan een fosfor-groep vastmaken aan andere eiwitten en aan zichzelf via het kinase domein, en het kan GTP-moleculen omzetten in GDP via het ROC-domein. Tot slot bevat LRRK2 een COR-domein dat essentieel is voor de stabiliteit van LRRK2 ‘dimeren’ (paartjes). Meer informatie over de functie van deze domeinen en hun onderlinge interacties is te vinden in hoofdstuk 1.



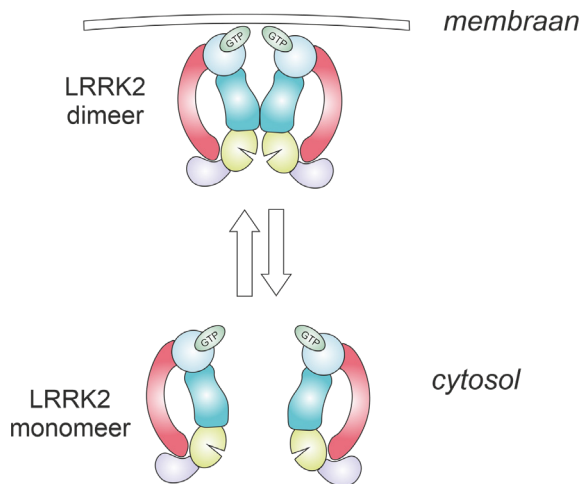
Figuur 1: Domeinstructuur van de Roco eiwitten LRRK2 en Roco4.

In al deze verschillende LRRK2-domeinen zijn mutaties (genetische veranderingen) gevonden die parkinson kunnen veroorzaken. Het onderliggende mechanisme hoe die

mutaties uiteindelijk leiden tot celdood is echter nog niet duidelijk. Eerder onderzoek heeft uitgewezen dat de meeste mutaties leiden tot een verhoogde activiteit van het kinase domein en ook tot een verhoogde dimerisatie. Interessant genoeg is deze kinase-activiteit ook bij parkinsonpatiënten zonder LRRK2-mutaties hoger dan in gezonde mensen. Opheldering van het activatiemechanisme van LRRK2 is dus belangrijk om het ontstaan van parkinson te begrijpen en voor het ontwerpen van nieuwe medicijnen die de hyperactiviteit van LRRK2 tegengaan.

Het LRRK2 activeringsmechanisme onderzoeken

Experimenten met opgezuiverd LRRK2 eiwit en in cel-lysaten (de inhoud van kapotgemaakte cellen) hebben aangetoond dat LRRK2 voornamelijk monomeer is in het cytosol en dimeer aan membraanstructuren van de cel (Figuur 2), en dat het dimeer een hogere kinase activiteit heeft dan een monomeer. Dimerisatie is dus relevant voor het activatiemechanisme van LRRK2, en medicijnen die ervoor zorgen dat er minder dimeer is kunnen dus worden gebruikt voor de verlaging van de kinaseactiviteit in parkinsonpatiënten.



Figuur 2: Model van het monomeer-dimeer evenwicht van LRRK2 (en Roco4). Dit model is gebaseerd op experimenten met opgezuiverde eiwitten en met de cel-inhoud van gebroken cellen. Experimenten in dit proefschrift suggereren dat dit model ook geldt voor LRRK2 in levende cellen.

Bij aanvang van dit onderzoek werd daarom de vraag gesteld hoe de transitie tussen inactief cytosolisch monomeer en actief membraangebonden dimeer gereguleerd wordt. We wilden daarbij met een lichtmicroscop in levende cellen kijken in plaats van te werken met lysaten of opgezuiverd eiwit, zodat de cellulaire processen tijdens het

experiment door konden gaan. Het bestuderen van de dimerisatie van LRRK2 in levende cellen was echter lastig en nog niet eerder gedaan. Wel was onze onderzoeksgroep al betrokken bij de ontwikkeling van een microscooptechniek waarbij dimeren van een klein opgezuiverd eiwit gemeten konden worden alsook monomeren in levende cellen. Het meten van dimeren in levende cellen was dan ook een logische volgende stap.

‘Spectroscopie’ met de microscoop

De betreffende techniek was een combinatie van Fluorescence Correlation Spectroscopy (FCS) en Photon Counting Histogram (PCH) analyse (FCS+PCH). Daarbij worden eiwitten (die te klein zijn om met de golflengte van zichtbaar licht waar te kunnen nemen) voorzien van een fluorescent label dat licht in een andere golflengte uitzendt dan waar je het mee beschijnt. Door met een filter alleen naar de uitgezonden golflengte te kijken neem je alleen de locatie van het betreffende eiwit waar. Hiermee kunnen plaatjes gemaakt worden, maar in het geval van FCS wordt de laser op één vast punt gericht in de cel – het confocale volume, schematisch afgebeeld op de cover – en wordt de lichtintensiteit in dat kleine punt gemeten in de tijd. Door diffusie (de beweging van moleculen door een vloeistof) bewegen de eiwitten in en uit het confocale volume, wat resulteert in fluctuaties in het fluorescente signaal. Met FCS en PCH kan vanuit dat signaal de *‘brightness per particle’* (helderheid per deeltje) berekend worden. Aangezien een dimeer bestaat uit twee eiwitten met twee fluorescente labels, schijnt het twee keer zo helder als een monomeer en is daarmee als zodanig te onderscheiden.

Ontwikkeling BDGA-methode

Bij aanvang van dit promotieonderzoek werden de beschreven metingen in eerste instantie uitgevoerd op Roco4 eiwitten in levende *Dictyostelium*-cellen. Roco4 en LRRK2 behoren tot dezelfde Roco eiwitfamilie en komen wat betreft domeinstructuur grotendeels overeen. *Dictyostelium* (specifiek *Dictyostelium discoideum*) is een eencellige amoëbe (slijmschimmel) die normaal gesproken in de grond voorkomt en bij voedselschaarste samentrekt tot een meercellige structuur met sporen die extreme omstandigheden kunnen doorstaan. *Dictyostelium* is gemakkelijker te kweken dan menselijke cellijnen en wordt daarom regelmatig als modelorganisme gebruikt. Onze groep had Roco4 al eerder gebruikt als model voor LRRK2, en *Dictyostelium* al eerder voor bovengenoemde ‘FCS+PCH’-experimenten. Het lag dan ook voor de hand om deze methode eerst toe te passen op *Dictyostelium* Roco-eiwitten alvorens LRRK2 zelf te onderzoeken. Het meten en vooral het analyseren van Roco4 (met Green Fluorescent Protein (GFP) als fluorescent label) in *Dictyostelium* was echter niet eenvoudig. Het beschreven promotieonderzoek heeft zich daarom met name gericht op het optimaliseren van deze metingen en analyses.

Het probleem met de gecombineerde 'FCS+PCH'-analyse van de Roco4 metingen in *Dictyostelium* was met name dat een goede 'fit' met het theoretische model (een set formules waarmee eigenschappen zoals diffusiesnelheid en *brightness per particle* berekend kunnen worden) niet gemaakt kon worden vanwege de ruis die onder andere optreedt als grote eiwitten in levende cellen interacties aangaan met hun omgeving en daardoor onvoorspelbare diffusiepatronen volgen. Een manier om informatie uit de FCS-data te halen zonder een dergelijke 'fit' is via *data mining* technieken zoals Principle Component Analysis (PCA). In hoofdstuk 3 wordt hiervoor een aanzet gegeven. Via PCA kon een onderscheid gemaakt worden tussen monomeer GFP (monoGFP) en dimeer GFP (diGFP) in cel-lysaten. Deze methode kan nu verder ontwikkeld worden voor de analyse van complexere eiwitten in levende cellen.

Om alsnog de gecombineerde FCS+PCH-analyse methode te ontwikkelen voor Roco eiwitten in levende cellen, hebben we deze techniek in hoofdstuk 4 eerst geoptimaliseerd voor simpelere eiwitten en deze Brightness and Diffusion Global Analysis (BDGA) genoemd. Door mengsels te maken van lysaten met een verschillende verhouding tussen monoGFP en diGFP, konden we aantonen dat met BDGA de gemiddelde fluorescentie per deeltje een goede maat is voor het aandeel dimeren in het monster. Dus bij het vergelijken van een onbekend monster met de helderheid van GFP en diGFP, kan het aandeel van het monomeer en dimeer kwantitatief worden geschat. Voor metingen van dimerisatie in levende *Dictyostelium*-cellen hebben we eerst aangetoond dat de cellen die diGFP tot expressie brengen een significant hogere helderheid per deeltje opleverden dan cellen met monoGFP.

Om een beter model te krijgen voor een monomeer-dimeer evenwicht (mengsel) zoals dat in de natuur voorkomt hebben we gebruikgemaakt van het FKBP12 eiwit met GFP label. Na toevoeging van een bepaald ligand (een molecuul dat aan een eiwit bindt) verandert FKBP12 gedeeltelijk van een monomeer in een dimeer, waardoor je de gemiddelde helderheid van de monomeren (zonder ligand) kunt vergelijken met dat van een mengsel/evenwicht van monomeer en dimeer. En inderdaad, met BDGA namen we een duidelijke toename waar in de gemiddelde helderheid per deeltje – tussen de controlemetingen van monoGFP en diGFP in – wat wijst op de aanwezigheid van een monomeer-dimeer evenwicht. Hieruit concluderen we dat de ontwikkelde BDGA-*methodology* (manier van toepassen van een bestaande methode) geschikt is om de dimerisatie van GFP-gelabelde eiwitten in levende cellen te bestuderen. Een meer gedetailleerde beschrijving van de optimalisatie van de FKBP12-experimenten in *Dictyostelium* is te vinden in het eerste deel van hoofdstuk 5. Beide technieken – BDGA en FKBP-geïnduceerde dimerisatie – kunnen onafhankelijk van elkaar worden gebruikt en zouden van groot nut zijn voor het 'Dicty'-onderzoeksveld.

Dimerisatie van Roco4 en LRRK2

Om initiële informatie te krijgen over de dimerisatie van LRRK2, hebben we in hoofdstuk 2 gebruik gemaakt van (niet-levende) lysaten van humane cellen en de FCS-metingen op een vereenvoudigde manier geanalyseerd. Hierbij werd gebruik gemaakt van een eenvoudiger methode dan BDGA: de totale fluorescentie-intensiteit in het confocale volume werd gedeeld door het gemiddelde aantal eiwitten in datzelfde volume. Hierbij vonden we dat het Roc-COR domein van LRRK2 in aanwezigheid van GppNHp (een analoog van GTP, dat door het Roc domein omgezet kan worden in GDP middels de zogenaamde GTPase cyclus) gemiddeld dezelfde helderheid heeft als GFP-monomeren, terwijl de gemiddelde helderheid per deeltje toenam in aanwezigheid van GDP, wat duidt op de aanwezigheid van dimeren of hogere orde oligomeren (deeltjes van meer dan twee eiwitten aan elkaar). Dit komt overeen met de andere bevindingen in hoofdstuk 2, waarbij opgezuiverde Roco eiwitten voornamelijk monomeer waren in de GTP-gebonden toestand en dimeer in de GDP-gebonden of nucleotide-vrije toestand. Aangezien de GTPase cyclus van het Roc domein belangrijk is voor de regulatie van het kinase domein, zou dit betekenen dat de dimerisatie tijdens deze cyclus een belangrijke stap is in het activatiemechanisme van LRRK2.

In het tweede deel van hoofdstuk 5 is de BDGA-methode verder ontwikkeld voor toepassing op Roco-eiwitten in levende cellen, met behulp van Roco4 in *Dictyostelium*. De belangrijkste aanpassing was de introductie van een zogenaamd twee-componentenmodel in de BDGA-analyse, waarmee de helderheid van langzaam en snel diffunderende deeltjes konden worden onderscheiden. Vervolgens konden we de ontwikkelde twee-componenten BDGA-methode rechtstreeks toepassen op de FCS-gegevens van LRRK2 in HEK293-cellen. Dit betekent dat we BDGA nu zo geoptimaliseerd hebben dat het mogelijk is om de dimerisatie van zowel *Dictyostelium* Roco als menselijk LRRK2 in levende cellen te bestuderen.

Nu de BDGA-methode geoptimaliseerd was, kon de dimerisatie van Roco4 en LRRK2 in levende cellen worden onderzocht. Interessant is dat Roco4 en LRRK2 vergelijkbare resultaten opleverden. In de twee-componentenanalyse vertegenwoordigde de eerste component de grootste fractie van vrij diffunderende monomere deeltjes, en de tweede component vertegenwoordigde de minderheid van diffusie-beperkte, dimere deeltjes, die waarschijnlijk gebonden zijn aan membraanstructuren zoals organellen en **vesicles** (blaasjes) in de cel. Dit bevestigt de eerdergenoemde hypothese dat LRRK2 monomeer is in het cytosol en dimeer aan membranen (Figuur 2).

Conclusies en vooruitzichten

Samengevat hebben we een techniek ontwikkeld om de dimerisatie van LRRK2 in levende cellen te bestuderen. De BDGA-methode die in dit proefschrift is ontwikkeld,

is nu klaar om te worden ingezet om de LRRK2-activeringscyclus in meer detail te bestuderen, bijvoorbeeld door in levende cellen de invloed te onderzoeken van mutaties, LRRK2-bindende eiwitten en andere moleculen die het monomeer-dimeer-evenwicht van LRRK2 zouden kunnen beïnvloeden.

Daarnaast kan onze ontwikkelde techniek ook worden toegepast op andere eiwitten dan LRRK2. Omdat onze opstelling alleen een standaard FCS-microscoop en standaard fluorescerende labels (zoals GFP) vereist, verwachten we dat deze breed toepasbaar zal zijn en dat in principe elk cytosolisch eiwit met een (potentieel) monomeer-dimeer evenwicht nu in levende cellen gemeten kan worden. Daarnaast verwachten we dat deze methode naast Roco4 en LRRK2 ook geschikt is voor andere complexe eiwitten die zich gedeeltelijk hechten aan cellulaire structuren. Dit maakt de BDGA-methode breed toepasbaar op een grote verscheidenheid aan eiwitten.

Met onze nieuwe BDGA-methode hebben we kunnen bevestigen dat LRRK2 in levende cellen een monomeer-dimeer evenwicht heeft. Aangezien de dimeren waarschijnlijk een rol spelen bij de verhoogde kinase-activiteit van LRRK2 in parkinsonpatiënten, is het van belang om stoffen te testen die de dimerisatie tegengaan. De hier ontwikkelde experimentele setup biedt de mogelijkheid om de invloed van zulke stoffen op LRRK2-dimerisatie rechtstreeks in levende cellen te testen. De ontwikkeling van medicijnen tegen de ziekte van Parkinson komt daarmee een stap dichterbij.

Acknowledgements

At the end of this dissertation, I would like to thank the many people who have helped me along the long road. It has not always been easy, both in terms of scientific results and personal circumstances. Particularly after the illness and death of my daughter Lisa I would not have been able to continue without the support of family, friends, colleagues and especially my supervisors, who gave me the space to return to work at my own pace. I enjoyed my time in the lab and although my scientific career will probably end here, I would not have chosen otherwise.

Arjan, bedankt dat je me de mogelijkheid gaf om onder jouw supervisie mijn PhD te doen. Je stond altijd klaar met tips en ideeën voor de projecten, en gaf me veel ruimte voor persoonlijke ontwikkeling, zoals cursussen in de richting van onderwijs en later wetenschapscommunicatie. Ook tijdens de slotfase voelde ik me echt gesteund, door je snelle reacties op aangeleverde tekst. Toen ik mijn proefschrift uiteindelijk indiende keurden jij en Peter hem diezelfde dag nog goed.

Peter, ik ben waarschijnlijk de laatste student die onder jou als eerste promotor promoveert (Arjan was toen ik in 2013 begon nog geen full professor). Bedankt voor je hulp, met vragen over statistiek kon ik altijd bij je binnenlopen. Ook toen je al met pensioen was stond je klaar met snelle feedback.

Ineke, moeder (en inmiddels oma) van de groep, dankjewel voor de goede zorgen, je enthousiasme en je luisterend oor. Ik leerde je kennen tijdens mijn bachelor, toen ik een research vak bij jullie groep volgde. Het is me altijd bijgebleven dat je me toen dringend adviseerde om te stoppen met computerwerk vanwege de ontstane RSI. Ik dacht dat stoppen geen optie was, maar dankzij de hulp van jou en Peter kon ik het vak toch mondeling afronden. Ik denk dat dit de sfeer typeert van een lab met veel oog voor elkaar, iets waar ik heel dankbaar voor ben en wat ik elke promovendus gun. Toen ik net begonnen was met mijn PhD nam je me op sleeptouw naar Wageningen. Bedankt dat je me een goede start gaf voor de samenwerking daar, en natuurlijk ook de gezellige reis. Ook tijdens de rest van mijn PhD heb ik altijd erg genoten van je aanwezigheid, Ineke. Bedankt voor al je hulp, zowel persoonlijk als met de experimenten en analyses, zelfs nog tijdens je pensioen.

Maarten, opa van de groep, bedankt voor de gezelligheid en het voorbeeld dat je voor me was als docent (al tijdens mijn eerste studiejaar in 2007 gaf je een inspirerend vak over DNA, en later begeleidde je – samen met Michael Chang – mijn bachelorscriptie). Tijdens de koffie (gelijk om 9 uur 's ochtends) sloot ik graag bij jou en de rest van de oude garde aan. De gesprekken over het onderwijs en de faculteit vond ik erg boeiend,

iets waar ik nu in mijn nieuwe functie zelfs wat aan heb.

Jannet, bedankt voor al je steun, zowel secretariael als door een keer op Lisa te passen terwijl ik een presentatie gaf. Maar vooral ook voor je persoonlijke medeleven toen het moeilijk was.

Jan Willem en Adrie, bedankt voor de goede samenwerking. Mijn verblijf in Wageningen was altijd gezellig en ook heel nuttig dankzij jullie. Bedankt dat ik gebruik mocht maken van jullie microscoop en op die manier nauwkeurige metingen kon doen.

Victor, thank you so much for our pleasant and fruitful collaboration. Without your help I would not have been able to get the analyses working. Thank you for your patience when we were figuring out how to combine the data of different measurement days in a statistically valid way and for explaining some biophysical concepts. Our collaboration was truly interdisciplinary and I enjoyed it.

Wouter, je was al bij de groep weg toen ik erbij kwam, maar goed onderzoek staat op de schouders van reuzen. Dankzij jouw onderzoek naar Roco eiwitten zijn we waar we zijn. Jouw thesis heb ik zo grondig gebruikt dat hij inmiddels uit elkaar valt.

Douwe, ook jij was al bij de groep weg toen ik kwam, maar later kwam je terug. Bedankt voor alle DNA-plasmides die je op zo'n systematische manier hebt gemaakt.

Bernd, you highly recommended to join the lab when I was applying for this job, because of the good atmosphere and good supervision. You were totally right. You weren't in Groningen much (you did most of your research in Dortmund), but when you were there, you often took the initiative to all go out for dinner, which was always a lot of fun.

Kasia, during the bachelor course I mentioned above you were my supervisor, which was very nice. Rama, dear Rama, your smile is still being missed. You initiated the goodbye movies, which became a very fun tradition. Suzanne, we have not seen each other a lot as you were based in Dortmund, but your drive and enthusiasm were an example for me. Thank you for writing a review together. Ankita, Liu, Panos, Dominika, Srishti, Gargi, Xiao and Asmaa, we were not very close but I enjoyed being your colleague. Karisha, Carmen, Mudrika, Emmanouil and other students, even though our projects ended up in a dead end and not in this thesis, it was a great pleasure to supervise you.

Ina, I think from all PhD students in the lab our trajectories had the longest overlap, as had our topics (for quite a while we were the only ones in Groningen working on

Roco4). Thank you for being such a nice and considerate colleague, and for all the tips and tricks you gave me while we were both looking for a job.

Richard, toen ik net op het lab begon kon ik niet zo goed hoogte van je krijgen, maar later veranderde dat in diep respect voor wie je bent. Geduld met mensen, no-nonsense en je helpt anderen graag. Bedankt voor het maken van plasmides en voor alles dat je voor het lab doet.

Ahmed, you have become to me a brother and a friend. Thank you for all the nice conversations about religion (and thank you Richard for tolerating them in our office). And for the tip to use FKBP12 to induce dimerization, this really moved my project forward.

Marjon, you probably don't know how excited I was when you joined our lab. And these expectations were not in vain, since we were not only colleagues but also became friends. Thank you for inviting me to your place, we should definitely have done that more often. When you moved to Nijmegen we still kept in touch, thank you for the nice phone conversations. I'm happy you will be my paranymp.

Pragya, you really drove my research further. You were the first (and only) student I supervised who was working on my main FCS project (instead of a side project), and you nailed it. After your Master thesis Arjan invited you to work on the FCS project for about a year before you started your PhD, and this was very helpful and it was very nice to collaborate on the same project. Together with Ineke we did many measurements and analyses and finally got results that made sense. It was also nice and 'gezellig' to have you as an office mate, we talked a lot (misschien soms iets té gezellig). I'm happy you will be my paranymp too.

Dear Karla, we were not direct colleagues but as a fellow PhD student you were my buddy in the PhD support group, which I joined in the final year. Thank you for all your support; sharing my weekly schedules with you gave structure to my weeks and really helped to stick to my planning.

Talking about fellow PhD students, I would also like to thank the MindMint committee. Jonne, Hector, Taichi and Aafke, during the pandemic you were like colleagues to me. Thank you for our collaboration and the good time we had together, albeit mostly online, it was a very welcome change from my thesis work.

I would also like to thank all colleagues from my new job at the Biology Cluster. Especially Dineke, my new 'boss'. Thank you so much for your moral support and encouragement to finish my PhD.

And then I come to my friends and family. So many names I could mention: Merel, Marindy, Maaïke, people from the Grace Church community, and many others whose friendship supported me along the way. Thank you so much.

Dana and Lotteke, as study friends and later as fellow PhD students, it has been very nice to talk about our PhD projects, both the research and PhD life. Marthe, it was very nice getting to know you as a fellow mom and researcher. You have been (and still are) a great example to me. Sam and Maartje, thank you for our weekly dinners, you have become like family. Benno and Berdine thank you for already becoming a community together. I hope and trust one day we will (finally) establish an actual living community. Thank you for your support to work on my thesis.

Heit en mama, dank voor jullie steun al die jaren. Van jongs af aan hebben jullie mij de basis meegegeven waar ik nu de vruchten van pluk. Jullie hebben me altijd gesteund het beste uit mezelf te halen, welke richting ik ook koos. De laatste twee jaar hebben jullie elke week op David gepast, waardoor ik nog meer aan mijn thesis kon werken. Dankjewel.

Dick en Lia, toen de kinderopvang dicht was door de lockdown konden we twee keer een week bij jullie bivakkeren, waardoor ik alsnog goed met mijn proefschrift bezig kon. Het was gezellig en David vond het leuk, ook zonder lockdown kunnen we dat best weer eens doen ;)

Lisa en David, jullie komst was een groot geschenk en gaf het leven (nog meer) kleur. David je energie en geweldige peuterhumor fleuren de dagen op.

En Dirk, hoe kan ik je bedanken. Dit proefschrift is echt een gezamenlijke inspanning geweest die ook veel van jou vroeg. Dankjewel dat je me de ruimte gaf om ook zonder contract verder te werken aan mijn proefschrift, zowel financieel als moreel. Je bent mijn steun en toeverlaat en ik kan me niet indenken hoe deze reis geweest zou zijn zonder jou. Ik hou van je.

Last but not least I thank God for his support. With his power and strength, I was able to continue, even in hard times.

List of publications

Nederveen-Schippers LM, Pathak P, Keizer-Gunnink I, Westphal AH, Haastert PJM van, Borst JW, Kortholt A, and Skakun V. (2021). Combined FCS and PCH Analysis to Quantify Protein Dimerization in Living Cells. *Int. J. Mol. Sci.*22(14):7300. doi: 10.3390/ijms22147300.

Yatskou MM, Skakun VV, Nederveen-Schippers LM, Kortholt A, and Apanasovich VV. (2020). Complex Analysis of Fluorescence Intensity Fluctuations of Molecular Compounds. *J. Appl. Spectrosc.*87(4):685–692. doi: 10.1007/s10812-020-01055-6.

Deyaert E, Wauters L, Guaitoli G, Konijnenberg A, Leemans M, Terheyden S, Petrovic A, Gallardo R, Nederveen-Schippers LM, Athanasopoulos PS, Pots H, Van Haastert PJM, Sobott F, Gloeckner CJ, Efremov R, Kortholt A, Versées W. (2017). A homologue of the Parkinson's disease-associated protein LRRK2 undergoes a monomer-dimer transition during GTP turnover. *Nat Commun.*8(1):1008. doi: 10.1038/s41467-017-01103-4.

Terheyden S, Nederveen-Schippers LM, Kortholt A.(2016). The unconventional G-protein cycle of LRRK2 and Roco proteins. *Biochem Soc Trans.* 44(6):1611-1616. Review. doi: 10.1042/BST20160224

Malina A, Mills JR, Cencic R, Yan Y, Fraser J, Schippers LM, Paquet M, Dostie J, Pelletier J. (2013). Repurposing CRISPR/Cas9 for in situ functional assays. *Genes Dev.*27(23):2602-14. doi: 10.1101/gad.227132.113.

Curriculum Vitae

Laura Myrna Schippers was born on December 17th, 1988 in Leeuwarden. In 2007 she completed her secondary education at the Christelijk Gymnasium Beyers Naudé in Leeuwarden and continued her education at the University of Groningen studying Life Science & Technology with a specialization in Molecular Biology. For her minor she took courses in microbial genetics and chemical biology at Uppsala University in Sweden. During her Masters in Molecular Biology and Biotechnology, which started in 2010, also in Groningen, she carried out her first research project in the Systems Biology group of Prof. Dr. Matthias Heinemann, studying yeast ageing using a microfluidics set-up. She conducted her second research project at McGill University in Montréal, Canada, under the supervision of Prof. Dr. Jerry Pelletier. During this project, she was involved in the early stages of applying the CRISPR-CAS system to human cell lines. From October 2013, she performed her PhD research at the Cell Biochemistry group of the University of Groningen, under the supervision of Prof. Dr. P.J.M. van Haastert and Prof. Dr. A. Kortholt, on the subject described in this thesis. In May 2021 Laura registered at the Chamber of Commerce as a science journalist, regularly writing for UKrant, and since October 2021 she has been working as a Research Project Coordinator for the MSc program Biomedical Sciences at the University of Groningen.