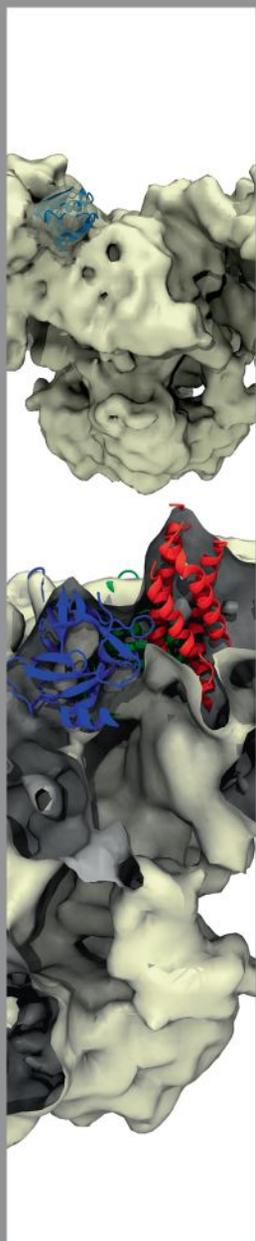
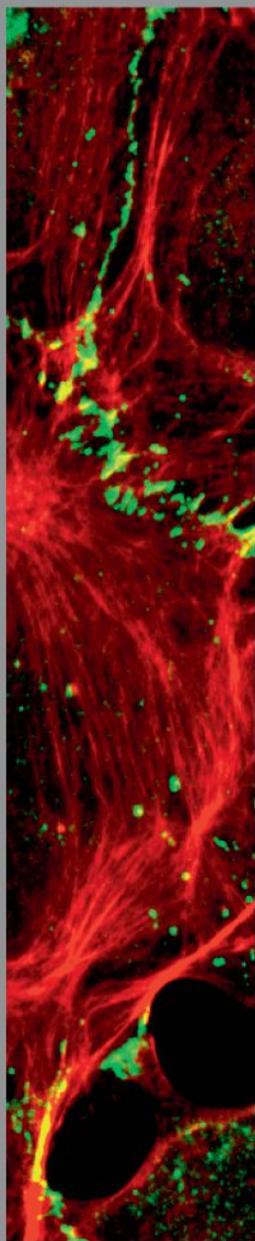
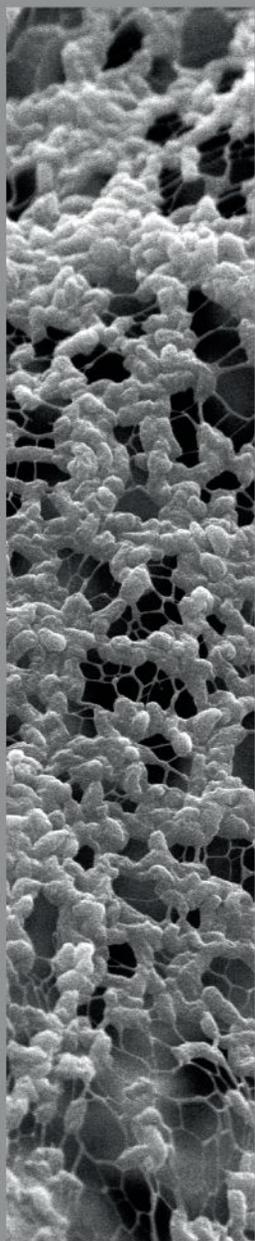
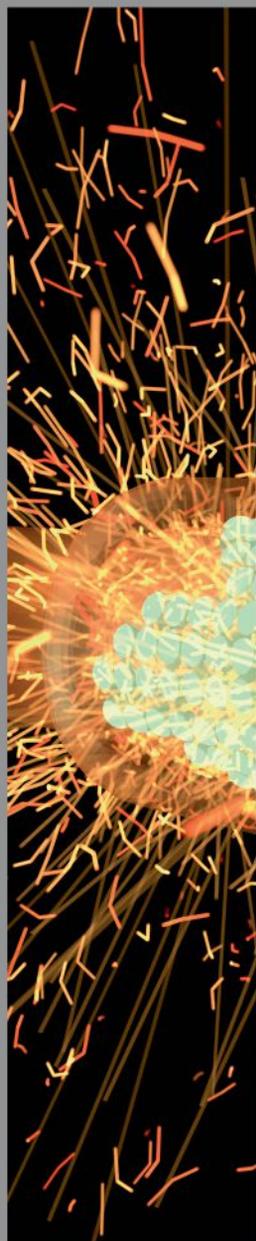
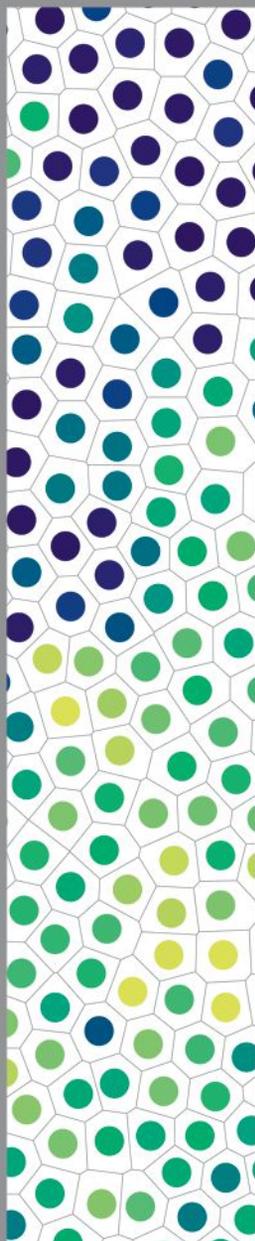


Biophysical *Journal*

Best of 2014





OptoSplit

Acquire simultaneous images with one camera.

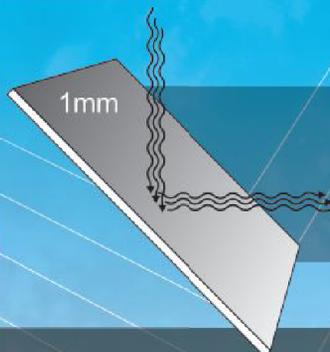
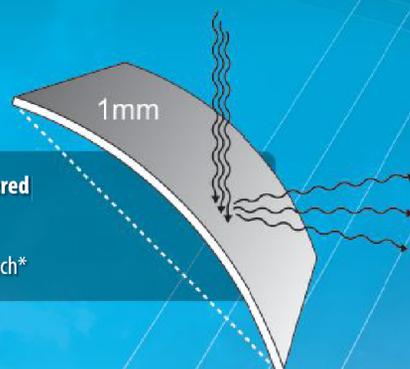
The OptoSplit II LS and OptoSplit III LS from Cairn Research enable simultaneous, multi-channel imaging by splitting the image into multiple components on the same camera. Record differences in color, polarization, ion concentrations or pH using the dual emission channel II LS and the triple emission channel III LS. Now you can perform:

- ▶ Simultaneous multicolor imaging
- ▶ Single molecule, TIRF, and Super-Resolution Microscopy
- ▶ Simultaneous FRET imaging
- ▶ Dual polarization imaging
- ▶ Calcium imaging (and other ion indicators), pH imaging
- ▶ Simultaneous phase contrast and fluorescence

This intuitive, modular system allows you to choose almost any combination of UltraFlat dichroics and ET emission filters to split the images with user-interchangeable cubes. Simple controls allow you to precisely register images pixel-by-pixel.

Eliminate temporal artifacts in your imaging data by acquiring images at the same time.

Conventional Sputtered Dichroic Mirror
 Unspecified flatness
 Typically >10 waves/inch*



UltraFlat Sputtered Dichroic Mirrors
 1mm ≤ 2.00 waves/inch
 2mm ≤ 0.50 waves/inch
 3mm ≤ 0.25 waves/inch

UltraFlat Dichroics

Superior surface flatness for distortion-free reflection.

Chroma's UltraFlat dichroics provide the flattest surfaces available for splitting wavelengths in critical applications, such as:

IMAGE SPLITTING ▶ Chroma's UltraFlat dichroics provide reflected images that are in focus from edge to edge, matching the transmitted images.

TIRF ▶ Eliminate spurious fluorescence signals by using dichroics with surfaces flat enough to reflect TIRF lasers without distortion, producing optimal evanescent waves. With other dichroics, some laser photons will propagate into the sample, producing unwanted fluorescence.

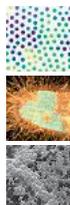
SIM ▶ Maintain the focus of reflected structured illumination and photoactivation micro-patterns.

STED ▶ Reflect well-defined, undistorted STED laser patterns onto your samples.

LASER BEAM COMBINING ▶ Maintain laser beam profile integrity when combining lasers.

Have confidence in the integrity of your reflected light without having to resort to dichroics >3mm thick.

Foreword



We are pleased to present this “Best of...” reprint collection, which provides a chance to reflect on what has caught the attention of *Biophysical Journal* readers in 2014. This collection includes a selection of 12 of the most accessed articles across a range of topics. Article selection is primarily based on the number of requests for PDF and full-text HTML versions of a given article. We acknowledge that no single measurement can truly be indicative of “the best” research papers over a given period of time. This is especially true when sufficient time has not necessarily passed to allow one to fully appreciate the relative importance of a discovery. That said, we think it is still informative to look back at the scientific community’s interests in what has been published over the past year.

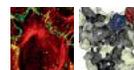
In this collection, you will see a range of the exciting topics, including cell biophysics, molecular machines, single-molecule microscopy, membrane biophysics, systems biophysics, biomolecular structure, biophysical methods, and channel electrophysiology, that have widely captured the attention and enthusiasm of our readers. They also represent several of the different types of papers that *BJ* publishes: one Biophysical Review, two Biophysical Letters, one Computational Tools article, and eight regular articles. We are particularly excited to welcome Computational Tools as a new feature of the journal; the article by Johnston et al., the first paper published in this category, is an excellent example of a valuable resource for the biophysics community.

We hope that you will enjoy reading this special collection and that you will visit <http://www.cell.com/biophysj/home> to check out the latest findings that we have had the privilege to publish. To stay on top of what your colleagues have been reading over the past 30 days, check out the “Most Read” category on our homepage. Also, be sure to visit <http://www.cell.com> to find other high-quality papers published in the full collection of Cell Press journals.

Finally, we are grateful for the generosity of our sponsors, who helped make this reprint collection possible.

Leslie M. Loew
Editor in Chief

Biophysical *Journal*



For information for the Best of Series, please contact:
Jonathan Christison
Program Director, Best of Cell Press
e: jchristison@cell.com
p: 617-397-2893
t: @CellPressBiz



F · S · T[®]
FINE SCIENCE TOOLS

Our Selection Speaks Volumes

- Scissors
- Retractors
- Magnifiers
- Probes & Hooks
- Bone Instruments
- Animal Identification
- Hemostats
- Forceps
- Surgical & Laboratory Equipment
- Feeding Needles
- Spatulae & Spoons
- Wound Closure
- Surgical Plates
- Instrument Care & Sterilization
- Rongeurs
- Scalpels & Knives
- Clamps
- Pins & Holders
- Needles & Needle Holders
- Student Quality Instruments & Much More

FINE SURGICAL INSTRUMENTS FOR RESEARCH[™]
Visit us at finescience.com or call **800 521 2109**

Scopus is the largest abstract and citation database of peer-reviewed literature and quality web sources with smart tools to track, analyze and visualize research.

**enrich
your
experience**

www.scopus.com

refine your research
SCOPUS[™]

Biophysical *Journal*

Best of 2014

Review

Super-resolution Microscopy Approaches for Live Cell Imaging

Antoine G. Godin, Brahim Lounis, and Laurent Cognet

Biophysical Letters

Indolicidin Binding Induces Thinning of a Lipid Bilayer

Chris Neale, Jenny C.Y. Hsu, Christopher M. Yip, and Régis Pomès

3D-SIM Super-resolution of FtsZ and Its Membrane Tethers in *Escherichia coli* Cells

Veronica Wells Rowlett and William Margolin

Articles

Explicit Tracking of Uncertainty Increases the Power of Quantitative Rule-of-Thumb Reasoning in Cell Biology

Iain G. Johnston, Benjamin C. Rickett, and Nick S. Jones

Imaging GFP-Based Reporters in Neurons with Multiwavelength Optogenetic Control

Veena Venkatachalam and Adam E. Cohen

Evolution of Specificity in Protein-Protein Interactions

Orit Peleg, Jeong-Mo Choi, and Eugene I. Shakhnovich

Evaluation of Fluorophores to Label SNAP-Tag Fused Proteins for Multicolor Single-Molecule Tracking Microscopy in Live Cells

Peter J. Bosch, Ivan R. Corrêa, Jr., Michael H. Sonntag, Jenny Ibach, Luc Brunsveld, Johannes S. Kanger, and Vinod Subramaniam

Selecting Ions by Size in a Calcium Channel: The Ryanodine Receptor Case Study

Dirk Gillespie, Le Xu, and Gerhard Meissner

Karyopherin-Centric Control of Nuclear Pores Based on Molecular Occupancy and Kinetic Analysis of Multivalent Binding with FG Nucleoporins

Larisa E. Kapinos, Rafael L. Schoch, Raphael S. Wagner, Kai D. Schleicher, and Roderick Y.H. Lim

Conformational Recognition of an Intrinsically Disordered Protein

James M. Krieger, Giuliana Fusco, Marc Lewitzky, Philip C. Simister, Jan Marchant, Carlo Camilloni, Stephan M. Feller, and Alfonso De Simone

The Berg-Purcell Limit Revisited

Kazunari Kaizu, Wiet de Ronde, Joris Pajmans, Koichi Takahashi, Filipe Tostevin, and Pieter Rein ten Wolde

Role of Denatured-State Properties in Chaperonin Action Probed by Single-Molecule Spectroscopy

Hagen Hofmann, Frank Hillger, Cyrille Delley, Armin Hoffmann, Shawn H. Pfeil, Daniel Nettels, Everett A. Lipman, and Benjamin Schuler

Be the first...

Biophysical Journal

PUBLISHED BY
Cell
PRESS

Biophysical Journal

Volume 96
Number 1
January 7, 2009
www.biophysj.org

Join the Biophysical Society today!

Benefits of membership:

- Free online access to *Biophysical Journal*, now published by Cell Press
- Reduced registration rates for the Annual Meeting
- Career development resources
- Funding opportunities
- Discounted subscriptions to Cell Press journals

[Become a member!](#)

Table of Contents for *Biophysical Journal*, Volume 96, Issue 1. January 2009

Biophysical Journal

Volume 96
Number 1
January 7, 2009
www.biophysj.org

Biophysical Society

Published by Cell Press
for the Biophysical Society

to read the latest issue of *Biophysical Journal*.

Receive *Biophysical Journal* Email Alerts – FREE!
Register Now at www.biophysj.org

Biophysical Review

Super-resolution Microscopy Approaches for Live Cell Imaging

Antoine G. Godin,^{1,2} Brahim Lounis,^{1,2} and Laurent Cagnet^{1,2,*}

¹University of Bordeaux, LP2N, UMR 5298, Talence, France; and ²Institut d'Optique Graduate School and Centre National de la Recherche Scientifique, LP2N, UMR 5298, Talence, France

ABSTRACT By delivering optical images with spatial resolutions below the diffraction limit, several super-resolution fluorescence microscopy techniques opened new opportunities to study biological structures with details approaching molecular structure sizes. They have now become methods of choice for imaging proteins and their nanoscale dynamic organizations in live cells. In this mini-review, we describe and compare the main far-field super-resolution approaches that allow studying endogenous or overexpressed proteins in live cells.

INTRODUCTION

The decryption of cell functions and subcellular processes has constantly benefited from advances in microscopy. In particular, the developments of fluorescence microscopy and of numerous fluorescent probes allowing the study of specific biomolecules at work in their native environment were instrumental to the advance of live cell mechanism investigations. The optical resolution of microscopes is limited by the diffraction of light, which commonly sets a limit of $\sim\lambda/2$ in far-field microscopy. By delivering optical images with spatial resolutions below the diffraction limit, super-resolution fluorescence microscopy offered new promises to study molecular processes with greater detail than with conventional microscopies (1,2). Most of these methods rely on the control of the number of emitting molecules in specific imaging volumes. This can be achieved by controlling local emitter fluorescent state populations or the labeling densities of fluorescing probes at any given time during the image acquisition process. In this mini-review, we will discuss the key features of super-resolution techniques used for live-cell studies. We schematically divide them into three major groups: those based on highly localized fluorescence emission volumes; those based on structured illumination; and those based on single-molecule localizations. A didactic representation of the three families of super-resolution approaches is presented in Fig. 1.

SUPER-RESOLUTION BASED ON HIGHLY LOCALIZED FLUORESCENCE EMISSION VOLUMES

Stimulated emission depletion (STED) and reversible saturable optical fluorescence transition (RESOLFT)

In a far-field confocal microscope, the effective fluorescence volume can be reduced below the diffraction limit (3) by

using saturable optical processes that deexcite emitters formerly excited by a focused laser beam. These processes work to prevent fluorescence emission from specific regions of the excitation beam by driving the molecules in these regions between bright and dark states using a depletion light beam. One elegant and efficient strategy consists of using stimulated emission by a high-intensity ($>MW/cm^2$), doughnut-shaped laser beam superimposed with the focused excitation laser beam, completely preventing fluorescence emission from emitters in peripheral regions of the excitation beam. This process was coined “stimulated emission depletion” (STED) (3). A doughnut-shaped depletion beam is the simplest design; however, in general, any depletion beam featuring a spatial intensity distribution with one or several intensity zeroes can be used to perform STED images.

To generate a super-resolved image with STED based on local excitation volumes, one must scan the excitation/depletion effective volumes over the sample in a deterministic point-by-point manner or by use of parallelized scanning schemes (4,5). STED was successfully applied in several live samples to study slow morphing and movements of organelles such as reticulum endoplasmic or microtubules (6), subcellular organization in live cells (7), and synaptic structures in live samples (7–9). For live cell studies, one should bear in mind that relatively high laser powers are needed in STED, especially when using continuous wave laser beams (e.g., $\sim MW/cm^2$ (10)). Using pulsed excitation beam together with time-gating detection allowed a ~ 2 – 3 -fold reduction in laser power (11). In addition, photobleaching is a limiting factor for long-term live sample imaging because each fluorescent molecule undergoes a large number of exciting/de-exciting cycles in the depletion beam.

An approach similar to STED using much lower intensities to deplete emitting molecular levels ($\sim kW/cm^2$) (12) is based on reversible photoswitching of marker proteins between a fluorescence-activated and a nonactivated state (13–15), whereby one of the transitions is accomplished by means of a spatial intensity distribution featuring a

Submitted June 18, 2014, and accepted for publication August 7, 2014.

*Correspondence: laurent.cagnet@u-bordeaux.fr

Editor: Brian Salzberg.

© 2014 by the Biophysical Society
0006-3495/14/10/1777/8 \$2.00



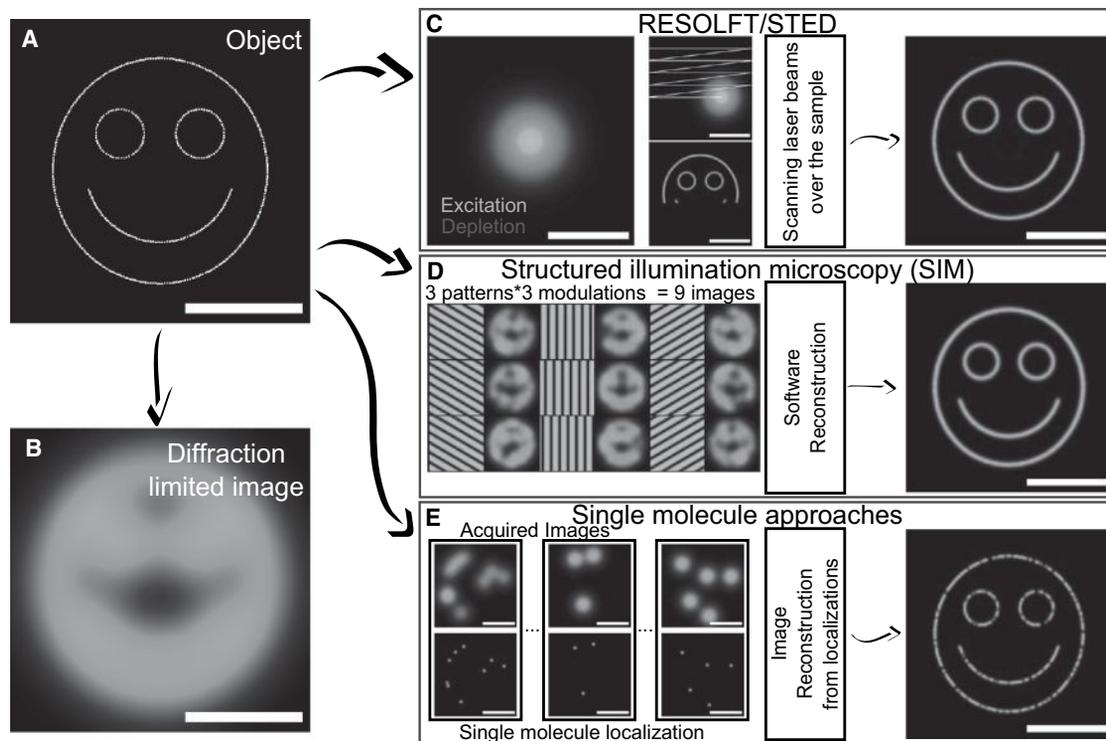


FIGURE 1 Schematic description of the superresolution microscopy approaches. All images for this didactic description are computer-generated. Object to be imaged consisted of fluorescent emitters (*A*) and corresponding diffraction-limited image (*B*). (*C*) In RESOLFT/STED, a focused excitation beam (cyan) superimposed with a doughnut-shaped depletion beam (red) are scanned over the sample to acquire an image at high resolution (down to ~50–80 nm in live cells). (*D*) In SIM, after the required software reconstruction, multiple wide-field images are acquired using sinusoidal illumination grid patterns to obtain high-resolution images (down to ~50–100 nm in live cells using nonlinear saturated illumination). (*E*) In single-molecule localization microscopy, a large number of wide-field images containing a few isolated single fluorescent emitters are successively acquired. A high-resolution image is reconstructed from the localizations of each individual molecule. Resolutions down to ~50 nm are commonly achieved in live cells. In the example provided, we considered the detection of 80% of the molecules present in the object image. Scale bar represents 1 μm . To see this figure in color, go online.

zero. This generalized approach was named after “reversible saturable optical fluorescence transition” (RESOLFT). Bright photostable switchable fluorophores and fluorescent proteins development were particularly instrumental in the development of these techniques (14–16). Importantly, fluorescent proteins provide specific 1:1 protein labeling and offer the possibility of intracellular live cell imaging.

STRUCTURED ILLUMINATION MICROSCOPY (SIM)

Structured illumination microscopy (SIM) is based on standard wide-field microscopy and is compatible with most standard fluorophores and labeling protocols. It uses nonuniform illuminations with known spatial patterns (e.g., originally a sinusoidal grid, but other illumination distributions can also be used (17)). From multiple acquisitions (e.g., nine images, incorporating three phase shifts for three pattern orientations (18)), high spatial frequency information is retrieved with a dedicated algorithm, comprising a method inaccessible with standard illumination schemes (19). Contrary to standard laser scanning modalities like STED/RESOLFT, SIM allows acquisition of a large field of view over limited times. However, SIM routinely pro-

vides only an approximately twofold resolution enhancement of standard wide-field microscopy as compared to other super-resolution methods (19). Nonlinear saturated SIM using fluorophore saturation or photoswitchable proteins as in RESOLFT can achieve higher resolution enhancement (~50 nm), but requires an increased number of image acquisitions (up to 63) and a complex reconstruction process (20,21). SIM has been demonstrated for long-term, live cell imaging in microtubules and other dynamic structures (21–23). Three-dimensional SIM imaging has been further achieved using 15 different pattern acquisitions per axial planes for reconstruction instead of nine images to reject the out-of-focus light (24). Whole-cell volume imaging has been performed using three-dimensional SIM in two colors (25). And, interestingly, fast SIM imaging (11 Hz) has even been developed with a 100-nm resolution for a small field of view ($\sim 8 \times 8 \mu\text{m}^2$) (18).

SINGLE-MOLECULE LOCALIZATION MICROSCOPY APPROACHES

It is well known that the position of isolated single fluorescent emitters can be determined by image analysis with

greater precision than is available from the diffraction limit alone. This feature, which has been used for more than 20 years in live cell, single-particle/-molecule studies (26), is key to providing today's super-resolved images. Super-resolution methods based on single-molecule localizations simply consist of reconstructing an image from single molecule localizations retrieved from a large number of movie frames (typically thousands of camera frames). The main requirement is that each frame contains the detection of spatially well-separated fluorescent emitters (27). Inasmuch as a large volume of single-molecule detections must to be registered to reconstruct a high-content super-resolved image, this acquisition process is inherently slow (typically more than a few seconds). Below, we describe three families of such approaches, distinguishing how fluorescent molecules are stochastically isolated from nonfluorescent ones in each camera frame.

Photoactivation localization microscopy (PALM)

The development of photoactivation localization microscopy (PALM) (28,29) is closely linked to the advent of photoactivatable proteins (16), which allows us to control, by light, the density of fluorescing proteins in each image. Although it is restricted to expression systems, preventing the study of endogenous proteins in their native environment, PALM takes advantage of the versatility and specificity of genetically encoded, fluorescently tagged molecules in cells, and has quickly become the tool of choice for super-resolution live cell imaging. PALM, by design, is not restricted to biomolecules present at the cell plasma membrane, and allows the study of intracellular biomolecules. By tracking the movement of each individual protein, PALM also allows measuring local diffusion properties in living cells on short timescales (30–32) and cellular structural changes in three dimensions on longer timescales (33).

Stochastic optical reconstruction microscopy (STORM) and ground state depletion microscopy followed by individual molecular return (GSDIM)

Stochastic optical reconstruction microscopy (STORM) (34) and direct STORM (dSTORM) (35) use switchable organic fluorophores placed in specific buffers (e.g., with reducing properties) instead of using fluorescent proteins as in PALM. Those probes can be targeted on genetically encoded or endogenous proteins using adequate ligands. STORM was first demonstrated using Cy3–Cy5 pairs (34) but was quickly extended in dSTORM to other synthetic fluorophores that can be stochastically and reversibly switched in the imaging buffers (36,37). STORM is particularly powerful for fixed cells applications (38–40), and can be extended in three dimensions (38). One caution is that live cell experiments should be performed with great care due to possible issues

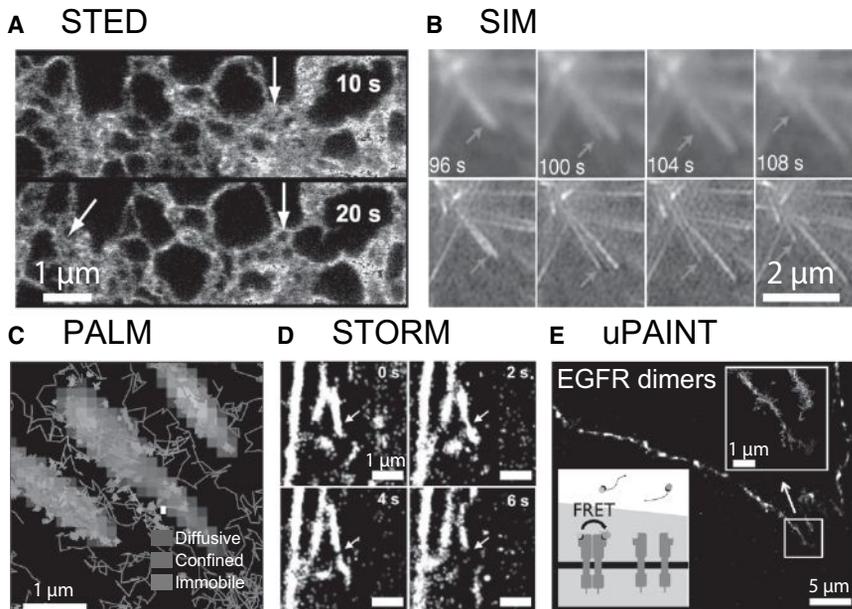
with the use of reducing/oxidizing buffers that can affect cell integrity (41). Of special interest is that STORM has been shown to take advantage of some reduction in thiol glutathione, which is naturally present at millimolar concentrations in bacteria (42) or in specific cell compartments such as the nuclei of eukaryote cells when using buffers with low cellular toxicity (43,44). Organelles from live cells such as the membrane and mitochondria have also been investigated using multicolor STORM (45–47).

Analogously to dSTORM, where stochastic photoswitching is used to control the number of emitting fluorophores, ground state depletion microscopy followed by individual molecular return (GSDIM) covers the techniques employing the transition between the fluorescent singlet state and the metastable triplet state as a stochastic on-off switch (48,49). More precisely, efficient transition to the long-lived triplet state is achieved in such techniques by using high-excitation intensities combined with an imaging buffer, similar to STORM, to allow obtaining triplet lifetimes just long enough to leave only a few emitting fluorophores at any time in each image. Under these conditions, GSDIM has been employed for imaging living cells using both fluorescent protein tags (as in PALM) or various organic fluorophores (as in STORM) that selectively bind to tagged proteins (50,51).

Universal point accumulation imaging in the nanoscale topography (uPAINT)

In contrast with PALM, STORM, and GSDIM, which are based on the emitters stochastic photoswitching, the method known as “universal point accumulation imaging in the nanoscale topography” (uPAINT) (52–55) captures real-time molecular interactions to control the density of fluorescent emitters suitable for single molecule identification in each image. In the uPAINT approach, target molecules are individually imaged when a specific ligand coupled to a fluorescent dye binds to the target molecule.

Unbound ligands freely diffuse in the imaging buffer (with typical diffusion constants of approximately tens of $\mu\text{m}^2/\text{s}$) and, due to an oblique illumination excitation, are not excited in an efficient manner. Therefore, unbound ligands are not detected efficiently by a detector operating at a typical video rate, in contrast to bound ligands, which diffuse together with a membrane receptor (with typical diffusion constants of $<1 \mu\text{m}^2/\text{s}$) in the oblique laser illumination. With uPAINT, any binding entity conjugated to fluorophores having high specificity toward a target molecule (e.g., natural/synthetic ligand, antibody) can be used as fluorescent probes to reveal the targeted molecules. Applications include receptors and GPI-anchored proteins diffusing on live cell membranes labeled with antibodies or synthetic ligands. uPAINT was also applied to image and track endogenous receptors such as glutamate receptors in neurons (54) and epidermal growth factor receptors



Each reconstruction was obtained using 1000 frames (2 ms per frame). Scale bar = 1 μm . This figure was adapted from Wilmes et al. (47). (E) uPAINT: live cell superresolution imaging of membrane epidermal growth factor receptor (EGFR) dimers based on single-molecule fluorescence resonance energy transfer induced by fluorescent ligand activation. (Inset) Preferential cell-edge localization of EGFR dimers. In addition, uPAINT provides numerous single-molecule trajectories on a single cell, allowing the extraction of the diffusion properties of the EGFR dimer population from the whole-ligand-activated EGFR population. This figure was adapted from Winckler et al. (56). To see this figure in color, go online.

(EGFRs) at high densities in culture living cells (56). Comparisons between transfected nonendogenous receptors and endogenous glutamate receptors were also performed (57).

A similar approach allowed tracking and imaging by continuously labeling sodium ion channels in live cells, using small fluorescently labeled molecule agents that display reversible binding to the sodium ion channel (58). Interestingly, combining single-molecule fluorescence resonance energy transfer and dual-color uPAINT allowed the specific super-resolution imaging and tracking of interacting receptors activated by their cognate ligand in live cells (56). This feature stems from the fact that in this uPAINT study, fluorescently tagged ligands are directly used as imaging probes, allowing us to extract and image the population of activated functional receptors upon ligand binding in real-time. Examples of high resolution images obtained based on highly localized fluorescence emission volumes, structured illumination microscopy and single-molecule localization microscopy are presented in Fig. 2.

DISCUSSION

Dynamics and resolution in live cells

In live cells or organisms, supramolecular structures and organelles morph often in reaction to stimuli (seconds to minutes and up to hours) over relatively short timescales. To fully understand the cell signaling that induces those processes, it is important to investigate therein the dynamics of

FIGURE 2 Examples of achievements obtained with superresolution microscopy in live biological samples. (A) STED: continuous-wave STED images of the yellow fluorescent protein (*citrine*) targeted to the endoplasmic reticulum in live cells revealing small tubules (~ 60 nm). Image sequences show morphing of the endoplasmic reticulum at arrows (pixel size = 20 nm, 10 s recording time per image). Scale bar = 1 μm . This figure was adapted from Hein et al. (6). (B) SIM: total-internal reflection microscopy image series of eGFP- α -tubulin in a live S2 cell and corresponding SIM images revealing the elongation followed by a rapid shrinking of a microtubule. Integration time of 270 ms per frame. This figure was adapted from Kner et al. (18). (C) PALM: numerous single trajectories of $\beta 3$ -integrin fused with mEOS2, obtained on a single MEF cell with PALM, revealing that $\beta 3$ -integrin undergo slower free-diffusion inside focal adhesions (*gray*) than outside, as well as confined diffusion and immobilization. Figure adapted from Rossier et al. (31). (D) STORM: spatial dynamics of cortical actin skeleton stained with Lifeact-HaloTag/ATTO655.

molecular (re)organizations. In particular, to grasp the full spectrum of mobility behaviors of biomolecules (up to 1 $\mu\text{m}^2/\text{s}$ for membrane receptors), fast video rate acquisition is required (20–100 Hz) on large fields of view. Yet, in all super-resolution methods, breaking the diffraction limit on a given field of view comes at the expense of the acquisition time.

In point scanning RESOLFT/STED methods, a compromise between imaging large fields of view and fast acquisition speed has to be made because of the requirement for a dense pixilation. Imaging rates in RESOLFT are rather slow because pixel integration times are limited by the protein photoswitching processes. Being based on stimulated emission, STED is not subject to this fundamental limit. However, increased resolution being achieved with high laser powers, care should be taken to ensure live cell integrity. RESOLFT/STED methods are able to resolve the movements of the slow structures (typically 10–30 s for 512×512 pixels) such as, for instance, microtubule networks organizations (22,59) and neuron morphology dynamics (60,61), in cell cultures and live animals. Interestingly, by acquiring small fields of view, the dynamics of nanoscale structures can be monitored with higher imaging rates (28 Hz for 60×82 pixels) (62). Recently, large parallelization of the depletion patterns combined with the use of matrix detectors drastically increased imaging speed over large field of view in RESOLFT (59) and STED (5).

In its standard form, SIM uses nine wide-field fluorescent images to build a super-resolved image in typically 1 s.

Many applications of SIM were described for imaging sub-cellular structures in living cells (18,23,25). With saturated SIM (the structured illumination approach giving better spatial resolution), additional illumination patterns are needed for the reconstruction (21), leading to degraded time resolution. Important to note for SIM techniques is that any aberration, sample movement, or fluorophore photobleaching during the image sequence will induce artifacts that will strongly affect the quality of reconstructed super-resolved images (18).

In single-molecule localization approaches, two time-scales are relevant:

1. Raw-images acquisition rate. This sets the individual molecules' tracking time resolution (1–10 of ms). This rate also sets the single-molecule pointing accuracy through its impact on the signal/noise of each molecular detection (63,64). The analysis of single-molecule trajectories provides local mobility maps on live cell regions with high spatiotemporal resolutions (30–32,54,57)
2. Total number of images needed to reconstruct a super-resolved image (which sets the rate). Indeed, in addition to the pointing accuracy, the local density of single-molecule detections obtained from a studied structure also plays a central role in the final spatial resolution (as announced by the Nyquist theorem). For instance, to obtain images with 10-nm resolution, local densities of at least 10,000 detection/ μm^2 are needed. Hence, thousands of image frames are commonly acquired, justifying global recording times of approximately seconds to minutes. This timescale directly defines the time resolution at which nanoscale organization of molecular assemblies (e.g., cellular organelles) can be analyzed.

Computer analysis requirements

Conversely to RESOLFT/STED methods that do not require any postacquisition analysis, the main source of SIM's complexity lies in the sophisticated algorithms required for image reconstruction. As in single-molecule localization-based techniques, positions of the emitting single molecules must be retrieved using cutting-edge software (described in a recent comparative study (65)).

Labeling strategies and consequences

Expression systems in concert with fluorescent protein engineering provide a method of choice to study, with high specificity, subcellular organizations in live cells, making RESOLFT, SIM, and PALM essential methods for the applications described in this review. However, despite their wide applicability, one should bear in mind that some signaling and structural artifacts can arise due to the use of fluorescent proteins (66). In this context, using fluorophores conjugated to specific ligands, the methods SIM,

uPAINT, STED, and to some extent, STORM/GSDIM, should be considered, inasmuch as they are compatible with studying endogenous receptors in living cells. Interestingly, by synchronizing single-molecule detection and ligand-induced receptor activation, uPAINT is, to date, the only super-resolution method that allows studying, in real-time, specifically activated functional receptors and their interactions at the membrane of living cells (56).

Of note, protein number quantification can in principle be performed using PALM and uPAINT, inasmuch as, in these methods, photobleaching irreversibly turns off fluorophores after their detection. In STORM/GSDIM, however, reversible stochastic switching of fluorophores can bias such quantitative analysis because observing the same fluorescent molecules more than once is plausible. Interestingly, recent advances in self-labeling proteins such as the SNAP, CLIP, and Halo tags, allow efficient live cell protein labeling, including intracellular ones (46,51,67), and provide valuable tools to perform multicolor super-resolution imaging that can, in principle, be applied to STED, STORM, or uPAINT.

Studying structures in three dimensions

Several methods were developed to improve axial resolution in fluorescence imaging. The most widely used strategies are based on single-molecule localization and provide axial information (~20–70 nm) of the position without severely altering either the radial or time precision. By shaping the detection point-spread function along the axial position, single-molecule position can be precisely determined along the optical axis (38,68,69). Detecting molecules using two objectives (70,71), by moving the sample in the axial direction (72) or by interfering the signals obtained from two objectives (73,74), could also yield similar resolution improvements along the optical axis. Although the last approach is the most precise (~20 nm), it is also the most complex to implement. Finally, STED/RESOLFT can also be extended in three dimensions by scanning distinct axial planes independently (9).

SUMMARY

Super-resolution approaches described in this mini-review were proven to deliver information on subcellular organization at different timescales using various labeling probes. Table 1 summarizes the main pros/cons of the approaches discussed here and outlines the different spatiotemporal fundamental limits.

CONCLUSION

During the last two decades, super-resolution approaches have provided new insights into subcellular organization at nanometer-scale resolutions. Several of these methods

TABLE 1 Comparison of the superresolution approaches presented in this review

Approach	STED	RESOLFT	SIM	Single-molecule approaches		
				PALM	STORM/GSDIM	uPAINT
Resolution (live samples)	50–70 nm	80–100 nm	50 ^c –100 nm	50 nm	50 nm	50 nm
Toxicity	++	+++	+++	+++	+	+++
Endogenous	+++	—	+++	—	+++	+++
Time for single image (a field of 50 × 50 μm is considered for comparison)	10–100 s ~0.1 s ^a	>500 s ~ 3 s ^b	~1 s	> 2 s	> 2 s	> 2 s
Intracellular labeling (live)	+	+++	+++	+++	+	—
Implementation complexity	+	+	+	+++	+++	+++
Reconstruction algorithm	N.A. ^a	N.A. ^b	+	++	++	++
Dynamics of large molecular structures	+++	+++	+++	++	++	++
Dynamics of single molecule	+	+	+	+++	++	+++
Multicolor imaging	+	++	+++	++	+++	+++

^aParallelization of STED nanoscopy using optical lattices was recently achieved with an imaging rate of 12.5 Hz for a 2.9 × 2.9 μm image (5). It requires a simple reconstruction algorithm.

^bLarge parallelization of the depletion patterns drastically increased imaging speed over a larger field of view in RESOLFT (120 × 100 μm in ~3 s (59)). It requires a simple reconstruction algorithm.

^cUsing nonlinear saturated SIM allows accuracy enhancement up to 50 nm (21).

have reached a sufficient level of maturity to make them routinely applicable to many biological systems. As described in this mini-review, each approach has its own advantages and drawbacks for live cell imaging. However, in combining these techniques, one can tackle specific biological questions to take advantage of the strengths of each method, such as the insurance of orthogonal labeling for multicolor imaging (47).

Other recent approaches can also be implemented to increase the image resolution by a factor of 2–4 upon applying software analytical tools to standard experimental imaging data sets. These include deconvolution microscopy (75), stochastic optical fluctuation imaging (76), Bayesian localization microscopy (77), and compressed sensing (78–80). We foresee that correlating the information gathered by such specific molecular imaging methods with structural information obtained by electron microscopy will provide new insights into molecular subcellular organization and interaction.

We acknowledge financial support from the Fondation pour la Recherche Medicale, Fonds de Recherche Nature et Technologies du Quebec (to A.G.G.), Agence Nationale de la Recherche, Region Aquitaine, the French Ministry of Education and Research, the Centre National de la Recherche Scientifique, the European Research Council, and FranceBioImaging (grant No. ANR-10-INSB-04-01).

REFERENCES

- Hell, S. W. 2007. Far-field optical nanoscopy. *Science*. 316:1153–1158.
- Huang, B., M. Bates, and X. Zhuang. 2009. Super-resolution fluorescence microscopy. *Annu. Rev. Biochem.* 78:993–1016.
- Hell, S. W., and J. Wichmann. 1994. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* 19:780–782.
- Bingen, P., M. Reuss, ..., S. W. Hell. 2011. Parallelized STED fluorescence nanoscopy. *Opt. Express*. 19:23716–23726.
- Yang, B., F. Przybilla, ..., B. Lounis. 2014. Large parallelization of STED nanoscopy using optical lattices. *Opt. Express*. 22:5581–5589.
- Hein, B., K. I. Willig, and S. W. Hell. 2008. Stimulated emission depletion (STED) nanoscopy of a fluorescent protein-labeled organelle inside a living cell. *Proc. Natl. Acad. Sci. USA*. 105:14271–14276.
- Westphal, V., S. O. Rizzoli, ..., S. W. Hell. 2008. Video-rate far-field optical nanoscopy dissects synaptic vesicle movement. *Science*. 320:246–249.
- Nägerl, U. V., K. I. Willig, ..., T. Bonhoeffer. 2008. Live-cell imaging of dendritic spines by STED microscopy. *Proc. Natl. Acad. Sci. USA*. 105:18982–18987.
- Tønnesen, J., G. Katona, ..., U. V. Nägerl. 2014. Spine neck plasticity regulates compartmentalization of synapses. *Nat. Neurosci.* 17:678–685.
- Willig, K. I., B. Harke, ..., S. W. Hell. 2007. STED microscopy with continuous wave beams. *Nat. Methods*. 4:915–918.
- Vicidomini, G., G. Moneron, ..., S. W. Hell. 2011. Sharper low-power STED nanoscopy by time gating. *Nat. Methods*. 8:571–573.
- Testa, I., N. T. Urban, ..., S. W. Hell. 2012. Nanoscopy of living brain slices with low light levels. *Neuron*. 75:992–1000.
- Hofmann, M., C. Eggeling, ..., S. W. Hell. 2005. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc. Natl. Acad. Sci. USA*. 102:17565–17569.
- Dedecker, P., J. Hotta, ..., J. Hofkens. 2007. Subdiffraction imaging through the selective donut-mode depletion of thermally stable photoswitchable fluorophores: numerical analysis and application to the fluorescent protein Dronpa. *J. Am. Chem. Soc.* 129:16132–16141.
- Grotjohann, T., I. Testa, ..., S. W. Hell. 2011. Diffraction-unlimited all-optical imaging and writing with a photochromic GFP. *Nature*. 478:204–208.
- Patterson, G. H., and J. Lippincott-Schwartz. 2002. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science*. 297:1873–1877.
- Gao, L., L. Shao, ..., E. Betzig. 2012. Noninvasive imaging beyond the diffraction limit of 3D dynamics in thickly fluorescent specimens. *Cell*. 151:1370–1385.
- Kner, P., B. B. Chhun, ..., M. G. Gustafsson. 2009. Super-resolution video microscopy of live cells by structured illumination. *Nat. Methods*. 6:339–342.
- Gustafsson, M. G. 2000. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* 198:82–87.

20. Gustafsson, M. G. 2005. Nonlinear structured-illumination microscopy: wide-field fluorescence imaging with theoretically unlimited resolution. *Proc. Natl. Acad. Sci. USA*. 102:13081–13086.
21. Rego, E. H., L. Shao, ..., M. G. Gustafsson. 2012. Nonlinear structured-illumination microscopy with a photoswitchable protein reveals cellular structures at 50-nm resolution. *Proc. Natl. Acad. Sci. USA*. 109:E135–E143.
22. York, A. G., S. H. Parekh, ..., H. Shroff. 2012. Resolution doubling in live, multicellular organisms via multifocal structured illumination microscopy. *Nat. Methods*. 9:749–754.
23. Olshausen, P. V., H. J. Defeu Soufo, ..., A. Rohrbach. 2013. Super-resolution imaging of dynamic MreB filaments in *B. subtilis*—a multiple-motor-driven transport? *Biophys. J.* 105:1171–1181.
24. Shao, L., P. Kner, ..., M. G. Gustafsson. 2011. Super-resolution 3D microscopy of live whole cells using structured illumination. *Nat. Methods*. 8:1044–1046.
25. Fiolka, R., L. Shao, ..., M. G. Gustafsson. 2012. Time-lapse two-color 3D imaging of live cells with doubled resolution using structured illumination. *Proc. Natl. Acad. Sci. USA*. 109:5311–5315.
26. Cognet, L., C. Leduc, and B. Lounis. 2014. Advances in live-cell single-particle tracking and dynamic super-resolution imaging. *Curr. Opin. Chem. Biol.* 20:78–85.
27. Betzig, E. 1995. Proposed method for molecular optical imaging. *Opt. Lett.* 20:237–239.
28. Betzig, E., G. H. Patterson, ..., H. F. Hess. 2006. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*. 313:1642–1645.
29. Hess, S. T., T. P. Girirajan, and M. D. Mason. 2006. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* 91:4258–4272.
30. Manley, S., J. M. Gillette, ..., J. Lippincott-Schwartz. 2008. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat. Methods*. 5:155–157.
31. Rossier, O., V. Oceau, ..., G. Giannone. 2012. Integrins $\beta 1$ and $\beta 3$ exhibit distinct dynamic nanoscale organizations inside focal adhesions. *Nat. Cell Biol.* 14:1057–1067.
32. Gudheti, M. V., N. M. Curthoys, ..., S. T. Hess. 2013. Actin mediates the nanoscale membrane organization of the clustered membrane protein influenza hemagglutinin. *Biophys. J.* 104:2182–2192.
33. Fu, G., T. Huang, ..., J. Xiao. 2010. In vivo structure of the *E. coli* FtsZ-ring revealed by photoactivated localization microscopy (PALM). *PLoS ONE*. 5:e12682.
34. Rust, M. J., M. Bates, and X. Zhuang. 2006. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods*. 3:793–795.
35. Wombacher, R., M. Heidbreder, ..., M. Sauer. 2010. Live-cell super-resolution imaging with trimethoprim conjugates. *Nat. Methods*. 7:717–719.
36. Heilemann, M., S. van de Linde, ..., M. Sauer. 2009. Super-resolution imaging with small organic fluorophores. *Angew. Chem. Int. Ed. Engl.* 48:6903–6908.
37. Schermelleh, L., R. Heintzmann, and H. Leonhardt. 2010. A guide to super-resolution fluorescence microscopy. *J. Cell Biol.* 190:165–175.
38. Huang, B., W. Wang, ..., X. Zhuang. 2008. Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. *Science*. 319:810–813.
39. Dani, A., B. Huang, ..., X. Zhuang. 2010. Super-resolution imaging of chemical synapses in the brain. *Neuron*. 68:843–856.
40. Löschberger, A., S. van de Linde, ..., M. Sauer. 2012. Super-resolution imaging visualizes the eightfold symmetry of gp210 proteins around the nuclear pore complex and resolves the central channel with nanometer resolution. *J. Cell Sci.* 125:570–575.
41. Jones, S. A., S. H. Shim, ..., X. Zhuang. 2011. Fast, three-dimensional super-resolution imaging of live cells. *Nat. Methods*. 8:499–508.
42. Wang, W., G. W. Li, ..., X. Zhuang. 2011. Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*. 333:1445–1449.
43. Klein, T., A. Löschberger, ..., M. Sauer. 2011. Live-cell dSTORM with SNAP-tag fusion proteins. *Nat. Methods*. 8:7–9.
44. Benke, A., and S. Manley. 2012. Live-cell dSTORM of cellular DNA based on direct DNA labeling. *ChemBioChem*. 13:298–301.
45. Shim, S. H., C. Xia, ..., X. Zhuang. 2012. Super-resolution fluorescence imaging of organelles in live cells with photoswitchable membrane probes. *Proc. Natl. Acad. Sci. USA*. 109:13978–13983.
46. Benke, A., N. Olivier, ..., S. Manley. 2012. Multicolor single molecule tracking of stochastically active synthetic dyes. *Nano Lett.* 12:2619–2624.
47. Wilmes, S., M. Staufenbiel, ..., J. Piehler. 2012. Triple-color super-resolution imaging of live cells: resolving submicroscopic receptor organization in the plasma membrane. *Angew. Chem. Int. Ed. Engl.* 51:4868–4871.
48. Fölling, J., M. Bossi, ..., S. W. Hell. 2008. Fluorescence nanoscopy by ground-state depletion and single-molecule return. *Nat. Methods*. 5:943–945.
49. Testa, I., C. A. Wurm, ..., C. Eggeling. 2010. Multicolor fluorescence nanoscopy in fixed and living cells by exciting conventional fluorophores with a single wavelength. *Biophys. J.* 99:2686–2694.
50. Lalkens, B., I. Testa, ..., S. W. Hell. 2012. MRT letter: Nanoscopy of protein colocalization in living cells by STED and GSDIM. *Microsc. Res. Tech.* 75:1–6.
51. Lukinavičius, G., K. Umezawa, ..., K. Johnsson. 2013. A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat. Chem.* 5:132–139.
52. Sharonov, A., and R. M. Hochstrasser. 2006. Wide-field subdiffraction imaging by accumulated binding of diffusing probes. *Proc. Natl. Acad. Sci. USA*. 103:18911–18916.
53. Cognet, L., D. A. Tsybouski, and R. B. Weisman. 2008. Subdiffraction far-field imaging of luminescent single-walled carbon nanotubes. *Nano Lett.* 8:749–753.
54. Giannone, G., E. Hossy, ..., L. Cognet. 2010. Dynamic super-resolution imaging of endogenous proteins on living cells at ultra-high density. *Biophys. J.* 99:1303–1310.
55. Giannone, G., E. Hossy, ..., L. Cognet. 2013. High-content super-resolution imaging of live cell by uPAINT. *Methods Mol. Biol.* 950:95–110.
56. Winckler, P., L. Lartigue, ..., L. Cognet. 2013. Identification and super-resolution imaging of ligand-activated receptor dimers in live cells. *Sci Rep.* 3:2387.
57. Nair, D., E. Hossy, ..., J. B. Sibarita. 2013. Super-resolution imaging reveals that AMPA receptors inside synapses are dynamically organized in nanodomains regulated by PSD95. *J. Neurosci.* 33:13204–13224.
58. Ondrus, A. E., H. L. Lee, ..., J. Du Bois. 2012. Fluorescent saxitoxins for live cell imaging of single voltage-gated sodium ion channels beyond the optical diffraction limit. *Chem. Biol.* 19:902–912.
59. Chmyrov, A., J. Keller, ..., S. W. Hell. 2013. Nanoscopy with more than 100,000 ‘doughnuts’. *Nat. Methods*. 10:737–740.
60. Rochefort, N. L., and A. Konnerth. 2012. Dendritic spines: from structure to in vivo function. *EMBO Rep.* 13:699–708.
61. Berning, S., K. I. Willig, ..., S. W. Hell. 2012. Nanoscopy in a living mouse brain. *Science*. 335:551.
62. Lauterbach, M. A., J. Keller, ..., S. W. Hell. 2010. Comparing video-rate STED nanoscopy and confocal microscopy of living neurons. *J. Biophotonics*. 3:417–424.
63. Bobroff, N. 1986. Position measurement with a resolution and noise-limited instrument. *Rev. Sci. Instrum.* 57:1152–1157.
64. Thompson, R. E., D. R. Larson, and W. W. Webb. 2002. Precise nanometer localization analysis for individual fluorescent probes. *Biophys. J.* 82:2775–2783.

65. Chenouard, N., I. Smal, ..., E. Meijering. 2014. Objective comparison of particle tracking methods. *Nat. Methods*. 11:281–289.
66. Swilius, M. T., and G. J. Jensen. 2012. The helical MreB cytoskeleton in *Escherichia coli* MC1000/pLE7 is an artifact of the N-Terminal yellow fluorescent protein tag. *J. Bacteriol.* 194:6382–6386.
67. Stagge, F., G. Y. Mitronova, ..., S. Jakobs. 2013. SNAP-, CLIP- and Halo-tag labelling of budding yeast cells. *PLoS ONE*. 8:e78745.
68. Pavani, S. R., M. A. Thompson, ..., W. E. Moerner. 2009. Three-dimensional, single-molecule fluorescence imaging beyond the diffraction limit by using a double-helix point spread function. *Proc. Natl. Acad. Sci. USA*. 106:2995–2999.
69. Izeddin, I., M. El Beheiry, ..., M. Dahan. 2012. PSF shaping using adaptive optics for three-dimensional single-molecule super-resolution imaging and tracking. *Opt. Express*. 20:4957–4967.
70. Juette, M. F., T. J. Gould, ..., J. Bewersdorf. 2008. Three-dimensional sub-100 nm resolution fluorescence microscopy of thick samples. *Nat. Methods*. 5:527–529.
71. Xu, K., H. P. Babcock, and X. Zhuang. 2012. Dual-objective STORM reveals three-dimensional filament organization in the actin cytoskeleton. *Nat. Methods*. 9:185–188.
72. Chasles, F., B. Dubertret, and A. C. Boccara. 2006. Full-field optical sectioning and three-dimensional localization of fluorescent particles using focal plane modulation. *Opt. Lett.* 31:1274–1276.
73. Shtengel, G., J. A. Galbraith, ..., H. F. Hess. 2009. Interferometric fluorescent super-resolution microscopy resolves 3D cellular ultrastructure. *Proc. Natl. Acad. Sci. USA*. 106:3125–3130.
74. Kanchanawong, P., G. Shtengel, ..., C. M. Waterman. 2010. Nanoscale architecture of integrin-based cell adhesions. *Nature*. 468:580–584.
75. Sibarita, J. B. 2005. Deconvolution microscopy. *Adv. Biochem. Eng. Biotechnol.* 95:201–243.
76. Dertinger, T., R. Colyer, ..., J. Enderlein. 2009. Fast, background-free, 3D super-resolution optical fluctuation imaging (SOFI). *Proc. Natl. Acad. Sci. USA*. 106:22287–22292.
77. Cox, S., E. Rosten, ..., R. Heintzmann. 2012. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nat. Methods*. 9:195–200.
78. Studer, V., J. Bobin, ..., M. Dahan. 2012. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Proc. Natl. Acad. Sci. USA*. 109:E1679–E1687.
79. Zhu, L., W. Zhang, ..., B. Huang. 2012. Faster STORM using compressed sensing. *Nat. Methods*. 9:721–723.
80. Zanella, R., G. Zanghirati, ..., G. Vicidomini. 2013. Towards real-time image deconvolution: application to confocal and STED microscopy. *Sci. Rep.* 3:2523.

Indolicidin Binding Induces Thinning of a Lipid Bilayer

Chris Neale,^{†‡} Jenny C. Y. Hsu,^{‡§} Christopher M. Yip,^{‡§} and Régis Pomès^{†*}

[†]Molecular Structure and Function, The Hospital for Sick Children, Toronto, Ontario, Canada; [‡]Department of Biochemistry and [§]Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Ontario, Canada

ABSTRACT We use all-atom molecular dynamics simulations on a massive scale to compute the standard binding free energy of the 13-residue antimicrobial peptide indolicidin to a lipid bilayer. The analysis of statistical convergence reveals systematic sampling errors that correlate with reorganization of the bilayer on the microsecond timescale and persist throughout a total of 1.4 ms of sampling. Consistent with experimental observations, indolicidin induces membrane thinning, although the simulations significantly overestimate the lipophilicity of the peptide.

Received for publication 11 October 2013 and in final form 21 February 2014.

*Correspondence: pomes@sickkids.ca

Antimicrobial peptides are a component of the innate immune system of eukaryotes (1). As such, they must interact with pathogenic membranes, either during translocation or by disrupting their structural integrity (2). Here we examine the binding of the 13-residue cationic antimicrobial peptide indolicidin (3) (ILPWKWPWWPWR-NH₂) to a lipid membrane as a first step towards elucidating its mechanism of action.

Molecular solutes interact with lipid membranes in many cellular processes (4). Computational approaches such as molecular dynamics simulations have been widely used to characterize these interactions (5). However, molecular dynamics simulations can require unfeasibly long times to reach equilibrium (6). Therefore, it is common to compute equilibrium properties of solute insertion into lipid bilayers using umbrella sampling (7) simulations in which the solute is restrained along the bilayer normal using harmonic restraining potentials, or umbrellas, centered at z_i^0 values distributed between bulk water and the bilayer center.

It is often assumed that equilibrium properties rapidly attain convergence in umbrella sampling simulations; accordingly, convergence measures are rarely published (8). However, we have recently shown that umbrella sampling simulations require up to 100 ns per umbrella (3 μ s in total) to eliminate systematic sampling errors in the standard free energy of binding, ΔG_{bind}^0 , of an arginine side-chain analog from bulk water to a lipid bilayer (8). The fact that umbrella sampling has been used to investigate the bilayer insertion of substantially larger solutes (9) motivates a systematic evaluation of statistical sampling convergence of ΔG_{bind}^0 for indolicidin in a lipid bilayer.

To estimate ΔG_{bind}^0 of indolicidin to a lipid bilayer, we conducted 60 sets of umbrella-sampling simulations while systematically varying the initial conformation. In each umbrella sampling simulation, each umbrella was simulated for 1.5 μ s, yielding a total simulation time of 1.4 ms and 60 independent free energy or potential of mean force (PMF) profiles from bulk water to the center

of a POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphatidylcholine) lipid bilayer.

The PMF profiles indicate that indolicidin strongly binds to the bilayer, partitioning inside the lipid headgroups (Fig. 1, *A* and *E*). Importantly, the mean estimate of ΔG_{bind}^0 decays exponentially with equilibration time t_{eq} , indicating that systematic sampling errors in individual simulations continued to decrease throughout the 1.5- μ s interval as rare events led to more favorable states (Fig. 1 *B*). The low frequency of transitions to more favorable states exacerbates the requirement for massive sampling using multiple independent simulations.

Computational limitations precluded extending all 60 sets of umbrella-sampling simulations to even longer times. Instead, we identified the two simulations at each umbrella that appeared to be most representative of equilibrium and extended each to 10 μ s per umbrella (see Methods in the Supporting Material). The resulting estimates of ΔG_{bind}^0 continued to decrease until $t_{\text{eq}} = 4 \mu$ s (68 μ s in total), after which they stabilized at the asymptotic limit of the exponential fit of the shorter simulation data ($\Delta G_{\text{bind}}^0 = -26 \pm 5$ kcal/mol; Fig. 1 *C*).

As indolicidin approaches the bilayer, it is drawn closer (Fig. 2 *A*) as salt bridges form between the peptide and the phospholipid headgroups (Fig. 2 *B*), inducing their protrusion (Figs. 1 *D* and 2 *C*). At large separation distances, this state is attained only when the peptide becomes highly extended (Fig. 2, *D* and *E*). As indolicidin is inserted more deeply, the surface of the lipid bilayer invaginates (Figs. 1 *E* and 2 *C*), maintaining peptide-lipid salt bridges (Fig. 2 *B*) and leading to the formation of a pore when the solute is near the bilayer center (Figs. 1 *F* and 2 *C*, and see Fig. S2, Fig. S3, Fig. S4, and Fig. S5 in the Supporting

Editor: Scott Feller.

© 2014 by the Biophysical Society

<http://dx.doi.org/10.1016/j.bpj.2014.02.031>



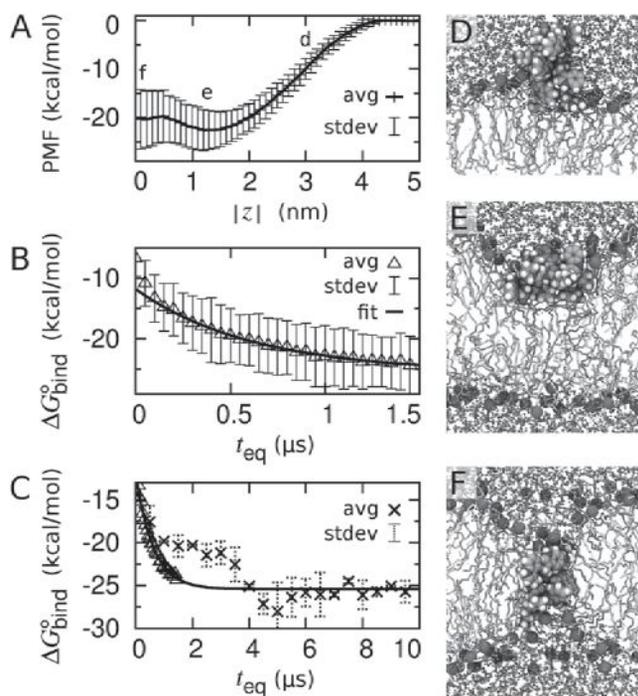


FIGURE 1 PMF for indolicidin partitioning into a POPC bilayer. (A) Average PMF from 60 independent umbrella-sampling simulations based on $1 < t \leq 1.5 \mu\text{s}$ /umbrella. (B) Average ΔG_{bind}^0 from 50-ns time intervals per umbrella ($t_{\text{eq}} < t \leq t_{\text{eq}} + 50 \text{ ns}$) as a function of equilibration time, t_{eq} . (Solid line) Single exponential fit to the mean over $0.5 < t_{\text{eq}} \leq 1.5 \mu\text{s}$. (C) Mean values of ΔG_{bind}^0 from the 10- μs /umbrella simulations (crosses) together with the mean values of ΔG_{bind}^0 (triangles) and exponential fit from panel B. PMF and ΔG_{bind}^0 profiles obtained from each of the 60 independent simulations are shown in Fig. S1 in the Supporting Material. (D–F) Representative conformations after 1.5 μs of simulation at $z_i^0 =$ (D) 3 nm, (E) 1.2 nm, and (F) 0.0 nm. To see this figure in color, go online.

Material). These Boltzmann-weighted ensemble averages may not be mechanistically representative of nonequilibrium binding events (8,10).

The reorganization of the peptide, the bilayer, and the ionic interactions between them became more pronounced with increasing simulation time at peptide insertion depths shallower than the global free energy minimum ($|z_i^0| > 1.4 \text{ nm}$; Fig. 1 A and Fig. 2, B–D). These conformational transitions are likely the source of the systematic drift of ΔG_{bind}^0 . Reorganization of the bilayer also controls the rate of equilibration during membrane insertion of an arginine side chain (8,9) and a cyclic arginine nonamer (11), suggesting that slow reorganization of lipids around cationic solutes presents a general impediment to simulation convergence.

Consistent with the perturbation of membrane thickness observed by in situ atomic force microscopy (12), our results suggest that indolicidin insertion induces local thinning of the bilayer (Fig. 1, E and F, and Fig. 2, C and F). The different conformational ensembles sampled by the

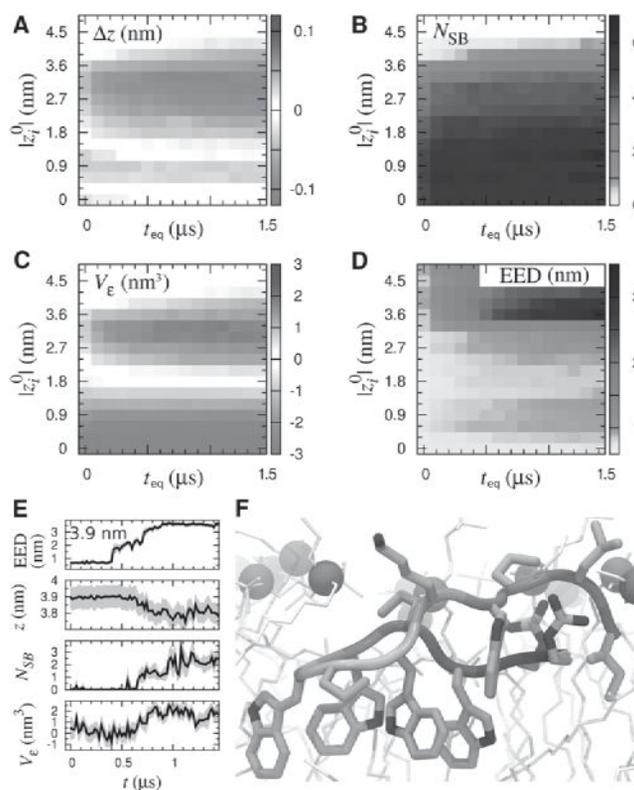


FIGURE 2 Slow equilibration of bilayer and peptide. (A–D) Color quantifies conformational reorganization for $t_{\text{eq}} < t \leq t_{\text{eq}} + 100 \text{ ns}$ as a function of t_{eq} and $|z_i^0|$. (A) Deviation of insertion depth, z , from z_i^0 , $\Delta z \equiv z - z_i^0$; (B) number of peptide-lipid salt bridges, N_{SB} ; (C) volume change of the bilayer's proximal leaflet in the radial vicinity of the solute, V_e ; and (D) peptide end-to-end distance (EED). There is no sampling for $t > 0.5 \mu\text{s}$ at $|z_i^0| \geq 4.5 \text{ nm}$. (E) Representative time-series of a trajectory at $z_i^0 = 3.9 \text{ nm}$. (F) Representative conformation at 10 μs for $|z_i^0| = 1.2 \text{ nm}$. To see this figure in color, go online.

peptide in water and in the lipid bilayer (Fig. 2 D) are consistent with the observations that indolicidin is disordered in solution (13) and adopts stable conformations in the presence of detergent (14). Although the peptide's conformation continued to change when it was deeply inserted (Fig. 2 D), the amount of water in the bilayer's hydrophobic core converged relatively rapidly (see Fig. S2). Indolicidin can induce the formation of hydrated, porelike defects (see Fig. S2, Fig. S3, Fig. S4, and Fig. S5) but does not act as a chloride carrier (see Fig. S6 and Fig. S7). Future studies of the mechanism of indolicidin action will examine the effect of multiple peptide binding.

The PMF profile presented in this Letter is strikingly different from that computed by Yeh et al. (15) using different force field parameters for indolicidin partitioning into a DMPC (1,2-dimyristoyl-*sn*-glycero-3-phosphatidylcholine) bilayer, from which the binding free energy was estimated to be 0 kcal/mol (15). However, that study comprised only 25 ns per umbrella and likely suffers from systematic sampling errors induced by initial conditions (see Fig. S8).

Our estimate of the binding affinity is much larger than the values obtained for indolicidin and large unilamellar POPC vesicles using isothermal titration calorimetry, -7.4 kcal/mol (16), and equilibrium dialysis, -8.8 kcal/mol (13). Such a discrepancy suggests that the relative accuracy of binding free energies for amino-acid side-chain analogs (8,9,17) does not necessarily extend to polypeptides. Although more work is needed to elucidate the source of this discrepancy, this study underlines the importance of attaining convergence before evaluating force-field accuracy.

Importantly, this work also highlights the extensive sampling required to remove systematic errors induced by initial conditions in atomistic simulations of peptides in membranes. Slow equilibration of the system is due to rare transitions across hidden free energy barriers involving reorganization of the membrane. Two simple recommendations are 1), evaluating the time-dependence of ensemble averages, and 2), conducting multiple simulations with different initial conditions. We have recently shown that by using enhanced sampling techniques it is possible to identify the locations of hidden free energy barriers without a priori knowledge (9). Future research will examine strategies for speeding up the crossing of these barriers, such as optimized order parameters including bilayer reorganization and enhanced sampling techniques including a random walk along the order parameter (9).

SUPPORTING MATERIAL

Supplemental Methods and Results, two tables, eight figures, three equations and references (18–24) are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)00275-6](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00275-6).

ACKNOWLEDGMENTS

Computations were performed at SciNet (25) and data was stored at SHARCNET, both of which are resources of Compute Canada.

This work was funded in part by National Sciences and Engineering Research Council discovery grant No. RGPIN-418679 and by Canadian Institutes of Health Research operating grants No. MOP-43998 and No. MOP-43949.

REFERENCES and FOOTNOTES

- Hancock, R. E. W. 2001. Cationic peptides: effectors in innate immunity and novel antimicrobials. *Lancet Infect. Dis.* 1:156–164.
- Epand, R. M., and H. J. Vogel. 1999. Diversity of antimicrobial peptides and their mechanisms of action. *Biophys. Biochim. Acta.* 1462: 11–28.
- Selsted, M. E., M. J. Novotny, ..., J. S. Cullor. 1992. Indolicidin, a novel bactericidal tridecapeptide amide from neutrophils. *J. Biol. Chem.* 267:4292–4295.
- Stein, W. 1986. Transport and Diffusion across Cell Membranes. Academic Press, Orlando, FL.
- Saiz, L., and M. L. Klein. 2002. Computer simulation studies of model biological membranes. *Acc. Chem. Res.* 35:482–489.
- Grossfield, A., S. E. Feller, and M. C. Pitman. 2007. Convergence of molecular dynamics simulations of membrane proteins. *Proteins.* 67:31–40.
- Torrie, G. M., and J. P. Valleau. 1977. Nonphysical sampling distributions in Monte Carlo free-energy estimation: umbrella sampling. *J. Comput. Phys.* 23:187–199.
- Neale, C., W. F. D. Bennett, ..., R. Pomès. 2011. Statistical convergence of equilibrium properties in simulations of molecular solutes embedded in lipid bilayers. *J. Chem. Theory Comput.* 7:4175–4188.
- Neale, C., C. Madill, ..., R. Pomès. 2013. Accelerating convergence in molecular dynamics simulations of solutes in lipid membranes by conducting a random walk along the bilayer normal. *J. Chem. Theory Comput.* 9:3686–3703.
- Kopelevich, D. I. 2013. One-dimensional potential of mean force underestimates activation barrier for transport across flexible lipid membranes. *J. Chem. Phys.* 139:134906.
- Huang, K., and A. E. García. 2013. Free energy of translocating an arginine-rich cell-penetrating peptide across a lipid bilayer suggests pore formation. *Biophys. J.* 104:412–420.
- Shaw, J. E., J.-R. Alattia, ..., C. M. Yip. 2006. Mechanisms of antimicrobial peptide action: studies of indolicidin assembly at model membrane interfaces by in situ atomic force microscopy. *J. Struct. Biol.* 154:42–58.
- Ladokhin, A. S., M. E. Selsted, and S. H. White. 1997. Bilayer interactions of indolicidin, a small antimicrobial peptide rich in tryptophan, proline, and basic amino acids. *Biophys. J.* 72:794–805.
- Rozeck, A., C. L. Friedrich, and R. E. W. Hancock. 2000. Structure of the bovine antimicrobial peptide indolicidin bound to dodecylphosphocholine and sodium dodecyl sulfate micelles. *Biochemistry.* 39:15765–15774.
- Yeh, I.-C., D. R. Ripoll, and A. Wallqvist. 2012. Free energy difference in indolicidin attraction to eukaryotic and prokaryotic model cell membranes. *J. Phys. Chem. B.* 116:3387–3396.
- Andrushchenko, V. V., M. H. Aarabi, ..., H. J. Vogel. 2008. Thermodynamics of the interactions of tryptophan-rich cathelicidin antimicrobial peptides with model and natural membranes. *Biochim. Biophys. Acta.* 1778:1004–1014.
- MacCallum, J. L., W. F. D. Bennett, and D. P. Tieleman. 2008. Distribution of amino acids in a lipid bilayer from computer simulations. *Biophys. J.* 94:3393–3404.
- Hess, B., C. Kutzner, ..., E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* 4:435–447.
- Jorgensen, W. L., J. Chandrasekhar, ..., M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
- Berger, O., O. Edholm, and F. Jähnig. 1997. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* 72:2002–2013.
- Kaminski, G. A., R. A. Friesner, ..., W. L. Jorgensen. 2001. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B.* 105:6474–6487.
- Kandt, C., W. L. Ash, and D. P. Tieleman. 2007. Setting up and running molecular dynamics simulations of membrane proteins. *Methods.* 41: 475–488.
- Ferrenberg, A. M., and R. H. Swendsen. 1988. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* 61:2635–2638.
- Rokitskaya, T. I., N. I. Kolodkin, ..., Y. N. Antonenko. 2011. Indolicidin action on membrane permeability: carrier mechanism versus pore formation. *Biochim. Biophys. Acta.* 1808:91–97.
- Loken, C., D. Gruner, ..., R. Van Zon. 2010. SciNet: lessons learned from building a power-efficient top-20 system and data centre. *J. Phys. Conf. Ser.* 256:012026.

Cell Career Network just got better!



- Additional reach via extended network
- Post a job within 5 minutes via self service
- More intuitive feel and easier navigation
- New vacancy packages and pricing

Your recruitment process just became even easier.
Get started today!

careers.cell.com/recruiters



Cell
PRESS

Biophysical Letter

3D-SIM Super-resolution of FtsZ and Its Membrane Tethers in *Escherichia coli* Cells

Veronica Wells Rowlett¹ and William Margolin^{1,*}

¹Department of Microbiology and Molecular Genetics, University of Texas Medical School at Houston, Houston, Texas

ABSTRACT FtsZ, a bacterial homolog of eukaryotic tubulin, assembles into the Z ring required for cytokinesis. In *Escherichia coli*, FtsZ interacts directly with FtsA and ZipA, which tether the Z ring to the membrane. We used three-dimensional structured illumination microscopy to compare the localization patterns of FtsZ, FtsA, and ZipA at high resolution in *Escherichia coli* cells. We found that FtsZ localizes in patches within a ring structure, similar to the pattern observed in other species, and discovered that FtsA and ZipA mostly colocalize in similar patches. Finally, we observed similar punctate and short polymeric structures of FtsZ distributed throughout the cell after Z rings were disassembled, either as a consequence of normal cytokinesis or upon induction of an endogenous cell division inhibitor.

Received for publication 16 May 2014 and in final form 21 August 2014.

*Correspondence: william.margolin@uth.tmc.edu

The assembly of the bacterial tubulin FtsZ has been well studied in vitro, but the fine structure of the cytokinetic Z ring it forms in vivo is not well defined. Super-resolution microscopy methods including photoactivated localization microscopy (PALM) and three-dimensional-structured illumination microscopy (3D-SIM) have recently provided a more detailed view of Z-ring structures. Two-dimensional PALM showed that Z rings in *Escherichia coli* are likely composed of loosely-bundled dynamic protofilaments (1,2). Three-dimensional PALM studies of *Caulobacter crescentus* initially showed that Z rings were comprised of loosely bundled protofilaments forming a continuous but dynamic ring (1–3). However, a more recent high-throughput study showed that the Z rings of this bacterium are patchy or discontinuous (4), similar to Z rings of *Bacillus subtilis* and *Staphylococcus aureus* using 3D-SIM (5). Strauss et al. (5) also demonstrated that the patches in *B. subtilis* Z rings are highly dynamic.

Assembly of the Z ring is modulated by several proteins that interact directly with FtsZ and enhance assembly or disassembly (6). For example, FtsA and ZipA promote ring assembly in *E. coli* by tethering it to the cytoplasmic membrane (7,8). Sula is an inhibitor of FtsZ assembly, induced only after DNA damage, which sequesters monomers of FtsZ to prevent its assembly into a Z ring (9). Our initial goals were to visualize Z rings in *E. coli* using 3D-SIM, and then examine whether any FtsZ polymeric structures remain after Sula induction. We also asked whether FtsA and ZipA localized in patchy patterns similar to those of FtsZ.

We used a DeltaVision OMX V4 Blaze microscope (Applied Precision, GE Healthcare, Issaquah, WA) to view the high-resolution localization patterns of FtsZ in *E. coli* cells producing FtsZ-GFP (Fig. 1). Three-dimensional views were reconstructed using softWoRx software

(Applied Precision). To rule out GFP artifacts, we also visualized native FtsZ from a wild-type strain (WM1074) by immunofluorescence (IF).

Both FtsZ-GFP (Fig. 1, A, B, and B1) and IF staining for FtsZ (Fig. 1, C, D, and D1) consistently localized to patches around the ring circumference, similar to the *B. subtilis* and *C. crescentus* FtsZ patterns (4,5). Analysis of fluorescence intensities (see Fig. S1, A and B, in the Supporting Material) revealed that the majority of Z rings contain one or more gaps in which intensity decreases to background levels (82% for FtsZ-GFP and 69% for IF). Most rings had 3–5 areas of lower intensity, but only a small percentage of these areas had fluorescence below background intensity (34% for FtsZ-GFP and 21% for IF), indicating that the majority of areas with lower intensity contain at least some FtsZ.

To elucidate how FtsZ transitions from a disassembled ring to a new ring, we imaged a few dividing daughter cells before they were able to form new Z rings (Fig. 1 E). Previous conventional microscopy had revealed dynamic FtsZ helical structures (10), but the resolution had been insufficient to see further details. Here, FtsZ visualized in dividing cells by 3D-SIM localized throughout as a mixture of patches and randomly-oriented short filaments (*asterisk* and *dashed oval* in Fig. 1, respectively). These structures may represent oligomeric precursors of Z ring assembly.

To visualize FtsZ after Z-ring disassembly another way, we overproduced Sula, a protein that blocks FtsZ assembly. We examined *E. coli* cells producing FtsZ-GFP after

Editor: Sean Sun.

© 2014 by the Biophysical Society
<http://dx.doi.org/10.1016/j.bpj.2014.08.024>



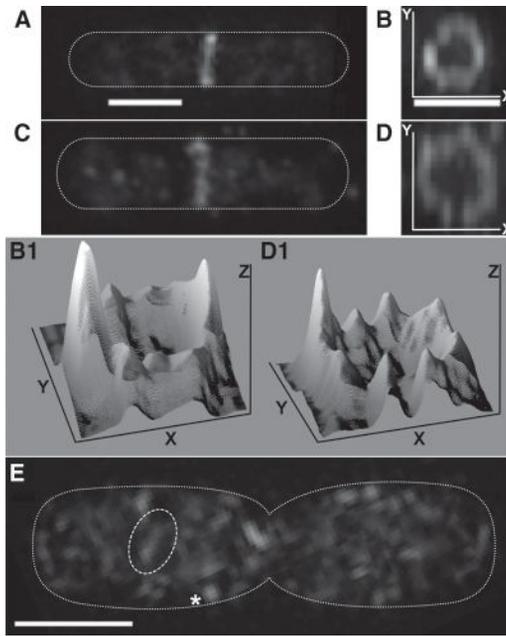


FIGURE 1 Localization of FtsZ in *E. coli*. (A) Cell with a Z ring labeled with FtsZ-GFP. (B) Rotated view of Z ring in panel A. (C) Cell with a Z ring labeled with DyLight 550 (Thermo Fisher Scientific, Waltham, MA). (D) Rotated view of Z ring in panel C. (B1 and D1) Three-dimensional surface intensity plots of Z rings in panels B and D, respectively. (E) A dividing cell producing FtsZ-GFP. The cell outline is shown in the schematic. (Asterisk) Focus of FtsZ localization; (open dashed ovals) filamentous structures of FtsZ. Three-dimensional surface intensity plots were created using the software ImageJ (19). Scale bars, 1 μm .

induction of *sulA* expression from a pBAD33-*sulA* plasmid (pWM1736) with 0.2% arabinose. After 30 min of *sulA* induction, Z rings remained intact in most cells (Fig. 2 A and data not shown). The proportion of cellular FtsZ-GFP in the ring before and after induction of *sulA* was consistent with previous data (data not shown) (1,11).

Notably, after 45 min of *sulA* induction, Z rings were gone (Fig. 2, B and B1), replaced by numerous patches and randomly-oriented short filaments (asterisk and dashed ovals in Fig. 2), similar to those observed in a dividing cell. FtsZ normally rapidly recycles from free monomers to ring-bound polymers (11), but a critical concentration of SulA reduces the pool of available FtsZ monomers, resulting in breakdown of the Z ring (9). The observed FtsZ-GFP patches and filaments are likely FtsZ polymers that disassemble before they can organize into a ring.

We confirmed this result by overproducing SulA in wild-type cells and detecting FtsZ localization by IF (Fig. 2, C, D, and D1). The overall fluorescence patterns in cells producing FtsZ-GFP versus cells producing only native FtsZ were similar (Fig. 2, B1 and D1), although we observed fewer filaments with IF, perhaps because FtsZ-GFP confers slight resistance to SulA, or because the increased amount of FtsZ in FtsZ-GFP producing cells might titrate the SulA more effectively.

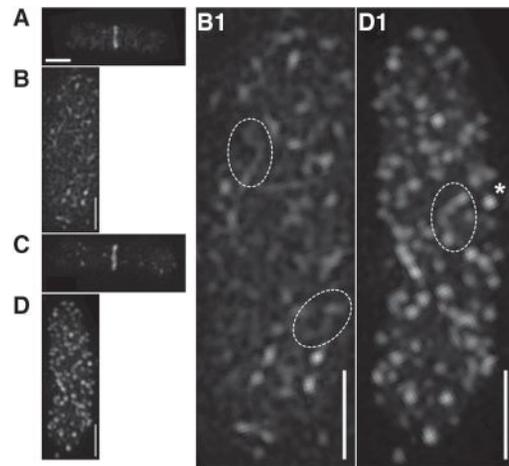


FIGURE 2 Localization of FtsZ after overproduction of SulA. (A) Cell producing FtsZ-GFP after 0.2% arabinose induction of SulA for 30 min. (B) After 45 min. (B1) Magnified cell shown in panel B. (C) Cell producing native FtsZ labeled with AlexaFluor 488 (Life Technologies, Carlsbad, CA) 30 min after induction; (D) 45 min after induction. (D1) Magnified cell shown in panel D. Scale bars, 1 μm . (Asterisk) Focus of FtsZ localization; (open dashed ovals) filamentous structures of FtsZ.

Additionally, we wanted to observe the localization patterns of the membrane tethers FtsA and ZipA. Inasmuch as both proteins bind to the same C-terminal conserved tail of FtsZ (12–14), they would be expected to colocalize with the circumferential FtsZ patches in the Z ring. We visualized FtsA using protein fusions to mCherry and GFP (data not shown) as well as IF using a wild-type strain (WM1074) (Fig. 3 A). We found that the patchy ring pattern of FtsA localization was similar to the FtsZ pattern. ZipA also displayed a similar patchy localization in WM1074 by IF (Fig. 3 B).

To determine whether FtsA and ZipA colocalized to these patches, we used a strain producing FtsA-GFP (WM4679) for IF staining of ZipA using a red secondary antibody. FtsA-GFP (Fig. 3 C) and ZipA (Fig. 3 D) had similar patterns of fluorescence, although the three-dimensional intensity profiles (Fig. 3, C1 and D1) reveal slight differences in intensity that are also visible in a merged image (Fig. 3 E). Quantitation of fluorescence intensities around the circumference of the rings revealed that FtsA and ZipA colocalized almost completely in approximately half of the rings analyzed (Fig. 3 F, and see Fig. S2 A), whereas in the other rings there were significant differences in localization in one or more areas (see Fig. S2 B). FtsA and ZipA bind to the same C-terminal peptide of FtsZ and may compete for binding. Cooperative self-assembly of FtsA or ZipA might result in large-scale differential localization visible by 3D-SIM.

In conclusion, our 3D-SIM analysis shows that the patchy localization of FtsZ is conserved in *E. coli* and suggests that it may be widespread among bacteria. After disassembly of the Z ring either in dividing cells or by excess levels of the cell division inhibitor SulA, FtsZ persisted as patches and short filamentous structures. This is consistent with a highly

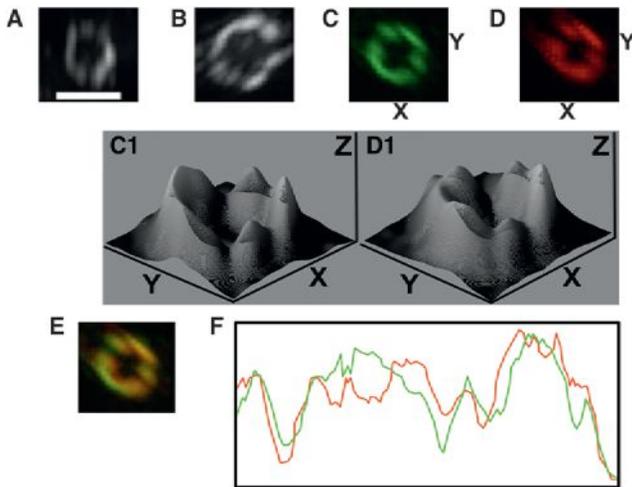


FIGURE 3 Localization of FtsA (A) and ZipA (B) by IF using AlexaFluor 488. (C) FtsA-GFP ring. (D) Same cell shown in panel C with ZipA labeled with DyLight 550. (C1 and D1) Three-dimensional surface intensity plots of FtsA ring from panel C or ZipA ring from panel D, respectively. (E) Merged image of FtsA (green) and ZipA (red) from the ring shown in panels C and D. (F) Intensity plot of FtsA (green) and ZipA (red) of ring shown in panel E. The plot represents intensity across a line drawn counterclockwise from the top of the ring around the circumference, then into its lumen. Red/green intensity plot and three-dimensional surface intensity plots were created using the software ImageJ (19). Scale bar, 1 μm .

dynamic population of FtsZ monomers and oligomers outside the ring, originally observed as mobile helices in *E. coli* by conventional fluorescence microscopy (10) and by photoactivation single-molecule tracking (15). FtsA and ZipA, which bind to the same segment of FtsZ and tether it to the cytoplasmic membrane, usually display a similar localization pattern to FtsZ and each other, although in addition to the differences we detect by 3D-SIM, there are also likely differences that are beyond its $\sim 100\text{-nm}$ resolution limit in the X,Y plane.

As proposed previously (16), gaps between FtsZ patches may be needed to accommodate a switch from a sparse Z ring to a more condensed ring, which would provide force to drive ring constriction (17). If this model is correct, the gaps should close upon ring constriction, although this may be beyond the resolution of 3D-SIM in constricted rings. Another role for patches could be to force molecular crowding of low-abundance septum synthesis proteins such as FtsI, which depend on FtsZ/FtsA/ZipA for their recruitment, into a few mobile supercomplexes.

How are FtsZ polymers organized within the Z-ring patches? Recent polarized fluorescence data suggest that FtsZ polymers are oriented both axially and circumferentially within the Z ring in *E. coli* (18). The seemingly random orientation of the non-ring FtsZ polymeric structures we observe here supports the idea that there is no strong constraint requiring FtsZ oligomers to follow a circumferential path around the cell cylinder. The patches of FtsZ in the

unperturbed *E. coli* Z ring likely represent randomly oriented clusters of FtsZ filaments that are associated with ZipA, FtsA, and essential septum synthesis proteins. New super-resolution microscopy methods should continue to shed light on the in vivo organization of these protein assemblies.

SUPPORTING MATERIAL

Preparation of Samples for 3D-SIM and two figures are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)00895-9](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00895-9).

ACKNOWLEDGMENTS

We thank Anna and Tomasz Zal for training on the OMX V4 Blaze system and technical suggestions.

This work was supported by National Institutes of Health grant No. GM61074 and the Graduate School of Biomedical Sciences.

SUPPORTING CITATIONS

Reference (20) appears in the Supporting Material.

REFERENCES and FOOTNOTES

1. Fu, G., T. Huang, ..., J. Xiao. 2010. In vivo structure of the *E. coli* FtsZ-ring revealed by photoactivated localization microscopy (PALM). *PLoS ONE*. 5:e12682.
2. Buss, J., C. Coltharp, ..., J. Xiao. 2013. In vivo organization of the FtsZ-ring by ZapA and ZapB revealed by quantitative super-resolution microscopy. *Mol. Microbiol.* 89:1099–1120.
3. Biteen, J. S., E. D. Goley, ..., W. E. Moerner. 2012. Three-dimensional super-resolution imaging of the midplane protein FtsZ in live *Caulobacter crescentus* cells using astigmatism. *ChemPhysChem*. 13:1007–1012.
4. Holden, S. J., T. Pengo, ..., S. Manley. 2014. High throughput 3D super-resolution microscopy reveals *Caulobacter crescentus* in vivo Z-ring organization. *Proc. Natl. Acad. Sci. USA*. 111:4566–4571.
5. Strauss, M. P., A. T. F. Liew, ..., E. J. Harry. 2012. 3D-SIM super resolution microscopy reveals a bead-like arrangement for FtsZ and the division machinery: implications for triggering cytokinesis. *PLoS Biol.* 10:e1001389.
6. Romberg, L., and P. A. Levin. 2003. Assembly dynamics of the bacterial cell division protein FtsZ: poised at the edge of stability. *Annu. Rev. Microbiol.* 57:125–154.
7. Pichoff, S., and J. Lutkenhaus. 2002. Unique and overlapping roles for ZipA and FtsA in septal ring assembly in *Escherichia coli*. *EMBO J.* 21:685–693.
8. Pichoff, S., and J. Lutkenhaus. 2005. Tethering the Z ring to the membrane through a conserved membrane targeting sequence in FtsA. *Mol. Microbiol.* 55:1722–1734.
9. Chen, Y., S. L. Milam, and H. P. Erickson. 2012. Sula inhibits assembly of FtsZ by a simple sequestration mechanism. *Biochemistry*. 51:3100–3109.
10. Thanedar, S., and W. Margolin. 2004. FtsZ exhibits rapid movement and oscillation waves in helix-like patterns in *Escherichia coli*. *Curr. Biol.* 14:1167–1173.
11. Anderson, D. E., F. J. Gueiros-Filho, and H. P. Erickson. 2004. Assembly dynamics of FtsZ rings in *Bacillus subtilis* and *Escherichia coli* and effects of FtsZ-regulating proteins. *J. Bacteriol.* 186:5775–5781.
12. Ma, X., and W. Margolin. 1999. Genetic and functional analyses of the conserved C-terminal core domain of *Escherichia coli* FtsZ. *J. Bacteriol.* 181:7531–7544.

13. Szwedziak, P., Q. Wang, ..., J. Löwe. 2012. FtsA forms actin-like protofilaments. *EMBO J.* 31:2249–2260.
14. Mosyak, L., Y. Zhang, ..., W. S. Somers. 2000. The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by x-ray crystallography. *EMBO J.* 19:3179–3191.
15. Niu, L., and J. Yu. 2008. Investigating intracellular dynamics of FtsZ cytoskeleton with photoactivation single-molecule tracking. *Biophys. J.* 95:2009–2016.
16. Lan, G., B. R. Daniels, ..., S. X. Sun. 2009. Condensation of FtsZ filaments can drive bacterial cell division. *Proc. Natl. Acad. Sci. USA.* 106:121–126.
17. Osawa, M., and H. P. Erickson. 2011. Inside-out Z rings—constriction with and without GTP hydrolysis. *Mol. Microbiol.* 81:571–579.
18. Si, F., K. Busiek, ..., S. X. Sun. 2013. Organization of FtsZ filaments in the bacterial division ring measured from polarized fluorescence microscopy. *Biophys. J.* 105:1976–1986.
19. Schneider, C. A., W. S. Rasband, and K. W. Eliceiri. 2012. NIH image to ImageJ: 25 years of image analysis. *Nat. Methods.* 9:671–675.
20. Levin, P. A. 2002. Light microscopy techniques for bacterial cell biology. *In* *Methods in Microbiology*, Vol. 31: Molecular Cellular Microbiology. P. Sansonetti and A. Zychlinsky, editors. Academic Press, London, United Kingdom, pp. 115–132.

Introducing the China Gateway

The home for Chinese scientists



The new Cell Press China Gateway is the ultimate destination for Chinese scientists. Now highlights of the latest research, regional events, and featured content from Cell Press are accessible online, on a single webpage.

The Cell Press China Gateway was designed with you in mind. We hope you'll make yourself at home.

Visit us at www.cell.com/china

CellPress

Article

Explicit Tracking of Uncertainty Increases the Power of Quantitative Rule-of-Thumb Reasoning in Cell Biology

Iain G. Johnston,¹ Benjamin C. Rickett,¹ and Nick S. Jones^{1,*}¹Department of Mathematics, Imperial College London, London, United Kingdom

ABSTRACT Back-of-the-envelope or rule-of-thumb calculations involving rough estimates of quantities play a central scientific role in developing intuition about the structure and behavior of physical systems, for example in so-called Fermi problems in the physical sciences. Such calculations can be used to powerfully and quantitatively reason about biological systems, particularly at the interface between physics and biology. However, substantial uncertainties are often associated with values in cell biology, and performing calculations without taking this uncertainty into account may limit the extent to which results can be interpreted for a given problem. We present a means to facilitate such calculations where uncertainties are explicitly tracked through the line of reasoning, and introduce a probabilistic calculator called CALADIS, a free web tool, designed to perform this tracking. This approach allows users to perform more statistically robust calculations in cell biology despite having uncertain values, and to identify which quantities need to be measured more precisely to make confident statements, facilitating efficient experimental design. We illustrate the use of our tool for tracking uncertainty in several example biological calculations, showing that the results yield powerful and interpretable statistics on the quantities of interest. We also demonstrate that the outcomes of calculations may differ from point estimates when uncertainty is accurately tracked. An integral link between CALADIS and the BioNumbers repository of biological quantities further facilitates the straightforward location, selection, and use of a wealth of experimental data in cell biological calculations.

INTRODUCTION

Rule-of-thumb, or back-of-the-envelope, calculations are of great utility across the sciences, allowing estimates of quantities to be obtained while gleaning intuition about the important numerical features of a system. In physics, the paradigm of the Fermi problem has been used for decades to develop intuition about the structure and behavior of systems by employing reasonable approximations, order-of-magnitude estimates, dimensional analysis, and clearly stated assumptions. The use of the napkin (often more readily available than an envelope in modern cafés and conferences) as a medium to perform rough calculations and gain understanding of a system given limited experimental information is well known in the physical sciences and has recently gained popular attention (1). Recent mathematical approaches to complex problems in wider scientific fields have employed these back-of-the-envelope approaches, including bioestimates in physical biology (2) and cell biology (3) and the popular “street-fighting mathematics” for use throughout the sciences (4).

However, these calculations currently do not have as central a role in cell biology as they do in the physical sciences,

despite receiving substantial recent attention as powerful tools for reasoning in quantitative biology (5,6), and being facilitated by quantitative resources like the excellent BioNumbers database (7). One reason for this absence is that many of the quantities involved in cell biology are either intrinsically highly variable or have large measurement errors. Calculations that do not take these uncertainties into account (yielding a mean value estimate without associated uncertainties), although powerful in their own right, may represent only part of the story (Fig. 1 A).

In some back-of-the-envelope circumstances, accuracy may be maintained without the explicit tracking of uncertainties. An example of this is in calculations involving the multiplication of several terms, each of which may be reasonably assumed to be normally distributed with similar coefficients of variation. In such a calculation, the logarithm of the error in an estimate scales with the square-root of the number of terms in the calculation. However, quantitative cell biology often involves distributions that cannot be assumed to be normally distributed, as well as calculations more general than simple multiplications of terms. In these circumstances, where individual uncertainties can differ between terms and may be over many orders of magnitude, the risk of inaccuracy associated with calculations without uncertainty is increased. If uncertainties are included in such calculations, it is often through standard propagation-of-uncertainty approaches (8), which typically track a limited

Submitted June 25, 2014, and accepted for publication August 29, 2014.

*Correspondence: nick.jones@imperial.ac.uk

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/3.0/>).

Editor: Daniel Beard.

© 2014 The Authors
0006-3495/14/12/2612/6 \$2.00

<http://dx.doi.org/10.1016/j.bpj.2014.08.040>



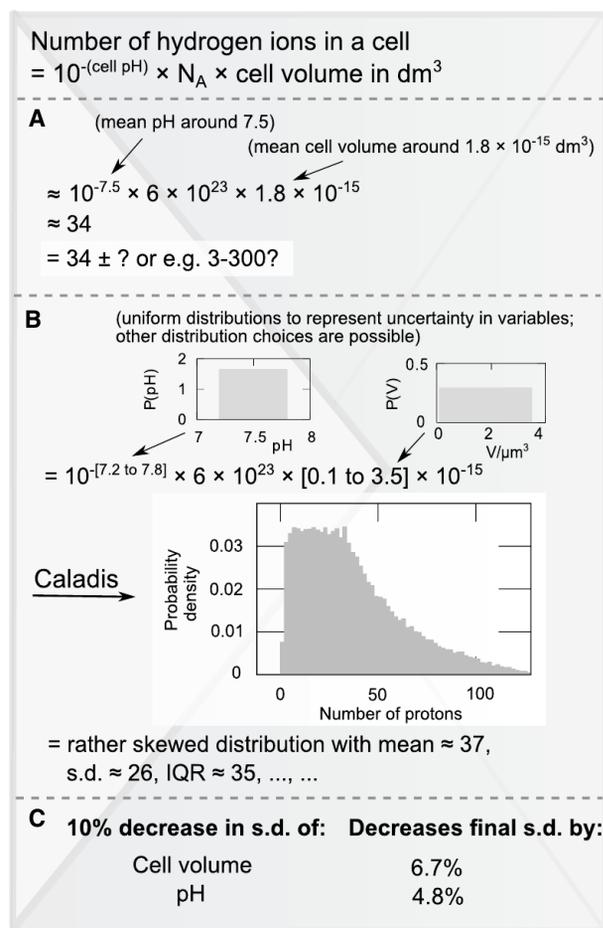


FIGURE 1 An example back-of-the-envelope calculation. Current technology is unable to measure the number of protons in a cell, so we estimate this number from measured quantities. (A) An estimate without uncertainty, combining rough estimates of pH and cell volume to obtain a guess for the number of protons. In this example, mean values are chosen to match the means of known measurements, but no associated uncertainty is analyzed. (B) An estimate using CALADIS to explicitly account for uncertainties in the measured quantities and reporting explanatory statistics about the final quantity, using uniform distributions to represent the uncertainty in the variables involved. Other representative distributions are possible and can be analyzed using our approach (see Results). (C) CALADIS also finds that in this example calculation, more of the final uncertainty arises from uncertainty in cell volume than pH: refining volume estimates is slightly preferred as the optimal experimental strategy to lower overall uncertainty. To see this figure in color, go online.

number of distribution moments and can thereby fail to accurately represent the distribution of the final result for nonnormal distributions. Of course, the process of performing rough calculations and obtaining estimated answers is immensely valuable in its own right, for the reasons discussed above. To complement this powerful process of Fermi reasoning in biology, we here suggest a complementary form of envelope reasoning, allowing for calculations including uncertain quantities.

METHODS

Explicitly tracking uncertainty in cell biological calculations

We propose an approach to biological rule-of-thumb calculations involving uncertain quantities that does not solely rely on point estimates of quantities of interest. Instead, our approach involves treating every uncertain quantity in a rule of thumb calculation as a probability distribution describing this uncertainty. The following iterative process is then performed: in each iteration, a sampled value is taken from each distribution of interest in the calculation. The value of the complete calculation is computed given this set of samples. This process is iterated many times to build up a distribution of values describing the output of the calculation. This output distribution then provides an interpretable and statistically rigorous answer to the rule-of-thumb question. We present this approach as a complement to, and not a substitute for, the valuable process of Fermi estimation, and stress again the value of “just having a go with the numbers”.

We emphasize that our approach, calculation of quantities using samples from distributions rather than point estimates, can be used to obtain interpretable results in cases where we do not have access to the full set of original measurements. This situation is likely to apply, for example, when using summarized results from previous independent experiments. In this case, our method can be viewed as a generalization of the resampling approaches that could be used if we had full access to the original data, such as bootstrapping or jackknifing (9).

In addition to adding statistical power to rule-of-thumb questions in cell biology, this approach can also be used to facilitate efficient design of experiments to reduce uncertainty in a given quantity. In the picture of calculations performed using probability distributions, this goal can be accomplished using a simple variant of a sensitivity analysis approach. Consider artificially decreasing the variance of each distribution in a calculation one-by-one. Decreasing the variance of each individual distribution will lead to a decrease in the overall variance of the output distribution, and the magnitudes of these induced overall decreases can be recorded. The quantity with the most power to decrease overall variance in the calculation output can then be identified, and its value further refined through experiment. Conceptually, this approach resembles performing a sensitivity analysis on the variance of the solution distribution with respect to the variances of individual input distributions.

An important point to consider when attempting to quantify uncertainty in scientific calculations is the source and meaning of the word “uncertainty”. A degree of measurement error may be associated with an experimental protocol, causing uncertainty in the resulting value due to imprecision. Alternatively, a given physical or biological quantity may exhibit genuine variability independent of the measurement process, in that its value fluctuates or changes with time and/or other controlling factors. The degree to which calculations involving uncertain quantities are interpretable is contingent on the types of uncertainty involved (see Discussion).

CALADIS: a probabilistic calculator for biology

We introduce a web-based calculator called CALADIS (from “calculate a distribution”), available for free use (and free source code download) at www.caladis.org. CALADIS, in addition to computing with constant quantities and standard mathematical operators and functions, naturally incorporates probability distributions as fundamental calculation elements, yielding as its output a probability distribution over the final answer. As described above, this probabilistic calculation approach allows uncertainties to be tracked throughout a calculation, providing a wealth of output data and allowing a complete view of the statistical details of the output of a probabilistic calculation (Fig. 1 B) and further information about the sources of uncertainty (Fig. 1 C; see later).

We underline that our web tool requires no knowledge of computer programming and no access to mathematical software tools, and, in addition to

functioning on desktop and laptop browsers, is compatible with a range of hand-held devices. Our aim in designing this tool is to facilitate fast and easy calculations involving uncertain biological quantities for users including those who lack the background or software to produce their own machinery for performing such calculations. The ability of our site to function on mobile devices makes it a plausible substitute for the well-known napkin over coffee or a conference dinner, facilitating informal but rigorous rough estimates of quantities as new ideas emerge.

CALADIS presents the user with a field (Fig. 2 A) to input calculation expressions, which may involve probability distributions identified with a prepended # symbol, where # functions as a sigil denoting a distribution, e.g.,

$$4/3 * \pi * \# \text{cellRadiusDist} ^ 3.$$

For every probability distribution found in the input expression, CALADIS prompts the user to choose a distribution type, and appropriate parameters to describe that distribution (for example, perhaps specifying that #cellRadiusDist is a uniform distribution between 1 and 1.5 μm), or, in the case of BioNumbers (see below), automatically populates the distribution details with the appropriate parameters (Fig. 2 B). Users may also use a built-in browser to input distributions corresponding to recorded quantities from biological experiments (Fig. 2 C; see BioNumbers below). The user may then click “Calculate”, whereupon CALADIS computes a probability distribution describing the final answer using the above approach, sampling many times from each distribution the user entered to build up a set of sam-

ples from the resultant distribution, which is then displayed graphically (Fig. 2 D). This interface includes a tool to estimate the probability mass between two given values, user-controlled display of the probability of lying in each bin, summary statistics of the distribution (Fig. 2 E), results from the optional standard deviation (SD) analysis (Fig. 2 F), and a URL that serves as a permanent link to that calculation. This collection of output statistics and graphics allows a complete overview of the probabilistic result of the user’s calculation.

CALADIS also facilitates the aforementioned efficient design of experimental strategies, through consideration of the contributions of different quantities to the overall uncertainty in a calculation. The user has the option of performing a standard deviation analysis for common types of input distribution in the web interface. In this analysis, the SD of each input distribution of this type is artificially reduced by 10%, and the resulting effect on the SD and interquartile range of the resultant distribution is recorded (Fig. 2 F). Intuition about the input variable with the most power to refine the overall output estimate can then be gained straightforwardly.

BioNumbers

We have embedded the data provided by the BioNumbers repository (7) within CALADIS. BioNumbers contains a huge range of biological measurements, spanning scales from microscopic chemical reaction rates and cellular concentrations to ecosystem- and planetwide statistics of

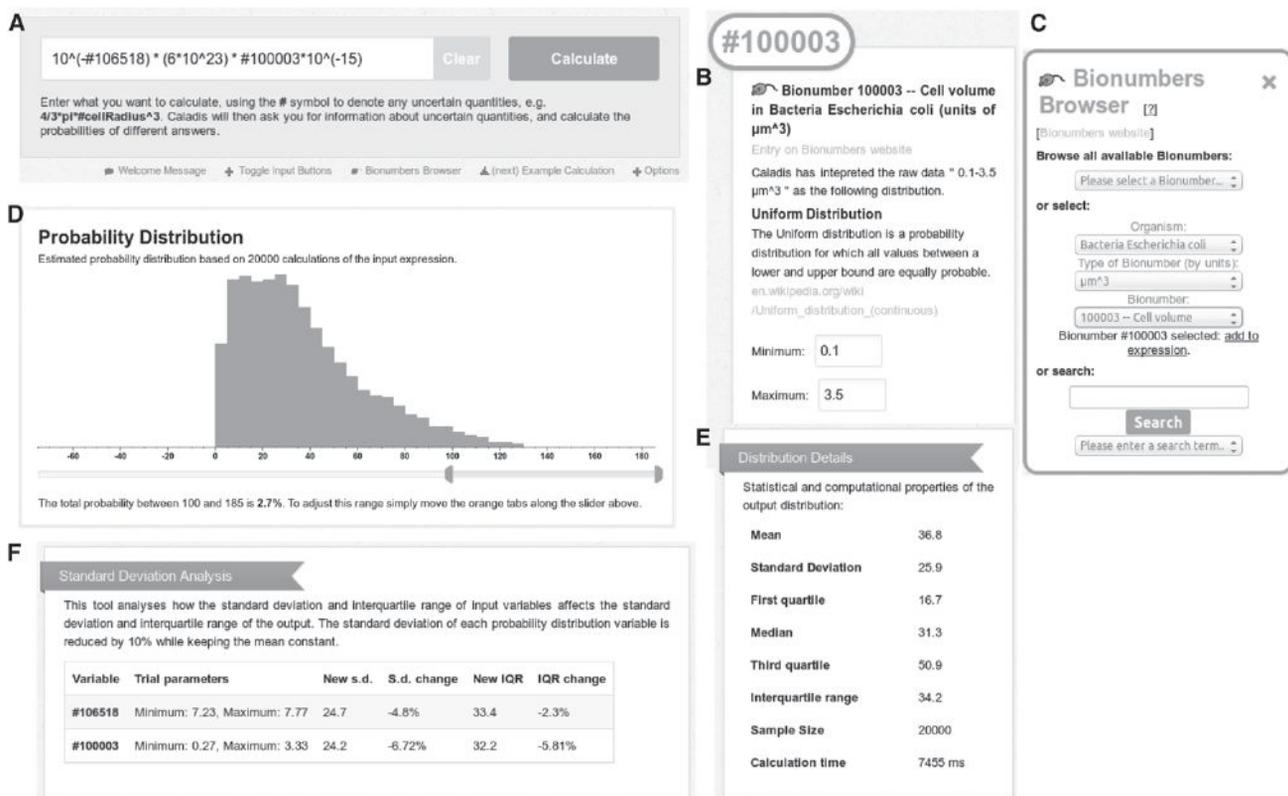


FIGURE 2 Elements of CALADIS interface. (A) The expression input box: a user enters a calculation here, providing any required information about distributions (for example, perhaps specifying that a certain distribution is uniform between 0 and 1, or normal with mean 1 and SD 0.1). (B) Each probability distribution in the input expression must then be characterized, either through the user’s entry of appropriate parameters, or (as depicted) through the automatic recognition of a BioNumber. (C) The BioNumbers Browser allows the identification, selection, and inclusion of values from the BioNumbers database. (D) The resultant distribution for the calculation is then displayed, along with summary statistics of the distribution (E) and (optionally) SD analysis (F) assessing the sensitivity of overall variance with respect to the variance of individual elements. This illustration involves, as an example calculation, the proton number calculation discussed in the Results. To see this figure in color, go online.

biological populations. Our link to the database allows us to perform powerful rule-of-thumb biophysical and cell-biological calculations with BioNumbers (5) while tracking uncertainties to estimate the ranges of the final answer.

Within our web tool, the BioNumbers database is parsed to obtain, for each BioNumber, a corresponding probability distribution, units, and a URL to the source data. Probability distributions are assigned based on the format of the source data and according to a user-defined protocol (see the Supporting Material). The units of each value are automatically obtained from the database. Users may then use a variety of approaches to identify and select BioNumbers for use in a probabilistic calculation, and the corresponding probability distributions are automatically included as calculation elements (see the Supporting Material).

RESULTS

Problems with reasoning with mean values in nonlinear contexts

We first illustrate how reasoning using only mean estimates may lead to incorrect results in calculations. Consider two measured quantities X and Y , perhaps corresponding to the abundance of two different types of entity in a population. We are interested in the proportion of X in the population $P = X/(X + Y)$.

Say we have the information that the measured quantities follow log-normal distributions, with X having mean $m_X = 0.1$ and SD (of the log-normal distribution itself, as opposed to the underlying normal distribution) $s_X = 0.1$, and Y having mean $m_Y = 0.9$ and SD $s_Y = 0.9$. In this artificial example, estimating the expected proportion of X in the population from the means alone would give

$$\hat{P} = m_X/(m_X + m_Y) = 0.1.$$

However, accurately tracking uncertainty in this calculation produces the counterintuitive result that $\mathbb{E}(P) \approx 0.144$, rather more than the population proportion estimated from mean values (Fig. 3 A).

This illustration contrasts with the cases where a calculation is straightforwardly additively or multiplicatively separable. In such cases, the fact that functions $f(X)$ and $g(X)$ of independent random variables X and Y are themselves independent leads to the results $\mathbb{E}(f(X)g(Y)) = \mathbb{E}(f(X))\mathbb{E}(g(Y))$ and $\mathbb{E}(f(X) + g(Y)) = \mathbb{E}(f(X)) + \mathbb{E}(g(Y))$, implying that calculations based on the individual means of X and Y will accurately estimate the overall mean. The error in the mean-based estimate \hat{P} in our example arises from the structure of the expression used to calculate the population proportion: the fraction cannot be separated into independent functions of X and Y . Generally in such inseparable cases, calculations based solely on mean values may not provide correct estimators. In such cases, explicitly tracking uncertainty not only provides a powerful characterization of the uncertainty in the final answer but also guarantees that such errors in the mean outcome are not made.

Next, we give two example calculations from the BioNumber of the Month website (3) to illustrate the process of explicitly tracking uncertainties in cell biological calculations with BioNumbers. The details of the BioNumber distributions used are shown in the Supporting Material.

The number of hydrogen ions in a cell

Given measurement of the pH and volume V of a system, the number of hydrogen ions in the system can be deduced as $n = 10^{-\text{pH}} N_A V$, where $N_A \approx 6 \times 10^{23}$ is Avogadro's number. In the December 2011 entry of Milo (3), measurements of pH and cell volume are used to estimate that an *Escherichia coli* cell contains ~60 hydrogen ions. Using CALADIS' BioNumbers browser to search for "cell volume" and "cytoplasm pH" identifies BioNumbers 100003 (*E. coli* cell volume) and 106518 (*E. coli* pH). These values appear in BioNumbers as (100003) 0.1–3.5 μm^3 , interpreted as $U(0.1, 3.5) \mu\text{m}^3$; and (106518) 7.2–7.8, interpreted as $U(7.2, 7.8)$. It is possible to interpret these results in terms of different probability distributions—a facility supported by CALADIS (see the Supporting Material). For example, the quantity 0.1–3.5 μm^3 could be interpreted as a log-normal distribution with 0.1 μm^3 and 3.5 μm^3 as $\pm 1\sigma$ points of the distribution. However, in this specific example, we use a uniform distribution, because the corresponding log-normal distribution exhibits extremely high variance with a range over more than an order of magnitude, which does not intuitively match the expected distribution of cell sizes in a population. Additionally, analytic results for the distribution of exponentially growing, dividing cells suggest a quadratic distribution that bears a stronger resemblance to the uniform than the log-normal picture (10). The ability to explore these different interpretations, and quantitatively debate the properties of each, are valuable scientific processes which our approach facilitates.

We can automatically access these BioNumbers and their associated uncertainties in CALADIS, then calculate the above equation while tracking uncertainties (this calculation forms the example used illustratively in Fig. 1 B). We find that the resultant distribution (see Fig. 3 B) easily spans an order of magnitude, with 14% of the density < 10 protons and 3% > 100 protons (statistics straightforwardly found using CALADIS' interface). Use of SD analysis suggests that more of this uncertainty originates from the spread of cell volumes. We now have a mean estimate at ~37 protons and a full characterization of the uncertainty associated with this answer, allowing a quantified degree of confidence to be associated with our reasoning.

Diffusion times in cells

In the March 2010 entry of Milo (3), the characteristic time-scales for diffusion through cells of various sizes are

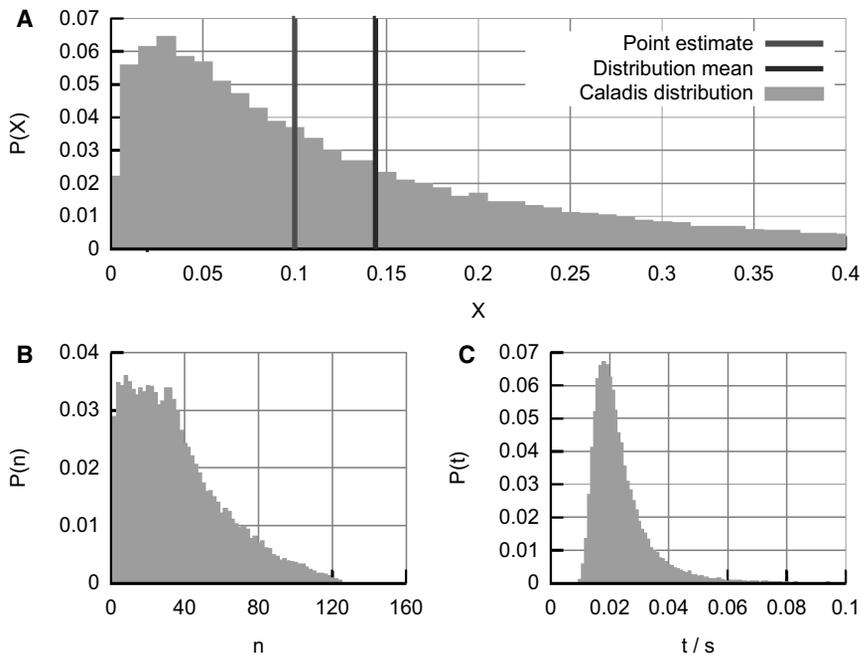


FIGURE 3 Point estimates and biological distributions. (A) The distribution resulting from the illustrative $X/(X + Y)$ calculation in the text. The value obtained by considering mean estimates alone differs from the mean of the true distribution, which is heavily skewed, highlighting the importance of explicitly tracking uncertainty. (B and C) Estimates, using data from biological experiments via the BioNumbers database, and tracking uncertainties for (B) number of protons in an *E. coli* cell and (C) time for GFP to diffuse $1 \mu\text{m}$ in *E. coli*. All distributions are direct outputs from CALADIS. To see this figure in color, go online.

explored, using the expression $t = x^2/6D$, where x is the length scale of diffusion and D is the diffusion constant of the species of interest. Milo (3) uses a rough estimate of the diffusion constant for GFP in *E. coli* and order-of-magnitude reasoning to obtain an estimate of 10 ms to traverse a root mean square distance of $1 \mu\text{m}$.

Using CALADIS' BioNumbers browser to search for "diffusion rate" identifies BioNumber 100193 (diffusion rate in *E. coli*), recorded as $7.7 \pm 2.5 \mu\text{m}^2 \text{s}^{-1}$ and interpreted as $N(7.7, 2.5) \text{m}^2 \text{s}^{-1}$. We follow the calculation in Milo (3) by including this BioNumber in the above equation, using $x = 1 \mu\text{m}$, and performing the probabilistic calculation of t in CALADIS, tracking uncertainties. We observe that the resultant distribution (see Fig. 3 C) is highly skewed, with an apparent coefficient of variation (the ratio of the SD to the mean, illustrating the spread of the distribution) of ~ 2.4 . This example, where a probability distribution appears in the denominator of an expression for a quantity of interest, illustrates how the resultant uncertainty can behave unintuitively when variables are combined even in relatively simple ways. Calculation of a resultant distribution provides a more robust method in these circumstances than traditional propagation-of-uncertainty approaches, and construction of a full probability distribution for the output of a calculation allows interpretation of details like skewness that are missed by a simple estimate of the SD alone.

DISCUSSION

We have described an approach for performing rule-of-thumb calculations in biophysics and cell biology while incorporating the considerable uncertainty often involved in such biological contexts. This approach, which does not

rely solely on point estimates of relevant quantities, allows the treatment and interpretation of the uncertainty involved in such calculations, increasing their trustworthiness and their power to assist intuitive reasoning. In addition, it may be used to optimize experimental design, by helping to identify measurements with the greatest power to refine knowledge of the overall quantity of interest.

To facilitate the straightforward use of this approach, both at a computer and on mobile devices, we have introduced CALADIS, an online tool for performing calculations involving probability distributions, available for free use and with its source code open and available to download. CALADIS has a particular link to rule-of-thumb calculations with BioNumbers in cell biology, and we have illustrated its use in deriving distributions of quantities of biophysical and cell biological interest. In employing these calculations in a scientific context, it is important to note that tracking the uncertainties in calculations is only useful if the underlying model is appropriately trusted: hygienic treatment of errors is a separate consideration from picking the right model for the world. It is unlikely that the model probability distributions employed in our approach (and many other analyses) represent the perfect description of a quantity arising in the real world; however, we hope that our approach, with the broad range of distributions supported by CALADIS, provides a means of reasonably estimating a wide range of real quantities. As we highlight above, the discussion of appropriate models for uncertainty, and their quick quantitative comparison, is a scientifically beneficial feature facilitated by our approach.

Back-of-the-envelope calculations (though used throughout history) have become increasingly popular recently as tools for developing quantitative reasoning and

intuition (1,4,5). Despite this increase in popularity, their use is not yet as prevalent in biology as in the physical sciences. We hope that this tool provides support for, and may increase trust in, the use of back-of-the-envelope calculations in quantitative cell biology (and across the biosciences) by exposing the role of uncertainties. We have shown that in some cases (for example, in calculating proportions), failing to track uncertainties can lead to rough guesses that do not represent the full truth of the calculation.

In our work with biological calculations, we have found that CALADIS plays a useful role in quality control for rule-of-thumb reasoning: after having made an approximate estimate on a real napkin, it is helpful to check whether the biological question at hand remains adequately answered if variability/imprecision is appropriately accommodated. We further note the reverse possibility: rather than serving as a sanity check for our envelope calculations, CALADIS can help create optimism about our estimates. For example, in settings where the uncertainty of some calculation elements is known to be very substantial, it might be the case that the final distribution of the estimated quantity is, in fact, sufficiently constrained for scientific advance. As discussed, an uncertainty appended to an estimated quantity needs to be treated with care (since it can depend on distribution choice) but it can still serve as a partial certificate for the relevance of the estimate. We suggest that researchers may present links to their calculations within CALADIS, so that readers are then free to use their prior beliefs to modify the component distributions (if, for example, a reader is less confident about a variable than the author) to see whether the conclusions are still robust.

As mentioned previously, the interpretation of calculations tracking uncertainty is contingent on the source of the uncertainty in the elements of the calculation, which may arise from imprecision (for example, measurement errors associated with an experimental protocol) or variability (the natural fluctuations intrinsic to a system of interest). Care must be taken in the interpretation of the resultant distribution depending on the sources of uncertainty in the calculation. For example, consider a quantity X which is subject to natural variability, stationary but fluctuating with time, and which has been characterized by a distribution involving a finite number N of measurements of X at different times. If we are interested in the behavior of X over an infinitesimally small time window, it makes sense to draw from this distribution of X , since this distribution represents plausible states of the system. If we are interested in the time-averaged behavior of X , we may instead consider the distribution of $\hat{E}(X)$, an estimate of the mean of X . $E(X)$ is a single number about which we are uncertain: the distribution of $\hat{E}(X)$ derived from our measurements will have a finite width (the standard error on the mean, dependent on N), corresponding to imprecision rather than natural variability. Mixing uncertainties due to imprecision with those due to variability may lead to results that are not trivial to interpret. We underline the importance

of transparency in the meaning of a probabilistic calculation to avoid misinterpretation—in the above example, it should be explicitly stated whether a calculation involves (variable) single instances of a measurement (X) or (imprecise) time-averaged behavior ($E(X)$).

The process of sampling from distributions describing individual quantities, performing a calculation using these samples, and building a final distribution is akin to several methodologies of use in Bayesian statistics (11). The difference between our approach and Bayesian sampling approaches is that after establishing our distributions we condition on no further data, instead assuming that the individual distributions (which could be pictured as priors) already contain all information on the likelihood of individual values. In this sense, the Bayesian interpretation of our approach is not as a method for extracting posteriors from priors given data, but is instead a method for performing calculations with priors without new data, thus constructing new prior distributions over more complicated quantities.

SUPPORTING MATERIAL

Additional supplemental information are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)01124-2](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)01124-2).

The authors thank Roslyn Lavery for help with web hosting and development, Ron Milo for helpful comments on the manuscript, and the valuable comments of three anonymous reviewers.

I.G.J. acknowledges funding from the UK Medical Research Council. N.S.J. acknowledges grant Nos. EP/K503733/1 and BBD0201901.

REFERENCES

- Weinstein, L., and J. Adam. 2009. *Guesstimation: Solving the World's Problems on the Back of a Cocktail Napkin*. Princeton University Press, Princeton, NJ.
- Phillips, R., J. Kondev, ..., H. Garcia. 2009. *Physical Biology of the Cell*. Garland Science, New York.
- Milo, R. 2009. BioNumber of the Month. http://www.weizmann.ac.il/plants/Milo/index.php?page_name=BioNumberOfTheMonth.
- Mahajan, S. 2010. *Street-Fighting Mathematics: The Art of Educated Guessing and Opportunistic Problem Solving*. MIT Press, Cambridge, UK.
- Phillips, R., and R. Milo. 2009. A feeling for the numbers in biology. *Proc. Natl. Acad. Sci. USA*. 106:21465–21471.
- Moran, U., R. Phillips, and R. Milo. 2010. SnapShot: key numbers in biology. *Cell*. 141:1262.
- Milo, R., P. Jorgensen, ..., M. Springer. 2010. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* 38:D750–D753.
- Regan, H., S. Ferson, and D. Berleant. 2004. Equivalence of methods for uncertainty propagation of real-valued random variables. *Int. J. Approx. Reason.* 36:1–30.
- Wasserman, L. 2004. *All of Statistics: A Concise Course in Statistical Inference*. Springer, New York.
- Rausenberger, J., and M. Kollmann. 2008. Quantifying origins of cell-to-cell variations in gene expression. *Biophys. J.* 95:4523–4528.
- Andrieu, C., N. De Freitas, ..., M. Jordan. 2003. An introduction to MCMC for machine learning. *Mach. Learn.* 50:5–43.

Article

Imaging GFP-Based Reporters in Neurons with Multiwavelength Optogenetic Control

Veena Venkatachalam¹ and Adam E. Cohen^{2,3,*}¹Biophysics Program and ²Departments of Chemistry and Chemical Biology and Physics, Harvard Stem Cell Institute, Harvard University, Cambridge, Massachusetts; and ³Howard Hughes Medical Institute, Chevy Chase, Maryland

ABSTRACT To study the impact of neural activity on cellular physiology, one would like to combine precise control of firing patterns with highly sensitive probes of cellular physiology. Light-gated ion channels, e.g., Channelrhodopsin-2, enable precise control of firing patterns; green fluorescent protein-based reporters, e.g., the GCaMP6f Ca²⁺ reporter, enable highly sensitive probing of cellular physiology. However, for most actuator-reporter combinations, spectral overlap prevents straightforward combination within a single cell. Here we explore multiwavelength control of channelrhodopsins to circumvent this limitation. The “stoplight” technique described in this article uses channelrhodopsin variants that are opened by blue light and closed by orange light. Cells are illuminated with constant blue light to excite fluorescence of a green fluorescent protein-based reporter. Modulated illumination with orange light negatively regulates activation of the channelrhodopsin. We performed detailed photo-physical characterization and kinetic modeling of four candidate stoplight channelrhodopsins. The variant with the highest contrast, sdChR(C138S,E154A), enabled all-optical measurements of activity-induced calcium transients in cultured rat hippocampal neurons, although cell-to-cell variation in expression levels presents a challenge for quantification.

INTRODUCTION

Combining optogenetic stimulation and fluorescence imaging

Activity-dependent changes in physiology are a hallmark of neural metabolism and information processing. Action potential generation and neurotransmitter release constitute significant metabolic loads, and maintenance of homeostasis in the presence of changing energy demands engages multiple metabolic pathways (1). Activity-dependent changes in neurons are integral in learning and memory, and contribute to the pathophysiology of diseases ranging from epilepsy to schizophrenia. Despite the fundamental importance of activity-dependent changes in neuronal physiology, we still do not fully understand many of the underlying pathways. A method to perturb neuronal activity with high spatiotemporal resolution while monitoring real-time cellular responses would be a valuable tool in this effort (2).

Optical tools for perturbing neural activity include photocaged glutamate (3,4), light-activated agonists of endogenous ion channels (5,6), azobenzene-derivatized glutamate receptors (7), and heterologously expressed microbial rhodopsins (8). Of these, the rhodopsins have been particularly effective because they can be genetically targeted to specific subclasses of cells; are readily activated with modest doses of visible light; and typically do not require an exogenous

cofactor (the retinal chromophore is present at sufficient levels in most vertebrate tissues). Channelrhodopsin-2 (ChR2), a light-gated cation channel from *Chlamydomonas reinhardtii*, allows optical control of neural activity in species ranging from worms to monkeys (9). New channelrhodopsins are frequently added to the optogenetic toolkit, distinguished by kinetic or spectroscopic features optimized for particular classes of experiments (10–14).

Protein-based fluorescent sensors have been developed for real-time measurements of membrane voltage, pH, calcium, ATP, NADH, cAMP, glutamate, reactive oxygen species, several redox potentials, activity of kinases and phosphatases, and many other modalities (15–18). Targeting of these reporters to subcellular domains in genetically specified subpopulations of neurons enables detailed studies of calcium fluxes, metabolic state, vesicle cycling, and signaling pathways. Although fluorescent proteins have been developed with excitation maxima throughout the visible spectrum (19), the vast majority of single-wavelength or FRET-based reporters contain a derivative of green fluorescent protein (GFP) (excitation, 488 nm; emission, 509 nm).

To achieve simultaneous optical perturbation and readout within the same cell, one must minimize the degree of optical crosstalk: the light used to trigger the actuator should not perturb the fluorescence of the reporter; and the light used to excite the reporter should not trigger the actuator (Fig. 1 *a*). Several approaches have been developed to achieve these goals. One- and two-photon glutamate uncaging is compatible with imaging of GFP-based reporters (20).

Submitted June 25, 2014, and accepted for publication August 7, 2014.

*Correspondence: cohen@chemistry.harvard.edu

Editor: Hagan Bayley.

© 2014 by the Biophysical Society
0006-3495/14/10/1554/10 \$2.00

<http://dx.doi.org/10.1016/j.bpj.2014.08.020>



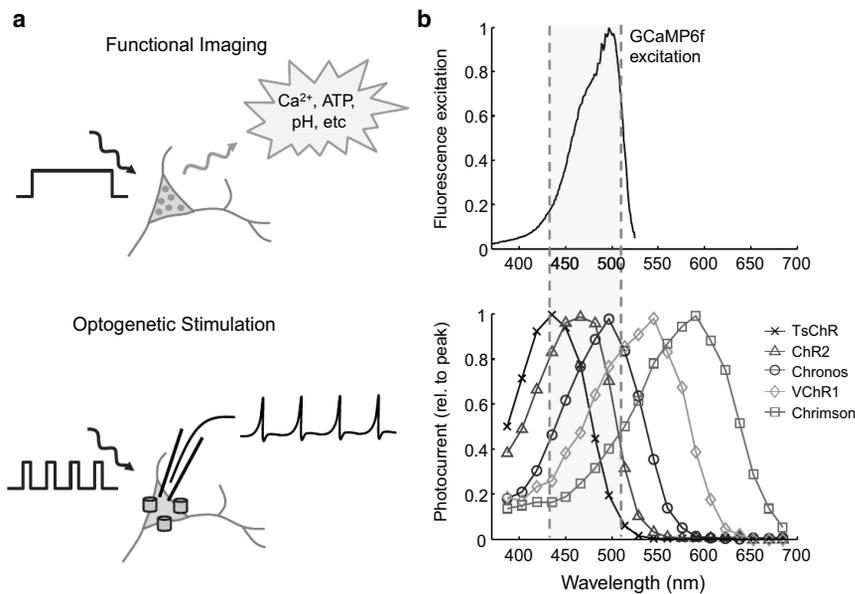


FIGURE 1 Spectral overlap limits combination of GFP-based reporters and channelrhodopsin-based actuators. (a) (Top) Continuous blue illumination is used to monitor GFP-based reporters of dynamic physiological quantities. (Bottom) Pulsed blue illumination is used for optogenetic stimulation. These two modalities cannot be combined in a single cell without optical crosstalk. (b) (Top) Fluorescence excitation spectrum of GCaMP6f, a sensor of Ca²⁺ (data from Chen et al. (31)). Other GFP-based reporters have a similar excitation spectrum. (Bottom) Action spectra of a panel of channelrhodopsins (data from Klapoetke et al. (10)). All channelrhodopsins are activated to some extent by the blue light used to excite a GFP-based reporter. To see this figure in color, go online.

Channelrhodopsin actuation can be paired with red-shifted voltage- or calcium-sensitive organic dyes (21,22) or proteins (13,23–25), although the number of reporters sufficiently red-shifted is limited.

Efforts to produce red-excited channelrhodopsins have shifted the excitation peaks as far as 590 nm, allowing independent optical control of distinct neural populations (10,26). However, these proteins retain 20–30% activation at the blue wavelengths used for excitation of GFP and thus experience spurious activation under fluorescence imaging conditions. Fig. 1 *b* shows the spectral overlap of GCaMP6f, a popular Ca²⁺ reporter, with a panel of channelrhodopsin actuators. Spectral overlap has remained a significant barrier to paired optical actuation and sensing in single neurons.

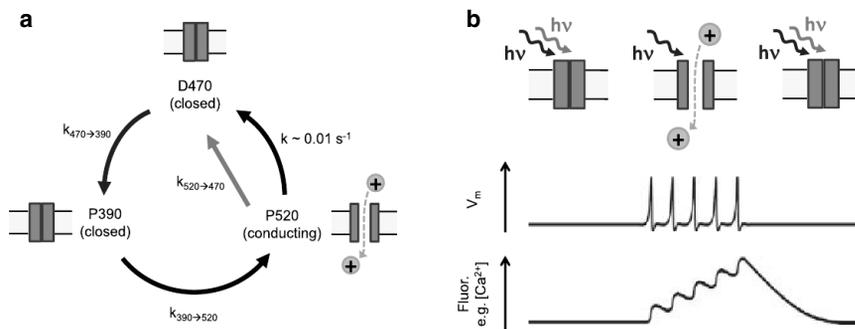
The complex photocycles of microbial rhodopsins open the possibility of sophisticated optical control (27). These proteins sample different states—each with its own function and absorption spectrum—with interstate transitions driven by illumination, membrane voltage, and thermal energy. Evolution has made use of these conformational and spectral gymnastics. For example, in *Halobacterium salinarum* a single photoreceptor, sensory rhodopsin I, mediates phototaxis both toward red light and away from ultraviolet (UV)/violet light. Red or orange light inhibits kinase activity in sensory rhodopsin I, while UV or near-UV light activates kinase activity, with opposite effects on cell motility (28). In the lab, one can use spectrally and temporally tuned pulses of light to interact with photocycle intermediates, thereby driving the population into states or distributions of states inaccessible under steady-state illumination. We previously applied this strategy to record stable photochemical imprints of membrane voltage (29) and to encode absolute values of membrane voltage into nonequilibrium dynamics of photocycle intermediates (30). Here we apply this strategy to

modulate channelrhodopsin photocurrents in the presence of continuous blue illumination, thereby enabling simultaneous optogenetic stimulation and fluorescence imaging of a GCaMP Ca²⁺ indicator (31).

Repurposing step-function opsins as stoplight channelrhodopsins

Step-function opsins (SFO) are channelrhodopsin variants that are opened by blue light and closed by orange or red light (32). The full SFO photocycle (33) is quite complex; Fig. 2 *a* shows a dramatically simplified version of this photocycle, which captures the features essential for the present discussion, but is not appropriate for a quantitative determination of rate constants. We reasoned that under continuous blue illumination, one could modulate the population in the open state by modulating the intensity of simultaneously applied orange light (Fig. 2 *b*). By collecting reporter fluorescence at wavelengths bracketed between the blue and orange illumination wavelengths, one could image the reporter with negligible crosstalk from the modulated orange beam. The challenge, then, was to identify an SFO and illumination conditions (intensities, wavelengths, times) that would facilitate robust crosstalk-free fluorescence imaging with simultaneous optogenetic stimulation of neurons. We call our proposed scheme the “stoplight” technique because the SFO photocurrent is stopped by red (or orange) light.

Due to the countervailing effects of blue and orange light on the open-state population, the orange intensity required to close the channel depends on the blue intensity. The blue intensity is set by the attributes of the fluorescent reporter and the demands for spatial and temporal resolution in the imaging. For instance, single-molecule or



3orange light opens the channels. In a neuron, this conductance could induce a train of action potentials and induce a fluorescence response in a GFP-based reporter, here represented by a Ca^{2+} indicator. Reapplication of orange light closes the channels and suppresses firing. To see this figure in color, go online.

high-magnification experiments require much higher intensity than population-average measurements; voltage imaging at a 1 kHz frame rate to detect neuronal action potentials requires higher illumination intensity than imaging at a 50 Hz frame rate to detect Ca^{2+} transients. To image a GFP-based reporter expressed in neurons under a strong constitutive promoter, e.g., CaMKIIa or hSynapsin, with $\sim 1\text{-}\mu\text{m}$ spatial resolution and $\sim 10\text{-ms}$ temporal resolution requires blue illumination (488 nm) at an intensity of $0.1\text{--}10\text{ W/cm}^2$. We thus measured photocurrents with blue illumination in this range, and with a simultaneously applied second beam of variable intensity and wavelength.

METHODS

Molecular biology

We made point mutants of ChR2, CoChR, and sdChR in a lentiviral vector (22217; Addgene, Cambridge, MA) under the CaMKIIa promoter. To make these point mutants, we modified the two-sided splicing by overlap extension technique of Horton et al. (34). Each gene was PCR-amplified in two fragments: fragment A, upstream of the mutation site; and fragment B, downstream of the mutation site. The mutation was introduced into both the reverse PCR primer used to amplify fragment A and the forward PCR primer used to amplify fragment B. We used Gibson Assembly (New England Biolabs, Ipswich, MA) to insert fragments A and B into the lentiviral vector (Addgene plasmid 22217 digested with the restriction enzymes *Bam*HI and *Eco*RI). We produced: ChR2(C128S)-eGFP, sdChR(C138S)-TS-eGFP-ER, sdChR(C138S,E154A)-TS-eGFP-ER, CoChR(C108S)-eGFP, sdChR(C138S,E154A)-TS-GCaMP6f-ER, and sdChR(C138S,E154A)-TS. The trafficking sequence (TS) used in Gradinaru et al. (35) was added to the C-terminus of sdChR to improve its membrane trafficking. The FCYENEV endoplasmic reticulum (ER) export motif of Gradinaru et al. (35) was also added to the C-terminus of GFP or GCaMP6f in these constructs to improve trafficking. The TS sequence did not improve membrane trafficking of ChR2(C128S). CoChR showed excellent membrane trafficking in HEK cells and neurons, so we did not add trafficking sequences to CoChR.

For expression in HEK cells, ChR2(C128S) was expressed under an ubiquitin promoter (using the plasmid backbone from Addgene 22051).

Fusions between stoplight rhodopsins and GCaMP6f were made by inserting a linker with the sequence AAPVVAVSKAAAKSRITSEGEYIPLDQIDINV, where the bold residues correspond to the TS. Extensive experience with this linker has shown independent function of the upstream rhodopsin and the downstream fluorescent reporter.

FIGURE 2 Stoplight optical control of a step-function opsin. (a) Simplified photocycle of a step-function opsin adapted from Berndt et al. (32), comprising a blue-absorbing ground state (D470), and an orange-absorbing open state (P520). The P390 intermediate limits the maximum rate at which the protein can go from the ground state to the open state. The spectra of the states are broad and overlapping, with maxima at the indicated wavelengths. (b) Proposed stoplight illumination protocol. (Cartoon) Anticipated data. Simultaneous application of weak ($\sim 1\text{ W/cm}^2$) blue and strong ($\sim 300\text{ W/cm}^2$) orange illumination leaves most channels closed. Removal of the

To generate a vector for expression of mitochondrial GCaMP5G in neurons, we added one copy of the Cox8 localization sequence (MSVLT PLLLRGLTGSARRLPVPRAKIHSLGDPVV, from Addgene 23348 (36)) to the N-terminus of GCaMP5G in a lentiviral expression vector under the CaMKII promoter.

All constructs were verified by sequencing.

Cell culture

HEK-293T cell culture and DNA transfection were performed as in MacLaurin et al. (37). Briefly, HEK293T cells were grown in DMEM supplemented with 10% FBS and penicillin/streptomycin in a 37°C incubator under 5% CO_2 . Forty-eight hours before measurement, cells were transfected using Transit-293 (Mirus) with the opsin of interest (see Molecular Biology above). These cells were replated at a density of $\sim 5000\text{--}10,000$ cells/ cm^2 on Matrigel-coated (Corning Life Sciences, Tewksbury, MA) cover-glass bottom dishes (P35G-1.5-14-C; MatTek, Ashland, MA) 24 h before measurement. All-*trans* retinal ($5\text{ }\mu\text{M}$) was added to each dish 1–2 h before imaging.

All experimental protocols involving use of animals were approved by the Harvard Institutional Animal Care and Use Committee. Primary neuronal cell culture was performed as in Venkatachalam et al. (29). Postnatal day 0 (P0) pups were euthanized and hippocampi were dissected following the procedure in Goslin (38). Briefly, isolated hippocampi were digested with papain and homogenized in Hank's Balanced Salt Solution containing MgCl_2 and kynurenic acid to prevent excitotoxicity. Cells were plated on glass-bottomed dishes (P35G-1.5-14-C; MatTek) coated with $20\text{ }\mu\text{g/mL}$ poly-D-lysine, and cultured in plating medium consisting of MEM (Life Technologies, Norwalk, CT) containing 10% fetal bovine serum, 0.5% glucose, 10 mM HEPES, 2 mM Glutamax (Life Technologies), 100 mg/L transferrin, insulin, and B27. After 60 h, the medium was replaced with NbActiv4 (Nb4-500; BrainBits, Springfield, IL). At 4 days in vitro, $2\text{ }\mu\text{M}$ AraC was added to suppress further glial growth. At 7 days in vitro, neurons were transfected using calcium phosphate (Cat. No. 631312; Clontech, Mountain View, CA). Each 3.5 cm dish was transfected with 200 ng of opsin DNA diluted in 1800 ng of nonexpressing pUC19 plasmid.

Patch-clamp electrophysiology and fluorescence imaging

The combined epifluorescence microscope and patch-clamp apparatus is described in Hou et al. (30). We imaged GFP-based fluorophores using excitation at 488 nm. Fluorescence was separated from scattered laser light by a 525/30 emission filter. Images were acquired at a 20-Hz frame rate on a scientific complementary metal-oxide semiconductor (CMOS) camera (Orca 4 Flash; Hamamatsu Photonics, Hamamatsu City, Japan).

All whole cell patch-clamp experiments in HEK cells were performed in voltage-clamp mode, while experiments in neurons were performed in current-clamp mode using a patch-clamp amplifier (model 2400; A-M Systems, Sequim, WA) and a micromanipulator (MP-285; Sutter Instrument, Novato, CA). Micropipettes were pulled from filamented borosilicate glass capillary tubes (1.5 mm OD, 0.84 mm ID; World Precision Instruments, Sarasota, FL) using a glass micropipette puller (P-1000; Sutter Instruments) to a tip resistance of 5–10 M Ω and filled with intracellular buffer (125 mM potassium gluconate, 8 mM NaCl, 0.6 mM MgCl₂, 0.1 mM CaCl₂, 1 mM EGTA, 10 mM HEPES, 4 mM Mg-ATP, and 0.4 mM Na-GTP at pH 7.3; adjusted to 295 mOsm with sucrose). The extracellular solution for all recordings was retinal-free Tyrode's buffer (125 mM NaCl, 2 mM KCl, 3 mM CaCl₂, 1 mM MgCl₂, 10 mM HEPES, and 30 mM glucose at pH 7.3; adjusted to 305–310 mOsm with sucrose). For experiments in HEK cells we added 50 μ M 2-APB to block electrical gap junctions. For experiments in neurons, we added 10 μ M NBQX, 20 μ M Gabazine, 25 μ M AP-5 to block synaptic transmission. All experiments were performed at 24°C.

Electrophysiology recordings were low-pass filtered at 5 kHz using the internal filter and then digitized using a data acquisition card (PCIe-6343; National Instruments, Austin, TX), at 10 kHz. Custom software routines written in MATLAB (The MathWorks, Natick, MA) and LABVIEW (National Instruments) were used to deliver illumination light of different wavelengths and intensities with submillisecond temporal precision.

RESULTS

Measurements were performed on a homemade system for simultaneous patch-clamp electrophysiology and multiwavelength fluorescence illumination and imaging, described in detail in Venkatachalam et al. (29) and Hou et al. (30). In brief, lasers at wavelengths of 488, 532, 561, 594, and 640 nm were combined by dichroic mirrors and modulated via an acousto-optic tunable filter to select the time-dependent intensity for each wavelength at the sample. Fluorescence imaging was performed in an inverted epifluorescence microscope equipped with a high numerical-aperture objective and a scientific CMOS camera. Membrane electrical properties were measured via whole-cell patch-clamp in either constant-current or constant-voltage mode. Illumination, imaging, and electrophysiology were synchronized via custom LABVIEW software (National Instruments). All experiments were conducted at 24°C.

We first tested the previously described SFO, ChR2(C128S) (32). We expressed ChR2(C128S)-eGFP in HEK293T cells (see Methods) and applied the illumination sequence shown in Fig. 3 *a*. We simultaneously monitored membrane current via whole-cell patch-clamp, maintaining $V_m = -70$ mV. A pulse of red light (640 nm, 700 W/cm², 300 ms) initialized the protein in the fully closed D470 state. A pulse of blue light of variable intensity (0.03–20 W/cm², 1000 ms) opened the channels and induced an inward (negative) photocurrent. During the middle 500 ms of the blue pulse, a stoplight pulse of orange light (594 nm, 300 W/cm²) was added to the illumination and the photocurrent decreased in magnitude. The photocurrent traces in Fig. 3 *a* illustrate the countervailing influences of blue and orange light. For the weakest blue illumination, the orange light largely suppressed the photocurrent (98% fractional inhibi-

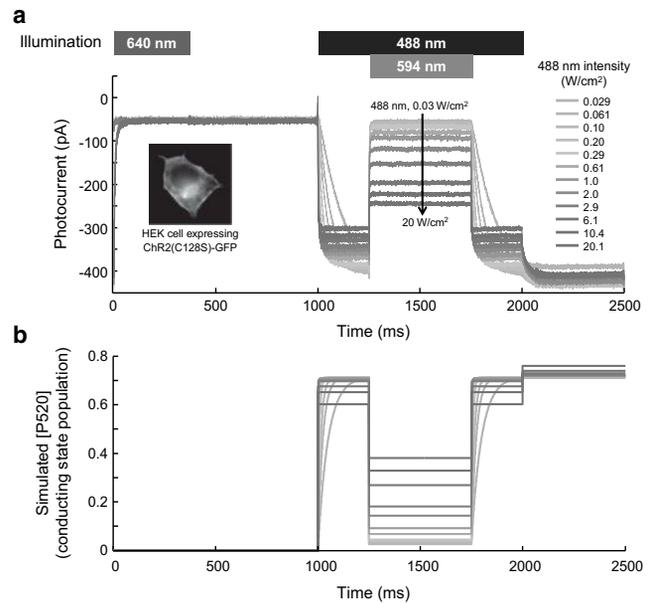


FIGURE 3 ChR2(C128S) can be closed under steady-state blue illumination by the addition of a stoplight. (*a*) A HEK-293T cell expressing ChR2(C128S) was illuminated with the indicated pulse sequence under whole-cell voltage clamp at -70 mV. The initial pulse of red illumination was responsible for closing channels left open by the previous illumination cycle. Open channelrhodopsin passed a negative current. Intense orange light (300 W/cm², 594 nm) suppressed $\sim 95\%$ of the photocurrent induced by moderate blue light (~ 300 mW/cm², 488 nm). (*b*) Results of a kinetic simulation of the photocycle model in Fig. 2 *a* using estimated rates. Each trace shows the expected population of the P520-conducting state at a different 488-nm illumination intensity (as in panel *a*). To see this figure in color, go online.

tion); however, this blue intensity was too low for fluorescence imaging, and only opened the channel slowly. For the strongest blue illumination, the orange light only partially closed the channel (23% fractional inhibition). In the absence of blue illumination, the orange light alone opened the channels to only 1.5% of maximum photocurrent, indicating minimal spurious channel activation by the stoplight.

To gain insight into the photocurrent dynamics of Fig. 3 *a*, we simulated the kinetic scheme of Fig. 2 *a* with time-varying illumination. By using this simplified model, we hoped to build intuition for the main qualitative features of the data. Due to the broad absorption spectrum of each state, we allowed each color of light to drive each optical transition with an efficiency proportional to the absorption coefficient of the source state. We neglected thermal isomerization from P520 to D470 because this rate is extremely slow, <0.01 s⁻¹ (32). Under these assumptions, the rates are

$$k_{470 \rightarrow 390} = k_1(I_{\text{blue}} + mI_{\text{SL}}),$$

$$k_{390 \rightarrow 520} = k_2,$$

$$k_{520 \rightarrow 470} = k_3(I_{\text{blue}} + nI_{\text{SL}}),$$

where I_{blue} is the blue intensity, I_{SL} is the stoplight intensity, m is the fractional absorption of D470 at λ_{SL} relative to λ_{blue} , and n is the fractional absorption of P520 at λ_{SL} relative to λ_{blue} . The governing equations are the following:

$$\frac{d}{dt} \begin{pmatrix} D470 \\ P390 \\ P520 \end{pmatrix} = \begin{pmatrix} -k_1(I_{\text{blue}} + mI_{\text{SL}}) & 0 & k_3(I_{\text{blue}} + nI_{\text{SL}}) \\ k_1(I_{\text{blue}} + mI_{\text{SL}}) & 0 & 0 \\ 0 & k_2 & -k_3(I_{\text{blue}} + nI_{\text{SL}}) \end{pmatrix} \begin{pmatrix} D470 \\ P390 \\ P520 \end{pmatrix}.$$

The calculated time-dependent population of the conducting state P520 recapitulated the main features of the photocurrent data (Fig. 3 *b*), with $k_1 = 0.5 \text{ cm}^2 \text{ W}^{-1} \text{ ms}^{-1}$, $k_2 = 15 \text{ ms}^{-1}$, $k_3 = 0.2 \text{ cm}^2 \text{ W}^{-1} \text{ ms}^{-1}$, $m = 0.001$, and $n = 0.05$. These kinetic parameters only approximately describe the dynamics of a more complex underlying photocycle and thus should not be ascribed to well-defined molecular transitions.

The simulations explained two nonintuitive features of the data:

First, the steady-state photocurrent under blue-only illumination (e.g., from $t = 1000$ – 1250 ms) was a decreasing function of blue intensity. Solving the kinetic equations for steady state with $I_{\text{SL}} = 0$ yields a population in the conducting P520 state,

$$[P520]_{\text{ss}} = \frac{A}{B + I_{\text{blue}}},$$

where $A = k_2/k_3$, and $B = k_2/k_1 + k_2/k_3$. Thus $[P520]_{\text{ss}}$ is a decreasing function of I_{blue} .

Second, when the blue light was stopped ($t = 2000$ ms) the photocurrent increased in magnitude. Under steady-state blue illumination there was a finite population in the P390 intermediate. Upon cessation of blue illumination, the rate $P520 \rightarrow D470$ became negligible ($< 0.01 \text{ s}^{-1}$), while population thermally transferred from $P390 \rightarrow P520$. Thus $[P520]$, and the photocurrent, increased. This simple model thus provided a framework for interpreting photocurrent data with ChR2(C128S) and other step-function opsins.

Optimization of stoplight wavelength and intensity for ChR2(C128S)

We next varied the wavelength and intensity of the stoplight to identify the optimal parameters for fast and high-contrast photoswitching. The illumination protocol was the same as in Fig. 3. We measured steady-state photocurrents (i_{ss}) at $V_m = -70$ mV as a function of blue illumination intensity (I_{blue}) either in the absence of a stoplight or with stoplight wavelength selected from $\lambda_{\text{SL}} = 532$, 594, or 640 nm. In all cases, the stoplight intensity was $I_{\text{SL}} = 200 \text{ W/cm}^2$ (Fig. 4 *a*).

A good stoplight would minimize the photocurrent under all blue illumination intensities. The red (640 nm) stoplight effectively shut the channel at very low I_{blue} , but was overpowered by the blue beam for $I_{\text{blue}} >$

0.1 W/cm^2 . The red beam was too far off resonance with the P520 transition to drive rapidly $P520 \rightarrow D470$. The green (532 nm) stoplight led to significant current at all values of I_{blue} . The green stoplight drove $D470 \rightarrow P390$ in addition to $P520 \rightarrow D470$, and thus was an ineffective stoplight. The orange (594 nm) stoplight achieved a balance between maximizing the rate of $P520 \rightarrow D470$, while minimizing crosstalk to $D470 \rightarrow P390$. For I_{blue} between 0.1 and 1 W/cm^2 , the orange stoplight suppressed photocurrent by 97–84%.

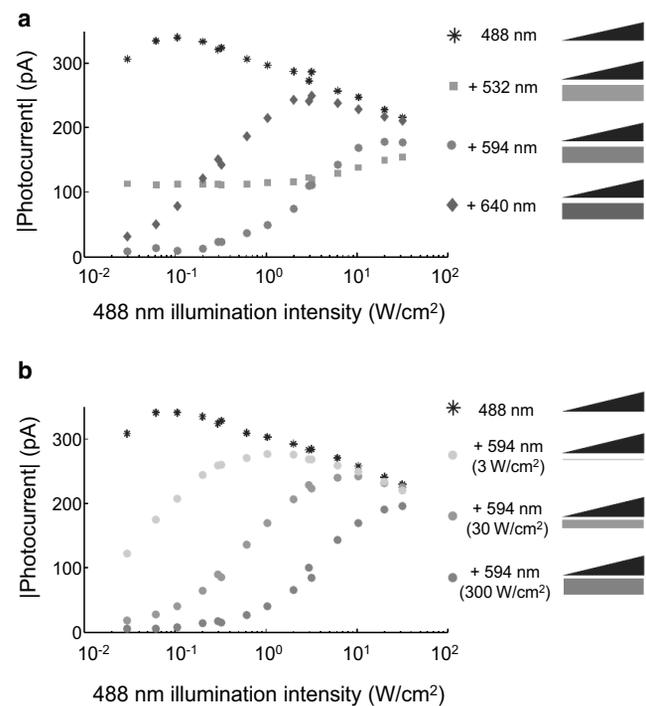


FIGURE 4 Optimization of illumination parameters for ChR2(C128S). (a) Steady-state photocurrents in the presence of blue illumination only (blue) and in the simultaneous presence of blue paired with green, orange, or red illumination (colored correspondingly). (b) Steady-state photocurrents in the presence of blue illumination only (blue) and with varying powers of orange stoplight. The aim is to maximize the blue-only photocurrent while minimizing the blue-plus-stoplight photocurrent. Data represents a single representative recording from $n = 5$ recordings. To see this figure in color, go online.

We next investigated the effect of stoplight intensity (Fig. 4 *b*). Ideally one would minimize the stoplight intensity to avoid risk of photodamage and to enable application over as wide a field of view as possible for a given illumination power. As expected, weaker stoplights were less effective at counteracting the blue photocurrent. At $I_{\text{blue}} = 300 \text{ mW/cm}^2$, the 300 W/cm^2 stoplight inhibited 95% of the blue photocurrent, compared with 75% inhibition at 30 W/cm^2 , and 25% inhibition at 3 W/cm^2 . Thus for ChR2(C128S), the stoplight must be ~ 1000 -fold more intense than the blue light.

Characterization of stoplight protocol in novel SFOs

Our spectroscopic explorations of ChR2(C128S) elucidated the critical parameters for stoplight performance. The closed and open states should have minimal spectral overlap, and the protein should have high conductance in the blue-only state. This conductance is the product of the expression level, the efficiency of trafficking to the plasma membrane, and the unit conductance of the open channel. Visual inspection of eGFP fluorescence in neurons expressing ChR2(C128S)-eGFP showed poor membrane trafficking. In our experiments and in previous reports (32), this protein did not pass sufficient photocurrent to induce robust spiking in cultured neurons.

We thus introduced mutations homologous to C128S into other channelrhodopsin variants that showed superior trafficking and sensitivity. The mutant ChR2(H134R) passes larger photocurrents than wild-type ChR2 (39), so we made ChR2(C128S, H134R). However, this construct did not “traffic” well in HEK cells, so it was not pursued further. A recent screen for improved channelrhodopsins identified two with greatly enhanced photocurrents: one from *Chloromonas oogama* (CoChR) and one from *Scherffelia dubia* (sdChR) (10). We thus made CoChR(C108S) and sdChR(C138S). We previously found that the mutant sdChR(E154A), termed “CheRiff”, had minimal activation by red or orange light (13), so we also made sdChR(C138S, E154A).

We expressed each mutant in HEK cells and characterized its photocurrents and kinetics under blue illumination (488 nm, 300 mW/cm^2) and simultaneously modulated orange illumination (594 nm, 300 W/cm^2) (Fig. 5 *a*). The ideal stoplight channelrhodopsin would show large photocurrent with blue-only illumination, and large fractional inhibition by orange light (i.e., reside in the *top-right* region of Fig. 5 *a*). CoChR(C108S) had the highest blue-only photocurrents ($748 \pm 129 \text{ pA}$, $n = 7$ cells), while sdChR(C138S,E154A) yielded the greatest fractional inhibition by orange light ($94.3 \pm 1.8\%$, $n = 5$ cells). All statistics are mean \pm SE.

We characterized the channel opening time (τ_{open}), corresponding to the orange light turning off, and the closing time

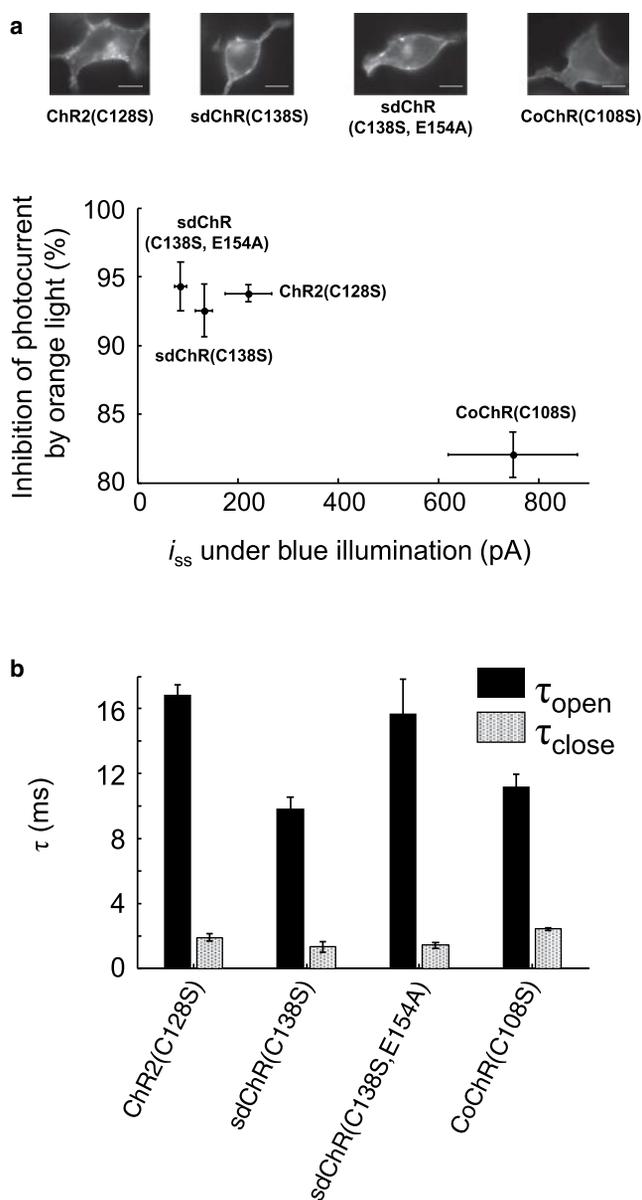


FIGURE 5 Gating properties of step-function opsin variants. (*a*) Photocurrents (at $V = -70 \text{ mV}$) in HEK cells were recorded with steady-state blue illumination (488 nm , 300 mW/cm^2) and with simultaneously applied orange illumination (594 nm , 300 W/cm^2). Ideal stoplight behavior would yield a large photocurrent under blue-only illumination, and a large fractional inhibition by orange light. (*Top*) Images of HEK cells expressing the constructs, imaged via eGFP fluorescence in an opsin-eGFP fusion. (*Scale bars*) $10 \mu\text{m}$. (*b*) Time constants of channel opening (τ_{open}) and closing (τ_{close}) under constant blue illumination and modulated orange illumination. (*Error bars*) Mean \pm SE of $n = 5$ –7 cells.

(τ_{close}), corresponding to the orange light turning on (Fig. 5 *b*). Photocurrent recordings were fit to a single exponential. All fits had $r^2 > 0.98$. The mutant sdChR(C138S) had the fastest opening time constant of $\tau_{\text{open}} = 9.8 \pm 0.7 \text{ ms}$ ($n = 7$ cells). All candidates had $\tau_{\text{close}} < 3 \text{ ms}$. The rapid closing of the step-function opsins could be useful for inducing high-frequency trains of action potentials, and is

much faster than the spontaneous closing of conventional channelrhodopsins.

Testing the stoplight technique in neurons

While no ChR mutant was optimal by all measures, for further characterization in neurons we selected sdChR(C138S, E154A) on account of its high (94%) suppression by orange light and its good membrane trafficking. We compared wild-type ChR2 and sdChR(C138S, E154A) in paired cultures of rat hippocampal neurons, using calcium phosphate transfection of the constructs in lentiviral vectors under the CaMKIIa promoter. Synaptic blockers were added to the imaging medium to ensure cell-autonomous dynamics. We used manual patch-clamp in whole-cell current-clamp mode to monitor the membrane voltage as we varied the blue and orange illumination (see Methods).

In a neuron expressing ChR2-eGFP, blue pulses (488 nm, 3 W/cm², 10 ms) on a dark background robustly induced single action potentials (Fig. 6 *a*). However, introduction of a steady-state blue background of 300 mW/cm², such as might be used for GFP imaging, depolarized the cell by 40 mV, leading to sodium channel inactivation, and suppression of firing. These results illustrate the need for a more sophisticated approach to simultaneous optical stimulation and imaging.

We expressed sdChR(C138S, E154A) fused to a TS, an eGFP, and an ER export motif (see Methods) in a neuron and monitored the membrane voltage in response to a stoplight illumination protocol (Fig. 6 *b*). At $I_{\text{blue}} = 0$, modulation of the orange stoplight ($I_{\text{stoplight}} = 300 \text{ W/cm}^2$) had no effect on membrane voltage. Simultaneous application of

blue light ($I_{\text{blue}} = 300 \text{ mW/cm}^2$) and orange light ($I_{\text{stoplight}} = 300 \text{ W/cm}^2$) depolarized the resting voltage by only 5 mV. Modulation of the orange beam ($t_{\text{on}} = 800 \text{ ms}$, $t_{\text{off}} = 200 \text{ ms}$) in the presence of blue light robustly triggered action potentials. These results demonstrate that the stoplight technique can trigger neuronal firing under continuous blue illumination appropriate for imaging a GFP-based reporter. Phototoxicity is a concern because of the high intensity of the orange stoplight. We found that neurons could typically tolerate ~2 min of stoplight illumination before they developed changes in action potential parameters such as were indicative of phototoxicity.

We then replaced the eGFP expression marker with GCaMP6f (31), a genetically encoded fluorescent reporter of Ca²⁺ (see Methods). We expressed sdChR(C138S, E154A)-TS-GCaMP6f-ER in cultured rat hippocampal neurons and applied the stoplight illumination protocol ($I_{\text{blue}} = 300 \text{ mW/cm}^2$, $I_{\text{stoplight}} = 300 \text{ W/cm}^2$, Fig. 7). We simultaneously recorded membrane potential via manual patch-clamp and GCaMP6f fluorescence (excited by the blue beam).

The 594-nm light was blocked every 6 s, with t_{off} increasing from 25 to 425 ms. Blockage of the stoplight led to action potentials, with longer off-times leading to larger numbers of action potentials as recorded electrically. The GCaMP6f fluorescence showed positive-going transients associated with blockage of the stoplight. Although t_{off} increased for each off-period, the amplitude of the Ca²⁺ transient grew in discrete steps, corresponding to the number of action potentials induced. This result demonstrates that the stoplight protocol can optically induce action potentials and simultaneously monitor action-potential-induced Ca²⁺ transients with single-spike resolution. The

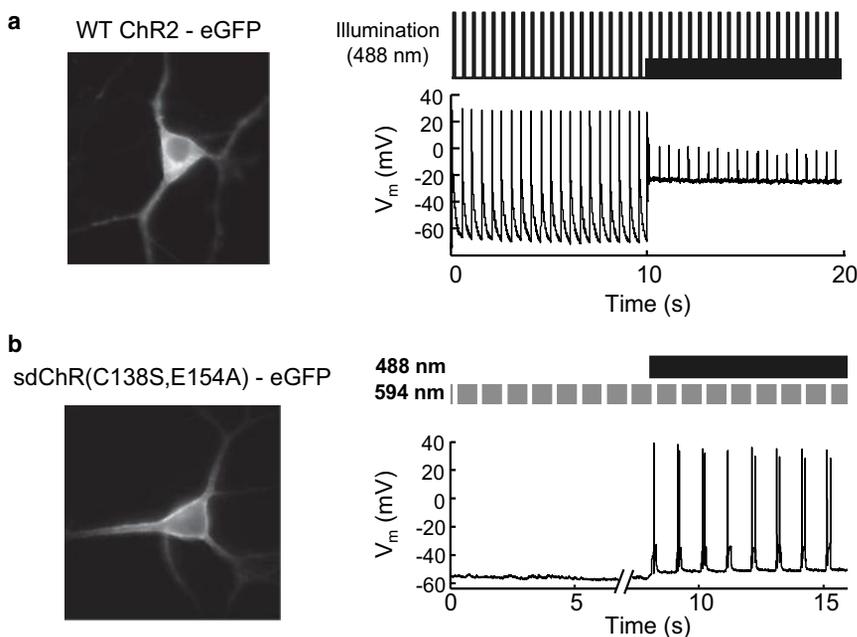


FIGURE 6 Comparison of conventional ChR2 versus stoplight protocol for simultaneous optogenetic stimulation and eGFP imaging. (*a*) Attempt to use wild-type ChR2 to optically induce action potentials in a cultured rat hippocampal neuron while simultaneously imaging eGFP. From $t = 0 \text{ s}$ to $t = 10 \text{ s}$, baseline blue intensity was zero, and blue pulses (488 nm, 10 ms, 3 W/cm²) reliably induced action potentials. From $t = 10 \text{ s}$ to $t = 20 \text{ s}$, baseline blue intensity was 300 mW/cm², appropriate for eGFP imaging. This illumination depolarized the neuron by ~40 mV. Blue pulses (10 ms, 3 W/cm²) failed to induce action potentials on this depolarizing background. (*b*) Optogenetic control of action potentials under continuous blue illumination using the stoplight technique with sdChR(C138S, E154A). Onset of orange illumination alone (594 nm, 300 W/cm²) depolarized the cell by <2 mV. Subsequent pulses of orange light did not detectably perturb membrane potential, with a measurement noise of 1 mV. Simultaneous blue (300 mW/cm²) and orange (300 W/cm²) illumination depolarized the cell by only 5 mV. Interruption of the orange stoplight ($t_{\text{off}} = 200 \text{ ms}$) reliably triggered action potentials. To see this figure in color, go online.

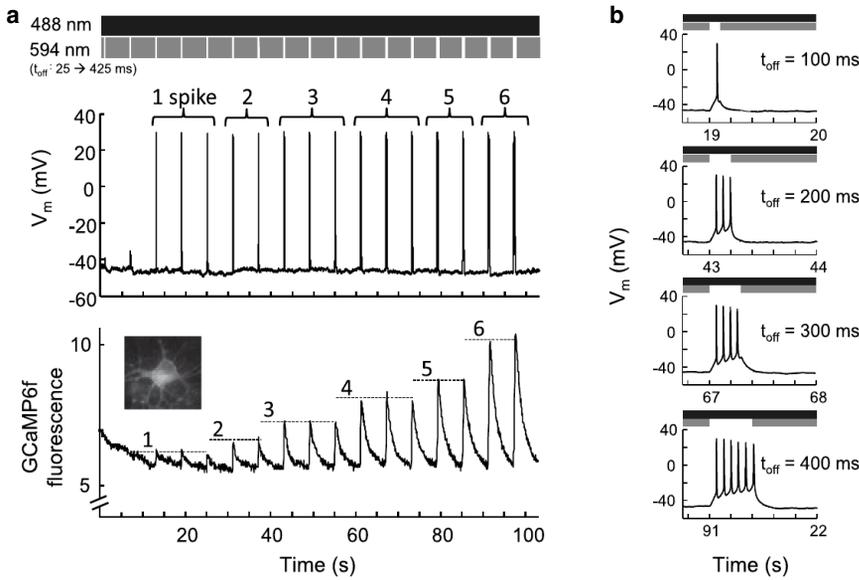


FIGURE 7 Crosstalk-free optical triggering of action potentials and optical monitoring of Ca^{2+} transients. (a) Calcium influx increased with the number of action potentials triggered by sdChR(C138S, E154A). (Top) Stoplight illumination scheme, with t_{off} increasing from 25 to 425 ms. (Middle) Membrane voltage recorded in whole-cell current clamp. The number of action potentials induced in each off-period is labeled above the spike. (Bottom) Simultaneously recorded GCaMP6f fluorescence, showing spikes of discrete amplitude that grew with number of action potentials. (b) Closeups of the membrane voltage during the off-periods of increasing length. To see this figure in color, go online.

patch pipette in this experiment served only as a passive probe of membrane potential, and was not necessary for the optical experiment.

A key merit of the all-optical stoplight protocol is its ability to measure neuronal responses with higher throughput than is practical with patch-clamp electrophysiology. We applied the stoplight protocol without simultaneous patch-clamp recordings to $n = 23$ neurons. Of these, nine (39%) showed clearly resolved unitary Ca^{2+} transients similar to those shown in Fig. 7. In the cells that showed a clear Ca^{2+} response, the minimum t_{off} to induce a single action potential varied between 10 and 400 ms (mean 123 ms, $n = 9$ cells). More frequently, we observed complex Ca^{2+} transients that did not clearly resemble single action potential responses. This cell-to-cell variability likely arose from two factors: 1), variations in resting potential, which caused variation in the amount of depolarization needed to reach the threshold voltage for action potential initiation; and 2), variations in opsin expression, which caused variations in opsin-mediated photocurrents. The small baseline depolarization under simultaneous blue and orange illumination was likely proportional to opsin expression level, and therefore also varied between cells. GCaMP6f shows a nonlinear response to Ca^{2+} concentration, with a K_d value of 375 nM (31). Thus, the amplitude of the fluorescence transients depended on the initial cytoplasmic Ca^{2+} concentration. These sources of variability, some biological and some technical, present a challenge for using the stoplight technique to deterministically induce defined numbers of action potentials.

Despite these limitations, we tested whether the stoplight technique could report associations between neural activity and response of GFP-based sensors. To compare Ca^{2+} handling in the cytoplasm and the mitochondria, we made mitochondrially targeted Ca^{2+} sensor by fusing a mitochon-

drial targeting domain (36) to GCaMP5G (see Methods). We cotransfected neurons with sdChR(C138S,E154A) and either cytosolic GCaMP6f or mitochondrial GCaMP5G. The orange stoplight was blocked for 3 s intervals with a 10 s period, to induce episodes of sustained depolarization. The cytosolic Ca^{2+} response showed a clear rise and fall and returned to baseline after each stimulus (Fig. 8 and Movie S1 in the Supporting Material). The mitochondrial Ca^{2+} response was much slower, and did not return to baseline between stimuli (Fig. 8) (see Movie S2). The mitochondrial

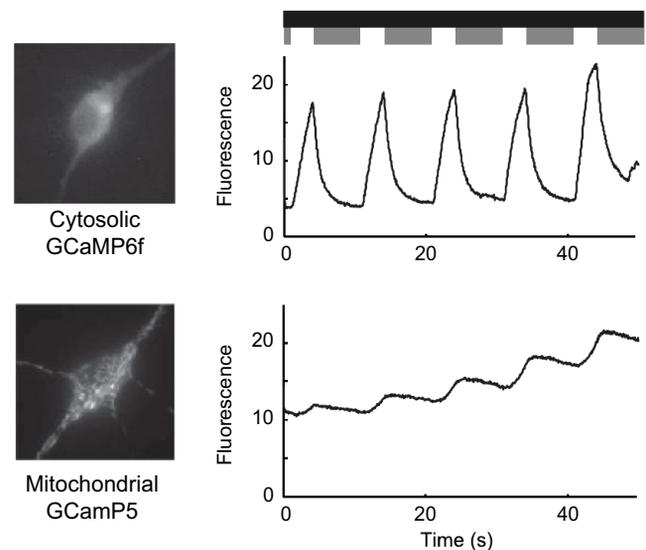


FIGURE 8 Stoplight protocol reveals activity-dependent physiological responses in neurons. Cultured rat hippocampal neurons were cotransfected with sdChR(C138S,E154A) and with a Ca^{2+} indicator targeted either to the cytosol (top) or the mitochondria (bottom). Fluorescence, corresponding to $[Ca^{2+}]_i$, was continuously monitored in response to optically induced activity. To see this figure in color, go online.

Ca²⁺ response also lagged stimulus onset or offset, reflecting the multistep process for Ca²⁺ to transfer between the mitochondrial matrix and the extracellular medium. The decay of Ca²⁺ in the mitochondria was much slower than the previously reported off-time for GCaMP5G (40), confirming that the difference in observed timing was not due to the difference between GCaMP5G and GCaMP6f.

DISCUSSION

Nonlinear control through interaction with photocycle intermediates provides a powerful means to steer opsin dynamics. This approach is readily implemented with moderate-intensity quasi-CW light sources and millisecond timescale modulation. The stoplight technique provides a means to control neuronal activation while simultaneously monitoring fluorescent reporters (either protein or small-molecule) spectrally similar to GFP. This technique relies on temporal modulation of an orange (594 nm) illumination source to control neuronal activity, resulting in minimal optical crosstalk between the light used for optogenetic actuation and the light used for imaging. The stoplight method also leads to faster bidirectional switching than is achieved with single-wavelength optogenetic control. Orange light closes step-function opsins in <2 ms, while thermally activated closing of channelrhodopsins occurs with a time constant of at least 15 ms (41,42). With a single gene, one can rapidly hyperpolarize or depolarize a cell relative to the membrane potential established in the photostationary state.

Several advances will be required for the stoplight technique to become generally applicable. To start, it is necessary to improve the dynamic range of the opsin photocurrent: to have a higher blue-only photocurrent, and a lower blue-plus-orange photocurrent. A high blue-only photocurrent is necessary to trigger action potentials in a robust manner. A low blue-plus-orange photocurrent is necessary to prevent baseline depolarization and sodium channel inactivation. Among the opsins we explored, there appeared to be a tradeoff between performance along these parameters (Fig. 5 a). We see no fundamental reason for this tradeoff. A broader search of opsin backbones or further mutation of the opsins presented here may identify variants that can reliably and precisely trigger single time-action potentials via the stoplight protocol, over a wide range of expression levels and natural neuronal variability. The CoChR variants show particular promise. We did not apply CoChR(C108S) in neurons due to the incomplete suppression of photocurrent (82%) by orange light. However, the exceptional membrane localization and high blue-only photocurrent (750 pA) of this mutant make it a promising candidate for further mutagenesis. Novel strategies for improving expression and membrane trafficking can dramatically increase photocurrents (42).

A related challenge is to decrease the minimum orange light intensity needed to suppress the photocurrent. The

stoplight intensity we used, 300 W/cm², borders on the photodamage threshold in neurons. This intensity can only be achieved with a high magnification objective and laser illumination focused to a region approximately the size of a single cell. Applications to studying neural circuit dynamics in vitro or in vivo will require photocurrent suppression at much lower stoplight intensity.

SUPPORTING MATERIAL

Two movies showing Ca²⁺ transients in neurons under Stoplight stimulation are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)00891-1](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00891-1).

We thank Christopher Werley for design and construction of the microscope; we thank Christopher Werley, Peng Zou, and Daan Brinks for software development. Vaibhav Joshi and Niklas Smedemark-Margulies performed tissue culture. Vedhavalli Nathan, Jonathan Gootenberg, and Katherine Williams helped with cloning. Nathan Klapoetke and Ed Boyden provided genes for sdChR and CoChR and channelrhodopsin action spectra. Douglas Kim and Eric Schreier provided the action spectrum of GCaMP6f.

This work was supported by the Howard Hughes Medical Institute; Office of Naval Research (ONR) Presidential Early Career Award for Scientists and Engineers award No. N00014-11-1-0549; the Harvard Center for Brain Science; National Institutes of Health (NIH) grant No. 1-R01-EB012498-01 and NIH New Innovator grant No. 1-DP2-OD007428; the Gordon and Betty Moore Foundation; a Herchel Smith Graduate Fellowship (to V.V.); and NIH MSTP grant No. T32GM07753-36 (to V.V.).

Conflict of interest: A.E.C. is a founder of Q-State Biosciences and a member of its scientific advisory board.

REFERENCES

- Rangaraju, V., N. Calloway, and T. A. Ryan. 2014. Activity-driven local ATP synthesis is required for synaptic function. *Cell*. 156:825–835.
- Cohen, A. E., and V. Venkatachalam. 2014. Bringing bioelectricity to light. *Annu. Rev. Biophys.* 43:11.
- Nikolenko, V., K. E. Poskanzer, and R. Yuste. 2007. Two-photon photostimulation and imaging of neural circuits. *Nat. Methods*. 4:943–950.
- Kwon, H. B., and B. L. Sabatini. 2011. Glutamate induces de novo growth of functional spines in developing cortex. *Nature*. 474:100–104.
- Kokel, D., C. Y. J. Cheung, ..., R. T. Peterson. 2013. Photochemical activation of TRPA1 channels in neurons and animals. *Nat. Chem. Biol.* 9:257–263.
- Kramer, R. H., A. Mouro, and H. Adesnik. 2013. Optogenetic pharmacology for control of native neuronal signaling proteins. *Nat. Neurosci.* 16:816–823.
- Levitz, J., C. Pantoja, ..., E. Y. Isacoff. 2013. Optical control of metabotropic glutamate receptors. *Nat. Neurosci.* 16:507–516.
- Packer, A. M., B. Roska, and M. Häusser. 2013. Targeting neurons and photons for optogenetics. *Nat. Neurosci.* 16:805–815.
- Yizhar, O., L. E. Fenno, ..., K. Deisseroth. 2011. Optogenetics in neural systems. *Neuron*. 71:9–34.
- Klapoetke, N. C., Y. Murata, ..., E. S. Boyden. 2014. Independent optical excitation of distinct neural populations. *Nat. Methods*. 11:338–346.
- Wietek, J., J. S. Wiegert, ..., P. Hegemann. 2014. Conversion of channelrhodopsin into a light-gated chloride channel. *Science*. 344:409–412.

12. Berndt, A., S. Y. Lee, ..., K. Deisseroth. 2014. Structure-guided transformation of channelrhodopsin into a light-activated chloride channel. *Science*. 344:420–424.
13. Hochbaum, D. R., Y. Zhao, ..., A. E. Cohen. 2014. All-optical electrophysiology in mammalian neurons using engineered microbial rhodopsins. *Nat. Methods*. 11:825–833. <http://dx.doi.org/10.1038/nmeth.3000>.
14. Govorunova, E. G., O. A. Sineshchekov, ..., J. L. Spudich. 2013. Characterization of a highly efficient blue-shifted channelrhodopsin from the marine alga *Platymonas subcordiformis*. *J. Biol. Chem.* 288:29911–29922.
15. Depry, C., S. Mehta, and J. Zhang. 2013. Multiplexed visualization of dynamic signaling networks using genetically encoded fluorescent protein-based biosensors. *Pflugers Arch.* 465:373–381.
16. Mehta, S., and J. Zhang. 2011. Reporting from the field: genetically encoded fluorescent reporters uncover signaling dynamics in living biological systems. *Annu. Rev. Biochem.* 80:375–401.
17. Tantama, M., J. R. Martínez-François, ..., G. Yellen. 2013. Imaging energy status in live cells with a fluorescent biosensor of the intracellular ATP-to-ADP ratio. *Nat. Commun.* 4:2550.
18. Hung, Y. P., J. G. Albeck, ..., G. Yellen. 2011. Imaging cytosolic NADH-NAD⁺ redox state with a genetically encoded fluorescent biosensor. *Cell Metab.* 14:545–554.
19. Shaner, N. C., P. A. Steinbach, and R. Y. Tsien. 2005. A guide to choosing fluorescent proteins. *Nat. Methods*. 2:905–909.
20. Vogt, K. E., S. Gerharz, ..., M. Canepari. 2011. Combining membrane potential imaging with L-glutamate or GABA photorelease. *PLoS ONE*. 6:e24911.
21. Tsuda, S., M. Z. Kee, ..., G. J. Augustine. 2013. Probing the function of neuronal populations: combining micromirror-based optogenetic photostimulation with voltage-sensitive dye imaging. *Neurosci. Res.* 75:76–81.
22. Lim, D. H., M. H. Mohajerani, ..., T. H. Murphy. 2012. In vivo large-scale cortical mapping using channelrhodopsin-2 stimulation in transgenic mice reveals asymmetric and reciprocal relationships between cortical areas. *Front Neural Circuits*. 6:11.
23. Wu, J., L. Liu, ..., R. E. Campbell. 2013. Improved orange and red Ca²⁺ indicators and photophysical considerations for optogenetic applications. *ACS Chem. Neurosci.* 4:963–972.
24. Zhao, Y., S. Araki, ..., R. E. Campbell. 2011. An expanded palette of genetically encoded Ca²⁺ indicators. *Science*. 333:1888–1891.
25. Canepari, M., D. Zecevic, ..., M. De Waard. 2013. Combining calcium imaging with other optical techniques. *Cold Spring Harb Protoc.* 2013:1125–1131. <http://dx.doi.org/10.1101/pdb.top066167>.
26. Ogren, J. I., S. Mamaev, ..., K. J. Rothschild. 2014. Retinal chromophore structure and Schiff base interactions in the red-shifted channelrhodopsin-1 from *Chlamydomonas augustae*. *Biochemistry (NY)*. 53:3961–3970.
27. Spudich, J. L., O. A. Sineshchekov, and E. G. Govorunova. 2014. Mechanism divergence in microbial rhodopsins. *Biochim. Biophys. Acta. Bioenergetics*. 1837:546–552.
28. Spudich, J. L., and R. A. Bogomolni. 1984. Mechanism of color discrimination by a bacterial sensory rhodopsin. *Nature*. 312:509–513.
29. Venkatachalam, V., D. Brinks, ..., A. E. Cohen. 2014. Flash memory: photochemical imprinting of neuronal action potentials onto a microbial rhodopsin. *J. Am. Chem. Soc.* 136:2529–2537.
30. Hou, J. H., V. Venkatachalam, and A. E. Cohen. 2014. Temporal dynamics of microbial rhodopsin fluorescence reports absolute membrane voltage. *Biophys. J.* 106:639–648.
31. Chen, T. W., T. J. Wardill, ..., D. S. Kim. 2013. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature*. 499:295–300.
32. Berndt, A., O. Yizhar, ..., K. Deisseroth. 2009. Bi-stable neural state switches. *Nat. Neurosci.* 12:229–234.
33. Ritter, E., P. Piwowarski, ..., F. J. Bartl. 2013. Light-dark adaptation of channelrhodopsin C128T mutant. *J. Biol. Chem.* 288:10451–10458.
34. Horton, R. M., H. D. Hunt, ..., L. R. Pease. 1989. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene*. 77:61–68.
35. Gradinaru, V., F. Zhang, ..., K. Deisseroth. 2010. Molecular and cellular approaches for diversifying and extending optogenetics. *Cell*. 141:154–165.
36. Karbowski, M., D. Arnoult, ..., R. J. Youle. 2004. Quantitation of mitochondrial dynamics by photolabeling of individual organelles shows that mitochondrial fusion is blocked during the Bax activation phase of apoptosis. *J. Cell Biol.* 164:493–499.
37. MacLaurin, D., V. Venkatachalam, ..., A. E. Cohen. 2013. Mechanism of voltage-sensitive fluorescence in a microbial rhodopsin. *Proc. Natl. Acad. Sci. USA*. 110:5939–5944.
38. Goslin, K. 1998. *Culturing Nerve Cells*. The MIT Press, Cambridge, MA.
39. Nagel, G., M. Brauner, ..., A. Gottschalk. 2005. Light activation of channelrhodopsin-2 in excitable cells of *Caenorhabditis elegans* triggers rapid behavioral responses. *Curr. Biol.* 15:2279–2284.
40. Akerboom, J., T. W. Chen, ..., L. L. Looger. 2012. Optimization of a GCaMP calcium indicator for neural activity imaging. *J. Neurosci.* 32:13819–13840.
41. Lin, J. Y., M. Z. Lin, ..., R. Y. Tsien. 2009. Characterization of engineered channelrhodopsin variants with improved properties and kinetics. *Biophys. J.* 96:1803–1814.
42. Mattis, J., K. M. Tye, ..., K. Deisseroth. 2012. Principles for applying optogenetic tools derived from direct comparative analysis of microbial opsins. *Nat. Methods*. 9:159–172.

Article

Evolution of Specificity in Protein-Protein Interactions

Orit Peleg,¹ Jeong-Mo Choi,² and Eugene I. Shakhnovich^{2,*}¹School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts; and ²Department of Chemistry and Chemical Biology, Harvard University, Cambridge Massachusetts

ABSTRACT Hub proteins are proteins that maintain promiscuous molecular recognition. Because they are reported to play essential roles in cellular control, there has been a special interest in the study of their structural and functional properties, yet the mechanisms by which they evolve to maintain functional interactions are poorly understood. By combining biophysical simulations of coarse-grained proteins and analysis of proteins-complex crystallographic structures, we seek to elucidate those mechanisms. We focus on two types of hub proteins: Multi hubs, which interact with their partners through different interfaces, and Singlish hubs, which do so through a single interface. We show that loss of structural stability is required for the evolution of protein-protein-interaction (PPI) networks, and it is more profound in Singlish hub systems. In addition, different ratios of hydrophobic to electrostatic interfacial amino acids are shown to support distinct network topologies (i.e., Singlish and Multi systems), and therefore underlie a fundamental design principle of PPI in a crowded environment. We argue that the physical nature of hydrophobic and electrostatic interactions, in particular, their favoring of either same-type interactions (hydrophobic-hydrophobic), or opposite-type interactions (negatively-positively charged) plays a key role in maintaining the network topology while allowing the protein amino acid sequence to evolve.

INTRODUCTION

Proteins that maintain promiscuous molecular recognition, i.e., the ability to maintain functional interactions with multiple partners (1), represent hubs in protein-protein-interaction (PPI) networks. Those highly connected proteins in the cell PPI network are known to play essential roles in cellular control (2–4), and many of them are encoded by essential genes. Therefore, there has been a special interest in the study of the structural and functional properties of hubs that differentiate them from nonhubs (2,5). These differences are manifested in 1), structural properties of the interfaces, and 2), their thermodynamic stability, as further described below.

The identification of structural properties of hubs is based on gene expression patterns and localization of proteins within the cell, as defined by Han et al. (2,3). The authors identified two classes of hubs. The first class contains hubs that interact with all their partners at the same time and in the same space (referred to as Party hubs), and the second consists of hubs that interact with their partners at different times or locations (referred to as Date hubs). The differences between different types of hubs could be manifested in molecular properties (structure and sequence) of Party- and Date-hub proteins. Indeed, based on crystallographic data, Kim et al. (8,9) found two classes of hubs: Multi hubs, which interact with their partners through different interfaces, and Single or Singlish hubs, which

interact with most of their partners through one interface. Therefore Singlish hubs cannot interact simultaneously with all their partners, and they generally tend to behave as Date hubs (10).

The next line of evidence highlighting the differences between hubs and nonhubs is based on structural disorder. A hypothetical folding-upon-binding mechanism to provide functional promiscuity in PPIs posits that hub proteins are disordered but get folded upon binding to a partner, potentially acquiring different tertiary structures with different interaction (11,12). Therefore, the study of hub proteins and their interaction partners is relevant for the understanding of the particular role that disorder might play in multifunctional molecular recognition. Bioinformatics analysis provided an initial support to the disorder hypothesis by showing that Singlish hubs and their partners have a higher level of predicted disorder than do Multi hubs, and higher than the proteome average (9). In addition, It has been shown that surfaces of some hubs are enriched in charged and polar amino acids and depleted of hydrophobic content (13–16), a signature of disordered proteins (17). However, despite significant efforts, a clear understanding, at the molecular level, of the physical mechanisms that use hydrophobicity and electrostatics to provide functional promiscuity of hubs in a PPI network remain elusive.

Recently, we developed a multiscale microscopic biophysics-based model to study evolution of functional and nonfunctional PPIs within a simple, yet nontrivial, functional PPI network (18,19). The analysis highlighted an intrinsic conflict in the evolution of hub proteins between

Submitted May 1, 2014, and accepted for publication August 1, 2014.

*Correspondence: shakhnovich@chemistry.harvard.edu

Editor: David Eliezer.

© 2014 by the Biophysical Society
0006-3495/14/10/1686/11 \$2.00

<http://dx.doi.org/10.1016/j.bpj.2014.08.004>



the requirement to maintain multiple functional interactions and the need to avoid ever more abundant nonfunctional ones (20–23). The evolutionary compromise between these factors was achieved in the model (19) by modulating the intracellular abundance of hub proteins and simultaneously decreasing their surface hydrophobic content. Here, we employed the modified biophysics-based model, along with bioinformatics analysis and simple arguments, to study physical principles of evolutionary design of different types of functional PPI hubs. We provide reasoning as to why proteins belonging to Singlish hub systems are more prone to disorder than those belonging to Multi hub systems (9). We found that PPIs of hubs of different types are stabilized by different kinds of physical interactions (e.g., electrostatic, hydrophobic, etc.), both in the biophysics-based model and in the analysis of protein crystallographic databases (Structural Interaction Network (SIN) database (8,9,24)). Finally, we use combinatoric arguments to show that the physical nature of hydrophobic and electrostatic interactions, in particular, their favoring of either same-type (hydrophobic-hydrophobic) or opposite-type interactions (negatively-positively charged), plays a key role in maintaining the network topology while allowing the protein amino acid sequence to evolve. For clarity, we include a flowchart in Fig. 1, depicting our workflow and how results from the lattice-model proteins, SIN database, and combinatoric model are utilized.

METHODS

Lattice protein simulations

We build on a recent biophysics-based multiscale model of evolution wherein simple-lattice-model globular proteins are encoded in genomes of

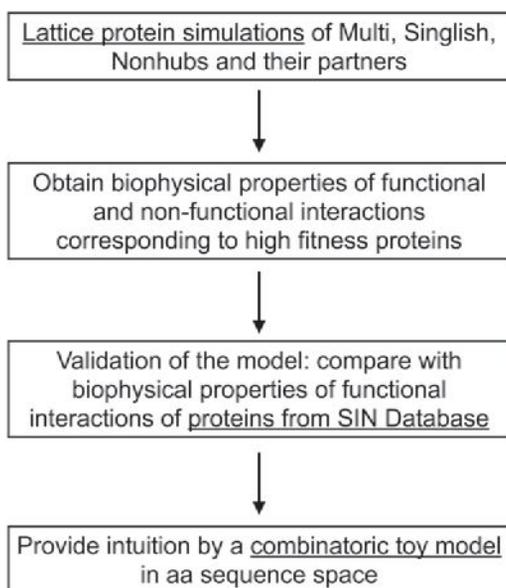


FIGURE 1 Flowchart depicting our workflow, and highlighting how results from each description level (lattice model proteins, SIN database, and combinatoric toy model) are utilized.

model cells. These genomes are subject to mutations, and the corresponding amino acid sequences determine the precise structure, stability, and interactions of the proteins involved (19,25). In previous work, these simulations have been successful in drawing a direct link between the genotype-phenotype relationship of an organism and several evolved biophysical properties of proteins in its cells, such as their stability, abundance, and PPI strength.

Each of our model proteins consists of 27 amino acid residues that fold into $3 \times 3 \times 3$ cubic lattice conformations (Fig. 2 A). This 27-mer lattice model protein has 103,346 maximally compact conformations (26). Following Heo et al. (19), for computational efficiency, we use a subset of 10,000 randomly selected conformations as our ensemble. Only amino acids occupying neighboring sites on the lattice can interact, and the interaction energy depends on amino acid types according to the Miyazawa-Jernigan (MJ) potential (27), both for intra- and intermolecular interactions. We calculate the Boltzmann probability of folding to a native state, P_{nat}^i , for each protein:

$$P_{\text{nat}}^i = \frac{e^{-E_0^i/T}}{\sum_{k=1}^{10,000} e^{-E_k^i/T}}, \quad (1)$$

where E_0^i is the energy of the most stable, or native, conformation out of 10,000 conformations and T is the temperature in arbitrary units. The probability of folding, P_{nat}^i , is therefore a proxy for the degree of disorder of protein i . We impose the condition that the native conformation is the minimum energy conformation in the conformational ensemble. Mutations that lead to violation of this condition are considered lethal: every protein in the model proteome is deemed essential, so that its failure to fold deprives the cell of an essential function, causing the lethal phenotype. We model the PPIs with a rigid docking scheme. Six faces of a cubic lattice provide six possible interaction surfaces, and there are four possible directions (which correspond to rotational degrees of freedom) to dock two lattice proteins through two interaction surfaces. Hence, in total, there exist $6 \times 6 \times 4 = 144$ docking modes for a binary protein complex. The Boltzmann probability of interaction in a functional binding mode between proteins i and j is

$$P_{\text{int}}^{ij} = \frac{e^{-E_f^{ij}/T}}{\sum_{k=1}^{144} e^{-E_k^{ij}/T}}, \quad (2)$$

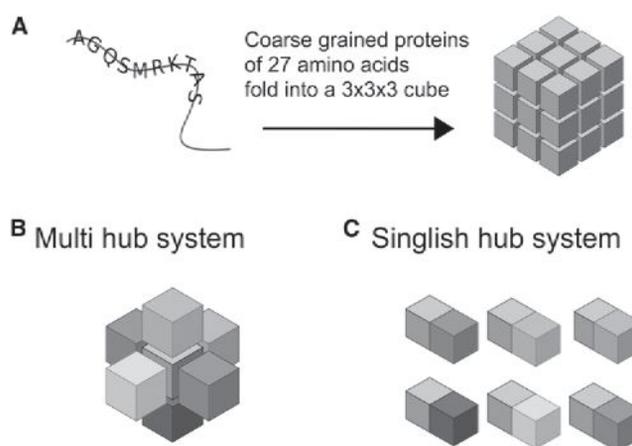


FIGURE 2 (A) Schematic representation of the lattice protein model. (B and C) Model proteins in the system, including the hub protein (gray) and up to six interaction partners (colors). For the Multi hub system (B), each face of the hub is designated to interact with a specific binding partner; For the Singlish hub system (C), only a single face is chosen as an interaction interface with all the partners. To see this figure in color, go online.

where E_j^{ij} is the interaction energy of a functional binding mode (defined below), and E_k^{ij} are the interaction energies for all 144 docking modes. We assume that each protein in the cell folds with a two-state folding kinetics,



where U_i denotes unfolded states and F_i the unique native state. k_f^i and k_u^i are the folding and unfolding rate constants, respectively. The steady-state solution is

$$F_i = \frac{k_f^i}{k_u^i} U_i = \frac{P_{\text{nat}}^i}{1 - P_{\text{nat}}^i} U_i. \quad (4)$$

We use the law of mass action (LMA) as described in Heo et al. (25) to calculate all possible functional and nonfunctional interactions between all proteins with folded conformations. We solve the LMA equations numerically (25) to obtain the concentrations of the complexes, with a permitted error of 10^{-15} for each protein concentration. Once the concentrations of protein complexes are obtained, we can define the concentration of functional complexes. The functional concentrations are

$$G_{i,j}^f = [F_i F_j] \times P_{\text{int}}^{ij}, \quad (5)$$

where $[F_i F_j]$ are the concentrations of dimers formed by folded proteins i and j in any configuration. We consider both Multi and Singlish hubs (Fig. 2, B and C, respectively). For a Multi hub, each face of the hub is designated to functionally interact with a specific binding partner. For a Singlish hub, a single face is chosen as a functional interaction surface available to all partners, and interactions with different partners are physically mutually exclusive and thus cannot occur simultaneously. We limit the number of partner proteins, N_p , to 6, i.e., the maximal proteome size (including the hub) is 7. Nonfunctional concentrations are defined as

$$G_i^{\text{nf}} = C_i - \sum_{j=1}^{N_p+1} G_{i,j}^f, \quad (6)$$

where C_i is the total concentration of protein i . The fitness of each cell in the population is given by the cell division rate,

$$b = b_0 \frac{\prod_{i=2}^{N_p+1} G_{\text{hub},i}}{1 + \alpha \left(\sum_{i=1}^{N_p+1} C_i - C_0 \right)^2}, \quad (7)$$

where b_0 is a constant parameter chosen at the beginning of simulations to scale the rate, and thus the timescale, of evolution, and α scales the overexpression penalty. To limit protein overexpression, we set the parameters $\alpha = 500$ and $C_0 = (N_p + 1) \times 0.1$, as the initial protein concentration was chosen to be $C = 0.1$. We use the Gillespie algorithm (28) with fixed population size ($N = 100$). Upon cell division, a mother cell gives birth to a daughter cell. To keep the population size constant, a newborn cell replaces a randomly chosen cell in the population. Upon replication, both the mother and daughter cells are subjected to either a mutational event with constant rate ($\mu = 10^{-3}$ /gene/replication) or to a change in the expression level of one protein in the cell. This protein is chosen randomly, and its expression level changes with a constant rate ($r = 0.01$ /cell division) such that its concentration in the new cell is obtained from the old one as $C^{\text{new}} = C^{\text{old}} + \epsilon$, where ϵ is a Gaussian random num-

ber with zero mean and variance 0.1. All simulations ran for 6×10^6 generations and were considered to be in a steady state when C_i , P_{nat}^i , and P_{int}^{ij} reached a plateau value. The random seed of each simulation is initialized with a unique value. The results described below are averaged over 50 different random simulations.

To calculate the evolutionary rates of proteins in the simulation, we count the number of fourfold, twofold, and nondegenerate sites (corresponding to four, two, and one amino acid representation in nucleotide space), and use the formula of Hartl and Clark ((29), page 340):

$$\begin{aligned} N_s^t &= (\text{fourfold degenerate sites}) \\ &+ 1/3 \times (\text{twofold degenerate sites}) \\ N_a^t &= (\text{nondegenerate sites}) \\ &+ 2/3 \times (\text{twofold degenerate sites}), \end{aligned} \quad (8)$$

where N_s^t is the number of synonymous sites at generation t , and N_a^t is the number of nonsynonymous sites at generation t . Next, we calculate

$$\begin{aligned} dN(t) &= 2A_s^t / (N_s^0 + N_s^t) \\ dS(t) &= 2A_a^t / (N_a^0 + N_a^t), \end{aligned} \quad (9)$$

where N_s^0 and N_a^0 are the numbers of synonymous and nonsynonymous sites, respectively, of the initial sequence. A_s^t and A_a^t are the numbers of synonymous and nonsynonymous amino acids, respectively, at generation t . The evolutionary rate is then defined to be dN/dS .

SIN database analysis

To compare the lattice proteins to natural proteins, we used the SIN database (8,9,24), which identified Multi and Singlish hubs and their available atomistic structures in complex with their partners at the Protein Data Bank (PDB). We used the information available for human proteins, in total 37 Multi, 36 Singlish, and 30 Nonhub interactions (see Table S1 in the Supporting Material for the list of PDB files considered). Nonhub interactions were defined as interactions between proteins that do not belong to a Singlish or Multi hub group. We identified and analyzed the interfaces, and studied the statistics of contacts (two residues are defined to be in contact when any pair of atoms from the two different amino acids are separated by $<4 \text{ \AA}$). To facilitate comparison with simulations we used the MJ potential to evaluate the contact energies.

Combinatoric toy model

The model consists of three proteins (a hub and two partners), where each protein has two interaction surfaces. The simulation space is one-dimensional (1D) and proteins can interact via a single interface. In total, there are six surfaces in the system, and surfaces can be hydrophobic, hydrophilic, or positively or negatively charged. This corresponds to the reduction of the amino acid pull from 20 to 4. We use an averaged MJ potential for the four types of amino acids, where hydrophobic amino acids were identified as M, F, I, L, V, W, P, and C, negatively charged as D and E, and positively charged as R and K (30) (as done for the lattice-model simulation and SIN database analysis). The model system has a countable sequence space. In particular, there are six surfaces, each of which can be one of four types; thus, there are a total of $4^6 = 4096$ possible configurations.

For each configuration, we solve the LMA equations

$$C_i + C_j \xrightleftharpoons{K_{ij}} C_i \times C_j, \quad (10)$$

where C_i and C_j are the concentrations of protein i and protein j , respectively. The initial concentrations were chosen to be $C_{\text{hub}} = 0.5$ and $C_{\text{par}} = 0.1$ (the latter set for both partners). The association constants are defined as

$$K_{i,j} = \frac{[C_i \times C_j]}{[C_i][C_j]} = \sum_{k,m=1}^{3 \text{ or } 4} e^{-E_{k,m}}, \quad (11)$$

where $E_{k,m}$ is the averaged MJ interaction energy mentioned above and k and m are the particular bound surfaces of proteins i and j , respectively. For interaction between two different proteins, there are four modes of interaction. For interaction between two identical proteins, there are three modes of interaction due to the 1D symmetry in the system. The functional concentrations are defined as

$$G_{\text{hub,par1}} = [C_{\text{hub}} \times C_{\text{par1}}] \frac{e^{-E_{\text{pair1}}}}{\sum_{k,m=1}^4 e^{-E_{k,m}}} \quad (12)$$

$$G_{\text{hub,par2}} = [C_{\text{hub}} \times C_{\text{par2}}] \frac{e^{-E_{\text{pair2}}}}{\sum_{k,m=1}^4 e^{-E_{k,m}}}$$

The functional pair varies between Multi, Singlish, and Nonhub systems. For Multi hubs, each surface of the hub is designated for a different partner. For Singlish hubs, the same surface of the hub interacts with both partners. For Nonhubs, we define a single functional interaction, whereas the second partner is required to remain in a monomeric form. For hub systems, the fitness of the configuration is

$$f = G_{\text{hub,par1}} \times G_{\text{hub,par2}}, \quad (13)$$

whereas for Nonhub systems, the fitness of the configuration is

$$f = G_{\text{hub,par1}} \quad (14)$$

RESULTS

Lattice protein simulations

Even though the simulated-lattice-protein Multi, Singlish, and Nonhub systems considered here differ in their functional interaction networks and numbers of partners, some trends are similar in most systems. 1), Shortly after the beginning of the simulations, the concentration of the hub proteins increases such that $C^{\text{hub}} = \sum_i C_i^{\text{partner}}$ (after $\sim 10^4$ generations; see Fig. 3 A). 2), Hub and partner proteins eventually evolve a stable structure (P_{nat} increases), whereas some proteins experience a sharp drop in P_{nat} beforehand (see Fig. 3 B). 3), The nonfunctional concentration, G_i^{nf} , of the hub increases at first due to the increase in hub concentration, but when the hub concentration, C , stabilizes, the G_i^{nf} of the hub proteins decreases. The G_i^{nf} of partners decreases monotonically throughout the simulation (see Fig. 3 C). 4), The hub-partner pairs develop functional surfaces with increasing probability of functional interaction, P_{int} (see Fig. 3 D).

Effect of number of partners

One particular goal was to understand the evolutionary processes that dominate changes in the thermodynamic stability of hub proteins and their level of disorder. In both Multi and Singlish hub systems, hub and partner proteins evolve a stable structure (P_{nat} increases). Although the P_{nat} of hubs tends to increase monotonically, those of partner proteins experience a drop before maximal stability is reached. We studied the effect of number of partners on the evolution of protein stability, as reflected in P_{nat} . Fig. 4 shows the P_{nat} of hubs

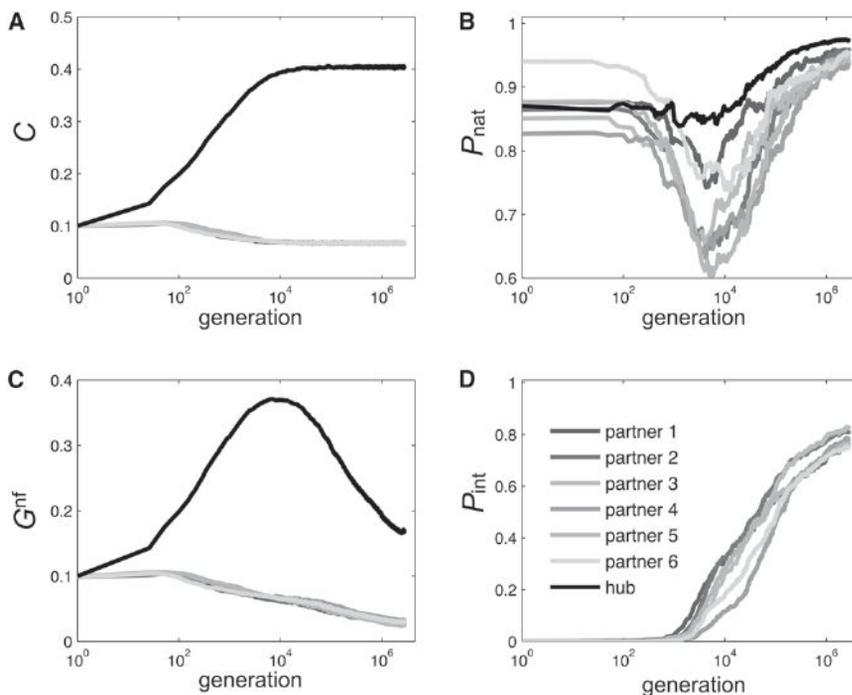


FIGURE 3 Evolution of molecular properties in Singlish hub systems with six interaction partners with regard to concentration, C , of hub (black) and partner proteins (colors; see legend in panel D) (A), thermodynamic stability, P_{nat}^i (B), nonfunctional concentration, G_i^{nf} (C), and functional interaction probabilities, P_{int}^{ij} (D). Data are plotted versus generation, averaged over 50 independent realizations (of different random-number seed), and shown in logarithmic scale for the x axis. The Multi hub system shows similar behavior for these quantities. To see this figure in color, go online.

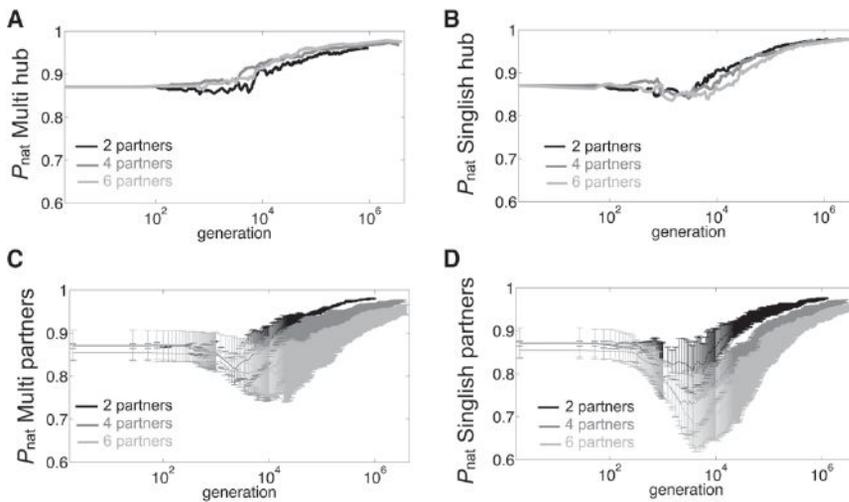


FIGURE 4 Effect of the number of partners on stability, P_{nat} , of proteins as a function of generation for Multi hubs (A), Singlish hubs (B), partners of Multi hubs (C), and partners of Singlish hubs (D). Systems with two, four, and six partners are shown in black, blue and magenta, respectively. For partner proteins, P_{nat} is averaged over all partners in the system, and error bars correspond to 1 Standard Deviation (SD). All plotted data (including averaged P_{nat} and SD) are averaged over 50 independent realizations and shown in logarithmic scale for the x axis. To see this figure in color, go online.

and partners for Multi (Fig. 4, A and C, respectively) and Singlish hub systems (Fig. 4, B and D, respectively) and compares systems with two, four, and six partners. The P_{nat} of partners of both Multi and Singlish hub systems drops with respect to its initial value. As the number of partners increases, both the amplitude of the drop and the time necessary for stabilization after reaching the minimal P_{nat} increase. The drop in P_{nat} is more profound for partners of Singlish hubs in comparison to those of Multi hubs. As partners lose stability, they open the sequence space needed to improve the interaction strength, P_{int} (Fig. 5). However, functional interaction surfaces in Multi hub systems evolve more slowly and poorly, as can be seen in Fig. 5 for systems with a maximal partner number of six.

Evolutionary rates

In previous studies, the evolutionary rate of proteins has been shown to correlate negatively with the number of functional interfaces (8), yet a clear understanding of the behavior of the evolutionary rates of hub proteins and partners throughout evolution remains elusive. Therefore, we calculated the evolutionary rates, dN/dS , by counting the number of synonymous and nonsynonymous sites in the nucleotide sequence (see further details and Eq. 9 in Methods).

In Fig. 6, we plot the evolutionary rate for hub proteins (Fig. 6 A) and partner proteins (Fig. 6 B) as a function of generation. Our results are in agreement with those of Kim et al. (8) and show that the more functional interfaces a protein has, the lower is its evolutionary rate in steady state (once the evolutionary rate reaches a plateau); Singlish hubs, which have a single interaction surface, reach a higher evolutionary rate than do Multi hubs, which have six functional interfaces (Fig. 6 A). Partners of both Multi and Singlish hubs reach a similar plateau value (Fig. 6 B). Both Singlish hubs and their partners experience an increase in evolutionary rate, $\sim 10^4$ generations before Multi hubs and their partners.

Biophysics of specificity

Next we turn to the question of how hubs use the hydrophobic and charge interactions to design their surfaces to maximize functional interactions. To that end, we examined the functional energy contributions from hydrophobic and electrostatic contacts once simulations reached a steady state (after 6×10^6 generations). Contacts between amino acid pairs were classified as hydrophobic contacts if both amino acids were hydrophobic (hydrophobic amino acids were identified as M, F, I, L, V, W, P, and C (30)), or as electrostatic contacts if the pair contained a positively and a

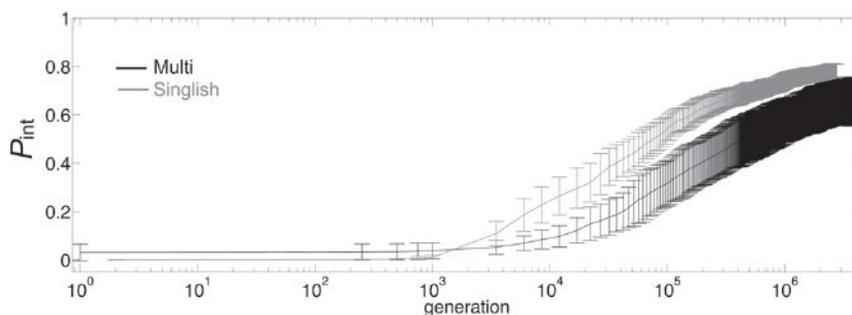
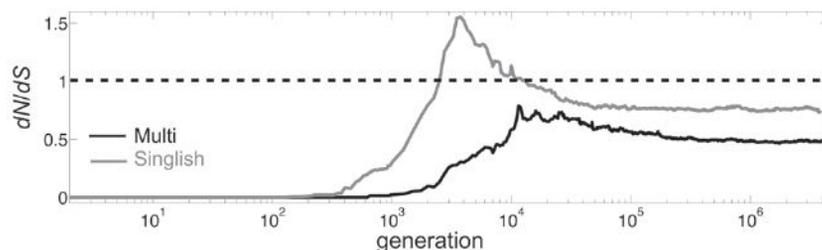


FIGURE 5 Comparing P_{int} for Multi and Singlish hub systems with the maximal number of partners (six). P_{int} is averaged over all partners in the system, and error bars correspond to 1 SD. P_{int} and SD values are averaged over 50 independent realizations and shown in logarithmic scale for the x axis. To see this figure in color, go online.

A Evolutionary rate of hub proteins



B Evolutionary rate of partner proteins

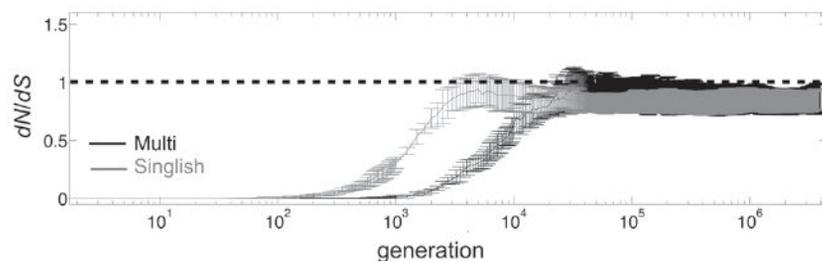


FIGURE 6 Evolutionary rates, dN/dS , versus generation, for hub proteins (A) and partner proteins (B). In B, the data are averaged over all six partners, and error bars correspond to 1 SD. Plotted data are averaged over 50 independent realizations and shown in logarithmic scale for the x axis. To see this figure in color, go online.

negatively charged amino acid (negatively charged amino acids were identified as D and E, and positively charged amino acids as R and K (30)). Values of the hydrophobic and electrostatic energies were normalized by the total interaction energy of their corresponding functional interface. We extracted those energies for 50 different realizations, and we plot the hydrophobic and electrostatic functional energy histograms in Fig. 7 A and B, respectively. The latter show that Multi hubs use electrostatic contacts more than Singlish do. In particular, the mean electrostatic contribution is $18.6 \pm 0.1\%$ and $13.2 \pm 0.1\%$ for Multi and Singlish hubs, respectively (\pm values represent the variance calculated over multiple realizations; Kolmogorov-Smirnov (KS) test, $p = 9.6 \times 10^{-18}$). Nonhubs have a lower mean electrostatic contribution $12.7 \pm 0.3\%$ compared to hub systems. The same analysis for hydrophobic contacts shows an opposite trend: mean contributions of $52.6 \pm 0.4\%$ and $66.2 \pm 0.3\%$ for Multi and Singlish hubs, respectively (KS test, $p = 3.9 \times 10^{-32}$). Nonhubs have a higher mean hydrophobic contribution ($69.4 \pm 0.4\%$) in comparison to hubs.

As the simulations gave us clear insights and predictions regarding the steady-state biophysical properties of hub-protein systems, the next step was to test these predictions for natural proteins. To that end, we used the SIN database (8,9,24), which identified Multi and Singlish hubs and their available atomistic structures in complex with their partners at the PDB (see further details in Methods). As in the simulations analysis, we compared the percentages of hydrophobic-hydrophobic and opposite-charge contacts for both Multi and Singlish hubs (see Fig. 7, C and D). This analysis verified our prediction that electrostatic contacts are more predominant in Multi hubs and hydrophobic contacts are

more predominant in Singlish hubs. The electrostatic contributions calculated using the SIN database are $15.5 \pm 1.1\%$ and $7.8 \pm 0.5\%$ for Multi and Singlish hubs, respectively (\pm values represent variance calculated over multiple PDB files; KS test, $p = 0.001$). Nonhubs have an electrostatic contribution ($8.0 \pm 0.3\%$) comparable to that of Singlish hubs. The same analysis for hydrophobic contacts shows average contributions of $27.8 \pm 2.1\%$ and $28.4 \pm 2.5\%$ for Multi and Singlish hubs, respectively (KS test indicates a nonsignificant difference between the two). However, Nonhubs have a higher hydrophobic contribution ($38.6 \pm 2.4\%$) in comparison to hubs (KS test, $p = 0.012$ and $p = 0.004$ for Multi and Singlish hubs, respectively).

Next, we analyzed the nonfunctional interaction energies. Due to the geometrical complexity of the crystallographic 3D structure of proteins, nonfunctional PPI interfaces are not clearly defined, and therefore, analysis of the nonfunctional energies is not feasible for proteins in the SIN database. However, nonfunctional interaction energies are accessible in the lattice-protein simulations, and their energetic contribution is calculated as for the functional ones. Again, we consider both hydrophobic and electrostatic contributions to the interaction energies. As we are interested in the negative design of nonfunctional interactions (i.e., minimizing nonfunctional interactions), we extract information about the following least favorable interactions: hydrophobic with nonhydrophobic (marked as the hydrophobic contribution), and electrostatic interactions between charges with the same sign (marked as the electrostatic contribution).

We plot the nonfunctional energy histograms in Fig. 8, showing the mean energetic contributions of both

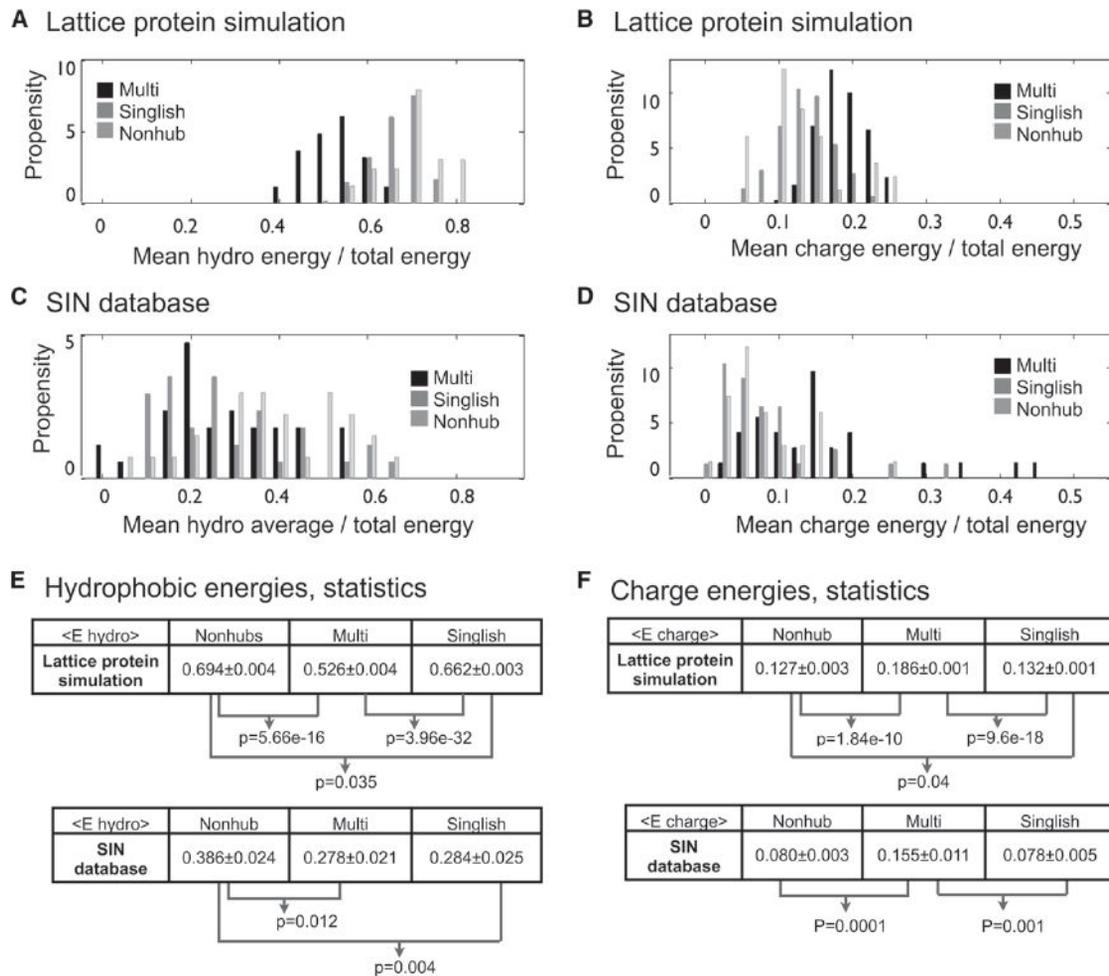


FIGURE 7 Mean interaction energy histograms showing fractions of hydrophobic and electrostatic contributions for lattice protein simulations (A and B, respectively) and SIN database (C and D, respectively). For the lattice protein, energies are calculated at the end of the simulations at generation 6×10^6 . Multi and Singlish hub systems are shown in black and blue, respectively. Nonhub systems are shown in magenta. Data were binned to generate hydrophobic energy histograms (bin size $dx = 0.05$), and charge energy histograms (bin size $dx = 0.025$). (E) Statistics of hydrophobicity histograms expressed as mean \pm variance. (F) Statistics of electrostatic histograms expressed as mean \pm variance. The p values were extracted using the KS test. Only significant values ($p < 0.05$) are noted. To see this figure in color, go online.

hydrophobic (Fig. 8 A) and electrostatic contacts (Fig. 8 B) divided by the total energy of the interaction. Similar to the case for functional interactions, Multi hubs use electrostatic contacts in their nonfunctional PPI network more than Singlish hubs and Nonhubs. In particular, the electrostatic contribution is $8.8 \pm 0.004\%$ for Multi hubs, and $7.8 \pm 0.005\%$ and $7.8 \pm 0.003\%$ for Singlish and Nonhubs, respectively (KS test, $p = 7.65 \times 10^{-20}$ and $p = 2.27 \times 10^{-9}$ for Multi-Singlish and Multi-Nonhub, respectively). The electrostatic contribution is not significantly different for Singlish hubs and Nonhubs. Therefore, in comparison to Singlish hubs and Nonhubs, Multi hubs use more electrostatic interactions to both maximize probabilities for functional interactions and minimize probabilities for nonfunctional interactions.

The same analysis for hydrophobic contacts shows that the distributions of Multi hubs, Singlish hubs, and Nonhubs

are significantly different (see p values in Fig. 8 C). Nonhubs use the smallest number of hydrophobic contacts for the nonfunctional interactions, $18.7 \pm 0.005\%$ on average. The average hydrophobic contribution is $25.5 \pm 0.022\%$ for Singlish hubs and $26.9 \pm 0.021\%$ for Multi hubs. The trend showing the average hydrophobic contribution is highest for Multi hubs, intermediate for Singlish hubs, and lowest for Nonhubs is exactly the opposite of their relationship when accounting for functional interactions. Therefore, Multi hubs are highly efficient in negative design of their nonfunctional surfaces for minimizing unfavorable interactions, both hydrophobic and electrostatic.

Combinatoric toy model

Finally, we turn to a simpler model to provide insight into why Multi hubs rely on electrostatics to design their

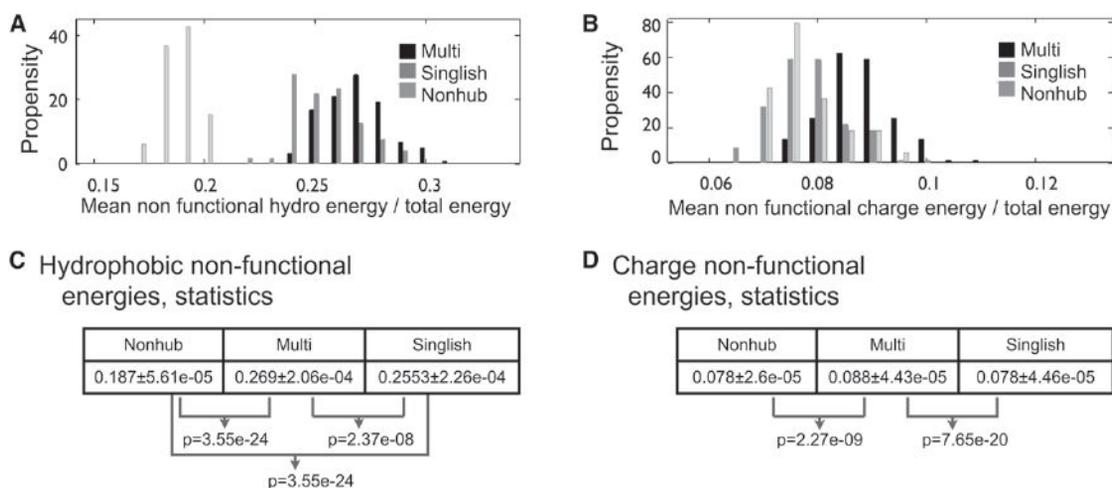


FIGURE 8 Negative design in the multiscale evolutionary model. Nonfunctional interaction energy histograms showing the contribution fraction of hydrophobic interactions (**A**) and electrostatic interactions (**B**) for lattice protein simulations. Multi and Singlish hub systems are shown in black and blue, respectively. Nonhub systems are shown in magenta. Bin sizes used to generate hydrophobic energy histograms and charge energy histograms were $dx = 0.01$ and $dx = 0.005$, respectively. (**C**) Statistics of hydrophobicity histograms expressed as mean \pm variance. (**D**) Statistics of electrostatic histograms expressed as mean \pm variance. The p values were extracted using the KS test. Only significant values ($p < 0.05$) are noted. To see this figure in color, go online.

functional interaction surfaces, whereas Singlish hubs make predominant use of hydrophobic interactions. One possible explanation involves combinatorics, or counting of states, where a state is a certain amino acid sequence along the surfaces of the proteins. In particular, we wish to identify high-fitness sequences for Multi hubs, Singlish hubs, and Nonhubs and to classify them according to their hydrophobicity and charge content. To that end, we created a simplified 1D model of a hub and two partners, where each protein has two surfaces. Each surface is assigned to be one of four types: hydrophobic, positively or negatively charged, or neutral. The fitness (defined in Eqs. 13 and 14) is calculated according to the probability of functional interactions (see Methods for further details). The advantages of the simplified model are 1), the feasibility of sequence space enumeration, and 2), the isolation of surface energetics from thermodynamic stability. The combination of these advantages allowed us to address the fitness-maximization problem as a tiling puzzle, where electrostatic and hydrophobic surfaces are distributed among the proteins.

We start by considering the set of 80 top fitness configurations shown in Fig. 9, A–C, for Multi, Singlish, and Nonhub systems. The vertical axis indicates fitness. For each system, functional interactions are indicated in magenta for the top configuration. Hydrophobic surfaces are shown in green, hydrophilic in black, positively charged in red, and negatively charged in blue. We find that the maximal fitness configurations for both Multi and Singlish hub systems contain hydrophobic functional surfaces. For Multi hub systems, the first ~ 40 configurations have two hydrophobic functional contacts, but the rest of the top configurations have one hydrophobic and one electrostatic functional

contact. This is different from the Singlish- and Nonhub systems, where all top configurations have two hydrophobic functional contacts.

Further understanding of the origin of the favoring of electrostatic functional interactions in Multi hub systems can be extracted from the fitness distributions plotted in Fig. 9, D–F, where we plot histograms of the number of configurations (in log scale) corresponding to a certain relative fitness (fitness divided by maximal fitness). The configuration count is divided for all hydrophobic functional contacts (*green*), all electrostatic functional contacts (*red*), and one hydrophobic and one electrostatic contact (*blue*, not available for the Nonhub systems, since they only have one functional contact). When fitness decreases, more configurations containing electrostatic functional surfaces become available, most profoundly for Multi hub systems. In situations where the maximal-fitness configuration is not easily approached (for example, due to limitations arising from thermodynamic stability considerations, i.e., stand-alone lattice proteins are known to be most stable when $\sim 30\%$ of their amino acid sequence is hydrophobic (19)), the next available configurations are more likely to contain electrostatic functional interfaces in Multi hub systems in comparison to Singlish-hub systems. This may explain why Multi hubs use more electrostatic than hydrophobic functional interactions and why the reverse is true for Singlish hubs in the lattice protein simulations and in the SIN database (see Biophysics of specificity). This effect can be quantified by the average fitness of the configurations considered, and indeed, the average fitness of configurations involving electrostatic functional interactions for Multi hub systems is $\langle f \rangle = 0.109 \pm 0.001$, which is

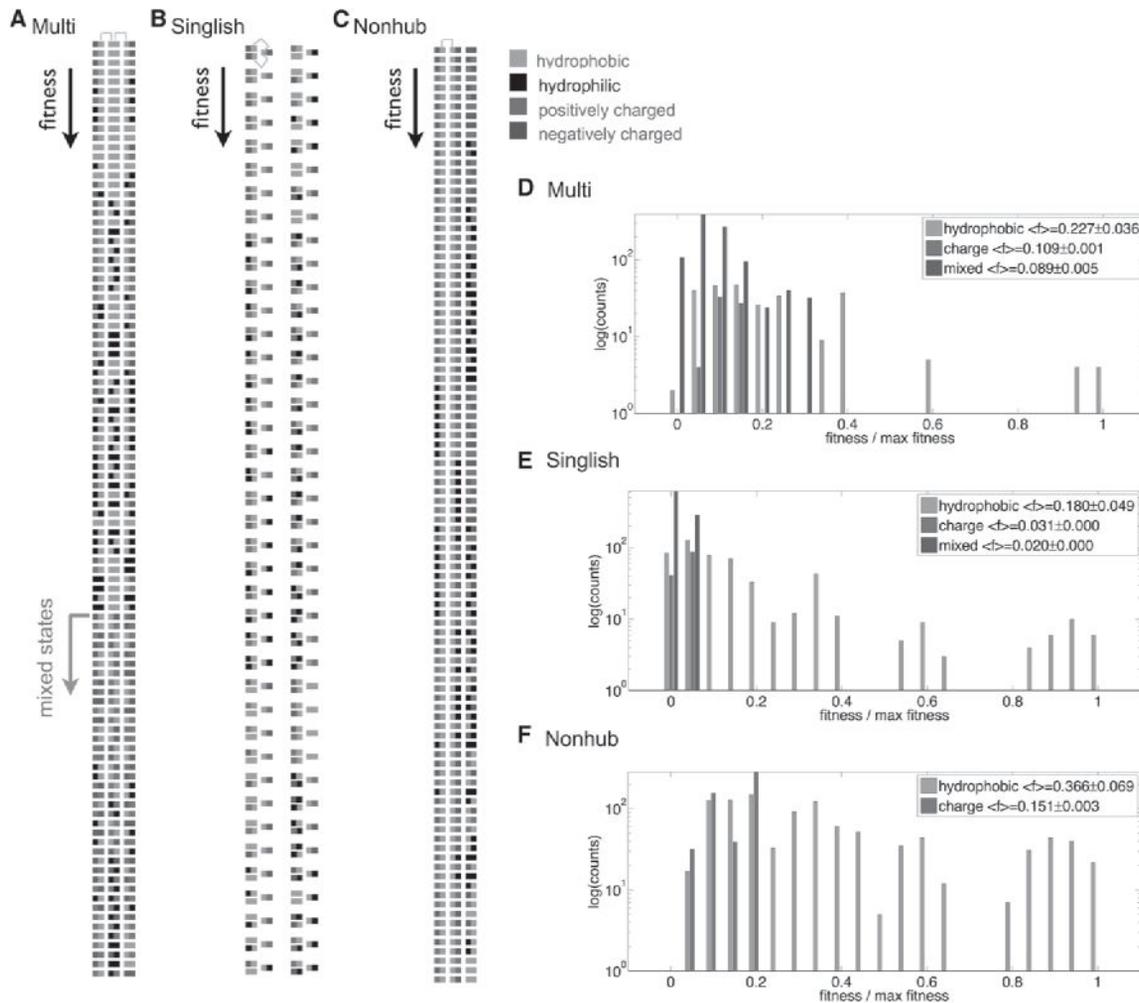


FIGURE 9 Results using a simplified 1D model. (A–C) The top 80 configurations for Multi (A), Singlish (B), and Nonhub systems (C). Configurations are arranged vertically according to their fitness (highest fitness at the top). For each system, functional interactions are indicated in magenta for the top configuration. Hydrophobic surfaces are shown in green, hydrophilic in black, positively charged in red, and negatively charged in blue. (D–F) Histograms of the number of configurations (in log scale) corresponding to a certain relative fitness (fitness divided by maximal fitness) are shown for Multi (D), Singlish (E), and Nonhub systems (F). The configuration count is divided for all hydrophobic functional contacts (green), all electrostatic functional contacts (red), and one hydrophobic and one electrostatic contact (blue; not available for the Nonhub system). Mean \pm variance values of the distributions are indicated in the legend. To see this figure in color, go online.

approximately three times larger than that for Singlish-hub systems, $\langle f \rangle = 0.031 \pm 0.0001$.

DISCUSSION

The findings presented here have implications for the conceptual understanding of the emergence of disorder in PPI networks, as well as the design of functional interactions within them. We start by providing reasoning as to why proteins belonging to Singlish-hub systems are more prone to disorder than those belonging to Multi hub systems (9). Our simulations show that throughout evolution, partners of Singlish hubs experience a larger drop in stability, P_{nat} , (i.e., they explore more disordered configurations), and a longer time to recover their maximal P_{nat} , in comparison to Multi hub partners (Fig. 4, C and D). Hubs in the

simulations experience a profoundly smaller drop in P_{nat} (Fig. 4, A and B), despite the bioinformatics analysis predicting a high level of disorder for Singlish hubs as well (9). This discrepancy could be attributed to the fact that in our simulations, proteins are assigned hub or partner exclusively, whereas in nature, hub proteins may become partners in other hub systems, particularly in Singlish-hub systems (9). Singlish-hub partners may utilize the drop in P_{nat} to open up a sequence space necessary for the evolution of functional interactions. Indeed, evolutionary rates of proteins in our simulations (Fig. 6) show a trend similar to that described by Kim et al. (8), who observed that evolutionary rates negatively correlate with the number of functional interfaces. Moreover, our simulations show that proteins in Singlish-hub systems spend more time in functional interaction complexes compared to proteins in Multi hub systems

and therefore evolve a more robust and efficient PPI (Fig. 5).

In addition, our simulations and analysis of the SIN database reveal two different mechanisms of surface design: whereas Multi hubs rely on electrostatics to design their functional interaction surfaces, Singlish hubs make predominant use of hydrophobic interactions (Fig. 7, *C* and *D*). Despite the simplified nature of the protein structure in the simulations, estimation of the contact energies for both simulated coarse-grained and atomistic real protein structures follow the trend mentioned above (Fig. 7, *A* and *B*) and suggest that our coarse-grained simulations capture the fundamental physics of hub protein systems. It is remarkable that different ratios of hydrophobic to electrostatic interactions of interfacial amino acids maintain distinct network topologies (i.e., Multi and Singlish) and underlie a fundamental design principle of PPI. Furthermore, the PPI takes place in a crowded environment, where functional interactions must overcome nonfunctional ones. Indeed, we show that hubs are highly efficient in designing their nonfunctional surfaces as well, for minimizing energetically favorable yet nonfunctional interactions; hubs use more hydrophobic mismatches (noncompatible amino acid pairs at the hub-partner interface) than do Nonhubs (Fig. 8 *A*), and Multi hubs use more electrostatic mismatches than do Singlish hubs and Nonhubs do (Fig. 8 *B*).

To further understand the origin of the two design mechanisms, we turned to a simplified 1D system of a hub and two partners, where each protein has two surfaces. Each surface is modeled as one of four types: hydrophobic, positively or negatively charged, or neutral. The model provided evidence that the different design mechanisms of Multi and Singlish hub systems arise from a limited number of configurations (i.e., sequences of surface types) in the system. As for the lattice protein simulations, high fitness requires both maximizing probabilities of functional PPIs and minimizing probabilities of nonfunctional PPIs. After enumerating all possible configurations and ranking them according to their fitness, we revealed that there are more high-fitness configurations employing charge in Multi hub systems than in Singlish hub systems (Fig. 9). This combinatoric reasoning implies that the physical nature of hydrophobic and electrostatic interactions, in particular their favoring of either same-type interactions (hydrophobic-hydrophobic) or opposite-type interactions (negatively-positively charged), plays a key role in maintaining the network topology while allowing the protein amino acid sequence to evolve.

CONCLUSIONS

In conclusion, our results highlight the origin of thermodynamic instabilities and several biophysical steady-state properties of hub protein systems. Although our findings reveal the fundamental design principles of these systems, there are opportunities to extend the scope of this work,

focusing, for example, on hierarchical systems composed of several hub networks and testing our predictions by experiments *in vitro* and possibly *in vivo*.

SUPPORTING MATERIAL

One table with details regarding the PPIs considered in the analysis of the SIN database is available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)00841-8](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00841-8).

This work was financially supported by Swiss National Science Foundation (SNSF) grant no. PBEZP3_140130 4 (O.P.) and National Science Foundation grant MCB-1243837 (E.S.).

REFERENCES

- Mosca, R., R. A. Pache, and P. Aloy. 2012. The role of structural disorder in the rewiring of protein interactions through evolution. *Mol. Cell. Proteomics*. 11: M111.014969.
- Han, J. D. J., N. Bertin, ..., M. Vidal. 2004. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*. 430:88–93.
- Krogan, N. J., G. Cagney, ..., J. F. Greenblatt. 2006. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*. 440:637–643.
- Jeong, H., S. P. Mason, ..., Z. N. Oltvai. 2001. Lethality and centrality in protein networks. *Nature*. 411:41–42.
- Zotenko, E., J. Mestre, ..., T. M. Przytycka. 2008. Why do hubs in the yeast protein interaction network tend to be essential: reexamining the connection between the network topology and essentiality. *PLOS Comput. Biol.* 4:e1000140.
- Reference deleted in proof.
- Reference deleted in proof.
- Kim, P. M., L. J. Lu, ..., M. B. Gerstein. 2006. Relating three-dimensional structures to protein networks provides evolutionary insights. *Science*. 314:1938–1941.
- Kim, P. M., A. Sboner, ..., M. Gerstein. 2008. The role of disorder in interaction networks: a structural analysis. *Mol. Syst. Biol.* 4:179. <http://dx.doi.org/10.1038/msb.2008.1016>.
- Andorf, C. M., V. Honavar, and T. Z. Sen. 2013. Predicting the binding patterns of hub proteins: a study using yeast protein interaction networks. *PLoS ONE*. 8:e56833.
- Dunker, A. K., M. S. Cortese, ..., V. N. Uversky. 2005. Flexible nets. The roles of intrinsic disorder in protein interaction networks. *FEBS J.* 272:5129–5148.
- Dyson, H. J., and P. E. Wright. 2002. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* 12:54–60.
- Patil, A., and H. Nakamura. 2005. Filtering high-throughput protein-protein interaction data using a combination of genomic features. *BMC Bioinformatics*. 6:100.
- Patil, A., and H. Nakamura. 2006. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett.* 580:2041–2045.
- Higurashi, M., T. Ishida, and K. Kinoshita. 2008. Identification of transient hub proteins and the possible structural basis for their multiple interactions. *Protein Sci.* 17:72–78.
- Higurashi, M., T. Ishida, and K. Kinoshita. 2009. PiSite: a database of protein interaction sites using multiple binding states in the PDB. *Nucleic Acids Res.* 37:D360–D364. <http://dx.doi.org/10.1093/Nar/Gkn659>.
- Uversky, V. N., J. R. Gillespie, and A. L. Fink. 2000. Why are “natively unfolded” proteins unstructured under physiologic conditions? *Proteins*. 41:415–427.

18. Cetinbaş, M., and E. I. Shakhnovich. 2013. Catalysis of protein folding by chaperones accelerates evolutionary dynamics in adapting cell populations. *PLOS Comput. Biol.* 9:e1003269.
19. Heo, M., S. Maslov, and E. Shakhnovich. 2011. Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions. *Proc. Natl. Acad. Sci. USA.* 108:4258–4263.
20. Zhang, J., S. Maslov, and E. I. Shakhnovich. 2008. Constraints imposed by non-functional protein-protein interactions on gene expression and proteome size. *Mol. Syst. Biol.* 4:210. <http://dx.doi.org/10.1038/msb.2008.1048>.
21. Deeds, E. J., O. Ashenberg, ..., E. I. Shakhnovich. 2007. Robust protein-protein interactions in crowded cellular environments. *Proc. Natl. Acad. Sci. USA.* 104:14952–14957.
22. Johnson, M. E., and G. Hummer. 2011. Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. *Proc. Natl. Acad. Sci. USA.* 108:603–608.
23. Levy, E. D., S. De, and S. A. Teichmann. 2012. Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc. Natl. Acad. Sci. USA.* 109:20461–20466.
24. Bhardwaj, N., A. Abyzov, ..., M. B. Gerstein. 2011. Integration of protein motions with molecular networks reveals different mechanisms for permanent and transient interactions. *Protein Sci.* 20:1745–1754.
25. Heo, M., L. Kang, and E. I. Shakhnovich. 2009. Emergence of species in evolutionary “simulated annealing”. *Proc. Natl. Acad. Sci. USA.* 106:1869–1874.
26. Shakhnovich, E., and A. Gutin. 1990. Enumeration of all compact conformations of copolymers with random sequence of links. *J. Chem. Phys.* 93:5967–5971.
27. Miyazawa, S., and R. L. Jernigan. 1996. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* 256:623–644.
28. Gillespie, D. T. 2007. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* 58:35–55.
29. Hartl, D. L., and A. G. Clark. 2007. Principles of Population Genetics. Sinauer Associates, Sunderland, MA.
30. Berezovsky, I. N., K. B. Zeldovich, and E. I. Shakhnovich. 2007. Positive and negative design in stability and thermal adaptation of natural proteins. *PLOS Comput. Biol.* 3:e52.

Article

Evaluation of Fluorophores to Label SNAP-Tag Fused Proteins for Multicolor Single-Molecule Tracking Microscopy in Live Cells

Peter J. Bosch,¹ Ivan R. Corrêa, Jr.,² Michael H. Sonntag,³ Jenny Ibach,⁴ Luc Brunsveld,³ Johannes S. Kanger,¹ and Vinod Subramaniam^{1,*}

¹Nanobiophysics, MESA+ Institute for Nanotechnology and MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, Enschede, The Netherlands; ²New England Biolabs, Ipswich, Massachusetts; ³Laboratory of Chemical Biology, Department of Biomedical Engineering, and Institute of Complex Molecular Systems, Eindhoven University of Technology, Eindhoven, The Netherlands; and ⁴Max Planck Institute of Molecular Physiology, Dortmund, Germany

ABSTRACT Single-molecule tracking has become a widely used technique for studying protein dynamics and their organization in the complex environment of the cell. In particular, the spatiotemporal distribution of membrane receptors is an active field of study due to its putative role in the regulation of signal transduction. The SNAP-tag is an intrinsically monovalent and highly specific genetic tag for attaching a fluorescent label to a protein of interest. Little information is currently available on the choice of optimal fluorescent dyes for single-molecule microscopy utilizing the SNAP-tag labeling system. We surveyed 6 green and 16 red excitable dyes for their suitability in single-molecule microscopy of SNAP-tag fusion proteins in live cells. We determined the nonspecific binding levels and photostability of these dye conjugates when bound to a SNAP-tag fused membrane protein in live cells. We found that only a limited subset of the dyes tested is suitable for single-molecule tracking microscopy. The results show that a careful choice of the dye to conjugate to the SNAP-substrate to label SNAP-tag fusion proteins is very important, as many dyes suffer from either rapid photobleaching or high nonspecific staining. These characteristics appear to be unpredictable, which motivated the need to perform the systematic survey presented here. We have developed a protocol for evaluating the best dyes, and for the conditions that we evaluated, we find that Dy 549 and CF 640 are the best choices tested for single-molecule tracking. Using an optimal dye pair, we also demonstrate the possibility of dual-color single-molecule imaging of SNAP-tag fusion proteins. This survey provides an overview of the photophysical and imaging properties of a range of SNAP-tag fluorescent substrates, enabling the selection of optimal dyes and conditions for single-molecule imaging of SNAP-tagged fusion proteins in eukaryotic cell lines.

INTRODUCTION

Single-molecule fluorescence microscopy has emerged in recent years as a powerful tool to investigate the structural dynamics and biological functions of proteins and macromolecular protein complexes (1–5). Single-molecule fluorescence approaches can reveal the dynamic interactions of individual proteins and heterogeneity in the spatial distribution of proteins that are difficult to detect using other fluorescence microscopy approaches (6–8). Despite the extraordinary advances in single-molecule fluorescence achieved to date, there remain many technical challenges that must be overcome to systematically study proteins in their native, highly complex, cellular environment. One of the challenges involves the specific and monovalent labeling of proteins of interest with a photostable fluorescent probe. In the last decade, several technologies have been developed that permit proteins to be specifically tagged with organic dyes in live cells (2,3,9–11). In this article, we focus on

the fluorescent labeling of proteins for single-molecule tracking.

Single-molecule fluorescence microscopy allows the tracking of proteins in a living cell at high resolution for a short period of time (12–15). The trajectories obtained contain valuable spatiotemporal information on interactions of proteins with their microenvironment (16–18). For instance, a protein may interact with other molecules, resulting in transient slowed diffusion or confinement by the cytoskeletal or other nanoscale compartmentalization structures in the plasma membrane (11,15,19–23). One of the main advantages of single-molecule fluorescence microscopy is the ability to track single protein molecules to provide details on the kinetics of protein association and dissociation. When the trajectories of a single protein species are recorded in multiple colors, they can reveal the kinetics of homodimerization interactions by comovement of the labeled molecules (11,24). For this comovement analysis, the protein species needs to be labeled with fluorophores emitting light at spectrally distinct wavelengths to allow simultaneous visualization at high resolution of two distinct proteins (of one protein species). Knowledge of protein interactions and their kinetics is important to understand the underlying signal

Submitted January 27, 2014, and accepted for publication June 10, 2014.

*Correspondence: subramaniam@amolf.nl

Vinod Subramaniam's present address is FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

Editor: David Piston.

© 2014 by the Biophysical Society
0006-3495/14/08/0803/12 \$2.00



<http://dx.doi.org/10.1016/j.bpj.2014.06.040>

transduction mechanisms and to model the cellular signal regulatory system (25–27).

A common approach to fluorescent labeling of proteins is to clone and express the protein of interest fused to an autofluorescent protein (FP). Several FPs are currently available that are suitable for single-molecule tracking, such as mCitrine, mCherry (28), and the infra-red iRFP (29). Although these genetically encoded labels allow multi-color tracking, FPs cannot match the photostability of small organic dyes (2,30), limiting the timescale over which a protein can be tracked and the accuracy with which it can be localized. To permit imaging of longer trajectories, fluorescent probes should ideally be bright and photostable (i.e., slow to photobleach) in addition to being specifically linkable to the protein of interest. The tools of choice in this case are organic dyes and quantum dots (Qdots). Although Qdots are extremely bright and photostable, they are larger than most proteins themselves, which might sterically hinder the movement of the protein (31). In addition, the use of monovalent quantum dots requires custom fabrication (32–34), and they might suffer from nonspecific labeling or aggregation (35,36).

More recent genetic techniques allow the specific and monovalent labeling of recombinant proteins with small organic fluorophores in live cells. Labeling by means of protein tags complements other approaches to labeling proteins with organic fluorophores, such as labeling with fluorescently labeled ligands and immunostaining with antibodies or Fab fragments. Since the tag does not compete with ligand binding and has no antagonistic function, the effects of ligands or inhibitors on ligand-free and fully functional receptors can be studied. The monovalency of the substrates ensures that no artificial clustering is induced. Among the most versatile of the protein tags is the SNAP-tag, a 20 kDa mutant of the human DNA repair protein O⁶-alkylguanine-DNA alkyltransferase (hAGT) that reacts specifically and rapidly with benzylguanine (BG) or benzylchloropyrimidine (CP) derivatives, leading to a covalent labeling of the SNAP-tag with a synthetic probe (37–41). The reaction occurs through a well-defined mechanism and predictable monova-

lent stoichiometry. For example, BG substrates derived from organic fluorophores react with SNAP-tag to provide specific labeling of a protein species with a fluorescent label at physiologically relevant conditions in the cell (Fig. 1). For labeling at the single-molecule level, SNAP-tag is especially suitable to label plasma membrane resident proteins using membrane-impermeable substrates.

There are a wide variety of reported applications in protein labeling for the SNAP-tag system, including super-resolution imaging (42–44), analysis of protein function (45) and protein half-life (46), observation of protein-protein interactions (47), sensing cell metabolites (48), and identification of drug targets (49). SNAP-tag labeling has also been utilized to study several protein complexes at the single-molecule level (50–59). Although the application of the SNAP-tag labeling system for single-molecule tracking had already been suggested (14), only recently did Calebiro et al. demonstrate the first example, to our knowledge, of this tagging technology for single-color fluorescence tracking (60). These authors used direct receptor labeling with SNAP-tag to dynamically monitor the adrenergic receptors β 1AR and β 2AR and the γ -aminobutyric acid GABAB receptor on intact cells, and compared their spatial arrangement, mobility, and supramolecular organization. Benke et al. have recently shown a new approach to single-molecule tracking by using the blinking properties of synthetic dyes attached to SNAP-tag (61). Although this approach optimizes the total number of observable diffusion steps, it does not facilitate the observation of a single protein for as extended a period of time as possible, a requirement for the study of binding associations.

Here we report a comprehensive survey on the photostability and binding specificity of several SNAP-tag fluorescent substrates using widefield and total internal reflection fluorescence (TIRF) single-molecule microscopy. We have investigated the suitability of 22 fluorescent substrates (BG dyes) by characterizing their properties in living cells using a C-terminal (extracellular) SNAP-tag fusion to the epidermal growth factor receptor (EGFR), a plasma membrane resident protein. Due to the significant autofluorescence of cells when

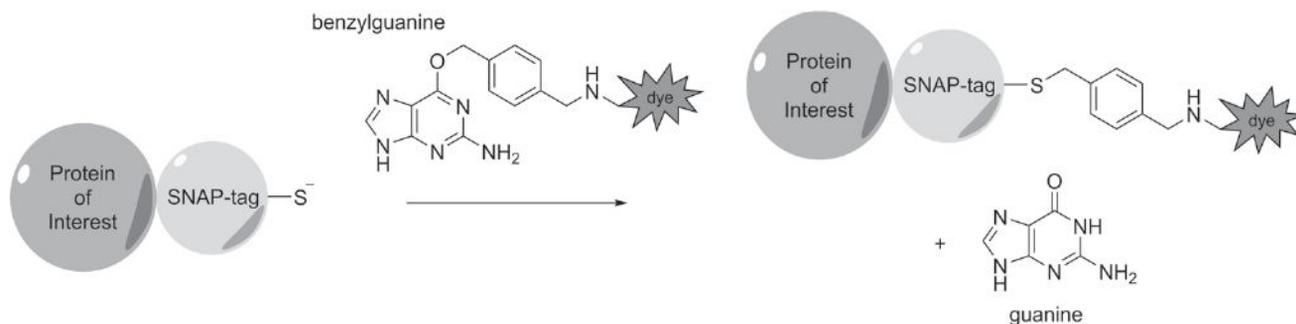


FIGURE 1 Schematic of the binding of a benzylguanine (BG) substrate to a SNAP-tag fusion protein. The SNAP-tag is fused to a protein of interest. Upon binding, the benzyl group reacts with a cysteine in the active site of SNAP-tag, releasing the guanine group. In this survey, the BG was conjugated to fluorescent dyes, but BG can in principle be coupled to any molecule of choice. To see this figure in color, go online.

using blue-excitable dyes, we limited our survey to green- and red-excitable dyes. Since a lot of dyes are available in this spectral range, we selected dyes from different manufacturers that are commonly available and used, trying to include dyes from various chromophore families. We emphasize that it is essential to study the fluorophores in the cellular setting, because photophysical properties are known to differ depending on the nature of their conjugate and their microenvironment. For example, different photostabilities have been observed for fluorescent proteins on different interfaces, due to the apparent role of the protein shell rigidity for each chromophore (62). In addition, the fluorescence of a number of fluorophores may be quenched by electron donors like guanine, tryptophan, etc. (63). Therefore, the photophysical properties of free substrates in solution or immobilized on a glass surface do not necessarily reflect their properties after reaction with the SNAP-tag fused protein. Very recently, the photostability of two red-excitable fluorescent substrates was measured for another protein tag (A-TMP) at the single-molecule level (64). The binding specificity for these substrates was not determined.

METHODS

Chemicals, purification, and analysis of SNAP-tag substrates

Commercially available compounds were used without further purification. SNAP-Surface Alexa Fluor 546 (BG-Alexa 546), SNAP-Surface 549 (BG-Dy 549), SNAP-Surface 632 (BG-Dy 632), SNAP-Surface 647 (BG-Dy 647), SNAP-Surface Alexa Fluor 647 (BG-Alexa 647), and SNAP-Surface 649 (BG-Dy 649) were obtained from New England Biolabs (Ipswich, MA).

BG-Atto 550, BG-Atto 565, BG-Atto 620, BG-Atto 633, BG-Atto 647N, BG-Atto 655, and BG-TF5 have been described previously (40–42,65). The remaining substrates for the labeling of SNAP-tag fusion proteins were prepared by reacting the building block BG-NH₂ (S9148, New England Biolabs) with commercially available *N*-hydroxysuccinimide esters (NHS) of the corresponding fluorophores. Atto Rho6G and Atto 532 were obtained from Atto-Tec (Siegen, Germany); Dy 549, Dy 630, Dy 634, Dy 648, and Dy 651 were obtained from Dyomics (Jena, Germany). CF633 and CF640R were obtained from Biotium (Hayward, CA), and Star635 was obtained from Abberior (Göttingen, Germany).

BG-549-549, BG-Dy 651, BG-CF 633, BG-CF 640R, and BG-Star 635 were purified and analyzed with the following equipment. Reverse-phase high-performance liquid chromatography (HPLC) was performed on an Agilent LC/MS Single Quad System 1200 Series (analytical) and Agilent 1100 Preparative-Scale Purification System (semi-preparative). Analytical HPLC was performed on a Waters Atlantis T3 C18 column (2.1 × 150 mm, 5 μm particle size) at a flow rate of 0.5 mL/min with a binary gradient from Phase A (0.1 M triethyl ammonium bicarbonate (TEAB) or 0.1% trifluoroacetic acid (TFA) in water) to Phase B (acetonitrile) and monitored by absorbance at 280 nm. Semipreparative HPLC was performed on VYDAC 218TP series C18 polymeric reverse-phase column (22 × 250 mm, 10 μm particle size) at a flow rate of 20 mL/min. Mass spectra were recorded by electrospray ionization (ESI) on an Agilent 6120 Quadrupole LC/MS system. BG-Atto Rho6G, BG-Dy 630, BG-Dy 634, and BG-Dy 648 were purified and analyzed as follows. Reverse-phase high-performance liquid chromatography (HPLC) was performed on the Shimadzu SCL-10 AD VP series (analytical) and the Shimadzu LC-20 AD System (preparative). Analytical HPLC was performed on a reverse-phase HPLC column (GraceSmart PP18, 50 mm × 2.1 mm, 3 μm) at a

flow rate of 0.20 mL/min and a binary gradient of acetonitrile in water (both containing 0.1% formic acid) at 298 K. Mass and ultraviolet-visible spectra were recorded with an ion trap (LCQ Fleet Ion Trap Mass Spectrometer, Thermo Scientific, Waltham, MA) and a diode array detector (Finnigan Surveyor PDA Plus detector, Thermo Electron, Waltham, MA). Preparative reverse-phase HPLC was performed on a reverse-phase HPLC column (GraceAlpha C18, 5 μm, 250 mm × 4.6 mm; Fisher Scientific, Waltham, MA) at a flow rate of 1 mL/min with an isocratic gradient of Phase A (0.1% formic acid in water or 25 mM ammonium acetate in water, pH 4) to Phase B (0.1% formic acid in acetonitrile) and monitored with an ultraviolet-visible detector (SPD-10AV VP series, Shimadzu, Kyoto, Japan).

Further details of the synthesis of the SNAP-tag substrates are described in the Supporting Material.

Cell culture

All cell culture materials were obtained from PAA Laboratories (Pasching, Austria) unless stated otherwise. MCF7 cells were cultured in high-glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum and penicillin streptomycin at 37°C with 5% CO₂. The H441 epithelial human lung adenocarcinoma cancer cell line was a gift from Anton Terwisscha van Scheltinga (Department of Medical Oncology, University of Groningen, Groningen, The Netherlands). These cells were cultured in Roswell Park Memorial Institute (RPMI) 1640 medium supplemented as above. The HeLa epithelial human cervix adenocarcinoma cancer cell line was a gift from Wilma Petersen (University of Twente, Twente, The Netherlands), and was cultured in Iscove's modified Dulbecco's media (IMDM) supplemented as above.

We created stably expressing SNAP-EGFR HeLa cells by transfecting HeLa cells in a 60 cm² well of 40%-confluent cells using 9 μg of SNAP-EGFR plasmid DNA plus 20 μL lipofectamine LTX and 9 μL Plus reagent (Invitrogen, Carlsbad, CA) in 15 mL penicillin-streptomycin-free cell medium, as described in the suppliers' protocol. Selection (1400 μg/mL of active G418) was applied after 24 h. After 5 days, the cells were split over two six-well plates. After 10 days, the wells were screened for expression of SNAP-EGFR by labeling with 500 nM BG-Alexa 546 for 15 min and fluorescence microscopy analysis on the single-molecule sensitive microscope, as described later. Note that the expression level can be very low at this stage, and the imaging required a single-molecule-sensitive fluorescence microscope. One well contained positive cells with an expression level slightly above the single-molecule level; these cells were further cultured. For culturing, the concentration of active G418 was 350 μg/mL.

Sample preparation

For each dye, video recordings were taken of four samples: the SNAP-tag-positive cells and the three negative cell lines. Before measurements, HeLa cells stably expressing SNAP-EGFR, HeLa cells, MCF7 cells, and H441 cells were plated in Greiner Bio CellView dishes (product no. 627870) in full medium, and left overnight to allow the cells to adhere to the glass. The HeLa cells stably expressing SNAP-EGFR were also starved in fetal-bovine-serum-free medium and left for another night to reduce the activity and internalization of the EGFR fusion protein. On the day of the experiment, cells were washed with starvation medium containing 0.5% bovine serum albumin. Labeling of the SNAP-EGFR proteins was carried out thereafter by incubating the cells for 2 min (± 10 s) with 400 nM of each BG dye in starvation medium containing 0.5% bovine serum albumin. Samples were washed immediately by replacing the labeling solution with phosphate-buffered saline supplemented with magnesium and calcium. This washing step was repeated at least three times. Incubation and washing of the SNAP-tag negative cells with the substrates was performed using the same conditions.

Microscopy settings

The microscope hardware is described in the Supporting Material. Measurements to determine nonspecific binding of the SNAP substrates were performed using widefield and TIRF illumination. Measurements to determine the photobleaching of the substrates were performed using widefield illumination. The illumination time differed for each fluorescent substrate, and was chosen in such a way that single molecules were clearly visible over the autofluorescence background of the cell. We sought to collect the same average number of photons per molecule in a frame for each fluorescent substrate. Since the quantum yield has not been previously determined for all fluorescent substrates, the illumination time to yield an equal number of emitted photons per molecule could not be calculated beforehand. Videos were recorded at 20–30 fps, which is the highest allowed frame rate of the camera at the maximum readout rate of 10 MHz and frame size of 512×512 pixels. Each video recording consisted of 800 frames. Before recording each video, a minimal number of frames (~10–30 frames) were used to focus on the basal membrane of the cell.

Single-molecule brightness

To allow conversion of pixel counts to photons, a calibration of the gain of the two EMCCD cameras was performed by the mean-variance method (Fig. S1 in the Supporting Material). The slope of the line in this curve is equal to the inverse gain of the camera. A gain of 49.9 ± 0.1 was found for the camera recording the green-excitable dyes, and a gain of 32.5 ± 0.1 was found for the camera recording the red-excitable dyes. The pixel intensities in counts are divided by the camera gain to convert the pixel intensities to photons. The brightness of one molecule (sometimes also termed spot intensity) was calculated as the integrated intensity of a single molecule using a Gaussian fit performed by the tracking algorithm used (66). This yields for all single molecules the number of detected photons per single molecule per frame. We defined the single-molecule brightness, B , as the average of these numbers in one recording.

The brightness of the dye conjugates can be compared between dyes by a relative brightness (Table S1), which is a normalized value given by the single-molecule brightness, B (Fig. S2), divided by the acquisition time, the excitation efficiency η_e , the emission efficiency, and the laser excitation power. The excitation efficiency, η_e , is equal to the fraction of the maximum value of the excitation spectrum of the dye at the wavelength of the lasers, i.e., 532 nm for the green dyes and 637 nm for the red dyes. The spectra of the dyes were downloaded from the SemRock website (<http://www.semrock.com>), except for the CF dyes, TF5, and Star635; we measured the spectra for these dyes with a Varian Cary Eclipse fluorescence spectrophotometer (Palo Alto, CA).

Tracking of single molecules

To obtain trajectories from the raw videos, we used previously described tracking software (11,66). The settings used for the cost matrices in this software can be found in the Supporting Material. For the initial detection of molecules, the tracking algorithm uses an intensity threshold. This threshold was taken as the same for all video series of one fluorescent substrate to obtain a fair comparison of the level of specific and nonspecific labeling. The threshold was determined in the situation where the substrates are incubated with SNAP-tag-expressing cells (specifically attached); we used the same threshold values in the detection of nonspecifically attached substrates.

After obtaining the single-molecule trajectories, two filtering operations were applied with the purpose of discarding very short tracks, and differentiating between completely immobile and (transiently) mobile molecules. Very short tracks (having fewer than seven localizations in total) were excluded, as they did not contain much significant information; there is also a higher chance that a fluorescent spot that is detected only for a few frames was attributable to a noise rather than to a specifically labeled fluores-

cent molecule. A segment of a track was defined as the subsequent positions of a fluorescent molecule in adjacent frames. This meant that blinking of a dye resulted in multiple segments within a track. Immobile tracks were discarded because they often represented dye molecules bound to the glass surface; they were detected using a radius-of-gyration algorithm (67). The threshold for the trajectory area was defined by a gyration radius of 40 nm, as this corresponded to the apparent area traveled by an immobile molecule due to the localization accuracy.

Analysis of single-molecule photobleaching rate

The number of fluorescent molecules, $N(i)$, in each frame i was determined for each recording. Since photobleaching follows an exponential decay profile, the photobleaching rates are obtained for each video recording by fitting the number of molecules over time with a one-component exponential function without offset:

$$N(i) = N(1) \times \exp(-1/\tau \times i), \quad (1)$$

where i is the frame number, τ is the mean photobleaching time (in frames), hence $1/\tau$ is the rate of photobleaching per frame, and $N(1)$ is the fitted number of molecules in the first frame ($i = 1$). The fit was performed over frame numbers 20–600. In the first few frames, the autofluorescence of the cells might obscure a proper detection of single molecules by the algorithm. Because the autofluorescence bleaches rapidly, the fluorescent molecules can be reliably detected after 20 frames. At frame 600, the number of molecules was reduced to a basal level in most recordings. The fluorophore's mean photobleaching time, τ , is multiplied by the single-molecule brightness, B . This yields the expected average number of detected photons per molecule, P .

Since the dye conjugates have different emission spectra, we corrected for the transmission efficiency of the filter set to obtain a precise comparison of the dyes. The most relevant parameter to compare is the photobleaching rate per emitted photon and not per detected photon. This is because not all the emitted photons pass the filters placed before the camera. Not all of the emitted photons are collected by the objective, but the fraction of photons collected is the same for all the dyes, and it is therefore not necessary to correct for this. Furthermore, the quantum efficiency of the CCD chip is similar around the measured wavelengths. Therefore, the photobleaching rates were only corrected for the efficiency of the filter set, η_f , which describes the efficiency with which the emitted fluorescence passed the filter set used. The expected number of detectable photons, P_{corr} is given by the expected detected number of photons, P , divided by the detection efficiency, η_f . The detection efficiency, η_f , of a dye was determined by integration over the combined transmission spectrum of the dichroic mirrors and the emission filter multiplied by the normalized emission spectrum of the dye. This efficiency is listed for each dye in Table S1. The expected number of detectable photons per molecule, P_{corr} , was calculated as

$$P_{corr} = \frac{1}{\eta_f} \times B \times \tau. \quad (2)$$

RESULTS

Nonspecific binding of the SNAP substrates

We first screened the dyes to assess the level of nonspecific staining of the dye conjugates in cells not expressing the SNAP-tag fusion protein (SNAP-tag-negative cells). We excluded from further analyses substrates leading to high nonspecific staining of intracellular structures. The microscopy video recordings of H441 cells incubated with each dye conjugate are shown in Fig. 2 (widefield illumination)

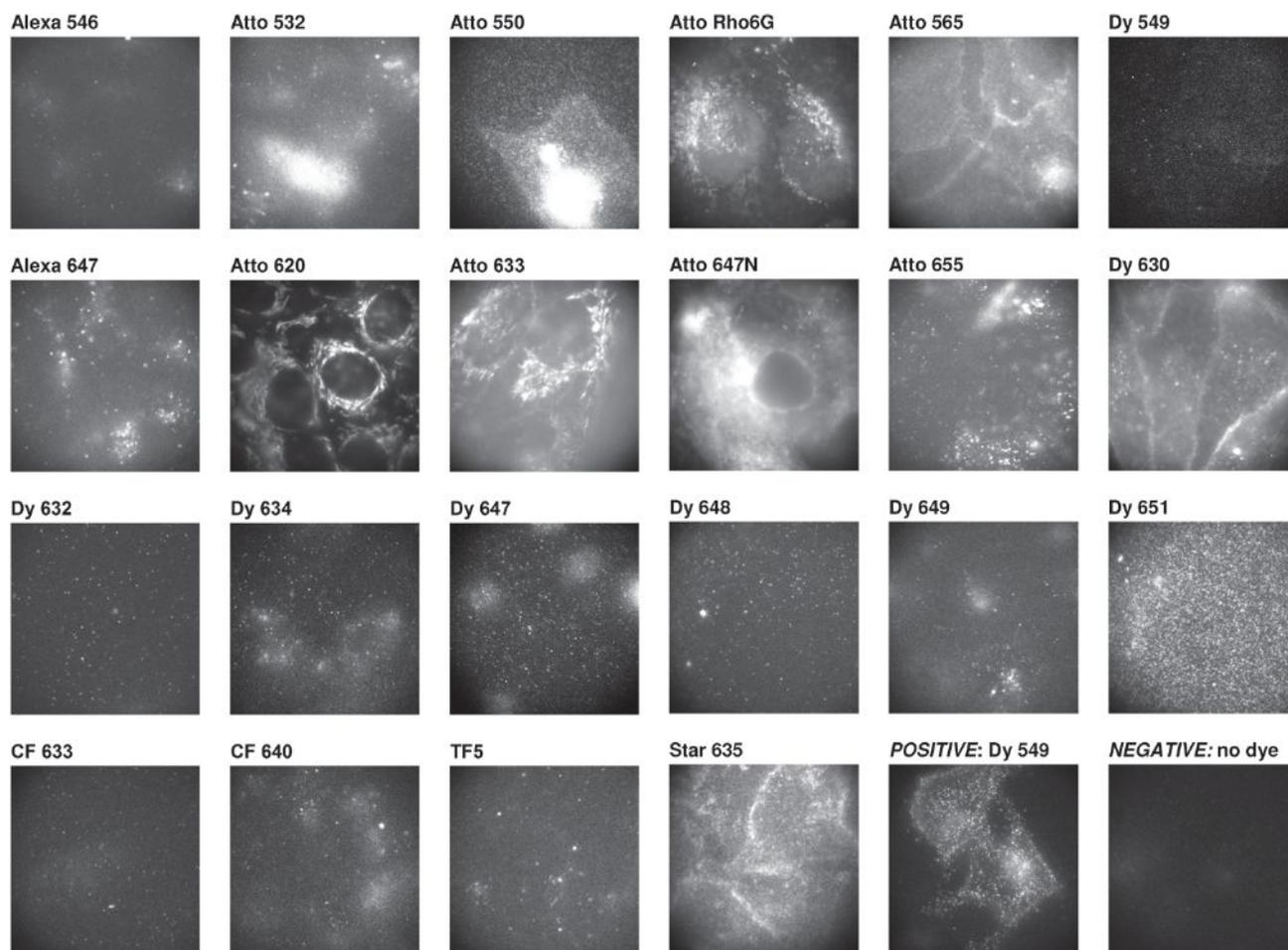


FIGURE 2 Fluorescence images of SNAP-tag-negative cells incubated with SNAP-tag fluorescent substrates. Incubation of the fluorescent substrates with SNAP-tag-negative cells reveals large differences in nonspecific binding to cellular components or the glass surface. An image showing the staining on SNAP-tag-positive cells is included for comparison. The images are recorded in widefield mode on a single-molecule-sensitive microscope. The field of view was completely confluent with cells. The size of the images is $61 \times 61 \mu\text{m}$. The photon intensity scale has not been determined, and varies between images.

and in Fig. S3 (TIRF illumination). In all the images, there was full confluence of cells in the field of view. Although TIRF images are often preferred over widefield images to record receptor proteins due to the reduced background level, likewise, only nonspecific staining near and at the plasma membrane of the cells will be observed with TIRF imaging. Nonspecifically stained intracellular structures were better observed using widefield imaging, and were used for screening of nonspecific staining. The screening for nonspecific binding was based on observations of at least 50 cells per sample, and resulted in the exclusion of the following substrates: Atto-550, Atto-565, Atto-620, Atto-633, Atto-647N, Dy-630, Dy-651, and Star-635.

The fluorescent substrates tested exhibited similar levels of nonspecific staining in HeLa cells (data not shown). The nonspecific staining observed did not appear to substantially vary among cells in the same sample, or between samples prepared on different days. Dead cells usually showed much more nonspecific staining than healthy cells.

The remaining dye conjugates were incubated with SNAP-tag-negative HeLa, MCF7, and H441 cells. Microscopy recordings were taken for each SNAP substrate in the different cell lines, with the focus of the microscope at the basal membrane of the cells. The tracking algorithm provided the number of detected molecules in each frame. For each substrate, the camera acquisition time used was the same as that used in the recordings with the SNAP-tag-positive cells (see Table S1). This ensured that the number of detectable nonspecific molecules was determined under the same imaging and tracking conditions as for the imaging of specifically bound molecules. Next we differentiated completely immobile molecules from (transiently) mobile molecules. Completely immobile molecules are often molecules bound to the glass substrate; these are typically of less concern, since they can usually be readily excluded before further analysis. In contrast, nonspecific mobile molecules obscure the analysis of the specifically labeled protein molecules.

The average number of mobile molecules, as well as the total number of molecules (mobile and immobile), detected in frame numbers 20–40 are shown in Fig. 3. The first 20 frames were excluded because the autofluorescence of the cells is then particularly high, which obscures the specific detection of labeled proteins. Only regions with full confluence of cells were recorded. The total number of nonspecific molecules per field of view is a measure of the expected number of molecules adsorbed on the glass substrate (underneath the cells), the immobile molecules, plus the number of false-positive molecules on the plasma membrane, the (transiently) mobile molecules. The number of nonspecific molecules in frame numbers 120–140 is also shown to gain insight into the photobleaching of nonspecifically bound substrates.

Photostability of the substrates bound to SNAP-tag

To determine the photostability of the dyes bound to SNAP-tag, we incubated them with cells expressing the SNAP-EGFR fusion protein. Microscopy recordings were taken for each dye conjugate to determine the photobleaching rate of the dyes bound to SNAP-tag. To avoid variance between cells of different samples as a result of transfection, we used a stably transfected HeLa cell line, which had low expression levels of SNAP-EGFR (single-molecule density). We optimized the incubation concentration and time for high labeling efficiency and low nonspecific binding using a titration series with BG-Alexa 546, and found that 2 min incubation with 400 nM of substrate was enough for a complete labeling with this dye (and also used these incubation conditions in the nonspecific binding assay). Note that incubation using elevated dye concentrations or prolonged incubation time might result in higher nonspecific binding levels. For each dye conjugate, we observed a similar percentage of labeled cells (estimated to be 15%) irrespective of the specific dye choice. We believe that this percentage of labeled cells was caused by a large population of cells that do not express the SNAP-tag. The fraction of SNAP-tag receptors labeled in cells appeared to vary slightly from dye to dye.

For an accurate comparison, we aimed to obtain the same number of detected photons per frame (single-molecule brightness, B) for all green and all red dyes. All the dyes were bright enough to be detected at a single-molecule level in a widefield setup in the presence of cellular autofluorescence background. A widefield setup is more appropriate than a TIRF setup for an accurate comparison as the single-molecule brightness, B , is very difficult to control in TIRF due to varying TIRF angles and the presence of molecules at different depths. Furthermore, the expected number of photons emitted from a fluorophore does not depend on the type of illumination. For the characterization procedure followed, we found that optimal single-molecule

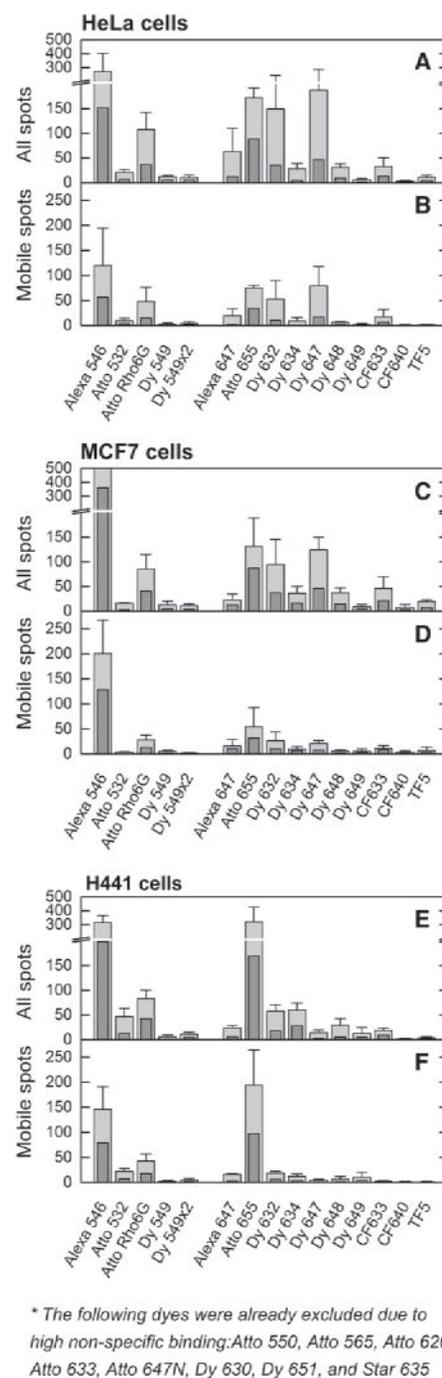


FIGURE 3 Quantification of the nonspecific binding of SNAP-tag fluorescent substrates in live cells. The values show the number of mobile fluorescent substrates and the total number of fluorescent substrates (mobile and immobile) that were nonspecifically bound to cells. The values were determined per field of view area in HeLa, MCF7, and H441 cell lines for each dye showing nonspecific binding on the single-molecule level. Some dyes had extremely high levels of nonspecific binding, and since no individual spots could be detected, these were excluded from this graph. Shown are the average number of molecules detected in frame numbers 20–40 (light gray) and 120–140 (dark gray). The field of view is a circular area ($1520 \mu\text{m}^2$) with a radius of $\sim 22 \mu\text{m}$. The values were determined in multiple recordings, and the average number is shown here, with the error representing the sample standard deviation.

brightness was $B = 150$ photons for red-excitable dyes and $B = 200$ photons for green-excitable dyes. Some dyes needed a relatively long acquisition time to obtain the targeted single-molecule brightness, B (see Table S1 for the acquisition times used and Fig. S2 for the resulting single-molecule brightness, B). Table S1 also lists the relative brightness of each dye conjugate to SNAP-tag. At least four movies of different cells per dye conjugate were recorded and analyzed (Fig. 4 A and Movie S1). The brightness (spot intensity) of the molecules follows a Poisson-like distribution, as shown in Fig. 4 B.

Due to photobleaching, the number of observed fluorescent molecules, $N(i)$, decreased over time (Fig. 4 C). We fitted the rate of photobleaching using Eq. 1 to extract the mean photobleaching time, τ , for each fluorophore. Using Eq. 2, the expected number of detectable photons per molecule, P_{corr} , was calculated. A basal level of detected molecules was observed even after a long imaging time. We believe that these remaining molecules are the result of molecules in an intermittent state (blinking) and a constant influx of molecules from out-of-focus areas into focus. The expected number of detectable photons per molecule, P_{corr} , was obtained from multiple recordings per fluorescent substrate, and the average value and standard deviation are shown in Fig. 5. The conversion from numbers of molecules to photons requires that single molecules be detected. This was checked by confirming that the number of emitted photons per molecule does not vary over time (Fig. 4 D).

In Fig. 6, we summarize the results for nonspecific binding versus the photostability for each dye. From this figure, it is clear that both green- (e.g., Dy 549) and red-excitable dyes (e.g., CF633 and CF640) are suitable for single-

molecule tracking. This result allowed us to examine the possibility of simultaneously labeling the SNAP-tag with two spectrally different dyes. The simultaneous incubation of a 1.0:0.67 mixture of BG-CF633 and BG-Dy 549 resulted in roughly equal labeling of the SNAP-tag receptor with these two dyes (Fig. 7). Movie S2 shows SNAP-EGFR proteins labeled with these two dyes diffusing in the plasma membrane of a live cell. The disappearance of receptors is due to photobleaching.

DISCUSSION

The results show that a careful choice of the dye to conjugate to the SNAP-substrate to label SNAP-tag fusion proteins is very important, as many fluorescent substrates suffer from either rapid photobleaching or high nonspecific staining. We found that of the 22 fluorescent substrates tested, three can be used for single-molecule tracking applications, as these substrates combine both a low level of nonspecific binding and a high photostability. Among the green-excitable fluorescent substrates, BG-Dy 549 showed the highest photostability with the lowest nonspecific staining (Fig. 6). As an alternative, BG-Alexa 546 could be used in ensemble measurements (e.g., FRET studies), as it is photostable and only results in detectable nonspecific binding at the single-molecule level. Among the red-excitable fluorescent substrates, BG-CF640 and BG-CF633 exhibited the best results (Fig. 6). Whereas BG-CF640 showed slightly lower nonspecific staining, CF633 might be relatively brighter depending on the filter sets available. Even though BG-Atto 655 showed the highest photostability among the substrates tested (Fig. 5), its use is limited to ensemble measurements,

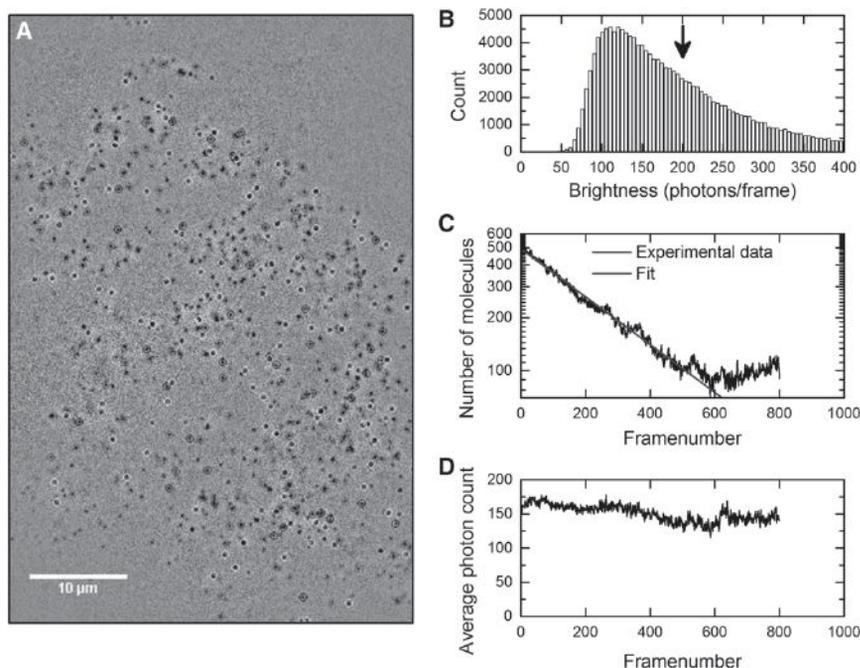


FIGURE 4 Example of the performed photobleaching analysis on one video recording. A fluorescence image series of SNAP-EGFR labeled with a BG-Dy 549 was recorded. (A) The tracking algorithm finds the molecules in the raw microscopy recording, and after exclusion of immobile molecules and very short trajectories, the detected molecules are encircled in the microscopy recording, where colors are used to differentiate tracks; see also Movie S1. (B) Histogram of the number of detected photons per frame of all the found molecules (brightness or spot intensity). The arrow indicates the average of the values, which we defined as the single-molecule brightness, B . (C) Number of detected molecules per frame, $N(i)$, as a function of frame number i . In red, a fit of the data using a single-exponential decay function according to Eq. 1 to yield the mean photobleaching time, τ , for each fluorophore. (D) The average brightness of the molecules in one frame does not change over time, confirming that we indeed looked at single molecules. To see this figure in color, go online.

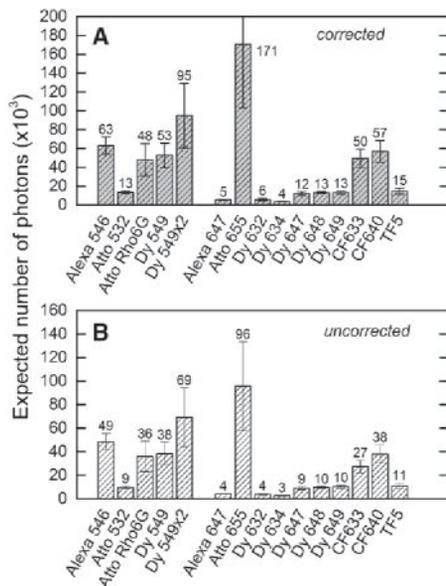


FIGURE 5 Expected number of detectable photons per molecule, P_{corr} , for each SNAP-tag fluorescent substrate. The expected number of photons provides a value for the photostability of a dye conjugate. The values were determined in multiple recordings, and the average number is shown here, with the error representing the sample standard deviation. (A) Values are corrected for the detection efficiency of the microscope for each dye. (B) Values are not corrected for the detection efficiency, and represent the expected number of photons detected in our setup.

since it showed high levels of nonspecific binding to membrane components of all the three cell lines (Fig. 3).

Nonspecific binding of the SNAP substrates

One of the main advantages of single-molecule tracking techniques is the ability to discriminate single mobile mol-

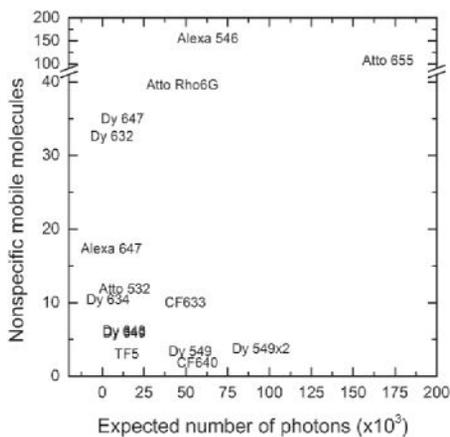


FIGURE 6 Comparison of the performance of the SNAP-tag fluorescent substrates for use in single-molecule tracking microscopy. The performance is shown in terms of photostability and nonspecific binding. BG-Dy 648 and BG-Dy 649 overlap in the graph. Fluorescent substrates in the lower right corner show little nonspecific attachment to cells, and the most emitted photons per molecule before photobleaching. These substrates are the preferred choice for single-molecule tracking microscopy.

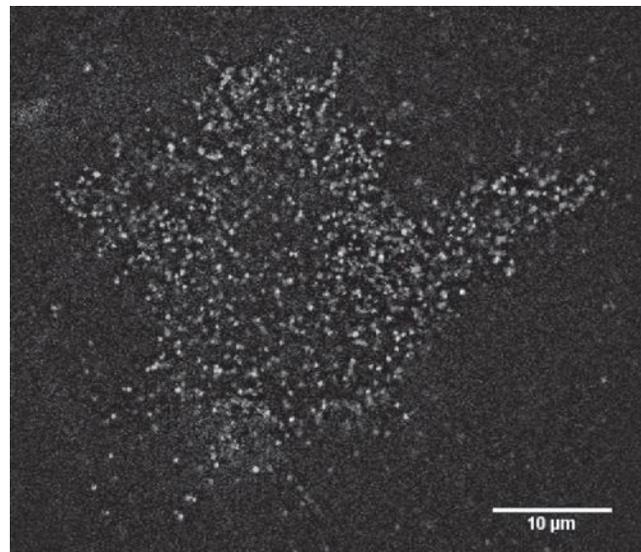


FIGURE 7 A TIRF image demonstrating dual-color labeling of SNAP-tag receptors at the single-molecule level. The labeling was performed on SNAP-EGFR with BG-Dy549 (green) and BG-CF633 (red). The combination of relatively photostable dye conjugates with little nonspecific staining allows multicolor single-molecule tracking microscopy. Using this technique, receptor homodimers can be directly visualized. See also Movie S2. To see this figure in color, go online.

ecules from cellular autofluorescence, immobile fluorescent molecules, and clusters of fluorescent molecules. We utilized this to characterize the nonspecific binding of the fluorescent substrates. Dealing with the nonspecific binding of fluorophores to any cellular components is one of the biggest challenges in microscopy. Several of the BG dyes tested showed high levels of nonspecific binding (Fig. 2). We found that the amount of nonspecific binding of the BG dyes is roughly the same among the cell lines tested (Fig. 3). None of the dyes that led to appreciable levels of nonspecific staining photobleached within a short period (Fig. 3); hence, differences in photobleaching of specifically and nonspecifically bound dyes cannot be used advantageously to discriminate between the two cases.

The cause of nonspecific binding might be explained from a molecular perspective. Several dyes contain long-chain hydrocarbons that are lipophilic; therefore, they easily incorporate into lipid-rich structures such as cellular membranes. Sulfonate acid groups are often added to dyes to enhance their solubility in water. These groups are negatively charged and electrostatically repelled away from the negatively charged lipid headgroups in cellular membranes. Negatively charged dyes include sulfonated fluorescein- and cyanine-based dyes (68). On the other hand, cationic (positively charged) dyes, such as many rhodamine-based dyes have been reported to bind to mitochondria in live cells (69). Therefore, the major factors influencing nonspecific binding might be the lipophilic character of a dye in combination with localized electronic charges. Furthermore, the inability of certain dyes to penetrate the plasma membrane

hinders access to intracellular structures. In general, neutral and anionic (negatively charged) dyes in this survey appeared to have less of a tendency to bind to cellular substructures (e.g., Alexa 546/647, Dy 632/634, Dy 648/649). Some dyes (e.g., Atto 647N and Dy 651) adhered to a large extent to the glass coverslip (which may be avoided by optimizing cleaning procedures), obscuring the detection of specifically bound single molecules in the adjacent plasma membrane of the cell. The complex effects of local charges in combination with polar and lipophilic groups in a dye molecule make it difficult to predict the nonspecific binding ability of dyes beforehand. For example, the net charge of a molecule does not completely explain the nonspecific interaction, such as for the negatively charged BG-Dy 651 and the neutrally charged BG-Dy 630. Both showed a considerable amount of nonspecific binding to cellular components. We also did not find a correlation between the chromophore family and the nonspecific labeling level. For example, the incubation of cells with the rhodamine-derived dyes BG-Alexa 546 and BG-Atto 532 resulted in low nonspecific levels, whereas BG-Atto 550 and BG-Atto 565 led to much higher nonspecific levels. Likewise, the cyanine-based BG-Alexa 647 showed almost no nonspecific binding, whereas the BG-Dy 630 exhibited extremely high nonspecific binding.

Benke et al. have reported the use of five BG dyes for single-molecule tracking in eukaryotic cells (61). In their approach, the fluorescence of these dyes was stochastically activated for superresolution microscopy; however, no data on nonspecific binding was provided. Stöhr et al. described the quenching of several dyes after conjugation to BG and subsequent SNAP-tag binding (63,70). Their data demonstrate that the photophysics (i.e., the photobleaching time and fluorescence quenching) of a given dye can be altered by its molecular environment. Furthermore, they conclude that it is impossible to predict the changes in fluorescence beforehand due to the complex effects of local charges in the dye molecule. Stöhr et al. also reported on the background levels of remaining unreacted dyes inside *Escherichia coli* after washing protocols. Interestingly, some substrates, such as BG-Atto 620 and BG-Atto 633, which reportedly exhibited a low background staining in *E. coli*, led to a surprisingly high nonspecific binding in our experiments with mammalian cell lines. Stöhr et al. also reported the labeling of 3T3 fibroblast cells with BG-Atto 550, BG-Atto 633, and BG-Atto 647N. In a similar way, we noticed that BG-Atto 550 and BG-Atto 647N produced high levels of nonspecific binding. However, in contrast to the results of Stöhr et al., in our case, BG-Atto 633 showed a very intense nonspecific staining of cytosolic and membrane structures (Fig. 2 and Fig. S2). This discrepancy could be caused by the difference in fluorescence intensity levels between the two studies, as we looked at nonspecific staining in the context of single molecules.

Photostability of the substrates bound to SNAP-tag

Whereas many red-excitable dye conjugates did not show any substantial nonspecific binding levels, these dyes appeared to be less photostable than the green-excitable dyes. Two dyes, CF633 and CF640, are photostable enough to permit prolonged imaging with low nonspecific staining (Fig. 6). Between these two dyes, CF640 showed marginally less nonspecific staining (Fig. 3).

Another noteworthy observation was that the photostabilities of the Dy dyes of relatively close excitation wavelengths were very similar (Fig. 5), for instance, those of Dy 647, Dy 648, and Dy 649, as well as Dy 632 and Dy 634. From a molecular perspective, Dy 647, Dy 648, and Dy 649 are typical cyanine dyes, whereas Dy 632 and 634 have one indole group with a polymethine chain linked to a benzopyrylium group. The slight differences in these chromophores did not seem to have a large effect on its photobleaching rate.

Complications and validity

We have performed the photobleaching experiments on SNAP-tag fused transmembrane EGFR proteins, which have a basal internalization rate even when the cells are starved (71,72). This might lead to a false enhanced bleaching detection. During the 30 s of imaging, however, the internalization rate of the receptor is small compared to the photobleaching rate (73). Even 1 h after the labeling, no significant decrease of receptor molecules was observed at the plasma membrane of the cells. However, in some instances, a small increase in fluorescence in the cytosol was noticeable, which could be attributed to the basal level of receptor internalization and the recycling process. Another complication stems from the fact that this receptor seems to localize more in filopodia and the periphery of the cells (74); these receptors can diffuse more easily in and out of focus. Because molecules diffusing in and out of the plane of focus will likely be in equilibrium, this should not influence the recorded bleaching rate at the beginning of a recording, when receptors in focus have not been bleached yet. Later, however, as bleached receptors leave the plane, unbleached receptors can enter the focal plane from outside the plane, causing the bleaching rate to appear slower than it actually is. Therefore, we derived the bleaching rate from that part of the recordings where the number of molecules is still decreasing. Furthermore, the rate of receptors entering the focal plane within the 30 s of imaging will be limited, and this rate will be independent of the dye used.

Improvements to fluorescent SNAP substrates

The attachment of two Dy 549 dyes on a single SNAP substrate (Dy 549 \times 2) seems to be an interesting approach to

prolonging imaging of the protein, as its photostability almost doubled in comparison to the substrate with single-loaded Dy 549 (Fig. 5). The brightness of the double-loaded SNAP substrate was similar to that of the single-loaded substrate (Table S1). This might be due to self-quenching, which is commonly observed when the number of fluorophores on a protein is increased, and which affects the fluorescence intensity but not the photobleaching rate per photon for the complex. Further studies are needed to confirm that the (single-molecule) brightness is indeed similar in SNAP substrates with one, two, or even more Dy 549 fluorophores.

Another interesting approach is the incorporation of a strong fluorescence quencher in the guanine group. Such a fluorogenic method ensures that the benzylguanine coupled fluorophore becomes dramatically more fluorescent upon binding to the SNAP-tag (40). Although the guanine itself already acts as a relatively good quencher for several dyes (63), the more dramatic fluorogenic approach could bypass the issue of nonspecific binding for extremely photostable dyes such as Atto-655. Another interesting idea is to use a SNAP-tag substrate derivatized with a fluorophore and a triplet-state quencher (e.g., a molecular oxygen reducing agent) (75). This strategy has led to an overall decrease in the number of dark-state transitions, which led to imaging periods up to 25-fold longer (75). Prolonged imaging may allow observation and tracking of many more interactions of the protein on its path through the cell.

CONCLUSIONS

We have screened and analyzed the photostability and nonspecific binding properties of a wide range of green- and red-excitable dyes for labeling proteins in cells by means of the SNAP-tag technology. The SNAP-tag labeling strategy is particularly useful for labeling proteins on the plasma membrane, since there are no restrictions on the membrane permeability of the fluorescent label. Properties of dyes have generally been determined in ensemble fluorescence imaging and on relatively large biomolecules such as antibodies. However, properties of dyes can be rather different at the single-molecule level and when conjugated to a small biomolecule, such as the SNAP substrate (BG), and in the local microenvironment of the SNAP-tag. We have characterized the photostability and specificity for several SNAP-substrate dye conjugates in different cell lines at the single-molecule level. We performed the characterization in widefield mode to prevent illumination variations experienced in a TIRF setup, and at high single-molecule brightness to adequately count most dye molecules in the recordings. To provide a meaningful comparison, we used similar photon counts per single molecule for each spectrally similar dye, corrected for the detection efficiency of our microscope for the dye's emission spectrum, and tracked the bound dyes to differentiate the motion of the nonspecifically bound molecules.

We found that in our system, the SNAP substrates labeled with Dy 549, CF633, and CF 640 are the best choices to label SNAP-tag fusion proteins for single-molecule tracking among the fluorescent substrates tested. Also, we show that the attachment of two Dy-549 dyes on one BG probe is an interesting approach for prolonging imaging of the protein. Finding two spectrally different SNAP-tag-labeling dyes that were suitable for single-molecule imaging proved to be a challenge, as most of the fluorescent substrates tested either showed a large amount of nonspecific fluorescence or were rapidly photobleached.

Since both green- and red-excitable fluorescent SNAP substrates have been identified, multicolor single-molecule imaging of the same protein species can become a routine experiment by simultaneously incubating these substrates with the SNAP-tag fusion proteins in live cells. This should allow direct observation of homodimers. For an extension to three-color single-molecule imaging, BG-Alexa 488 could be used as the third dye conjugate, since it is already known to be a suitable dye for single-molecule tracking (61), although the intense cellular autofluorescence at this excitation wavelength limits its use to TIRF microscopy. In addition, we anticipate that our conclusions could be applied to the chemically similar tagging technology CLIP-tag, which also has the guanine moiety in its substrate. Our results are probably not directly translatable to chemically different molecular tags, such as Halo Tag, or the acyl carrier protein based ACP and MCP tags. The combination of SNAP-tag with another molecular labeling tag allows orthogonal labeling on two different protein species. Thus, an interesting extension to single-protein-species studies is the direct visualization of two interacting proteins of different species, as occurs, for example, in heterodimer formation.

SUPPORTING MATERIAL

Three figures, two tables, two movies, and Supporting Methods are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)00686-9](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00686-9).

We thank Yvonne Kraan for assistance in cell culturing and sample preparations; and we thank Peter Relich and Keith Lidke of the University of New Mexico for sharing their tracking software with us.

PB and MS were supported by an ERA-NET NanoSci E+ grant through Stichting Technische Wetenschappen grant 11022-NanoActuate. JI is supported by the same ERANET NanoSci-E+ grant through DFG grant VE 579/1-1 and by the German Research Foundation via DFG grant VE 579/3-1.

REFERENCES

1. Kusumi, A., C. Nakada, ..., T. Fujiwara. 2005. Paradigm shift of the plasma membrane concept from the two-dimensional continuum fluid to the partitioned fluid: high-speed single-molecule tracking of membrane molecules. *Annu. Rev. Biophys. Biomol. Struct.* 34:351–378.
2. Xia, T., N. Li, and X. Fang. 2013. Single-molecule fluorescence imaging in living cells. *Annu. Rev. Phys. Chem.* 64:459–480.

3. Joo, C., H. Balci, ..., T. Ha. 2008. Advances in single-molecule fluorescence methods for molecular biology. *Annu. Rev. Biochem.* 77:51–76.
4. Hinterdorfer, P., and A. van Oijen. 2009. Handbook of Single-Molecule Biophysics. Springer, New York.
5. Joo, C., M. Fareh, and V. N. Kim. 2013. Bringing single-molecule spectroscopy to macromolecular protein complexes. *Trends Biochem. Sci.* 38:30–37.
6. Lippincott-Schwartz, J., E. Snapp, and A. Kenworthy. 2001. Studying protein dynamics in living cells. *Nat. Rev. Mol. Cell Biol.* 2:444–456.
7. Miyawaki, A. 2011. Proteins on the move: insights gained from fluorescent protein technologies. *Nat. Rev. Mol. Cell Biol.* 12:656–668.
8. Giepmans, B. N. G., S. R. Adams, ..., R. Y. Tsien. 2006. The fluorescent toolbox for assessing protein location and function. *Science.* 312:217–224.
9. Pantoja, R., E. A. Rodriguez, ..., H. A. Lester. 2009. Single-molecule imaging of a fluorescent unnatural amino acid incorporated into nicotinic receptors. *Biophys. J.* 96:226–237.
10. Callegari, A., S. Luin, ..., F. Beltram. 2012. Single particle tracking of acyl carrier protein (ACP)-tagged TrkA receptors in PC12nnr5 cells. *J. Neurosci. Methods.* 204:82–86.
11. Low-Nam, S. T., K. A. Lidke, ..., D. S. Lidke. 2011. ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding. *Nat. Struct. Mol. Biol.* 18:1244–1249.
12. Goulian, M., and S. M. Simon. 2000. Tracking single proteins within cells. *Biophys. J.* 79:2188–2198.
13. Sako, Y., S. Minoghchi, and T. Yanagida. 2000. Single-molecule imaging of EGFR signalling on the surface of living cells. *Nat. Cell Biol.* 2:168–172.
14. Wieser, S., and G. J. Schütz. 2008. Tracking single molecules in the live cell plasma membrane: Do's and Don't's. *Methods.* 46:131–140.
15. Sergé, A., N. Bertaux, ..., D. Marguet. 2008. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods.* 5:687–694.
16. Kholodenko, B. N., J. F. Hancock, and W. Kolch. 2010. Signalling ballet in space and time. *Nat. Rev. Mol. Cell Biol.* 11:414–426.
17. Cambi, A., and D. S. Lidke. 2012. Nanoscale membrane organization: where biochemistry meets advanced microscopy. *ACS Chem. Biol.* 7:139–149.
18. Saxton, M. J., and K. Jacobson. 1997. Single-particle tracking: applications to membrane dynamics. *Annu. Rev. Biophys. Biomol. Struct.* 26:373–399.
19. Kapanidis, A. N., and T. Strick. 2009. Biology, one molecule at a time. *Trends Biochem. Sci.* 34:234–243.
20. Schütz, G. J., H. Schindler, and T. Schmidt. 1997. Single-molecule microscopy on model membranes reveals anomalous diffusion. *Biophys. J.* 73:1073–1080.
21. Bramshuber, M., and G. J. Schütz. 2012. In Vivo Tracking of Single Biomolecules: What Trajectories Tell Us About the Acting Forces. Springer, Berlin/Heidelberg.
22. Dietrich, C., B. Yang, ..., K. Jacobson. 2002. Relationship of lipid rafts to transient confinement zones detected by single particle tracking. *Biophys. J.* 82:274–284.
23. de Keijzer, S., A. Sergé, ..., B. E. Snaar-Jagalska. 2008. A spatially restricted increase in receptor mobility is involved in directional sensing during *Dictyostelium discoideum* chemotaxis. *J. Cell Sci.* 121:1750–1757.
24. Cutler, P. J., M. D. Malik, ..., K. A. Lidke. 2013. Multi-color quantum dot tracking using a high-speed hyperspectral line-scanning microscope. *PLoS ONE.* 8:e64320.
25. Cebecauer, M., M. Spitaler, ..., A. I. Magee. 2010. Signalling complexes and clusters: functional advantages and methodological hurdles. *J. Cell Sci.* 123:309–320.
26. Harding, A. S., and J. F. Hancock. 2008. Using plasma membrane nanoclusters to build better signaling circuits. *Trends Cell Biol.* 18:364–371.
27. Radhakrishnan, K., Á. Halász, ..., B. S. Wilson. 2012. Mathematical simulation of membrane protein clustering for efficient signal transduction. *Ann. Biomed. Eng.* 40:2307–2318.
28. Shaner, N. C., P. A. Steinbach, and R. Y. Tsien. 2005. A guide to choosing fluorescent proteins. *Nat. Methods.* 2:905–909.
29. Filonov, G. S., K. D. Piatkevich, ..., V. V. Verkhusha. 2011. Bright and stable near-infrared fluorescent protein for in vivo imaging. *Nat. Biotechnol.* 29:757–761.
30. Harms, G. S., L. Cognet, ..., T. Schmidt. 2001. Autofluorescent proteins in single-molecule research: applications to live cell imaging microscopy. *Biophys. J.* 80:2396–2408.
31. Wieser, S., M. Axmann, and G. J. Schütz. 2008. Versatile analysis of single-molecule tracking data by comprehensive testing against Monte Carlo simulations. *Biophys. J.* 95:5988–6001.
32. Howarth, M., W. Liu, ..., A. Y. Ting. 2008. Monovalent, reduced-size quantum dots for imaging receptors on living cells. *Nat. Methods.* 5:397–399.
33. Liu, H. Y., and X. Gao. 2011. Engineering monovalent quantum dot-antibody bioconjugates with a hybrid gel system. *Bioconjug. Chem.* 22:510–517.
34. Farlow, J., D. Seo, ..., Y.-W. Jun. 2013. Formation of targeted monovalent quantum dots by steric exclusion. *Nat. Methods.* 10:1203–1205.
35. Michalet, X., F. F. Pinaud, ..., S. Weiss. 2005. Quantum dots for live cells, in vivo imaging, and diagnostics. *Science.* 307:538–544.
36. Pinaud, F., S. Clarke, ..., M. Dahan. 2010. Probing cellular events, one quantum dot at a time. *Nat. Methods.* 7:275–285.
37. Keppler, A., S. Gendreizig, ..., K. Johnsson. 2003. A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* 21:86–89.
38. Keppler, A., H. Pick, ..., K. Johnsson. 2004. Labeling of fusion proteins with synthetic fluorophores in live cells. *Proc. Natl. Acad. Sci. USA.* 101:9955–9959.
39. Gautier, A., A. Juillerat, ..., K. Johnsson. 2008. An engineered protein tag for multiprotein labeling in living cells. *Chem. Biol.* 15:128–136.
40. Sun, X., A. Zhang, ..., I. R. Corrêa, Jr. 2011. Development of SNAP-tag fluorogenic probes for wash-free fluorescence imaging. *ChemBioChem.* 12:2217–2226.
41. Corrêa, Jr., I. R., B. Baker, ..., M.-Q. Xu. 2013. Substrates for improved live-cell fluorescence labeling of SNAP-tag. *Curr. Pharm. Des.* 19:5414–5420.
42. Pellett, P. A., X. Sun, ..., J. Bewersdorf. 2011. Two-color STED microscopy in living cells. *Biomed. Opt. Express.* 2:2364–2371.
43. Dempsey, G. T., J. C. Vaughan, ..., X. Zhuang. 2011. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat. Methods.* 8:1027–1036.
44. Klein, T., A. Löffelberger, ..., M. Sauer. 2011. Live-cell dSTORM with SNAP-tag fusion proteins. *Nat. Methods.* 8:7–9.
45. Foraker, A. B., S. M. Camus, ..., F. M. Brodsky. 2012. Clathrin promotes centrosome integrity in early mitosis through stabilization of centrosomal ch-TOG. *J. Cell Biol.* 198:591–605.
46. Bojkowska, K., F. Santoni de Sio, ..., D. Trono. 2011. Measuring in vivo protein half-life. *Chem. Biol.* 18:805–815.
47. Ville, D., J. DE Bellescize, ..., V. DES Portes. 2009. Ring 14 chromosome presenting as early-onset isolated partial epilepsy. *Dev. Med. Child Neurol.* 51:917–922.
48. Masharina, A., L. Reymond, ..., K. Johnsson. 2012. A fluorescent sensor for GABA and synthetic GABA(B) receptor ligands. *J. Am. Chem. Soc.* 134:19026–19034.
49. Chidley, C., H. Haruki, ..., K. Johnsson. 2011. A yeast-based screen reveals that sulfasalazine inhibits tetrahydrobiopterin biosynthesis. *Nat. Chem. Biol.* 7:375–383.
50. Hoskins, A. A., L. J. Friedman, ..., M. J. Moore. 2011. Ordered and dynamic assembly of single spliceosomes. *Science.* 331:1289–1295.

51. Breitsprecher, D., R. Jaiswal, ..., B. L. Goode. 2012. Rocket launcher mechanism of collaborative actin assembly defined by single-molecule imaging. *Science*. 336:1164–1168.
52. Jaiswal, R., D. Breitsprecher, ..., B. L. Goode. 2013. The formin Daam1 and fascin directly collaborate to promote filopodia formation. *Curr. Biol.* 23:1373–1379.
53. Smith, B. A., K. Daugherty-Clarke, ..., J. Gelles. 2013. Pathway of actin filament branch formation by Arp2/3 complex revealed by single-molecule imaging. *Proc. Natl. Acad. Sci. USA*. 110:1285–1290.
54. Smith, B. A., S. B. Padrick, ..., J. Gelles. 2013. Three-color single molecule imaging shows WASP detachment from Arp2/3 complex triggers actin filament branch formation. *eLife*. 2:e01008.
55. Peisley, A., M. H. Jo, ..., S. Hur. 2012. Kinetic mechanism for viral dsRNA length discrimination by MDA5 filaments. *Proc. Natl. Acad. Sci. USA*. 109:E3340–E3349.
56. Qiu, W., N. D. Derr, ..., S. L. Reck-Peterson. 2012. Dynein achieves processive motion using both stochastic and coordinated stepping. *Nat. Struct. Mol. Biol.* 19:193–200.
57. Derr, N. D., B. S. Goodman, ..., S. L. Reck-Peterson. 2012. Tug-of-war in motor protein ensembles revealed with a programmable DNA origami scaffold. *Science*. 338:662–665.
58. Numata, N., T. Shima, ..., K. Sutoh. 2011. C-sequence of the *Dictyostelium* cytoplasmic dynein participates in processivity modulation. *FEBS Lett.* 585:1185–1190.
59. Yang, Y., and C. Y. Zhang. 2013. Simultaneous measurement of SUMOylation using SNAP/CLIP-tag-mediated translation at the single-molecule level. *Angew. Chem. Int. Ed. Engl.* 52:691–694.
60. Calebiro, D., F. Rieken, ..., M. J. Lohse. 2013. Single-molecule analysis of fluorescently labeled G-protein-coupled receptors reveals complexes with distinct dynamics and organization. *Proc. Natl. Acad. Sci. USA*. 110:743–748.
61. Benke, A., N. Olivier, ..., S. Manley. 2012. Multicolor single molecule tracking of stochastically active synthetic dyes. *Nano Lett.* 12:2619–2624.
62. Schenk, A., S. Ivanchenko, ..., G. U. Nienhaus. 2004. Photodynamics of red fluorescent proteins studied by fluorescence correlation spectroscopy. *Biophys. J.* 86:384–394.
63. Stöhr, K., D. Siegberg, ..., D.-P. Herten. 2010. Quenched substrates for live-cell labeling of SNAP-tagged fusion proteins with improved fluorescent background. *Anal. Chem.* 82:8186–8193.
64. Wang, T. Y., L. J. Friedman, ..., V. W. Cornish. 2014. The covalent trimethoprim chemical tag facilitates single molecule imaging with organic fluorophores. *Biophys. J.* 106:272–278.
65. Lukinavičius, G., K. Umezawa, ..., K. Johnsson. 2013. A near-infrared fluorophore for live-cell super-resolution microscopy of cellular proteins. *Nat. Chem.* 5:132–139.
66. Smith, C. S., N. Joseph, ..., K. A. Lidke. 2010. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat. Methods*. 7:373–375.
67. Bosch, P. J., J. S. Kanger, and V. Subramaniam. 2014. Classification of dynamical diffusion states in single molecule tracking microscopy. *Biophys. J.* 107. <http://dx.doi.org/10.1016/j.bpj.2014.05.049>.
68. Panchuk-Voloshina, N., R. P. Haugland, ..., R. P. Haugland. 1999. Alexa dyes, a series of new fluorescent dyes that yield exceptionally bright, photostable conjugates. *J. Histochem. Cytochem.* 47:1179–1188.
69. Johnson, L. V., M. L. Walsh, ..., L. B. Chen. 1981. Monitoring of relative mitochondrial membrane potential in living cells by fluorescence microscopy. *J. Cell Biol.* 88:526–535.
70. Stöhr, K. 2008. Fluoreszenzgelöschte Sonder für dianostische und analytische Anwendungen, PhD thesis, Heidelberg University, Heidelberg, Germany.
71. Citri, A., and Y. Yarden. 2006. EGF-ERBB signalling: towards the systems level. *Nat. Rev. Mol. Cell Biol.* 7:505–516.
72. Wang, Q., G. Villeneuve, and Z. Wang. 2005. Control of epidermal growth factor receptor endocytosis by receptor dimerization, rather than receptor kinase activation. *EMBO Rep.* 6:942–948.
73. Hofman, E. G., A. N. Bader, ..., P. M. P. Van Bergen En Henegouwen. 2010. Ligand-induced epidermal growth factor receptor (EGFR) oligomerization is kinase-dependent and enhances internalization. *J. Biol. Chem.* 285:39481–39489.
74. Lidke, D. S., P. Nagy, ..., T. M. Jovin. 2004. Quantum dot ligands provide new insights into erbB/HER receptor-mediated signal transduction. *Nat. Biotechnol.* 22:198–203.
75. Altman, R. B., Q. Zheng, ..., S. C. Blanchard. 2012. Enhanced photostability of cyanine fluorophores across the visible spectrum. *Nat. Methods*. 9:428–429.

Article

Selecting Ions by Size in a Calcium Channel: The Ryanodine Receptor Case Study

Dirk Gillespie,^{1,*} Le Xu,² and Gerhard Meissner²¹Department of Molecular Biophysics and Physiology, Rush University Medical Center, Chicago, Illinois; and ²Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina

ABSTRACT Many calcium channels can distinguish between ions of the same charge but different size. For example, when cations are in direct competition with each other, the ryanodine receptor (RyR) calcium channel preferentially conducts smaller cations such as Li⁺ and Na⁺ over larger ones such as K⁺ and Cs⁺. Here, we analyze the physical basis for this preference using a previously established model of RyR permeation and selectivity. Like other calcium channels, RyR has four aspartate residues in its GGGIGDE selectivity filter. These aspartates have their terminal carboxyl group in the pore lumen, which take up much of the available space for permeating ions. We find that small ions are preferred by RyR because they can fit into this crowded environment more easily.

INTRODUCTION

Calcium-selective ion channels have a wide range of selectivity, permeation, and gating properties. From the selectivity and permeation point of view, there are roughly two general classes of calcium channels: 1), the surface-membrane calcium channels (e.g., L-, T-, P/Q-, and N-type channels), which tend to have a micromolar Ca²⁺ affinity (1,2) and relatively smaller conductance (3); and 2), the intracellular calcium channels (e.g., the ryanodine receptor (RyR) and the inositol triphosphate receptor (IP₃R)), which have a much weaker millimolar Ca²⁺ affinity (4–7). These two calcium channel types share very little homology, but both have a mixture of four glutamates and aspartates in their selectivity filters (8–13).

While selectivity and permeation of Ca²⁺ is the main physiological function of these channels, monovalent cations also play an important role. For example, RyR and IP₃R's poor Ca²⁺ selectivity allows them to simultaneously conduct K⁺ countercurrent to prevent large changes in membrane potential during Ca²⁺ release (14,15). Nonphysiological monovalent cations are also important because by studying them, we can improve our understanding of ion channel selectivity and permeation.

Selectivity among different monovalents has become an important topic in the years since the structure of the potassium channel was published (16). The potassium channel's physiological selectivity hinges on distinguishing between ions of different size, with one of the primary determinants being the dehydration energy difference between Na⁺ and K⁺ (17–20). Interestingly, RyR shares a relatively large homology with the potassium channel in the selectivity filter

(21–24). However, there are significant differences between the two pores, namely, the eight negative charges of RyR (11–13) and the 2–3 times larger diameter of RyR (25,26). This leads to a RyR selectivity in favor of the smaller Na⁺ over the larger K⁺ (27,28) (this study), the opposite of what happens in the potassium channel. Therefore, understanding size selectivity in RyR gives us a broader picture of selectivity in all channels.

In the last few years, selectivity in calcium channels has been studied with various methods, and two distinctly different models of the selectivity filter have emerged. The model proposed by Nonner et al. (29) describes the glutamates and aspartates, as well as their negatively charged, terminal carboxyl (COO⁻) groups, as protruding into the lumen of the selectivity filter, whereas the model of Corry et al. (30) describes the carboxyl groups as being in the protein, outside of the permeation pathway. In the first model, the pore is a liquid-like environment because the glutamates and aspartates are flexible and so the carboxyl groups can rearrange around the permeating cations. The role of the fluctuations of the carboxyl groups is to coordinate the ions; the better the coordination, the better the selectivity. In the second model, the lumen of the selectivity filter contains only the permeating cations and the carboxyl groups cannot rearrange to screen them. Selectivity in this model is almost exclusively the result of electrostatics (30) and therefore it cannot distinguish between two ion species of the same valence (31,32). The model pore of Nonner et al. (29), however, predicts large differences in affinity between monovalent cations (28,29,33–37).

The difference in selectivity between these models of the L-type calcium channel is due to the direct (i.e., very close-range) interactions of the carboxyl groups with the permeating cations; only with the carboxyl groups in the

Submitted April 10, 2014, and accepted for publication September 30, 2014.

*Correspondence: dirk_gillespie@rush.edu

Editor: Nathan Baker.

© 2014 by the Biophysical Society
0006-3495/14/11/2263/11 \$2.00



permeation pathway does the pore have micromolar Ca^{2+} affinity (38). The carboxyls in the pore make a highly charged environment that attracts cations, while at the same time making a very crowded, liquid-like environment where ions must compete for space with each other and the carboxyls. In this charge/space competition (CSC) mechanism, selectivity is a balance of electrostatic and excluded-volume forces (i.e., two ions cannot overlap) (28,29,33–42).

The importance of this balance of forces has recently been studied in RyR using a Poisson-Nernst-Planck/density functional theory (PNP/DFT) model. That analysis showed the bulk of the energetics that favors Ca^{2+} over monovalent cations is provided by the screening of the permeant ions by the carboxyls (28). It was also found that, with all else being equal, when monovalents compete with Ca^{2+} for the pore, the smaller ones compete much more effectively because it is energetically easier to insert a small ion than a bigger one (28). In this work, we continue the study of selectivity in RyR by considering monovalent versus monovalent selectivity using the same RyR model.

The advantage of using this model is that it reproduces all the known experimental permeation and selectivity data of RyR from two major labs in more than 180 ionic mixtures and several mutants, including the new experiments described here (28,37,43,44). Other calcium channel models, at best, reproduce a very small amount of experimental data (30,38) and in many cases only consider ion binding in the selectivity filter, not current (29,33–36,38,40–42). The RyR model also predicted (before experiments confirmed them) all of the known anomalous mole fraction effects of RyR (28,37) and how they change with voltage and concentrations (43). In addition, the model predicted and experiments confirmed that Ca^{2+} selectivity of RyR actually decreases with increasing luminal Ca^{2+} concentration (44). Collectively, these results strongly suggest that the model captures the physics of both permeation and selectivity when one, two, or (as we show here) three cation species compete for the pore. The PNP/DFT approach has an additional advantage over the particle simulations used in the other calcium channel model: it naturally decomposes the energetics of selectivity into distinct energy terms that allow one to understand selectivity.

Using this approach, we find that there are large differences only for the nonelectrostatic energy terms; the electrostatic energies, from both the mean electrostatic potential and the ions' ability to screen other ions, were very similar across the ion species we studied. The only large energetic difference between monovalent cations is in the term that quantifies how easily the ions fit between the carboxyl groups that are inside the selectivity filter lumen. Small ions are preferred by RyR because they fit much more easily than larger ions. While small ions can get closer to other ions to screen their charge better, this advantage is secondary.

THEORY AND METHODS

Experiments

Single channel measurements were performed as previously described using Mueller-Rudin-type planar lipid bilayers containing a 5:3:2 mixture of bovine brain phosphatidylethanolamine, phosphatidylserine, and phosphatidylcholine (25 mg of total phospholipid/ml n-decane) (12). Proteoliposomes containing the purified RyRs were added to the *cis* (cytosolic) chamber of the bilayer apparatus and fused with the lipid bilayer. Single-channel currents were measured with the indicated buffer solutions on both sides of the lipid bilayer, and 2–20 μM Ca^{2+} and 1 mM or no ATP in the *cis* chamber of the bilayer apparatus. The *trans* (luminal) side of the bilayer was defined as ground. Electrical signals were filtered at 2 kHz, digitized at 10 kHz, and analyzed as described previously (12). All data are from the RyR2 (cardiac muscle) isoform.

Models

In this work, we used our previously published model (28) without any changes. The geometry of the pore and the locations of the amino acids used in the model are shown in Fig. 1. Further details may be found in Gillespie (28).

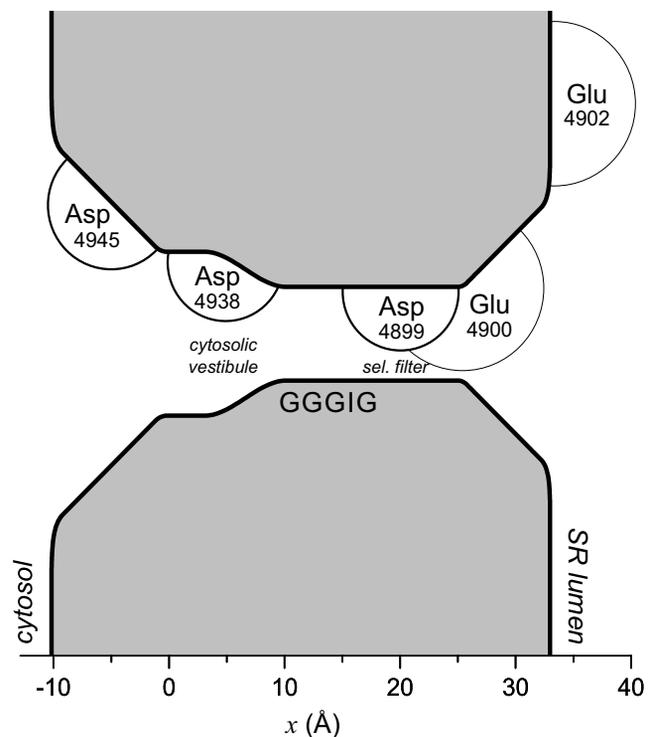


FIGURE 1 Geometry of the model RyR pore. In the experiments and calculations, the lumen of the sarcoplasmic reticulum (SR) is electrically grounded. The circle around each labeled amino acid is meant to illustrate the range of the motion of the terminal carboxyl group. Aspartates (*thick circles*) and glutamates (*thin circles*) are given a radius of 5 Å and 7 Å, respectively. Only the amino acids of one of the four identical RyR subunits is shown. Asp-4945, Asp-4938, Asp-4899, Glu-4900, and Glu-4902 are the only amino acids explicitly modeled in the theory. The GGGIG sequence (4894–4898 in the numbering) at the cytosolic end of the selectivity filter is only a reference point for readers familiar with the RyR sequence. Throughout, we use the RyR1 numbering scheme of the amino acids (34).

The ions are modeled as charged, hard spheres and their flux through the pore is described by a combination of 1D Poisson-Nernst-Planck theory and density functional theory (PNP/DFT) of fluids (45):

$$-J_i = \frac{1}{kT} D_i(x) A(x) \rho_i(x) \frac{d\mu_i}{dx} \quad (1)$$

$$-\frac{\epsilon\epsilon_0}{A(x)} \frac{d}{dx} \left(A(x) \frac{d\phi}{dx} \right) = e \sum_i z_i \rho_i(x) \quad (2)$$

where ρ_i and μ_i are the concentration and electrochemical potential, respectively, of ion species i throughout the pore and baths; J_i is the flux of species i ; and $A(x)$ is the area of the equi-chemical potential surfaces that is estimated as previously described (46,47). In the pore, this corresponds to the cross-sectional area whose radius is shown in Fig. 1. The dielectric constant ϵ of the system is 78.4. ϵ_0 is the permittivity of free space, k is the Boltzmann constant, and $T = 298.15$ K is the temperature. The functions ϕ and D_i are the mean electrostatic potential and the diffusion coefficient of species i , respectively.

The 1D description of Eqs. 1 and 2 assumes that there are no significant radial effects on the concentrations, for example those caused by a hard-wall model of the lumen/protein interface. Such a simple ion size/rigid pore model is not used here. Instead, we assume that the walls are relatively flexible. This is supported by experiment; large cations with diameters of ~ 7 Å flow through RyR with low conductance (25), but the pore can stretch to conduct the ~ 10 -Å-diameter neomycin with sufficiently high voltage (26).

DFT of ions

The electrochemical potentials μ_i are described by DFT of electrolytes (39,48), which decomposes them into four terms:

$$\mu_i = \underbrace{kT \cdot \ln(A_i^3 \rho_i(x))}_{\text{ideal gas}} + \underbrace{z_i e \phi(x)}_{\text{mean electrostatic}} + \underbrace{\mu_i^{\text{SC}}(x)}_{\text{screening}} + \underbrace{\mu_i^{\text{HS}}(x)}_{\text{excluded volume}} \quad (3)$$

where A_i is the thermal de Broglie wavelength, a constant that will drop out later since potentials are only defined up to a constant. These terms represent the different contributions to the energy required to insert an ion of species i at location x .

If ions were point charges, the ideal gas and mean electrostatic potential terms would make the familiar Poisson-Nernst-Planck theory of drift-diffusion (49,50). The size of the ions is described by the screening and excluded-volume terms. The excluded-volume term describes the energetic penalty to find space for an ion among all the other ions. It is a penalty (i.e., a positive energy) because other ions (and waters) will overlap with the new ion, which is not allowed, and therefore must move out of the way. The screening term, on the other hand, is a favorable term (i.e., a negative energy) because it measures how well ions rearrange to screen each other to minimize the instantaneous electrical potential. Roughly speaking, the more perfectly screened the ions are, the easier it is to insert another ion, and small ions do a better job of screening because they approach closer to the charge. It found that the screening term is important for RyR Ca^{2+} versus monovalent cation selectivity (28). It is also what gives ions an activity coefficient that is generally < 1 (51).

One important aspect of all of the terms in Eq. 3 is that they are not instantaneous measures of these energies. Rather, they are long-time averages at each location, just like the concentrations on which they depend. Therefore, the mean electrostatic potential, given by the Poisson equation (Eq. 2), is the average electrostatic potential, not an instantaneous electrostatic potential. Because of this, $\phi(x)$ affects all of the monovalent ion concentrations equally strongly (Eq. 3).

Some important aspects of the instantaneous electrostatic potential are lost in a mean-field theory where only the mean electrostatic potential is considered. For example, small ions will bind closer than large ions to the side chains of the aspartate and glutamate residues, and therefore they will experience a stronger electrostatic interaction and at the same time screen the charge more effectively. While mean-field theories such as PNP miss this effect, this is exactly the interaction that is captured in the screening term: $\mu_i^{\text{SC}}(x)$ averages the Coulombic interactions using the radial distribution function (which includes ion sizes) (51), whereas PNP does not.

RyR and ion dehydration

Something not included in the model is ion dehydration/resolution. In many channels, ions must shed their hydration shells of nearest waters to enter the channel; this is an energetic penalty. They are then resolvated by the channel protein when the side chains or backbone carbonyl oxygens interact directly with the ions as the waters would have; this is generally energetically favorable because the protein side chains or carbonyl oxygens tend to be at least partially charged. Since it takes a lot of energy to dehydrate an ion, this can be a very important part of the energetics of selectivity (e.g., in the potassium channel (17–20)).

In RyR, however, empirical evidence shows that ion dehydration is not a significant component of selectivity or permeation. For example, Mg^{2+} permeates the RyR equally as well as Ca^{2+} (5). This is significant because the Gibbs energy of solvation of Mg^{2+} is 130 kT larger than that for Ca^{2+} (52). This very large difference is enough to prevent Mg^{2+} permeation through the L-type calcium channel (3). The fact that the large dehydration energy of Mg^{2+} is not a hindrance to its conduction is one indication that ion dehydration is not a major factor in RyR.

This may seem somewhat surprising given the well-established homology between the RyR selectivity filter and that of the potassium channel (21–24). RyR is, however, significantly wider; large-ion conduction experiments indicate a minimum RyR pore radius of ~ 7 Å (25,27) that can stretch to 10 Å (26). This makes the need to strip waters off ions significantly less than in the much narrower potassium channel (53).

The model also lends credence to this idea. Because the model is based on physics and not data fitting of individual data points, if ion dehydration/resolution played a major role, then the model should fail to reproduce the experimental data. Only nine data points were used to establish the model parameters, so all of the other hundreds of data points are purely the result of the physics that is in the model (28,37). By ignoring ion dehydration, we are hypothesizing that ion dehydration is much smaller than the other terms in Eq. 3, for example, that the electrostatic energy of being surrounded by carboxyl groups in the selectivity filter is much more favorable than ion dehydration is a penalty.

As described in the Introduction, this PNP/DFT model reproduces and predicts the current/voltage relations of native RyR in more than 120 previously published ionic solutions (28,37,43,44) and another 60 ionic solutions presented here, as well as the current/voltage relations of various charge-neutralizing mutations (28,37). If ion dehydration/resolution were a significant contributor to either permeation or selectivity, one would surmise that a model that did not include this would not be able to reproduce or predict data over the wide range of conditions as this model has (e.g., micromolar to molar concentrations, ± 150 mV voltage ranges, and hundreds of various mixtures of Li^+ , Na^+ , K^+ , Rb^+ , Cs^+ , Mg^{2+} , and Ca^{2+}), even if under some conditions there may have been a cancellation of errors. This indicates that the model correctly describes, at least to first order, the physics of RyR permeation and selectivity.

Energetics of selectivity

To study binding selectivity (i.e., the amount of an ion species that accumulates at one location in the pore), we consider identical baths and zero applied voltage so that the ions are in equilibrium. Then, all of the electrochemical potentials are constant everywhere and there are no ion fluxes from any species. In Eq. 3, the left-hand side is then a constant, namely,

the electrochemical potential of the baths. The right-hand side contains the components of the electrochemical potential as they change with location in the baths and into the channel and selectivity filter. Because the left-hand side is constant at equilibrium, any change with location in one term on the right-hand side must be countered with a change in the opposite direction by the other terms.

We will examine how the components of the electrochemical potential compare among monovalent cations of different size. To do that, Eq. 3 for ion species i is rewritten as

$$-kT \ln \left(\frac{\rho_i(x)}{\rho_i^{\text{bath}}} \right) = \overbrace{z_i e \phi(x)}^{\text{mean electrostatic}} + \overbrace{\Delta\mu_i^{\text{SC}}(x)}^{\text{screening}} + \overbrace{\Delta\mu_i^{\text{HS}}(x)}^{\text{excluded volume}} \quad (4)$$

where $\Delta\mu_i^{\text{SC}}(x)$ indicates the screening chemical potential at x minus the bath value, and similarly for the HS term. To compare two different ion species (e.g., K^+ and Cs^+), Eq. 4 for the two species can be subtracted:

$$\begin{aligned} \ln \left(\frac{\rho_{\text{Cs}}(x)}{\rho_{\text{K}}(x)} \right) &= \overbrace{\ln \left(\frac{[\text{Cs}^+]}{[\text{K}^+]} \right)}^{\text{number advantage}} + \overbrace{(z_{\text{K}} - z_{\text{Cs}}) \frac{e\phi(x)}{kT}}^{\text{mean electrostatic advantage}} \\ &+ \overbrace{\frac{1}{kT} (\Delta\mu_{\text{K}}^{\text{SC}}(x) - \Delta\mu_{\text{Cs}}^{\text{SC}}(x))}^{\text{screening advantage}} \\ &+ \overbrace{\frac{1}{kT} (\Delta\mu_{\text{K}}^{\text{HS}}(x) - \Delta\mu_{\text{Cs}}^{\text{HS}}(x))}^{\text{excluded-volume advantage}}. \end{aligned} \quad (5)$$

Eq. 5 defines the binding selectivity (the left-hand side) in terms of the ratio of concentrations at the same location x . (In this example, a positive term favors the binding of Cs^+ whereas a negative term favors K^+ .) To have one ion species at a higher concentration, at least one of the four components on the right-hand side must favor that ion species. In line with previous work (28,54), we call these terms advantages because they reflect which ion species has an energetic advantage for ion binding in each term. The number advantage describes which ion species has a higher bath concentration and is therefore more likely to enter the channel. The mean (i.e., long-time averaged) electrostatic advantage favors the ion species with the higher valence. Here, all of the ions are monovalents, so this term is always zero. The screening advantage favors small, high-valence ions because they neutralize the charge of other ions (screen) better than large, low-valence ions. This is one way in which ion size comes into selectivity; the other way is through the excluded-volume advantage.

This same type of analysis has been used previously to understand Ca^{2+} versus monovalent selectivity in RyR (28). That work showed that RyR is a Ca^{2+} channel because of its electrostatic advantages, both in the mean electrostatic advantage and in the screening advantage. Moreover, the model predicted that RyR calcium selectivity diminishes as $[\text{Ca}^{2+}]$ increases because the pore becomes more charge-neutral. This decreases the mean electrostatic advantage, whereas the screening advantage was predicted to remain unchanged. This was recently verified experimentally (44), showing that breaking the energetics into these contributions correctly captures the reasons for selectivity.

Lastly, we describe an aspect of the electrochemical potential that we will use later. Each component in Eq. 3 (except the ideal gas term) for species i depends on the concentrations of all ion species, not just species i . This means that even if an ion species is present at only trace concentrations, its electrochemical potential at any point in space is not zero; it takes energy to insert even one ion into the mix of ions already present at that point. Later, we will use this idea to compare the electrochemical potentials of three different ions (Na^+ , K^+ , and Cs^+) even if only two of them (e.g., Na^+ and K^+) are in the baths.

Assumptions and approximations

Like all models, this model includes assumptions and approximations. They generally fall into two categories: the description of the channel and the physics used to describe the ions and the ion current.

The structure of the channel was inferred from experimental mutation data as previously described in detail (28,37). This was necessary (and continues to be so) because although a low-resolution structure of RyR has been well established for some time (21–23), a high-resolution structure has not yet been determined (55) (Montserrat Samsó, Virginia Commonwealth University, personal communication, 2014), especially not for the selectivity filter. So far, structures that include the pore have a resolution of ~ 10 Å and only certain isolated domains (e.g., the N-terminal domain (56–58)) have yielded high-resolution x-ray crystallographic structures. Given these limitations, the overall structure of the selectivity filter and pore shown in Fig. 1 was chosen to reflect the well-established structural homology with the potassium channel (21–24) and the fact, as described above, that the RyR pore is significantly wider, in line with a previous molecular dynamics study by Shirvanyants et al. (59). That study showed large fluctuations in the selectivity filter diameter from 0 to 6 Å. One shortcoming of our model is that these fluctuations are not included in the 1D model used here, which necessarily requires a fixed diameter.

On the physics side, the use of mean-field equations, like the PNP equations, has been criticized as not being valid in narrow channels (60). However, in a recent study by one author of Ref. 60, the opposite was found when the size of the ions was taken into account (61). Specifically, when the density profiles were taken from Brownian dynamics simulations (whose description of ions is very similar to that used here), even a 1D Nernst-Planck theory reproduced the full current-voltage curve of the Brownian dynamics simulation. Therefore, the permeation physics of the Nernst-Planck equation is probably correct because we are not considering single-file channels where conservation of momentum is important. Comparisons with simulations show that DFT correctly computes the profiles of hard-sphere ions in many systems (62,63). Moreover, our 1D DFT RyR profiles are very similar to cross-sectional averages of full 3D simulations (Dezsó Boda, University of Pannonia, personal communication, 2014), so it is probable that the 1D Nernst-Planck equation used here is applicable. However, more studies on crowded channels should be done, which is work we plan to continue (64–66).

One also cannot rule out the possibility that approximations that underlie the DFT (e.g., the mean spherical approximation) may scale similarly to physics that is not included (e.g., ion dehydration), resulting in a cancellation of errors. We have tested this previously (67) and are continuing to do so.

RESULTS

In this work, we examine the competition of monovalent cations for the RyR pore. Specifically, we analyze the energetic differences between Na^+ , K^+ , and Cs^+ , which differ in both size and RyR conductance; their unhydrated crystal diameters are 2.04, 2.76, and 3.40 Å (68), respectively, and their experimentally measured conductances in 250 mM symmetric conditions are 481, 800, and 519 pS (37), respectively. We also conducted experiments using Li^+ , which has a crystal diameter of 1.33 Å and conductance of 210 pS.

Comparing the model with experiments

The first thing to know is whether the model can successfully compute the competition between multiple monovalent cations. Previous studies using the same RyR model compared model and experimental results when two cation

species competed for the pore, but only under a limited set of conditions. Specifically, they considered bi-ionic conditions and two mole fraction experiments where the relative concentrations of two cation species was changed (28,37).

In Fig. 2, we show more comparisons of the model with experiments. For all of the results shown in the figure (described below), none of the previously published parameters of the model (28) were changed in any way.

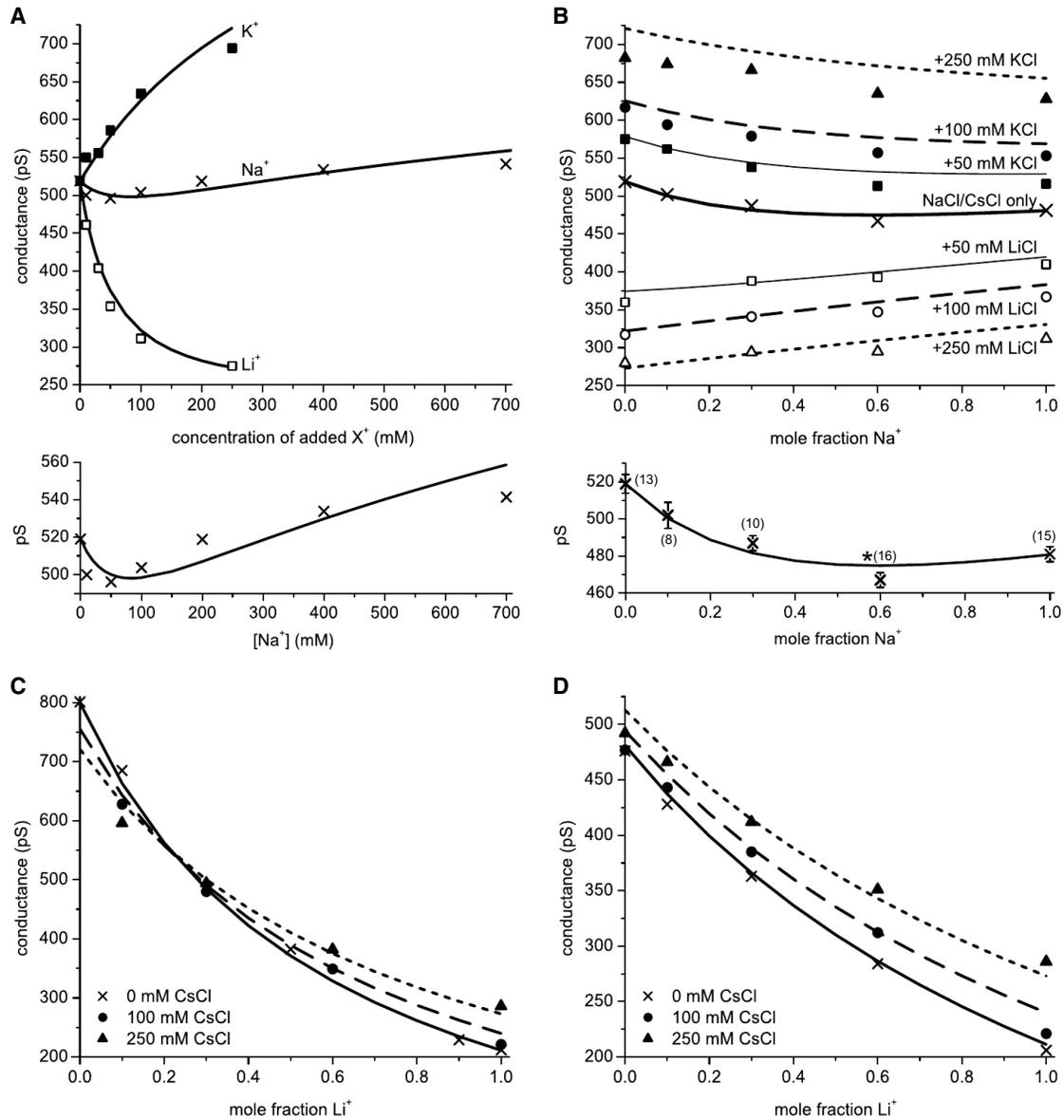


FIGURE 2 (A–D) Comparison of the model with experiments in which two (A) or three (B–D) monovalent cations compete for the pore. (A) Salting-out experiment in which XCl is added to 250 mM CsCl. Conductance is plotted as $[X^+]$ is changed for $X^+ = \text{Li}^+$ (\square), Na^+ (\times), and K^+ (\blacksquare). Below the main figure, the Na^+ curve (\times) is redrawn on a different scale to show its minimum. (B) Two sets of experiments in which Na^+ , Cs^+ , and Li^+ compete for the pore or Na^+ , Cs^+ , and K^+ compete for pore while all three ion species' concentrations are changed. The mole fraction of Na^+ is changed as $[\text{Na}^+] + [\text{Cs}^+]$ is constant at 250 mM (\times and thick solid line). For the other curves, 50 mM (box symbols), 100 mM (circles), or 250 mM (triangles) of either Li^+ (open symbols) or K^+ (solid symbols) is in the bath in addition to the Na^+/Cs^+ mixture. The error between theory and experiment is largest for the top dotted line, but the relative error for that line is always less than 6%. Below the main figure, the Na^+ curve (\times) is redrawn on a different scale to show its minimum. It also shows the number of recordings in parentheses and error bars. The point at mole fraction 0.6 was shown earlier to be statistically significantly smaller than its two neighbors (denoted with *), experimentally verifying the minimum. (C) Li^+ , K^+ , and Cs^+ compete for the pore while all three ion species' concentrations are changed. The mole fraction of Li^+ is changed as $[\text{Li}^+] + [\text{K}^+]$ is constant at 250 mM (\times and solid line). For the other curves, 100 mM (\bullet) or 250 mM (\blacktriangle) of Cs^+ is in the bath in addition to the Li^+/K^+ mixture. (D) Li^+ , Na^+ , and Cs^+ compete for the pore while all three ion species' concentrations are changed. The mole fraction of Li^+ is changed as $[\text{Li}^+] + [\text{Na}^+]$ is constant at 250 mM (\times and solid line). For the other curves, 100 mM (\bullet) or 250 mM (\blacktriangle) of Cs^+ is in the bath in addition to the Li^+/Na^+ mixture. In all panels, both the cytosolic and luminal baths are identical. All standard error bars are approximately the size of the symbol or smaller.

One kind of experiment we conducted was a salting-out experiment in which [CsCl] was kept at 250 mM while more and more LiCl, NaCl, or KCl was added to the baths. This is the experiment shown in Fig. 2 A. Not only does the model reproduce the large changes in conductance when Li^+ or K^+ are added, but the model also reproduces the minimum in the conductance when Na^+ is added to the Cs^+ . This minimum is similar to the anomalous mole fraction effect (AMFE) that was predicted by the model for this mixture of these two ion species (37), shown as the thick solid line in Fig. 2 B. Both of these curves are redrawn below the main graphs in Fig. 2, A and B, to explicitly show their minima.

In Fig. 2, B–D, three monovalents are competing for the pore. The model predicts all of the features of the experimental results. For example, the Na^+/Cs^+ AMFE (Fig. 2 B, *thick solid line*) disappears when either Li^+ or K^+ is present in addition to the Na^+/Cs^+ mixture (Fig. 2 B, *box symbols*). Also, the model reproduces the crossing of the curves in Fig. 2 C and the crossover concentration. The relative error between the model and experiments is always less than 6% when three cations are competing for the pore.

All experiments were done after the calculations were performed. Therefore, the curves shown in the Fig. 2 are predictions of the model based on the physics of ion permeation and selectivity contained in the model. While one can never eliminate the possibility of canceling errors, one can infer that if the physics in the model is wrong (e.g., if ion dehydration plays a major role or if drift-diffusion does not correctly describe ion permeation), then the model should fail to reproduce experiments. Fig. 2 shows that the model seems to get the proportions of the three cations in the pore correct, which would not be possible if many of the energies were several kT off. Given how well the model predicted these challenging and nonlinear results, we will now use the model to analyze how RyR selects among monovalent cations of different sizes.

One cation species competing for the pore

To start the analysis of selectivity, let us first consider the concentration profiles of the individual ions when they are not competing with each other for the pore, that is, when there is only one cation species in the baths. This is shown in Fig. 3 A for 250 mM symmetric conditions.

Overall, the profiles are very similar. They reach a maximum of ~ 13 M in the selectivity filter where the protein charge is the highest. Other local bumps in concentration of ~ 5 M occur near the other charged amino acids (shown in Fig. 1). All of the ions have the same concentration in the charged regions because the negative charges of the amino acids must be neutralized as much as possible. As previously shown, there are always approximately three monovalents in the filter (37).

A more nuanced examination of the profiles shows that the larger the ion, the more oscillations there are. Cs^+ , espe-

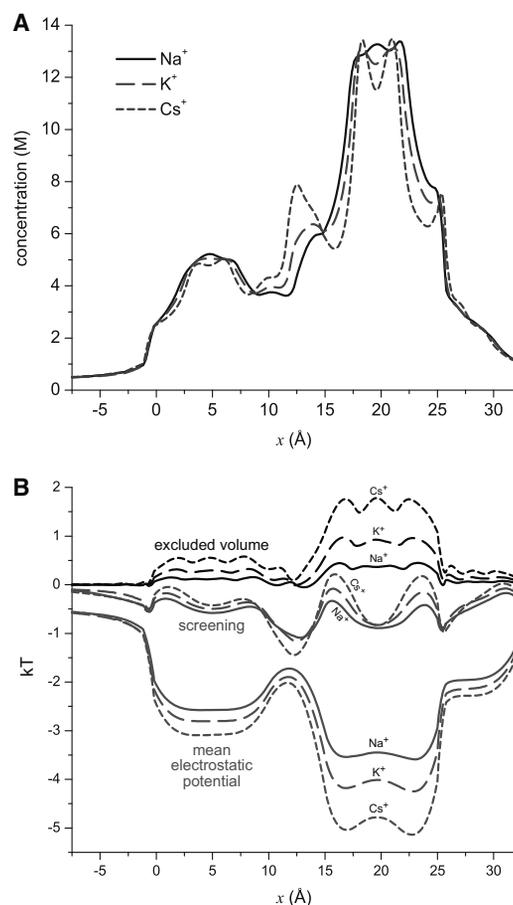


FIGURE 3 (A) Concentration profiles throughout the pore of Na^+ (*solid, black*), K^+ (*long-dashed, red*), and Cs^+ (*dotted, blue*). For each curve, XCl ($X = \text{Na}, \text{K}, \text{or Cs}$) is at 250 mM in both baths and the other two cations are not present. (B) Profiles of the terms in Eq. 4: excluded volume $\Delta\mu_i^{\text{HS}}(x)$ (*black*), screening $\Delta\mu_i^{\text{SC}}(x)$ (*purple*), and mean electrostatic potential $z_i e \phi(x)$ (*green*). Each group has a curve for the three cations: Na^+ (*solid*), K^+ (*long-dashed*), and Cs^+ (*dotted*). To see this figure in color, go online.

cially, has two distinct peaks in the selectivity filter (between $x = 15$ Å and 25 Å) and another on the cytosolic side of the filter (between $x = 10$ Å and 15 Å). This is a commonly found feature of large ions. In regions where ions must be present to neutralize a charge, their large size prevents other ions from being near them. This results in distinct layers of ions, seen in the profiles as peaks of high concentration and troughs of low concentration (69–71).

The idea that it is more difficult for large ions to find space in regions crowded with charged amino acid groups (the COO^- terminal groups of the aspartates and glutamates) and other ions is reflected in the energetics. This is shown in Fig. 3 B, where the excluded volume, screening, and mean electrostatic potential terms from Eq. 4 are plotted.

For all of the ions, the excluded volume term is positive because the pore is more crowded than the bath. Therefore, it is more difficult for the ions to find space in the pore than in the bath, resulting in an energetic penalty (a positive

energy in Fig. 3 *B*). Conversely, because the pore is charged, there is a negative electrostatic potential pulling the cations in. Similarly, screening is easier when more negative charges are around, and thus the screening term is negative (favoring ions accumulating in the pore).

The biggest difference among the three cations is in the excluded volume and mean electrostatic potential terms; the differences in the screening term are relatively small in comparison. In the excluded volume term, the largest ion (Cs^+) has the largest energetic penalty, as expected, whereas the smallest ion (Na^+) barely has any excluded volume penalty.

Since all of the ions have roughly the same concentrations throughout the pore as they try to neutralize the protein charge, it follows from Eq. 4 that any differences between the cations in one energetic term must be counteracted by another term. In this case, the difference in the excluded volume terms is made up by the mean electrostatic potential. Specifically, because the large Cs^+ has more difficulty in finding space, a larger (more negative) electrostatic potential is necessary to pull in sufficient numbers of Cs^+ .

Two and three cation species competing for the pore

Now, let us examine what occurs when the ions compete with each other for the pore. Fig. 4 shows the concentration profiles for the three cations as the bath K^+ concentration is increased from 0 mM (*A*) to 125 mM (*B*) and 250 mM (*C*) whereas the Na^+ and Cs^+ concentrations remain constant at 125 mM each.

In the absence of K^+ , only Na^+ and Cs^+ are competing for the pore, and even though they are at the same concentration in the baths, there is a large discrepancy in their concentrations in the pore; there are far more Na^+ than Cs^+ (Fig. 4 *A*). This is also true when all three cations have the same bath concentration (Fig. 4 *B*); the ions' concentrations are inversely related to their size. In fact, only when $[\text{K}^+]$ is twice $[\text{Na}^+]$ is the K^+ concentration in the selectivity filter the same as the Na^+ concentration (Fig. 4 *C*).

The energetic origin of this difference is shown in Fig. 5. There, the excluded-volume and screening terms of Eq. 4 are shown for all six mixtures of cations examined so far in Figs. 3 *A* and 4. In this case, only these two terms are shown because now we want to consider the differences in energetics between ion species, as in Eq. 5. Since all of the ions have the same charge, their mean electrostatic potential is the same in each experiment. Fig. 5 *A* shows the profiles along the pore of the two terms, whereas Fig. 5 *B* shows the values in the middle of the selectivity filter for individual experiments.

The aggregate results of Fig. 5 *A* show how the excluded-volume and screening terms behave in general. Specifically, it can be seen that the screening terms are approximately the same for all cations under all conditions. This is significant because smaller ions are better screeners, something that usually gives them an energetic advantage. However, here this advantage is relatively small compared to the excluded-volume term. Unlike the screening term, this term stratifies into three layers, one for each cation, indicating significant differences among the ions. Collectively, this shows that the only substantive difference between the cations is how their size makes it difficult for them to find space, rather than how their size makes it difficult to screen; this screening effect is secondary.

This is borne out in the results of the individual experiments shown in Fig. 5 *B*. In each case, the smaller ions are marginally better screeners, but more importantly, they have a significantly smaller energetic penalty for finding space in the selectivity filter. Specifically, the screening advantage of the small Na^+ over the large Cs^+ is always <0.25 kT. This is about five times larger than bulk, indicating that Na^+ has a significant increase in screening ability because its charge can get closer to the negative carboxyl groups. However, its excluded volume advantage is consistently larger at ~ 1 kT. These energetic differences may seem small, but the ratio of Na^+ to Cs^+ concentration is the exponential of this difference when these two ions compete head to head, creating an ~ 3.5 -fold difference in concentration in the selectivity filter (Fig. 4).

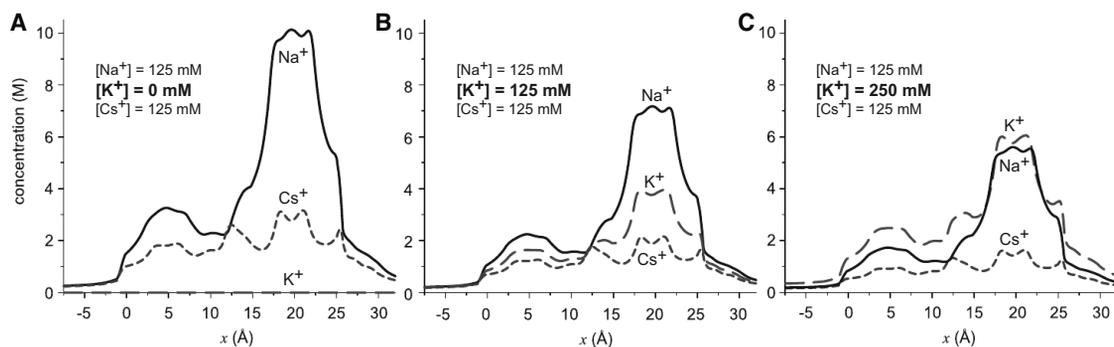


FIGURE 4 The concentration profiles throughout the pore of Na^+ (solid, black), K^+ (long-dashed, red), and Cs^+ (dotted, blue) as $[\text{K}^+]$ in both baths is increased from 0 mM (*left*) to 125 mM (*middle*) to 250 mM (*right*). In all panels, $[\text{Na}^+]$ and $[\text{Cs}^+]$ are 125 mM. To see this figure in color, go online.

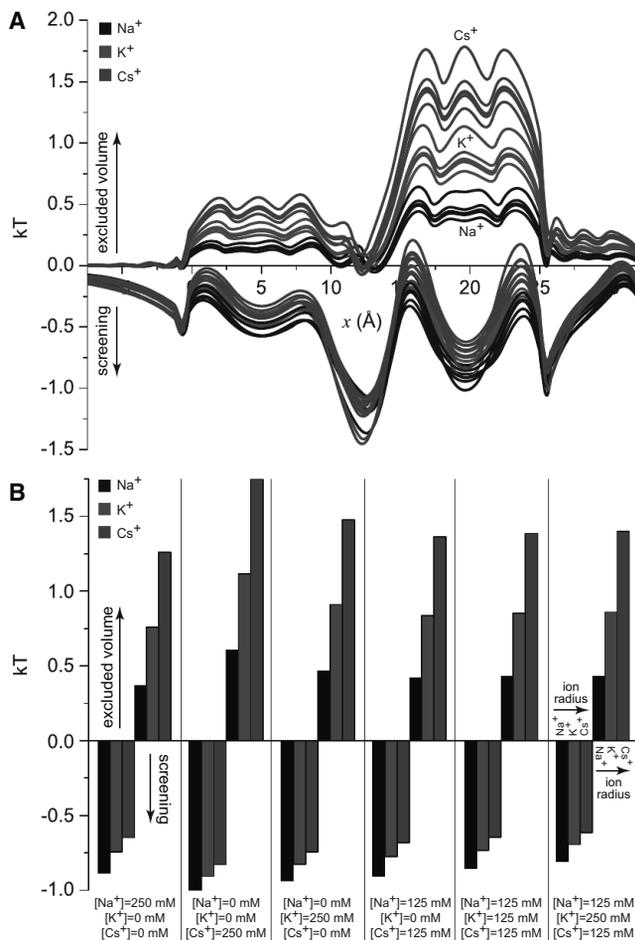


FIGURE 5 (A) Profiles of the excluded-volume and screening terms in Eq. 4 for Na^+ (black), K^+ (red), and Cs^+ (blue) for the six conditions whose concentration profiles are shown in Figs. 3 A and 4 (enumerated at the bottom of panel B). The group of lines for the excluded-volume term is positive, whereas the group for the screening term is negative. (B) Bar graph of the values of the lines in panel A in the middle of the selectivity filter ($x = 20 \text{\AA}$). To see this figure in color, go online.

DISCUSSION

Whether an ion channel spans intra- or extracellular membranes, it faces a sea of ions from which it must pick ions to conduct. The atomic cations tend to be monovalent (Na^+ and K^+) or divalent (Ca^{2+} and Mg^{2+}) and so a channel must be able to select among ions of the same charge whose main difference is size. Many monovalent cation-selective channels, such as the neuronal sodium and potassium channels, exclude essentially all other physiological ions; their selectivity is for their namesake ion only (72). Similarly, many extracellular calcium channels have such high Ca^{2+} affinity that they exclude virtually all other ions under physiological conditions, even if they do conduct monovalent cations in the absence of Ca^{2+} (3,27).

Intracellular calcium channels like RyR and IP_3R offer a more nuanced view of size selectivity. They conduct both monovalent and divalent cations, even in mixtures. Both

RyR and IP_3R have a Ca^{2+} to K^+ permeability ratio of only ~ 7 , and both conduct Mg^{2+} (5,6). In fact, the conductance of K^+ and Mg^{2+} at the same time as Ca^{2+} is important for their physiological roles in intracellular calcium release (14,15,44). Moreover, these channels can distinguish between monovalents of different size, preferring small ones to large ones (27).

This can be seen, for example, in experiments with RyR where Ca^{2+} was added to symmetric 100 mM NaCl and separately to 100 mM CsCl (28). Individually, Na^+ and Cs^+ have approximately the same conductance through RyR (480 and 520 pS, respectively, as shown in Fig. 2 B), so there is little difference in how fast the ions move through the channel. However, only $10 \mu\text{M}$ Ca^{2+} significantly affects Cs^+ current, while ten times more is needed to affect Na^+ current (28). Viewed in terms of ion competition for the pore, Cs^+ is more easily displaced by Ca^{2+} than Na^+ . Consistent with this idea, the block of these monovalent cation currents by Ca^{2+} was significantly weaker with Na^+ (28) and virtually nonexistent with the smaller Li^+ (43).

In RyR experiments involving only combinations of monovalent cations under bi-ionic conditions, there is consistently a current from the small cation at zero applied potential (73). In other words, even though there are equal concentrations of, say, K^+ on one side of the channel and Cs^+ on the other, there is a net K^+ current unless a voltage is applied that opposes K^+ movement. With Li^+ instead of Cs^+ , Li^+ current flows. In both cases, it is the smaller cation that is favored. Also, as shown in Fig. 2 A, the addition of Li^+ to Cs^+ produces a much steeper change in conductance than the addition of K^+ , which is larger than Li^+ , suggesting that small ions more easily displace larger ones.

By what mechanism do smaller cations displace larger ones? To illustrate, consider the addition of Na^+ to 250 mM Cs^+ . For reference, the energetics profiles of Cs^+ in the absence of Na^+ are shown in Fig. 3 B. Initially, at very low $[\text{Na}^+]$, very few Na^+ are present in any part of the channel (e.g., the selectivity filter or the cytosolic vestibule). As $[\text{Na}^+]$ increases, more Na^+ are drawn in by the electrostatic potential. This mean electrostatic potential, however, applies equally strongly to both the Na^+ and Cs^+ ions, and therefore this force will not favor one ion species over the other (it disappears in Eq. 5).

The one difference between these ion species is that Na^+ has a crystal diameter that is 1.36\AA smaller than that of the Cs^+ (68) and therefore has only 18.2% the volume of Cs^+ . The smaller radius allows the charge of the Na^+ to get closer to another charge and therefore screen it better. This difference in the instantaneous electrostatic potential felt by Na^+ (compared with Cs^+) gives Na^+ an advantage in the SC term of Eq. 5. Moreover, the smaller volume allows the Na^+ to squeeze into a crowded space more easily. This gives it an even larger advantage in the HS term of Eq. 5. To understand how these terms change as the $[\text{Na}^+]$ increases

continuously, we can also look at the ratio of the Na^+ and Cs^+ concentrations in different parts of the channel (Fig. 6). Specifically, we consider crowded and relatively uncrowded parts of the channel, namely, the selectivity filter and the cytosolic vestibule (Fig. 1).

The blue line in Fig. 6 depicts what would happen if there were no differences between Na^+ and Cs^+ , for example, if electrostatic forces dominated as in the model of Corry et al. (30). In that case, the concentration ratios in each part of the channel would be the same as in the baths; only the number advantage term in the Eq. 5 would matter. However, both the excluded volume and screening advantages of Na^+ affect the final Na^+ to Cs^+ ratios. Specifically, in Fig. 6, the red line/symbols show the effect of having only the screening term and ignoring the excluded-volume term, while the black line/symbols include the effects of both the excluded-volume and screening terms.

In both the crowded selectivity filter (*lines* in Fig. 6) and the much less crowded vestibule (*symbols*), the screening term favors Na^+ ; the *red line* and *symbols* in Fig. 6 are above the *blue line*. Moreover, the screening is almost identical in both parts of the channel; the *red line* and *red symbols* are virtually the same. This means that the screening is more or less independent of crowding. On the other hand, the excluded-volume term favors Na^+ a lot more in the selectivity filter than in the vestibule; the *black line* and *symbols* in Fig. 6 are very different. This means that the ability to find space in the selectivity filter crowded with three permeating ions and eight oxygens leads to large discrimination between Na^+ and Cs^+ . Put another way, in the densely packed selectivity filter, Na^+ having a much smaller volume is a substantial advantage over Cs^+ , whereas in the less crowded vestibule this is less so.

Overall, the same conclusions hold for all combinations of different monovalent cations (data not shown). The larger

the size difference between these ions, the more the smaller one is favored in the selectivity filter because it can find space in the selectivity filter more easily, and not because it screens the carboxyl groups of the channel protein better. This preferential accumulation of small ions in the selectivity filter is then reflected in the current because the selectivity filter is the bottleneck for permeation. This makes reproducing the experimental results in Fig. 2 significant, because it requires the model to compute the correct proportions of three species, something that has not been done by other models (30).

The fact that RyR by and large selects monovalent cations by size (27) is due to both this mechanism and RyR's apparent small dehydration/resolvation penalty for ions, unlike the much narrower potassium channel. This probably makes RyR relatively unique among selective ion channels, since sodium, potassium, and other calcium channels tend to be narrow compared to nonselective channels like α -hemolysin. RyR is wide enough to conduct a large current, but still narrow enough to have substantial Ca^{2+} selectivity (albeit much weaker compared to the L-type calcium channel).

In channels where ion dehydration/resolvation is important, the selectivity sequence will not be smallest to largest. Because it is significantly easier to strip waters off a large ion than a small one, channels in which this is important can have a selectivity sequence of largest to smallest. Therefore, crowded channels are quite likely to have a range of very different selectivity sequences because each will balance small ion selectivity based on finding space with the penalty for ion dehydration in a different way. This was recently explored in a study that showed that almost all Eisenman selectivity sequences were possible, with the addition of several new ones (74). RyR is then very useful because it allows one to strip away the competing factor of ion dehydration and understand the physics of a relatively simpler system.

CONCLUSIONS

Experiments show that RyR can distinguish among many kinds of monovalent cations by size. Here, we used a model of RyR permeation and selectivity to understand why this happens, and found that it occurs because the selectivity filter has much of the lumen taken up by amino acid groups from the protein and the smaller ions fit better. Other kinds of calcium channels, as well as many other channel types, can also distinguish between ions of the same charge but different size. Therefore, this study may also help to illuminate the physics of selectivity and permeation in those channels.

D.G. was supported by NIH grant R01-AR054098 (Michael Fill, principal investigator). G.M. was supported by NIH grants AR 018687 and HL 073051.

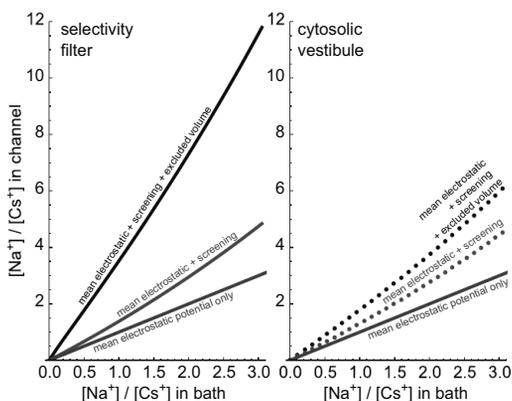


FIGURE 6 Ratio of Na^+ to Cs^+ concentrations in the selectivity filter (*solid lines*) and in the cytosolic vestibule (*symbols*) as a function of their bath concentration ratio. The effect of the screening advantage (SC) is shown in red and the effect of the excluded-volume term (HS) is shown in black. The blue line, shown for reference, is the bath concentration ratio. To see this figure in color, go online.

REFERENCES

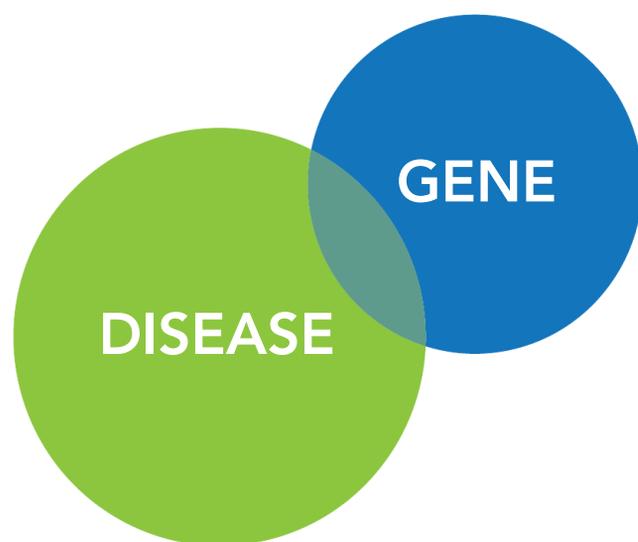
- Almers, W., E. W. McCleskey, and P. T. Palade. 1984. A non-selective cation conductance in frog muscle membrane blocked by micromolar external calcium ions. *J. Physiol.* 353:565–583.
- Almers, W., and E. W. McCleskey. 1984. Non-selective conductance in calcium channels of frog muscle: calcium selectivity in a single-file pore. *J. Physiol.* 353:585–608.
- Hess, P., J. B. Lansman, and R. W. Tsien. 1986. Calcium channel selectivity for divalent and monovalent cations. Voltage and concentration dependence of single channel current in ventricular heart cells. *J. Gen. Physiol.* 88:293–319.
- Smith, J. S., R. Coronado, and G. Meissner. 1985. Sarcoplasmic reticulum contains adenine nucleotide-activated calcium channels. *Nature.* 316:446–449.
- Tinker, A., and A. J. Williams. 1992. Divalent cation conduction in the ryanodine receptor channel of sheep cardiac muscle sarcoplasmic reticulum. *J. Gen. Physiol.* 100:479–493.
- Bezprozvanny, I., and B. E. Ehrlich. 1994. InsP₃ receptor: functional properties and regulation. In *Handbook of Membrane Channels*. C. Peracchia, editor. Academic Press, New York, pp. 511–526.
- Ramos-Franco, J., D. Galvan, ..., M. Fill. 1999. Location of the permeation pathway in the recombinant type 1 inositol 1,4,5-trisphosphate receptor. *J. Gen. Physiol.* 114:243–250.
- Yang, J., P. T. Ellinor, ..., R. W. Tsien. 1993. Molecular determinants of Ca²⁺ selectivity and ion permeation in L-type Ca²⁺ channels. *Nature.* 366:158–161.
- Ellinor, P. T., J. Yang, ..., R. W. Tsien. 1995. Ca²⁺ channel selectivity at a single locus for high-affinity Ca²⁺ interactions. *Neuron.* 15:1121–1132.
- Talavera, K., M. Staes, ..., B. Nilius. 2001. Aspartate residues of the Glu-Glu-Asp-Asp (EEDD) pore locus control selectivity and permeation of the T-type Ca(2+) channel α_{1G} . *J. Biol. Chem.* 276:45628–45635.
- Gao, L., D. Balshaw, ..., G. Meissner. 2000. Evidence for a role of the luminal M3-M4 loop in skeletal muscle Ca(2+) release channel (ryanodine receptor) activity and conductance. *Biophys. J.* 79:828–840.
- Wang, Y., L. Xu, ..., G. Meissner. 2005. Probing the role of negatively charged amino acid residues in ion permeation of skeletal muscle ryanodine receptor. *Biophys. J.* 89:256–265.
- Xu, L., Y. Wang, ..., G. Meissner. 2006. Two rings of negative charges in the cytosolic vestibule of type-1 ryanodine receptor modulate ion fluxes. *Biophys. J.* 90:443–453.
- Gillespie, D., and M. Fill. 2008. Intracellular calcium release channels mediate their own countercurrent: the ryanodine receptor case study. *Biophys. J.* 95:3706–3714.
- Guo, T., A. Nani, ..., M. Fill. 2013. Sarcoplasmic reticulum K(+) (TRIC) channel does not carry essential countercurrent during Ca(2+) release. *Biophys. J.* 105:1151–1160.
- Doyle, D. A., J. Morais Cabral, ..., R. MacKinnon. 1998. The structure of the potassium channel: molecular basis of K⁺ conduction and selectivity. *Science.* 280:69–77.
- Varma, S., and S. B. Rempe. 2007. Tuning ion coordination architectures to enable selective partitioning. *Biophys. J.* 93:1093–1099.
- Varma, S., D. Sabo, and S. B. Rempe. 2008. K⁺/Na⁺ selectivity in K channels and valinomycin: over-coordination versus cavity-size constraints. *J. Mol. Biol.* 376:13–22.
- Varma, S., D. M. Rogers, ..., S. B. Rempe. 2011. Perspectives on: ion selectivity: design principles for K⁺ selectivity in membrane transport. *J. Gen. Physiol.* 137:479–488.
- Noskov, S. Y., and B. Roux. 2007. Importance of hydration and dynamics on the selectivity of the KcsA and NaK channels. *J. Gen. Physiol.* 129:135–143.
- Ludtke, S. J., I. I. Serysheva, ..., W. Chiu. 2005. The pore structure of the closed RyR1 channel. *Structure.* 13:1203–1211.
- Samsó, M., T. Wagenknecht, and P. D. Allen. 2005. Internal structure and visualization of transmembrane domains of the RyR1 calcium release channel by cryo-EM. *Nat. Struct. Mol. Biol.* 12:539–544.
- Samsó, M., W. Feng, ..., P. D. Allen. 2009. Coordinated movement of cytoplasmic and transmembrane domains of RyR1 upon gating. *PLoS Biol.* 7:e85.
- Ramachandran, S., A. Chakraborty, ..., G. Meissner. 2013. Structural determinants of skeletal muscle ryanodine receptor gating. *J. Biol. Chem.* 288:6154–6165.
- Tinker, A., and A. J. Williams. 1993. Probing the structure of the conduction pathway of the sheep cardiac sarcoplasmic reticulum calcium-release channel with permeant and impermeant organic cations. *J. Gen. Physiol.* 102:1107–1129.
- Mead, F., and A. J. Williams. 2002. Block of the ryanodine receptor channel by neomycin is relieved at high holding potentials. *Biophys. J.* 82:1953–1963.
- Lindsay, A. R., S. D. Manning, and A. J. Williams. 1991. Monovalent cation conductance in the ryanodine receptor-channel of sheep cardiac muscle sarcoplasmic reticulum. *J. Physiol.* 439:463–480.
- Gillespie, D. 2008. Energetics of divalent selectivity in a calcium channel: the ryanodine receptor case study. *Biophys. J.* 94:1169–1184.
- Nonner, W., L. Catacuzzeno, and B. Eisenberg. 2000. Binding and selectivity in L-type calcium channels: a mean spherical approximation. *Biophys. J.* 79:1976–1992.
- Corry, B., T. W. Allen, ..., S.-H. Chung. 2001. Mechanisms of permeation and selectivity in calcium channels. *Biophys. J.* 80:195–214.
- Corry, B., T. Vora, and S.-H. Chung. 2005. Electrostatic basis of valence selectivity in cationic channels. *Biochim. Biophys. Acta.* 1711:72–86.
- Chung, S.-H., and B. Corry. 2005. Three computational methods for studying permeation, selectivity and dynamics in biological ion channels. *Soft Matter.* 1:417–427.
- Nonner, W., D. Gillespie, ..., B. Eisenberg. 2001. Ion accumulation in a biological calcium channel: effects of solvent and confining pressure. *J. Phys. Chem. B.* 105:6427–6436.
- Boda, D., D. D. Busath, ..., S. Sokolowski. 2000. Monte Carlo simulations of the mechanism of channel selectivity: the competition between volume exclusion and charge neutrality. *J. Phys. Chem. B.* 104:8903–8910.
- Boda, D., D. Henderson, and D. D. Busath. 2001. Monte Carlo study of the effect of ion and channel size on the selectivity of a model calcium channel. *J. Phys. Chem. B.* 105:11574–11577.
- Boda, D., D. Henderson, and D. D. Busath. 2002. Monte Carlo study of the selectivity of calcium channels: improved geometry. *Mol. Phys.* 100:2361–2368.
- Gillespie, D., L. Xu, ..., G. Meissner. 2005. (De)constructing the ryanodine receptor: modeling ion permeation and selectivity of the calcium release channel. *J. Phys. Chem. B.* 109:15598–15610.
- Boda, D., W. Nonner, ..., D. Gillespie. 2008. Volume exclusion in calcium selective channels. *Biophys. J.* 94:3486–3496.
- Gillespie, D., W. Nonner, and R. S. Eisenberg. 2002. Coupling Poisson-Nernst-Planck and density functional theory to calculate ion flux. *J. Phys. Condens. Matter.* 14:12129–12145.
- Boda, D., M. Valiskó, ..., D. Gillespie. 2006. The effect of protein dielectric coefficient on the ionic selectivity of a calcium channel. *J. Chem. Phys.* 125:34901.
- Boda, D., M. Valiskó, ..., D. Gillespie. 2007. Combined effect of pore radius and protein dielectric coefficient on the selectivity of a calcium channel. *Phys. Rev. Lett.* 98:168102.
- Boda, D., W. Nonner, ..., D. Gillespie. 2007. Steric selectivity in Na channels arising from protein polarization and mobile side chains. *Biophys. J.* 93:1960–1980.
- Gillespie, D., J. Giri, and M. Fill. 2009. Reinterpreting the anomalous mole fraction effect: the ryanodine receptor case study. *Biophys. J.* 97:2212–2221.

44. Gillespie, D., H. Chen, and M. Fill. 2012. Is ryanodine receptor a calcium or magnesium channel? Roles of K^+ and Mg^{2+} during Ca^{2+} release. *Cell Calcium*. 51:427–433.
45. Evans, R. 1992. Density functionals in the theory of nonuniform fluids. In *Fundamentals of Inhomogeneous Fluids*. D. Henderson, editor. Marcel Dekker, New York, pp. 85–176.
46. Nonner, W., and B. Eisenberg. 1998. Ion permeation and glutamate residues linked by Poisson-Nernst-Planck theory in L-type calcium channels. *Biophys. J.* 75:1287–1305.
47. Gillespie, D. 1999. A Singular Perturbation Analysis of the Poisson-Nernst-Planck System: Applications to Ionic Channels. Rush University, Chicago, IL.
48. Gillespie, D., W. Nonner, and R. S. Eisenberg. 2003. Density functional theory of charged, hard-sphere fluids. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 68:031503.
49. Chen, D., J. Lear, and B. Eisenberg. 1997. Permeation through an open channel: Poisson-Nernst-Planck theory of a synthetic ionic channel. *Biophys. J.* 72:97–116.
50. Kurnikova, M. G., R. D. Coalson, ..., A. Nitzan. 1999. A lattice relaxation algorithm for three-dimensional Poisson-Nernst-Planck theory with application to ion transport through the gramicidin A channel. *Biophys. J.* 76:642–656.
51. Barthel, J. M. G., H. Krienke, and W. Kunz. 1998. *Physical Chemistry of Electrolyte Solutions: Modern Aspects*. Springer, New York.
52. Fawcett, W. R. 1999. Thermodynamic parameters for the solvation of monatomic ions in water. *J. Phys. Chem. B.* 103:11181–11185.
53. Lipkind, G. M., and H. A. Fozzard. 2000. KcsA crystal structure as framework for a molecular model of the $Na(+)$ channel pore. *Biochemistry*. 39:8161–8170.
54. Boda, D., J. Giri, ..., D. Gillespie. 2011. Analyzing the components of the free-energy landscape in a calcium selective ion channel by Widom's particle insertion method. *J. Chem. Phys.* 134:055102.
55. Ludtke, S. J., and I. I. Serysheva. 2013. Single-particle cryo-EM of calcium release channels: structural validation. *Curr. Opin. Struct. Biol.* 23:755–762.
56. Lobo, P. A., and F. Van Petegem. 2009. Crystal structures of the N-terminal domains of cardiac and skeletal muscle ryanodine receptors: insights into disease mutations. *Structure*. 17:1505–1514.
57. Tung, C.-C., P. A. Lobo, ..., F. Van Petegem. 2010. The amino-terminal disease hotspot of ryanodine receptors forms a cytoplasmic vestibule. *Nature*. 468:585–588.
58. Kimlicka, L., K. Lau, ..., F. Van Petegem. 2013. Disease mutations in the ryanodine receptor N-terminal region couple to a mobile intersubunit interface. *Nat. Commun.* 4:1506.
59. Shirvanyants, D., S. Ramachandran, ..., N. V. Dokholyan. 2014. Pore dynamics and conductance of RyR1 transmembrane domain. *Biophys. J.* 106:2375–2384.
60. Corry, B., S. Kuyucak, and S.-H. Chung. 2000. Tests of continuum theories as models of ion channels. II. Poisson-Nernst-Planck theory versus Brownian dynamics. *Biophys. J.* 78:2364–2381.
61. Song, C., and B. Corry. 2011. Testing the applicability of Nernst-Planck theory in ion channels: comparisons with Brownian dynamics simulations. *PLoS ONE*. 6:e21204.
62. Gillespie, D., M. Valiskó, and D. Boda. 2005. Density functional theory of the electrical double layer: the RFD functional. *J. Phys. Condens. Matter*. 17:6609–6626.
63. Valiskó, M., D. Boda, and D. Gillespie. 2007. Selective adsorption of ions with different diameter and valence at highly-charged interfaces. *J. Phys. Chem. C*. 111:15575–15585.
64. Boda, D., R. Kovács, ..., T. Kristóf. 2014. Selective transport through a model calcium channel studied by local equilibrium Monte Carlo simulations coupled to the Nernst-Planck equation. *J. Mol. Liq.* 189:100–112.
65. Boda, D., É. Csányi, ..., T. Kristóf. 2014. Dynamic Monte Carlo simulation of coupled transport through a narrow multiply-occupied pore. *J. Chem. Phys.* 118:700–707.
66. Berti, C., S. Furini, ..., C. Fiegna. 2014. Three-dimensional Brownian dynamics simulator for the study of ion permeation through membrane pores. *J. Chem. Theory Comput.* 10:2911–2926.
67. Boda, D., D. Henderson, and D. Gillespie. 2013. The role of solvation in the binding selectivity of the L-type calcium channel. *J. Chem. Phys.* 139:055103.
68. Shannon, R. D., and C. T. Prewitt. 1969. Effective ionic radii in oxides and fluorides. *Acta Crystallogr. B.* 25:925–946.
69. Nielaba, P., and F. Forstmann. 1985. Packing of ions near an electrolyte-electrode interface in the HNC/LMSA approximation to the RPM model. *Chem. Phys. Lett.* 117:46–48.
70. Lamperski, S., and C. W. Outhwaite. 2002. Exclusion volume term in the inhomogeneous Poisson-Boltzmann theory for high surface charge. *Langmuir*. 18:3423–3424.
71. Lamperski, S., and L. B. Bhuiyan. 2003. Counterion layering at high surface charge in an electric double layer. Effect of local concentration approximation. *J. Electroanal. Chem.* 540:79–87.
72. Hille, B. 2001. *Ion Channels of Excitable Membranes*. Sinauer Associates, Sunderland, MA.
73. Chen, D. P., L. Xu, ..., B. Eisenberg. 1999. Selectivity and permeation in calcium release channel of cardiac muscle: alkali metal ions. *Biophys. J.* 76:1346–1366.
74. Krauss, D., B. Eisenberg, and D. Gillespie. 2011. Selectivity sequences in a model calcium channel: role of electrostatic field strength. *Eur. Biophys. J.* 40:775–782.

UNCOVER NEW INSIGHTS TO HELP YOU IDENTIFY POTENTIAL BIOLOGICAL TARGETS

Target Insights is a support tool that helps users find important facts that other search tools may miss, because it is based upon an automated analysis of the full text of scientific articles. Target Insights gets you to the right target quicker and with more confidence than other tools.

Developed by Elsevier to meet increasing demands on speed and efficiency it enables target identification and validation teams to discover new insights and update their understanding of the disease model by revealing crucial information which is normally obscured and hard-to-find, in three steps... [Identify](#), [Prioritize](#), [Validate](#).



For the latest information visit
targetinsights.com/info

Copyright © 2012 Elsevier Properties SA. All rights reserved.
Target Insights™ is owned by Elsevier Properties SA and used under license.



Karyopherin-Centric Control of Nuclear Pores Based on Molecular Occupancy and Kinetic Analysis of Multivalent Binding with FG Nucleoporins

Larisa E. Kapinos, Rafael L. Schoch, Raphael S. Wagner, Kai D. Schleicher, and Roderick Y. H. Lim*
Biozentrum and the Swiss Nanoscience Institute, University of Basel, Basel, Switzerland

ABSTRACT Intrinsically disordered Phe-Gly nucleoporins (FG Nups) within nuclear pore complexes exert multivalent interactions with transport receptors (Karyopherins (Kaps)) that orchestrate nucleocytoplasmic transport. Current FG-centric views reason that selective Kap translocation is promoted by alterations in the barrier-like FG Nup conformations. However, the strong binding of Kaps with the FG Nups due to avidity contradicts rapid Kap translocation *in vivo*. Here, using surface plasmon resonance, we innovate a means to correlate *in situ* mechanistic (molecular occupancy and conformational changes) with equilibrium (binding affinity) and kinetic (multivalent binding kinetics) aspects of Karyopherin β 1 (Kap β 1) binding to four different FG Nups. A general feature of the FxFG domains of Nup214, Nup62, and Nup153 is their capacity to extend and accommodate large numbers of Kap β 1 molecules at physiological Kap β 1 concentrations. A notable exception is the GLFG domain of Nup98, which forms a partially penetrable cohesive layer. Interestingly, we find that a slowly exchanging Kap β 1 phase forms an integral constituent within the FG Nups that coexists with a fast phase, which dominates transport kinetics due to limited binding with the pre-occupied FG Nups at physiological Kap β 1 concentrations. Altogether, our data reveal an emergent Kap-centric barrier mechanism that may underlie mechanistic and kinetic control in the nuclear pore complex.

INTRODUCTION

Nuclear pore complexes (NPCs) perforate the nuclear envelope that separates the nucleus and cytoplasm in eukaryotic cells (1–3). Each NPC facilitates the continuous bidirectional exchange of specific cargoes for maintaining cellular order and function. Despite its putative ~50 nm diameter (2), the upper limit for non-signal-mediated passive transport through the NPC is ~40 kDa (4). Thus, small molecules diffuse freely through the NPC, whereas macromolecules >~5 nm in size are withheld (5). Remarkably, soluble transport receptors (6), such as the 97 kDa import receptor karyopherin β 1 (Kap β 1), also known as importin β (7), gain rapid and exclusive NPC access despite exceeding the passive limit. On this basis, nucleocytoplasmic transport is orchestrated by Kaps that identify and shuttle signal-specific cargoes from the complex biological milieu (sometimes using Kap α /importin α as an adaptor) through NPCs (8). In the absence of Kaps, the exquisite selectivity of the NPC is demonstrated in the rejection of even signal-specific cargoes, which are smaller entities than entire Kap-cargo complexes (9). Still, because the size of a legitimate Kap-cargo complex far exceeds the passive transport limit, it is generally accepted that a molecular gating mechanism alleviates spatial constraints and underlies NPC functionality and transport control (10).

Located within the NPC interior are 11 distinct nucleoporins (Nups) that bear large numbers of phenylalanine-

glycine (FG)-repeat motifs (FG Nups) (11). Current estimates indicate that a total of ~200 FG Nups circumscribe the entire central channel in multiples of eight from the cytoplasmic periphery to the central plane to the distal ring of the nuclear basket. The FG Nups are tethered to the inner walls of the NPC by anchor domains from which FG-rich domains emanate to occupy the aqueous space within the central channel (12). The FG domains are large intrinsically disordered polypeptides (13) that are fundamental to the NPC gating mechanism for two apparent reasons: 1), the FG-repeat motifs exert binding interactions with Kaps (7,14–16); and 2), their collective barrier-forming properties exclude passive molecules (10). The FG domains can be categorized by their FG-repeat motifs (i.e., GLFG, FxFG, and FG) (17), hydrophobicity, and overall net charge (18–21). Solution biochemical analyses, as well as *in vivo* studies, generally show that FxFG domains exhibit noncohesive properties (17,22). As an example, the surface-tethered FxFG domains of Nup153 and Nup62 form extensible brush-like layers that support this view (23,24). On the other hand, GLFG domains are more cohesive (17,22). Subsequent studies show that both FG-domain types can cohere into macroscopic hydrogels under nonphysiological conditions (25–28).

Although it remains a formidable problem to visualize FG-domain morphology inside the NPC, the contrasting properties of the FG domains largely dominate the basis of mechanistic FG-barrier-centric models. Given the notion that a lack of binding implies NPC rejection, sufficient Kap-FG binding is thought to cause a transient breach or opening in the FG-domain barrier to make space for translocation to

Submitted December 3, 2013, and accepted for publication February 12, 2014.

*Correspondence: roderick.lim@unibas.ch

Larisa E. Kapinos and Rafael L. Schoch contributed equally to this work.

Editor: Daniel Muller.

© 2014 by the Biophysical Society
0006-3495/14/04/1751/12 \$2.00



proceed. The selective phase model derives from the characteristics of macroscopic FG hydrogels whereby the FG domains form a sieve-like meshwork that only Kaps can dissolve or melt through (25–28). The virtual gating/polymer brush model is based on the brush-like behavior of surface-tethered FG domains that entropically exclude nonspecific cargoes (3,23,29) while promoting Kap access by ‘reversibly collapsing’ (30). On this basis, it has been postulated that Kap-cargo complexes diffuse on a hydrophobic FG-rich layer of permanently collapsed FG domains that coat the NPC walls, also referred to as ‘reduction of dimensionality’ (31,32). Finally the two-gate/forest model describes how inter- and intra-FG-domain cohesion together with other noncohesive regions might define a particular barrier arrangement that demarcates distinct zones of traffic through the NPC (17,21).

Further details regarding possible NPC-barrier entry mechanisms are sparse. Each Kap β 1 molecule consists of ~10 hydrophobic grooves on its outer HEAT-repeat surface that can all potentially bind FG repeats (7,33,34). Given that the number of FG repeats per FG domain also varies from 5 to ~50 in vertebrates, Kap-FG-domain binding involves highly multivalent interactions (10), which are generally known to impart a strong avidity that enhances stability and specificity (35). However, this is paradoxical in the context of the NPC (36), because the high submicromolar Kap β 1-FG-domain binding affinities (33,37,38) predict slow off rates (given a diffusion-limited on rate) that contradict the rapid (~5 ms) in vivo dwell time (39). As this implies, Kap-FG binding ought to be sufficiently strong to ensure selectivity but also weak enough to promote fast translocation through the NPC. Nonetheless, an explanation as to how Kap-FG binding kinetics is balanced against the mechanistic control of the FG-domain barrier is still lacking.

In this work, we use surface plasmon resonance (SPR) to compare and contrast Kap β 1 binding to the FG domains from different NPC sites: the cytoplasmic periphery (Nup214) (40), the central channel (Nup62 and Nup98) (41,42), and the nuclear basket (Nup153) (43). At high surface densities, Nup214, Nup62, and Nup153 are able to form extended molecular brushes, whereas a lack of extensibility in Nup98 indicates that it forms a single cohesive surface layer. We find that Nup214, Nup62, and Nup153 accommodate Kap β 1 with a high molecular occupancy at physiological concentrations (44) (i.e., 4–20 μ M Kap β 1). In comparison, Nup98 is only partially penetrable to Kap β 1. Finally, we implement a kinetic analysis based on the regularization method of Svitel et al. (45,46) that uses singular value decomposition to extract the effective on rates and off rates associated with heterogeneous multivalent binding. These results are complemented by equilibrium analyses, which reveal that Kap β 1-FG-domain binding is characterized by high- and low-affinity phases that vary depending on Kap β 1 occupancy within the FG-domain layer. This suggests that a Kap-centric rather than an FG-

domain-centric barrier mechanism regulates transport selectivity and speed through the NPC.

METHODS

Cloning and expression of recombinant proteins

Full-length human Kap β 1 was cloned, expressed, and purified as described previously (47). Protein purity was analyzed by 12% polyacrylamide gel electrophoresis at 0.1% sodium dodecyl sulfate and Kap β 1 concentration was determined by measuring its absorption at 280 nm. The FG domains of human cNup153, cNup98, and cNup62 were cloned, expressed, and purified as described before (24,47). A construct containing the FG domain of cNup214 in pETM-11 was kindly provided by B. Fahrenkrog. To allow attachment of cNup214 to the Au surface, the first amino acid preceding the fragment’s original sequence was mutated into cysteine by site-directed mutagenesis using the primers

5'-TTT CAG GGC GCC ATG TGT ATG AGT CCT GGC TTT
(primer 1)

and

3'-AAA GTC CCG CGG TAC ACA TAC TCA GGA CCG AAA
(primer 2).

cNup214 was then expressed and purified as described (24). The final purity of the His₆-tag-free FG domains was analyzed by 12% polyacrylamide gel electrophoresis at 0.1% sodium dodecyl sulfate (Supporting Material).

Dynamic light-scattering measurements

The hydrodynamic radii of the purified proteins were measured by dynamic light scattering (Zetasizer Nano, Malvern Instruments Ltd, Worcestershire, United Kingdom), as previously described (24).

SPR measurements

All SPR measurements were performed at 25°C in phosphate-buffered saline (PBS) at pH7.2 (GIBCO, Life Technologies, Carlsbad, CA) in a four-flow-cell Biacore T100 instrument (GE Healthcare Life Sciences/Biacore, Uppsala, Sweden) as described before (24).

Kinetic analysis of multivalent interactions

Binding data were analyzed using a model that calculates a discrete distribution of kinetic states ($k_{on,i}$, $k_{off,i}$), as introduced by Svitel et al. (46). For the interaction maps, we populated a set of 36×36 ($k_{on,i}$, $k_{off,i}$) pairs and constructed k_{on} and k_{off} versus K_D maps, where the color intensity corresponds to the fractional abundance. The accompanying histograms are summed over the respective axis values. Tikhonov-regularized solutions were obtained using the Regularization Tools package by Per Christian Hansen (48) and an active set method was applied to provide nonnegativity (49). All calculations and visualizations were performed using Matlab (MathWorks, Natick, MA). See the Supporting Material for details.

RESULTS

SPR characterization of surface-tethered FG domains

Fig. 1 superimposes the known FG-domain anchoring sites of Nup214, Nup62, Nup98, and Nup153 at the cytoplasmic

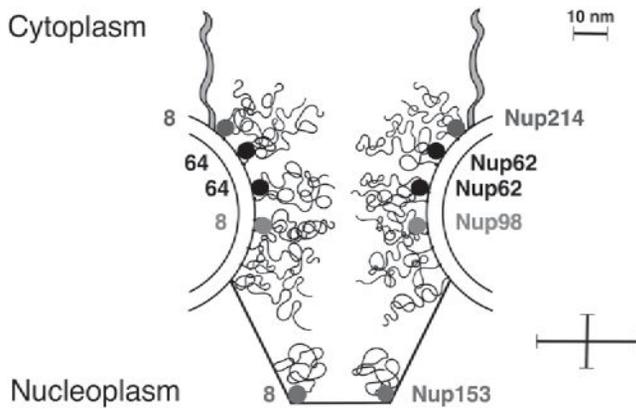


FIGURE 1 Surface-tethered FG domains within the NPC. Estimated FG-domain copy numbers are shown with their known anchoring sites within the NPC central channel, as determined from *Xenopus laevis* oocytes. Error bars denote the uncertainty in their exact locations. Scale bar, 10 nm. To see this figure in color, go online.

periphery, central channel, and distal ring, respectively, on an outline of the NPC obtained from *Xenopus laevis* oocytes (50). This accounts for their estimated copy numbers based on the eightfold rotational symmetry of the NPC (51,52), including the uncertainty in their relative positions based on immunogold localization studies (40–43), which is $\pm \sim 15$ nm in the NPC axis and ± 20 nm in the radial axis. Needless to say, it remains difficult to calculate exactly how closely spaced (i.e., axially and circumferentially) the FG domains are in the NPC (53), which in part is the motivation for this work. Recent evidence suggests that close packing is expected within the central channel for Nup62 based on the presence of 128 copies per NPC (52) (see Fig. 1). The cysteine-modified FG-domain constructs used in this study are Nup214 (amino acids (aa) 1809–2090; partial FxFG domain), Nup62 (aa 1–240; full-length FxFG domain), Nup98 (aa 1–498; full-length GLFG domain), and Nup153 (aa 874–1475; full-length FxFG domain); for details, see the Supporting Material. For clarity, these are termed cNup214, cNup62, cNup98, and cNup153, respectively. From dynamic light-scattering analyses, their measured hydrodynamic radii (r_h) are 3.4 ± 1.5 nm (cNup214), 3.7 ± 1.7 nm (cNup62), 5.6 ± 1.6 nm (cNup98), and 5.1 ± 3.2 nm (cNup153).

Briefly, our SPR experiments use bovine serum albumin (BSA) molecules, which act as inert noninteracting probes that naturally feel the intrinsic exclusion volume of the FG domains (i.e., their thickness) (Fig. 2 A). A comprehensive description of the method and related calculations can be found in studies by Schoch and colleagues (24,54). A first BSA injection provides the initial FG-domain-layer thickness (d_{cNup} ; Fig. 2 B), using the expression

$$d_{cNup} = \frac{l_d}{2} \ln \left(\frac{R_{ref} m_{BSA}}{R_{BSA} m_{ref}} \right) + d_{ref}, \quad (1)$$

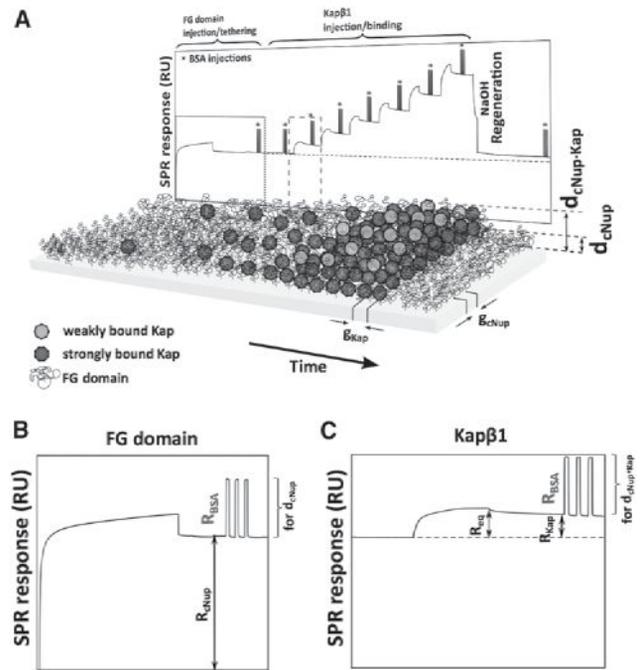


FIGURE 2 SPR methodology and measured parameters. (A) A typical experimental sequence starts with the surface tethering of the FG domains followed by titrations of increasing Kap β 1 concentrations. Triple BSA injections (*) are used to determine FG-domain layer thickness before (d_{cNup}) and after ($d_{cNup-Kap}$) each Kap β 1 injection. An increase in RU corresponds to increased Kap β 1-FG-domain binding, as depicted. A terminal NaOH regeneration step ensures that Kap β 1 is biochemically bound to the FG domains. (B) Zoom-in of the dotted box in A. R_{cNup} corresponds to the surface-tethered FG domains and is used to calculate the inter-FG-domain grafting distance (g_{cNup}). d_{cNup} is calculated from the BSA response (R_{BSA}). (C) Zoom-in of the dashed box in A. R_{eq} corresponds to the Kap β 1-FG-domain binding equilibrium at each respective Kap β 1 concentration. R_{Kap} is used to extract the next-neighbor distance of bound Kap β 1 molecules (g_{Kap}) before the BSA injections (R_{BSA}), which is then correlated to $d_{cNup-Kap}$. To see this figure in color, go online.

where $l_d = 350$ nm is the characteristic evanescent field decay length; R_{BSA} is the BSA-SPR response from the FG-domain measurement cell, where m_{BSA} is a calibration constant; R_{ref} is the BSA-SPR response for a reference cell (not shown), where m_{ref} is a calibration constant; and $d_{ref} = 2$ nm is the thickness of a 1-mercapto-11-undecyltetra(ethyleneglycol) (i.e., HS-(CH₂)₁₁-(OCH₂CH₂)₃-OH) passivating layer in the reference cell. To calculate the inter-FG-domain grafting distance from the SPR response (R_{cNup}), the expression (24)

$$g_{cNup} = \sqrt{\frac{1300 \times M_W \times 10^{21}}{N_A \times R_{cNup}}} \quad (2)$$

is applied, where M_W is the molecular weight of the FG domain and N_A is Avogadro's constant. In this manner, we are able to determine how the grafting distance between FG domains (g_{cNup}) affects d_{cNup} .

Once d_{cNup} and g_{cNup} are established, subsequent BSA injections are used to monitor in situ changes in the layer thickness as caused by Kap β 1-FG binding interactions at increasing Kap β 1 concentrations. This is defined as $d_{cNup \cdot Kap}$ and follows directly from Eq. 1 (replacing d_{cNup}) by measuring R_{BSA} after each consecutive Kap β 1 injection (Fig. 2 C). Likewise, the average distance between bound Kap β 1 molecules, g_{Kap} , at each respective Kap β 1 concentration can be obtained using Eq. 2 (substituting R_{Kap} for R_{cNup}).

Conformational characterization of surface-tethered FG domains

Fig. 3 summarizes the dependence of d_{cNup} on g_{cNup} . A general feature of all four FG domains is that their layer thicknesses increase as g_{cNup} decreases. For brevity, our analysis considers two regimes: 1), close-packed, where $g_{cNup} < r_h$; and 2), sparse, where $g_{cNup} > r_h$. In the close-packed regime, the FG domains have a tendency to form extended molecular brushes that obey the scaling behavior of polyelectrolytic intrinsically disordered proteins (55). By definition, brush formation occurs when surface-tethered polymeric chains stretch away from their anchoring sites due to lateral crowding. This does not preclude the existence of intra- or inter-FG-domain interactions, which depend on the intrinsic physicochemical properties of each FG domain (e.g., hydrophobicity and charge; see Table S1 in the Supporting Material). At their smallest obtainable grafting distances, the noncohesive FxFG domains of cNup214, cNup62, and

cNup153 form molecular brushes which reach maximal thicknesses of $d_{cNup} \sim 20$ nm in the close-packed regime. In contrast, cNup98 exhibits a limited extension in the close-packed regime despite its relatively large amino acid composition (498 aa) and r_h , which comes closest to that of cNup153 (602 aa). Yet cNup98 exhibits a maximum thickness of ~ 11 nm when $g_{cNup} = 4$ nm, which is approximately half the thickness of cNup153 at the same grafting distance. This indicates that cNup98 is inherently more compact and cohesive compared to the other FxFG domains (22,28).

Molecular occupancy of Kap β 1 bound within close-packed FG-domain layers

Subsequent changes in close-packed FG-domain thickness due to Kap β 1 binding ($d_{cNup \cdot Kap}$) can be correlated to the relative arrangement of Kap β 1 molecules bound within the layer (Supporting Material). This was monitored by titrating Kap β 1 in the sequence 0.1, 0.5, 1.0, 10, 20, 30, 40, 62.5, and 125 nM, and 0.25, 0.5, 1, 2, 4, 6 to 13.6 μ M. Fig. 4 summarizes the Kap β 1-binding response across all experiments for each close-packed FG domain (5–10 experiments per FG domain), which we describe in three ways. First, the relative change in layer thickness ($\Delta d_{cNup \cdot Kap}/d_{cNup}$) is shown as a function of surface density of bound Kap β 1 ($\rho_{Kap\beta 1}$), which is related to the number of Kap β 1 layers formed (Fig. 4 A). This relation is given as 2200 RU or 1000 Da/nm², based on the amount of material that corresponds to the equivalent of one (net) Kap β 1 layer (24). Second, the change in the total protein mass density within the layer ($\Delta \nu_{total}$ in Da/nm³, which accounts for both Kap β 1 and the FG-domain mass per unit volume) is plotted as a function of $\Delta d_{cNup \cdot Kap}/d_{cNup}$ (Fig. 4 B) (see the Supporting Material). Based on this plot, three different conformational responses can be distinguished: compaction (I), where $\Delta \nu_{total}$ is positive and $\Delta d_{cNup \cdot Kap}/d_{cNup}$ is negative; compact extension (II), where $\Delta \nu_{total}$ and $\Delta d_{cNup \cdot Kap}/d_{cNup}$ are positive; and porous extension (III), where $\Delta \nu_{total}$ is negative and $\Delta d_{cNup \cdot Kap}/d_{cNup}$ is positive (i.e., the layer becomes more porous). Third, the representative level of molecular occupancy of Kap β 1 that is reached is illustrated within each close-packed FG-domain layer at the highest applied Kap β 1 concentrations (Fig. 4 C).

Overall, each FG domain exhibits its own characteristic response upon binding Kap β 1. cNup214 almost doubles its initial thickness value ($\Delta d_{cNup214 \cdot Kap}/d_{cNup214} = 0.8$) at the point where one Kap β 1 layer is bound. From here, Kap β 1 occupancy increases up to 2.5 layers (at maximum titration) without any further increase in thickness. We find from $\Delta \nu_{total}$ that cNup214 reaches a maximum porous extension after initial Kap β 1 binding followed by a filling of the layer as subsequent Kap β 1 molecules bind. In comparison, cNup62 collapses into a more compact layer upon initial Kap β 1 binding, as indicated by the decrease in

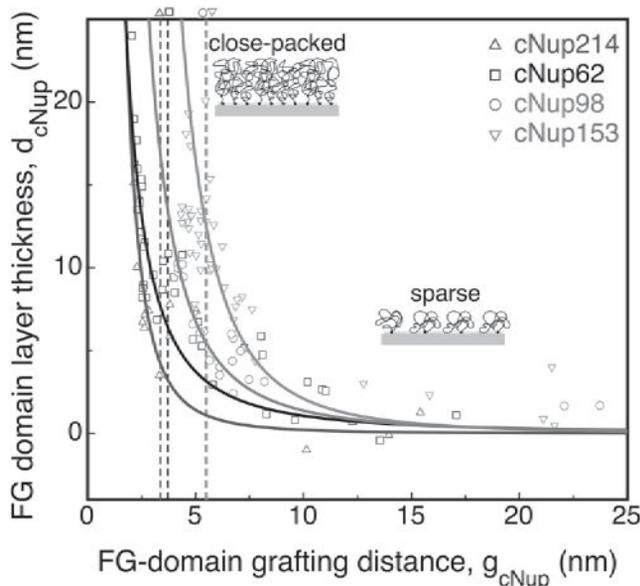


FIGURE 3 Dependence of layer thickness (d_{cNup}) on FG-domain grafting distance (g_{cNup}) for cNup214, cNup62, cNup98, and cNup153. Color-coded dashed vertical lines labeled by their respective symbols correspond to the measured hydrodynamic radii (r_h) of the FG domains (see Methods). To see this figure in color, go online.

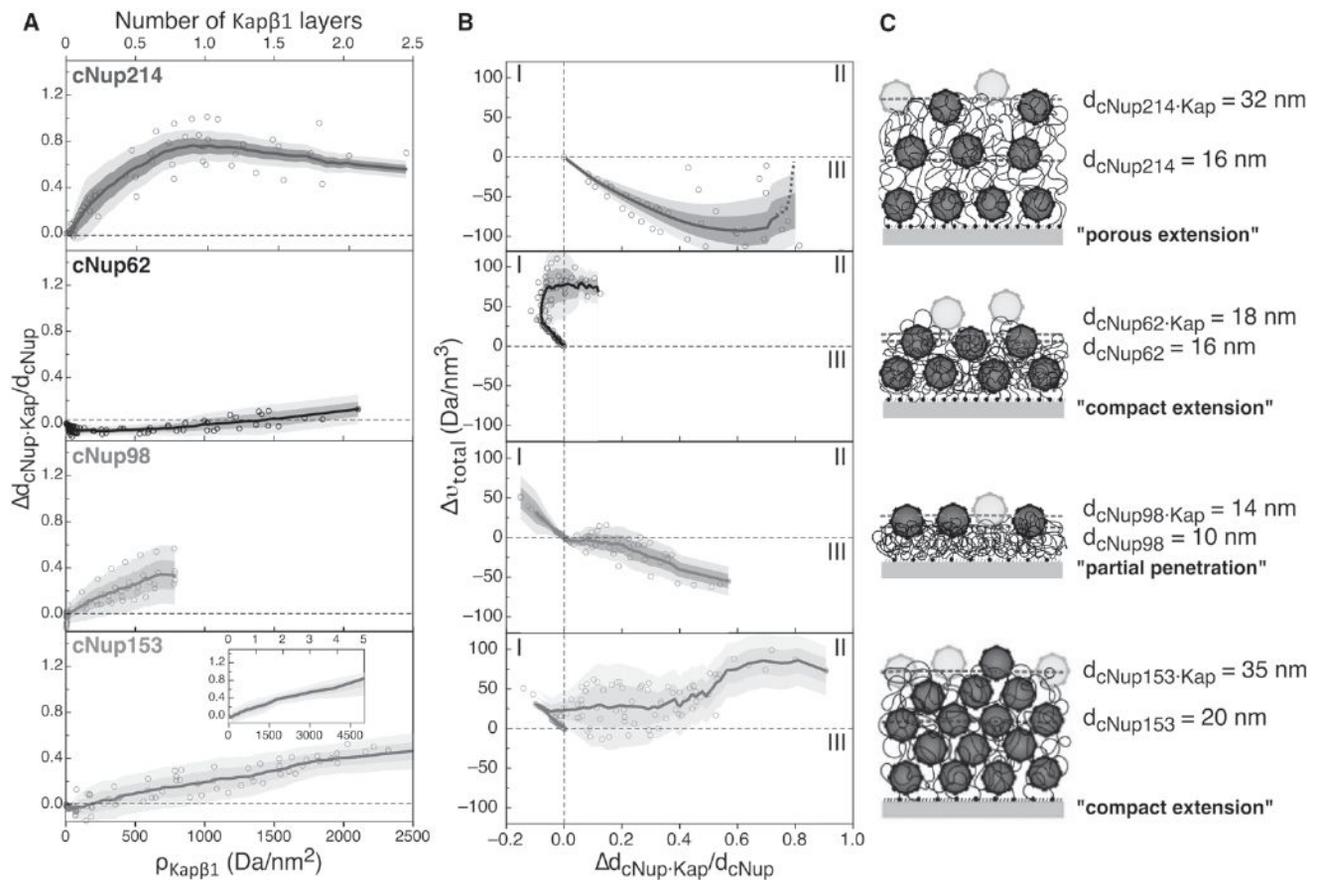


FIGURE 4 Different close-packed FG domains respond to Kap β 1-binding differently. (A) Relative changes in FG-domain layer thickness ($\Delta d_{cNup \cdot Kap} / d_{cNup}$) plotted as a function of Kap β 1 surface density ($\rho_{Kap\beta 1}$) for cNup214, cNup62, cNup98, and cNup153. $\rho_{Kap\beta 1} = 1000$ Da/nm 2 for a single Kap β 1 layer. The data account for the full range of Kap β 1 injections (0–13.6 μ M) except in the case of cNup153, where data for the full range of injections are shown in the inset. (B) Corresponding changes in total protein density (Kap β 1 and FG domains; $\Delta \nu_{total}$) plotted as a function of $\Delta d_{cNup \cdot Kap} / d_{cNup}$. The three characteristic responses are compaction (I), compact extension (II), and porous extension (III). See text for details. (C) Sketch of Kap β 1 occupancy within each FG-domain layer before BSA injection (dark green) and at equilibrium in the presence of physiological Kap β 1 concentration (light and dark green). Values d_{cNup} and $d_{cNup \cdot Kap}$ highlight the change in thickness before and after Kap β 1 binding. Note that the dark and light shaded areas in A and B correspond to 1 and 2 standard deviations (SD), respectively. To see this figure in color, go online.

$\Delta d_{cNup62 \cdot Kap} / d_{cNup62}$ with increasing $\Delta \nu_{total}$. This is followed by a recovery phase and a compact extension that accommodates a maximum of two Kap β 1 layers (24). In a similar way, cNup153 first undergoes a reversible collapse transition (30) (Kap β 1 < 60 nM) before exhibiting compact extension at higher Kap β 1 concentrations. Indeed, this reaches $1.75 d_{cNup153}$ (~35 nm), which accommodates up to five Kap β 1 layers at 13.6 μ M Kap β 1.

Interestingly, cNup98 undergoes a compaction at very low Kap β 1 concentrations (<30 nM) followed by a small increase in thickness, incorporating less than one Kap β 1 layer at physiological Kap β 1 concentration. Given that $d_{cNup98 \cdot Kap}$ reaches a maximum of 14 nm and that Kap β 1 can be approximated as a sphere with a diameter of 9.9 nm (24), the decrease in $\Delta \nu_{total}$ likely results from a partial (i.e., incomplete) penetration of Kap β 1 into the compact cNup98 layer, which cannot extend further due to intrinsic cohesion. Based on the above analysis, the differing extensibilities upon Kap β 1 binding of each close-packed FG

domain are given as cNup214 > cNup153 > cNup62 > cNup98.

Molecular occupancy of Kap β 1 bound to sparse FG-domain layers

The thickness of a sparse FG-domain layer is underestimated due to an interdigitation or penetration by BSA molecules into the gaps or spaces between individual FG-domain mushrooms. Instead, we compare g_{cNup} to g_{Kap} , which shows that the maximal Kap β 1 occupancy does not exceed one Kap β 1 bound per FG-domain molecule (Supporting Material).

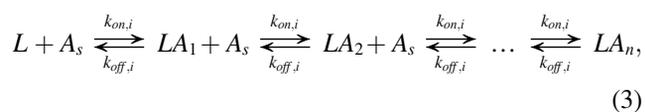
Kap β 1 binding avidity depends on FG-domain surface density

In terms of Kap β 1-FG-domain binding equilibrium, the results of Langmuir isotherm analyses vary widely, depending

on FG-domain surface density (Fig. 5 A). The resulting equilibrium dissociation constants (K_D) (Fig. 5 B) show that Kap β 1 binds moderately at $\sim 1 \mu\text{M}$ K_D for all FG domains in a sparse configuration where a single isolated FG-domain molecule effectively binds one Kap β 1 molecule (see previous section). In comparison, the experimental Kap β 1 binding data obtained in the close-packed FG-domain regime requires a two-component Langmuir isotherm fit (Supporting Material). Here, a high-affinity species ($K_D \sim 100 \text{ nM}$ to $1 \mu\text{M}$) represents tight Kap β 1 binding given the high FG-repeat density in each close-packed FG-domain layer. The increase in Kap β 1 occupancy at higher concentrations leads to a reduction of free FG repeats within the layer, which results in a second, low-affinity species ($K_D \sim 10 \mu\text{M}$). It is important to note that this increase in K_D represents a general hallmark of binding avidity that reflects a reduction of multivalent interactions when a close-packed FG-domain layer becomes saturated with Kap β 1. At that stage, fewer FG repeats are available, and subsequent Kap β 1 binding becomes weak as avidity is diminished.

Kinetic analysis of multivalent Kap β 1 binding to close-packed FG domains

To obtain a parsimonious distribution of kinetic constants, we applied the analysis introduced by Svitel et al. (45,46), which uses a two-dimensional distribution of association and dissociation rate constants ($k_{on,i}$, $k_{off,i}$) to describe Kap β 1-FG-domain binding as a superposition of pseudo-first-order reactions (Supporting Material). Briefly, this circumvents difficulties associated with analyzing heterogeneous interfacial interactions that are often encountered in multivalent systems (56–58). In the context of this study, Kap β 1 binding depends on the FG-domain surface density and the number of Kap β 1 molecules already bound. Here, k_{on} describes how quickly a Kap β 1 molecule locates and binds to FG repeats within an FG-domain layer, whereas k_{off} correlates to Kap β 1-FG-domain binding strength and stability. Rather than modeling the binding of Kap β 1 to individual FG repeats, the analysis (Fig. 6 A) considers a two-dimensional lattice containing $10 \times 10 \text{ nm}^2$ surface sites (given the 10 nm size of Kap β 1 (24)). Each surface site is sticky (in analogy to containing FG domains), and allows for analyte (Kap β 1) molecules to bind, as kinetically modeled by



where L denotes the empty surface sites, A_s is the analyte concentration in solution, and LA_j to LA_n correspond to surface sites where n represents the number of analytes that can bind per surface site. Here, we define $n = 3, 3, 1,$ and 3 for

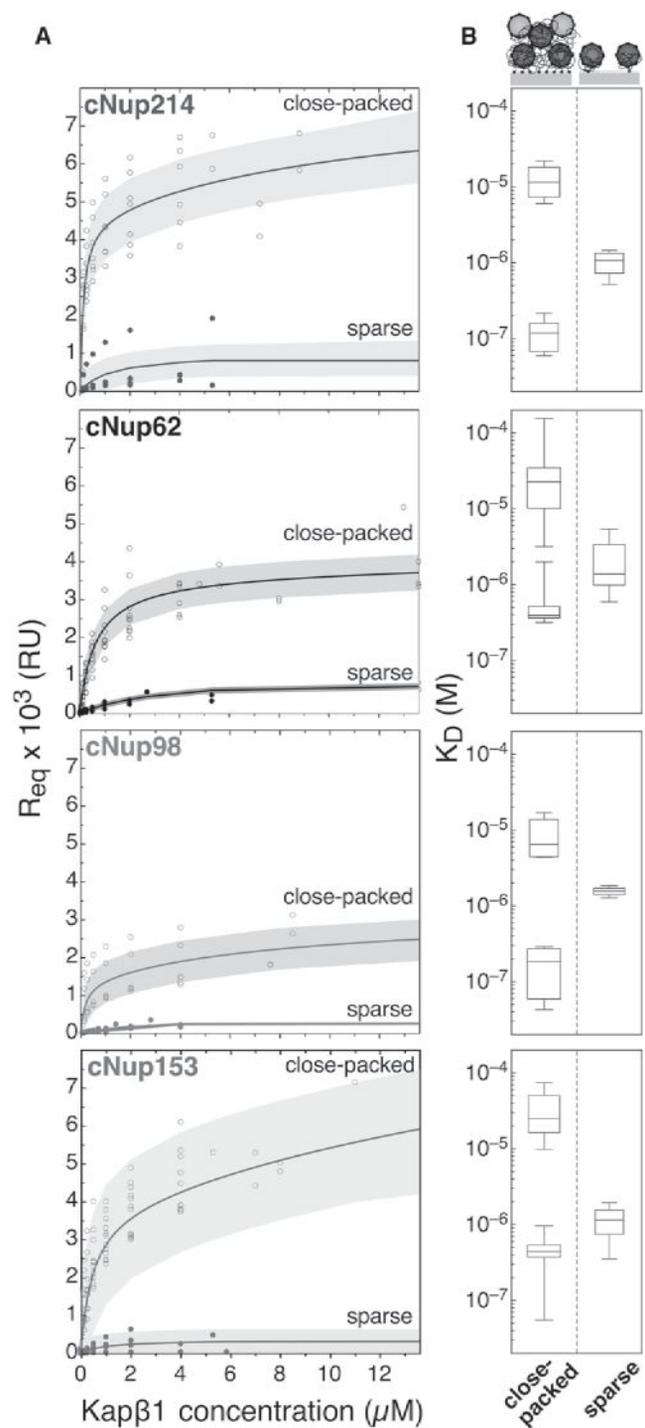


FIGURE 5 Equilibrium analysis of Kap β 1-FG-domain binding. (A) Dependence of Kap β 1-FG-domain equilibrium binding response (R_{eq}) on the bulk Kap β 1 concentration for cNup214, cNup62, cNup98, and cNup153 in the close-packed ($g_{cNup} < r_h$) and sparse ($g_{cNup} > 2r_h$) regimes. Solid lines represent a two-component Langmuir isotherm fit in the close-packed regime and a one-component Langmuir isotherm fit in the sparse regime. The shaded area corresponds to 1 SD in both cases. (B) Equilibrium dissociation constants (K_D) obtained from A in the close-packed and sparse regimes. Box plots show the median and first and third quartiles (values >6 SD are considered to be outliers and are not shown). To see this figure in color, go online.

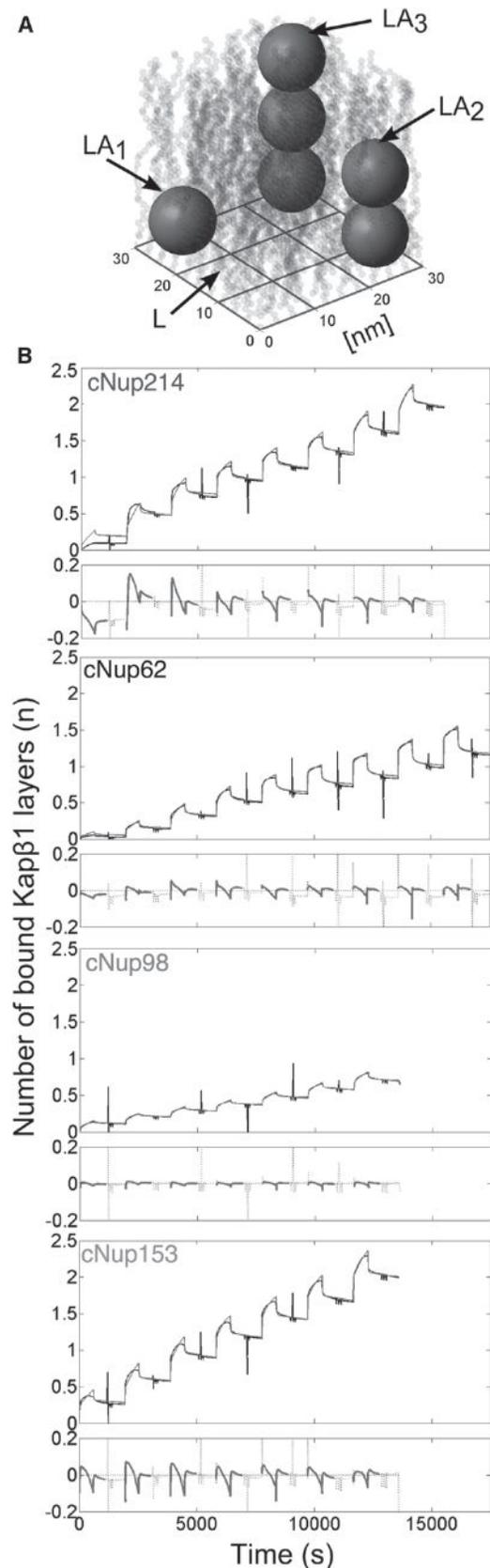


FIGURE 6 Kinetic analysis of Kap β 1-FG-domain binding. (A) Schematic representation of FG-domain layer occupancy for $n = 1, 2, 3$. The

cNup214, cNup62, cNup98, and cNup153, respectively. This imposes a sequential binding constraint that mimics the multilayered Kap β 1-FG-domain binding characteristics observed experimentally (Fig. 4).

Fitting the SPR sensograms by this method (Fig. 6 B) provides a constellation of k_{on} and k_{off} values in the interaction maps shown for each FG domain (Fig. 7, A–D). Distribution analysis reveals distinct populations of k_{on} and k_{off} that can be grouped into slow (low k_{off}) and fast (high k_{off}) kinetic phases irrespective of the FG domain (Fig. 7 E). The slow phase is manifested as a band with intense peaks for $k_{off} < 10^{-3} \text{ s}^{-1}$ and $k_{on} < 10^6 \text{ M}^{-1} \text{ s}^{-1}$, whereas the fast phase has a speckled distribution around $k_{off} \sim 10^0 \text{ s}^{-1}$ and $k_{on} \sim 10^6 \text{ M}^{-1} \text{ s}^{-1}$ and a lower overall population, as seen from the accompanying histograms. This fast phase accounts for $\sim 10\%$ of bound Kap β 1 molecules that exhibit a quick dissociation, as is found from the SPR sensograms (Supporting Material). An interesting feature of the slow phase for all FG domains is the occurrence of two to three peaks in k_{on} that decrease from $\sim 10^5 \text{ M}^{-1} \text{ s}^{-1}$ at low K_D (i.e., high affinity) to $\sim 10^1 \text{ M}^{-1} \text{ s}^{-1}$ at high K_D (i.e., low affinity). This likely denotes the transition from a moderate rate of penetration to a largely vacant FG-domain layer at low Kap β 1 concentrations, to a slower rate of entry at higher concentrations due to a reduction of accessible binding sites within the layer and to other effects, such as layer extension and steric hindrance caused by increasing Kap β 1 occupancy. A slow release then follows in both cases once a stable complex is formed due to binding avidity. It is important to point out that this is accompanied by the emergence of the fast phase, which becomes more prominent at higher Kap β 1 concentrations (i.e., high K_D) and can be correlated to limited binding at the periphery of the FG-domain layer due to the onset of saturation inside it.

Interestingly, we find that k_{on} and k_{off} provide a broad range of K_D values. The maxima in the distributions as seen from the histograms on the top of each map correlate well with $K_{D1} \sim 100 \text{ nM}$ from equilibrium analysis (Fig. 5 B). In addition, we find high K_D distributions with peaks at $\sim 1 \mu\text{M}$, $3.4 \mu\text{M}$, $8.5 \mu\text{M}$, and $2.3 \mu\text{M}$ for cNup214, cNup62, cNup98, and Nup153, respectively. Nevertheless, we note that the observed high K_D values have contributions from both slow and fast binding species

model considers a two-dimensional lattice of $10 \times 10 \text{ nm}^2$ binding sites (based on the size of Kap β 1). L denotes the empty surface sites (i.e., containing FG domains) and LA_n corresponds to the number (n) of analyte molecules bound per surface site. (B) Representative fits (gray) to SPR sensograms (black) for Kap β 1 binding to close-packed cNup214, cNup62, cNup98, and cNup153. The residuals of the fits are included below the curves. Neglecting the SPR signals from BSA injections, the RMSD values (bold residuals) are 0.13 (cNup214) > 0.094 (cNup153) > 0.063 (cNup62) > 0.026 (cNup98) (in terms of bound Kap β 1 layers). Note that fitting errors may arise from structural changes that occur in the FG-domain layer as Kap β 1 binding progresses (e.g., layer extension). To see this figure in color, go online.

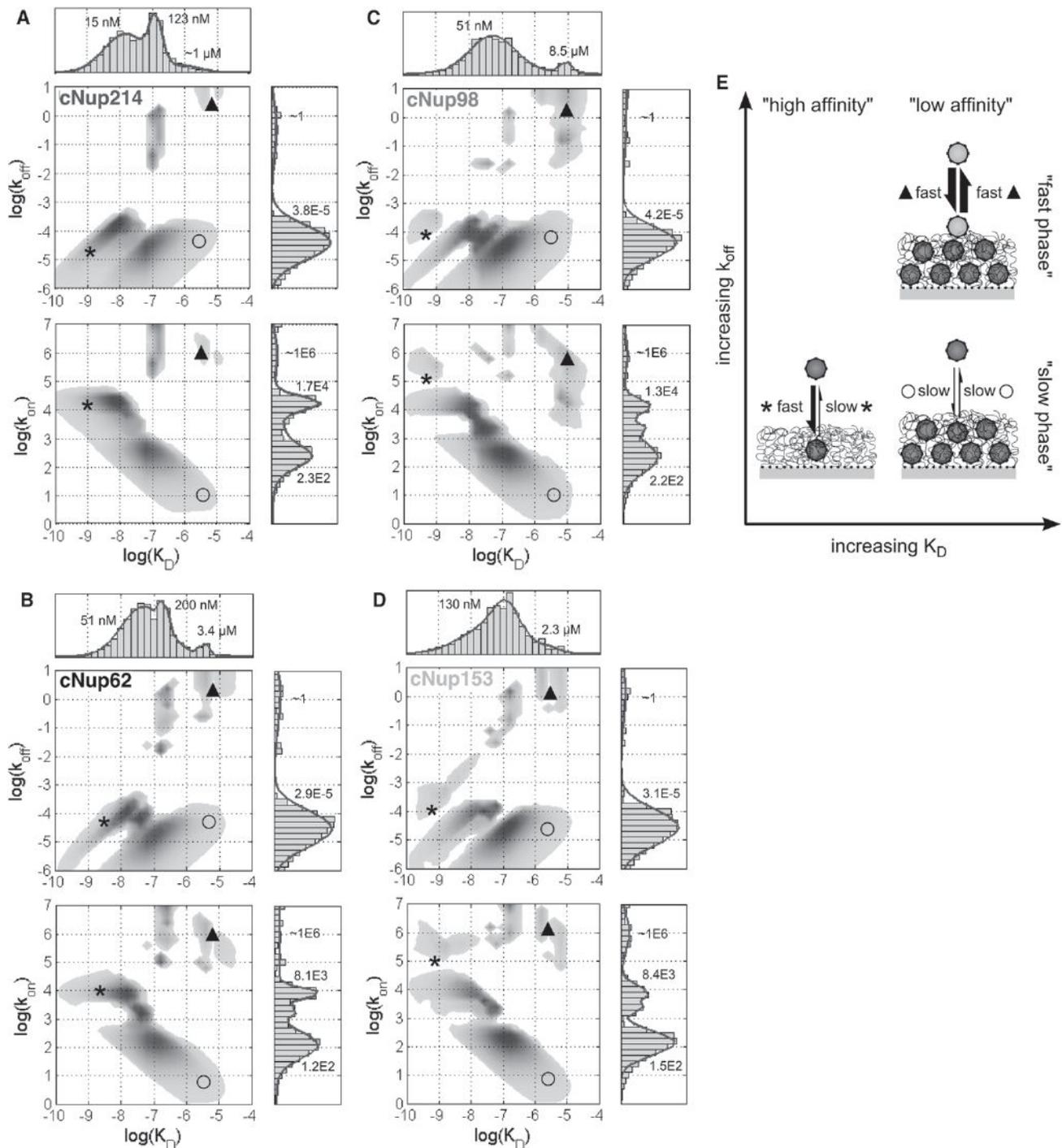


FIGURE 7 (A–D) Kinetic maps for Kap β 1 binding to close-packed cNup214 (A), cNup62 (B), cNup98 (C), and cNup153 (D). Each interaction map is averaged over four sensograms for each FG domain. The color intensity indicates the fractional abundance of a kinetic state. The histograms above and to the right of each map sum over all values in a given axis. (E) State diagrams separating slow (dark; low k_{off}) and fast (light; high k_{off}) kinetic phases of Kap β 1 binding. Moderate to fast k_{on} into a largely vacant FG-domain layer and slow k_{off} due to stable multivalent interactions results in high-affinity binding (*). Two low-affinity phases emerge due to FG-domain layer saturation, a fast phase, characterized by high (k_{off} , k_{on}) pairs due to limited binding at the layer periphery (\blacktriangle), and a slow phase characterized by low (k_{off} , k_{on}) pairs due to slow penetration into a preoccupied layer (\circ). To see this figure in color, go online.

with short or long half-lives, respectively (defined from k_{off}). Yet, in kinetic terms, the vast Kap β 1 majority that interacts with an already saturated FG-domain layer is in the fast

phase. This is because they bind and unbind in a dynamic fashion with much higher attempt frequencies. Here, the fraction of interactions that lead into the slow phase can

be estimated from $k_{on}(\text{slow})/k_{on}(\text{fast})$ (i.e., $10^1/10^6$), which corresponds to 0.001% at high Kap β 1 concentrations. To state this more clearly, the existence of the slow phase is not diminished at the highest Kap β 1 concentrations, nor is the existence of the fast phase diminished at low Kap β 1 concentrations. It is just that at high concentrations, the slow phase becomes less accessible due to hindered penetration into a saturated layer, which on the other hand enhances the fast phase.

Kinetic analysis of multivalent Kap β 1 binding to sparse FG domains

The K_D distribution in the sparse FG-domain regime is narrow, in agreement with equilibrium Langmuir isotherm analysis (Fig. 5 B). Here, the fast phase is still present but is reduced in population, i.e., a lower fraction of complexes with high $k_{off} \geq 0.001$ is observed. It is also striking that the slow k_{on} phase is well defined, with a single dominant peak of similar magnitude to that of saturated close-packed FG-domain layers, i.e., $10^1\text{--}10^3 \text{ M}^{-1} \text{ s}^{-1}$. Nonetheless, the reduction in k_{on} that is correlated to the level of Kap β 1 occupancy in the close-packed regime does not feature in the sparse regime (Supporting Material).

DISCUSSION

Relevance of FG-domain Kap β 1 occupancy for the NPC

The following principles generally apply to FG-centric barrier models. First, FG-domain behavior must exhibit barrier-like functionality. Second, sufficiently strong Kap-FG interactions are required to ensure NPC transport selectivity, because insufficient binding implies barrier rejection. Third, Kap-FG interactions cause conformational FG-domain changes (e.g., by ‘reversibly collapsing’ or ‘rapidly dissolving’) to alleviate spatial constraints imposed by the barrier. Fourth, and paradoxically, high Kap mobility follows from strong binding.

With the exception of cNup98, a key finding here is the high molecular occupancy of Kap β 1 in the FxFG domains of cNup214, cNup62, and cNup153 at physiological concentrations of Kap β 1. Indeed, this has been observed in cells where ~ 100 Kap β 1 molecules populate the NPC at steady state (44,59). By accounting for Kap β 1 at physiological concentrations, we find that most of the bound Kap β 1 molecules penetrate and occupy the FG-domain layers due to Kap β 1-FG binding avidity. The low K_D values (i.e., ~ 100 nM) in combination with slow unbinding kinetics ($k_{off} < 10^{-3} \text{ s}^{-1}$ or, in terms of half-lives, > 693 s) reflect that Kap β 1 forms stable multivalent complexes inside the FG-domain layer at steady state. Although such strong binding might ensure biochemical selectivity, it cannot account for the rapid ~ 5 ms NPC translocation times (39). Hence, this

predicts that only a slow transport phase would proceed through a close-packed FG-domain barrier if it were to span the entire NPC (e.g., meshwork). This is because equal numbers of FG repeats would be accessible to bind individual Kap β 1 molecules given their homogenous distribution in the NPC channel.

How then might Kaps proceed? Our results reveal that a low-affinity fraction of Kap β 1 ($K_D \geq 1 \mu\text{M}$) dominates at physiological concentrations once an FG-domain layer is saturated and pre-occupied with Kap β 1. This low-affinity species experiences hindered penetration due to increased steric effects arising from FG-domain layer extension and saturation to promote a fast phase with limited access to FG repeats at the layer periphery. Our kinetic analysis shows that over time only an estimated 0.001% of the total interactions would lead to the slow phase at physiological concentrations. Thus, in the scenario of a pre-occupied NPC, most Kap β 1 molecules entering the central channel would remain in the fast phase. Meanwhile, cNup98 is able to promote fast transport because its intralayer cohesion makes it the least penetrable to Kap β 1. Indeed, a similar situation can be expected if nonspecific proteins from cell lysate are able to bind and occupy the FG domains (36). In accordance with theory (60), our findings show that optimal Kap-FG-domain interaction strength can be tuned by saturating the FG domains with Kap β 1. With their fast off rates ($k_{off} \sim 1 \text{ s}^{-1}$) and short half-lives (< 100 ms), these species would dominate fast transport through the NPC at physiological Kap concentrations. Nevertheless, this would require an unobstructed path to support transport by the fast phase, such as a single central channel that would be surrounded by the peripheries of Kap-occupied FG-domain layers, as recently shown by single-molecule fluorescence detection (61,62).

Model of Kap-centric NPC control

A major consequence of our findings is that at physiological concentrations, the tightly bound slow-phase Kap β 1 molecules likely form an essential barrier component of the NPC that acts against nonspecific cargoes. This represents a shift in paradigm with respect to FG-centric barrier models (e.g., brush, meshwork) because it highlights the role of interacting FG domains and Kaps rather than FG domains alone. We find that the FG domains appear to act as a flexible velcro-like scaffold that can extend and contract with increasing or decreasing Kap occupancy, as illustrated by the ability of infiltrating nanoparticles to control molecular brush morphology (63). In addition, we predict that Kap occupancy dictates NPC barrier conformation, transport selectivity, and speed in the NPC.

Such a Kap-centric barrier model is shown in Fig. 8. At physiological concentration, the NPC mechanism consists of a majority of slow-phase Kap β 1 molecules that are incorporated within extended FxFG domains that line the central channel toward the NPC periphery, with the exception

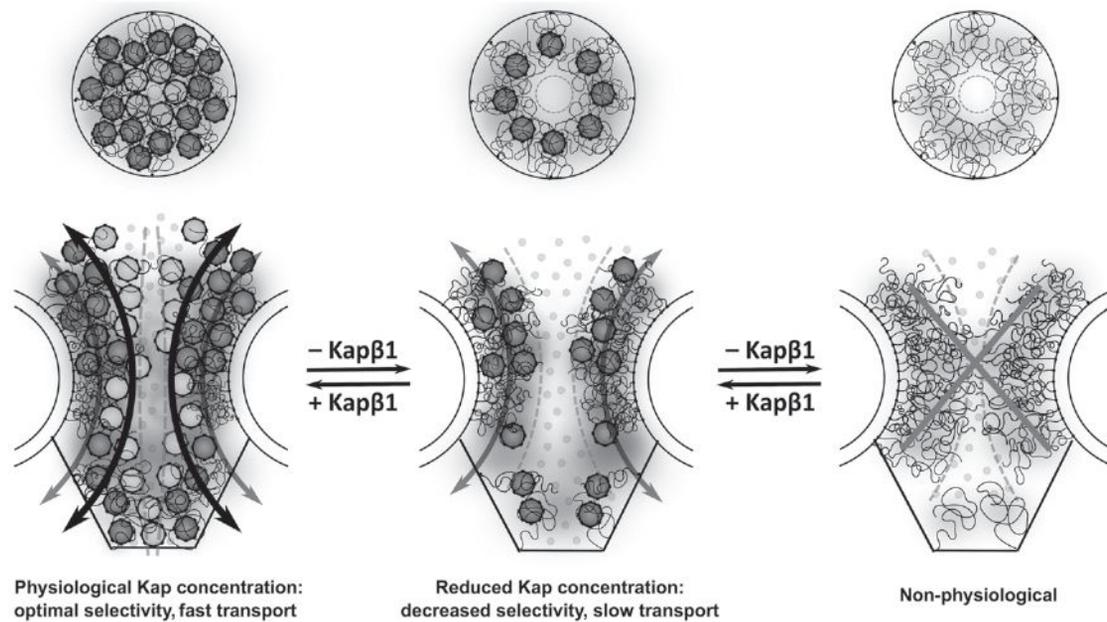


FIGURE 8 Kap-centric control in NPCs. Kap β 1 is an integral constituent of the NPC at physiological steady-state conditions. Selective barrier functionality against nonspecific cargoes is provided by slow-phase Kap β 1 molecules (*dark green*) that saturate and engorge peripheral FG domains (e.g., Nup214 and Nup153). Fast-phase Kap β 1 molecules (*light green*) promote fast transport through a narrow central channel (e.g., Nup62) due to the limited availability of FG repeats. Nup98 coheres into a narrow annular ring or bottleneck at the central plane of the NPC. Kap β 1 reduction leads to barrier contraction and a decrease in selectivity because of a widening of the aqueous channel. The availability of free FG repeats slows down the transport of Kap β 1 molecules, which eventually repopulate the FG domains to reinstate normal Kap occupancy and, thereafter, selectivity and speed control. NPCs devoid of Kap β 1 are likely unphysiological. To see this figure in color, go online.

of Nup98 (i.e., GLFG), which forms a cohesive annular ring around the central plane. Together, the Kaps and the engorged FG domains surround a narrow aqueous channel along the NPC axis that enforces the passive size limit while remaining selective to fast-phase Kap β 1 molecules that bind weakly enough to diffuse along peripheral regions of the FG domains, e.g., by a reduction of dimensionality (31,32). Any perturbation that reduces Kap β 1 occupancy would lead to a retraction of the barrier and a decrease in selectivity (i.e., the pore would become more leaky) because of a widening of the aqueous channel. This leads to a slowdown of subsequent Kap β 1 transport given the increase of free FG repeats. We remark, however, that this repopulation of the NPC by slow Kap β 1 species provides a feedback mechanism that reinstates Kap occupancy, self-heals FG-domain conformation (24), and thereafter normalizes transport selectivity and speed control. This may explain how the mechanistic (occupancy) and kinetic (FG-repeat availability) characteristics of the barrier are balanced to accommodate local perturbations in the NPC (i.e., higher Kap β 1 occupancy \Rightarrow higher selectivity/less leaky, fast Kap β 1 transport, and lower Kap β 1 occupancy \Rightarrow lower selectivity/more leaky, slow Kap β 1 transport).

It is important to note that the molecular view we propose directly agrees with the preferential binding of Kap β 1 along the NPC walls due to FG-domain binding and may further embody the interactions that underlie the NPC transport pathway as a ‘self-regulated viscous channel’ (61,62).

Further validations can be found from the inverse correlation between (decreasing) Kap β 1 interaction time and (increasing) import efficiency with increasing Kap β 1 concentrations where efficient nuclear transport commences only at sufficiently high Kap β 1 concentrations ($>1.5 \mu\text{M}$) (64). Our work also predicts that a continuum of different transport rates can exist depending on local NPC Kap concentrations. This scenario explains why the NPC cannot be devoid of Kaps, because any Kap that encounters an FG-domain-only barrier would suffer from high FG-binding avidity and slow down. A final provocative consequence of a Kap-centric barrier mechanism is that a reduction of Kaps rather than FG domains (65) would result in NPC leakiness. As a case in point, an increased presence of Kaps seems to tighten barrier functionality in both FG-domain gels (26) and artificial NPCs (66).

Relevance of surface-tethered FG domains

In contrast to the in-solution behavior of nontethered FG domains, the NPC interior presents many closely tethered FG domains that display collective functional characteristics *in vivo* (67). Our work indicates that FG-domain surface tethering is an essential contextual consideration for the NPC, because it defines the pore boundary, establishes FG-domain orientation with respect to an interface, and enforces a limit on Kap occupancy (and how far FG domains can extend). As shown, Kap β 1 binding avidity depends

on FG-domain surface density as would be defined by 1), FG-domain copy numbers (51, 2), FG-domain tethering sites within the central NPC channel (1), and 3), the corresponding distances between neighboring tethering sites (i.e., FG-domain surface density). In the close-packed regime, all the FxFG domains studied here (cNup214, cNup62, and cNup153) exhibit molecular brush behavior and have a large capacity to incorporate up to two layers or more of Kap β 1 molecules at physiological Kap β 1 concentrations, albeit with varying degrees of extensibility.

On the other hand, cNup98 forms a short compact GLFG domain layer that is only partially penetrable to Kap β 1. This suggests that the close-packed, surface-tethered form of cNup98 may cohere more strongly than pure Nup98 hydrogels, where Kap β 1 penetrated a depth of a few micrometers (28). Further functional correlations are difficult to establish, because a hydrogel can be comprised of fibrous meshworks and sub-micrometer-sized porous channels with unique morphological and sieving properties (68).

CONCLUSIONS

To our knowledge, these findings provide novel evidence for a paradigm where Kaps serve as integral, possibly regulatory constituents of the NPC by balancing mechanistic with kinetic control over the NPC barrier mechanism. This is consistent with the observation that NPCs are not devoid of Kaps at physiological concentrations. Here, the role of multivalency is twofold: it allows for a slow phase of Kap β 1 to penetrate and drive the extension of FG domains (i.e., the barrier) due to high binding avidity followed by a fast transient phase of Kap β 1 that proceeds with reduced avidity along the peripheral regions of the FG domains. As such, a provocative ramification is that the effectiveness of the NPC barrier changes with local Kap concentration. Nevertheless, it is unclear how the effect of confinement within a cylindrical geometry will influence the observed effects. Therefore, it will be important to test for Kap-centric control in NPCs and biomimetic nanopores (47,66) in the presence of different Kaps, cargoes, and nuclear transport factors (e.g., RanGTP).

SUPPORTING MATERIAL

One table and seven figures are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)00227-6](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00227-6).

This work is supported by the Swiss National Science Foundation, the Biozentrum, and the Swiss Nanoscience Institute.

REFERENCES

- Alber, F., S. Dokudovskaya, ..., M. P. Rout. 2007. The molecular architecture of the nuclear pore complex. *Nature*. 450:695–701.
- Beck, M., F. Förster, ..., O. Medalia. 2004. Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science*. 306:1387–1390.
- Rout, M. P., J. D. Aitchison, ..., B. T. Chait. 2000. The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* 148:635–651.
- Keminer, O., and R. Peters. 1999. Permeability of single nuclear pores. *Biophys. J.* 77:217–228.
- Paine, P. L., L. C. Moore, and S. B. Horowitz. 1975. Nuclear envelope permeability. *Nature*. 254:109–114.
- Chook, Y. M., and K. E. Süel. 2011. Nuclear import by karyopherin- β s: recognition and inhibition. *Biochim. Biophys. Acta*. 1813:1593–1606.
- Bayliss, R., T. Littlewood, and M. Stewart. 2000. Structural basis for the interaction between FxFG nucleoporin repeats and importin- β in nuclear trafficking. *Cell*. 102:99–108.
- Stewart, M. 2007. Molecular mechanism of the nuclear protein import cycle. *Nat. Rev. Mol. Cell Biol.* 8:195–208.
- Görlisch, D., F. Vogel, ..., R. A. Laskey. 1995. Distinct functions for the two importin subunits in nuclear protein import. *Nature*. 377:246–248.
- Terry, L. J., and S. R. Wente. 2009. Flexible gates: dynamic topologies and functions for FG nucleoporins in nucleocytoplasmic transport. *Eukaryot. Cell*. 8:1814–1827.
- Grossman, E., O. Medalia, and M. Zwirger. 2012. Functional architecture of the nuclear pore complex. *Annu. Rev. Biophys.* 41:557–584.
- Devos, D., S. Dokudovskaya, ..., A. Sali. 2006. Simple fold composition and modular architecture of the nuclear pore complex. *Proc. Natl. Acad. Sci. USA*. 103:2172–2177.
- Denning, D. P., S. S. Patel, ..., M. Rexach. 2003. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proc. Natl. Acad. Sci. USA*. 100:2450–2455.
- Bayliss, R., S. W. Leung, ..., M. Stewart. 2002. Structural basis for the interaction between NTF2 and nucleoporin FxFG repeats. *EMBO J.* 21:2843–2853.
- Bayliss, R., T. Littlewood, ..., M. Stewart. 2002. GLFG and FxFG nucleoporins bind to overlapping sites on importin- β . *J. Biol. Chem.* 277:50597–50606.
- Bayliss, R., K. Ribbeck, ..., M. Stewart. 1999. Interaction between NTF2 and xFxFG-containing nucleoporins is required to mediate nuclear import of RanGDP. *J. Mol. Biol.* 293:579–593.
- Patel, S. S., B. J. Belmont, ..., M. F. Rexach. 2007. Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex. *Cell*. 129:83–96.
- Ando, D., M. Colvin, ..., A. Gopinathan. 2013. Physical motif clustering within intrinsically disordered nucleoporin sequences reveals universal functional features. *PLoS ONE*. 8:e73831–e73842.
- Colwell, L. J., M. P. Brenner, and K. Ribbeck. 2010. Charge as a selection criterion for translocation through the nuclear pore complex. *PLoS Comput. Biol.* 6:e1000747.
- Tagliazucchi, M., O. Peleg, ..., I. Szleifer. 2013. Effect of charge, hydrophobicity, and sequence of nucleoporins on the translocation of model particles through the nuclear pore complex. *Proc. Natl. Acad. Sci. USA*. 110:3363–3368.
- Yamada, J., J. L. Phillips, ..., M. F. Rexach. 2010. A bimodal distribution of two distinct categories of intrinsically disordered structures with separate functions in FG nucleoporins. *Mol. Cell. Proteomics*. 9:2205–2224.
- Xu, S., and M. A. Powers. 2013. In vivo analysis of human nucleoporin repeat domain interactions. *Mol. Biol. Cell*. 24:1222–1231.
- Lim, R. Y. H., N. P. Huang, ..., U. Aebi. 2006. Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proc. Natl. Acad. Sci. USA*. 103:9512–9517.
- Schoch, R. L., L. E. Kapinos, and R. Y. H. Lim. 2012. Nuclear transport receptor binding avidity triggers a self-healing collapse transition in FG-nucleoporin molecular brushes. *Proc. Natl. Acad. Sci. USA*. 109:16911–16916.

25. Frey, S., and D. Görlich. 2007. A saturated FG-repeat hydrogel can reproduce the permeability properties of nuclear pore complexes. *Cell*. 130:512–523.
26. Frey, S., and D. Görlich. 2009. FG/FxFG as well as GLFG repeats form a selective permeability barrier with self-healing properties. *EMBO J*. 28:2554–2567.
27. Hülsmann, B. B., A. A. Labokha, and D. Görlich. 2012. The permeability of reconstituted nuclear pores provides direct evidence for the selective phase model. *Cell*. 150:738–751.
28. Labokha, A. A., S. Gradmann, ..., D. Görlich. 2013. Systematic analysis of barrier-forming FG hydrogels from *Xenopus* nuclear pore complexes. *EMBO J*. 32:204–218.
29. Rout, M. P., J. D. Aitchison, ..., B. T. Chait. 2003. Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol*. 13:622–628.
30. Lim, R. Y. H., B. Fahrenkrog, ..., U. Aebi. 2007. Nanomechanical basis of selective gating by the nuclear pore complex. *Science*. 318:640–643.
31. Peters, R. 2005. Translocation through the nuclear pore complex: selectivity and speed by reduction-of-dimensionality. *Traffic*. 6:421–427.
32. Peters, R. 2009. Translocation through the nuclear pore: Kaps pave the way. *Bioessays*. 31:466–477.
33. Bednenko, J., G. Cingolani, and L. Gerace. 2003. Importin β contains a COOH-terminal nucleoporin binding region important for nuclear transport. *J. Cell Biol*. 162:391–401.
34. Isgro, T. A., and K. Schulten. 2005. Binding dynamics of isolated nucleoporin repeat regions to importin- β . *Structure*. 13:1869–1879.
35. Mammen, M., S. K. Choi, and G. M. Whitesides. 1998. Polyvalent interactions in biological systems: Implications for design and use of multivalent ligands and inhibitors. *Angew. Chem. Int. Ed.* 37:2754–2794.
36. Tetenbaum-Novatt, J., L. E. Hough, ..., M. P. Rout. 2012. Nucleocytoplasmic transport: a role for nonspecific competition in karyopherin-nucleoporin interactions. *Mol. Cell. Proteomics*. 11:31–46.
37. Ben-Efraim, I., and L. Gerace. 2001. Gradient of increasing affinity of importin β for nucleoporins along the pathway of nuclear import. *J. Cell Biol*. 152:411–417.
38. Lott, K., A. Bhardwaj, ..., G. Cingolani. 2010. The importin β binding domain modulates the avidity of importin β for the nuclear pore complex. *J. Biol. Chem*. 285:13769–13780.
39. Dange, T., D. Grünwald, ..., U. Kubitschek. 2008. Autonomy and robustness of translocation through the nuclear pore complex: a single-molecule study. *J. Cell Biol*. 183:77–86.
40. Paulillo, S. M., E. M. Phillips, ..., B. Fahrenkrog. 2005. Nucleoporin domain topology is linked to the transport status of the nuclear pore complex. *J. Mol. Biol*. 351:784–798.
41. Chatel, G., S. H. Desai, ..., B. Fahrenkrog. 2012. Domain topology of nucleoporin Nup98 within the nuclear pore complex. *J. Struct. Biol*. 177:81–89.
42. Schwarz-Herion, K., B. Maco, ..., B. Fahrenkrog. 2007. Domain topology of the p62 complex within the 3-D architecture of the nuclear pore complex. *J. Mol. Biol*. 370:796–806.
43. Fahrenkrog, B., B. Maco, ..., U. Aebi. 2002. Domain-specific antibodies reveal multiple-site topology of Nup153 within the nuclear pore complex. *J. Struct. Biol*. 140:254–267.
44. Paradise, A., M. K. Levin, ..., J. H. Carson. 2007. Significant proportions of nuclear transport proteins with reduced intracellular mobilities resolved by fluorescence correlation spectroscopy. *J. Mol. Biol*. 365:50–65.
45. Svitel, J., A. Balbo, ..., P. Schuck. 2003. Combined affinity and rate constant distributions of ligand populations from experimental surface binding kinetics and equilibria. *Biophys. J*. 84:4062–4077.
46. Svitel, J., H. Boukari, ..., P. Schuck. 2007. Probing the functional heterogeneity of surface binding sites by analysis of experimental binding traces and the effect of mass transport limitation. *Biophys. J*. 92:1742–1758.
47. Kowalczyk, S. W., L. Kapinos, ..., C. Dekker. 2011. Single-molecule transport across an individual biomimetic nuclear pore complex. *Nat. Nanotechnol*. 6:433–438.
48. Hansen, P. C. 2007. Regularization tools version 4.0 for Matlab 7.3. *Numer. Algorithms*. 46:189–194.
49. Landi, G., and F. Zama. 2006. The active-set method for nonnegative regularization of linear ill-posed problems. *Appl. Math. Comput*. 175:715–729.
50. Frenkiel-Krispin, D., B. Maco, ..., O. Medalia. 2010. Structural analysis of a metazoan nuclear pore complex reveals a fused concentric ring architecture. *J. Mol. Biol*. 395:578–586.
51. Cronshaw, J. M., A. N. Krutchinsky, ..., M. J. Matunis. 2002. Proteomic analysis of the mammalian nuclear pore complex. *J. Cell Biol*. 158:915–927.
52. Solmaz, S. R., R. Chauhan, ..., I. Melčák. 2011. Molecular architecture of the transport channel of the nuclear pore complex. *Cell*. 147:590–602.
53. Peleg, O., and R. Y. H. Lim. 2010. Converging on the function of intrinsically disordered nucleoporins in the nuclear pore complex. *Biol. Chem*. 391:719–730.
54. Schoch, R. L., and R. Y. H. Lim. 2013. Non-interacting molecules as innate structural probes in surface plasmon resonance. *Langmuir*. 29:4068–4076.
55. Bright, J. N., T. B. Woolf, and J. H. Hoh. 2001. Predicting properties of intrinsically unstructured proteins. *Prog. Biophys. Mol. Biol*. 76:131–173.
56. Fasting, C., C. A. Schalley, ..., R. Haag. 2012. Multivalency as a chemical organization and action principle. *Angew. Chem. Int. Ed. Engl*. 51:10472–10498.
57. Gjelstrup, L. C., J. D. Kaspersen, ..., T. Vorup-Jensen. 2012. The role of nanometer-scaled ligand patterns in polyvalent binding by large mannan-binding lectin oligomers. *J. Immunol*. 188:1292–1306.
58. Munoz, E. M., J. Correa, ..., E. Fernandez-Megia. 2013. Real-time evaluation of binding mechanisms in multivalent interactions: a surface plasmon resonance kinetic approach. *J. Am. Chem. Soc*. 135:5966–5969.
59. Tokunaga, M., N. Imamoto, and K. Sakata-Sogawa. 2008. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods*. 5:159–161.
60. Zilman, A., S. Di Talia, ..., M. O. Magnasco. 2007. Efficiency, selectivity, and robustness of nucleocytoplasmic transport. *PLOS Comput. Biol*. 3:e125.
61. Ma, J., A. Goryaynov, ..., W. Yang. 2012. Self-regulated viscous channel in the nuclear pore complex. *Proc. Natl. Acad. Sci. USA*. 109:7326–7331.
62. Ma, J., and W. Yang. 2010. Three-dimensional distribution of transient interactions in the nuclear pore complex obtained from single-molecule snapshots. *Proc. Natl. Acad. Sci. USA*. 107:7305–7310.
63. Opferman, M. G., R. D. Coalson, ..., A. Zilman. 2013. Morphology of polymer brushes infiltrated by attractive nanoinclusions of various sizes. *Langmuir*. 29:8584–8591.
64. Yang, W. D., and S. M. Musser. 2006. Nuclear import time and transport efficiency depend on importin β concentration. *J. Cell Biol*. 174:951–961.
65. Strawn, L. A., T. X. Shen, ..., S. R. Wentz. 2004. Minimal nuclear pore complexes define FG repeat domains essential for transport. *Nat. Cell Biol*. 6:197–206.
66. Jovanovic-Talisman, T., J. Tetenbaum-Novatt, ..., B. T. Chait. 2009. Artificial nanopores that mimic the transport selectivity of the nuclear pore complex. *Nature*. 457:1023–1027.
67. Atkinson, C. E., A. L. Mattheyses, ..., S. M. Simon. 2013. Conserved spatial organization of FG domains in the nuclear pore complex. *Biophys. J*. 104:37–50.
68. Milles, S., K. Huy Bui, ..., E. A. Lemke. 2013. Facilitated aggregation of FG nucleoporins under molecular crowding conditions. *EMBO Rep*. 14:178–183.

Conformational Recognition of an Intrinsically Disordered Protein

James M. Krieger,[†] Giuliana Fusco,[‡] Marc Lewitzky,^{§¶} Philip C. Simister,[§] Jan Marchant,[†] Carlo Camilloni,[‡] Stephan M. Feller,^{§¶} and Alfonso De Simone^{†*}

[†]Department of Life Sciences, Imperial College London, London, UK; [‡]Department of Chemistry, University of Cambridge, Cambridge, UK;

[§]Department of Oncology, University of Oxford, Oxford, UK; and [¶]Institute of Molecular Medicine, Martin Luther University Halle-Wittenberg, Halle, Germany

ABSTRACT There is a growing interest in understanding the properties of intrinsically disordered proteins (IDPs); however, the characterization of these states remains an open challenge. IDPs appear to have functional roles that diverge from those of folded proteins and revolve around their ability to act as hubs for protein-protein interactions. To gain a better understanding of the modes of binding of IDPs, we combined statistical mechanics, calorimetry, and NMR spectroscopy to investigate the recognition and binding of a fragment from the disordered protein Gab2 by the growth factor receptor-bound protein 2 (Grb2), a key interaction for normal cell signaling and cancer development. Structural ensemble refinement by NMR chemical shifts, thermodynamics measurements, and analysis of point mutations indicated that the population of preexisting bound conformations in the free-state ensemble of Gab2 is an essential determinant for recognition and binding by Grb2. A key role was found for transient polyproline II (PPII) structures and extended conformations. Our findings are likely to have very general implications for the biological behavior of IDPs in light of the evidence that a large fraction of these proteins possess a specific propensity to form PPII and to adopt conformations that are more extended than the typical random-coil states.

INTRODUCTION

Protein-protein interactions (PPint) regulate a large number of biomolecular processes. It is becoming increasingly evident that to achieve a comprehensive understanding of partner recognition and binding affinity in PPint, in addition to the structures of the macromolecular complexes, it is necessary to characterize in detail the conformations that proteins and peptides adopt before binding (1). Dynamical behavior is indeed a key modulator of the energetics of macromolecular interactions. Biomolecular structural fluctuations of backbone and side chain atoms are finely tuned to allow processes such as partner selection (2), allosteric modulation (3), and cellular signaling (4–8). Structural dynamics are particularly relevant in weak bindings and in molecular interactions that involve intrinsically disordered proteins (IDPs) (9–11). It is now clear that disordered protein states are highly abundant in all living organisms and have biological activities that can be distinct from those of folded proteins (12–15). In this context, understanding the mechanisms by which IDPs recognize and interact with their partners remains a major challenge. This task requires a structural characterization based on probability distributions to account for the conformational heterogeneity of IDPs rather than single structures.

Efforts to define a relationship among the local dynamics, residual structure, and binding properties of IDPs are hampered by the intrinsic limits of standard experimental techniques. High-resolution structural techniques such as

x-ray crystallography are able to indirectly imply structural disorder, but they cannot quantitatively characterize transient structural states and populations adopted in IDP ensembles. Such information can be partially obtained from single-molecule experiments (16), but at a significantly lower resolution than atomic details. Recent advances in biomolecular NMR have enabled new approaches to probe the conformational preferences of disordered protein states (17–21). These methods provide new opportunities to characterize transient structures and their populations in the conformational ensembles of IDPs, with significant accuracy (22–24). This study illustrates the latter point very clearly by addressing the solution properties of a disordered protein fragment and by defining their relationship with the thermodynamic affinity for a binding partner. We obtained this result by measuring backbone chemical shifts (CSs) and employing the resulting data to study transient secondary-structure populations. To that end, we used the δ 2D method (22,25) and backbone dynamics (using the random coil index (RCI)) (26), as well as ensemble-averaged restrained molecular-dynamics (MD) simulations with full representation of all protein and solvent atoms. Using this approach, we characterized the conformational determinants of complex formation between the growth factor receptor-bound protein 2 (Grb2) and the disordered protein Grb2-associated binder 2 (Gab2) (27–31).

Grb2 is a 25 kDa adaptor protein that is involved in signal transduction and cell communication, and is composed of an SH2 domain flanked by N- and C-terminal SH3 domains. Gab proteins belong to a family of large multisite docking (LMD) proteins that play scaffolding roles in the assembly

Submitted January 16, 2014, and accepted for publication March 6, 2014.

*Correspondence: adesimon@imperial.ac.uk

Editor: David Eliezer.

© 2014 by the Biophysical Society
0006-3495/14/04/1771/9 \$2.00



of large multimolecular complexes (32–34). These proteins are often implicated in signal integration and serve as processing platforms for several signaling pathways. We focused on Gab2, a largely intrinsically disordered protein of ~74 kDa with a folded N-terminal PH domain whose interaction with Grb2 has been implicated in normal cell signaling (35) and cancer development (27). Peptide arrays showed two binding sites in Gab2 (Gab2a and Gab2b) interacting with the C-terminal domain of Grb2 (Grb2 SH3C). Both epitopes host an RxxK motif, which is essential for Grb2 SH3C binding, with the highest affinity found in the Gab2b motif (27).

We employed the fragment 503–524 of Gab2 (Gab₂₅₀₃₋₅₂₄), i.e., spanning the Gab2b motif, to characterize the binding affinity for the SH3C domain of Grb2, which occurs with a low micromolar affinity (Table 1). Our NMR CS analyses showed that Gab₂₅₀₃₋₅₂₄ is mainly disordered in solution, which is in agreement with the conformational properties adopted in the full-length Gab2 sequence (36), but adopts some residual structure of extended β and polyproline type II (PPII) character. Structural ensemble refinement, isothermal titration calorimetry (ITC), and analysis of the effects of point mutations revealed that the population of preexisting bound conformations in the free-state ensemble is an essential factor in the recognition and binding of Gab2b by Grb2 SH3C. Transient PPII segments were shown to play a key role in the affinity of this PPint. This structural motif is a well-known recognition element for SH3 domains and has been implicated in the interaction properties of IDPs. The data presented here therefore address the molecular determinants of binding modes in IDPs and show how a description of their structural propensities as probability distributions derived from the information contained in NMR data can lead to a thorough understanding of their biological properties.

MATERIALS AND METHODS

Peptide synthesis and purification

Peptides spanning the wild-type (WT) sequence (SRGSEIQPPVNRNLKPDRKAK) of Gab₂₅₀₃₋₅₂₄ or point mutations (see Table 1) were produced by 9-fluorenylmethoxycarbonyl (Fmoc) solid-phase peptide synthesis (37) using a Respep SL tabletop peptide synthesizer (Intavis Bio-

analytical Instruments, Koeln, Germany). Briefly, amino acids with amines protected by Fmoc groups were coupled via their carboxyl groups to the amino groups of a growing chain starting with an Fmoc-lysine Wang resin (two wells, each containing 20 μ mol). Before each round of coupling, the unreacted amino acids were washed out and the N-terminus was deprotected with piperidine/dimethylformamide. After the final deprotection step, the resins were removed from the synthesizer and the peptides were cleaved off with trifluoroacetic acid (TFA) in a sealed syringe.

TFA cleavage products were precipitated with ice-cold tert-butyl methyl ether (TBME) and centrifuged three times, with replacement of the supernatant with fresh TBME at 6000 $\times g$ for 10 min to wash away the TFA. These pellets were then desiccated overnight and redissolved in 4 ml of a 12.5% methanol/water solution. Then 20 μ l of these solutions was analyzed by electrospray ionization (ESI) coupled with liquid chromatography/mass spectrometry (LC-MS) on a system comprised of a Waters (Waters, Milford, MA) 515 HPLC pump, a Waters Micromass ZQ MS detector, a Waters 600 controller, a Waters 2998 photodiode array (PDA) detector, and a Waters sample manager. Then 4 ml was taken for preparative LC-MS and used with mass-triggered collection of LC fractions for peptide purification. Methanol was removed from the purified peptide solutions on an EZ-2 plus centrifugal evaporator (Genevac, Ipswich, UK) and water was removed by lyophilization in preweighed microcentrifuge tubes (Eppendorf, Hamburg, Germany).

Purification of Grb2 SH3C

The pGEX vector encoding Grb2 SH3C has been described elsewhere (38). For protein expression, bacteria were grown in terrific broth with 100 μ g/ml ampicillin. When OD₆₀₀ reached 0.8, expression was induced with 0.05 mM isopropyl- β -D-thiogalactopyranoside overnight at 18°C, and bacteria were then sedimented by centrifugation. To generate protein for ITC, chilled bacterial pellets were lysed in cold TPE (1% (v/v) Triton X-100, PBS (pH 7.4), 100 mM EDTA, 10 μ g/ml aprotinin, 0.7 μ g/ml pepstatin A, 0.5 μ g/ml leupeptin, and 5 μ g/ml antipain) and sonicated, and the lysate was clarified by centrifugation at 20,000 $\times g$ for 1 h at 4°C. GST-fusion protein was purified by incubation with glutathione (GSH)-sepharose beads overnight at 4°C on a nutator. Beads were then washed extensively with 50 mM TrisHCl (pH 7.5), 100 mM EDTA, and 0.1% (v/v) Tween 20. Bound GST-fusion protein was eluted with 100 mM GSH, pH adjusted to approximately ~7.5 with TrisHCl (pH 8.8), and the eluate was dialyzed against 5 mM TrisHCl (pH 7.5). The integrity of the dialyzed protein was analyzed by SDS-PAGE and Coomassie Blue staining, and the protein concentration was assayed by the Bradford method. Purified GST-Grb2 SH3C fusion protein was snap-frozen in aliquots and stored at –80°C until further use in ITC.

ITC

ITC was performed with a VP-ITC MicroCalorimeter (MicroCal, Northampton, MA). Peptides were dissolved at 250 μ M (750 μ M for the

TABLE 1 ITC measurements of Grb2 SH3C interactions with WT and mutant peptides derived from Gab2

	Sequence	n^a	K_d (μ M)	ΔH (kcal \cdot mol ⁻¹)	$-T\Delta S$ (kcal \cdot mol ⁻¹)
WT	SRGSEIQPPVNRNLKPDRKAK	0.91 \pm 0.06	3.95 \pm 0.51	-11.8 \pm 1.33	4.46 \pm 1.42
P510A	SRGSEIQAPPVNRNLKPDRKAK	1.11 \pm 0.07	5.09 \pm 0.34	-10.4 \pm 0.94	3.15 \pm 0.99
P511A	SRGSEIQPAPVNRNLKPDRKAK	–	TLQ ^b	–	–
P512A	SRGSEIQPPAVNRNLKPDRKAK	0.99 \pm 0.06	6.00 \pm 0.25	-11.3 \pm 0.34	4.22 \pm 0.31
R515A	SRGSEIQPPVNRNLKPDRKAK	–	NBD ^c	–	–
K518A	SRGSEIQPPVNRNLAPDRKAK	–	NBD ^c	–	–
P519A	SRGSEIQPPVNRNLKADRKAK	0.84 \pm 0.01	57.85 \pm 1.7	-17.2 \pm 0.36	11.4 \pm 0.39

^aStoichiometry.

^bAffinity was too low for exact quantification.

^cNo binding was detected.

P519A peptide) in ITC buffer (25 mM HEPES (pH 7.5, adjusted with potassium hydroxide), 100 mM potassium acetate, and 5 mM magnesium acetate), clarified by centrifugation for 10 min at $20,000 \times g$ before use, degassed (ThermoVac; MicroCal), and titrated from a syringe (300 μ l total volume) into a sample chamber holding 1.43 ml of clarified and degassed 25 μ M GST-Grb2 SH3C (75 μ M when analyzing the P519A peptide). Upon reaching the equilibrium temperature of 25°C, peptide solutions were titrated into the sample chamber by 18 injections of 15 μ l each. The resulting peaks of measured deviations from the equilibrium temperature were integrated to yield the quantity of heat generated. The best fit to the data to calculate the binding affinity K_d was obtained using χ^2 minimization on a model assuming a single set of binding sites. All steps of the data analysis were performed using ORIGIN (V5.0) software provided by the manufacturer.

NMR spectroscopy

Initial assignment of peptide resonances ($C\alpha$, $C\beta$, $H\alpha$, HN , and N) was performed under acidic conditions by dissolving the peptides in H_2O using 10% v/v D_2O for resonance locking. The acidic pH improved the assignment convergence owing to the limited broadening of the resonances of exchangeable protons. Subsequently, the assignments were transferred to spectra recorded under buffer conditions relevant to this investigation (25 mM HEPES pH 7.5 (adjusted with potassium hydroxide), 100 mM potassium acetate) at a temperature of 303 K. All NMR spectra were referenced using the resonance of the 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) molecule.

Experiments were carried out using 600 MHz and 800 MHz spectrometers (Bruker Avance III, Bruker Biospin, Billerica, MA). Proton assignment was carried out using a combination of total correlation spectroscopy (TOCSY) and rotating-frame nuclear Overhauser enhancement spectroscopy (ROESY) spectra. These were used for resonance assignment according to highly established protocols (39). Having completed the proton assignment, we were able to assign carbon and nitrogen resonances by using natural abundance ^{13}C and ^{15}N and recording heteronuclear single quantum coherence (HSQC) spectra. Random-coil CSs from the CamCoil method (25) were used to help assign the ^{13}C CSs.

Determination of the secondary-structure populations and dynamics of IDPs

To address the secondary-structure populations of IDPs, we employed the δ 2D method (22,25), which is based on the structural information provided by NMR CSs, by using the measured backbone CSs ($C\alpha$, $C\beta$, $H\alpha$, HN , and N). CSs were also employed to infer the local dynamics using the RCI (26).

Structural ensemble refinement by CS-restrained MD

To refine accurate structural ensembles of Gab2₅₀₃₋₅₂₄, we employed CSs in restrained MD simulations using the CamShift method (40). Briefly, CS restraints were imposed by adding a pseudo-energy term (E^{CS}) to a standard molecular-mechanics force field (E^{FF}):

$$E^{Tot} = E^{FF} + E^{CS} \quad (1)$$

The resulting force field (E^{Tot}) was employed in MD simulations, where the pseudo-energy term is given by

$$E^{CS} = \alpha \sum_i (\delta_i^{Exp} - \delta_i^{Calc})^2 \quad (2)$$

where the i sum runs over all the CSs employed in the refinement, α is the weight of the restraint term, and δ^{exp} and δ^{calc} are the experimental and calculated CSs, respectively. We employed the replica-averaged scheme, which implies that a given CS is calculated by

$$\delta^{Calc} = \frac{1}{m} \sum_m^4 \delta_m^{Calc} \quad (3)$$

where m runs over four replicas and δ_m^{Calc} is the CS of replica m .

We performed replica-averaged restrained MD simulations by using an implementation of GROMACS (41) as described previously (42) and employing a recent protocol based on four replicas (43). Each of the four replicas was equilibrated separately. Starting conformations were modeled as linear chains and accommodated in a dodecahedron box volume of 276 nm³ of volume. The box was filled with explicit waters and energy minimized. For each replica, the system was thermalized during an NVT simulation of 250 ps in which the temperature was increased up to 300 K. Subsequently, the pressure was equilibrated using a MD simulation of 200 ps under the Berendsen barostat. Finally, the individual replicas were equilibrated for 1-ns-long simulations.

The scheme of the replica-averaged restrained MD simulations is as follows: The four replicas evolve through a series of annealing cycles between 300 K and 400 K, with each cycle being composed of 100 ps at 300 K, 100 ps of linear increase in the temperature up to 400 K, 100 ps of constant-temperature MD simulations at 400 K, and 300 ps of a linear decrease in the temperature to 300 K. During these cycles, the experimental restraints are imposed as averages over the four replicas according to Eqs. 1–3. The total amount of sampling in each Gab2b peptide analyzed was 1 μ s (250.2 ns per replica equivalent to 417 cycles).

The simulations were carried out using the Amber03W force field (44) and the TIP4P2005 (45) water model. This force-field setting was optimized to account for the conformational properties of IDPs. The protonation states of pH-sensitive residues were as follows: Arg and Lys were positively charged, Asp and Glu were negatively charged, and His was neutral. The net charge of the system was neutralized by the addition of four Cl^- ions (three in the case of R515A and K518A). A time step of 2 fs was used together with LINCS constraints (46). van der Waals and electrostatic interactions were cut off at 0.9 nm, and long-range electrostatic effects were treated with the particle mesh Ewald method (47). All of the simulations were done in the canonical ensemble by keeping the volume fixed and setting the system temperature with a V-rescale thermostat (48).

We collected the final samplings from the 300 K portions of the replica simulations after discarding the initial 50 ns in each replica (in this time lapse, the simulations are allowed to converge). The total number of conformations collected in each sampling was 13,320. Convergence of the simulations is reported in Fig. S4 in the Supporting Material.

RESULTS

Conformational properties of Gab2₅₀₃₋₅₂₄

We used NMR spectroscopy, calorimetry, and statistical thermodynamics to characterize the molecular determinants of the Gab2₅₀₃₋₅₂₄/Grb2 SH3C interaction. Multiple lines of evidence have shown that the binding between these two domains is a key event in the interaction of the respective full-length proteins (27,36). ITC measurements show that the Gab2₅₀₃₋₅₂₄ fragment, which includes the RxxK binding motif of Gab2b, has micromolar affinity for Grb2 SH3C (see Table 1, Fig. S1, and the Materials and Methods section for details regarding the peptide synthesis and ITC measurements). Previous x-ray studies revealed that the

conformation adopted by the Gab2b motif when complexed to Grb2 SH3C (27) includes a short 3_{10} helix (residues 515–517) positioning the arginine and lysine of the RxxK motif in parallel orientation and a PPII fragment (residues 510–512; Fig. 1). Although the conformation of Gab2b bound to Grb2 SH3C is known, we focused on the conformational properties of Gab2₅₀₃₋₅₂₄ when isolated in solution to gain insight into the determinants for the binding affinity. To this end, we measured NMR CSs of backbone atoms (in particular C α , C β , H α , HN, and N) for Gab2₅₀₃₋₅₂₄ at pH 7.5 and 303 K (Fig. S2). NMR CSs of C α , C β , and H α atoms, which are fine probes of secondary structures in proteins, show values that are in the range of theoretical random-coil standards, as calculated using the CamCoil method (25) (Fig. S3). This finding indicates that Gab2₅₀₃₋₅₂₄ maintains the unstructured nature that it adopts within the full-length sequence (36). Although unstructured proteins and peptides are difficult to characterize with standard techniques of structural biology (i.e., x-ray crystallography and NOE/ 3 J-coupling-based NMR (49–52)), recent approaches exploiting the information extracted from CSs allow us to address, with residue-specific resolution, the population of transient secondary structures and coil regions.

We exploited the information contained in the CSs using the δ 2D method (22,25), which has proved to be a powerful tool for studying the conformational properties of IDPs (53–58), and calculated in this manner the secondary-structure populations in the disordered Gab2₅₀₃₋₅₂₄ fragment. This analysis revealed nonnegligible amounts of residual structure in the free state (Fig. 2). In particular, a region spanning residues 508–511 (designated throughout this article as residual structure 1 (RS1)), adopted partial PPII and extended β character with population peaks of up to

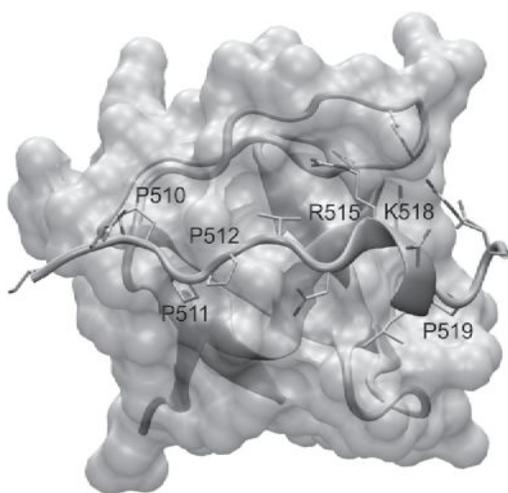


FIGURE 1 Binding interface between Grb2 SH3C (represented as a white surface/green cartoon) and Gab2b peptide (orange cartoon). The figure was produced using the x-ray coordinates (PDB code: 2vwf). To see this figure in color, go online.

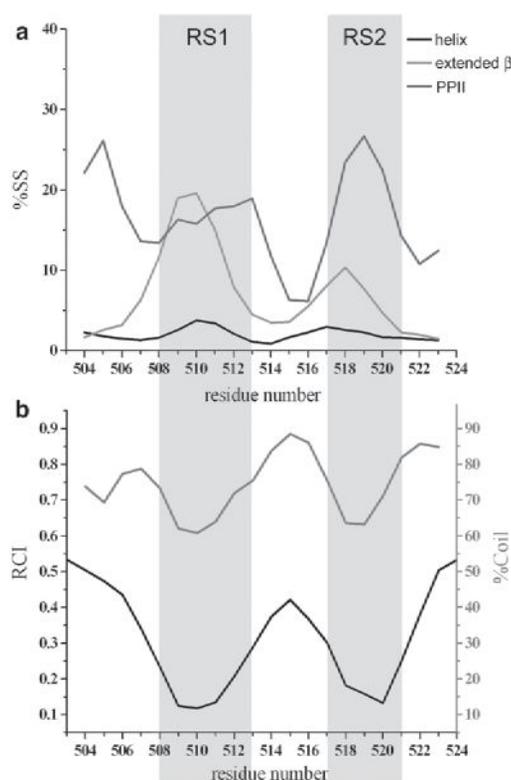


FIGURE 2 Conformational properties of the free state of Gab2₅₀₃₋₅₂₄ by analysis of CSs. (A) Populations of secondary-structure elements using δ 2D (22,25). Black, orange, and green lines report the populations of helix, extended β , and PPII conformations, respectively. (B) Contents of random-coil regions estimated by δ 2D (red) and RCI (26) (black line). The scale for the δ 2D profile is reported on the right side (red) and the scale for RCI appears on the left side (black). To see this figure in color, go online.

20% of the ensemble (Fig. 2 A). This finding suggests that, although RS1 is flexible and unstructured, it has a significant propensity to be highly extended when unbound in solution. Because of the extended β and PPII characters, RS1, which also shows a minor helical population (which is considered negligible because it is at the limit of the methodological error), is associated with a reduced coil character (Fig. 2 B). This finding was independently confirmed by the RCI (26) analysis (Fig. 2 B). Another region of the fragment that showed a significant content of residual PPII structure and partial character of an extended β conformation, namely RS2, spanned residues 517–521.

Residual structure in both RS1 and RS2 might have relevance in the context of binding the Grb2 SH3C. In RS1, the bound state of Gab2b consists of a PPII fragment (residues 510–512) that follows residue Q509, which is in an extended β conformation (backbone ϕ and ψ angles of -158.1 and 138.5 , respectively). On the other hand, a population of residual PPII in RS2 might play a role in the bound state by favoring an important kink starting at residue 518, and allowing R521 to fold back in the binding cavity to establish

a key salt bridge with Grb2 SH3C (Fig. 1). Notably K518 adopts backbone ϕ and ψ angles of -68.65 and 140.06 , respectively, which are typical values for the PPII region. Accordingly, the residual PPII conformation in the unbound state might preorganize the formation of the kink that stabilizes the complex.

Effect of point mutations on conformational selection

To further investigate the role that residual structure in the unbound state plays in the recognition of Gab2₅₀₃₋₅₂₄, we performed extensive refinement of its structural ensemble by using CS-restrained MD (42). The samplings (totaling 1 μ s for each construct) were performed by imposing CSs as ensemble averages over four replicas and were carried out until convergence was proved (Fig. S4). The resulting ensemble showed high agreement between experimental CSs and those calculated using an independent and highly accurate predictor, SPARTA+ (59), which is based on a totally different approach compared with the CamShift method (42) used in our structural refinement (Fig. S5). This back validation suggests that the refined ensembles are highly accurate and match the experimental data with standard deviations (SDs) that are lower than the statistical error of SPARTA+ (Fig. S5).

Our analyses focused on a comparison between the conformations of the free state (NMR ensemble) and those of the bound state (x-ray structure). We computed a free-energy projection of 13,320 conformations collected from the NMR sampling by using two independent coordinates. The first was built by calculating the root mean-square deviation (RMSD) of the ϕ and ψ backbone angles between the free-state structures and the bound conformation. The second coordinate was based on the gyration radius calculated using the C α atoms. To be consistent with the x-ray structure, only residues 508–521 were employed in this calculation. The analysis resulted in a spread free-energy landscape (Fig. 3 A), with the first coordinate ranging from approximately 20° to 100° in dihedral RMSD, and the second coordinate (gyration radius) ranging from 5.5 to 11.5 Å. In addition to a main basin of random-coil conformations (basin b1), the energy landscape showed a second minimum (basin b2) including conformations with low dihedral RMSD values ($\sim 40^\circ$) and gyration radii similar to that adopted in the bound structure (8.5 Å). The conformations corresponding to basin b2 are highly similar to that of the bound state (Fig. 3 A). Accordingly, in line with the CS analyses of $\delta 2D$ (22,25) and RCI (26), the energy landscape showed a set of preorganized conformations populating the free-state ensemble.

We then analyzed the influence of the basin b2 population on the ability of Grb2 to recognize and bind Gab2₅₀₃₋₅₂₄. In particular, we mutated a set of residues in Gab2₅₀₃₋₅₂₄ that are directly involved in the binding site of Grb2 SH3C

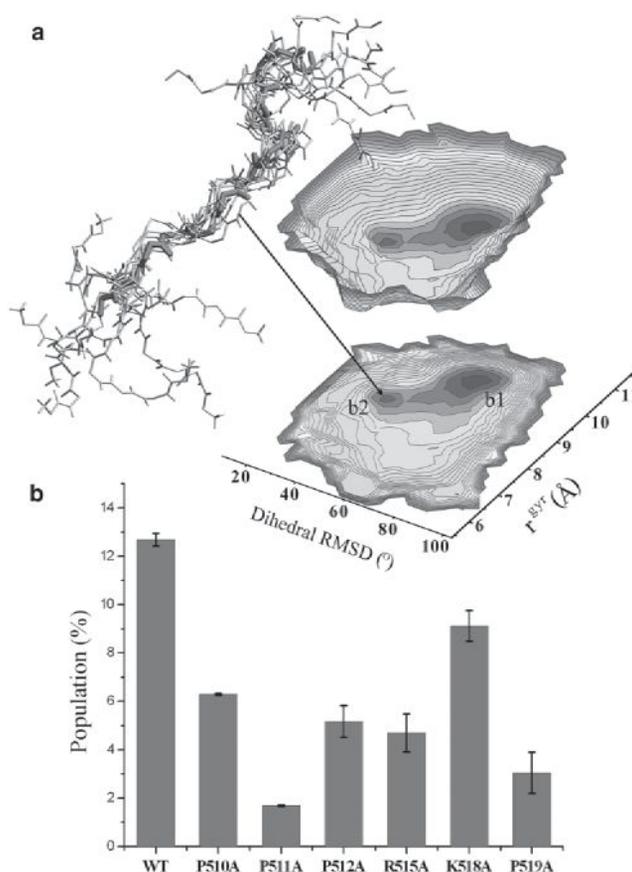


FIGURE 3 Structural ensembles of the free state of Gab2₅₀₃₋₅₂₄ sampled by CS-restrained simulations. (A) Energy landscape of the WT sequence projected on two coordinates: the dihedral RMSD from the x-ray structure of the bound state and the gyration radius. The landscape is evaluated in discrete points using the Boltzmann equation and then interpolated using spline functions (see De Simone et al. (2,60) for additional details regarding the free-energy projections). The plot shows a basin composed of structures that are very similar to the bound state, corresponding to a low RMSD and a radius of gyration that is close to 8.5 Å (the value adopted in the bound structure). A bundle of conformations from the basin b2 is shown (backbone atoms only). Gray conformations are from the NMR ensemble (free state) and the red conformation corresponds to the crystal structure (bound state). (B) Based on this basin, we could estimate the population of bound conformations in the structural ensembles of the free state of the Gab2₅₀₃₋₅₂₄ sequences (WT and mutants) considered here. To see this figure in color, go online.

(Fig. 1). ITC measurements provided direct evidence of the effects of these point mutations on the binding affinity. All mutations showed a decrease in the affinity for binding Grb2 SH3C, and a general trend that included an unfavorable entropy contribution and favorable enthalpy terms for the binding (Table 1). The strongest affinity impairment was measured for the mutations that affected the RxxK core motif, namely, R515A and K518A, for which no binding was detected. The other mutations tested targeted the proline residues in the binding site. Of these, P511A induced the strongest perturbation of the binding affinity (with a resulting K_d lower than the reliably quantifiable range), followed by P519A.

The $\delta 2D$ (22,25) analysis showed that all of the mutants retain a degree of disorder with P511A and K518A having an increased coil content in RS1 and RS2, respectively, and P519A being less structured in both RS1 and RS2 (Fig. 4 A). Similarly, the RCI (26) suggests that all the peptides exhibit a similar disorder, with P519A being the most dynamical mutant (Fig. 4 B). An interesting result is associated with the analysis of PPII content in the various mutants. P \rightarrow A mutations in RS1 affect mainly the PPII content in this region (Fig. 4 C). Of these, P511A shows the highest reduction in PPII structure, whereas in P510A the reduction of PPII is partially compensated for by a local increase in extended β content (Fig. 4 D). Another strong effect on the PPII content of RS1 is shown by R515A. Finally, K518A and P519A affect mainly the PPII in RS2, with P519A having the most significant perturbation, which is in part also transferred to RS1. No significant effects were detected for the helical residual structure (Fig. 4 E).

Overall, the residual structure in the mutants of Gab2₅₀₃₋₅₂₄ shows that the partial impairment of binding affinity for Grb2 is correlated with a reduced amount of pre-organized bound conformations in the free-state ensemble.

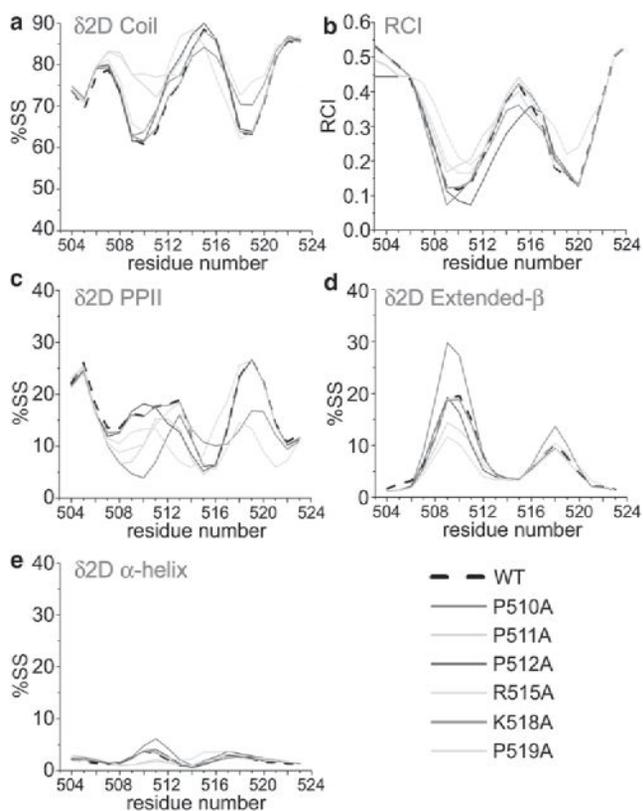


FIGURE 4 Effects of point mutations on the conformational properties of the free state of Gab2₅₀₃₋₅₂₄ by analysis of CSs. (A) Populations of coil conformations from $\delta 2D$ (22,25). (B) RCI (26). (C) Populations of PPII conformations from $\delta 2D$. (D) Populations of extended β conformations from $\delta 2D$. (E) Populations of helical conformations from $\delta 2D$. To see this figure in color, go online.

This important finding is underlined by the analysis of the populations in basin b2 of the projected energy landscape (Fig. 3). If we exclude R515A and K518A, which in addition to the conformational perturbation of the backbone lack the key side chains for the RxxK-binding motif, the resulting trend of populations is consistent with the ranking obtained from ITC measurements (Table 1). Indeed, the ranking of populations in the region occupied by basin b2 is WT > P510A > P512A > P519A > P511A (Fig. 3 B). This remarkable correspondence indicates that one of the key factors governing the interaction propensities of IDPs is the population of preexisting bound conformations in the free-state ensemble. It is worth noting that other factors may also contribute to the decrease of affinity between Grb2 and Gab2₅₀₃₋₅₂₄, including the loss of favorable native interactions in the protein-protein complex. For instance, these factors might contribute to the ~ 15 -fold loss of binding affinity associated with the P519A mutation (see K_d values in Table 1), which is not entirely accounted for by the decrease in the population of preexisting bound conformations in the energy landscape of Gab2₅₀₃₋₅₂₄ (Fig. 3 B).

DISCUSSION

It is now well established that a significant fraction of all proteins feature a partial or total degree of structural disorder (11,12,15,61). Disordered regions can influence intrinsic biological properties, such as the propensity to self-assemble (62) or allosteric modulation (63). It is also becoming increasingly evident that IDPs are actively employed in fundamental biological processes, including cellular signaling through disordered linkers (64) and protein translation or transcriptional regulation (65). Their remarkable ability to interact with multiple cellular partners is likely promoted by their inherent flexibility and energy landscapes encoding for multiple conformational minima, enabling shape adaptation during partner selection. Although the functional role of IDPs is attracting increasing interest, understanding the underlying structural and physical principles of their biological activity remains a difficult challenge. In this context, NMR spectroscopy is emerging as a major tool (17,18,22,66), and we illustrate this point very clearly by showing that it is possible to use new NMR approaches to address the functional determinants of the biological activity of IDPs.

One prominent representative of functional IDPs is Gab2, a platform protein involved in complex molecular signaling. Gab2 is a multisite docking molecule with a conserved, folded N-terminal domain that enables docking to the plasma membrane, and a large disordered region that mediates interactions with many signaling proteins. The long disordered region of Gab2 acts as a substrate region for multiple phosphorylations by tyrosine kinases. It also forms a multivalent docking station for SH2 and SH3 domains, which allows it to interact with a large number of

proteins from different families (e.g., Crk/CRKL, Nck1, Grb2, PI3 kinase, and SHP2). Thus, as found for other IDPs, Gab2 behaves as a versatile hub for signaling pathway cross talk in several different cell types. It acts in many physiological processes, including growth of the bone marrow and cardiac function, as well as in pathological conditions such as leukemia and Alzheimer's disease (67). Besides representing a convenient system in which to address the binding and recognition of IDPs, Gab2 is therefore a central node in cellular signaling of major biological relevance.

Using the information contained in the backbone CSs, the RCI (26) showed two regions of Gab2₅₀₃₋₅₂₄, RS1 and RS2, with reduced conformational flexibility. The high occurrence of PPII and extended β structure in these regions, as revealed by δ 2D (22,25), suggests a nonnegligible intrinsic tendency toward extended conformations. PPII conformations in RS1 are directly relevant to the conformation adopted in the complex with Grb2 SH3C (PDB code: 2vwf). Similarly, we propose that the marked residual content of PPII in RS2 has a role in nucleating an essential kink that conveys to the backbone an optimal conformation for a salt bridge made by residue R521. The intrinsic propensity to adopt residual structures converges with the evidence that the energy landscape of Gab2₅₀₃₋₅₂₄ includes a basin of conformers with similar structural properties to the bound state. As shown by the mutant analysis of Gab2₅₀₃₋₅₂₄, changes in the population of these conformers correlate with the impairment in the Grb2 binding affinity.

Overall, our analysis of transient conformations by NMR and statistical mechanics indicates that the presence of prearranged bound conformations in the heterogeneous structural ensemble of the disordered Gab2₅₀₃₋₅₂₄ is an essential determinant of the binding affinity for its protein partner Grb2, as tested by mutational analysis and calorimetric measurements. Therefore, the results reported here suggest that the binding of Gab2₅₀₃₋₅₂₄ requires a selection of conformations that appear to be intrinsically encoded in the energy landscape of this disordered state.

In conclusion, by using recently developed methods that exploit the information contained in NMR CSs, we investigated the conformational properties of disordered Gab2₅₀₃₋₅₂₄, and found intrinsic conformations that influence the affinity for protein partner recognition and binding. Thus, the physical principles that emerge from this study add to our understanding of the recognition of IDPs by showing that preexisting bound conformations have a significant influence on the binding affinity. The elements reported in this work are likely to have very general implications for the biological behavior of IDPs in light of the fact that a large fraction of these proteins possess a specific propensity to form the PPII helix and adopt conformations that are more extended than the typical random-coil states.

SUPPORTING MATERIAL

Five figures are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)00271-9](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00271-9).

We are grateful to the Tate group (Department of Chemistry, Imperial College London) for assisting with the peptide synthesis and purification.

This research was supported by the Centre for Structural Biology at Imperial College, Leverhulme Trust (A.De S.), Parkinson's UK (G.F.), and FP7 Marie Curie Actions (C.C.).

REFERENCES

- Boehr, D. D., R. Nussinov, and P. E. Wright. 2009. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* 5:789–796.
- De Simone, A., R. W. Montalvo, ..., M. Vendruscolo. 2013. Characterization of the interdomain motions in hen lysozyme using residual dipolar couplings as replica-averaged structural restraints in molecular dynamics simulations. *Biochemistry.* 52:6480–6486.
- Boehr, D. D., H. J. Dyson, and P. E. Wright. 2006. An NMR perspective on enzyme dynamics. *Chem. Rev.* 106:3055–3079.
- Masterson, L. R., A. Mascioni, ..., G. Veglia. 2008. Allosteric cooperativity in protein kinase A. *Proc. Natl. Acad. Sci. USA.* 105:506–511.
- Das, R., V. Esposito, ..., G. Melacini. 2007. cAMP activation of PKA defines an ancient signaling mechanism. *Proc. Natl. Acad. Sci. USA.* 104:93–98.
- Tzeng, S. R., and C. G. Kalodimos. 2013. Allosteric inhibition through suppression of transient conformational states. *Nat. Chem. Biol.* 9:462–465.
- Tzeng, S. R., and C. G. Kalodimos. 2012. Protein activity regulation by conformational entropy. *Nature.* 488:236–240.
- Selvaratnam, R., S. Chowdhury, ..., G. Melacini. 2011. Mapping allostery through the covariance analysis of NMR chemical shifts. *Proc. Natl. Acad. Sci. USA.* 108:6133–6138.
- Arai, M., J. C. Ferreon, and P. E. Wright. 2012. Quantitative analysis of multisite protein-ligand interactions by NMR: binding of intrinsically disordered p53 transactivation subdomains with the TAZ2 domain of CBP. *J. Am. Chem. Soc.* 134:3792–3803.
- Tompa, P. 2012. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* 37:509–516.
- Uversky, V. N. 2013. A decade and a half of protein intrinsic disorder: biology still waits for physics. *Protein Sci.* 22:693–724.
- Dunker, A. K., Z. Obradovic, ..., C. J. Brown. 2000. Intrinsic protein disorder in complete genomes. *Genome Inform. Ser. Workshop Genome Inform.* 11:161–171.
- Dunker, A. K., I. Silman, ..., J. L. Sussman. 2008. Function and structure of inherently disordered proteins. *Curr. Opin. Struct. Biol.* 18:756–764.
- Uversky, V. N. 2011. Multitude of binding modes attainable by intrinsically disordered proteins: a portrait gallery of disorder-based complexes. *Chem. Soc. Rev.* 40:1623–1634.
- Tompa, P. 2011. Unstructural biology coming of age. *Curr. Opin. Struct. Biol.* 21:419–425.
- Soranno, A., B. Buchli, ..., B. Schuler. 2012. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA.* 109:17800–17806.
- Dyson, H. J., and P. E. Wright. 2004. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* 104:3607–3622.
- Mittag, T., and J. D. Forman-Kay. 2007. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* 17:3–14.
- Schneider, R., J. R. Huang, ..., M. Blackledge. 2012. Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol. Biosyst.* 8:58–68.

20. Milojevic, J., V. Esposito, ..., G. Melacini. 2007. Understanding the molecular basis for the inhibition of the Alzheimer's A β -peptide oligomerization by human serum albumin using saturation transfer difference and off-resonance relaxation NMR spectroscopy. *J. Am. Chem. Soc.* 129:4282–4290.
21. De Simone, A., M. Gustavsson, ..., M. Vendruscolo. 2013. Structures of the excited states of phospholamban and shifts in their populations upon phosphorylation. *Biochemistry.* 52:6684–6694.
22. Camilloni, C., A. De Simone, ..., M. Vendruscolo. 2012. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry.* 51:2224–2231.
23. Mittag, T., J. Marsh, ..., J. D. Forman-Kay. 2010. Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure.* 18:494–506.
24. Jensen, M. R., P. R. L. Markwick, ..., M. Blackledge. 2009. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure.* 17:1169–1185.
25. De Simone, A., A. Cavalli, ..., M. Vendruscolo. 2009. Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins. *J. Am. Chem. Soc.* 131:16332–16333.
26. Berjanskii, M. V., and D. S. Wishart. 2007. The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Res.* 35 (Web Server issue):W531–W537.
27. Harkiolaki, M., T. Tsirka, ..., S. M. Feller. 2009. Distinct binding modes of two epitopes in Gab2 that interact with the SH3C domain of Grb2. *Structure.* 17:809–822.
28. Lewitzky, M., C. Kardinal, ..., S. M. Feller. 2001. The C-terminal SH3 domain of the adapter protein Grb2 binds with high affinity to sequences in Gab1 and SLP-76 which lack the SH3-typical P-x-x-P core motif. *Oncogene.* 20:1052–1062.
29. Lock, L. S., I. Royal, ..., M. Park. 2000. Identification of an atypical Grb2 carboxyl-terminal SH3 domain binding site in Gab docking proteins reveals Grb2-dependent and -independent recruitment of Gab1 to receptor tyrosine kinases. *J. Biol. Chem.* 275:31536–31545.
30. McDonald, C. B., K. L. Seldeen, ..., A. Farooq. 2011. Binding of the cSH3 domain of Grb2 adaptor to two distinct RXXK motifs within Gab1 docker employs differential mechanisms. *J. Mol. Recognit.* 24:585–596.
31. Eulendorf, R., and F. Schaper. 2009. A new mechanism for the regulation of Gab1 recruitment to the plasma membrane. *J. Cell Sci.* 122:55–64.
32. Nishida, K., L. Wang, ..., T. Hirano. 2002. Requirement of Gab2 for mast cell development and KitL/c-Kit signaling. *Blood.* 99:1866–1869.
33. Nishida, K., Y. Yoshida, ..., T. Hirano. 1999. Gab-family adapter proteins act downstream of cytokine and growth factor receptors and T- and B-cell antigen receptors. *Blood.* 93:1809–1816.
34. McDonald, C. B., V. Bhat, ..., A. Farooq. 2012. Bivalent binding drives the formation of the Grb2-Gab1 signaling complex in a noncooperative manner. *FEBS J.* 279:2156–2173.
35. McDonald, C. B., K. L. Seldeen, ..., A. Farooq. 2010. Assembly of the Sos1-Grb2-Gab1 ternary signaling complex is under allosteric control. *Arch. Biochem. Biophys.* 494:216–225.
36. Simister, P. C., and S. M. Feller. 2012. Order and disorder in large multi-site docking proteins of the Gab family—implications for signaling complex formation and inhibitor design strategies. *Mol. Biosyst.* 8:33–46.
37. Chan, W. C., and P. D. White. 2000. Fmoc Solid Phase Peptide Synthesis: A Practical Approach. Oxford University Press, Oxford, UK.
38. Pendergast, A. M., L. A. Quilliam, L. D. Cripe, C. H. Bassing, Z. Dai, N. Li, A. Batzer, K. M. Rabun, C. J. Der, J. Schlessinger, ..., 1993. BCR-ABL-induced oncogenesis is mediated by direct interaction with the SH2 domain of the GRB-2 adaptor protein. *Cell.* 75:175–185.
39. Wüthrich, K. 1989. Protein structure determination in solution by nuclear magnetic resonance spectroscopy. *Science.* 243:45–50.
40. Kohlhoff, K. J., P. Robustelli, ..., M. Vendruscolo. 2009. Fast and accurate predictions of protein NMR chemical shifts from interatomic distances. *J. Am. Chem. Soc.* 131:13894–13895.
41. Hess, B., C. Kutzner, ..., E. Lindahl. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular dynamics simulations. *J. Chem. Theory Comput.* 4:435–437.
42. Camilloni, C., P. Robustelli, ..., M. Vendruscolo. 2012. Characterization of the conformational equilibrium between the two major substates of RNase A using NMR chemical shifts. *J. Am. Chem. Soc.* 134:3968–3971.
43. Camilloni, C., A. Cavalli, and M. Vendruscolo. 2013. Assessment of the use of NMR chemical shifts as replica-averaged structural restraints in molecular dynamics simulations to characterize the dynamics of proteins. *J. Phys. Chem. B.* 117:1838–1843.
44. Best, R. B., and J. Mittal. 2010. Protein simulations with an optimized water model: cooperative helix formation and temperature-induced unfolded state collapse. *J. Phys. Chem. B.* 114:14916–14923.
45. Abascal, J. L., E. Sanz, ..., C. Vega. 2005. A potential model for the study of ices and amorphous water: TIP4P/Ice. *J. Chem. Phys.* 122:234511.
46. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* 18:1463–1472.
47. Darden, T., L. Perera, ..., L. Pedersen. 1999. New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations. *Structure.* 7:R55–R60.
48. Bussi, G., D. Donadio, and M. Parrinello. 2007. Canonical sampling through velocity rescaling. *J. Chem. Phys.* 126:014101.
49. Gil, S., T. Hošek, ..., I. C. Felli. 2013. NMR spectroscopic studies of intrinsically disordered proteins at near-physiological conditions. *Angew. Chem. Int. Ed. Engl.* 52:11808–11812.
50. Camilloni, C., D. Schaal, ..., A. De Simone. 2012. Energy landscape of the prion protein helix 1 probed by metadynamics and NMR. *Biophys. J.* 102:158–167.
51. Dyson, H. J., and P. E. Wright. 1991. Defining solution conformations of small linear peptides. *Annu. Rev. Biophys. Biophys. Chem.* 20:519–538.
52. Dyson, H. J., and P. E. Wright. 2001. Nuclear magnetic resonance methods for elucidation of structure and dynamics in disordered states. *Methods Enzymol.* 339:258–270.
53. Maltsev, A. S., J. Ying, and A. Bax. 2012. Impact of N-terminal acetylation of α -synuclein on its random coil and lipid binding properties. *Biochemistry.* 51:5004–5013.
54. Voelz, V. A., M. Jäger, ..., V. S. Pande. 2012. Slow unfolded-state structuring in Acyl-CoA binding protein folding revealed by simulation and experiment. *J. Am. Chem. Soc.* 134:12565–12577.
55. Roche, J., J. Ying, ..., A. Bax. 2013. Impact of hydrostatic pressure on an intrinsically disordered protein: a high-pressure NMR study of α -synuclein. *ChemBioChem.* 14:1754–1761.
56. Ozenne, V., R. Schneider, ..., M. Blackledge. 2012. Mapping the potential energy landscape of intrinsically disordered proteins at amino acid resolution. *J. Am. Chem. Soc.* 134:15138–15148.
57. Kang, L., G. M. Moriarty, ..., J. Baum. 2012. N-terminal acetylation of α -synuclein induces increased transient helical propensity and decreased aggregation rates in the intrinsically disordered monomer. *Protein Sci.* 21:911–917.
58. Libich, D. S., N. L. Fawzi, ..., G. M. Clore. 2013. Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR. *Proc. Natl. Acad. Sci. USA.* 110:11361–11366.
59. Shen, Y., and A. Bax. 2010. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J. Biomol. NMR.* 48:13–22.

60. De Simone, A., A. Dhulesia, ..., C. M. Dobson. 2011. Experimental free energy surfaces reveal the mechanisms of maintenance of protein solubility. *Proc. Natl. Acad. Sci. USA*. 108:21057–21062.
61. Tompa, P., and D. Kovacs. 2010. Intrinsically disordered chaperones in plants and animals. *Biochem. Cell Biol.* 88:167–174.
62. De Simone, A., C. Kitchen, ..., D. Frenkel. 2012. Intrinsic disorder modulates protein self-assembly and aggregation. *Proc. Natl. Acad. Sci. USA*. 109:6951–6956.
63. Ferreon, A. C., J. C. Ferreon, ..., A. A. Deniz. 2013. Modulation of allostery by protein intrinsic disorder. *Nature*. 498:390–394.
64. Akimoto, M., R. Selvaratnam, ..., G. Melacini. 2013. Signaling through dynamic linkers as revealed by PKA. *Proc. Natl. Acad. Sci. USA*. 110:14231–14236.
65. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197–208.
66. Jensen, M. R., K. Houben, ..., M. Blackledge. 2008. Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J. Am. Chem. Soc.* 130:8055–8061.
67. Reiman, E. M., J. A. Webster, ..., D. A. Stephan. 2007. GAB2 alleles modify Alzheimer's risk in APOE epsilon4 carriers. *Neuron*. 54:713–720.

The Berg-Purcell Limit Revisited

Kazunari Kaizu,[†] Wiet de Ronde,[‡] Joris Pajmians,[‡] Koichi Takahashi,[†] Filipe Tostevin,[‡] and Pieter Rein ten Wolde^{†*}

[†]RIKEN Quantitative Biology Center (QBIC), RIKEN, 6-2-3 Furuedai, Suita, Osaka 565-0874, Japan; and [‡]FOM Institute AMOLF, Science Park 104, 1098 XG, Amsterdam, The Netherlands

ABSTRACT Biological systems often have to measure extremely low concentrations of chemicals with high precision. When dealing with such small numbers of molecules, the inevitable randomness of physical transport processes and binding reactions will limit the precision with which measurements can be made. An important question is what the lower bound on the noise would be in such measurements. Using the theory of diffusion-influenced reactions, we derive an analytical expression for the precision of concentration estimates that are obtained by monitoring the state of a receptor to which a diffusing ligand can bind. The variance in the estimate consists of two terms, one resulting from the intrinsic binding kinetics and the other from the diffusive arrival of ligand at the receptor. The latter term is identical to the fundamental limit derived by Berg and Purcell (Biophys. J., 1977), but disagrees with a more recent expression by Bialek and Setayeshgar. Comparing the theoretical predictions against results from particle-based simulations confirms the accuracy of the resulting expression and reaffirms the fundamental limit established by Berg and Purcell.

INTRODUCTION

The evidence is accumulating that sensory systems in biology often operate near the fundamental limit set by the noise of counting signal molecules. Receptors in our visual system can detect single photons (1), some animals can smell single molecules (2), swimming bacteria can respond to the binding and unbinding of only a limited number of molecules (3,4), and eukaryotic cells can respond to a difference in ~10 molecules between the front and the back of the cell (5). Recent experiments suggest that the precision of the embryonic development of the fruitfly *Drosophila* is close to the limit set by the available number of regulatory proteins (6–8). This raises the question of what is the fundamental limit to the precision of chemical concentration measurements.

In their classic article, Berg and Purcell (3) considered a scenario in which a cell measures the concentration c of a ligand by monitoring the occupation state of the receptor molecules to which the ligand molecules bind and unbind. A central result is the precision with which the ligand concentration c can be inferred from the time-averaged occupancy of a single receptor. The analysis of Berg and Purcell predicts that in the limit at which the integration time T is much longer than the correlation time of the receptor state, the expected uncertainty in the time-averaged occupancy resulting from the random nature of diffusion is given by

$$\frac{\delta n}{\bar{n}} = \sqrt{\frac{(1 - \bar{n})}{2D\sigma cT}}, \quad (1)$$

where \bar{n} is the true mean occupancy, σ is the receptor-ligand binding cross section, and D is the diffusion constant of the ligand. The uncertainty or expected error in a corresponding estimate of the concentration is related to the noise in the observed occupancy via the gain $d\bar{n}/dc$,

$$\delta c = \frac{dc}{d\bar{n}} \delta n, \quad (2)$$

yielding Berg and Purcell's expression for the limit to the precision of concentration measurements by a single receptor (see Eq. 52 in Berg and Purcell (3)):

$$\frac{\delta c}{c} = \sqrt{\frac{2}{4D\sigma c(1 - \bar{n})T}}. \quad (3)$$

This result can be understood intuitively by noting that $4D\sigma c$ is the flux of ligand molecules arriving at the receptor and $1 - \bar{n}$ is the probability that the receptor is free (3). Therefore, $4D\sigma c(1 - \bar{n})$ is the effective rate of ligand binding if every collision between ligand and free receptor leads to successful binding. Berg and Purcell (3) argue that their result also holds for reactions that are not deeply in the diffusion-limited regime. After an unsuccessful receptor-ligand encounter, they argue, the ligand will rapidly collide with the receptor again and again until it binds the receptor, and these rounds of encounters can be captured by renormalizing σ .

The argument of Berg and Purcell ignores, however, that after an unsuccessful collision with the receptor the ligand molecule may diffuse back into the bulk, and a different ligand molecule may subsequently bind. Moreover, a ligand molecule that has just dissociated from the receptor may

Submitted September 22, 2013, and accepted for publication December 19, 2013.

*Correspondence: tenwolde@amolf.nl

Editor: Sean Sun.

© 2014 by the Biophysical Society
0006-3495/14/02/0976/10 \$2.00

<http://dx.doi.org/10.1016/j.bpj.2013.12.030>



rapidly rebind, or it may diffuse away from the receptor into the bulk. It thus remains unclear to what extent the result of Berg and Purcell applies to binding reactions that are not diffusion-limited.

Bialek and Setayeshgar (9) sought to generalize the result of Berg and Purcell by taking into account ligand-receptor binding dynamics. They considered a model in which the ligand molecules can diffuse, bind the receptor upon contact with an intrinsic association rate k_a , and unbind from the receptor with an intrinsic dissociation rate k_d . Invoking the fluctuation-dissipation theorem, they linearized the nonlinear reaction-diffusion equation, to obtain the following result for the fractional uncertainty in the estimate for the concentration (Eq. 32 in (9)):

$$\frac{\delta c}{c} = \sqrt{\frac{1}{\pi D \sigma c T} + \frac{2}{k_a c (1 - \bar{n}) T}}. \quad (4)$$

The first term arises from the stochastic arrival of the ligand molecules at the receptor by diffusion, whereas the second term is due to the intrinsic stochasticity of the binding kinetics of the receptor. Indeed, even in the limit that $D \rightarrow \infty$, such that the concentration at the receptor is constant, this concentration can still not be measured with infinite precision because the receptor stochastically switches between the bound and unbound states, leading to noise in the estimate of the receptor occupancy. This term is absent in Eq. 3 because Berg and Purcell assume that the binding reaction is fully diffusion-limited, meaning that the intrinsic rates k_a and k_d go to infinity.

The first term of Eq. 4 should be compared with Eq. 3. This term is considered to be the fundamental limit to the accuracy of measuring chemical concentrations via a single receptor, because it presents a noise floor that is solely due to the physics of diffusion, independent of the binding kinetics (9). Indeed, in the limit that the reaction is diffusion-limited, the second term in Eq. 4 is zero, and both theories should yield the same result. However, it is clear that in addition to the geometrical factor π (which comes from the fact that Berg and Purcell model the receptor as a reactive disk, whereas Bialek and Setayeshgar take the receptor to be a spherical particle), the expressions differ by a factor $1/(2(1 - \bar{n}))$. This difference can have marked implications. Although the Bialek-Setayeshgar expression predicts that the uncertainty remains bounded even in the limit that $\bar{n} \rightarrow 1$, the Berg-Purcell expression suggests that it diverges in this limit.

Here, we rederive the limit to the accuracy of sensing via a single receptor (10), borrowing heavily from the work of Agmon and Szabo (11) on diffusion-influenced reactions. Our expression is identical to that of Berezhkovskii and Szabo (12), who have recently independently derived this limit for an arbitrary number of receptors, when there is one receptor. Like the expression of Bialek and Setayeshgar

(Eq. 4), our expression consists of two terms: One term describes the effect of the diffusive transport of the ligand molecules to and from the receptor, and the other describes the effect of the intrinsic binding and unbinding kinetics of the receptor. Although the second term agrees with that of Bialek and Setayeshgar, the first does not agree with their expression but does agree with the expression of Berg and Purcell (again apart from the geometric factor).

We then perform extensive tests of these expressions by performing particle-based simulations using Green's function reaction dynamics (GFRD), which is an exact scheme for simulating reaction-diffusion systems at the particle level (13–15). The simulation results agree very well with our expression and that of Berezhkovskii and Szabo (12) for the full range of conditions that we considered, which spans the biologically relevant regime. This means that the Berg-Purcell limit is the most accurate expression for the fundamental limit to measuring chemical concentrations.

We end by examining the assumptions of our theory under biologically relevant conditions. This naturally suggests a simple but intuitive model. This model not only explains the origin of the factor $1/(2(1 - \bar{n}))$ in the Berg-Purcell expression, but also shows how their expression can be generalized to reactions that are not diffusion-limited by integrating out the rapid rebindings of dissociated molecules. The model also elucidates that rebindings do not contribute to the accuracy of sensing, because their likelihood does not depend on the concentration.

METHODS AND THEORY

We consider a single receptor A in a volume V that is surrounded by a large number N_B of noninteracting ligand molecules B at concentration $c = N_B/V$. We consider the pseudo first-order limit, meaning that $N_B \gg N_A = 1$ and $V \rightarrow \infty$. Without loss of generality, we may assume that the receptor is static and located at the origin, while the ligand molecules diffuse with diffusion constant D . A ligand molecule can bind a free receptor with an intrinsic association rate k_a when the two come in contact at the contact distance σ , which is the sum of the radii of the two respective molecules. A bound ligand molecule can dissociate from the receptor with an intrinsic dissociation rate k_d . The state of the receptor is denoted by the binary variable $n(t)$, which is one if the receptor is bound to a ligand at time t and zero otherwise. We note that this model is identical to that of Bialek and Setayeshgar (9) for the scenario of a single receptor molecule.

Following Berg and Purcell (3) and Bialek and Setayeshgar (9), we imagine that the cell estimates the concentration c from the receptor occupancy $n(t)$ integrated over an integration time T , $n_T = T^{-1} \int_0^T n(t) dt$. In the limit that the integration time T is much longer than the correlation time of $n(t)$, τ_n , the variance in our estimate n_T of the true mean occupancy \bar{n} is given by

$$(\delta n)^2 \approx \frac{2\sigma_n^2 \tau_n}{T} = \frac{P_n(\omega = 0)}{T} = \frac{2\text{Re}[\widehat{C}_n(s = 0)]}{T}, \quad (5)$$

where $\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2 = \bar{n}(1 - \bar{n})$ is the instantaneous variance and $P_n(\omega)$ and $\widehat{C}_n(s)$ are, respectively, the power spectrum and the Laplace transform of the correlation function $C_n(t)$ of $n(t)$. The uncertainty in the estimate for the concentration c can then be obtained from Eqs. 2 and 5. In Eq. 2, the

gain is $dc/d\bar{n} = c/(\bar{n} - \bar{n}^2)$, because $\bar{n} = c/(c + K_D)$ with K_D the receptor-ligand dissociation constant.

The correlation function of any binary switching process is given by

$$C_n(\tau) = p_*^0 \left(p_{*|*}(\tau) - p_*^0 \right), \quad (6)$$

where $p_*^0 \equiv \bar{n}$ is the equilibrium probability for the bound state ($*$) and $p_{*|*}(\tau) = \langle n(\tau)n(0) \rangle / \bar{n}$ is the probability the receptor is bound at $t = \tau$, given it was bound at $t = 0$. To obtain the correlation function, we thus need $p_{*|*}(\tau)$. It is convenient to focus on the conjugate probability

$$\mathcal{P}_{\text{rev}}(t|*) = 1 - p_{*|*}(t), \quad (7)$$

which is the probability that the receptor is free at time t given that it was bound at $t = 0$. Following Agmon and Szabo (11), we use the subscript “rev” to indicate that we consider a reversible reaction, meaning that in between $t = 0$ and t the receptor may bind and unbind ligand a number of times. The probability that a receptor-ligand pair dissociates between t' and $t' + dt'$ to form an unbound pair at contact is $k_d[1 - \mathcal{P}_{\text{rev}}(t'|*)]dt'$, whereas the probability that the free receptor with a ligand molecule at contact at time t' is still unbound at time $t > t'$ is $\mathcal{P}_{\text{rad}}(t - t'|\sigma)$. The subscript “rad” means that we now consider an irreversible reaction ($k_d = 0$), which can be obtained by solving the diffusion equation using a radiation boundary condition (11). Hence, $\mathcal{P}_{\text{rev}}(t|*)$ is given by (11)

$$\mathcal{P}_{\text{rev}}(t|*) = k_d \int_0^t [1 - \mathcal{P}_{\text{rev}}(t'|*)] \mathcal{P}_{\text{rad}}(t - t'|\sigma) dt'. \quad (8)$$

We emphasize that up to this point no approximation has been made. The question now is what is $\mathcal{P}_{\text{rad}}(t|\sigma)$, which is the quantity needed to solve Eq. 8. To address this, we introduce two new quantities:

1. $\mathcal{P}_{\text{rad}}(t|\text{eq})$, which is the probability that a receptor initially is free and surrounded by an equilibrium distribution of ligand molecules, and it remains free until at least a later time t ; and
2. $S_{\text{rad}}(t|\sigma)$, which is the probability that a free receptor initially surrounded by only one single ligand molecule at contact is still unbound at a later time t .

The quantity $S_{\text{rad}}(t|\sigma)$ thus refers to a system consisting of a receptor with only one ligand molecule, which initially is at contact with the receptor, whereas $\mathcal{P}_{\text{rad}}(t|\sigma)$ refers to a system of a receptor with $N_B \geq 1$ ligand molecules, one of which is at contact initially.

In general, it is not possible to obtain an exact analytical expression for $\mathcal{P}_{\text{rad}}(t|\sigma)$, the quantity that we need (Eq. 8). To illustrate this, imagine a bound receptor-ligand pair that is surrounded by an equilibrium, i.e., a statistically uniform distribution of ligand particles. When this receptor-ligand pair dissociates to form a receptor-ligand pair at contact surrounded by an equilibrium distribution of ligand molecules, then the probability that the receptor is still unbound at a later time t is given by (11)

$$\mathcal{P}_{\text{rad}}(t|\sigma) = \mathcal{P}_{\text{rad}}(t|\text{eq}) S_{\text{rad}}(t|\sigma). \quad (9)$$

Now, the ligand molecule at contact may either rebind the receptor or diffuse away from it. If it rebinds the receptor, then after the next dissociation event, the probability that the receptor will remain free for at least another time t will again be given by Eq. 9. Equation 9 breaks down when the ligand molecule at contact instead diffuses away from the receptor and another ligand molecule binds the receptor before the first ligand has relaxed to equilibrium. Indeed, the process of receptor binding generates nontrivial spatio-temporal correlations between the positions of the ligand molecules, which depend on the history of the association and dissociation events. This impedes an exact solution of the problem. However, if the dissociation rate k_d is low then it becomes reasonable to assume that after

each dissociation event, the unbound receptor-ligand pair at contact is surrounded by an equilibrium distribution of ligand (11), in which case the survival probability is given by Eq. 9. This is the crucial assumption that we make in our analysis.

With the assumption of Eq. 9, Eq. 8 can now be solved. For a pseudo-first-order irreversible reaction with a static target, $\mathcal{P}_{\text{rad}}(t|\text{eq})$ of Eq. 9 is given by Rice (16) (see also the Supporting Material) as

$$\mathcal{P}_{\text{rad}}(t|\text{eq}) = e^{-c \int_0^t k_{\text{rad}}(t') dt'}, \quad (10)$$

where $k_{\text{rad}}(t)$ is the time-dependent rate coefficient. Here, $ck_{\text{rad}}(t)$ is the rate at which ligand molecules will bind a free receptor at time t , given that the ligand distribution at time $t = 0$ was the uniform, equilibrium, distribution. As before, the subscript “rad” refers to a radiation boundary condition, meaning that if a ligand molecule and the receptor come into contact, they react with a finite rate k_a (11). The quantity $S_{\text{rad}}(t|\sigma)$ in Eq. 9 is via detailed balance (and the backward Smoluchowski equation) related to $k_{\text{rad}}(t)$: $k_{\text{rad}}(t) = k_a S_{\text{rad}}(t|\sigma)$ (11) (see also the Supporting Material). Together, these relations yield a simple expression for the Laplace transform of $\mathcal{P}_{\text{rad}}(t|\sigma)$ in terms of the Laplace transform $\hat{\mathcal{P}}_{\text{rad}}(s|\text{eq})$ of $\mathcal{P}_{\text{rad}}(t|\text{eq})$ (see the Supporting Material). Substituting this in the solution of Eq. 8 in the Laplace domain allows us to obtain the following expression for the Laplace transform of the correlation function in terms of $\hat{\mathcal{P}}_{\text{rad}}(s|\text{eq})$ (see the Supporting Material):

$$\hat{C}_n(s) = \sigma_n^2 \frac{ck_a \tau_c \hat{\mathcal{P}}_{\text{rad}}(s|\text{eq})}{1 - k_d \tau_c s \hat{\mathcal{P}}_{\text{rad}}(s|\text{eq})}, \quad (11)$$

where $\tau_c = (k_a c + k_d)^{-1}$ is the correlation time of the intrinsic receptor switching dynamics, i.e., the correlation time of the receptor occupancy when receptor-ligand association is reaction-limited and the effect of diffusion can be neglected.

To obtain an analytically closed form for the correlation function, we require an expression for $\hat{\mathcal{P}}_{\text{rad}}(s|\text{eq})$. We use

$$s \hat{\mathcal{P}}_{\text{rad}}(s|\text{eq}) \approx (1 + c \hat{k}_{\text{rad}}(s))^{-1}, \quad (12)$$

which correctly captures the short- and long-time limit of $\mathcal{P}_{\text{rad}}(t|\text{eq})$ (11) (see the Supporting Material). We exploit that

$$\hat{k}_{\text{rad}}(s) = k_a \hat{k}_{\text{abs}}(s) / (k_a + s \hat{k}_{\text{abs}}(s))$$

(see the Supporting Material), where

$$\hat{k}_{\text{abs}}(s) = 4\pi\sigma D \left(1 + \sigma \sqrt{s/D} \right) / s$$

is the Laplace transform of the time-dependent diffusion-limited rate constant $k_{\text{abs}}(t)$ (16). Here, the subscript “abs” refers to the absorbing boundary condition, meaning that $k_a \rightarrow \infty$, where each ligand-receptor collision will immediately lead to binding; $ck_{\text{abs}}(t)$ is thus the rate at which ligand molecules collide and associate with the receptor at time t , given that they start from the equilibrium (uniform) distribution. Substituting Eq. 12 in Eq. 11 yields (see the Supporting Material)

$$\hat{C}_n(s) = \sigma_n^2 \frac{\tau'_c(s)}{s \tau'_c(s) + 1}, \quad (13)$$

where $\tau'_c(s)$ is the intrinsic correlation time τ_c renormalized by the concentration fluctuations (9),

$$\tau_c'(s) = \tau_c(1 + \Sigma(s)); \Sigma(s) = \frac{k_a}{k_D(1 + \sqrt{s\tau_m})}, \quad (14)$$

with the diffusion-limited rate constant $k_D = k_{\text{abs}}(t \rightarrow \infty) = 4\pi\sigma D$ and the molecular timescale $\tau_m = \sigma^2/D$. The correlation time τ_n of the receptor is then given by $\tau_n = (\sigma_n^2)^{-1} \widehat{C}_n(s=0)$ (Eq. 5) as

$$\tau_n = \frac{1}{k_{\text{on}}c + k_{\text{off}}}, \quad (15)$$

where k_{on} and k_{off} are the renormalized association and dissociation rates

$$k_{\text{on}} = \left(\frac{1}{k_a} + \frac{1}{k_D} \right)^{-1} = \frac{k_a k_D}{k_a + k_D}, \quad (16)$$

$$k_{\text{off}} = \left(\frac{1}{k_d} + \frac{K_{\text{eq}}}{k_D} \right)^{-1} = \frac{k_d k_D}{k_a + k_D}, \quad (17)$$

and $K_{\text{eq}} = k_a/k_d = K_D^{-1}$ is the equilibrium constant (11). The effective association rate k_{on} is the long-time limit of the time-dependent rate coefficient $k_{\text{rad}}(t)$: $k_{\text{on}} = k_{\text{rad}}(t \rightarrow \infty)$; it takes into account the finite rate of diffusion and the finite probability of binding when receptor and ligand are at contact. Similarly, k_{off} is the effective rate at which a ligand dissociates from the receptor and diffuses into the bulk. Our simple coarse-grained model presented below (see Results) gives an intuitive derivation of these effective rate constants.

The uncertainty in the estimate of the concentration can be obtained by combining Eq. 13 with Eqs. 2 and 5, yielding our principal result

$$\frac{\delta c}{c} = \sqrt{\frac{1}{2\pi\sigma D c(1-\bar{n})T} + \frac{2}{k_a c(1-\bar{n})T}}. \quad (18)$$

The first term describes the uncertainty in the estimate of c that stems from the stochastic diffusive arrival of the ligand molecules, whereas the second term describes variability that results from the intrinsic binding dynamics of the receptor. If the receptor-ligand association reaction is fully reaction-limited, i.e., $k_a, k_d \rightarrow 0$ or $D \rightarrow \infty$, then the uncertainty in the concentration estimate is dominated by the latter term. Conversely, if the reaction is diffusion-limited, $k_a, k_d \rightarrow \infty$ or $D \rightarrow 0$, then the first term dominates the uncertainty, which is limited by the diffusive arrival and departure of the ligand molecules to and from the receptor.

It is clear that the second term in Eq. 18 is identical to that in the expression of Bialek and Setayeshgar (9), Eq. 4. Yet, the first term, which determines the fundamental limit, is different: the expression of Bialek and Setayeshgar misses a factor $1/(2(1-\bar{n}))$. The Berg-Purcell expression does contain this factor, and indeed, apart from a geometrical factor, our expression is identical to theirs in the limit that the reaction is fully diffusion-limited.

RESULTS

Simulation results

To test our theory, we have performed particle-based simulations. A key quantity of our theory is $\widehat{C}_n(s)$, Eq. 13, because the precision of our concentration estimate directly follows from this quantity and the gain $d\bar{n}/dc$ (see Eqs. 2 and 5). We therefore compare the power spectrum,

$$P_n(\omega) = 2\text{Re}[\widehat{C}_n(s=i\omega)]$$

with $\widehat{C}_n(s)$ given by Eq. 13, to that obtained from simulations. The simulation scheme should not only describe the diffusive transport at large length and timescales, but also capture the (re)binding dynamics at short scales. Moreover, to obtain an accurate estimate for the zero-frequency limit of the power spectrum, which is computationally challenging, the scheme should also be efficient. We have therefore employed Green's function reaction dynamics (GFRD) (13,15,17). Like Brownian dynamics (overdamped Langevin dynamics), GFRD simulates reaction-diffusion systems at the particle level; in essence, both are numerical procedures for solving the Smoluchowski equation (17). However, while Brownian dynamics uses a fixed time-step to propagate the particles, GFRD is an asynchronous, event-driven kinetic Monte Carlo scheme.

The central idea of GFRD is to decompose the many-body reaction-diffusion problem, which cannot be solved analytically, into sets of one- and two-body problems that can be solved analytically using Green's functions (13,17). In the recent version of GFRD, this decomposition is performed by putting single particles and pairs of particles in mathematical domains (18), for which the reaction-diffusion problem can be solved exactly (15). This yields for each domain an event-type, which is either a reaction or a particle leaving the domain, and an event-time, which is when this event will happen. These events are then executed in chronological order. Importantly, the mathematical domains are nonoverlapping, which means that the stochastic reaction-diffusion processes of the respective domains are independent. This makes GFRD an exact scheme for simulating reaction-diffusion problems at the particle level (15). Because the scheme is event-driven, it is also very fast: at the concentrations considered here, GFRD is up to 4–6 orders-of-magnitude faster than Brownian dynamics (13,15) (for more details, see www.GFRD.org). For this study, we exploited the spherical symmetry of the system and that the ligand molecules only interact with the receptor, but not among themselves.

The computational model is identical to that of our theory, albeit in a finite volume. It consists of a static single receptor in the center of a spherical simulation box with diameter L , surrounded by ligand molecules that diffuse with diffusion constant D . A ligand molecule that is in contact with a free receptor at the contact distance σ can associate with the receptor with an intrinsic association rate k_a and then dissociate from it with an intrinsic dissociation rate k_d ; after dissociation, the ligand molecule is put at contact.

Fig. 1 shows the power spectra as obtained from the simulations (*black circles*) together with the prediction of our theory (*solid red line*; Eq. 13), for $\bar{n} = 0.5$ and $c = 0.4 \mu\text{M}$. In the Supporting Material, we show results for

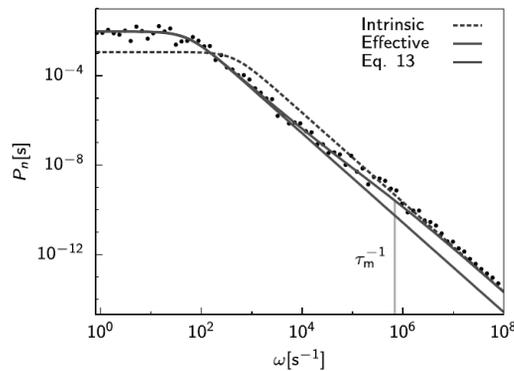


FIGURE 1 The power spectrum of the receptor state $P_n(\omega)$ for $c = 0.4 \mu\text{M}$. The simulation results (black circles) agree well with the theoretical prediction of Eq. 13 (solid red line). At high frequencies $\omega > 1/\tau_m = D/\sigma^2$, the effect of diffusion is negligible and the receptor dynamics is that of a Markovian switching process with intrinsic rates k_{ac} and k_d (dashed red line), while at low frequencies it is that of a Markovian switching process with effective rates $k_{on}c$ and k_{off} , respectively (solid gray line). The zero-frequency limit determines the precision of the concentration estimate. Parameters: $\bar{n} = 0.5$, $D = 1 \mu\text{m}^2 \text{s}^{-1}$, $\sigma = 10 \text{ nm}$, $L = 1 \mu\text{m}$, and $k_a = 552 \mu\text{M}^{-1} \text{s}^{-1}$. To see this figure in color, go online.

higher concentrations. The parameters have been chosen such that they are biologically meaningful, yet bring the system in the diffusion-limited regime—this makes it possible to probe the fundamental limit set by diffusion. Specifically, the diffusion constant $D = 1 \mu\text{m}^2/\text{s}$ is comparable to that of intracellular proteins (19,20).

Fig. 1 shows that the agreement between theory and simulation is very good over essentially the full frequency range. The high-frequency regime corresponds to the intrinsic switching dynamics of the receptor. In this regime, diffusion hardly plays any role and the receptor dynamics is dominated by the binding of ligand molecules that are essentially in contact with the receptor; consequently, the power spectrum is well approximated by that of a binary switching process with uncorrelated exponentially distributed waiting times with the intrinsic correlation time $\tau_c = (k_{ac} + k_d)^{-1}$ (red dashed line). The theory also accurately describes the intermediate frequency regime, which starts at $\omega_m = 1/\tau_m = D/\sigma^2$. In this regime a ligand molecule, after dissociation from the receptor, manages to diffuse away from the receptor over a few molecular distances σ , but then rebinds the receptor before another ligand molecule from the bulk does. The low-frequency regime of the power spectrum corresponds to the regime in which, after receptor dissociation, the ligand molecule diffuses into the bulk and, most likely, another molecule from the bulk binds the receptor. In this regime, the spectrum is well approximated by that of a memoryless switching process with the same effective correlation time as that of our theory, $\tau_n = \tau'_c(s=0) = (k_{on}c + k_{off})^{-1}$ (gray solid line).

The most important point of the power spectrum is at zero-frequency, $P_n(\omega = 0)$, because this determines the correlation time of the receptor and hence the uncertainty

in our estimate of the average receptor occupancy \bar{n} and the concentration c , following Eqs. 2 and 5. Fig. 2 shows $P_n(\omega = 0)$ as a function of the average receptor occupancy \bar{n} at a concentration of $0.4 \mu\text{M}$. (For higher concentrations, see the Supporting Material.) It is seen that the agreement between the theory and simulations is excellent. Fig. 2 also shows the prediction of Bialek and Setayeshgar for $P_n(\omega = 0)$ (9). Although their analysis predicts that $P_n(\omega = 0) = P_{1-n}(\omega = 0)$, our results show that the dependence of $P_n(\omega)$ on n is nonsymmetric, which reflects the fact that if the receptor is free more often, more binding events can be counted, leading to a more accurate estimate of the concentration. Because the Berg-Purcell formula in Eq. 3 directly follows from our expression for $P_n(\omega = 0) = 2\text{Re}[\hat{C}_n(s=0)]$, via Eqs. 2, 5, and 13, we conclude that the Berg-Purcell limit provides an accurate upper bound on the precision with which chemical concentrations can be measured.

Validity of assumptions under biological conditions

The central assumption of our theory is Eq. 9, which states that after dissociation the unbound receptor-ligand pair is surrounded by a uniform distribution of ligand molecules. This assumption breaks down when the following conditions apply:

Condition 1

After receptor dissociation, the rebinding of the ligand molecule to the receptor is preempted by the receptor binding of another, second, ligand molecule.

Condition 2

This second ligand molecule dissociates from the receptor before the first has diffused into the bulk.

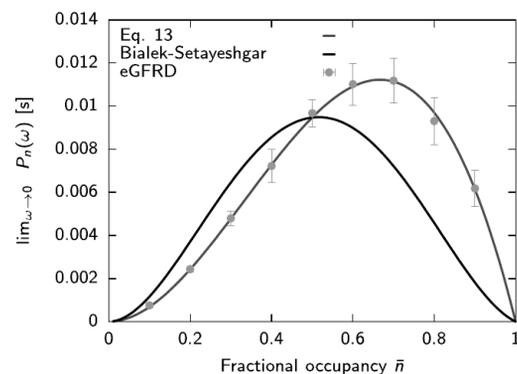


FIGURE 2 The zero-frequency limit of the power spectrum as a function of the average receptor occupancy \bar{n} for $c = 0.4 \mu\text{M}$; \bar{n} is varied by changing k_d . It is seen that the agreement between the theoretical prediction of Eq. 13 and the simulation results is very good (red line). In contrast, the prediction of Bialek and Setayeshgar (9) (black line) differs markedly from our results. Parameters: see Fig. 1. To see this figure in color, go online.

We argue that under biologically relevant conditions, neither condition arises and, therefore, the key assumption of our analysis holds.

A rebinding trajectory of a ligand molecule that has just dissociated from the receptor is very short on the timescale at which molecules arrive from the bulk at the concentrations considered here (see Fig. S1 in the Supporting Material). Consequently, the likelihood that another molecule interferes with such a rebinding event is negligible (see Fig. S3); a dissociated ligand molecule rebinds the receptor before it diffuses into the bulk as often as when it was the only ligand molecule present in the system. Condition 1 is thus not met and the central assumption, Eq. 9, holds.

Occasionally rebinding interferences will occur, and Condition 1 is met. However, we argue that Eq. 9 is still likely to hold, because Condition 2 is not met: a ligand molecule is typically bound long enough for the previously bound molecules to diffuse into the bulk. Consider a detector that binds a ligand with a cross-section $\sigma = 10$ nm, a diffusion constant $D = 1 \mu\text{m}^2 \text{s}^{-1}$, and an intrinsic rate k_a that equals the diffusion-limited rate $k_D = 75 \mu\text{M}^{-1} \text{s}^{-1}$, yielding an effective association rate $k_{\text{on}} = 38 \mu\text{M}^{-1} \text{s}^{-1}$, consistent with experimentally measured association rates (21). If the ligand is present at a biologically relevant concentration of $c = 1 \mu\text{M}$, then, for $\bar{n} = 0.5$, the time a ligand molecule is bound to the receptor is $t_d = k_d^{-1} \approx 0.01$ s. During this time, the previously bound ligand molecule, on average, has traveled at least a distance

$$\sqrt{6Dt_d} \approx 0.3 \mu\text{m}.$$

This corresponds to ~ 4 times the average distance between the ligand molecules at this concentration, meaning that, effectively, the ligand has diffused into the bulk. For lower concentrations, the dissociation time will be longer at constant \bar{n} , and the previously bound ligand molecule will have penetrated the bulk even deeper by the time that the newly bound ligand molecule dissociates. We thus expect that for concentrations up to micromolar, Condition

2 is not met—meaning that even when rebinding interferences do occasionally arise, and Eq. 9 still holds.

The other approximation of our theory, Eq. 12, ensures that the short- and long-time behavior of $\mathcal{P}_{\text{rad}}(t|\text{eq})$ is described correctly. Importantly, however, under the biologically relevant concentrations considered here, the receptor-binding rate of ligand molecules starting from a uniform distribution is so low that to a good approximation $\mathcal{P}_{\text{rad}}(t|\text{eq})$ is given by its long-time behavior, $\mathcal{P}_{\text{rad}}(t|\text{eq}) \approx e^{-k_{\text{on}}ct}$ (see Fig. S1). The picture that thus emerges is that after a receptor-ligand dissociation event, the molecules that are not in contact with the receptor truly form a bulk reservoir:

1. They have a uniform distribution (Eq. 9), and
2. They bind the receptor in a memoryless fashion with a constant rate $k_{\text{on}}c$.

A simple coarse-grained model

Ultimately, the success of Eq. 9 is due to the fact that the time a ligand molecule spends near the receptor is very short compared to the timescale on which ligand molecules arrive at the receptor from the bulk, $(k_Dc)^{-1}$ (see Fig. S1). On this timescale a ligand molecule at contact with the receptor effectively either instantly (re)binds the receptor with splitting probability p_{reb} or escapes into the bulk with probability $p_{\text{esc}} = 1 - p_{\text{reb}}$. This observation naturally suggests the following simple two-state model (17), in which the system switches between a receptor-bound and a receptor-unbound state with effective association and dissociation rates (see Fig. 3).

To derive the effective dissociation rate, we note that for a ligand molecule that has just dissociated from the receptor, the probability that it will rebound the receptor rather than diffuse away into the bulk is $p_{\text{reb}} = 1 - S_{\text{rad}}(\infty|\sigma) = k_a/(k_a + k_D)$. The mean number of rounds of rebinding and dissociation before the molecule escapes into the bulk is then

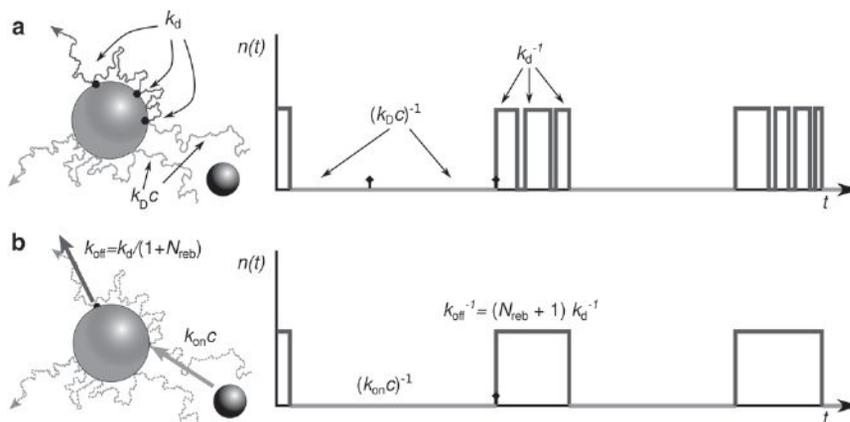


FIGURE 3 Cartoon of the coarse-grained model. (a) A typical time trace of the receptor state $n(t)$ of the original system. (b) Time trace of the coarse-grained model. (Top-left cartoon; red) a successful and an unsuccessful binding trajectory; (blue) a trajectory in which a ligand molecule undergoes a number of rounds of receptor dissociation and rebinding before it escapes into the bulk. The key observation is that the time a molecule spends near the receptor is very short on the timescale at which molecules arrive from the bulk. This makes it possible to integrate out the receptor rebindings and the unsuccessful arrivals of molecules from the bulk, giving the two-state model of Eq. 19. Fig. S1 in the Supporting Material quantifies the timescale separation. To see this figure in color, go online.

$$N_{\text{reb}} = (1 - p_{\text{reb}}) \sum_{i=0}^{\infty} i p_{\text{reb}}^i = p_{\text{reb}} / (1 - p_{\text{reb}}) = k_a / k_D.$$

The total time t_{on} the ligand is bound to the receptor before it diffuses away is on average $t_{\text{on}} = t_d(1 + N_{\text{reb}})$ and the effective dissociation rate is $k_{\text{off}} = 1/t_{\text{on}} = k_d k_D / (k_a + k_D)$, which is precisely the effective dissociation rate of our theory, Eq. 17.

A molecule that arrives at the receptor from the bulk at a rate k_D may either bind the receptor or escape back into the bulk. The escape probability is

$$p_{\text{esc}} = 1 - p_{\text{reb}} = S_{\text{rad}}(\infty | \sigma) = k_D / (k_a + k_D),$$

and the average number of times a molecule from the bulk encounters the receptor before it actually binds is

$$\begin{aligned} N_{\text{esc}} + 1 &= (1 - p_{\text{esc}}) \sum_{i=0}^{\infty} i p_{\text{esc}}^i + 1 = k_D / k_a + 1 \\ &= 1 / N_{\text{reb}} + 1. \end{aligned}$$

The effective rate at which a molecule binds from the bulk is then

$$k_{\text{on}} = k_D / (1 + N_{\text{esc}}) = k_a k_D / (k_a + k_D),$$

which is the diffusion-limited rate k_D times the probability

$$1 - S(\infty | \sigma) = k_a / (k_a + k_D)$$

that a molecule at contact binds the receptor instead of diffusing back into the bulk. This is the rate as predicted by our theory, Eq. 16. We note that the renormalization of the association and dissociation rates preserves the detailed-balance condition:

$$\bar{n} / (1 - \bar{n}) = k_a c / k_d = k_{\text{on}} c / k_{\text{off}}.$$

The dynamics of this two-state model is given by

$$\frac{dn(t)}{dt} = k_{\text{on}} c (1 - n(t)) - k_{\text{off}} n(t), \quad (19)$$

which, as Fig. 1 shows, correctly describes the relevant low-frequency dynamics of the receptor state. From Eq. 19, the power spectrum and hence, the uncertainty in our concentration estimate can be obtained straightforwardly:

$$\frac{\delta c}{c} = \sqrt{\frac{2}{k_{\text{on}} c (1 - \bar{n}) T}}. \quad (20)$$

Three points are worthy of note:

1. Noting that k_{on} is given by Eq. 16, the principal result of this article, Eq. 18, is reproduced.
2. For diffusion-limited reactions $k_{\text{on}} = k_D$ and the expression of Berg and Purcell, namely, Eq. 3, is recovered.

3. Arguably, however, the most important result is that the Berg-Purcell expression can also be applied to reactions that are not diffusion-limited, by replacing k_D with k_{on} .

Point 3 is a nontrivial result, because Eq. 20 takes into account that not every arrival of ligand at a free receptor leads to binding, and also that upon dissociation a ligand molecule may rebind the receptor many times before it escapes into the bulk. This can be seen by exploiting the detailed-balance condition and rewriting Eq. 20 as

$$\frac{\delta c}{c} = \sqrt{\frac{2}{T} \left(\frac{1}{k_{\text{off}}} + \frac{1}{k_{\text{on}} c} \right)} = \sqrt{\frac{2\tau_w}{T}}. \quad (21)$$

This expression has a clear interpretation (see Fig. 3). A receptor-bound molecule that dissociates from the receptor may rebind the receptor, but the probability for this to happen does not depend on the concentration (15). As a result, a rebinding event does not provide information on the concentration and should therefore not be counted as a concentration measurement; it merely increases the receptor correlation time τ_n by increasing its on-time from k_d^{-1} to $k_{\text{off}}^{-1} = k_d^{-1} / (1 + N_{\text{reb}})$. After $(1 + N_{\text{reb}})$ rounds of dissociation and rebinding, the molecule escapes into the bulk, and then another molecule will arrive at the receptor with a rate $k_D c$. This molecule may return to the bulk or bind the receptor—in either case, possibly after a number of unsuccessful but rapid collisions—such that the net rate at which a molecule from the bulk binds the receptor is $k_{\text{on}} c$. Importantly, this binding event occurs in a memoryless fashion and with a rate that depends on the concentration. Consequently, this event does provide an independent measurement of the concentration. The time $\tau_w = 1/k_{\text{off}} + 1/(k_{\text{on}} c)$ is thus the average time between independent concentration measurements, and T/τ_w is the total number of such measurements in the integration time T .

DISCUSSION

Using results from the theory of diffusion-influenced reactions by Agmon and Szabo (11), we have derived the fundamental limit for the precision of chemical concentration measurements via the reversible binding of ligand to a single receptor, a common motif in cell signaling. We have compared our expression to that of Berg and Purcell (3) and Bialek and Setayeshgar (9) and tested it against particle-based simulations. The premise of our study and that of Berg and Purcell (3) and Bialek and Setayeshgar (9) is that the concentration is estimated from the average receptor occupancy n_T over an integration time T set by the downstream network, and inverting the input-output relation $\bar{n}(c)$ (3,9,22–25). Recently, Endres and Wingreen (26) and Mora and Wingreen (27) showed that maximum likelihood estimation can improve this estimate, but it is not clear whether typical networks do this.

Our theoretical analysis, as well as that of Berg and Purcell (3) and Bialek and Setayeshgar (9), assumes that the total number of ligand molecules in the system is fixed; the total concentration c is thus constant. Indeed, the only sources of fluctuations in the receptor state are the diffusion of ligand to and from the receptor, and the stochastic binding and unbinding of ligand to and from the receptor. Moreover, all three analyses assume that the ligand molecules are noninteracting, which is a very reasonable assumption at the low concentrations considered here. The reason that this is nonetheless a nontrivial many-body reaction-diffusion problem is due to the interactions of the ligand molecules with the receptor.

As described in the surrounding text for Eq. 9, due to these receptor-ligand interactions, the spatial distribution of ligand molecules, conditioned on the state of the receptor (or the time since the last receptor binding or unbinding event), can deviate from the equilibrium uniform distribution. Only by averaging over the entire ensemble of ligand distributions, taking into account those corresponding to both the bound and unbound states of the receptor, is the uniform equilibrium distribution obtained. The central assumption of our theory is that after each receptor dissociation event, the receptor and ligand molecule at contact are surrounded by a uniform, equilibrium distribution of ligand molecules (see Eq. 9). This assumption makes it possible to reduce the many-body problem to a pair problem, in which the ligand molecules bind the receptor independently of one another.

Also in the analysis of Berg and Purcell (3), the many-body reaction-diffusion problem is reduced to a pair problem. However, the key difference with our analysis is that their analysis strictly holds only for diffusion-limited reactions, but they argue that it also holds for reactions that are not diffusion-limited. In essence, they start from the assumption that the receptor switches between the ligand-bound and unbound states according to a random telegraph process with exponentially distributed waiting times, thus ignoring the rebinding trajectories with algebraic waiting times. The receptor correlation time τ_n , needed to obtain the fundamental limit (see Eq. 5), is in this Markovian model given by the receptor-ligand association and dissociation rates. Berg and Purcell argue that the association rate k_f in the presence of unsuccessful ligand-receptor collisions can be obtained from the diffusion-limited binding rate $k_D = 4\pi\sigma D$ via a rescaling of the cross-section σ , while the dissociation rate k_b can then be obtained from the detailed balance relation $k_f c(1 - \bar{n}) = k_b \bar{n}$. However, the validity of this Markovian model with the ad hoc rescaling of k_D to get k_f remained unclear.

Bialek and Setayeshgar (9) do not assume that the receptor-ligand association rate is diffusion-limited. Indeed, their physical model is identical to ours: ligand molecules are noninteracting; their overall concentration is constant; ligand and receptor associate with a rate k_a when at contact,

and dissociate with a rate k_d when bound. However, they analyze their model by writing down the reaction-diffusion equation, and then solve this equation invoking the fluctuation-dissipation theorem.

Our principal result (Eq. 18) consists of two terms, like the expression of Bialek and Setayeshgar (Eq. 4): one describes the effect of the diffusive transport of ligand to and from the receptor and the other describes the effect of the intrinsic binding kinetics when the ligand is in contact. However, whereas the intrinsic binding term agrees with that of Bialek and Setayeshgar, the diffusive term does not: it contains an additional factor $1/(2(1 - \bar{n}))$. We believe that this is because, by invoking the fluctuation-dissipation theorem, Bialek and Setayeshgar (9) linearize the reaction-diffusion problem, thereby ignoring correlations between the state of the receptor and the ligand concentration. The expression of Berg and Purcell (Eq. 3) does not feature a term that arises from the intrinsic binding. However, their term agrees with our diffusive term, which is considered to be the fundamental limit. Because both terms in our expression contain the factor $1/(2(1 - \bar{n}))$, our principal result Eq. 18 can be rewritten as Eq. 20, which shows that the expression of Berg and Purcell (Eq. 3) can be generalized to reactions that are not diffusion-limited by simply replacing k_D with k_{on} . Our simple coarse-grained model elucidates why this is possible: $k_{on}c(1 - \bar{n}) = \tau_w^{-1}$ is the rate at which molecules independently bind the receptor from the bulk; via detailed balance, this expression not only captures the unsuccessful arrivals of molecules from the bulk, but also the receptor rebindings, which do not provide information about the concentration.

The purpose of sensing is to enable the detection of changes in the ligand concentration. If the concentration varies over only a small range around the dissociation constant K_D , then \bar{n} varies in a small interval around $\bar{n} = 0.5$, for which the expressions of Berg and Purcell (3) and Bialek and Setayeshgar (9) become equal. However, for most signaling systems it is not known how the ligand concentration varies under physiologically relevant conditions. Yet, information theory tells us that information transmission is maximized when the input distribution spans the full dynamic range of the signaling system, rather than a narrow range around K_D (28,29). Or, vice versa, in the case where the input distribution is fixed by the environment, the response curve that maximizes information transmission is one whose dynamic range matches the input distribution. This suggests that \bar{n} can be well above and below $\bar{n} = 0.5$ —meaning that the difference between the respective expressions can become significant.

A comparison of the power spectrum of the receptor occupancy as predicted by our theory against results from particle-based simulations shows that our theory is very accurate under biologically relevant conditions. As Fig. S6 shows, the theory accurately predicts the zero-frequency limit of the power spectrum for concentrations up to at least

36 μM . Given the complicated algebraic diffusion dynamics of rebinding trajectories, it is perhaps surprising that the error in the concentration estimate is so accurately predicted by a simple expression like Eq. 18 or 20. The success of this expression lies in the fact that biologically relevant concentrations are low, which means that the time a ligand molecule spends near the receptor is very short both on the timescale at which molecules arrive from the bulk and on the timescale a ligand molecule is bound to the receptor. This is the reason why both the rapid rebindings and the unsuccessful bulk arrivals can be integrated out, and the complicated many-body reaction-diffusion problem can be reduced to a pair problem in which ligand molecules interact with the receptor with renormalized association and dissociation rates (see Eq. 19).

Whether this simple approach can also be used when the ligand can bind to multiple receptors that are in close physical proximity remains an open question (3,9,22). Ligand rebinding between neighboring receptors may lead to nontrivial spatio-temporal correlations between the receptor states. Berezhkovskii and Szabo (12) have recently derived an expression for the accuracy of sensing via multiple receptors on a sphere, ignoring these spatio-temporal correlations. In the limit that the number of receptors goes to infinity, their expression for the accuracy of sensing in the diffusion-limited regime reduces to

$$\delta c/c = 1/\sqrt{2\pi DRcT}.$$

Paradoxically, by replacing the cross section of the receptor σ with twice the radius of the cell R , this is the result of Bialek and Setayeshgar (9) for a single receptor, Eq. 4: the expression for $\delta c/c$ does not contain the factor $1 - \bar{n}$. The absence of this factor can be understood intuitively by noting that in the limit that the number of receptors goes to infinity, there will always be receptors available for binding the ligand. The fact that in this limit the result of Bialek and Setayeshgar (9) for a single receptor is recovered supports the idea that their analysis is essentially a mean-field analysis that ignores receptor-ligand correlations. For a finite number of receptors (including the scenario that there is only one receptor), the occupancy of the receptor will, however, affect the precision of sensing, and these correlations are important. Clearly, it will be of interest to test the accuracy of sensing via multiple receptors, and quantify the importance of spatio-temporal correlations, via simulations. We leave this for future work.

Finally, our observations have implications for the modeling of biochemical networks. They underscore our earlier observation (17) that when the cell does not exhibit concentration gradients on cellular length scales, the effect of diffusion can often be captured in a well-stirred model, which can then be simulated using the Gillespie algorithm (30) instead of a much more computationally demanding particle-based algorithm (15). In such a well-stirred model,

the rapid rebindings are integrated out and association and dissociation occur in a memoryless fashion, with exponentially distributed waiting times with mean $(k_{\text{on}}c)^{-1}$ and k_{off}^{-1} , respectively. This is a simplification—at short times, the association-time distribution is algebraic due to the rebindings—but it is an accurate one. It not only preserves the equilibria of the association-dissociation reactions, but also the important noise characteristics of the network. Indeed, the high-frequency noise from the rapid rebindings is typically filtered by the network downstream and only the low-frequency noise, obeying exponential statistics, is significantly propagated downstream (17). On the other hand, when rebindings can qualitatively change the macroscopic behavior of the system, as in systems employing multisite protein modification, rebindings cannot simply be integrated out to yield the well-known macroscopic effective rate constants k_{on} and k_{off} (15,31). Yet, Gopich and Szabo (32) have recently shown that, because rebindings remain fast, it is nonetheless possible to arrive at an alternative simplified description for these systems that captures the effect of rebindings.

SUPPORTING MATERIAL

Six figures, 73 equations, References (33, 34) and Supplemental information are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(13\)05847-5](http://www.biophysj.org/biophysj/supplemental/S0006-3495(13)05847-5).

We thank Attila Szabo, Howard Berg, and Bill Bialek for stimulating discussions and Bela Mulder for a critical reading of the manuscript. Calculations were performed on the RIKEN Integrated Cluster of Clusters (RICC) facility in Saitama, Japan.

This work is part of the research program of the Foundation for Fundamental Research on Matter, which is part of the Netherlands Organization for Scientific Research (NOW).

REFERENCES

1. Rieke, F., and D. Baylor. 1998. Single-photon detection by rod cells of the retina. *Rev. Mod. Phys.* 70:1027–1036.
2. Boeckh, J., K.-E. Kaissling, and D. Schneider. 1965. Insect olfactory receptors. *Cold Spring Harb. Symp. Quant. Biol.* 30:263–280.
3. Berg, H. C., and E. M. Purcell. 1977. Physics of chemoreception. *Biophys. J.* 20:193–219.
4. Sourjik, V., and H. C. Berg. 2002. Receptor sensitivity in bacterial chemotaxis. *Proc. Natl. Acad. Sci. USA.* 99:123–127.
5. Ueda, M., and T. Shibata. 2007. Stochastic signal processing and transduction in chemotactic response of eukaryotic cells. *Biophys. J.* 93:11–20.
6. Gregor, T., D. W. Tank, ..., W. Bialek. 2007. Probing the limits to positional information. *Cell.* 130:153–164.
7. Erdmann, T., M. Howard, and P. R. ten Wolde. 2009. Role of spatial averaging in the precision of gene expression patterns. *Phys. Rev. Lett.* 103:258101.
8. Dubuis, J. O., G. Tkačik, ..., W. Bialek. 2013. Positional information, in bits. *Proc. Natl. Acad. Sci. USA.* 110:16301–16308.
9. Bialek, W., and S. Setayeshgar. 2005. Physical limits to biochemical signaling. *Proc. Natl. Acad. Sci. USA.* 102:10040–10045.

10. De Ronde, W. H. 2012. PhD thesis, Multiplexing Biochemical Signals. FOM Institute AMOLF, Amsterdam, The Netherlands.
11. Agmon, N., and A. Szabo. 1990. Theory of reversible diffusion-influenced reactions. *J. Chem. Phys.* 92:5270–5284.
12. Berezhkovskii, A. M., and A. Szabo. 2013. Effect of ligand diffusion on occupancy fluctuations of cell-surface receptors. *J. Chem. Phys.* 139:121910.
13. van Zon, J. S., and P. R. ten Wolde. 2005. Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics. *Phys. Rev. Lett.* 94:128103.
14. van Zon, J. S., and P. R. ten Wolde. 2005. Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space. *J. Chem. Phys.* 123:234910.
15. Takahashi, K., S. Tănase-Nicola, and P. R. ten Wolde. 2010. Spatio-temporal correlations can drastically change the response of a MAPK pathway. *Proc. Natl. Acad. Sci. USA.* 107:2473–2478.
16. Rice, S. A. 1985. Diffusion-Limited Reactions. Elsevier, Amsterdam, The Netherlands.
17. van Zon, J. S., M. J. Morelli, ..., P. R. ten Wolde. 2006. Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys. J.* 91:4350–4367.
18. Opplestrup, T., V. V. Bulatov, ..., B. Sadigh. 2006. First-passage Monte Carlo algorithm: diffusion without all the hops. *Phys. Rev. Lett.* 97:230602.
19. Elowitz, M. B., M. G. Surette, ..., S. Leibler. 1999. Protein mobility in the cytoplasm of *Escherichia coli*. *J. Bacteriol.* 181:197–203.
20. Elf, J., G.-W. Li, and X. S. Xie. 2007. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science.* 316:1191–1194.
21. Northrup, S. H., and H. P. Erickson. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. *Proc. Natl. Acad. Sci. USA.* 89:3338–3342.
22. Wang, K., W.-J. Rappel, ..., H. Levine. 2007. Quantifying noise levels of intercellular signals. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 75:061905.
23. van Haastert, P. J. M. 2010. A stochastic model for chemotaxis based on the ordered extension of pseudopods. *Biophys. J.* 99:3345–3354.
24. Hu, B., W. Chen, ..., H. Levine. 2010. Physical limits on cellular sensing of spatial gradients. *Phys. Rev. Lett.* 105:048104.
25. Govern, C. C., and P. R. ten Wolde. 2012. Fundamental limits on sensing chemical concentrations with linear biochemical networks. *Phys. Rev. Lett.* 109:218103.
26. Endres, R. G., and N. S. Wingreen. 2009. Maximum likelihood and the single receptor. *Phys. Rev. Lett.* 103:158101.
27. Mora, T., and N. S. Wingreen. 2010. Limits of sensing temporal concentration changes by single cells. *Phys. Rev. Lett.* 104:248101.
28. Tkačik, G., C. G. Callan, Jr., and W. Bialek. 2008. Information flow and optimization in transcriptional regulation. *Proc. Natl. Acad. Sci. USA.* 105:12265–12270.
29. Tostevin, F., and P. R. ten Wolde. 2009. Mutual information between input and output trajectories of biochemical networks. *Phys. Rev. Lett.* 102:218101.
30. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.
31. Mugler, A., and P. R. ten Wolde. 2013. The macroscopic effects of microscopic heterogeneity in cell signaling. *Adv. Chem. Phys.* 153:373–396.
32. Gopich, I. V., and A. Szabo. 2013. Diffusion modifies the connectivity of kinetic schemes for multisite binding and catalysis. *Proc. Natl. Acad. Sci. USA.* 110:19784–19789.
33. Popov, A. V., and N. Agmon. 2001. Three-dimensional simulations of reversible bimolecular reactions: the simple target problem. *J. Chem. Phys.* 115:8921–8931.
34. Gopich, I. V., and A. Szabo. 2002. Kinetics of reversible diffusion influenced reactions: the self-consistent relaxation time approximation. *J. Chem. Phys.* 117:507–517.

Article

Role of Denatured-State Properties in Chaperonin Action Probed by Single-Molecule Spectroscopy

Hagen Hofmann,^{1,*} Frank Hillger,¹ Cyrille Delley,¹ Armin Hoffmann,¹ Shawn H. Pfeil,² Daniel Nettels,¹ Everett A. Lipman,² and Benjamin Schuler^{1,*}

¹Department of Biochemistry, University of Zurich, Zurich, Switzerland; and ²Department of Physics, University of California, Santa Barbara, Santa Barbara, CA

ABSTRACT The bacterial chaperonin GroEL/GroES assists folding of a broad spectrum of denatured and misfolded proteins. Here, we explore the limits of this remarkable promiscuity by mapping two denatured proteins with very different conformational properties, rhodanese and cyclophilin A, during binding and encapsulation by GroEL/GroES with single-molecule spectroscopy, microfluidic mixing, and ensemble kinetics. We find that both proteins bind to GroEL with high affinity in a reaction involving substantial conformational adaptation. However, whereas the compact denatured state of rhodanese is encapsulated efficiently upon addition of GroES and ATP, the more expanded and unstructured denatured cyclophilin A is not encapsulated but is expelled into solution. The origin of this surprising disparity is the weaker interactions of cyclophilin A with a transiently formed GroEL-GroES complex, which may serve as a crucial checkpoint for substrate discrimination.

INTRODUCTION

The cellular machinery of molecular chaperones (1–4), quality control systems (3), and components that translocate and degrade proteins reflects the importance of protein homeostasis in the cell. To control protein folding, misfolding, and aggregation, these cellular factors can exert forces that will affect the conformation and dynamics of proteins (4,5). However, the complexity and promiscuity (6–8) of this machinery often complicate detailed mechanistic investigations of the underlying physical principles. For example, it has been suggested that the confinement of proteins inside a cavity, as formed by the heptameric rings of GroEL and GroES (1–4,9), can stabilize proteins (10,11) and accelerate protein folding reactions (12–16). However, results from experiments and simulation also show that this accelerating effect can be countered by interactions of the substrate with the GroEL-GroES cavity wall (12,17–20). To delineate these contributions we need to better understand the effect of GroEL on the conformation and dynamics of denatured proteins. Here, we compare two proteins with very different denatured-state properties, a destabilized variant of human cyclophilin A (CypA) (21–23) and bovine rhodanese

(Rho) (24) (Fig. 1 *a*) (25,26). Both proteins are known to bind to GroEL in vitro (23,24,27,28) and are identified as GroEL-binding-competent based on a sequence comparison with 284 GroEL substrates identified in *Escherichia coli* (Fig. 1, *b* and *c*) (6,7,26). Although human CypA and bovine Rho have very similar mean net charge and hydrophobicity (Fig. 1 *b*), the two proteins differ substantially in their length; bovine Rho (296 amino acids (aa)) is almost twice as large as human CypA (167 aa). Of most importance, however, previous experiments indicate that their GroEL-bound states are very different: whereas CypA bound to GroEL was shown to exhibit substantial dynamics with a lack of stable secondary structure (23,29), recent NMR experiments indicate that 85% of the amino acids in Rho are immobile in complex with GroEL (30). Thus, the two proteins represent different parts of the broad substrate spectrum of the promiscuous GroEL/ES chaperone (Fig. 1 *c*). To investigate the conformational properties and dynamics of both proteins along the encapsulation pathway, we employ a combination of single-molecule Förster resonance energy transfer (smFRET), microfluidic mixing, fluorescence correlation spectroscopy, and ensemble kinetic experiments.

MATERIALS AND METHODS

Single-molecule fluorescence spectroscopy

Measurements were performed at 22°C in 50 mM TrisHCl, 10 mM MgCl₂, 5 mM KCl, 100 mM 2-mercaptoethanol, 0.001% Tween 20, pH 7.5, using either a custom-built confocal microscope, as described previously (17), or a Micro Time 200 confocal microscope equipped with a HydraHarp 400 counting module (Picoquant, Berlin, Germany). The donor dye was excited with a diode laser at 485 nm (dual mode: continuous wave and

Submitted August 13, 2014, and accepted for publication November 3, 2014.

*Correspondence: hagen.hofmann@weizmann.ac.il or schuler@bioc.uzh.ch

Hagen Hofmann's present address is Department of Structural Biology, Weizmann Institute of Science, Rehovot 7600, Israel.

Frank Hillger's present address is Novartis Institutes for BioMedical Research, Novartis International AG, 4002 Basel, Switzerland.

Armin Hoffmann's present address is Celonic AG, 8051 Basel, Switzerland.

Shawn H. Pfeil's present address is Department of Physics, West Chester University of Pennsylvania, West Chester, PA 19383

Editor: Ashok Deniz.

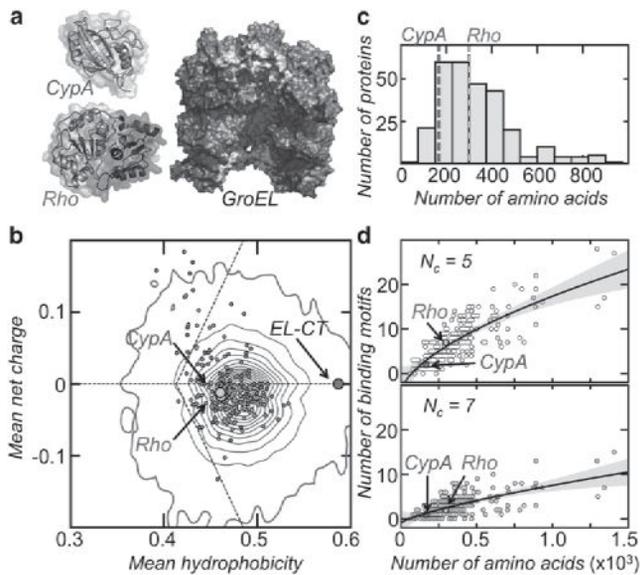


FIGURE 1 Structure and sequence characteristics of Rho and CypA. (a) Structure of cyclophilin (CypA; Protein Data Bank (PDB) 1OCA) and rhodanese (Rho; PDB 1RHS), with a surface representation of GroEL in its ADP state (PDB 1XCK). For clarity, four subunits of each ring have been removed. The substrate-binding region in the apical domains is shown in yellow. The size of the structures is not to scale. (b) Mean net charge versus mean hydrophobicity for 50,000 amino acid sequences drawn from the natural abundance of amino acids (35) (contours), for 284 identified GroEL substrates (gray circles) (6,7,18), CypA (green circle), Rho (blue circle), and the last 17 amino acids of the C-terminal tail of GroEL (red circle). (c) Sequence length distribution of 284 identified GroEL substrates (6). Dashed lines indicate the lengths of CypA (green) and Rho (blue). (d) Scaling of the number of GroEL binding motifs (18) PXXXXH ($N_c = 5$) and PXXXXHX (P, polar amino acid; H, hydrophobic amino acid; X, any amino acid) with the protein length for 284 identified GroEL substrates from *E. coli* (6,7,18). The solid line is a fit according to $y = a^x + y_0$ and indicates the average of the sequence length distribution of GroEL binding motifs. CypA (green) and Rho (blue) are close to this average. To see this figure in color, go online.

pulsed; LDH-D-C-485, PicoQuant) at an average power of 200 μ W. smFRET efficiency histograms were acquired in samples with a protein concentration of ~ 20 –50 pM, with the laser in continuous-wave mode; photon counts were recorded at a resolution of 16 ps by the counting electronics (time resolution was thus limited by the timing jitter of the detectors). For dual-color excitation, the acceptor was excited in addition to the donor with picosecond pulses at a wavelength range selected by a z582/15 (Chroma, Bellows Falls, VT) band-pass filter and a pulse frequency of 20 MHz (Optical Supercontinuum Systems SCF450-4-20MHz, Fianium, Southampton, United Kingdom). Successive photons detected in either channel and separated by $<100 \mu$ s were combined in one burst. A burst was retained as a significant event if the total number of counts exceeded 50 for free CypA or CypA-SR1 complexes or 25 for experiments using the microfluidic mixing device. For Rho-SR1 complexes or denatured Rho, a threshold of 20 for Rho-SR1 was used. Identified bursts were corrected for background, differences in quantum yields of donor and acceptor, the different collection efficiencies in the detection channels, cross talk, and direct acceptor excitation, as described previously (31). In addition, bursts during which acceptor photobleaching was likely to have occurred were discarded (27).

The quantum yields of the attached fluorophores are substantially lowered in all denatured Rho variants at 0.4 M guanidinium chloride (GdmCl), indicative of photoinduced electron transfer (PET) between exposed aro-

matic amino acids and the attached fluorophores (17). However, PET affects $\langle E \rangle$ only via quenching of the acceptor fluorophore, and the observed transfer efficiencies are thus lower bounds for the true values (32). Static quenching (PET) of the donor fluorophore does not affect the transfer efficiencies, because neither the donor nor the acceptor emits photons in this case. In general, uncertainties in observed transfer efficiencies of ~ 0.03 were estimated from the variations in the experimentally determined correction factor, which takes into account differences in detection efficiencies and quantum yields of the two fluorophores (31) over a period of 3 years.

Microfluidic mixing experiments

For rapid mixing experiments, microfluidic mixers fabricated by replica molding in polydimethylsiloxane were used as described previously (17,33). For experiments using the microfluidic device, the Tween 20 concentration was increased to 0.01% to prevent surface adhesion of the proteins. The transfer efficiency histograms for denatured CypA and Rho at 0.4 M GdmCl were obtained by mixing denatured protein in 4 M GdmCl with buffer at a flow rate of 0.8 mm/s by placing the confocal volume at position 100 μ m (125 ms) downstream of the mixing region. The applied pressures were 10.4 kPa (1.5 psi) in the two buffer channels and 6.8 kPa (1 psi) in the sample channel. The kinetics of CypA dissociation from GroEL-SR1 on mixing with ATP and GroES were obtained at a flow rate of 1.2 mm/s with a pressure of 2 psi in all channels.

For detection of the GroES-ATP-mediated release of the SR1-bound CypA variants, the binary CypA-SR1 complex was mixed at a ratio of 1:5.7 with 2.4 mM ATP and varying concentrations of GroES, resulting in final concentrations of 0.5 μ M, 1 μ M, 1.6 μ M, and 3 μ M GroES, and 2 mM ATP. The experiments were performed at pressures of 13.8 kPa (2.0 psi) applied to all channels, resulting in a mean velocity of 1.2 mm/s that was used to convert distances to times, as described by Pfeil et al. (33). The calculated velocities were checked by analyzing the donor-acceptor fluorescence intensity cross-correlation functions and the obtained rate constants k_1 and k_2 were corrected for variations in the flow between the different microfluidic chips used for the experiments. To determine the transfer efficiency histogram at $t = 0$, the binary CypA-SR1 complex was measured in the observation channel of the mixing device without ATP and GroES in the buffer channels.

Two-focus fluorescence correlation spectroscopy

Two-focus fluorescence correlation spectroscopy (2f-FCS) measurements (34) were performed at 22°C on a Micro Time 200 confocal microscope (PicoQuant) equipped with a differential interference contrast prism. The donor dye was excited alternatively with two orthogonally polarized diode lasers at 483 nm (LDH-D-C-485, PicoQuant) at a repetition rate of 20 MHz and a laser power of 30 μ W each. The distance between the two foci was determined as described previously (35).

Binding isotherms of denatured CypA or carbamidomethylated rhodanese (CAM-Rho) were obtained at a substrate concentration of 0.725 nM for CAM-Rho58 labeled with AlexaFluor 488 and AlexaFluor 594 or 0.125 nM for CypA-V2C-A488, respectively. Before measurement, the samples were incubated at 22°C for 2 h (CypA-SR1) and 12 h (CypA-SR1+ATP γ S, Rho-SR1, and RhoSR1-ATP γ S). The change in the Stokes radii on addition of GroEL-SR1 were fitted according to

$$R_H = \frac{\Delta R_H}{2P_0} \left(P_0 + [\text{SR1}]_0 + K_D - \sqrt{(P_0 + K_D + [\text{SR1}]_0)^2 - 4P_0[\text{SR1}]_0} \right) + R_{H0} \quad (1)$$

Here, ΔR_H is the change in R_H between free and completely bound substrate, P_0 is the total concentration of CypA and Rho, $[SR1]_0$ is the total concentration of GroEL-SR1 and R_{H0} is the Stokes radius in the absence of chaperone.

GroEL-SR1 binding kinetics of CypA and Rho using stopped-flow mixing

The nonlinear dependence of the pseudo-first-order binding rate constant (k) on GroEL-SR1 concentration was described by a kinetic model including the fast formation of a transient encounter complex using

$$k = \frac{k_{\max}[SR1]}{K + [SR1]}, \quad (2)$$

where k_{\max} is the rate constant for the rate-limiting conformational adaptation to the GroEL-SR1 surface and K is the dissociation constant for the encounter complex. The temperature dependence of K is given by

$$K = \exp\left(\frac{\Delta H_{\text{eq}} - T\Delta S_{\text{eq}}}{RT}\right), \quad (3)$$

where ΔH_{eq} is the enthalpy change and ΔS_{eq} is the entropy change on formation of the encounter complex relative to free substrate and free GroEL-SR1, R is the ideal gas constant, and T is the temperature. The temperature dependence of k_{\max} is given by

$$k_{\max} = k_0 \exp\left[-\frac{\Delta H_0^\ddagger + \Delta C^\ddagger(T - T_0) - T(\Delta S_0^\ddagger + \Delta C^\ddagger \ln(T/T_0))}{RT}\right], \quad (4)$$

with ΔH_0^\ddagger and ΔS_0^\ddagger being the change in activation enthalpy and entropy at $T_0 = 298$ K and ΔC^\ddagger being the change in heat capacity between the encounter complex and the top of the adaptation barrier. The preexponential factor k_0 was given by

$$k_0(T) = \tau_0^{-1} \frac{\eta(T_0)}{\eta(T)} \quad (5-1)$$

with

$$\eta(T) = \eta_0 \exp\left[\frac{B}{R(T - \theta)}\right], \quad (5-2)$$

with $\tau_0^{-1} = 10^6$ s⁻¹ as estimated from the ns-FCS measurements ($\tau_0 = 1$ μ s) and $T_0 = 298$ K (see the Supporting Material). Under the assumption that internal friction does not dominate the reaction (36), the temperature dependence of water viscosity was taken into account using the empirical equation (Eq. 5-2) with $\eta_0 = 2.4152 \cdot 10^{-5}$ Pa s, $B = 4.7428$ kJ mol⁻¹, and $\theta = 139.86$ K (37).

Chemical modification of CypA (K28C) labeled with AlexaFluor 488

For cross-linking with Ru³⁺ (38), 2 μ M of the CypA variant K28C-A488 was incubated with 1 mM tris-bipyridylruthenium chloride (Ru(bpy)₃Cl₂) and 20 mM ammonium persulfate in 40 μ L of 50% methanol/water. The sample was illuminated for 2–4 s by a continuous-wave laser at 488 nm

with a power of 12 mW. Immediately after illumination, the reaction was quenched by the addition of dithiothreitol to a final concentration of 50 mM. To suppress aggregation, GdmCl was also added to a final concentration of 3 M. The modification of the amino groups of 20 μ M of the K28C-A488 variant with 2 mM LC-SPDP (succinimidyl 6-[3'-(2-pyridyldithio)-propionamido] hexonate) was performed for 4 h in 50 mM sodium phosphate, pH 7.0, and 6 M GdmCl at 25°C. The reaction was stopped by the addition of TrisHCl-buffer (pH 7) with a final concentration of 50 mM. Without further purification, the reaction mixtures were diluted to a concentration of 1 nM in 50 mM TrisHCl, 10 mM MgCl₂, and 5 mM KCl, pH 7.5, with varying concentrations of GdmCl to obtain the Stokes radii of the two chemically modified CypA-variants. The complexes of the chemically modified CypA variants with GroEL-SR1 were purified with analytical size-exclusion chromatography before measurement.

Calculation of the entropy cost of confinement

The partition function for a Gaussian chain confined inside a cylinder (10) with height h and diameter d is

$$Z_{\text{Chain}} = \pi r^2 h \left(\frac{32}{\pi^2}\right) \left[\sum_{k=1,2,3,\dots} \frac{1}{x_k^2} \exp\left(-\frac{x_k^2 l_p N b}{3r^2}\right) \right] \times \left[\sum_{k=1,3,5,\dots} \frac{1}{k^2} \exp\left(-\frac{\pi^2 k^2 l_p N b}{3h^2}\right) \right]. \quad (6)$$

Here, r is the radius of the confining cylinder, h is the height of the cylinder, l_p is the persistence length of the confined chain, N is the number of bonds in the chain, b is the bond length, in our case the distance between two successive C α -atoms (0.38 nm), and x_k are the roots of $J_0(x)$, the Bessel function of the first kind of order zero ($x_1 = 2.4$, $x_2 = 5.52$, $x_3 = 8.65$, ...) (10). The term $\pi r^2 h$ in Eq. 6 accounts for the translational degrees of freedom of the chain. The free energy of the chain (in $k_B T$) inside a cylinder is given by

$$F_{\text{Chain}} = -\ln\left(\frac{Z_{\text{Chain}}}{\pi r^2 h}\right). \quad (7)$$

We calculated F_{Chain} for CypA with $N = 166$, $b = 0.38$ nm, and $l_p = 0.2$ nm (see also Hofmann et al. (35)) as parameters for the chain and $h = 4$ nm and $r = 2.25$ nm, as given by Horwich et al. (39), as the dimensions of the cylinder. Since the sums in Eq. 6 are dominated by small values of k (40), we included only the first five terms of the two sums in Eq. 7. The calculation results in $F_{\text{Chain}} = 8 k_B T$.

To compute the entropy of a hard sphere inside a cylinder we used an expression for the volume fraction, f , of the confining cylinder with the volume, V_C , that is accessible to a sphere with radius a . The volume fraction is given by

$$f = \frac{\pi(r-a)^2(h-a)}{\pi r^2 h}. \quad (8)$$

The radius of the compact denatured rhodanese (a) was obtained from the Stokes radius of denatured CAM-Rho (see Results). Finally, the free energy

of confinement (in $k_B T$) is given by $F_{\text{Sphere}} = -\ln(f)$. For rhodanese, we obtain $F_{\text{Sphere}} = 5 k_B T$.

RESULTS

The conformation of denatured CypA and Rho free in solution

To probe the conformational ensembles of denatured CypA and Rho, five variants of each protein with different inter-dye separation were investigated. Alexa Fluor 488 and Alexa Fluor 594 were attached to each variant as donor and acceptor, respectively, to obtain information on the distance between amino acid residues i and j of fluorophore attachment. The distance-related change of the mean transfer efficiency, $\langle E \rangle$, as a function of the number of peptide bonds between the fluorophores, $|i - j|$, can be used to characterize the conformational distribution of the denatured proteins (41,42). At high concentrations of GdmCl, which effectively suppresses interactions within denatured proteins (35,43), $\langle E \rangle$ decreases with increasing sequence separation of the dyes for both proteins (Fig. 2 *b*), as expected for fully unfolded proteins based on the length scaling of polymer models for expanded and unstructured chains (41,44).

However, to understand the effect of GroEL on the conformation of proteins, it is crucial to compare denatured CypA and Rho under near-physiological conditions in the absence of GroEL. To this end, we transiently populate the denatured proteins at low denaturant concentrations (0.4 M GdmCl) in a microfluidic mixing device designed specifically for kinetic single-molecule experiments (17,33). The proteins in 4 M GdmCl were mixed with physiological buffer within a dead time of 4 ms (33), and both proteins were still fully denatured 125 ms after transfer to 0.4 M GdmCl (Fig. 2 *a* and Fig. S1 in the Supporting Material) (19,44). Correspondingly, only two peaks are observed in the transfer efficiency histograms for each variant: the peak at high E results from double-labeled denatured protein, and the peak near $E = 0$ results from molecules lacking an active acceptor dye, a peak that can be eliminated by alternating excitation of donor and acceptor (45) (Fig. 2 *a*).

For denatured CypA at 0.4 M GdmCl, $\langle E \rangle$ of all variants is substantially higher than at 7.3 M GdmCl owing to the formation of interactions within the polypeptide that result in a compaction of the denatured protein (44,46,47). However, the transfer efficiencies of the different CypA variants still decrease with increasing sequence separation of the fluorophores, indicating that no or little specific tertiary structure is formed in denatured CypA (Fig. 2 *b*). In contrast to CypA, the denatured Rho variants at 0.4 M GdmCl show unusually high transfer efficiencies of $\langle E \rangle > 0.8$, independent of sequence separation from 39 to 159 peptide bonds (Fig. 2 *b*). The same result was found for CAM-Rho, a chemically modified version of rhodanese that resembles the denatured state of rhodanese but is unable to fold (Fig. S2). These high

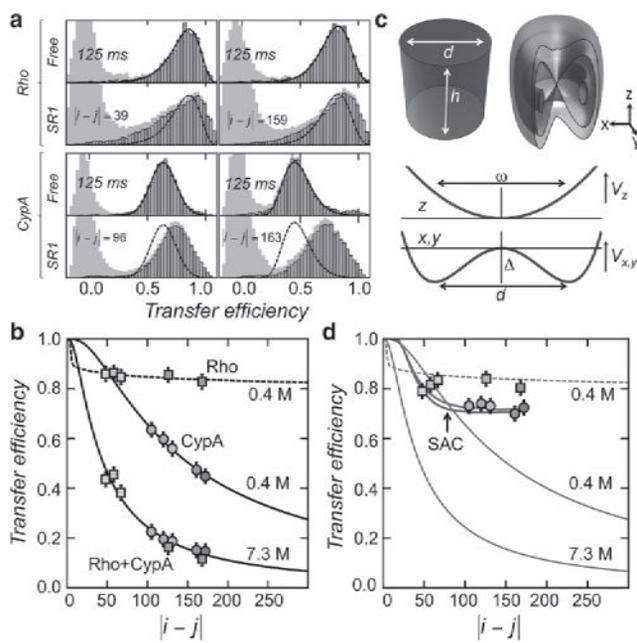


FIGURE 2 Transfer efficiency histograms and mean transfer efficiencies for denatured CypA and Rho free in solution and bound to GroEL-SR1. (*a*) Histograms for the shortest (blue) and longest (red) inter-dye variants of denatured Rho and CypA free in solution at 125 ms after mixing with native buffer in the microfluidic device (final GdmCl concentration, 0.4 M) and bound to GroEL-SR1. Solid lines are fits with a log-normal distribution, which are also shown as dashed lines in the histograms of CypA and Rho bound to GroEL-SR1. (*b*) Mean transfer efficiency of denatured Rho (squares) and CypA (circles) variants as a function of the sequence separation between the fluorophores at two different GdmCl concentrations. The length of the fluorophore linker was estimated to be equivalent to nine additional peptide bonds (35). Solid lines are fits with the ideal chain model with persistence length l_p as the only free parameter. The dashed line is a fit with the ideal chain model with the persistence length and the length-scaling exponent (ν) as free parameters (see the Supporting Material). (*c*) Confining geometries used in SAC simulations. Shown are the repulsive cylinder (upper left) and the potential with attractive interactions between the chain and the confining walls (upper right). The potential is harmonic in the z -direction (V_z) with spring constant ω^{-2} , and the potential in the xy -axis ($V_{x,y}$) is a double-well potential with barrier height Δ and distance d between the two wells. (*d*) Mean transfer efficiencies of denatured Rho (squares) and CypA (circles) variants bound to SR1 are shown as a function of the sequence separation between the fluorophores. Red and blue lines are the result of SAC simulations of a chain with 167 amino acids inside a cylinder (red) ($d = 7$ nm and $h = 8$ nm) and inside a potential (blue) ($\Delta = 1 k_B T$, $d = 4$ nm, and $\omega = 2$ nm) that best describe the measured transfer efficiencies (see Supporting Material). The gray lines are identical to the fits in *b* and are shown here for comparison. Error bars (± 0.03) in *b* and *d* represent our estimate of the uncertainty in the determination of transfer efficiencies (see Materials and Methods). To see this figure in color, go online.

transfer efficiencies and their independence of sequence separation cannot be reconciled with the properties of an expanded unstructured chain but rather suggest a very compact conformation (42) of denatured Rho under near-physiological conditions. Indeed, previous experiments have even suggested the formation of rather specific structure in denatured Rho at 0.4 M GdmCl (27). CypA and Rho can therefore be considered representatives of two

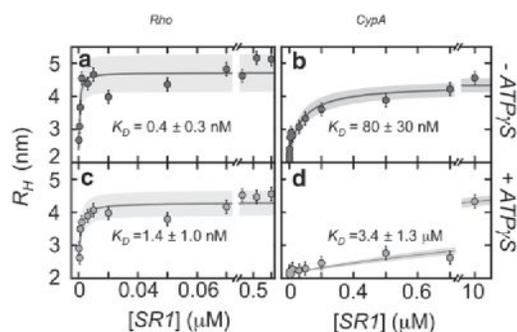


FIGURE 3 Binding of CypA and Rho by GroEL-SR1. Average Stokes radii (R_H) of CAM-Rho (E77C/K135C-A488/A594) (a) and CypA (V2C-A488) (b) as a function of the concentration of GroEL-SR1, determined using 2f-FCS at 22°C. (c and d) Same as in a and b, but in the presence of 1 mM ATP γ S. The shaded regions indicate the error band resulting from the propagation of the errors in the fit parameters (Eq. 2 in Materials and Methods). To see this figure in color, go online.

extremes in the substrate spectrum of GroEL: random-coil-like, with large conformational entropy on the one hand (CypA), and compact, partially structured, with low conformational entropy on the other (Rho). This characteristic difference between CypA and Rho is our starting point for investigating the role of denatured state properties on binding and encapsulation in the chaperonin GroEL.

The conformational distribution of CypA and Rho in complex with single-ring GroEL

In the absence of ATP and under physiological conditions, both CypA and Rho bind strongly to GroEL with dissociation constants of $K_D = 0.4 \pm 0.3$ nM for CAM-Rho and $K_D = 80 \pm 30$ nM for CypA, as determined by 2f-FCS (34) on the single-ring variant of GroEL (SR1) (Fig. 3, a and b). Comparable binding affinities have been found for DM-MBP ($K_D = 60$ nM) (48), a variant of maltose binding protein, and α -lactalbumin ($K_D = 27$ nM) (49). In complex with SR1, the transfer efficiency histograms of all variants of CypA and Rho are broader than those of the free denatured proteins under physiological conditions (Fig. 2 a), predominantly as a result of the increased fluorescence anisotropy of the donor and acceptor fluorophores (17,27) (Fig. S3).

A comparison of the mean transfer efficiency before (Fig. 2 b) and after binding to SR1 (Fig. 2 d) reveals a striking difference between CypA and Rho. The values of $\langle E \rangle$ for Rho are virtually unaltered on binding to SR1 (Fig. 2, b and d), implying that Rho remains compact in complex with the chaperone. The values of $\langle E \rangle$ for the CypA variants bound to SR1, however, are strongly increased compared to those of free denatured CypA (Fig. 2, b and d) and become virtually independent of $|i - j|$ (Fig. 2 d), reminiscent of the behavior of compact denatured Rho. This change in the length scaling of $\langle E \rangle$ on binding to GroEL indicates a substantial effect of confinement on denatured CypA. Both

the higher value of the transfer efficiencies and their independence of sequence separation indicate that the CypA chain is more compact in complex with GroEL than free in solution. However, the transfer efficiencies are still lower than those of Rho in complex with SR1, indicating a looser conformation of CypA compared to Rho (Fig. 2 d).

To elucidate whether the observed transfer efficiencies of CypA bound to GroEL are still in accord with a largely unstructured polypeptide, we used simulations of self-avoiding random chains (SACs) inside confining geometries with the dimensions of the cavity of GroEL with and without including attractive interactions with the cavity walls (Fig. 2 c and Fig. S5). Indeed, confining a SAC with the length of CypA (167 aa) to a small volume qualitatively reproduces the effect of $|i - j|$ -independent transfer efficiencies (Fig. 2 d), suggesting that confinement inside the central hole of GroEL, the location of the substrate binding sites, sufficiently explains the altered chain statistics of CypA in complex with SR1.

Even though both CypA and Rho bind to SR1 with nanomolar affinity (Fig. 3, a and b), the difference between their free energies of binding is significant ($\Delta\Delta G_{\text{Rho-CypA}} = 5 \pm 1 k_B T$). Considering the difference in the conformations of both substrates free and bound to SR1 (Fig. 2, b and d), we estimated the impact of the different denatured-state properties on the free energies of SR1 binding by computing the entropy cost of confining Rho and CypA in the interior of GroEL. Since denatured Rho is very compact, both in solution and also in complex with SR1 (Fig. 2, b and d), a lower limit for the entropy cost of GroEL binding is estimated from the process of trapping a sphere in the interior of GroEL. We modeled GroEL as a cylindrical cavity with a radius of 2.25 nm and a height of 4 nm (39) (see Materials and Methods) and estimated the radius of gyration of denatured Rho based on the Stokes radius of denatured CAM-Rho ($R_H = 2.7$ nm). Assuming that compact denatured Rho has a spherical shape, we obtain $R_G \approx 2$ nm using $R_G/R_H = (3/5)^{1/2}$ (50). In contrast, CypA is approximated as an ideal chain, which allows a straightforward calculation of the entropy cost on confinement in the same cylinder (see Materials and Methods). This approximation is supported by the length-scaling exponent (ν) of CypA, which has recently been found to be close to the value expected for an ideal chain ($\nu = 0.5$) (35). We note that neither estimate includes effects arising from solvent entropy, such as the displacement of water molecules during binding to GroEL. With these approximations, the difference in conformational entropy ($-T\Delta\Delta S_{\text{conf}}$) between the confinement of CypA and Rho inside the central hole of SR1 is estimated to be $\sim 3 k_B T$ and varies between $2.1 k_B T$ and $4.5 k_B T$ with a change in the cylinder volume of $\pm 25\%$. The comparison of the experimentally observed free-energy difference for SR1 binding ($\Delta\Delta G_{\text{Rho-CypA}} = 6 \pm 1 k_B T$) with our estimate of $\sim 3 k_B T$ obtained from polymer theory indicates that about half of the destabilization of the CypA-SR1 complex relative to

the Rho-SR1 can already be explained by the loss in conformational entropy upon confinement, which suggests that the polymeric properties of denatured substrates can significantly affect the binding of denatured substrates to GroEL. The remaining free-energy difference of $2 k_B T$ is likely to result from a greater interaction enthalpy of Rho with GroEL owing to its longer sequence, which allows more contacts with GroEL to be made.

Dynamics and thermodynamics of CypA and Rho interactions with GroEL

The different effects of GroEL on the conformation of CypA and Rho (Fig. 2, *b* and *d*) suggest that the kinetics of binding may also be different for the two denatured proteins. We used ensemble stopped-flow fluorescence of donor-labeled CypA and Rho to monitor the binding kinetics in the time regime of milliseconds to seconds. The binding to SR1 changes the fluorescence intensity of AlexaFluor 488 due to PET (51) to aromatic amino acids such as tryptophan and tyrosine in Rho and CypA, respectively. Compared to single-molecule FRET, PET is sensitive for local distance changes, which is especially advantageous for the study of protein dynamics in confined spaces where large distance changes cannot occur.

Mixing of rhodanese and cyclophilin with SR1 under pseudo-first-order conditions results in single-exponential binding kinetics with increasing fluorescence intensity for Rho and rising fluorescence intensity for Cyp (Fig. 4, *b* and *c*). However, the observed rate constant for binding, k , increases nonlinearly with increasing concentration of SR1 and saturates at high SR1 concentrations (Fig. 4, *d* and *e*), which indicates the presence of a second kinetic step that is rate-limiting under these conditions. A generalized binding model (52) that includes two steps, the diffusion-controlled formation of a transient-encounter complex between GroEL and the denatured substrate (S), $[\text{GroEL} \cdot \text{S}]^*$, followed by a conformational adaptation of the denatured proteins to the GroEL surface, describes this behavior quantitatively (Fig. 4 *a*). The rate constant for the adaptation process, k_{max} , can then be obtained as the asymptotic limit of the binding constant, k , approached at high SR1 concentrations (see Materials and Methods). Whereas $k_{\text{max}} = 105 \pm 8 \text{ s}^{-1}$ at 25°C for CypA, the adaptation process is slower for Rho, with $k_{\text{max}} = 27 \pm 3 \text{ s}^{-1}$ at 25°C . The diffusive chain reconfiguration time of denatured CypA free in solution was found to be 150 ns ns-FCS (Fig. S4). Since the conformational adaptation of both denatured proteins to SR1 is several orders of magnitude slower than the diffusive chain dynamics, we conclude that a large activation barrier must be involved in the adaptation to the SR1 surface. To obtain structural information from the activation parameters of this process, we express the adaptation rate constant in terms of a generalized reaction rate equation $k_{\text{max}} = k_0 \exp(-\Delta G^\ddagger/RT)$, and estimate the attempt fre-

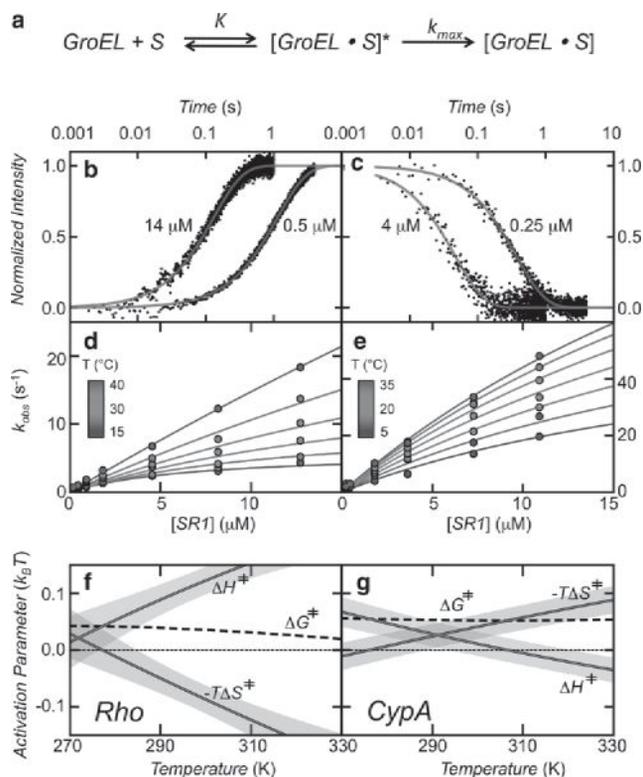


FIGURE 4 Binding and adaptation kinetics measured with stopped-flow fluorescence at 0.4 M GdmCl. (*a*) Generalized mechanism for the bimolecular reaction of substrate binding to GroEL-SR1. (*b* and *c*) Examples of progress curves for Alexa 488-labeled Rho (E285C) (*b*) and CypA (D13C/G124C) (*c*) at two concentrations of GroEL-SR1 indicated. (*d* and *e*) Change of the apparent pseudo-first-order rate constants for binding of Rho (*d*) and CypA (*e*) at different temperatures and SR1 concentrations and fits according to the model in Scheme 1 (see Materials and Methods) with the parameters $\Delta H_{\text{eq}} = -90 \pm 25 \text{ kJ mol}^{-1}$, $\Delta S_{\text{eq}} = -0.44 \pm 0.08 \text{ kJ mol}^{-1} \text{ K}^{-1}$, $\Delta H^\ddagger = 85 \pm 18 \text{ kJ mol}^{-1}$, $\Delta S^\ddagger = 0.2 \pm 0.1 \text{ kJ mol}^{-1} \text{ K}^{-1}$, and $\Delta C^\ddagger = 2.7 \pm 0.9 \text{ kJ mol}^{-1} \text{ K}^{-1}$ for Rho, and with the parameters $\Delta H_{\text{eq}} = -8 \pm 11 \text{ kJ mol}^{-1}$, $\Delta S_{\text{eq}} = -0.17 \pm 0.04 \text{ kJ mol}^{-1} \text{ K}^{-1}$, $\Delta H^\ddagger = 6 \pm 9 \text{ kJ mol}^{-1}$, $\Delta S^\ddagger = -0.05 \pm 0.03 \text{ kJ mol}^{-1} \text{ K}^{-1}$, and $\Delta C^\ddagger = -0.7 \pm 0.2 \text{ kJ mol}^{-1} \text{ K}^{-1}$ for CypA. The values for the activation parameters are given at $T_0 = 298 \text{ K}$ (see Eq. 4 in Materials and Methods). (*f* and *g*) Temperature dependence of the activation parameters for the adaptation processes for Rho (*f*) and CypA (*g*). All activation parameters are given in units of $k_B T$ per amino acid residue. Shaded regions indicate the error of the fits in *c* and *d*, which result from the fitting-parameter error propagation. The change in the viscosity of water with increasing temperature is taken into account (see Materials and Methods). To see this figure in color, go online.

quency (k_0) for crossing the barrier (ΔG^\ddagger) based on the chain reconfiguration times of denatured Cyp and Rho determined in free solution to be $k_0 \approx 1 \mu\text{s}^{-1}$ (Supporting Material) (53,54). By combining this estimate for the preexponential factor with the binding kinetics as a function of temperature (Fig. 4, *d* and *e*), we obtain the change in activation enthalpy (ΔH^\ddagger), activation entropy (ΔS^\ddagger), and heat capacity (ΔC_p^\ddagger) during adaptation (Fig. 4, *f* and *g*) (see Materials and Methods).

The changes in ΔH^\ddagger and ΔS^\ddagger are diametrically opposed for Rho and CypA. With increasing temperature, the

adaptation barrier for Rho is increasingly determined by enthalpy rather than entropy because of a positive heat capacity change (Fig. 4 *f*). This behavior is typical of the interaction of nonpolar solutes with water (55,56). With increasing temperature, the disorder of water clusters around hydrophobic residues increases, leading to a decrease in the entropy change upon mixing. At high temperatures, the aversion of nonpolar solutes to water is therefore mainly enthalpy-driven (55,56). For Rho, the temperature dependence of ΔH^\ddagger and ΔS^\ddagger therefore points to an exposure of hydrophobic residues in the adaptation process, indicating that changes in the hydration of these residues dominate the observed entropy change upon binding to GroEL. Although the change in quantum yield of Alexa-Fluor 488 upon binding of Rho to SR1 clearly demonstrates the presence of local conformational rearrangements in denatured Rho, the absence of an increase in activation entropy with increasing temperature, which could indicate an increasing loss in conformational entropy, suggests that confinement does not contribute much to the reaction, which is in line with the small change in the transfer efficiencies of all Rho variants on binding to GroEL. For CypA, on the other hand, the heat capacity change is negative and ΔS^\ddagger more and more dominates the adaptation barrier with increasing temperature (Fig. 4 *g*). Both the increasing cost of restricting the conformational distribution by GroEL-induced confinement and the net burial of hydrophobic side chains of CypA can explain this decrease in ΔS^\ddagger . Both interpretations are in accord with the compaction of the initially well-solvated and expanded CypA chain upon binding to GroEL-SR1, as revealed by the single-molecule FRET experiments (Fig. 2, *b* and *d*).

In summary, the differences in the kinetics, thermodynamics, and transfer efficiency of GroEL binding reflect differences in the degree of compaction of denatured Rho and CypA free in solution. How do these differences affect the key step in chaperonin action, the encapsulation of both proteins upon addition of ATP and GroES?

Encapsulation of Rho versus expulsion of CypA by GroEL-GroES

We monitored the encapsulation of CypA and Rho by 2f-FCS (34) and size-exclusion chromatography (Fig. S7). The average Stokes radii of complexes of donor-labeled CypA or Rho with SR1 are dominated by the size of SR1, resulting in values of 4.5–5.3 nm (Fig. 5 *a*). Binding of ATP to GroEL is known to trigger a rotation of the apical domains, followed by an upward movement on binding of GroES (57), which concludes the encapsulation reaction. The entire process takes place within seconds (58). Upon addition of 2 mM ATP and 1 μ M GroES (reflecting physiological concentrations) to the preformed Rho-SR1 complex, R_H of the fluorescent species increases from 5.3 nm to 5.9 nm, indicating that Rho is encapsulated in the central

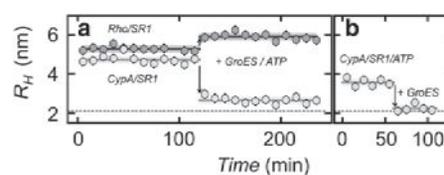


FIGURE 5 Encapsulation of Rho and release of CypA by GroEL-SR1. (a) Effect of the addition of 1 μ M GroES and 2 mM ATP on R_H of preformed complexes of GroEL-SR1 with CypA (green) and Rho (blue) at a final concentration of 25 nM GroEL-SR1. Arrows indicate the time of addition of ATP and GroES. (b) R_H of CypA in the presence of 275 nM GroEL-SR1 and 2 mM ATP before and after the addition of 1 μ M GroES (arrow). Solid lines indicate the mean of R_H . Dashed line indicates the R_H of free CypA. To see this figure in color, go online.

cavity formed by SR1 and GroES (Fig. 5 *a*). Surprisingly, however, the addition of ATP and GroES to CypA-SR1 complexes leads to a decrease of the Stokes radius to 2.6 nm, close to the Stokes radius of free CypA (2 nm) (Fig. 5 *a*). Evidently, the majority of CypA is not stably encapsulated in the cavity formed by SR1 and GroES, as also confirmed by size-exclusion chromatography (Fig. S7).

Why is CypA not encapsulated even though it binds to SR1 with nanomolar affinity? In contrast to Rho-SR1, we found that the stability of the CypA-SR1 complex is strongly diminished, by $\sim 4 k_B T$, in the presence of 1 mM ATP γ S, a nonhydrolyzable ATP analog (Fig. 3, *c* and *d*). This effect has been observed for other substrate proteins (59–61), such as DM-MBP, a variant of maltose-binding protein (48). Taking this decrease in affinity into account, a simple explanation for the failure of GroEL-GroES to encapsulate CypA would be a kinetic competition of GroES binding with the ATP-induced dissociation of CypA from GroEL. Encapsulation will only be successful for those GroEL-substrate complexes where GroES happens to bind before the substrate dissociates after ATP binding. In this case, preforming the CypA-SR1-ATP complex should lead to more efficient encapsulation of CypA. At a concentration of 275 nM SR1, the ternary CypA-SR1-ATP complex is populated significantly, as revealed by the Stokes radius of 3.8 nm compared to the value of free CypA (2 nm) (Fig. 5 *b*). However, the addition of GroES does not lead to an increased encapsulation yield of CypA. Instead, the Stokes radius again decreases to 2 nm, the value obtained for free CypA (Fig. 5 *b*). These results suggest that it is not the binding of ATP that is responsible for the release of CypA but the binding of GroES.

Assuming that the binding of ATP is faster than the binding and dissociation of substrate (62), three different models can potentially explain the results. In the first model (Fig. 6 *a*, Model 1), binding of GroES is sterically hindered by the presence of CypA. Consequently, CypA has to leave SR1-ATP for GroES to bind. In the second model (Fig. 6 *a*, Model 2), GroES can associate with the CypA-SR1-ATP complex and CypA dissociates during GroES binding. This model assumes the existence of an intermediate in

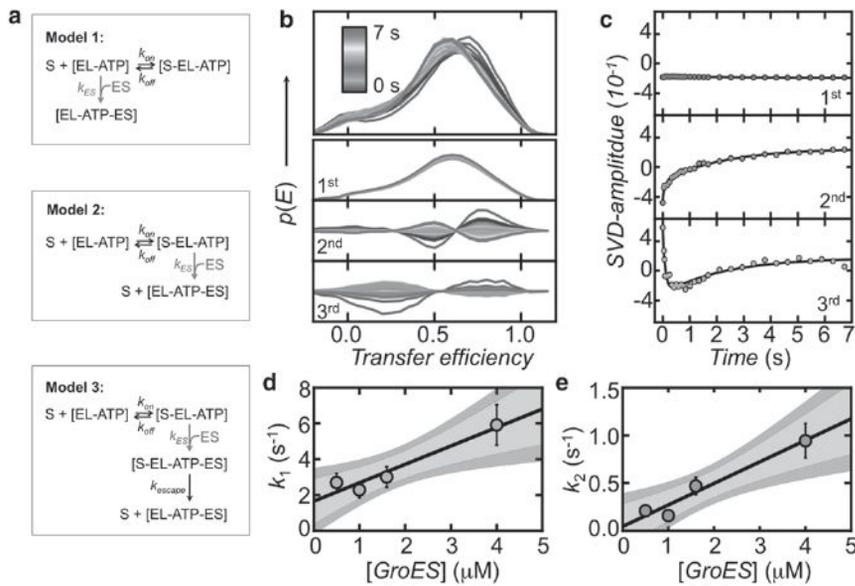


FIGURE 6 Release kinetics of CypA on addition of GroES and ATP. (a) Three different models for CypA release, with binding of GroES highlighted in red. (b) Transfer efficiency histograms (upper) for FRET-labeled CypA (V2C/K154C) bound to GroEL-SR1 at different times after mixing with 2 mM ATP and 1 μ M GroES in the microfluidic device (see Materials and Methods) and the first three resulting singular value decomposition components (lower). The peak at zero transfer efficiency was removed using pulsed interleaved excitation (45). (c) First three amplitude vectors of the SVD shown reflect the progress of the reaction. Solid lines are global fits of the three components with double-exponential decays. (d and e) GroES dependence of the fast (d) and slow (e) phases of CypA release, obtained for CypA variant D13C/G124C. To see this figure in color, go online.

which GroES is in a predocking position before it finally binds tightly to the R-state of SR1. Evidence for such an intermediate comes from recent cryo-electron microscopy studies, which identified at least two different conformations of GroEL-ATP, one of which would allow an interaction with GroES (63). As an alternative (Fig. 6 a, Model 3), denatured CypA is successfully encapsulated but leaves the SR1-ATP-GroES complex at a later stage. This mechanism is supported by the observation that substrate proteins can escape the cage formed by GroEL-GroES even after the encapsulation has taken place (64). To distinguish between these mechanisms, we investigated the kinetics of CypA dissociation from the GroEL-GroES complex using smFRET in combination with microfluidic mixing. Whereas a GroES-driven expulsion of CypA (Models 1 and 2) would occur within 10 s after the addition of GroES and ATP, an escape from the SR1-ATP-GroES cavity (Model 3) is expected to occur in the time regime of several minutes (64).

Kinetics of CypA expulsion from the SR1-ATP-GroES complex

In a time-resolved experiment using microfluidic mixing, we can monitor the kinetics of CypA dissociation during binding of ATP and GroES using smFRET (Fig. 6, b and c). A sample solution containing CypA-SR1 complexes was rapidly mixed with buffer containing ATP and GroES, and transfer efficiency histograms were obtained at different times after initiating the reaction (Fig. 6 b). For all CypA variants, a model-free analysis of the data using multidimensional singular value decomposition (17) (Fig. 6, b and c) resulted in three significant components (Fig. S6) with clearly double-exponential kinetics and two apparent rate constants ($k_1 = 10.5 \text{ s}^{-1}$, $k_2 = 0.44 \text{ s}^{-1}$) (Fig. 6 c), implying at least a three-state mechanism for the dissociation

of CypA. Most important, the rate constant (k_2) for the slow process suggests that $\sim 90\%$ of the SR1-bound CypA is already dissociated 5 s after mixing with GroES and ATP. The formation of the ternary SR1-ATP-GroES complex in the presence of substrate has previously been shown to occur on the same timescale (58), suggesting that CypA leaves GroEL during GroES binding. A slow escape after encapsulation (Fig. 6 a, Model 3) can therefore be excluded.

A distinction between Models 1 and 2 is possible from the GroES-concentration dependence of the CypA release kinetics. In Model 1, GroES only binds to SR1 without CypA bound. Correspondingly, the kinetics of CypA release are determined by the spontaneous dissociation of CypA and the rate constants for the two observed kinetic phases (k_1 , k_2) should not depend on the GroES concentration. In contrast, if CypA is released during binding of GroES (Fig. 6 a, Model 2), CypA dissociation will accelerate with increasing GroES concentrations. Experimentally, we find an increase in the rate constants for the fast (k_1) and slow (k_2) processes with increasing concentration of GroES (Fig. 6, d and e), indicating that GroES binding and CypA release are coupled processes. Hence, only Model 2 is in accord with the experimental data, i.e., CypA is released during the process of GroES binding.

Encapsulation versus expulsion: the role of substrate properties

Since GroES binding triggers the release of CypA, the interaction free energy between CypA and SR1 is apparently insufficient to resist the forces associated with the rearrangement of the apical domains of SR1 on GroES binding (63). However, what are the relative contributions of enthalpic or hydrophobic interactions between chaperone and substrate

protein and the conformational entropy for encapsulation of a disordered chain like CypA? To test the importance of the interaction strength between CypA and SR1 for encapsulation, we increased the hydrophobicity of the CypA sequence by decorating it with the amino-reactive hydrophobic linker LC-SPDP, thus forming LC-CypA. As expected, LC-CypA binds stably to SR1, resulting in a Stokes radius of 5.3 nm for the LC-CypA-SR1 complex (Fig. 7 *e*). It is important to point out that after the addition of GroES and ATP to preformed LC-CypA-SR1 complexes, we found a significantly increased encapsulation efficiency compared to that observed for CypA (Fig. 7, *d* and *e*). The Stokes radius of the free protein measured with 2f-FCS revealed that LC-CypA is more compact by ~ 0.6 nm than unmodified CypA in 0.5 M GdmCl (Fig. 7, *a* and *b*). This reduction in the Stokes radius is explained by the greater hydrophobicity of the sequence, which promotes a hydrophobic collapse of the chain (44,47). These observations raise the question of whether the compaction of the LC-CypA chain, as compared to CypA, lowers the entropic cost for encapsulation and is thus responsible for the increased encapsulation yield. We therefore cross-linked CypA intramolecularly by oxidizing tris-bipyridylruthenium(II) to tris-bipyridylruthenium(III) using laser-induced photolysis (see Materials and Methods) (38). The oxidized Ru^{3+} preferentially oxidizes aromatic amino acid residues that then react nonspecifically with nucleophilic groups such as the primary ϵ -amino groups of lysine residues, which should only lead to a marginal change in overall hydrophobicity. 2fFCS shows that at low concentrations of GdmCl (0.5 M), cross-linked CypA (Ru-CypA) (Fig. 7 *c*) is more compact

by ~ 0.4 nm than unmodified CypA (Fig. 7 *a*) and forms a stable complex with SR1 (Fig. 7 *f*). However, the addition of ATP and GroES leads to a rapid drop in the Stokes radius similar to that of non-cross-linked CypA (Fig. 7, *d* and *f*), indicating that the encapsulation efficiency of Ru-CypA is not markedly increased compared to CypA. The result therefore suggests that a reduction in the entropy cost of encapsulation is not the key determinant for greater encapsulation efficiency. Even though LC-CypA is more collapsed owing to its greater hydrophobicity, it is the increased hydrophobicity, and not the smaller dimensions, that facilitate its encapsulation.

DISCUSSION

CypA has a low conformational stability, is aggregation-prone at high protein concentrations (29), and binds to GroEL with high affinity (Fig. 3 *b*). However, it fails to become encapsulated in the chaperonin cavity (Fig. 5). This observation is very surprising, since CypA meets all sequence criteria for a GroEL substrate. With a length of 167 amino acids (18 kD), it is well within the size limits of typical substrate proteins (Fig. 1 *c*), and it has a mean hydrophobicity and net charge similar to those of Rho and other identified GroEL substrates (Fig. 1 *b*). Stan and co-workers identified hydrophobic GroEL-binding motifs based on a sequence comparison of 284 GroEL substrates (18) (Fig. 1 *d*), and recent NMR-relaxation studies have demonstrated that the amyloid peptide ($A\beta$ -40) indeed interacts with GroEL via the predicted consensus sequence (65). Based on sequence analysis, we identified three hydrophobic GroEL-binding motifs in CypA, which corresponds to a total of 17 amino acids relevant for the CypA-GroEL interactions, which is within the range reported for typical GroEL substrates (6,7,18) (Fig. 1 *d*). The question therefore arises, which aspects determine the success of chaperonin encapsulation?

Rho and CypA differ substantially in their chain lengths and in the compactness of their denatured-state ensembles. The more expanded conformation of CypA is expected to result in an entropy cost for binding to GroEL that is higher by $\sim 3 k_B T$ than that of Rho, as reflected in the entropy-dominated adaptation barrier that has to be crossed for a stable interaction of CypA with SR1 (Fig. 4 *f*). Correspondingly, rearrangements in the apical domains of GroEL upon GroES binding (57) might affect the interactions between CypA and GroEL more than those between Rho and GroEL, which would favor the release of CypA. However, the intramolecularly cross-linked and more compact Ru-CypA variant demonstrates that the encapsulation efficiency does not increase with increasing compaction of CypA (Fig. 7 *f*), making the higher entropy cost unlikely to be the dominant reason for the failing encapsulation of CypA. Only if the hydrophobicity of CypA is increased, as realized in the LC-CypA variant, can the encapsulation efficiency be improved

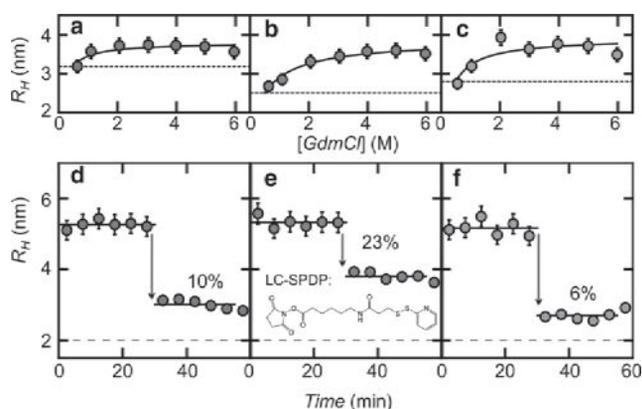


FIGURE 7 Effect of intramolecular cross-linking and chemical modifications of CypA on encapsulation efficiency. (*a–c*) The Stokes radii (R_H) from 2f-FCS of Alexa 488-labeled CypA (*a*), LC-CypA (*b*), and Ru-CypA (*c*) (variant K28C) are shown as a function of the GdmCl concentration. Black solid lines are fits with a denaturant binding isotherm. Black dashed lines indicate the Stokes radius extrapolated to 0.5 M GdmCl. (*d–f*) Effect of the addition of 2 μ M GroES and 1 mM ATP on the R_H of preformed SR1 in complex with CypA (*d*), LC-CypA (*e*), and Ru-CypA (*f*). Arrows indicate the time of addition of 1 mM ATP and 2 μ M GroES. The chemical structure of the LC-SPDP is shown as an inset in *e*. To see this figure in color, go online.

(Fig. 7 *e*). Thus, hydrophobicity facilitates both collapse and encapsulation, but compaction of the substrate protein alone is insufficient for increasing the encapsulation yield.

Our kinetic analysis shows that unmodified CypA leaves the GroEL-GroES complex in a transient state in which GroES is already associated but not yet tightly bound to GroEL, a type of conformation that has recently been suggested based on cryo-electron microscopy (63) and may serve as a checkpoint for substrate discrimination. For GroES to complete the binding process, its mobile loops have to interact with the apical domains of GroEL (63), and the interactions with the substrate have to be released. This leads to the dissociation of CypA; but how is premature substrate release prevented for Rho and other GroEL substrates? A recent model for the encapsulation process of denatured substrate proteins suggests that the C-terminal tails at the base of the GroEL cavity reduce premature substrate protein escape while the apical domains move to interact strongly with GroES (20). Indeed, the last 17 amino acids of the disordered C-terminal tail of GroEL (23 aa) show a high mean hydrophobicity of 0.587 (Fig. 1 *b*). Clearly, the interactions between these disordered tails and substrate proteins will depend on the total number and accessibility of nonpolar groups in the denatured protein. Given the similar mean hydrophobicities of Rho and CypA, the longer Rho chain is expected to interact more strongly with the C-terminal tails than the shorter CypA chain. Thus, our findings are in accord with the model of Chen et al. (20) and suggest that in addition to the presence of an upper limit for the interaction strength between a substrate protein and GroEL, as suggested by the iterative annealing model (66,67), there also exists a lower limit.

In summary, our results demonstrate that high affinity of a substrate protein for GroEL does not necessarily lead to its encapsulation and folding inside the chaperonin cage. Rather, the strength of the hydrophobic interactions of substrate proteins with a transient GroEL-GroES complex is decisive for avoiding premature release of the denatured protein.

SUPPORTING MATERIAL

Supporting Materials and Methods, seven figures, and one table are available at [http://www.biophysj.org/biophysj/supplemental/S0006-3495\(14\)01153-9](http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)01153-9).

We thank G. Lorimer for the gift of the SR1 plasmid. We also thank E. Shakhnovich and A. Horwich for helpful comments and Stephan Benke for help with cross-linking experiments.

This work was supported by a Starting Independent Researcher grant from the European Research Council (to B.S.) the Swiss National Center for Competence in Research for Structural Biology (to B.S.), the Swiss National Science Foundation (to B.S.), the Human Frontier Science Program (to E.A.L. and B.S.), and the National Science Foundation (to E.A.L.).

SUPPORTING CITATIONS

References (68–73) appear in the Supporting Material.

REFERENCES

1. Thirumalai, D., and G. H. Lorimer. 2001. Chaperonin-mediated protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 30:245–269.
2. Fenton, W. A., and A. L. Horwich. 2003. Chaperonin-mediated protein folding: fate of substrate polypeptide. *Q. Rev. Biophys.* 36:229–256.
3. Bukau, B., J. Weissman, and A. Horwich. 2006. Molecular chaperones and protein quality control. *Cell.* 125:443–451.
4. Hartl, F. U., A. Bracher, and M. Hayer-Hartl. 2011. Molecular chaperones in protein folding and proteostasis. *Nature.* 475:324–332.
5. Coyle, J. E., F. L. Texter, ..., S. E. Radford. 1999. GroEL accelerates the refolding of hen lysozyme without changing its folding mechanism. *Nat. Struct. Biol.* 6:683–690.
6. Houry, W. A., D. Frishman, ..., F. U. Hartl. 1999. Identification of in vivo substrates of the chaperonin GroEL. *Nature.* 402:147–154.
7. Kerner, M. J., D. J. Naylor, ..., F. U. Hartl. 2005. Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli*. *Cell.* 122:209–220.
8. Shimamura, T., A. Koike-Takeshita, ..., S. Iwata. 2004. Crystal structure of the native chaperonin complex from *Thermus thermophilus* revealed unexpected asymmetry at the *cis*-cavity. *Structure.* 12:1471–1480.
9. Hartl, F. U., and M. Hayer-Hartl. 2002. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science.* 295:1852–1858.
10. Zhou, H. X., and K. A. Dill. 2001. Stabilization of proteins in confined spaces. *Biochemistry.* 40:11289–11293.
11. Takagi, F., N. Koga, and S. Takada. 2003. How protein thermodynamics and folding mechanisms are altered by the chaperonin cage: molecular simulations. *Proc. Natl. Acad. Sci. USA.* 100:11367–11372.
12. Betancourt, M. R., and D. Thirumalai. 1999. Exploring the kinetic requirements for enhancement of protein folding rates in the GroEL cavity. *J. Mol. Biol.* 287:627–644.
13. Klimov, D. K., D. Newfield, and D. Thirumalai. 2002. Simulations of β -hairpin folding confined to spherical pores using distributed computing. *Proc. Natl. Acad. Sci. USA.* 99:8019–8024.
14. Mittal, J., and R. B. Best. 2008. Thermodynamics and kinetics of protein folding under confinement. *Proc. Natl. Acad. Sci. USA.* 105:20233–20238.
15. Hayer-Hartl, M., and A. P. Minton. 2006. A simple semiempirical model for the effect of molecular confinement upon the rate of protein folding. *Biochemistry.* 45:13356–13360.
16. Tang, Y.-C., H.-C. Chang, ..., M. Hayer-Hartl. 2006. Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell.* 125:903–914.
17. Hofmann, H., F. Hillger, ..., B. Schuler. 2010. Single-molecule spectroscopy of protein folding in a chaperonin cage. *Proc. Natl. Acad. Sci. USA.* 107:11793–11798.
18. Jewett, A. I., A. Baumketner, and J. E. Shea. 2004. Accelerated folding in the weak hydrophobic environment of a chaperonin cavity: creation of an alternate fast folding pathway. *Proc. Natl. Acad. Sci. USA.* 101:13192–13197.
19. Sirur, A., and R. B. Best. 2013. Effects of interactions with the GroEL cavity on protein folding rates. *Biophys. J.* 104:1098–1106.
20. Chen, D.-H., D. Madan, ..., H. S. Rye. 2013. Visualizing GroEL/ES in the act of encapsulating a folding protein. *Cell.* 153:1354–1365.
21. Fischer, G., B. Wittmann-Liebold, ..., F. X. Schmid. 1989. Cyclophilin and peptidyl-prolyl *cis-trans* isomerase are probably identical proteins. *Nature.* 337:476–478.
22. Kallen, J., C. Spitzfaden, ..., M. D. Walkinshaw. 1991. Structure of human cyclophilin and its binding site for cyclosporin A determined by X-ray crystallography and NMR spectroscopy. *Nature.* 353:276–279.
23. Nieba-Axmann, S. E., M. Ottiger, ..., A. Plückthun. 1997. Multiple cycles of global unfolding of GroEL-bound cyclophilin A evidenced by NMR. *J. Mol. Biol.* 271:803–818.

24. Mendoza, J. A., E. Rogers, ..., P. M. Horowitz. 1991. Chaperonins facilitate the in vitro folding of monomeric mitochondrial rhodanese. *J. Biol. Chem.* 266:13044–13049.
25. Jordan, I. K., F. A. Kondrashov, ..., S. Sunyaev. 2005. A universal trend of amino acid gain and loss in protein evolution. *Nature.* 433:633–638.
26. Stan, G., B. R. Brooks, ..., D. Thirumalai. 2006. Residues in substrate proteins that interact with GroEL in the capture process are buried in the native state. *Proc. Natl. Acad. Sci. USA.* 103:4433–4438.
27. Hillger, F., D. Hänni, ..., B. Schuler. 2008. Probing protein-chaperone interactions with single-molecule fluorescence spectroscopy. *Angew. Chem. Int. Ed. Engl.* 47:6184–6188.
28. Langer, T., G. Pfeifer, ..., F. U. Hartl. 1992. Chaperonin-mediated protein folding: GroES binds to one end of the GroEL cylinder, which accommodates the protein substrate within its central cavity. *EMBO J.* 11:4757–4765.
29. Zahn, R., C. Spitzfaden, ..., A. Plückthun. 1994. Destabilization of the complete protein secondary structure on binding to the chaperone GroEL. *Nature.* 368:261–265.
30. Koculi, E., R. Horst, ..., K. Wüthrich. 2011. Nuclear magnetic resonance spectroscopy with the stringent substrate rhodanese bound to the single-ring variant SR1 of the E. coli chaperonin GroEL. *Protein Sci.* 20:1380–1386.
31. Schuler, B. 2007. Application of single molecule Förster resonance energy transfer to protein folding. *Methods Mol. Biol.* 350:115–138.
32. Chen, H., S. S. Ahsan, ..., W. W. Webb. 2010. Mechanisms of quenching of Alexa fluorophores by natural amino acids. *J. Am. Chem. Soc.* 132:7244–7245.
33. Pfeil, S. H., C. E. Wickersham, ..., E. A. Lipman. 2009. A microfluidic mixing system for single-molecule measurements. *Rev. Sci. Instrum.* 80:05105.
34. Dertinger, T., V. Pacheco, ..., J. Enderlein. 2007. Two-focus fluorescence correlation spectroscopy: a new tool for accurate and absolute diffusion measurements. *ChemPhysChem.* 8:433–443.
35. Hofmann, H., A. Soranno, ..., B. Schuler. 2012. Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA.* 109:16155–16160.
36. Soranno, A., B. Buchli, ..., B. Schuler. 2012. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proc. Natl. Acad. Sci. USA.* 109:17800–17806.
37. Likhachev, E. R. 2003. Dependence of water viscosity on temperature and pressure. *Tech. Phys.* 48:514–515.
38. Fancy, D. A., and T. Kodadek. 1999. Chemistry for the analysis of protein-protein interactions: rapid and efficient cross-linking triggered by long wavelength light. *Proc. Natl. Acad. Sci. USA.* 96:6020–6024.
39. Horwich, A. L., G. W. Farr, and W. A. Fenton. 2006. GroEL-GroES-mediated protein folding. *Chem. Rev.* 106:1917–1930.
40. Zhou, H. X. 2008. Protein folding in confined and crowded environments. *Arch. Biochem. Biophys.* 469:76–82.
41. Hoffmann, A., A. Kane, ..., B. Schuler. 2007. Mapping protein collapse with single-molecule fluorescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. USA.* 104:105–110.
42. Tran, H. T., A. Mao, and R. V. Pappu. 2008. Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins. *J. Am. Chem. Soc.* 130:7380–7392.
43. Sherman, E., and G. Haran. 2006. Coil-globule transition in the denatured state of a small protein. *Proc. Natl. Acad. Sci. USA.* 103:11539–11543.
44. Reference deleted in proof.
45. Müller, B. K., E. Zaychikov, ..., D. C. Lamb. 2005. Pulsed interleaved excitation. *Biophys. J.* 89:3508–3522.
46. O'Brien, E. P., G. Ziv, ..., D. Thirumalai. 2008. Effects of denaturants and osmolytes on proteins are accurately predicted by the molecular transfer model. *Proc. Natl. Acad. Sci. USA.* 105:13403–13408.
47. Haran, G. 2012. How, when and why proteins collapse: the relation to folding. *Curr. Opin. Struct. Biol.* 22:14–20.
48. Sharma, S., K. Chakraborty, ..., F. U. Hartl. 2008. Monitoring protein conformation along the pathway of chaperonin-assisted folding. *Cell.* 133:142–153.
49. Li, Y., Z. Zheng, ..., L. Chen. 2010. Analysis of peptides and proteins in their binding to GroEL. *J. Pept. Sci.* 16:693–700.
50. Grosberg, A., and D. Kuznetsov. 1992. Quantitative theory of the globule-to-coil transition. 4. Comparison of theoretical results with experimental data. *Macromolecules.* 25:1996–2003.
51. Doose, S., H. Neuweiler, and M. Sauer. 2009. Fluorescence quenching by photoinduced electron transfer: a reporter for conformational dynamics of macromolecules. *ChemPhysChem.* 10:1389–1398.
52. Schreiber, G. 2002. Kinetic studies of protein-protein interactions. *Curr. Opin. Struct. Biol.* 12:41–47.
53. Kubelka, J., J. Hofrichter, and W. A. Eaton. 2004. The protein folding “speed limit”. *Curr. Opin. Struct. Biol.* 14:76–88.
54. Nettels, D., I. V. Gopich, ..., B. Schuler. 2007. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. USA.* 104:2655–2660.
55. Dill, K. A. 1990. Dominant forces in protein folding. *Biochemistry.* 29:7133–7155.
56. Privalov, P. L., and S. J. Gill. 1988. Stability of protein structure and hydrophobic interaction. *Adv. Protein Chem.* 39:191–234.
57. Clare, D. K., D. Vasishtan, ..., H. R. Saibil. 2012. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell.* 149:113–123.
58. Motojima, F., C. Chaudhry, ..., A. L. Horwich. 2004. Substrate polypeptide presents a load on the apical domains of the chaperonin GroEL. *Proc. Natl. Acad. Sci. USA.* 101:15005–15012.
59. Viitanen, P. V., G. K. Donaldson, ..., A. A. Gatenby. 1991. Complex interactions between the chaperonin 60 molecular chaperone and dihydrofolate reductase. *Biochemistry.* 30:9716–9723.
60. Martin, J., T. Langer, ..., F. U. Hartl. 1991. Chaperonin-mediated protein folding at the surface of groEL through a “molten globule”-like intermediate. *Nature.* 352:36–42.
61. Badcoe, I. G., C. J. Smith, ..., A. R. Clarke. 1991. Binding of a chaperonin to the folding intermediates of lactate dehydrogenase. *Biochemistry.* 30:9195–9200.
62. Tyagi, N. K., W. A. Fenton, and A. L. Horwich. 2009. GroEL/GroES cycling: ATP binds to an open ring before substrate protein favoring protein binding and production of the native state. *Proc. Natl. Acad. Sci. USA.* 106:20264–20269.
63. Clare, D. K., D. Vasishtan, ..., H. R. Saibil. 2012. ATP-triggered conformational changes delineate substrate-binding and -folding mechanics of the GroEL chaperonin. *Cell.* 149:113–123.
64. Motojima, F., and M. Yoshida. 2010. Polypeptide in the chaperonin cage partly protrudes out and then folds inside or escapes outside. *EMBO J.* 29:4008–4019.
65. Libich, D. S., N. L. Fawzi, ..., G. M. Clore. 2013. Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR. *Proc. Natl. Acad. Sci. USA.* 110:11361–11366.
66. Orland, H., and D. Thirumalai. 1997. A kinetic model for the chaperonin assisted folding of proteins. *J. Phys. I France.* 7:553–560.
67. Thirumalai, D. 1994. Theoretical Perspectives on In Vitro and In Vivo Protein Folding. Plenum Press, New York.
68. Lipari, G., and A. Szabo. 1980. Effect of librational motion on fluorescence depolarization and nuclear magnetic resonance relaxation in macromolecules and membranes. *Biophys. J.* 30:489–506.
69. Reference deleted in proof.

70. Gopich, I. V., D. Nettels, ..., A. Szabo. 2009. Protein dynamics from single-molecule fluorescence intensity correlation functions. *J. Chem. Phys.* 131:095102.
71. Henry, E., and J. Hofrichter. 1992. Singular value decomposition: application to analysis of experimental data. *Methods Enzymol.* 210:129–192.
72. Rosenbluth, M., and A. Rosenbluth. 1955. Monte Carlo calculation of the average extension of molecular chains. *J. Chem. Phys.* 23: 356–359.
73. Schuler, B., E. A. Lipman, ..., W. A. Eaton. 2005. Polyproline and the “spectroscopic ruler” revisited with single-molecule fluorescence. *Proc. Natl. Acad. Sci. USA.* 102:2754–2759.

Cell Press content is widely accessible

At Cell Press we place a high priority on ensuring that all of our journal content is widely accessible and on working with the community to develop the best ways to achieve that goal.

Here are just some of those initiatives...

Open archives

We provide free access to Cell Press research journals 12 months following publication

Access for developing nations

We provide free & low-cost access through programs like Research4Life

Open access journal

We launched Cell Reports - a new Open Access journal spanning the life sciences

Funding body agreements

We work cooperatively and successfully with major funding bodies

Public access

Full-text online via ScienceDirect is also available to the public via walk in user access from any participating library

Submission to PubMed Central

Cell Press deposits accepted manuscripts on our authors' behalf for a variety of funding bodies, including NIH and HHMI, to PubMed Central (PMC)

www.cell.com/cellpress/access





LOST IN PAPERS?

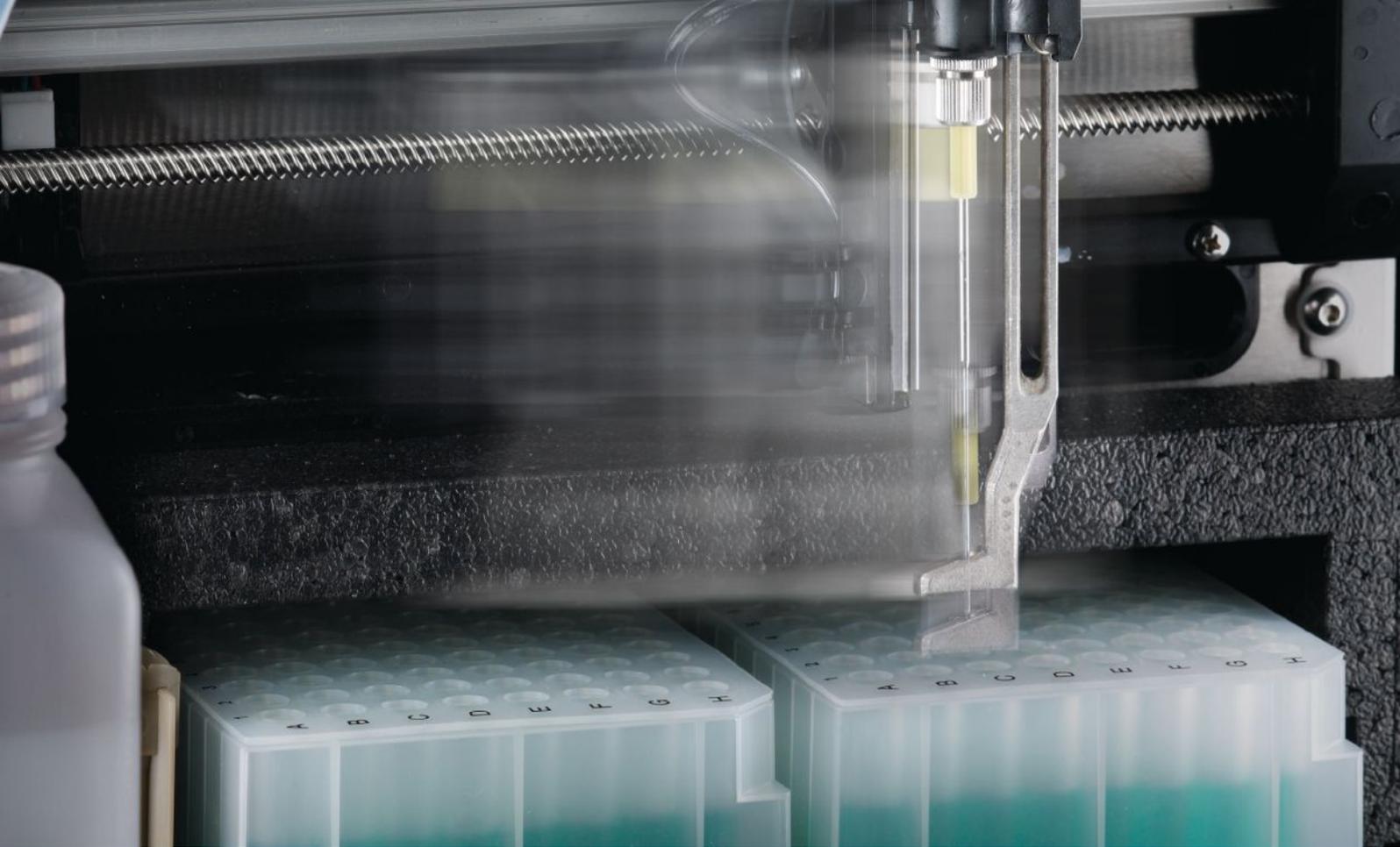
Find your way with Cell Press Reviews

Whether you're looking for answers to fundamental questions, brushing up on a new subject or looking for a synthesis on a hot topic, turn to Cell Press Reviews. We publish reviews in a vast array of disciplines spanning the life sciences.

Propel your research forward, faster with Cell Press Reviews. Our insightful, authoritative reviews—published in our primary research and Trends journals—go beyond synthesis and offer a point of view.

www.cell.com/reviews

CellPress
Your work is our life



Discover More Molecular Structures and Interactions

Nano ITC

Protein - Protein Interactions

- Prioritize Drug Candidate Target Interactions
- Validate Ligand Binding to Nucleic Acid
- Quantify both Enthalpy and Entropy in One Titration
- No labeling or immobilization required

Nano DSC

Protein Structural Domains and Stability

- Excipient Influence on Molecular Stability
- Stability of Biopharmaceuticals
- Direct Measure of Molecular Thermodynamics



www.tainstruments.com