Single Molecule Imaging Simulations with Full Fluorophore Photophysics.

Dominique Bourgeois

Institut de Biologie Structurale, Université Grenoble Alpes, CNRS, CEA, IBS, 38044 Grenoble, France

Correspondence: dominique.bourgeois@ibs.fr

Abstract

Advanced fluorescence imaging techniques such as single-molecule localization microscopy (SMLM) fundamentally rely on the photophysical behavior of the employed fluorophores. This behavior is generally complex and impacts data quality in a subtle manner. A simulation software named SMIS (Single-Molecule Imaging Simulator) is introduced that simulates a widefield microscope and incorporates fluorophores with arbitrarily complex photophysical properties. With SMIS, intricate data collection schemes combining 3D, multicolor, single-particle-tracking or quantitative SMLM can be implemented virtually and the influence of imaging and environmental parameters evaluated.

Main text

Single Molecule Localization Microscopy (SMLM) has become a central tool to decipher the mechanisms of life at the nanoscale[1]. It is now well established that the successful application of SMLM techniques such as Photoactivated Localization Microscopy (PALM), Stochastic Optical Reconstruction Microscopy (STORM) or even Point Accumulation in Nanoscale Topography (PAINT) relies on the proper photophysical behavior of the employed fluorophores, notably when quantitative information is sought, for example in the case of single-particle-tracking[2] or oligomer counting experiments[1]. In general, the results provided by many advanced fluorescence imaging methods may be severely biased by poorly controlled fluorophore phototransformations.

Several simulation packages have been developed in the last years to aid in setting up and interpreting SMLM experiments[3–8]. However, only relatively simplistic descriptions of fluorophore photophysics have been used, generally based on 3- or 4-state models (typically an inactive state, an active state, a dark state and a bleached state). While simulations obtained with these packages are extremely useful to describe many experimental cases and reproduce some of the artifacts commonly observed in SMLM, they do not incorporate the genuine photophysical behavior of the employed fluorophores. Moreover, they do not allow to thoroughly investigate how a fluorophore responds to specific imaging or environmental parameters employed for data collection. Recently the "Virtual SMLM" software has been introduced, which partly addresses this problem[9]. Yet, the photophysical complexity of the fluorophores and their spectral response to light or environmental conditions is not fully incorporated so that crosstalk effects or e.g. pH- or triplet-state-dependent mechanisms[10,11] cannot be addressed. More generally, biologists seeking to predict data quality for a planned experiment with specific

fluorophores and imaging conditions, chemists interested in the design of new dyes or studying their photophysical behavior, and physicists aiming at improved data acquisition schemes would benefit from the possibility to simulate their experiments, models or schemes *in silico* and confront the outcome with real data. A tool to do so has been lacking thus far.

Here I introduce SMIS (Single Molecule Imaging Simulator) a package able to simulate a wide variety of SMLM experiments with arbitrarily complex photophysical behavior of the fluorophores. Central to SMIS is the computation of the photophysical state of each fluorophore with sub-frame time-resolution in response to any data acquisition scheme specified by the user, possibly including multiple lasers. Evolution of the photophysical state is based on a thorough description of fluorophores that includes spectral data of photoactive states and photo- or thermally-induced transformations between all states. Beyond the description of fluorophore photophysics, SMIS incorporates a virtual widefield microscope, able to operate in multicolor, 3D and single-particle tracking mode (Methods). To highlight the potential of SMIS, I present below four different applications where fluorophore photophysics are shown to influence SMLM data in a complex manner.

The first application aims at investigating how green-state photophysics of the green-to-red photoconvertible fluorescent protein (FP) mEos4b affects the quality of PALM images. Nuclear pore complexes (NPCs) are virtually labeled in 2D with 32 mEos4b molecules fused to Nup96 nucleoporins, assuming a FP maturation level of 90%. A mEos4b model comprising 10 photophysical states is employed[12] (Fig. 1A, Supplementary Table S1), with each of the fluorescent green or red states (anionic chromophore) being in rapid equilibrium with their nonfluorescent (neutral chromophore) counterparts. Sample consisting of 100 NPCs were submitted to data acquisition protocols (Supplementary Table S2) similar to that recently described[13]. An example of ground-truth photophysical data generated by the SMIS data-analysis tool is shown in Supplementary Fig. S1. The retrieved SMLM images upon processing the SMIS-generated image stacks are shown on Fig. 1B. Correlations between those images and a ground-truth image assuming 100% labeling efficiency of Nup96 are shown in Supplementary Fig. S2 and Fig. 1D. In comparison to a case where green-state photophysics are deliberately omitted while red-state photophysics are included (Fig. 1B, Panel 1), the quality of the image obtained at 3.5 kW/cm² 561-nm laser power density when green-state photophysics are considered is reduced (Fig. 1B, Panel 2), with a number of Nup96 sites clearly missing, similar to what was observed experimentally with the protein mMaple[13.] Of note, ~10 % of green-to-red photoconversion events resulted from readout photoconversion induced by the 561-nm laser (Supplementary Fig. S3). Contrary to expectations based on ensemble-level calculations (Fig. 1C), reducing the 561-nm laser power-density to 0.5 kW/cm² produced a degraded image (Fig. 1B, Panel 3). Despite the superior photoconversion efficiency at low intensity (Fig. 1C, Fig. 1F), the lower quality of the image originates from both reduced localization precision (Fig. 1E) and longer on-times of individual mEos4b molecules that induced a higher localization density throughout data collection resulting in more mislocalizations (Supplementary Fig. S4B). In contrast, the superior quality of the NPC image obtained at 3.5 kW/cm² (Fig. 1B, Panel 2), despite only ~60% photoconversion efficiency, benefits from the fact that NPCs are composed of two layers of nearly superposing, indistinguishable molecules, compensating for the strong green-state bleaching of mEos4b molecules at this high laser power density. SMIS also allowed investigating the influence of pH on the

retrieved NPC image quality and to study the possible interest of using Fermi profiles[14] of the 405-nm laser to better distribute localizations throughout data collection. The results are described in Supplementary Text 1 (Fig. 1, Supplementary Fig. S4-S5).

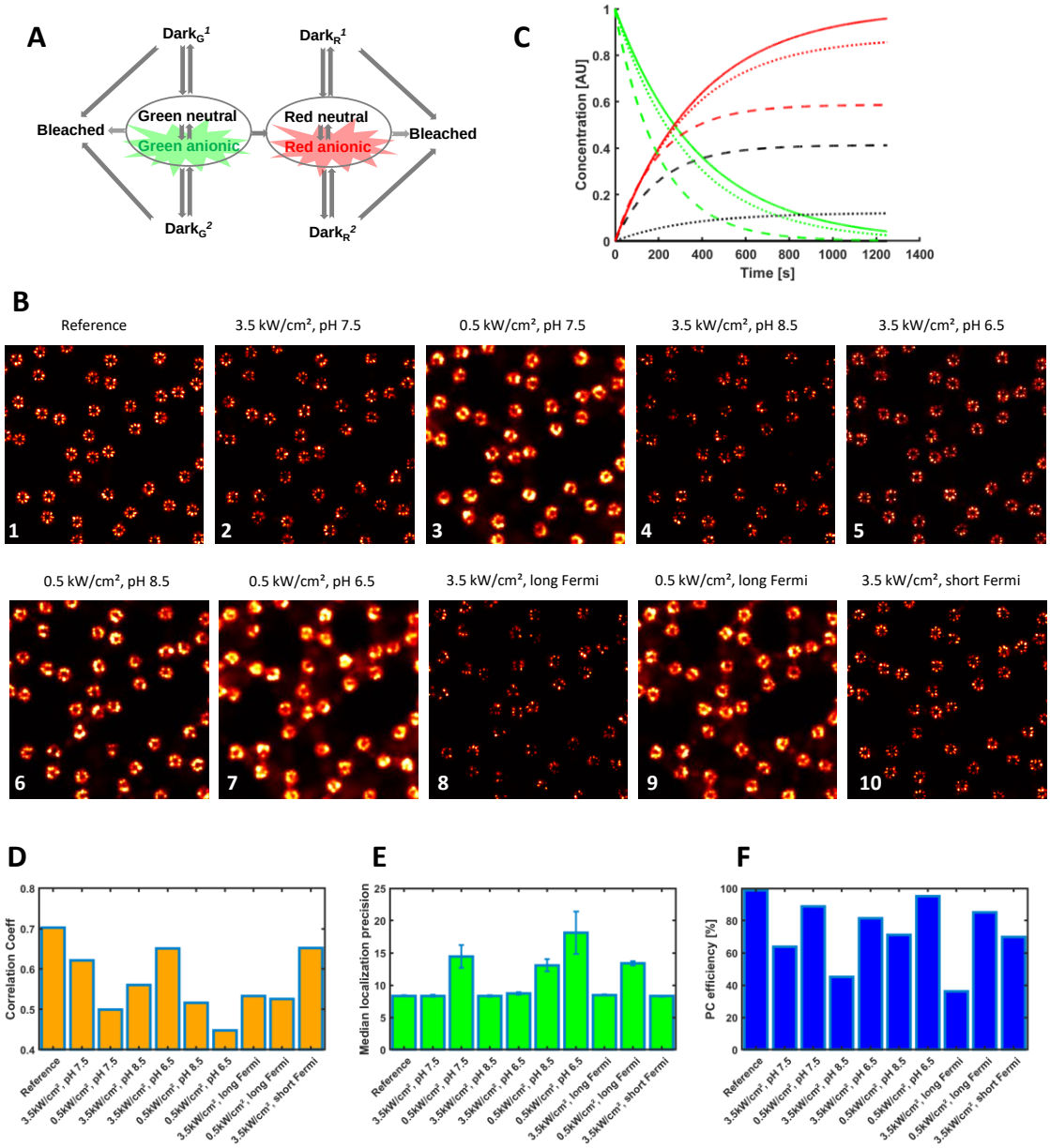

Fig. 1 PALM SMIS simulations of mEos4b-labeled Nup96 in the Nuclear Pore Complex. (A) Employed photophysical model of mEos4b. For both green and red mEos4b, the fluorescent state (anionic chromophore) is in rapid equilibrium with a nonfluorescent state (neutral chromophore). Each colored state can transition to two reversible dark state and one nonreversible bleached state. Dark states may also transition to bleached states. (B) Predicted green-to-red photoconversion efficiency and green-state photobleaching as a function of readout laser intensity. The decay of the green-state population and the rise of the red-state and bleached populations are shown in green, red and black, respectively. Curves are shown when green-state photophysics is neglected (plain lines), or when it is considered with 3.5 kW/cm² (dashed lines) or 0.5 kW/cm² (dotted lines) readout laser power density, respectively. (C) Rendered NPC images (normalized Gaussian) obtained upon processing of SMIS-generated data for various scenarios. Detailed parameters are given in Supplementary Table S2. Green-state photophysics is not considered in the simulation of Panel 1, shown as a reference case. (D) Correlation coefficients between rendered images and a ground-truth NPC image with 100% labeling efficiency. (E) Median merged localization precision retrieved upon processing the SMIS-generated image stacks. Error bars where computed by splitting

the dataset in three subsets along time and calculating standard deviations between the subsets. (F) Ground-truth photoconversion efficiency.

The example above shows that SMIS simulations have the potential to hint at optimal SMLM data collection strategies. Moreover, experimental data deviating from SMIS predictions could be used to refine the employed photophysical model, serving to increase knowledge on fluorescent protein photophysics.

In the second application, SMIS is used to simulate pseudo two-color PALM experiments employing a combination of so-called "primed-photoconversion" with the fluorescent protein Dendra2 and regular photoconversion with mEos4b or photoactivation with PAmCherry[10,15,16]. In such experiments, primed-photoconversion of Dendra2 is first obtained by simultaneous illumination with "priming" light at 488 nm and red-shifted light at 730 nm ("priming phase"), and regular photoconversion of mEos4b or PAmCherry is subsequently obtained with 405 nm laser light ("UV-phase")(Supplementary Table S4). The photoconversion priming mechanism in Dendra2 and other engineered priming variants[17] was proposed to involve excitation of the anionic green-state chromophore at 488 nm, followed by intersystem crossing to a long-lived triplet state[15] which would then be excited by far-red light to finally generate the red-emitting chromophore by a yet unknown photochemical process. Although two-color PALM imaging based on primed photoconversion was recently demonstrated to be successful experimentally[15,16], potential caveats such as premature photobleaching or residual photoconversion of the non-priming partner by 488-nm light were not quantified. Such quantification can be investigated with SMIS, as shown in Fig. 2. Here, a photophysical model of Dendra2 comprising 12 states was employed (Fig. 2A, Supplementary Table S3)[18], including 2 triplet states accessible from either the green or the red chromophore. The virtual sample was based on a neuronal cell, randomly decorated with 5000 of each fluorescent proteins, assumed to exhibit a maturation efficiency of 0.9 (Fig. 2C and Supplementary Fig. S6). The FP-labeled targets were assumed to colocalize within a 5 nm radius. An exemplary trace of a Dendra2 molecule experiencing primed photoconversion is shown in Fig. 2B, demonstrating the complex photophysical trajectory typically followed until photobleaching in the red state. With the chosen photophysical parameters and experimental set up, it can be seen that, as compared to the ground-truth reference, the mean coordinate based colocalization (CBC) and nearest neighbor distance (NND) retrieved upon processing of the SMIS-generated data are significantly degraded (Fig. 2E and Supplementary Fig. S7). The clear but non-optimal colocalization in the case of the Dendra2/mEos4b pair is apparent in Fig. 2D (see also Supplementary Fig. S6). Fig. 2F shows that the origin of the non-ideal colocalization mostly lies in incomplete photoconversion efficiency due to photobleaching in the green state, with a slight contribution from crosstalk errors due to readout photoconversion of the non-priming partner during the priming phase. Further comparison between the photophysical behaviors of mEos4b and PAmCherry in the described experiment are discussed in Supplementary Text 2 (Fig.2, Supplementary Fig. S6-S9)

The example above shows that SMIS is a useful tool to investigate the photophysical origin of nonideal data, and notably crosstalk effects, obtained from complex multicolor SMLM experiments.

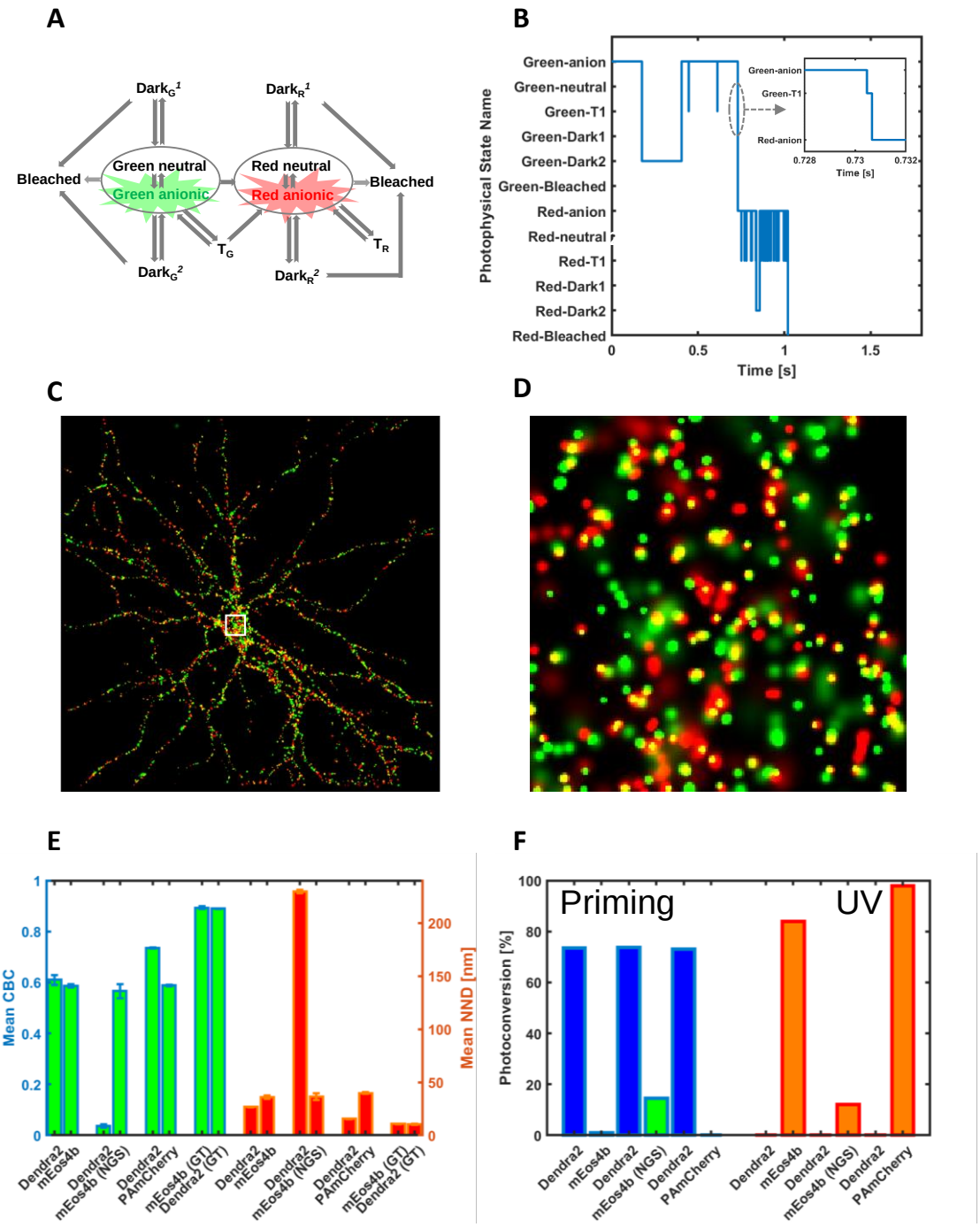

Fig. 2 SMIS simulations of pseudo 2-color PALM imaging using primed photoconversion. (A) Employed photophysical model of Dendra2. In addition to the photophysical states described for mEos4b (Fig. 1), triplet states accessible from both the green and red states are considered. (B) Example of a Dendra2 time trace showing the sequentially visited photophysical states. The inset shows that conversion to the red state goes via the triplet state through primed photoconversion. (C) Rendered 2 color image (green: mEos4b, red: Dendra2) following priming phase and UV phase illumination schemes. Area within the white square is enlarged in (D) for inspection of colocalization events. (E) Mean coordinate based colocalization (CBC) and mean nearest neighbor distance for the various fluorescent protein pairs tested. Error bars show standard deviation for n=3 simulations. (F) Ground truth photoconversion efficiency of the different fluorescent proteins tested during priming phase (left) or UV phase (right). NGS: no green state photophysics considered; GT: ground truth.

In the third application, I aimed to investigate whether a recently proposed photocycle of Cy5[19] properly accounts for the experimental dependence on laser power-density of dSTORM data acquired with the closely

related dye Alexa647[11]. Based on titration experiments, a nonlinear photobleaching effect of Alexa647 was reported, resulting in a reduced effective labeling efficiency of the NPC-Nup96 when the level of red light was raised. Here, the photophysical model of Cy5, described in Fig. 3A, comprises the fluorescence state, the μs-lived triplet state T1, the long-lived sulfur-adduct state RS^- (the main dark state in dSTORM), as well as ms-lived radical anionic and cationic states that can be accessed through T1. All these states can photobleach or interconvert through light-induced or thermal processes as reported by Gidi et al[19]. The *cis*-state was however not included in the scheme, as *cis-trans* isomerization is a very rapid process that can be considered as a non-radiative relaxation pathway lowering the fluorescence quantum yield of Cy5. Absorption spectra, interconversion rates and quantum yields were taken from Gidi et al[19], according to the experimental conditions used in both studies (no oxygen, β−mercaptoethanol concentration = 143 mM, pH 8) (Supplementary Table S5-8, Supplementary Fig. S10). In this application, NPCs were simulated in 3D. A first titration experiment was performed with SMIS (Supplementary Table S9), and representative rendered images collected at 6.4 and 480 kW/cm² of 647 nm light are shown in Fig. 3B. A three-dimensional view of an NPC reconstructed from SMIS data generated with an astigmatic point spread function shows the capability of SMIS for 3D simulations (Fig. 3C). The images of Fig. 3B are in line with the expected lower quality of the reconstruction at the higher power density. Photobleaching quantum yields from the fluorescent and the triplet state (which, to the author's knowledge, have not been reported experimentally) were adjusted to match the experimentally observed labeling efficiency of dSTORM data at the two laser power densities, providing values of ~5×10^{-7} and ~5×10^{-3}, respectively. Fig. 3D shows the evolution of the photophysical state's populations during the first 1000 frames of the two simulations (discarded in the reconstructed dSTORM images[11]), using the ensemble simulation tool of SMIS. At 480 kW/cm² of 647 nm laser light, the populations of the triplet state, and to a lower extent of both the anionic and cationic radical states quickly exceed that of the fluorescence state. As a consequence, intense photobleaching occurs at the early stage of data collection, affecting close to ~60% of the Alexa647 population after 1000 frames, to be compared with ~20% at 6.4 kW/cm². Yet, the single-molecule simulated data (Fig. 3E-F, plain lines) are only partially in line with the experimental data of ref[11] (Fig. 3E-F, dashed-dotted lines) (Supplementary Text 3, Supplementary Fig. S11-S12). Aiming at a better matching of the simulated and experimental data, we used SMIS to guide further refinement of the Cy5 photophysical model of ref[11] notably in view of the possible influence of cellular environment (Supplementary Text 3, Fig. 4, Supplementary Fig. S12). The results suggest that in addition to T1, a long-lived (ms) dark state yet to be identified is likely involved in controlling Cy5 photobleaching under dSTORM conditions. Hints towards the existence of such a state has been provided experimentally[20,21].

The application above shows that SMIS can be a powerful tool to assess and refine the subtle photophysical behavior of fluorescent dyes and even guide future mechanistic investigations.

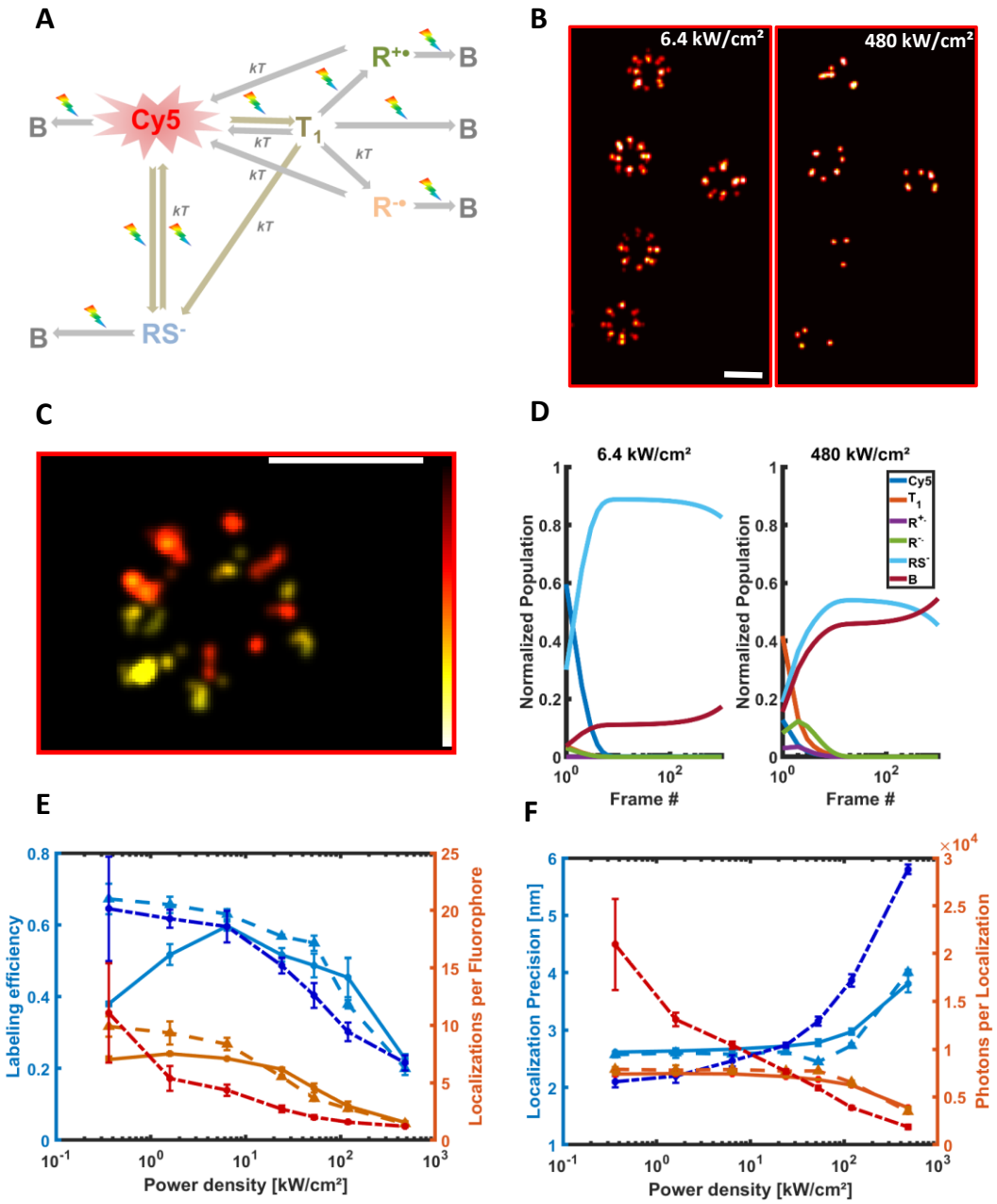

Fig. 3 dSTORM SMIS simulations of the Nuclear Pore Complex to investigate Cy5 photophysics. (A) Employed photophysical model of Cy5. T_1: triplet state; B: bleached state; RS^-: sulfur anion adduct; $R^{+\bullet}$: radical cation; $R^{-\bullet}$: radical anion; kT: thermal relaxation; rainbow arrow: light sensitivity (B) rendered NPC's using 6.4 kW/cm² (left) or 480 kW/cm² (right) 647-nm laser power density. Scale bar: 100 nm. (C) 3D view of a reconstructed NPC (color map encodes position along optical axis, scale bar: 100 nm). (D) Ensemble SMIS simulations showing the evolution of fractional populations of all photophysical states at 6.4 kW/cm² (left) or 480 kW/cm² (right) 647-nm laser power density. (E) Measured labeling efficiency (SMIS data, blue; experimental data, dark blue), and number of localizations per fluorophore (SMIS data, orange; experimental data, red), as a function of the used 640 nm laser power density. Experimental data were taken from ref[11]. Error bars for SMIS data show standard deviation for n=3 simulations. Plain lines: results using photophysical values of ref[19]; Dashed lines: corrected photophysical values (see text) (F) Measured localization precision and number of photons per localization. Color and line coding identical to (E).

The fourth application involves single-particle-tracking PALM (sptPALM), and seeks to investigate how fluorophore photophysics as well as exchange kinetics and molecular confinement may influence accurate retrieval of diffusion coefficients and exchange rates between populations of diffusing molecules. We consider

the scenario depicted in Fig. 4A, inspired from the recent study by Stracy et al[22] of nucleoid associated proteins (NAPs) diffusion dynamics in *E. coli*. Here, NAPs are assumed to slowly exchange between a tightly DNA-bound state and a searching state which is itself a combination of rapidly exchanging freely-diffusing and weakly DNA-bound states. SptPALM data were generated with SMIS considering a virtual 3D sample made of 100 randomly oriented bacteria (Fig. 4B, Supplementary Fig. S13A). The bacteria were assumed to generate slowly bleaching fluorescence background (Supplementary Fig. S13B). The NAPS were labeled with either mEos3.2 or PA-JF549[23], assuming that this photoactivatable dye can be successfully targeted to the bacterial nucleoid to bind e.g. a HaloTag-NAP construct. The photophysical properties of the fluorophores where derived from ref[23] (Supplementary Table S10-11). The study was also aimed at evaluating the potential benefit of using weak 488 nm light in addition to standard 561 nm and 405 nm illumination to increase the tracklengths of mEos3.2-labeled molecules[24]. Data collection parameters were set essentially as in Stracy et al[22] (Supplementary Table S12), except that a stroboscopic mode was employed to reduce temporal averaging[25] (Supplementary Fig. S13C). Care was taken in adjusting the activation light (405 nm) to achieve stringent localization sparsity, ie ~80% of the tracks not overlapping in time with other tracks in single nucleoids. Using Trackit[26] and UTrack[27] the expected significant difference in average tracklength between the 2 types of fluorophores was observed (Supplementary Fig. S13D). However it could be seen that the fidelity of track reconstruction only reached about 80% (Supplementary Fig. S14), which is mostly assigned to the difficulty in disentangling crossing tracks within a highly-confined volume. Next, apparent individual diffusion coefficients based on mean jump distances from tracks containing at least 5 localizations were calculated as described in Stracy et al[22], and histograms were produced (Fig. 4C) from which fractional populations and associated overall apparent diffusion coefficients could be fitted[28]. Based on calibration data (Supplementary Fig. S13E), and correcting for localizations uncertainties, unbiased diffusion coefficients and populations of the 3 involved states where finally derived and compared with the ground truth (see Methods) (Fig. 4D). Interestingly it can be seen that the retrieved diffusion coefficients for the population of searching NAPs are underestimated relative to the ground truth, although the effect is slightly less pronounced for PA-JF549. As a consequence, the relative populations of free and weakly DNA-bound molecules are also incorrectly estimated. Further simulations show that this essentially results from the distortion of the apparent diffusion coefficient histogram (Fig. 4C) in the presence of both kinetic exchange and confinement (Supplementary Fig. S15).

Next we attempted to extract the slow exchange rates between the tightly DNA-bound and the searching states using vbSPT[29]. Imposing a 2-state model in vbSPT allowed accurate retrieving of unbiased diffusion coefficients and populations (Fig. 4E). Interestingly, the accuracy of the recovered exchange rates (Fig. 4F) was slightly better for mEos3.2 in the presence of 488 nm light and significantly better for PA-JF549, showing the advantage of longer tracks even though the number of tracks was lower in PA-JF549 datasets as compared to mEos3.2 datasets. However when the number of diffusing states was not specified in vbSPT, a 3-state model was found (Supplementary Fig. S16A), with exchange rates significantly departing from the ground truth values, suggesting that vbSPT was biased by the strong confinement. Further simulations in which exchange kinetics were suppressed confirmed that this was indeed the case (Supplementary Fig. S17A). To attempt overcoming the 3-state-model bias due to confinement, SMIS simulations were repeated with a 5 ms total frametime. This

time, a 3-state-model bias was observed due to the fast exchange process (Supplementary Text 4, Supplementary Fig. S17-S19). Overall, the results suggest that collecting data with both frametimes is advisable, because consistent data are only produced for a 2-state model, strongly arguing in favor of such model.

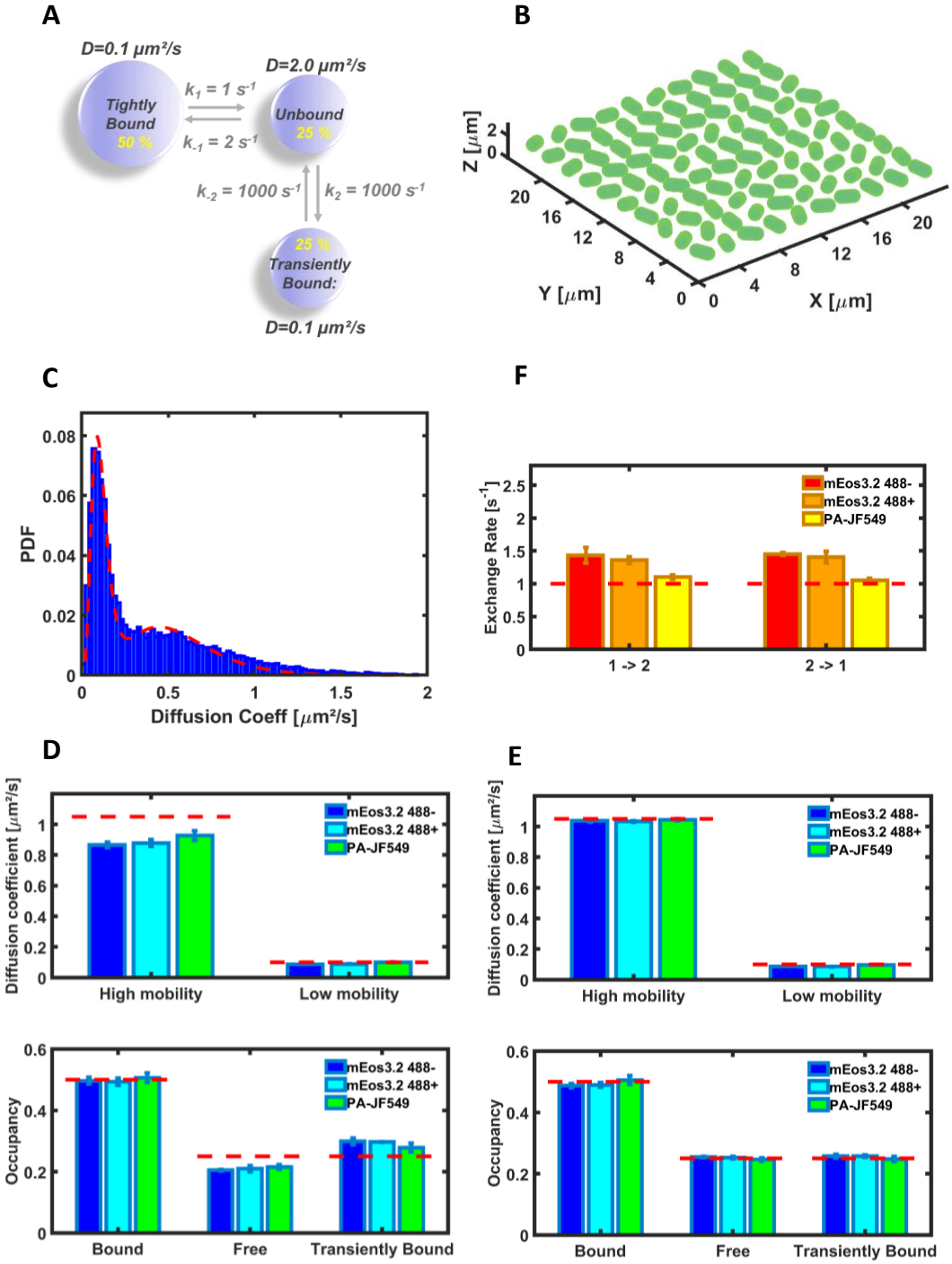

Fig. 4 sptPALM SMIS simulations of mEos3.2- or PA-JF549-labeled NAPs diffusing within the nucleoid of *E. coli* bacteria. (A) Employed model of NAP dynamics. (B) 3D view of the bacterial field used as the virtual SMIS sample. (C) Histogram of apparent diffusion coefficients from a mEos3.2 sample calculated from mean jump distances using 5-frames subtracks (blue), and fitted 2-state model (red). (D) Unbiased diffusion coefficients (upper panel) and diffusion state's occupancies (lower panel) recovered from MJD's-derived diffusion coefficients for mEos3.2 in the absence (blue) or presence (cyan) of 488 nm light, and for PA-JF549 (green). Ground truth values are shown as dashed red lines. (E) Unbiased diffusion coefficients (upper panel) and diffusion state's occupancies (lower panel) recovered from vbSPT-derived diffusion coefficients. Color coding as

in (D). (F) Slow exchange rates recovered from vbSPT, for mEos3.2 in the absence (red) or presence (orange) of 488 nm light, and for PA-JF549 (yellow). 1 → 2: Tight DNA-bound state to searching state; 2 → 1: Searching state to tight DNA-bound state.

The application above shows that SMIS can be a useful tool to investigate intertwining effects between complex diffusion dynamics, data acquisition scheme and fluorophore photophysics in single-particle-tracking experiments.

The 4 examples presented above demonstrate that SMIS is able to simulate a wide variety of complex single-molecule imaging scenarios, highlighting subtle behaviors that are not straightforward to extract from real experiments, but may yet strongly bias their proper interpretation. Many other experimental situations can be simulated by SMIS, including for example pulse chase experiments (Supplementary Fig. S20), PAINT experiments with fluorogenic markers using HILO illumination (Supplementary Fig. S21), or photochromic single-molecule FRET (Supplementary Fig. S22). In conclusion, beyond its pedagogical value, SMIS represents a valuable tool to help the proper setting and/or interpretation of advanced fluorescence imaging experiments in widefield mode. One of its major advantage is the possibility to directly use fluorophores defined in the SMIS fluorophore database, enabling biologists to test their planned experiments. Developers, on their side, can test arbitrarily complex photophysical models of current or new fluorophores, and investigate the outcome of novel laser illumination schemes, aimed at improving the power of SMLM.

Acknowledgements

I am indebted to the many people that helped me throughout this work thanks to inspiring discussions, in particular Jonas Ries, Viktorija Glembockyte, Philip Tinnefeld, Mark Bates, Ulrike Endesfelder, Martin Linden, Jochem Vink, Florian Levet, Daniel Sage and Juliette Griffié. I also thank my lab mates for multiple discussions and for testing SMIS and providing suggestions, in particular Jip Wulffele. The SMIS GUI was essentially developed during Covid confinement, but the virus is not acknowledged. This work was supported by the Agence Nationale de la Recherche (grants no. ANR-17-CE11-0047-01 and ANR-20-CE11-0013-01).

References

1. Lelek, M. *et al.* Single-molecule localization microscopy. *Nat Rev Methods Primers* **1**, 1–27 (2021).
2. Shen, H. *et al.* Single Particle Tracking: From Theory to Biophysical Applications. *Chem. Rev.* **117**, 7331–7376 (2017).
3. Sage, D. *et al.* Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software. *Nat Methods* **16**, 387–395 (2019).
4. Novák, T., Gajdos, T., Sinkó, J., Szabó, G. & Erdélyi, M. TestSTORM: Versatile simulator software for multimodal super-resolution localization fluorescence microscopy. *Sci Rep* **7**, 951 (2017).
5. Lindén, M., Ćurić, V., Boucharin, A., Fange, D. & Elf, J. Simulated single molecule microscopy with SMeagol. *Bioinformatics* **32**, 2394–2395 (2016).

6. Venkataramani, V., Herrmannsdörfer, F., Heilemann, M. & Kuner, T. SuReSim: simulating localization microscopy experiments from ground truth models. *Nat Methods* **13**, 319–321 (2016).

7. Lagardère, M., Chamma, I., Bouilhol, E., Nikolski, M. & Thoumine, O. FluoSim: simulator of single molecule dynamics for fluorescence live-cell and super-resolution imaging of membrane proteins. *Sci Rep* **10**, 19954 (2020).

8. Heydarian, H. *et al.* Template-free 2D particle fusion in localization microscopy. *Nat Methods* **15**, 781–784 (2018).

9. Griffié, J. *et al. Virtual-SMLM, a virtual environment for real-time interactive SMLM acquisition*. 2020.03.05.967893 https://www.biorxiv.org/content/10.1101/2020.03.05.967893v1 (2020).

10. Dempsey, W. P. *et al.* In vivo single-cell labeling by confined primed conversion. *Nat. Methods* **12**, 645–648 (2015).

11. Diekmann, R. *et al.* Optimizing imaging speed and excitation intensity for single-molecule localization microscopy. *Nat Methods* **17**, 909–912 (2020).

12. Thédié, D., Berardozzi, R., Adam, V. & Bourgeois, D. Photoswitching of Green mEos2 by Intense 561 nm Light Perturbs Efficient Green-to-Red Photoconversion in Localization Microscopy. *J. Phys. Chem. Lett.* **8**, 4424–4430 (2017).

13. Thevathasan, J. V. *et al.* Nuclear pores as versatile reference standards for quantitative superresolution microscopy. *Nat. Methods* **16**, 1045–1053 (2019).

14. Lee, S. H., Shin, J. Y., Lee, A. & Bustamante, C. Counting single photoactivatable fluorescent molecules by photoactivated localization microscopy (PALM). *Proc Natl Acad Sci U S A* **109**, 17436–41 (2012).

15. Mohr, M. A. *et al.* Rational Engineering of Photoconvertible Fluorescent Proteins for Dual-Color Fluorescence Nanoscopy Enabled by a Triplet-State Mechanism of Primed Conversion. *Angew. Chem. Int. Ed. Engl.* **56**, 11628–11633 (2017).

16. Virant, D., Turkowyd, B., Balinovic, A. & Endesfelder, U. Combining Primed Photoconversion and UV-Photoactivation for Aberration-Free, Live-Cell Compliant Multi-Color Single-Molecule Localization Microscopy Imaging. *Int J Mol Sci* **18**, E1524 (2017).

17. Turkowyd, B. *et al.* A General Mechanism of Photoconversion of Green-to-Red Fluorescent Proteins Based on Blue and Infrared Light Reduces Phototoxicity in Live-Cell Single-Molecule Imaging. *Angew. Chem. Int. Ed. Engl.* **56**, 11634–11639 (2017).

18. Berardozzi, R., Adam, V., Martins, A. & Bourgeois, D. Arginine 66 Controls Dark-State Formation in Green-to-Red Photoconvertible Fluorescent Proteins. *J. Am. Chem. Soc.* **138**, 558–565 (2016).

19. Gidi, Y. *et al.* Unifying Mechanism for Thiol-Induced Photoswitching and Photostability of Cyanine Dyes. *J. Am. Chem. Soc.* **142**, 12681–12689 (2020).

20. Kasper, R., Heilemann, M., Tinnefeld, P. & Sauer, M. Towards ultra-stable fluorescent dyes for single-molecule spectroscopy. in *Biophotonics 2007: Optics in Life Science* vol. 6633 405–416 (SPIE, 2007).

21. Baibakov, M. & Wenger, J. Laser-induced fluorescence quenching of red fluorescent dyes with green excitation: Avoiding artifacts in PIE-FRET and FCCS analysis. *Chemical Physics Letters* **706**, 669–674 (2018).

22. Stracy, M. *et al.* Transient non-specific DNA binding dominates the target search of bacterial DNA-binding proteins. *Molecular Cell* **81**, 1499-1514.e6 (2021).

23. Grimm, J. B. *et al.* Bright photoactivatable fluorophores for single-molecule imaging. *Nat. Methods* **13**, 985–988 (2016).

24. Zitter, E. D. *et al.* Mechanistic investigation of mEos4b reveals a strategy to reduce track interruptions in sptPALM. *Nat Methods* **16**, 707–710 (2019).

25. Hansen, A. S. *et al.* Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife* **7**, e33125 (2018).

26. Kuhn, T., Hettich, J., Davtyan, R. & Gebhardt, J. C. M. Single molecule tracking and analysis framework including theory-predicted parameter settings. *Sci Rep* **11**, 9465 (2021).

27. Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat Methods* **5**, 695–702 (2008).

28. Vrljic, M., Nishimura, S. Y., Brasselet, S., Moerner, W. E. & McConnell, H. M. Translational Diffusion of Individual Class II MHC Membrane Proteinsin Cells. *Biophysical Journal* **83**, 2681–2692 (2002).

29. Persson, F., Lindén, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat Methods* **10**, 265–269 (2013).

Methods

Fluorophore's definition and photophysical states

Fluorophores are defined in a SMIS database and can be created or modified by the user in an interactive manner. Four types of photophysical states can be specified: fluorescent states, photoactive dark states, non-photoactive dark states, and photo-bleached states. For each fluorescent state, excitation and fluorescence emission spectra are defined, generally from experimental data available for example in www.fpbase.org. Absorption spectra from photoactive dark states also need to be specified. If experimental spectra are not available, a tool is available in SMIS to generate tentative model spectra *in silico*. Extinction coefficients at reference wavelengths and fluorescence quantum yields of the fluorescent states are also defined. As is often the case for fluorescent proteins, a fluorescent state (typically the anionic form of a FP chromophore) can be specified to be in a pH-dependent rapid equilibrium with a dark state (typically the corresponding neutral form of the FP chromophore), according to given pKa and Hill coefficients.

Thermally- and light-induced transformations between all photophysical states are specified in the form of rates or phototransformation quantum yields, which can be interactively modified by the user.

Fluorophores can be set to freely tumble (representative of e.g. a live cell), or to adopt a defined orientation (representative of e.g. a frozen or fixed cell), in which case photon absorption and emission will depend on the chosen polarization of the lasers and microscope objective properties, respectively. Fluorophores can be specified to be fluorogenic (see below the section on virtual sample and fluorophore labelling). Their maturation level or labelling efficiency can be set, for example to account for the often non-fully efficient maturation of fluorescent proteins.

Photophysical transition matrix

Within SMIS, transition matrices for thermally-induced and light-induced transformations between photophysical states are defined for each fluorophore, and can be manipulated by the user in an interactive manner to modify the behavior of a fluorophore or to define new fluorophores. Thermally-induced transformations are defined with rates k_{ij}, whereas light induced transformations are defined with quantum yields q_{ij}.

At each step of the data collection (ie for each frame or inter-frame), the rate of photon absorption $k_i^{h\upsilon}$ by each photoactive state of the fluorophore and due to all defined lasers is defined as:

$$k_i^{h\upsilon} = \sum_j k_i^{h\upsilon_j}$$

Where $k_i^{h\upsilon_j}$ is the excitation rate by laser j at the position of the fluorophore, as defined below.

Let's consider a sampling time step Δt associated to a sampling rate $S = 1/\Delta t$, and with $\Delta t \ll T$ where T is the imaging frame time or the interframe time.

From the knowledge of the matrices $\{k_{ij}\}$ and $\{q_{ij}\}$, a global transition probability matrix $\{p_{ij}\}$ is then calculated at each step of the data collection in the following manner:

First, for a thermal process, the probability $p(k)$ of a transition at rate k to occur during Δt is given by $p(k) = k\Delta t = k/S$. Note that for a meaningful transition probability to be obtained, the property $S \gg k$ must be fulfilled, that is, a sufficiently high sampling frequency must be chosen (see below).

For a photo-induced process, let's consider $N = \Delta t/k^{h\nu}$, the number of photons absorbed by the fluorophore during Δt. As N can be very high, to release the requirements on the used sampling frequency S, we compute the probability of a transition occurring during Δt as:

$$\mathrm{p}(q) = q \times (1 + (1-q)^1 + (1-q)^2 + \ldots + (1-q)^{N-1}) = 1 - (1-q)^N$$

Note that if $N < 1$ (case of low photon absorption rate), the formula above still holds.

Finally for a transformation process which might be either thermally- or photo-induced, the transition probability is given by

$$\mathrm{p} = \mathrm{p}(q) + \mathrm{p}(k) - \mathrm{p}(k) \times \mathrm{p}(q)$$

From the knowledge of the matrix $\{p_{ij}\}$, the highest probability p_{Max} is determined and it is checked whether the criterium $p_{Max} < \propto$ is fulfilled where $\propto$ relates to the oversampling factor specified in SMIS (and set by default to a value of 10, ie $p_{Max} < 0.1$). If the criterium is not satisfied, the initially specified sampling frequency S is raised and the $\{p_{ij}\}$matrix recalculated until success.

Once the frame- or interframe-specific transition probability matrix is obtained, calculation of the fluorophore photophysical trace is performed as follows: based on random number generation, occurrences along time of transitions from states $\{s_i\}$ to states $\{s_j\}$ are recorded and the first transition from the starting photophysical state s_1, if any, is selected, that converts s_1 to s_2 at time t_1. Then, the first transition from s_2 occurring at $t > t_1$, if any, is selected, that converts s_2 to s_3 at time t_2. The process is repeated until the end of the frame- or interframe-period, or until photobleaching occurs.

Whenever a fluorophore is found in one or several fluorescence states $\{s_i^F\}$ for certain times τ_i during a frametime the number of photons it emits is calculated based on computed excitation rates k_l by all lasers (see below) that are on during the current frame, and the corresponding fluorescence quantum yields Q_i.

$$\mathrm{N} = \mathrm{P}\left(\sum_i \left\{Q_i \times \tau_i \times \sum_l k_l\right\}\right)$$

where P denotes the Poissonian operator.

Virtual sample and fluorophore labeling

A virtual sample need to be loaded for every fluorophore involved in the simulation. Virtual samples are typically segmented 2D images (in .tif format) or 3D stack of images (in Matlab format), in which different parts of the sample (called "subpatterns" in SMIS) are addressed by specific pixel values. This allows to distribute fractions of the fluorophore ensemble to the individual subpatterns. When fluorophores are defined to be fluorogenic, a level of fluorogenicity can be defined for each sub pattern, a useful feature to simulate e.g PAINT experiments with fluorogenic chromophores.

Computation of single molecule trajectories

Single fluorophores may diffuse in 2D or 3D throughout their associated virtual sample, according to specified diffusion coefficients and/or velocities (in the case of directed motion). Typically, a fluorophore diffuses in a confined manner within the subpattern it belongs to. However, fluorophores can be set to switch between several diffusion regimes within a specific subpattern, or to exchange between subpatterns according to a transition matrix defined by the user. This feature allows simulating complex diffusion scenarios, with fluorophores binding and unbinding from specific sample features at defined rates. Importantly, SMIS offers the option to calculate diffusion with sub-frame resolution, which provides motion-blurred fluorescence emission spots on the detector, proper temporal averaging as well as a more faithful description of the explored space, notably in the case of rapid motion.

The SMIS single-molecule diffusion modelling capabilities open the possibility to simulate complex sptPALM experiments, as well as ensemble techniques such as FRAP or pulse-chase experiments. PAINT experiments can also be simulated where fluorophores can transiently bind to their target to provide sparse single-molecule fluorescence emission.

Computation of excitation rates

The excitation rate of a rapidly tumbling molecule located at position (x, y) is given by: $k = \varepsilon P \lambda \frac{(10^{-6}) Ln(10)}{\mathcal{N} hc}$ where ε is the extinction coefficient of the considered photophysical state at the excitation wavelength λ, $P(x, y)$ is the laser power-density at (x, y), $\mathcal{N}$ is the Avogadro number, h is the Planck constant and c is the speed of light[1].

For a fixed oriented molecule with spherical coordinates θ and φ, the excitation rate of the absorbance dipole is angle dependent. For a circularly polarized laser beam whose electromagnetic field is assumed to be parallel to the objective focal plane, the excitation rate is given by $k_\theta = \frac{3}{2}\, k \cos^2 \theta$, where θ is measured from the focal plane. For a linearly polarized laser beam whose electromagnetic field is in the objective focal plane and aligned with the x axis, the excitation rate becomes $k_{\theta,\varphi} = 3\, k \cos^2 \theta \cos^2 \varphi$ (ref[1]).

Virtual microscope description

The experimental setup is simulated by a number of parameters describing laser illumination, microscope optical properties and detector characteristics.

Any number of laser beams with Gaussian or flat spatial profiles within the focal plane can be defined, with chosen wavelengths, powers, polarization (circular or linear) and FWHM (full width at half maximum). Complex laser sequence patterns can be generated during data collection, such as for example pre-bleaching steps, alternate excitation, and ramping-up or a Fermi profile for the activation laser.

The microscope is described by an objective of given numerical aperture NA, which is assumed to give rise to a Gaussian point spread function (PSF) with standard full width at half maximum (*FWHM = 1.22λ/NA* where λ is the emission wavelength). The photon collection efficiency takes into account the orientation of the emitting dipoles[2].

Transmission efficiencies of inserted emission filters (possibly multi-band), as well as overall transmission efficiency of the optical setup are also taken into account.

The EMCCD detector is described by the effective pixel size, dark current noise, readout noise and gain (counts /detected photon)[3].

SMLM data sets with a defined number of frames and frametime and inter-frametime can then be generated. For each single-molecule emission, the spatial (throughout the microscope PSF) and spectral (throughout the emission spectrum) distributions of emitted photons are calculated based on Poissonian statistics.

A defined level of Poissonian autofluorescence background, optionally proportional to the local illumination power density. Autofluorescence can be set to very spatially according to given input patterns, and temporally according to given decay rates.

The program outputs the stack of acquisition frames (.tif), diffraction limited images (.tif), laser illumination profiles (.tif), and for each activated molecule, its x and y coordinates, fluorescence emission trace, the number of detected photons and an estimate of the signal-to-noise ratio (.mat).

SMIS includes possibilities to simulate dual-channel detection, multi-color experiments (with N different dyes), sample drift, single-molecule Förster resonance energy transfer (FRET) between two fluorophores, and localized activation/bleaching similar to what can be normally achieved with a Fluorescence Recovery After Photobleaching (FRAP) module.

Widefield laser illumination can be defined as Gaussian, flat, and in 3D mode can be simulated in TIRF or HILO configurations.

SMIS performance

SMIS was developed under Matlab 2020a. The SMIS Graphical User Interface was developed with Matlab App Designer (Supplementary Fig. S23). Although SMIS is not meant to behave as a real-time simulation software[4], it performs reasonably fast (running time of ~1s/frame for a typical 3D SMLM experiment involving full photophysics of 3200 mEos4b molecules decorating 100 NPCs on a 2015 high-grade PC (HP Z-book 15 G2 equipped with a i7-4810MQ processor). Complex scenarios may take more time, but future developments including C++ code and parallel processing will improve the speed performance of SMIS.

As for any simulation software, the produced results merely reflect the input models, and accuracy needs to be traded to some extent for speed. SMIS mainly focuses on the impact of fluorophore photophysics, and thus, although it is able to simulate a complex virtual microscope environment, it is not designed to provide the most accurate optical response, e.g. it doesn't use for example vectorial Point Spread Functions (PSFs). Although SMIS is fully compatible with 3D imaging, only the astigmatism-based 3D localization method is currently implemented.

Software download and source code

The full code as well as a freely usable standalone version of SMIS will be made available on GitHub upon acceptance for publication of the present article.

Processing of SMIS data

SMIS data generated in Fig. 1 were localized, filtered and merged with Thunderstorm[5]. Image rendering used ImageJ (https://imagej.nih.gov/ij/).

SMIS data generated in Fig. 2 were localized, filtered and merged with Thunderstorm[5]. Coordinate-based colocalization and nearest neighbor measurements used the CBC tool of Thunderstorm.

SMIS data generated in Fig. 3 were processed with SMAP[6] using identical parameters as those employed in Diekmann et al[7].

SMIS data generated in Fig. 4 were processed with Trackit[8], using either the nearest neighbor track search option, or the more advanced U-Track-based option[9]. Exchange rates were extracted with vbSPT[10]. No blinking gap was allowed in the reconstituted tracks, as imposed by the 2 used evaluation procedures (histograms of diffusion coefficients based on mean jump distances (MJDs) and vbSPT). Fitting of the MJDs-based diffusion coefficient histograms exactly followed the procedures described in Stracy et al[11]. Retrieval of the fractional population's p1, p2 and p3 of tightly DNA-bound, weakly DNA-bound and freely diffusing molecules was based on the following: first, fitting of the MJD-based diffusion coefficient histograms or vbSPT results, assuming a 2-state model in both cases, provided the relative populations p1 and p23 of tightly DNA-bound and searching molecules, as well as their associated apparent diffusion coefficients Dapp,1 and Dapp,23. Those apparent diffusion coefficients were first corrected by subtracting the contribution of localization error ($\sigma^2/\Delta t$, where Δt is the frame time and σ is the median localization uncertainty over the considered dataset (~25 nm and ~12 nm for mEos3.2 and PA-JF549 data, respectively, estimated with Thunderstorm)). Then, unbiased diffusion coefficients D1 and D23 were extracted based on the calibration curves of Supplementary Fig. S13E. The fractional population p of weakly DNA-bound molecules relative to all searching molecules was given by:

$$D_{23} = p \times (D_1) + (1 - p) \times (D_{free})$$

where Dfree = 2.0 $\mu m^2/s$ was assumed to be extracted, as described in Stracy et al[11], from calibrated measurements in perturbed cells devoid of nucleoid where NAPs are freely diffusing. Finally $p_2 = p \times p_{23}$ and $p_3 = (1 - p) \times p_{23}$.

All figures were generated using ImageJ (https://imagej.nih.gov/ij/) or with homemade routines written in Matlab.

References

1. Avilov, S. *et al.* In cellulo evaluation of phototransformation quantum yields in fluorescent proteins used as markers for single-molecule localization microscopy. *PLoS ONE* **9**, e98362 (2014).
2. Fourkas, J. T. Rapid determination of the three-dimensional orientation of single molecules. *Opt Lett* **26**, 211–3 (2001).

3. Sage, D. *et al.* Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software. *Nat Methods* **16**, 387–395 (2019).

4. Griffié, J. *et al. Virtual-SMLM, a virtual environment for real-time interactive SMLM acquisition*. 2020.03.05.967893 https://www.biorxiv.org/content/10.1101/2020.03.05.967893v1 (2020).

5. Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G. M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389–2390 (2014).

6. Ries, J. SMAP: a modular super-resolution microscopy analysis platform for SMLM data. *Nat Methods* **17**, 870–872 (2020).

7. Diekmann, R. *et al.* Optimizing imaging speed and excitation intensity for single-molecule localization microscopy. *Nat Methods* **17**, 909–912 (2020).

8. Kuhn, T., Hettich, J., Davtyan, R. & Gebhardt, J. C. M. Single molecule tracking and analysis framework including theory-predicted parameter settings. *Sci Rep* **11**, 9465 (2021).

9. Jaqaman, K. *et al.* Robust single-particle tracking in live-cell time-lapse sequences. *Nat Methods* **5**, 695–702 (2008).

10. Persson, F., Lindén, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat Methods* **10**, 265–269 (2013).

11. Stracy, M. *et al.* Transient non-specific DNA binding dominates the target search of bacterial DNA-binding proteins. *Molecular Cell* **81**, 1499-1514.e6 (2021).