Analysis of binding site dependent labelling efficiency for DNA-PAINT using particle fusion

Heydarian, Hamidreza; Stallinga, Sjoerd; Rieger, Bernd

DOI
10.1016/j.optcom.2024.130834

Publication date
2024

Document Version
Final published version

Published in
Optics Communications

Optics Communications 569 (2024) 130834

Contents lists available at ScienceDirect

Optics Communications

journal homepage: www.elsevier.com/locate/optcom

Analysis of binding site dependent labelling efficiency for DNA-PAINT using particle fusion

Hamidreza Heydarian, Sjoerd Stallinga, Bernd Rieger *

Department of Imaging Physics, Lorentzweg 1, Delft, 2628 CJ, The Netherlands

ARTICLE INFO

MSC:
62H35
68U10

Keywords:
Single molecule localization microscopy
DNA-PAINT
Clustering
Labelling efficiency
Particle fusion

ABSTRACT

DNA-origami nanostructures have shown promising applications in single molecule localization microscopy. They have become a reference standard for benchmarking and for developing new techniques for nanoscopy. Here, we present a pipeline for quantifying the quality of these nano-structures when imaging multiple instances of them using DNA-PAINT technique. We show on several experimental datasets that these structures can have deformations and that the designed binding sites are not equally accessible for the labelled imager strands during the image acquisition process. These limitations result in non-uniform activation of the sites over the origami pattern when fusing the instances into a single reconstruction.

1. Introduction

Single molecule localization microscopy (SMLM) is the most widely used superresolution light microscopy method to study macromolecules, subcellular or synthetic structures and complexes at nanometre resolution. Among other techniques, DNA Points Accumulation In Nanoscale Topography (DNA-PAINT) has emerged as a powerful tool for optical nanoscopy with typically high photon count per emitting molecule, enabling sub−10 nm spatial resolution [1]. In DNA-PAINT, the needed stochastic blinking for sparse localization of individual molecules occurs via repeated transient binding and unbinding events of dye-labelled imager strands to the docking strands (binding sites) on the structure of interest. Therefore, a high degree of labelling (DOL) can only be achieved if all the available binding sites on the target structure are accessible and have equal binding affinity.

Previously, Strauss et al. [2] showed that strand incorporation is strongly dependent on the binding site positions in the structure when imaging DNA-origami nanostructures. To this end they used a template-based registration approach to align repeated measurements of the same nanostructure (“particles”) and a nearest neighbour search for the assignment of localizations to their corresponding binding sites. Their approach, however, suffers from two problems. First of all, template-based registration of individual particle is susceptible to the so-called template-bias problem [3]. This means that the assumed geometric features of the template structure are likely to appear in the final outcome of the registration. The implication is that the registration outcome does not provide independent evidence for the existence of such geometric features in reality. The use of a template-free registration would thus provide more certainty about the final result. Secondly, hard assignment of the localizations to their nearest neighbour binding sites is based on the assumption that the origami design corresponds to the physically realized nanostructure. However, this may not be a realistic assumption and it does not allow investigation of the possible deformation of the DNA-origami nanostructures.

Here, we present a workflow based on our previously developed template-free particle fusion algorithm [3] to investigate structural deformation of the particles with respect to the ground-truth model, quantitative assessment of the labelling efficiency of the used fluorophores and probing the accessibility of the binding sites on the target structure (Fig. 1).

2. Methods

2.1. Single molecule localization microscopy datasets

Our test dataset consists of several experimental DNA-origami nanostructures with different DOL and geometries imaged using DNA-PAINT. It includes nanostructures in the shape of the ‘TUD’ logo [4], a 3 × 4 grid pattern, and letters ‘O’ and ‘T’ [5]. The designs of the DNA-origami nanostructures are shown in Fig. 2a-d, and have a minimum lattice spacing of 5 nm. Sample preparation and SMLM imaging of these data were previously described in [3,6].

* Corresponding author.
E-mail address: b.rieger@tudelft.nl (B. Rieger).

https://doi.org/10.1016/j.optcom.2024.130834
Received 17 November 2022; Received in revised form 26 June 2024; Accepted 28 June 2024
Available online 6 July 2024

2.2. Data analysis

Our workflow starts with all-to-all registration [3] of all N segmented particles (Fig. 1). With the all-to-all registration, each individual particle is registered to all other particles by optimizing a distance measure which works directly on localization data and their corresponding uncertainties. This will provide us with N(N-1)/2 relative transformations (rotations and translations) which are subsequently averaged using Lie-algebra averaging. Lie-algebra averaging utilizes the redundancy in the number of required relative transformations to compute the absolute pose of each individual particle. With this the initial reconstruction is computed and this reconstruction then acts as a data-driven template which is used at the final step of the registration pipeline to register pre-aligned particles to and to obtain the final superparticle [3].

After particle fusion, we register the obtained superparticle to the DNA-origami design. We will make use of this alignment later when we want to assign each cluster of localizations (Gaussian mixture component) to its corresponding binding site. Then, we cluster each individual aligned particle using a weighted-data Gaussian mixture model clustering approach [7]. The advantage of this clustering algorithm over other existing methods is threefold. Firstly, it correctly models the image formation model in a SMLM acquisition by assuming that the localization distribution around each binding sites follows a multivariate normal distribution. Secondly, it has only one intuitive free parameter that needs to be set manually, namely the number of components of the GMM which is equivalent to the effective (DOL weighted) number of binding sites in the origami design. Lastly, it computes a saliency scores (weight) per mixture component which allows a mechanism for filtering of outlier Gaussian components that are often being formed due to false positive localizations. For all the available datasets in this work, we discard the outlier Gaussian components with less than 0.02 weight.

After clustering of each particle, the mixture component centres are assigned to their nearest neighbour binding sites in the origami design. This reveals the localization correspondence to the available sites. Once these correspondences are found, we count the number of events per site in the underlying structure, fit an anisotropic bivariate normal distribution using the MATLAB built-in function fitgmdist with the number of components $k = 1$ and used this to assess how uniform the localizations were distributed over the structure.

3. Results

We applied our algorithm on two 'TUD' datasets with 30% and 50% estimated labelling density with 549 and 456 particles, respectively. These two datasets were acquired separately with different concentrations of extended staples for DNA-PAINT to achieve different DOL [3]. Fig. 2e–f show the superparticles of the 'TUD' logo for the two datasets in which the so-called 'hotspot' problem, i.e. nonuniform distribution of the localizations over the available binding sites, is clearly visible. While the reported average localization uncertainty for both datasets is about 1 nm [3], the spread of the fitted Gaussian components around the binding sites for 50% DOL data is smaller than for 30% DOL. Furthermore, it seems that, for both datasets, the sites on the edge of the structure had a lower chance (three times in the worst case) of activation than those in the middle.

In another experiment, we fused 1233 DNA-origami nanostructures forming a 3 × 4 grid pattern with 20 nm raster spacing (Fig. 2b). Although this structure has almost the same dimensions as the 'TUD' logo, we observed a more uniform binding site activation on it (Fig. 2 g and l). In fact, this pattern has the most uniform distribution of localizations around the binding sites which can be related to better accessibility of the docking strands as these are well-separated in this case (Fig. 2o–s).

Lastly, we applied our algorithm to letters 'O' and 'T' from the letter dataset which was previously published by Huijben et al. [6]. Fig. 2c–d and Fig. 2h–i show the DNA-origami designs and their corresponding superparticles as a result of fusing 238 and 170 segmented particles, respectively. The two letters have both 24 binding sites which are packed in four dense clusters of six docking sites that are 5 nm apart. While the minimum site spacing (5 nm) for these two nanostructures is the same as for the 'TUD' logo, the individual binding sites are not resolvable in any of them.

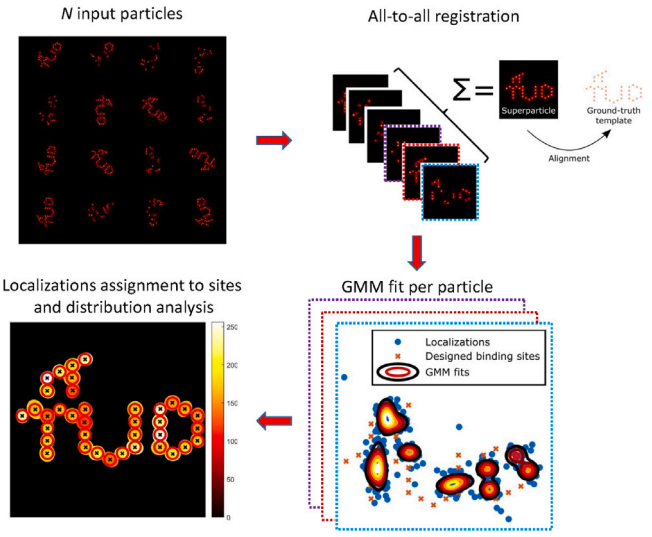

Fig. 1. Schematic workflow for binding site analysis of DNA-origami nanostructures. Input segmented particles are first aligned using the all-to-all registration method. Next, the resulting superparticle is aligned to the DNA-origami design pattern. This will bring all particles in line with the design. Each particle is then individually clustered using Gaussian mixture model clustering, followed by filtering of insignificant mixture components, outlier removal and the assignment of the mixture components to their nearest neighbour binding site in the DNA-origami pattern. Once the localization correspondence to the binding sites is found, a multivariate normal distribution is fitted to the distribution of localizations per site in order to find the mean position and the spread of localizations in the vicinity of that site.

4. Discussion

DNA-origami nanostructures have become a key tool for evaluating resolution performance of new methods and imaging techniques in SMLM. Example applications include drift correction [8], evaluating the resolution enhancement of new SMLM techniques [9] and providing an assay of the performance of fluorescent dye molecules [10]. Therefore, it is necessary to identify and characterize the practical limitations of these assemblies. In this study, we provided a workflow for quantification of localization events per defined binding site when imaging multiple instances of these nanostructures using DNA-PAINT.

Our analysis on several experimental DNA-origami structures with different binding site layouts revealed that the site activation is not always uniform and is highly dependent on the geometry of the structure. In particular, the sites on the edges of the structure had a lower chance of activation (Fig. 2e–f). This could be due to the flexibility of the origami which is maximum around its edges. This is in accordance with the problem of strand incorporation efficiency in a DNA-origami structure previously reported by Strauss et al. [2]. Additionally, the results indicate that the difference in the binding site intensities is physical due to the variable imager strand incorporation and not merely or exclusively a registration artefact.

In contrast, the analysis of the 3 × 4 grid nanostructure demonstrated a more uniform activation of the docking sites over the span of a similar-size DNA-origami like 'TUD' (Fig. 2o-s). In addition to the two-fold rotation and the reflection symmetry of the lattice, this, however, can be linked to the higher accessibility of the binding sites which

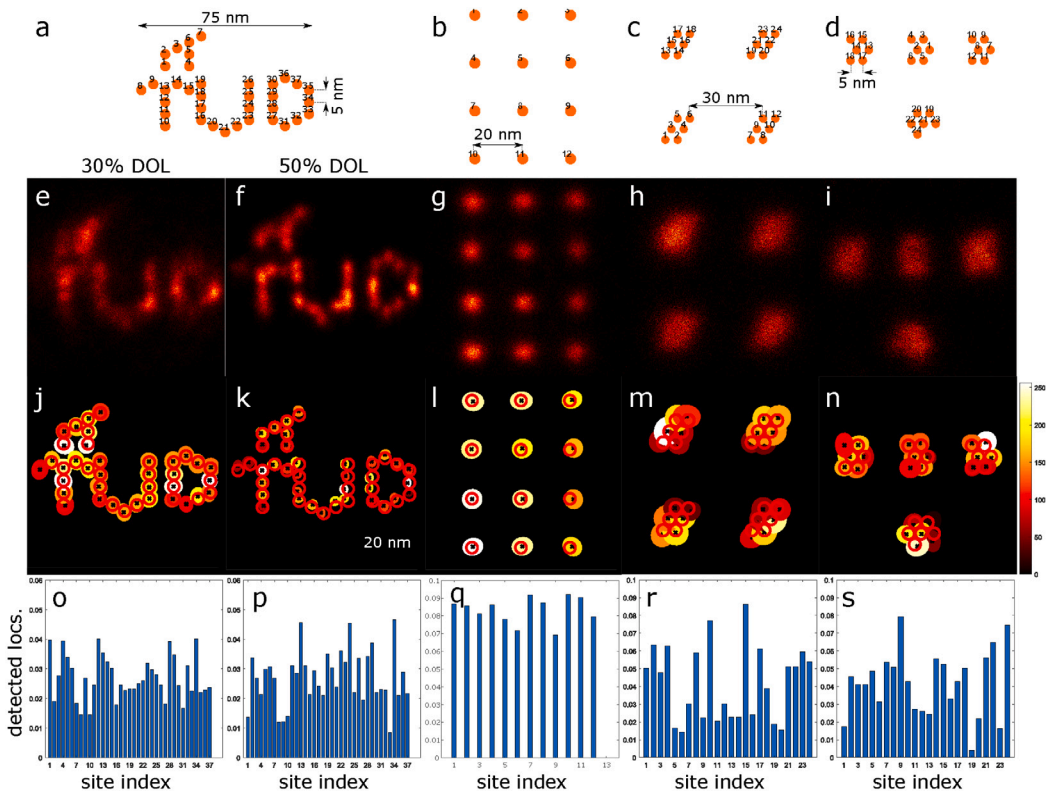

Fig. 2. Analysis of binding site dependent labelling efficiency for several DNA-origami nanostructures imaged using DNA-PAINT. (a–d) Schematics of the DNA-origami design patterns. Each cross represents a docking site in the design. (e–i) Superparticles as a result of fusing 549, 456, 1233, 238 and 170 'TUD' particles with 30% (e) and 50% DOL (f) and 3 × 4 grid structure, letter 'O' and letter 'T', respectively. (j–n) Cluster analysis and binding site activation maps. Each ellipse in these figures represents one standard deviation of the fitted Gaussians for each binding site. The black crosses show the estimated centre of these clusters which are quite often away from the positions in the origami design (centre of the red circles). Furthermore, the number of times each binding site is activated is colour-coded to show how often each site is activated. (o–s) The number of localizations per binding sites normalized to sum to unity. The analysis shows non-uniform activation of the binding sites in addition to deviations of the cluster centres (red circles) from the ground-truth binding sites (black crosses). Among others, the grid structure which has the largest binding site spacing, demonstrates the most uniform distribution of localizations per binding site (n).

are very well separated (20 nm) in this case and also to an increased robustness of the grid structure when compared to the 'TUD' logo.

Although the minimum binding site distances for the letter dataset is 5 nm and equal to the 'TUD' logo, these sites were not visible in the final reconstruction. The main difference of the 'O' and 'T' assemblies in this dataset and the 'TUD' logo is the sparse occupation of the DNA-origami area compared to the more continuous geometry of the 'TUD'. This sparsity has probably made the complex more flexible which subsequently resulted in structural deformation. In [6], we observed a similar behaviour when analysing structural heterogeneity of 3D tetrahedron DNA-origami nanostructures, which are is also large and sparse DNA-origami structures.

The impact of the quality of registration on the analysis could potentially be investigated by computing the distribution of the Bhattacharya metric over all particle pairs after the all to all registration. The average or median value of this alignment metric, normalized by the numbers of localization events in the particle pair, could serve as (relative) quality measure for alignment success.

In conclusion, our computational approach demonstrates that DNA-origami nanostructures undergo deformations which depend on the arrangement of the designed docking sites on the scaffold. Furthermore, our approach shows that even in DNA-PAINT, with high acquisition time, the distribution of the imager strands over the docking strands is not uniform and depends on the geometry, some docking sites have a higher binding affinity than others. Our approach identifies a potential source of uncertainty in DNA-origami test samples and indicates a limitation of these structures for sub-nanometre studies.

Funding

This work was supported by the European Research Council Nano@cryo, grant no. 648580 to H.H. and B.R.

CRediT authorship contribution statement

Hamidreza Heydarian: Investigation, Methodology, Software. **Sjoerd Stallinga:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing. **Bernd Rieger:** Conceptualization, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Bernd Rieger reports financial support was provided by European Research Council. Hamidreza Heydarian reports financial support was provided by European Research Council.

Data availability

The 'TUD' logo dataset can be found in the 'Single-Molecule Localization Microscopy (SMLM) 2D TU Delft logos' [4]. The 'grid', letter 'O' and 'T' datasets for this study can be found in the 'Single-Molecule Localization Microscopy (SMLM) 2D Digits 123 and TOL letters datasets' [5].

Acknowledgements

We thank Teun Huijben for providing particle averaging results for the letter dataset.

References

[1] R. Jungmann, M. Avendaño, J. Woehrstein, M. Dai, W. Shih, P. Yin, Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and exchange-PAINT, Nature Methods 11 (2014) 313–318.

[2] M. Strauss, F. Schueder, D. Haas, P. Nickels, R. Jungmann, Quantifying absolute addressability in DNA origami with molecular resolution, Nature Commun. 9 (2018) 1600.

[3] H. Heydarian, F. Schueder, M. Strauss, B.v. Werkhoven, M. Fazel, K. Lidke, R. Jungmann, S. Stallinga, B. Rieger, Template-free 2D particle fusion in localization microscopy, Nature Methods 15 (2018) 781–784.

[4] H. Heydarian, F. Schueder, M. Strauss, B.v. Werkhoven, M. Fazel, K. Lidke, R. Jungmann, S. Stallinga, B. Rieger, Data from: Single-molecule localization microscopy (SMLM) 2D TUD logos, 2018, 4TU.ResearchData repository, URL https://doi.org/10.4121/uuid:0d42a28f-f625-41a3-ba77-25e397685466.

[5] T. Huijben, H. Heydarian, A. Auer, F. Schueder, R. Jungmann, S. Stallinga, B. Rieger, Data from: Single-molecule localization microscopy (SMLM) 2D digits 123 and TOL letters datasets, 2021, 4TU.ResearchData repository, URL https://doi.org/10.4121/14074091.v1.

[6] T. Huijben, H. Heydarian, A. Auer, F. Schueder, R. Jungmann, S. Stallinga, B. Rieger, Detecting structural heterogeneity in single-molecule localization microscopy data particle averaging, Nature Commun. 12 (2021) 3791.

[7] I.D. Gebru, X. Alameda-Pineda, F. Forbes, R. Horaud, EM algorithms for weighted-data clustering with application to audio-visual scene analysis, IEEE Trans. Pattern Anal. Mach. Intell. 38 (2016) 2402–2415.

[8] M. Dai, R. Jungmann, P. Yin, Optical imaging of individual biomolecules in densely packed clusters, Nature Nanotechnol. 11 (2016) 798–807.

[9] J. Cnossen, T. Hinsdale, R. Thorsen, M. Siemons, F. Schueder, R. Jungmann, C. Smith, B. Rieger, S. Stallinga, Localization microscopy at doubled precision with patterned illumination, Nature Methods 17 (2020) 59–63.

[10] F. Schueder, M. Strauss, D. Hoerl, J. Schnitzbauer, T. Schlichthaerle, S. Strauss, P. Yin, H. Harz, H. Leonhardt, R. Jungmann, Universal super-resolution multiplexing by DNA exchange, Angew. Chem. Int. Ed. 56 (2017) 4052–4055.