Received: 26 June 2023 | Revised: 19 September 2023 | Accepted: 7 October 2023

DOI: 10.1111/jmi.13236

Journal of Microscopy RMS

THEMED ISSUE ARTICLE

Modelling 3D supramolecular structure from sparse single-molecule localisation microscopy data

Alistair Curd[1] | **Alexa Cleasby**[1] | **Michelle Baird**[2] | **Michelle Peckham**[1]

[1]Faculty of Biological Sciences, Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds, UK

[2]Cell and Developmental Biology Centre, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland

Correspondence
Alistair Curd, Faculty of Biological Sciences, Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, UK.
Email: a.curd@leeds.ac.uk

Funding information
UK Medical Research, Grant/Award Number: MR/K015613/1; UK Biotechnology and Biological Sciences Research Councils, Grant/Award Number: BB/S015787/1; Wellcome Trust, Grant/Award Number: 091108/Z/10/Z; Intramural Research Program of the National Heart, Lung and Blood Institute; US National Institutes of Health

Abstract

Single-molecule localisation microscopy (SMLM) has the potential to reveal the underlying organisation of specific molecules within supramolecular complexes and their conformations, which is not possible with conventional microscope resolution. However, the detection efficiency for fluorescent molecules in cells can be limited in SMLM, even to below 1% in thick and dense samples. Segmentation of individual complexes can also be challenging. To overcome these problems, we have developed a software package termed PERPL: Pattern Extraction from Relative Positions of Localisations. This software assesses the relative likelihoods of models for underlying patterns behind incomplete SMLM data, based on the relative positions of pairs of localisations. We review its principles and demonstrate its use on the 3D lattice of Z-disk proteins in mammalian cardiomyocytes. We find known and novel features at ~20 nm with localisations of less than 1% of the target proteins, using mEos fluorescent protein constructs.

KEYWORDS

relative positions, single-molecule localisation microscopy, sparse data analysis, structural modelling, superresolution microscopy, symmetry, Z-disk

1 | INTRODUCTION

Proteins are often found in subcellular arrangements much smaller than the resolution of a conventional fluorescence microscope (~250 nm). Single molecule localisation microscopy (SMLM) can localise individual proteins with a precision of 1–10 nm, and therefore has the potential to reveal a wealth of new information about ordered supramolecular complexes. However, in the dense, 3D, disordered environment of many protein complexes, background and stochastic noise prevent accurate, high-precision localisation of many of the labelled proteins, even to below 1%. In this situation, most spatial analysis approaches to determine their organisation can be rendered ineffective. Segmentation of instances of a complex, a typical requirement of these approaches, is also often challenging.

To overcome these problems, we have developed a software package termed PERPL: Pattern Extraction from Relative Positions of Localisations.[1-3] PERPL combines the relative positions of pairs of localisations of specific molecules from SMLM images of multiple complexes into a 3D relative position distribution (RPD). This allows us to identify characteristic underlying features of a complex in

3D from as few as two localisations per complex, in principle. To aid in the interpretation of the experimental RPD, it is possible to generate a model RPD for a candidate structure/underlying molecular organisation, and then fit the RPD from a range of such models to the experimental RPD. The relative likelihoods of the candidate models to be the true molecular organisation can be calculated, providing a quantitative level of confidence in selecting the most likely model.[1,2,4] PERPL can also be used to analyse distributions containing more than one target molecule.[5]

Practitioner Points

- In single-molecule localisation microscopy, limited detections of fluorescently labelled molecules can make the determination of cellular structures very challenging.
- Structural information is contained in the distribution of the relative positions of pairs of molecules; these relative positions can be aggregated over many instances of a cellular complex, even if few molecular localisations are obtained in a single complex.
- We provide software to aggregate these relative positions, fit structural models to their distribution, and select the most likely structure for the underlying molecular arrangement, even when less than 1% of molecules in a complex are detected.

2 | BACKGROUND: LOCALISATION EFFICIENCY

It is often difficult to estimate detection efficiency for single molecules in a sample, which requires prior knowledge of their organisation. However, in some cases, such as imaging of Nup107 in nuclear pores and α-actinin-2 in Z-disks as used here, we do have this knowledge. Before filtering for the most precise localisations, in a well-optimised sample that lends itself to SMLM, detection efficiencies of 50%−70% can be obtained.[6] However, dense structures, especially extending in Z, and high background seriously impact the ability of localisation algorithms to detect individual blinks and therefore reduce detection efficiency.[7]

To obtain information about a molecular arrangement to a given precision, it may also be useful to select only a subset of localisations for spatial analysis that have each have a correspondingly precise estimate of their position (e.g. localisation precision < desired precision on distances $\div \sqrt{2}$). This may be necessary to resolve structural features, but further reduces the effective detection efficiency for the molecules that can be included in analysis. Sufficient data may still be acquired for modelling the structure with PERPL by acquiring data from more instances of the complex (e.g. more fields of view), since PERPL can use as little as one relative position, or a pair of localisations, per complex over many complexes.

3 | ILLUSTRATION: NUP107 LOCALISATIONS IN THE NUCLEAR PORE COMPLEX

Many of the proteins within the nuclear pore complex (NPC) have an eightfold radially symmetric arrangement, which is repeated in two layers, including Nup107.[8] SMLM can achieve detection efficiencies of 50%−70% for NPC proteins,[6] making the NPC a relatively simple test case for a structural SMLM technique (Figure **1A**).

Here, we use 3D localisations of Nup107, which gave a detection efficiency of ~50% in a 2D dSTORM dataset with an identical labelling strategy.[6] Our localisation precision filter ($\sigma Z < 10$ nm) included 31% of the localisations in the 3D Nup107 dataset.[2] Assuming some reduction in the detection efficiency compared with the previous 2D data, owing to the larger astigmatic PSF, we estimate that the detection efficiency in the Nup107 data used for analysis was ~10%.

An *XY*-projection of 3D Nup107-SNAP-Alexa Fluor 647 localisations from dSTORM (direct stochastic optical reconstruction microscopy)[2] displays broken ring-like features (**i**), while an *XZ*-projection appears to contain two less well-defined layers of localisations in Z. From the 3D RPD, distances in *XY* are used to analyse the ring-like features, independent of their orientation in the plane (**ii**), and distances in Z are analysed separately. In *XY*, PERPL generates a parametric distance distribution for rotationally symmetric candidate structures (**iii**, **iv**). This model includes the degree of radial symmetry; terms for repeated localisations of a single fluorescent molecule, clustering at vertices of the structure and intervertex distances; and a background contribution. The corrected Akaike Information Criterion (*AICc*) estimates the relative likelihoods of the different *XY*-distance models from their fits to the experimental RPD. An eightfold symmetry (**v**) is found to be four times more likely than the next most likely model (ninefold), with the relative likelihood of other symmetry orders being negligible. Relative positions in Z are very well fit by a two layer model (**v**), and dimensions in *XY* and Z match previous electron microscopy (EM) data[2,8] (**v**).

The analysis can be refined further, so that only within-layer relative positions ($\Delta Z < 20$ nm) are included in the *XY* analysis. The relative likelihood for the eightfold model

Modelling 3D supramolecular structure from sparse SMLM data

Alistair Curd[1,*], Alexa Cleasby[1], Michelle Baird[2], Michelle Peckham[1]

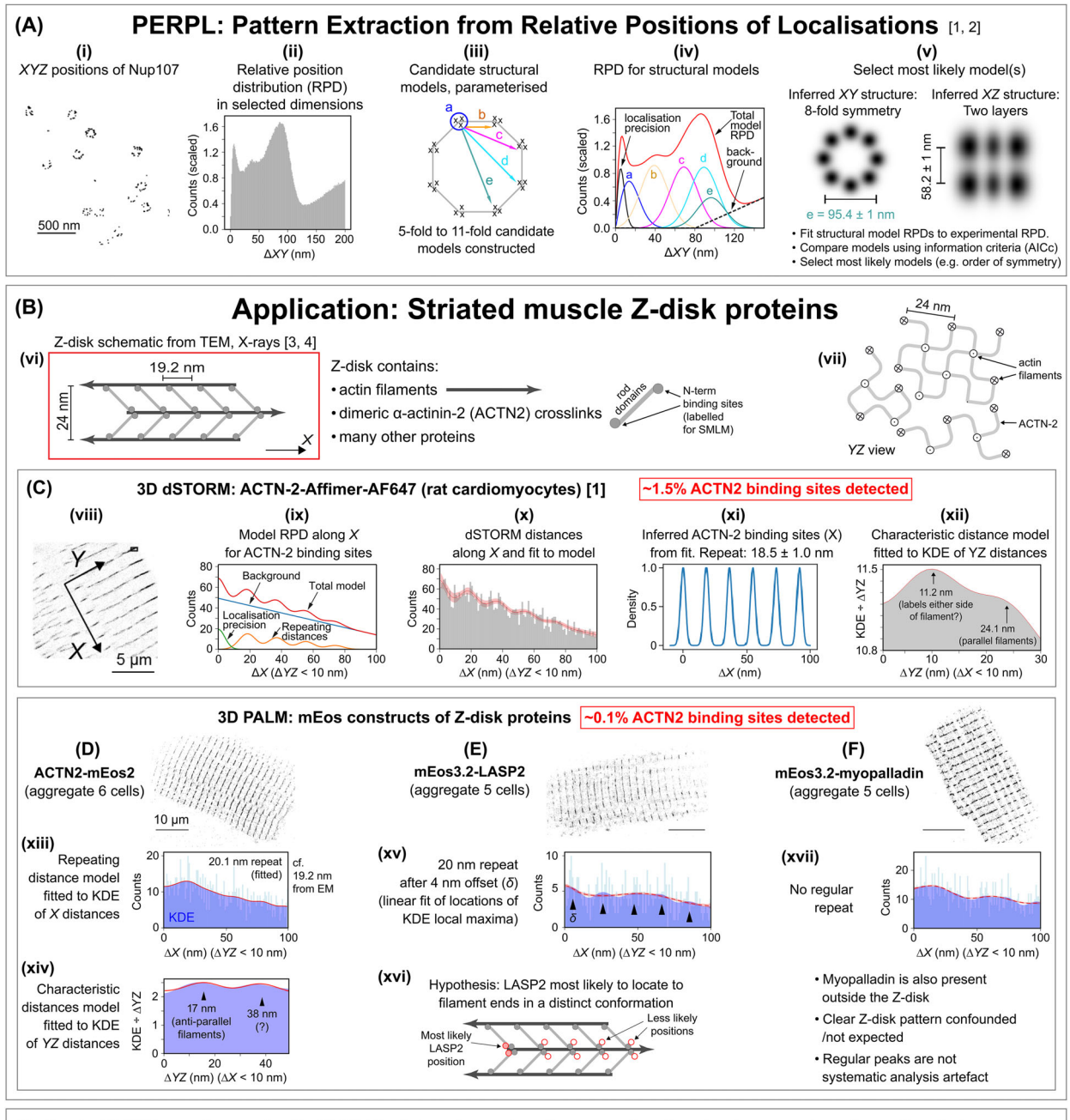
[1] Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK

[2] Cell and Developmental Biology Centre, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

*email: a.curd@leeds.ac.uk

Problem: Proteins are found in ordered complexes smaller than conventional microscope resolution.
SMLM can localise proteins precisely enough to observe supramolecular structures.
But detection efficiency in SMLM can be ~1% of proteins or even less, especially for thick and dense samples.

Solution: Aggregate pairwise relative positions over many instances of a complex and analyse patterns.



Outlook: PERPL can be used with SMLM to determine ordered supramolecular structures and conformational changes.
Overcomes difficulties when detection efficiency is low and segmentation for particle averaging is challenging.

[1] Curd et al., *Nano Letters* 21, 1213-20 (2021), [2] github.com/AlistairCurd/PERPL-Python3,
[3] Burgoyne et al., *J Mol Biol* 427, 3527-37 (2015), [4] Perz-Edwards et al., *Biophys J* 101, 709-17 (2011)

FIGURE 1 Modelling 3D supramolecular structure from sparse SMLM data. Principles (A) and applications (B-F) of PERPL analysis to overcome the challenge of low detection efficiency in SMLM experiments on molecular arrangements.

is then more than 100 times greater than for all other symmetry orders. This improvement in structural clarity arises from elimination of the effect of the small relative twist of the two layers of the NPC.[2,8]

4 | APPLICATION: STRIATED MUSCLE Z-DISK PROTEINS

4.1 | Z-disk structure

The Z-disk of striated muscle cells (Figure 1B) connects and anchors their sarcomeres by crosslinking actin filaments of opposite polarity that extend into adjacent sarcomeres (**vi**). The main component of the crosslinks is an antiparallel dimer of α-actinin-2 (ACTN2) which binds at 19.2 nm intervals along actin filaments, resulting in a tetragonal lattice structure for the Z-disk[9] (**vi**, **vii**). Many other proteins are present in the Z-disk and their arrangement and functions remain to be understood. Therefore, the arrangement of ACTN2 in the Z-disk is a more challenging test case for PERPL, and PERPL may help to determine the arrangement of other proteins in new experiments.

4.2 | 3D dSTORM data on α-actinin-2 (ACTN2)

We acquired 3D dSTORM localisations of ACTN2 in rat cardiomyocytes, labelled at the actin-binding N-terminus with an Affimer conjugated with Alexa Fluor 647[2] (Figure 1C, **viii**). The thick, dense structure of the Z-disks results in high background in SMLM, reducing detection efficiency. Together with incomplete labelling, only ~1.5% of ACTN2 molecules were detected, after filtering for localisation precision < 5 nm in XY for analysis of the nanoscale patterns.

To analyse the data, we defined the axis of the roughly cylindrical cell as X (**viii**) and filtered the 3D RPD by ΔYZ for relative positions likely to be along one actin filament ($\Delta YZ < 10$ nm) (**ix**, **x**). The distribution of X-distances appeared noisy around a linearly decreasing trend (**x**), but a kernel density estimate (KDE) indicated a regular repeat distance for local maxima. A model RPD including repeating distances in X (**ix**) fit well to the ΔX distribution (**x**) and was more than 20 times more likely to be correct than a model with no regular repeat. The repeating distance obtained for ACTN2 binding sites is 18.5 ± 1.0 nm, compared with 19.2 nm from EM[2,9] (**vi**, **xi**).

In the YZ-direction, radial in the cell (**vii**, **xii**), we used a model for short characteristic distances and high background (at longer distances, regularity is increasingly disrupted by discontinuities between rotated lattice domains, **vii**). The ΔYZ data had a lower signal-to-noise ratio, mitigated by fitting to a KDE. We also divided the KDE by ΔYZ to avoid fit results being dominated by a parameter for a monotonous increase in density.[2] A characteristic distance with a mean of 24 nm was obtained, corresponding with the distance between parallel actin filaments,[9] and another at 11 nm, which we interpret to be the YZ-distance between a pair of labels on ACTN2 binding domains on either side of the same actin filament (**xii**).

4.3 | 3D PALM data on three Z-disk proteins

We tested PERPL on even more challenging data and began new experiments on Z-disk proteins by acquiring 3D PALM (photoactivated localisation microscopy) data on mEos-fusion proteins expressed in the cardiomyocytes, using an adenovirus-based approach (Figure 1D–F). In this case, labelling density is lower than that achieved with Affimers. Moreover, mEos emits fewer photons on average than Alexa Fluor 647 in SMLM, resulting in less precise localisations. This reduced the detection efficiency for the ACTN2 N-terminus binding sites to ~0.1%, after filtering for localisation precision < 5 nm in XY. We obtained 1.3×10^5 localisations of ACTN2-mEos2 over 93 Z-lines across 6 cells, or ~1400 localisations per Z-line, defined as the visible band of Z-disks from myofibrils across the width of the cell (Figure 1D). Data for a Z-line extended over ~100 nm (Z-disk width) × ~15 μm (cell width) × ~2 μm (acquisition depth), with the Z-disks comprising unit cells ~20 × 17 × 17 nm, each containing two ACTN2 molecules.[9]

$$\text{Detection efficiency} = \frac{\#\text{ localisations acquired and used in analysis}}{\#\ Z\text{-disk lattice unit cells imaged} \times 2} \approx 0.001.$$

However, as long as sufficient numbers of characteristic pairwise relative positions can be obtained, even if detected pairs of target molecules are very sparsely distributed in a sample, underlying structural features should be observable in the aggregated experimental RPD.

To increase the number of relevant relative positions beyond those obtainable from a single field of view, PERPL allows these relative positions to be combined over multiple fields of view from the same sample type (Figure 1D–F). In the case of cardiomyocytes, the long axis of the cell can be designated X for every cell (**viii**), allowing transformation of the relative positions into a common coordinate system between cells. Aggregation of the relative positions into a combined 3D RPD is then straightforward and improves signal arising from structural features compared with counting noise. Use of KDEs can also mitigate the effect of noise on model fitting results (**xii**, **xiii**, **xiv**).

Testing this analysis on ACTN2-mEos2 localisations, we aggregated a 3D RPD from six cells (Figure 1D). A repeating distance model for ACTN2 binding sites on actin fitted the ΔX KDE well (**xiii**), although the ratios of the peaks corresponding to the repeated distance needed to become independent in this case. In the dSTORM data, with higher signal-to-noise ratio, the most likely model contained regular peaks with specific ratios expected from a fixed number of binding sites along a filament. The ACTN2 repeat from PALM was 20.1 nm, which is close to the EM and dSTORM results (**vi**, **xi**, **xiii**)

Characteristic distances in *YZ* were fitted at 17 and 38 nm (**xiv**). The 17-nm peak may be connected with the distance between antiparallel actin filaments, but the source of the 38-nm peak is not clear.

We additionally investigated the unknown arrangement of LASP2 (LIM and SH3 domain protein-2), using 3D PALM data obtained from mEos3.2-LASP2, aggregating a 3D RPD from five cells (Figure 1E). Our model RPDs did not fit the data well, but the KDE of ΔX had strikingly regular peaks, at intervals of 20.2 ± 0.8 nm after an initial offset of 4 ± 2 nm (linear regression of KDE local maximum positions against peak number from 0 to 4) (**xv**). This repeat is again similar to the 19.2, 18.5 and 20.1 nm repeats of ACTN2 from EM,[9] dSTORM[2] and PALM (this work), respectively (**vi**, **xi**, **xiii**).

This result leads us to propose a new arrangement for LASP2 in the Z-disk (**xvi**). LASP2 is known to bind to actin and ACTN2.[10] Moreover, the ends of the actin filaments, which are embedded within the Z-disk, are unique sites for some Z-disk proteins[11] and, therefore, are potentially sites for distinct roles in complexes for other Z-disk proteins. Therefore we propose that LASP2 is most likely to locate to the ACTN2 binding site at one edge of the Z-disk (agreeing with previous imaging[12]), but also binds at other ACTN2 binding sites with a lower relative frequency, with a distinct complex and role at the ends of actin filaments (**xvi**).

Lastly, we aggregated a 3D RPD from PALM data on mEos3.2-myopalladin in 5 cells (Figure 1F). Myopalladin also binds with actin and ACTN2, but has additional roles outside the Z-disk, including communication with the nucleus.[13] Therefore a regular pattern within the Z-disk may be confounded by the myopalladin distribution elsewhere in the cell in our analysis. We did not find the same regularity of peaks in the KDE as for ACTN2 and LASP2 (**xvii**), which may be a result of the combination of different cellular distributions of myopalladin in our analysis, and which demonstrates that the regularity of peaks in the KDE of ΔX is not a systematic artefact of our method. Segmentation of the Z-disk localisations[14] may be possible to clarify the data and may also lead to extraction of similar Z-disk lattice features as found for ACTN2 and LASP2.

With information obtained for the arrangement of single Z-disk proteins, we can go on to explore their positions relative to one another, using the same software[3,5] and through further development of model RPDs.

5 | OUTLOOK

PERPL (Pattern Extraction from Relative Positions of Localisations) analysis and software[1–3] can be used to determine structural features of supramolecular complexes from SMLM data. If suitable structural models can be constructed, it deals with issues of low detection efficiency and the difficulty of segmenting complexes for particle averaging in biological SMLM experiments. It can also be used to explore combined distributions of more than one target molecule.[5] PERPL may also be used to analyse image data when protein distributions can be converted to a set of coordinates, for example, the centres of mass for protein clusters.[5]

To investigate complexes or biological systems with candidate structural models beyond those already implemented (radial symmetry, linear repeats, one or two independent characteristic distances),[2,5] users need to derive new models for relative position distributions. These models (new Python functions) can use functions in the library *modelling_general.py* as building blocks and must then be fit in a new function that calls the model fitting function and uncertainty estimation functions from *modelling_general.py*.

In the models used to date, PERPL assumes that localisations are missing at random from molecular positions. However, it may also be useful to consider that there may be preferential sites in a complex for labelling, or for incomplete labelling. This would suggest alternative hypothetical relative position distributions for an underlying structural model.

Another recently developed approach to structural modelling from localisation distributions is LocMoFit,[15] which fits structural models to incomplete localisation distributions directly, rather than to relative position distributions. LocMoFit therefore allows more intuitive and simpler model construction and structural inference but requires the segmentation of individual complexes and fits each segmented complex. In contrast, PERPL does not require a segmentation step and avoids a bias in estimates of structural features at lower detection efficiencies.

Finally, PERPL can provide useful measurements of variability. On the one hand, it can give information on the variability of the structure of a supramolecular complex. On the other, it can be used to assess the performance of SMLM data acquisition and processing methods, given a test complex.[4]

ACKNOWLEDGEMENTS

The work was funded by the UK Medical Research and the UK Biotechnology and Biological Sciences Research Councils (MR/K015613/1 and BB/S015787/1, to M.P. and A. Curd); the Wellcome Trust (Institutional Strategic Support Fund at University of Leeds to A. Cleasby and travel award to M.P., 091108/Z/10/Z); and the Intramural Research Program of the National Heart, Lung and Blood Institute, US National Institutes of Health. The SMLM system was funded by alumnus M. Beverley, in support of the University of Leeds 'making a world of difference' campaign. Isolated cardiomyocytes were a kind gift from the Steele Group, University of Leeds. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

ORCID

Alistair Curd https://orcid.org/0000-0002-3949-7523
Michelle Peckham https://orcid.org/0000-0002-3754-2028

REFERENCES

1. Curd, A. (2021). 3D supramolecular structure from single-molecule localisation—Imaging ONEWORLD. https://www.youtube.com/watch?v=uB1aU8QC_Z4 (Accessed: 21/06/2023)
2. Curd, A. P., Leng, J., Hughes, R. E., Cleasby, A. J., Rogers, B., Trinh, C. H., Baird, M. A., Takagi, Y., Tiede, C., Sieben, C., Manley, S., Schlichthaerle, T., Jungmann, R., Ries, J., Shroff, H., & Peckham, M. (2021). Nanoscale pattern extraction from relative positions of sparse 3D localizations. *Nano Letters*, *21*, 1213–1220.
3. Curd, A. P. (2023). *AlistairCurd/PERPL-Python3: Two colour PERPL, updated to Python 3.11* (v0.2.1). https://doi.org/10.5281/zenodo.8359800
4. Prakash, K., & Curd, A. P. (2023). Assessment of 3D MINFLUX data for quantitative structural biology in cells. *Nature Methods*, *20*, 48–51.
5. Chuntharpursat-Bon, E., Povstyan, O. V., Ludlow, M. J., Carrier, D. J., Debant, M., Shi, J., Gaunt, H. J., Bauer, C. C., Curd, A., Simon Futers, T., Baxter, P. D., Peckham, M., Muench, S. P., Adamson, A., Humphreys, N., Tumova, S., Bon, R. S., Cubbon, R., Lichtenstein, L., & Beech, D. J. (2023). PIEZO1 and PECAM1 interact at cell-cell junctions and partner in endothelial force sensing. *Communications Biology*, *6*, 358.
6. Thevathasan, J. V., Kahnwald, M., Cieśliński, K., Hoess, P., Peneti, S. K., Reitberger, M., Heid, D., Kasuba, K. C., Hoerner, S. J., Li, Y., Wu, Y.-L., Mund, M., Matti, U., Pereira, P. M., Henriques, R., Nijmeijer, B., Kueblbeck, M., Sabinina, V. J., Ellenberg, J., & Ries, J. (2019). Nuclear pores as versatile reference standards for quantitative superresolution microscopy. *Nature Methods*, *16*, 1045–1053.
7. Baddeley, D., & Bewersdorf, J. (2018). Biological insight from super-resolution microscopy: What we can learn from localization-based images. *Annual Review of Biochemistry*, *87*, 965–989.
8. Von Appen, A., Kosinski, J., Sparks, L., Ori, A., Diguilio, A. L., Vollmer, B., Mackmull, M.-T., Banterle, N., Parca, L., Kastritis, P., Buczak, K., Mosalaganti, S., Hagen, W., Andres-Pons, A., Lemke, E. A., Bork, P., Antonin, W., Glavy, J. S., Bui, K. H., & Beck, M. (2015). In situ structural analysis of the human nuclear pore complex. *Nature*, *526*, 140.
9. Burgoyne, T., Morris, E. P., & Luther, P. K. (2015). Three-dimensional structure of vertebrate muscle Z-band: The small-square lattice Z-band in rat cardiac muscle. *Journal of Molecular Biology*, *427*, 3527–3537.
10. Zieseniss, A., Terasaki, A. G., & Gregorio, C. C. (2008). Lasp-2 expression, localization, and ligand interactions: A new Z-disc scaffolding protein. *Cell Motility and the Cytoskeleton*, *65*, 59–72.
11. Bang, M.-L., & Chen, J. (2015). Roles of nebulin family members in the heart. *Circulation Journal*, *79*, 2081–2087.
12. Fernandes, I., & Schöck, F. (2014). The nebulin repeat protein Lasp regulates I-band architecture and filament spacing in myofibrils. *Journal of Cell Biology*, *206*, 559–572.
13. Bang, M.-L., Mudry, R. E., Mcelhinny, A. S., Trombitás, K., Geach, A. J., Yamasaki, R., Sorimachi, H., Granzier, H., Gregorio, C. C., & Labeit, S. (2001). Myopalladin, a novel 145-Kilodalton sarcomeric protein with multiple roles in Z-disc and I-band protein assemblies. *Journal of Cell Biology*, *153*, 413–428.
14. Varga, D., Szikora, S., Novák, T., Pap, G., Lékó, G., Mihály, J., & Erdélyi, M. (2023). Machine learning framework to segment sarcomeric structures in SMLM data. *Scientific Reports*, *13*, 1582.
15. Wu, Y.-L., Hoess, P., Tschanz, A., Matti, U., Mund, M., & Ries, J. (2023). Maximum-likelihood model fitting for quantitative analysis of SMLM data. *Nature Methods*, *20*, 139–148.

How to cite this article: Curd, A., Cleasby, A., Baird, M., & Peckham, M. (2023). Modelling 3D supramolecular structure from sparse single-molecule localisation microscopy data. *Journal of Microscopy*, 1–6. https://doi.org/10.1111/jmi.13236