Superresolution microscopy: simulating and analyzing spatial point patterns of receptors

In this practical, you will analyze spatial point patterns of the epidermal growth factor receptor (EGFR) measured with the super-resolution microscopy technique DNA-PAINT. You will simulate spatial point patterns and compare them to experimental data to study how the dimerization of EGFR is affected by epidermal growth factor (EGF) treatment.

EGFR is a transmembrane protein involved in key cell signaling pathways. Its overexpression and dysregulation are associated with the development and progression of various cancers. Stanley Cohen received a Nobel Prize in Physiology and Medicine in 1984 for discovering this protein. Upon binding to EGF, EGFR dimerizes (see Figure 1), which is a critical step in signal transduction. EGFR can also form higher-order oligomers, but we will focus on dimers only in this practical. For more information, please check: https://en.wikipedia.org/wiki/Epidermal_growth_factor_receptor.

To visualize EGFR, DNA-PAINT was used, a super-resolution fluorescence microscopy method that detects single molecules through DNA hybridization. DNA-conjugated GFP nanobodies were used to label intracellularly tagged GFP fused to EGFR (see Figure 1).

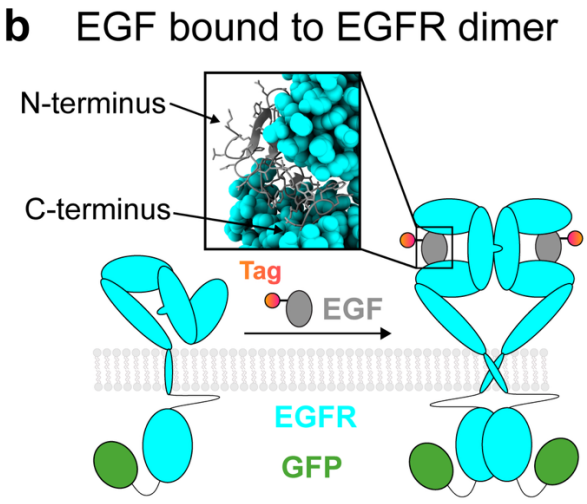

Figure 1. EGFR structure and dimerization. EGFR monomer (left) in the resting state dimerizes (right) upon binding to EGF. GFP at the intracellular side is shown, which is subsequently labeled with DNA-conjugated GFP nanobody, ultimately imaged with DNA-PAINT. From Honsa, et al. Small methods, 2025.

Preparation **before** the coursework

We assume you have some experience programming in Python for this course.

Before we start, please install Picasso on your computer as explained here: https://github.com/jungmannlab/picasso?tab=readme-ov-file#via-pypi. Please make sure you install Picasso with PyPI.

If you cannot bring your own laptop, please let us know in advance, we probably can organize one for you.

Additionally, please read the following:

- Coursework for the wet lab course to understand how DNA-PAINT works, which will make it much easier to understand what we look at
- "Statistical Methods" chapter of Khater, et al. Patterns, 2020.
- L.A. Masullo et al. Spatial and stoichiometric in situ analysis of biomolecular oligomerization at single-protein resolution. Nat Commun 16, 4202 (2025). (https://doi.org/10.1038/s41467-025-59500-z) → describes the general case of the kind of analysis we conduct in this course.
- M. Honsa et al. Imaging Ligand-Receptor Interactions at Single-Protein Resolution with DNA-PAINT. Small Methods 2025, 2401799. (https://doi.org/10.1002/smtd.202401799) → talks in more detail about EGFR and DNA-PAINT imaging thereof.
- https://www.youtube.com/watch?v=BK5x7IUTIyU&ab_channel=Computerphile – useful data structure that we will rely on heavily.

Your tasks

In this coursework, we will analyze how dimerization of EGFR changes upon EGF treatment. You are provided with the following files:

- `egfr_wt.hdf5` and `egfr_wt.yaml` – molecule positions of EGFR at the resting state (untreated).
- `egfr_treated.hdf5` and `egfr_treated.yaml` – molecule positions of EGFR after EGF treatment.
- `ripley.py` – Python script containing a function used to study interaction of a point pattern.

1. Open data in Picasso: Render (once your conda environment is on, type `picasso render` to open the app, simply drag and drop the files).
    a. What difference do you see between the two datasets?
2. In your Python script / jupyter notebook, open the same files using `picasso.io.load_locs`
    a. Note that x and y coordinates are saved in the units of camera pixels. Downstream, we need to convert that to nm (1 cam. pixel = 130 nm).
3. *(Extra points) As preliminary analysis, create Ripley curves (K(r), L(r) or H(r)) for both datasets. What do you see?*
    a. *Hint: Use the Python script supplied by us.*
4. Determine the first nearest neighbor distances (NNDs) for both experimental datasets and plot them in a histogram. *Hint: Use* `scipy.spatial.KDTree`
    a. What do you observe? How do the two datasets differ?
5. Simulate randomly distributed points (remember to keep the same density as in the data!).
    a. How do the NNDs of such simulated data compare to the experimental data? Does one dataset look more similar to the simulation than the other?

6. Simulate dimers (remember to keep the same density as in the data):
    a. Assume that the mean distance between the molecules in a dimer is 21 nm.
    b. Randomly distribute and randomly rotate the dimers around the simulated region of interest.
    c. Add uncertainty in the position of the simulated molecules – this mimicks the fact that the labels that we image are not fixed in place and can be found at different positions. Use `numpy.random.normal` with scale of 6 nm, i.e., apply Gaussian noise with sigma = 6 nm.
    d. Calculate NNDs of the simulated dimers. What do you observe?
7. Mix simulated monomers and dimers at different proportions. Can you deduce which stoichiometry is observed in both datasets? How does it compare to the introduction in this manuscript? *For extra points, write a function that will fit the proportion to any dataset.*
8. *(Extra points): Look at higher order NNDs (second, third, fourth nearest neighbors) of your simulated data and experimental data. What do you observe? Is our model, consisting of monomers and dimers, enough to explain the underlying structure of the data? What could be changed?*