Supporting Information

Super-resolution imaging of structure, molecular composition and stability of single oligonucleotide polyplexes

Natalia Feiner-Gracia1‡, R. Alis Olea1,2‡, Robert Fitzner3, Najoua El Boujnouni 2,

Alexander H. van Asbeck2, Roland Brock2*, Lorenzo Albertazzi1,4*

1Nanoscopy for Nanomedicine Group, Institute for Bioengineering of Catalonia (IBEC), The Barcelona Institute of Science and Technology (BIST), Carrer Baldiri Reixac 15-21, 08024 Barcelona, Spain.

2Dept. of Biochemistry, Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, Nijmegen

3Department of Mathematics and Computer Science, Eindhoven University of Technology, Post Office Box 513, 5600 MD Eindhoven, Netherlands.
MATERIALS AND METHODS

Reagents

N-terminally acetylated R9 and AlexaFluor488-R9 were purchased as peptide amides from EMC microcollections (Tübingen, Germany). Firefly luciferase-coding mRNA (L6107 and L6401 from Trilink, San Diego, USA) was 1929 nucleotides in length, capped with Cap 0, polyadenylated and modified with 5-methylcytidine and pseudouridine, with or without Cyanine 5 fluorescent labeling (Cyanine5-UTP:Pseudo-UTP 1:3). The mRNA was kept in the original 10 mM Tris-HCl buffer, pH 7.5. The approximate number of Cy5 labels per mRNA was determined from the ratio of Cy5 and mRNA concentrations extrapolated from measuring the extinction of a 10x dilution in water using a NanoDrop ND-1000 (Thermo Fisher, Massachusetts, USA) detailed information can be found in the Supplementary Information. The mRNA and Cy5 concentrations in the sample were
calculated from the absorption profile using the MicroArray function of the NanoDrop 1000 (Thermo Fisher).

**Polyplex preparation**

The N/P ratio was calculated by dividing the positive charges of the side-chain amino groups of the peptides (9 for R9) through the number of negative charges of the mRNA phosphate backbone (1929). For R9 polyplexes of N/P 5, the final peptide concentration was set to 50 µM. The N/P ratio for nona-arginine polyplexes was varied by maintaining the mRNA concentration at 0.045 µM (1160 ng).

For polyplex formation, equal volumes of diluted solutions of CPP and mRNA were prepared in water. The mixing was performed by simultaneous expulsion of the fluid from two identical micro-pipettes arranged to create an obtuse angle on the wall of an Eppendorf tube. For optimal imaging of R9 polyplexes, the AlexaFluor488 labeled R9 was pre-mixed with unlabeled R9 at 1.25% final labeling percentage.

**DLS measurements**

Several dilutions were prepared from non-fluorescently labeled polyplexes in RNase-free water and measured using a Zetasizer Nano-ZS ZEN 3600 (Malvern Instruments, Germany).
**Microscopy sample preparation**

Oxygen depleting dSTORM buffer was prepared fresh for every experiment using RNase-free MilliQ water, glucose (5% w/v), glucose oxidase (0.5 mg/mL), catalase (40 µg/µL) and β-mercaptomethylamine (100 mM; all Sigma Aldrich).

The microscopy slide was prepared using two stripes of double-sided adhesive tape that secured the coverslip on the glass slide. Polyplex solution was pipetted in the remaining space, requiring ~ 35 µL of sample solution. After 10-15 min incubation at room temperature, unbound polyplexes were washed by flushing 2 times 40 µL STORM buffer, followed by the addition 35 µL water-based solution of TetraSpeck Microspheres (0.1 µm, Life Technologies/Thermo Fisher, Massachusetts, USA) for drift-tracking. After 5-10 min of incubation, unbound particles were removed by flushing again 2 times 40 µL of dSTORM buffer. The sample was placed in the microscope upside-down, for imaging through the cover-slip (150 µm thickness, 24x24 mm size).

**dSTORM microscopy**

Microscopy was performed on a Nikon N-STORM system (Nikon Europe, Amsterdam) using three laser-lines: 488 nm (~80 mW) for AlexaFluor488, 561 nm (~80 mW) for TetraSpec Microspheres, and 647 nm (~160 mW) for Cy5 fluorescence. The sample was
illuminated using a total internal reflection fluorescence (TIRF) alignment system and the z-level was kept constant by Nikon perfect focus system. Fluorescence was captured using a 100x Nikon oil-immersion objective with 1.49 NA and passed through a quad-band pass dichroic filter (9733 Nikon). Images were captured by the in-built Hamamatsu ORCA-Flash 4.0 camera on a 256x256 pixels field (170 nm/pixel) at a rate of 20 ms/frame. dSTORM images contained 20,000 and 40,000 frames for the 647 nm and 488 nm channels respectively at 100% laser power. TetraSpec Microspheres were imaged once every 100 frames using 5% laser power in the 561 nm channel. Total imaging time was approx. 25 min/image.

*Image analysis*

Identification of individual blinks in each frame was performed by fitting a 2D Gaussian function on the intensity profile in NIS Elements software (Nikon). For all experiments, the identification thresholds (the difference between the number of photons of the detected pixel and the surrounding pixels) were kept constant: 250 photons for the 647 nm channel (mRNA-Cy5), 200 photons for the 488 nm channel (R9-Alexa Fluor 488) and 700 photons for the 561 nm channel (TetraSpec Microspheres for drift correction). In order to exclude
non-blinking molecules from consecutive frames, the trace length was set from 1 to 5, meaning that a molecule “on” for less than five consecutive frames was counted as a single localization, while a molecule “on” for more than five consecutive frames was discarded. No filter for the number of photons was applied.

The dSTORM image was exported as a list of localizations in “.txt” format and analyzed using a MatLab script previously described in Feiner-Gracia et.al34. The first 500 frames of the Cy5-mRNA and 1000 of the Alexa488-R9 were excluded to let the blinking equilibrate after the initial sample illumination. The mRNA-Cy5 localizations were clustered using a kernel density estimation with a bandwidth of 65 nm. An ellipse was fitted on the obtained clusters in order to apply filtering with the following parameters: minimum 10 localizations, minimum 10 nm for the small diameter, maximum 225 nm diameter for the longest axis, maximum elongation factor of 2.3 (representing the ratio of the long and short axes of the ellipse) and minimum 150 nm distance between cluster density centers. Using the cluster centers, circles to determine the polyplex size were fitted on the mRNA-Cy5 localizations to contain 90% of the points in each cluster. The circle radius was further checked to be minimum 5 nm and maximum 125.5 nm. The analysis output consisted of the number of localizations per cluster for mRNA-Cy5 (all clustered points) and Alexa488-R9 (the localizations within 1.2 times the circle radius) as well as the radius of the fitted circle for each identified cluster.
Polyplex internalization experiments

For cell biological experiments, a 10-fold lower mRNA concentration than for the dSTORM analysis was used. Polyplexes were formed at an mRNA concentration of 45 pM and peptide concentrations of 5, 15, 25 and 35 μM in increasing order of N/P ratio. For cellular uptake studies, the polyplexes were diluted 1:10. Fluorescein-labeled R9 was pre-mixed with unlabeled R9 at a ratio of 1:10. Polyplexes were formed as described previously and a 1:10 dilution of polyplex solution in RNase-free water was used for DLS measurements.

HeLa cells (DSMZ ACC-57) were maintained in sterile conditions in Dulbecco’s Modified Eagle’s medium (DMEM, Gibco) supplemented with 10% fetal calf serum (FCS, PAN-biotech, Aidenbach, Germany). HeLa cells were seeded in an 8-well chambered coverslip one day prior to the experiment (4.5*10⁴ cells/well). On the day of the experiment, cells were incubated with 250 μL of polyplex solution in FCS-free DMEM for 60 minutes at 37 °C, 5% CO₂. Cells were then washed three times with FCS-free DMEM and kept in Opti-MEM 1 reduced serum medium. Live cells were imaged directly using a Leica SP5 + FCS confocal microscope (Leica Microsystems), equipped with an HCX PL APO 63x 1.2 water immersion lens. Fluorescein was excited at 488 nm using an Argon ion laser and emission was collected between 495 and 545 nm. Cy5 was excited at 633 nm using a HeNe laser and emission was collected between 640 and 720 nm.
Images were quantitated with Fiji image analysis software. In brief, noise was reduced with a low pass filter followed by generation of threshold-based binary masks for extraction of vesicular fluorescence. Thresholds were adjusted independently for both channels. The pixel numbers of the binary masks were divided by the number of cells to determine average cellular pixel numbers. Average extracellular intensities were extracted from regions-of-interest outside cells from images acquired at higher detector sensitivity. Data for both independent experiments were normalized to the sum of all conditions and averaged. Error bars correspond to standard deviations of the means.

Quantification of the number of Cy5 molecules per mRNA molecule

First, the theoretical number of Cy5 molecules per mRNA molecules were calculated based on the data obtained from the provider. The mRNA contains a ratio of 1 pseudo-UTP: 3 UTP; the total number of mRNA bases is 1929 and considering 1 in every 4 bases is UTP we obtain 482 UTP bases. Therefore, theoretically we should have 125 Cy5 per mRNA molecule.

To experimentally verify the number of Cy5 dyes per molecule, a solution of mRNA-Cy5 at 20 ng/μL in MiliQ water was prepared and the absorbance at 260 nm and 280 nm (to calculate the real mRNA concentration) and at 647 nm (to calculate the Cy5 concentration) was measured in triplicates using a Nanodrop. We measured to have 0.973 μg/mL mRNA in the stock solution, which was close to the theoretical 1 μg/mL, and
40.5 μM Cy5. Then, using these concentrations we calculated the mean number of Cy5 molecules per mRNA molecule to be 27, therefore we have one labeled base every 70 bases.

*Polyplexes with single labelled-molecules*

To get to quantitative molecules count per polyplexes using dSTORM, one must consider that the dye can be localized several times during acquisition since fluorophores are subject to blinking. To take this effect into account it is necessary to calibrate the measurements by estimating the statistical distribution of these blinking events. Since this phenomenon is known to highly depend on the environment the number of localizations per individual molecule were estimated from measurements performed under similar conditions than during the main experiments. In the present work, to independent measurements were done, the first to detect single mRNA molecules and the second to detect single R9 molecules. In these experiments, the labeled mRNA-Cy5 or Alexa488-R9 respectively were diluted with unlabeled mRNA or R9 molecules until reaching a concentration where only one labeled molecule remained per polyplex. In each sample, the opposite compound, which was not diluted, was used as a reference for polyplex identification during image quantification.
**Figure S1.** Localizations of single labeled molecules inside polyplexes. Localizations were quantified inside polyplexes optimally labeled with the other color. (A) Distribution of localizations given by single Alexa488-R9 molecules inside polyplexes, bin 1, fitted with an exponential decay ($\tau = 2.7$); (B) distribution of localizations from single mRNA-Cy5, bin 5.

**SIMULATION**

To gain further insight we designed and implement a program that can simulate the underlying process of stochastic photoactivation to compare its output to the measurements. Agreement of simulation and experiment for a particular condition demonstrates that the correct assumptions on likelihood of photoactivation were made, a prerequisite for a quantitative analysis. We will briefly explain the simulation in four steps: temporal resolution, spatial resolution, the model, parameter estimation.
**Temporal resolution:** We want to simulate the activation/deactivation of light emission by dyes as measured. We could either simulate the real physical object as continuous-time process or just simulate the signal as processed by the STORM as individual frames. In fact, we implemented both versions, using exponential event times for the continuous-time version, and found no qualitative difference between both versions. In the following we only explain the time-discrete/frame-wise version, as it is substantially easier to explain.

**Spatial resolution:** We simulate the emission of light by dyes that are allocated within polyplexes. The radii of the polyplexes are smaller than the wavelength. While STORM can distinguish the location of individual dyes, it can do so only when just one is emitting light within one frame. If more than one dye in the same polyplex is emitting light in the same frame, the image processing will register that the detected light comes from more than one sources and will discard the signal. For this reason, the precise location of a dye within the polyplex is actually irrelevant for our result which only focusses on the detection likelihood and will not be modelled.

**The model:** We model each polyplex as a cluster of dyes that are all at the same spot. We make the reasonable assumptions that each dye within the cluster behaves independently and each will randomly be active, deactivate or bleach, as described in Figure S2.
Figure S2. Possible states and transitions between the states of a dye.

To replicate the analyzed experimental data the program output consists of three numbers for each frame: number of polyplexes with one active dye, number of polyplexes with more than one active dye, and the number of bleached dyes until this frame.

**Parameter estimation**: The simulation requires the input of parameters, that should be chosen in a reasonable manner. Here we discuss how we estimated the required parameters, see Figure S2.

Figure S3. Screenshot of the Input interface of the simulator. The parameters Nr cluster, Nr of RNA/cluster, Nr of red dyes, and Nr of green dyes can be chosen to be deterministic or to follow a passion, or a geometric distribution where the input value corresponds to the mean.
Most of these parameters are straightforward to estimate, such as
- number of frames for each kind of dyes,
- number of clusters = polyplexes to be simulated in one run,
- number of green-dyes = labeled peptides in each cluster
- number of mRNA on average in each cluster
- average number of red-dyes = Alexa488-R9 dyes on each mRNA

As can be seen in Figure S3 these quantities can be chosen to be deterministic for each polyplexes or to follow a poisson distribution, where in this case the input corresponds to mean. In all our simulations of polyplexes we use a Poisson distribution to account for the heterogeneity across different polyplexes. Also, we assumed the number of peptides and mRNA are independent from one another to simplify the simulation. Only the probabilities of changing the state of a dye remain to be estimated.

Simulation · Parameters estimation

For the peptide for which each peptide carried exactly one label, the probability of activation of the dye was calculated by using the data obtained from imaging polyplexes containing a single AlexaFluor488-R9. This was achieved by diluting the AlexaFluor488-R9 sample with non-labelled R9 down to 0.0002%. We compute the average number of
times these single dyes were active and divide this by the total number of frames to obtain a probability of activation of 0.0000548.

\[
P_{activation} = \frac{Mean \ number \ of \ localizations}{Total \ number \ of \ frames}
\]

The probability of deactivation was estimated using the average number of consecutive frames a dye stays ON once it is activated, which is one of the outputs of the NIS-elements software used to process the images. This number was estimated from the experimental data of the single AlexaFluor488-R9 and found to be 1.85 frames.

Then, we used the simulator to simulate the blinking behavior of single peptides and we calculated the cumulative mean of this same value (number of consecutive frames a dye stays ON once it is activated) using different deactivation probabilities from 0.25 to 0.8. After comparing the different simulations, the probability of deactivation was estimated to be 0.37, as the cumulative mean number of consecutive frames was stable at 1.855 close to the experimentally calculated one 1.85.
Figure S4. Estimation of the deactivation rate for the simulation. The cumulative mean number of consecutive frames a dye is ON was calculated and plotted after simulating single polyplexes having single AlexaFluor488-R9 using different probabilities of activation (A) 0.35, (B) 0.37 and (C) 0.38.

Having establish the probability of activation and deactivation the probability of bleaching was determined. The number of detected dyes per time frame of experimental and simulated data considering 0.1 as the probability of bleaching were plotted and compared. The data was binned every 1000 frames and a linear fitting was applied to compare simulation and experimental data. The slope of the curve was similar for both being -0.008 and -0.006 for the simulated and experimental data, respectively, Therefore 0.1 was considered the bleaching probability of the Alexa488 dye in r9 molecules.
Figure S5. Estimation of the bleaching rate of AlexaFluo488 conjugated to R9.

Supplementary Figures
Figure S6. STORM image of polyplexes with non-labelled molecules compared to a polyplexes labelled. Few sparse localizations can be observed associated to the background noise in the negative control.
Figure S7. Measurement of the polyplexes size at N/P 5 using dynamic light scattering (DLS), distribution by intensity.

Figure S8. 3D STORM imaging of polyplexes N/P 5. A. Localizations of a center Z slice of 100 nm
C. Localizations of a center Z slice of 50 nm
Figure S9. Simulation vs experimental data of polyplexes N/P 1. The number of localizations during imaging was simulated considering that each polyplex had either 2, 3 or 4 peptides. The frequency histogram of the number of localizations for each simulation is plotted together with the experimental data. The most similar data was the simulation of 2 peptides per polyplex.
Figure S10. Simulation vs experimental data of polyplexes N/P 3. The number of localizations during imaging was simulated considering each polyplex had either 3 or 4 peptides. The frequency histogram of the number of localizations for each simulation is plotted together with the experimental data. The most similar data was the simulation of 3 peptides per polyplex.
Figure S11. Simulation vs experimental data of polyplexes N/P 5. The number of localizations during imaging was simulated considering that each polyplex had either 46, 35 or 30 peptides. The frequency histogram of the number of localizations for each simulation is plotted together with the experimental data. The most similar data was the simulation of 30 peptides per polyplex, where the peak of both distributions fit.
**Figure S12.** Simulation vs experimental data of polyplexes N/P 7. The number of localizations during imaging was simulated considering each polyplex had either 35 or 25 peptides. The frequency histogram of the number of localizations for each simulation is plotted together with the experimental data. The most similar data was the simulation of 25 peptides per polyplex.
Figure S13. Mean number of R9 localizations/mRNA per polyplex as a function of the radius. (A) N/P 1 (B) N/P 3 (C) N/P 5 and (D) N/P 7. For each polyplex the number of R9 localizations was divided by the number of mRNA localizations in the same polyplexes; then all the polyplexes with the same radius were clustered and the mean number of R9 localizations/mRNA localizations was calculated and plotted.

Figure S14. Cytotoxicity study of polyplexes at different N/P. The polyplexes were incubated with cells for 1 hour and cytotoxicity was tested after 24 hours, controls were added where cells were incubated with PBS (C-) and 30% ethanol (C+).
Figure S15. Stability of N/P 5 polyplexes in serum at different concentrations. Number of localizations after 1 and 60 min of incubation with (a) 1% FBS, (b) 20% FBS, and (c) 100% FBS in comparison to the initial polyplexes. Left graphs correspond to the median number of mRNA localizations and the right graphs to the median number of R9 localizations.
