Electronic Supplementary Material (ESI) for Nanoscale.

Supplementary Information to 'Concentrating and labelling genomic DNA in a nanofluidic array'

Rodolphe Marie,[*,†] Jonas N. Pedersen,[†] Kalim U. Mir,[‡] Brian Bilenberg,[¶] and Anders Kristensen[†]

Department of Micro and Nanotechnology, Technical University of Denmark, Kongens Lyngby, Denmark, XGenomes, Pagliuca Harvard Life Lab, 127 Western Ave, Boston, USA, and NIL Technology ApS. Kongens Lyngby, Denmark

E-mail: rodolphe.marie@nanotech.dtu.dk

Detection of molecules and enzymatic reactions

Figure S1a,b show examples of the YOYO-1 and Cy3 signals from DNA molecules in pits, and Figure S1c,d show histograms of the mean intensities. Dashed, red lines mark the cutoff values used for detection of molecules in pits (Figure S1c) and enzymatic reactions (Figure S1d). The number of counts above the cutoffs are 1156 and 24, respectively.

[*]To whom correspondence should be addressed
[†]Department of Micro and Nanotechnology, Technical University of Denmark, Kongens Lyngby, Denmark
[‡]XGenomes, Pagliuca Harvard Life Lab, 127 Western Ave, Boston, USA
[¶]NIL Technology ApS. Kongens Lyngby, Denmark

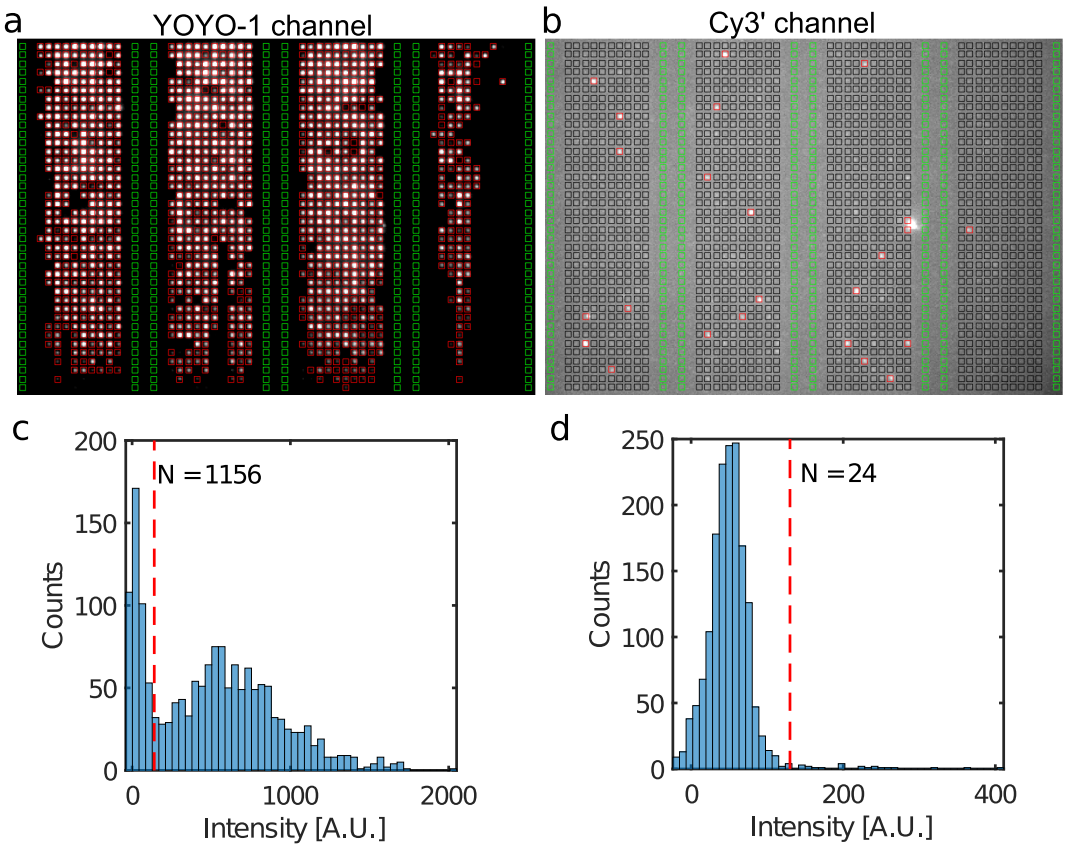


Figure S1: (a) Fluorescence images of YOYO-1 and (b) Cy3 signals. Regions used for measurements of background intensities are marked with green boxes. (c-d) Histograms of mean intensities of the YOYO-1 and Cy3 signals from individual pits, respectively. Dashed, red lines mark the threshold values for pit occupancy and enzymatic reactions, respectively.

Occupancy of nanopits

In equilibrium, a DNA molecule may have all its contour length in a single pit or distribute it over several, neighbouring pits. With the framework developed by Reisner *et al.* it is possible to calculate the probability P_N that a DNA molecule in equilibrium spans N square nanopits in a square nanopit array.[1] The different configurations are due to the balance between the free energy of DNA in the nanoslit and in the nanopits. The input parameters are the contour length of the DNA molecule, its persistence length and effective width, the height of the nanoslit, and the dimensions and the spacing of the nanopits.[1]

We minimize the free energy defined in Eqs. (1-7) in Ref. 1 and calculate the occupancy probabilities P_N [Eqs. (8-9) in Ref. 1]. The scaling prefactors in Eqs. (3,5,6) in Ref. 1 are set

to unity. With our experimental parameters (contour length $L = 21\,\mu$m, persistence length $P = 64\,$nm, height of the nanoslit $h = 100\,$nm, side length of pit $a = 300\,$nm, pit depth $h = 200\,$mn, distance between pits ℓ=2000 nm-a, and the effective width of DNA $w = 9\,$nm), we find for the 800 and 1000 nm nanopits that $P_1 = 1$. So in equilibrium only a monomer state is formed for these pit sizes. For the 600 nm pits, we get $P_1 = 0.93$ and $P_2 = 0.07$. So in equilibrium, a DNA molecule switches between a monomer and a dimer configuration.

Waiting time measurements for $1 \to 0$ events

Figure S2 shows the fluorescence intensity versus frame number for a single nanopit in a nanopit array. The intensity increases significantly when a DNA molecule enters the nanopit, e.g., at frame number 10. A pit is classified as occupied by DNA if the intensity is larger than a predefined threshold value for at least three consecutive frames. If a molecule enters an unoccupied pit and then leaves it again, the event is counted as a '$1 \to 0$' event. The waiting time $t_{1\to0}$ between the molecule enters and leaves the pit is recorded. Measurements of $t_{1\to0}$ are performed in an array where only a few molecules are present in order to minimise the number of pits occupied by more than a single molecule. This also reduces the chance that a molecule replaces another, undetected molecule between two frames.

Based on transition state theory,[2,3] we assume that the waiting times $t_{1\to0}$ are exponentially distributed,

$$p(t_{1\to0}) = \frac{\exp\left(-\frac{t_{1\to0}}{\tau_{1\to0}}\right)}{\tau_{1\to0}} \quad \text{for } t > 0, \tag{1}$$

with a characteristic time scale $\tau_{1\to0}$. As we cannot measure the duration of events shorter than a single time lapse Δt, we introduce a lower cutoff $K = 2\Delta t$. So the measured distribution of waiting times is assumed to follow a truncated exponential distribution,

$$p(t_{1\to0}) = \frac{\exp\left(-\frac{t_{1\to0}-K}{\tau_{1\to0}}\right)}{\tau_{1\to0}} \quad \text{for } t > K, \text{and zero otherwise.} \tag{2}$$

For N measured events with durations $t_{1\to0,i}$ $(i = 1, 2, 3, \ldots, N)$, the maximum likelihood estimator for the characteristic time scale is $\hat{\tau}_{1\to0} = \bar{t}_{1\to0} - K$ with $\bar{t}_{1\to0} = \frac{1}{N}\sum_{i=1}^{N} t_{1\to0,i}$. From the Fisher Information we get the standard error on the estimated parameter $\hat{\tau}_{1\to0}$ as $\hat{\sigma}_{\hat{\tau}_{1\to0}} = (\bar{t}_{1\to0} - K)/\sqrt{N}$.

Figure S3 shows a histogram of waiting times $t_{1\to0}$. The waiting times are consistent with a truncated exponential distribution with $\tau_{1\to0} = 26\,\mathrm{s}$, and the p-value for the fit is $p = 0.40$ (chi-squared goodness-of-fit test).

Events were recorded from measurement series for different pit sizes at various applied pressure drops. Only measurement series with more than 40 events were further analyzed, and the value $\hat{\tau}_{1\to0}$ and its standard error $\hat{\sigma}_{\hat{\tau}_{1\to0}}$ was calculated for each of them. Figure S4 shows the corresponding p-values for fits to the truncated exponential distributions. Only results from measurement series that returned p-value greater than 1% (dashed, horizontal line) were included in Figure 3b in the main text.

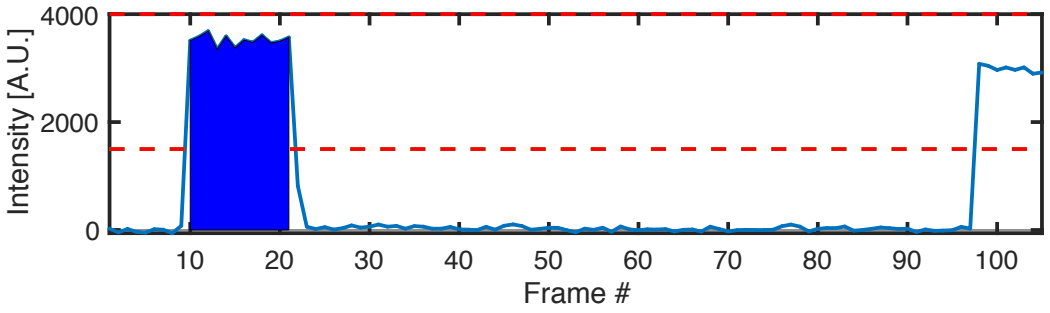


Figure S2: Intensity trace for a 1000 nm pit with an applied pressure drop $\Delta P = 0.3\,\mathrm{bar}$ and a time lapse $\Delta t = 5\,\mathrm{s}$. Dashed lines show the threshold value used for detection of events. Here a single $1 \to 0$ event is detected (filled, blue area). The last increase in intensity is not counted as an event because the intensity does not drop before the end of the recorded movie.

Waiting time measurement for $2 \to 1$ events

Figure S5a shows an intensity trace for a single pit, which is either empty, occupied by a single DNA molecule, or double-occupied. Figure S5b shows an intensity histogram used to

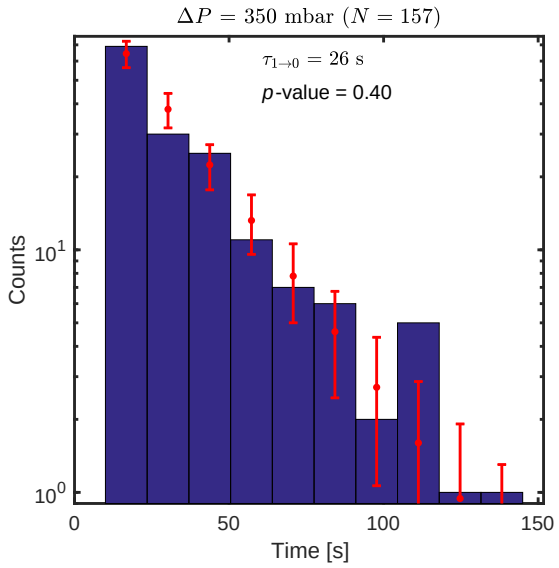


Figure S3: Histogram of waiting times for '$1 \to 0$' events for a pit size of 1000 nm and an applied pressure drop $\Delta P = 0.35$ bar. The total number of events is $N = 157$, and the time lapse is $\Delta t = 5$ s. The distribution of waiting times is fitted to a truncated exponential distribution, see Eq. (2). Red data points mark the expected number of counts in each bin $\langle n_i \rangle$ and its standard deviation ($\sqrt{\langle n_i \rangle}$) for a truncated exponential distribution with $\tau_{1\to0} = 26$ s and $K = 2\Delta t$. The p-value for the fit is 0.40.

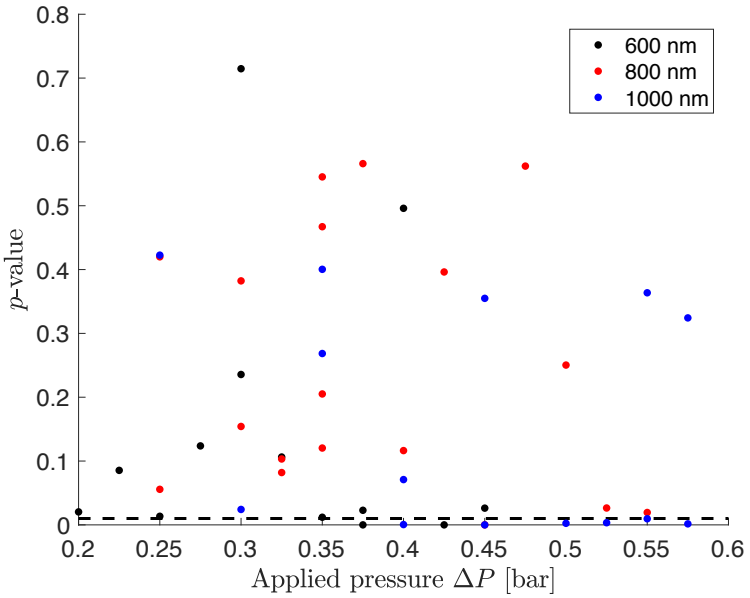


Figure S4: p-values for fits of the waiting time distributions to truncated exponential distributions for '$1 \to 0$' events for various values of the applied pressure ΔP. Pit sizes are 600 nm, 800 nm and 1000 nm. Dashed line shows $p = 0.01$.

distinguish between the different states. Events where a molecule enters an already occupied pit and then leaves it again are denoted a '$2 \rightarrow 1$' event. These are detected from the intensity trace. The threshold values (dashed lines in figure S5) ensure that the intensity corresponds to a whole molecule (intensity between the green and the lower red line) before and after the event, and two molecules (intensity greater than the upper red line) during the event. A pit has to be occupied by two molecules for at least three consecutive frames, i.e., for 4 s, before the event is counted as a '$2 \rightarrow 1$' event.

As for the '$1 \rightarrow 0$' events, we assume that the waiting times are exponentially distributed. We determine the characteristic time scales $\tau_{2\rightarrow1}$ and the standard errors as outlined for '$1 \rightarrow 0$'. Figure 3c in the main text shows the characteristic waiting times $\tau_{2\rightarrow1}$ for 1000 nm pits for various values of the applied pressure drop ΔP.

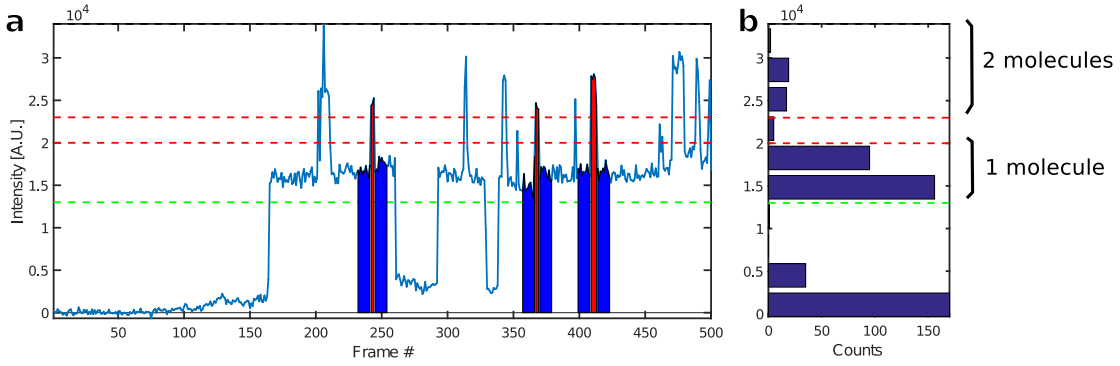


Figure S5: (a) Example of an intensity trace for a single pit with three $2 \rightarrow 1$ events (1000 nm pit, $\Delta P = 0.3$ bar, time lapse Δt=2 s). Dashed lines indicate the thresholds used to distinguish between single- and double-occupancy of the pit. (b) Intensity histogram with thresholds indicated by dashed lines.

Rate of incoming molecules

To simulate the filling of a pit array, we need the rate of incoming molecules r_{in} (see also the section 'Simulation of occupancy versus pressure'). This value can be extracted from the average number of molecules in a row in the array $\langle n(t) \rangle$ as the array fills up. The average is an ensemble average over the rows in the array.

As an example, Figure S6 shows the average number of molecules per pit in the array $\langle n(t)\rangle/N_{\text{pits}}$. In this particular case, the experiment corresponds to the array shown in Figure 3a in the main text where only 9 out of 10 identical rows are considered (40 pits long i.e. $N_{\text{pits}} = 40$). After the onset of the flow at $t = 0$, we assume that $\langle n(t)\rangle = r_{\text{in}}\, t$.* So we determine r_{in} from the experimental data by estimating the slope of the $\langle n(t)\rangle$-curve well before steady-state.

Figure S7 shows the extracted values for r_{in} for the 1000 nm pit array. We assume that the rate of incoming molecules is proportional to the flow velocity, and, consequently, the applied pressure drop ΔP. Thus we fit a straight line to data (dashed line in figure S7). This linear fit of r_{in} versus pressure drop ΔP is used as input for the simulations.

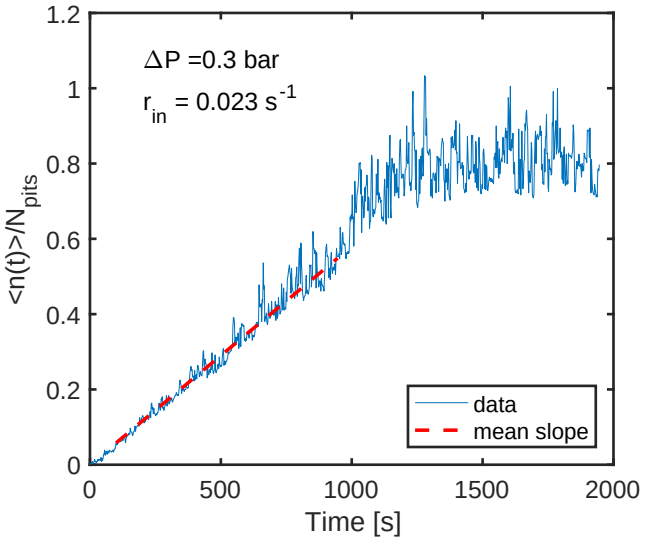


Figure S6: Number of molecules per pit in the array of 1000nm pits at ΔP=0.3 bar. We extract the rate of incoming molecules r_{in} from the mean slope of the linear part of the curve.

Simulation of occupancy versus pressure

We simulate the filling of a row in the nanopit array. The number of pits in the row is N_{pits}, and each pit can be occupied by at most two molecules. The state of the nanopit row is

*Note that the number of molecules in the array $n(t)$ is not identical to the number of occupied pits (see, e.g., Figure 2c in the main text) as the 1000 nm pits can be occupied by more than a single molecule.

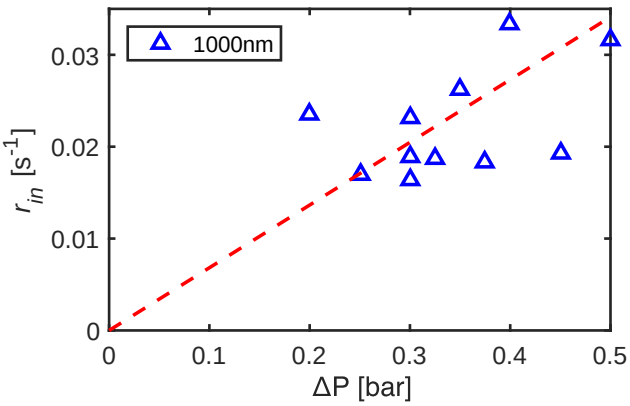


Figure S7: Rate of incoming molecules r_{in} versus applied pressure drop ΔP. Dashed line shows a linear fit that are subsequently used as input for the simulations of the fractional fillings of the pit array for different applied pressure drops, see Figure S8.

the number of molecules in the different pits, i.e., $(n_1, n_2, n_3, \ldots, n_{N_{\text{pits}}})$, where $n_i = 0, 1$, or 2, and $i = 1, 2, \ldots N_{\text{pits}}$. We simulate the state of the array and extract the number of pits occupied by at least one molecule $n_{\text{occ}}(t)$, and the total number of molecules in the row $n(t) = \sum_{i=1}^{N_{\text{pits}}} n_i$.

At $t = 0$, the array is empty, so $n_{\text{occ}}(0) = 0$. Molecules enters the row of pits from the reservoir with a rate r_{in}, leaves a single-occupied pit with a rate $r_{1\to0} = 1/\tau_{1\to0}$, and leaves a double-occupied pit with a rate $r_{2\to1} = 1/\tau_{2\to1}$. When a molecule leaves a pit, we assume that it settles in the next pit downstream occupied by at most a single molecule. If no such pit is available, the molecule leaves the row of nanopits.

Thus the state of the array can be changed in three ways:

- a molecule enters the row of nanopits from the reservoir,
- a molecule in a single-occupied pit leaves it (i.e., a '$1 \to 0$' event) and moves to the next pit occupied by at most a single molecule,
- a molecule in a double-occupied pit leaves it (i.e., a '$2 \to 1$' event) and moves to the next pit occupied by at most a single molecule.

We use the fit in Figure S7 for the rate of incoming molecules r_{in} for the different pressure

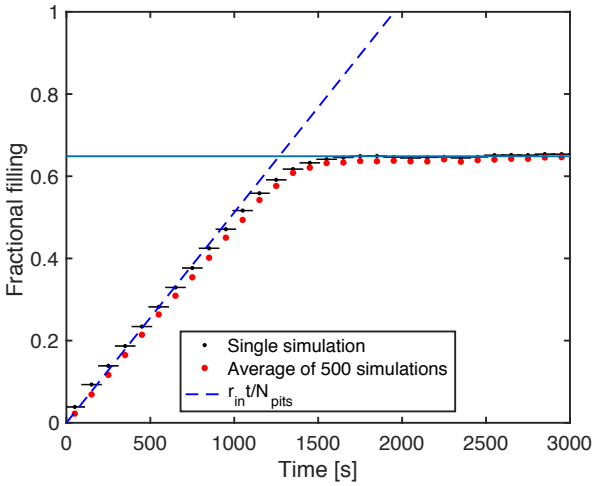


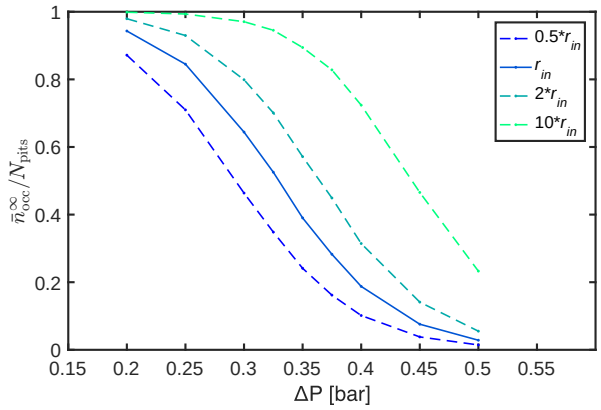
Figure S8: Simulation of the filling of a 1000 nm pit array at ΔP=0.3 bar for a row with $N_{\text{pits}} = 40$ (small, black data points), and the average over 500 rows (large, black data points). Dashed line shows $(r_{\text{in}}\, t)/N_{\text{pits}}$. Parameters in the simulation are read off from the fits in Figure 4c in the main text and Figure S7.

drops ΔP, and similarly we get $r_{1\to0}$ and $r_{2\to1}$ from the fits of the waiting times in Figure 4c in the main text.

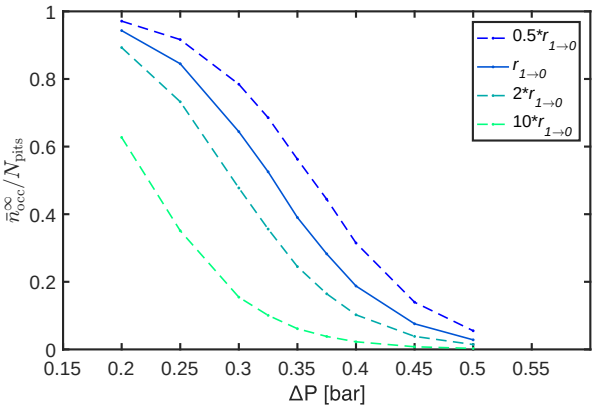
Starting from an empty array, we Monte Carlo simulate the state of the row of pits with the algorithm described in Ref.[4] and record $n_{\text{occ}}(t)$. We update the state of the row of pits until steady state. The time-averaged occupancy at steady-state is denoted $\bar{n}^{\infty}_{\text{occ}}$.

Figure S8 shows a simulation of the fractional filling $n_{\text{occ}}(t)/N_{\text{pits}}$ for a row in an array with 1000 nm pits for a pressure drop $\Delta P = 0.3$ bar. The number of pits in the row is the same as in the experiment, $N_{\text{pits}} = 40$. Red data points are the average of 500 simulations $\langle n_{\text{occ}}(t)\rangle/N_{\text{pits}}$, and the horizontal line show the fractional occupancy at steady-state $\bar{n}^{\infty}_{\text{occ}}$. The dashed line is $\langle n(t)\rangle/N_{\text{pits}} = (r_{\text{in}}\, t)/N_{\text{pits}}$, i.e., the normalized expected number of molecules in the row well before molecules start to leave it. Notice that $\langle n(t)\rangle/N_{\text{pits}} > \langle n_{\text{occ}}(t)\rangle/N_{\text{pits}}$ as pits can be occupied by more than a single molecule.

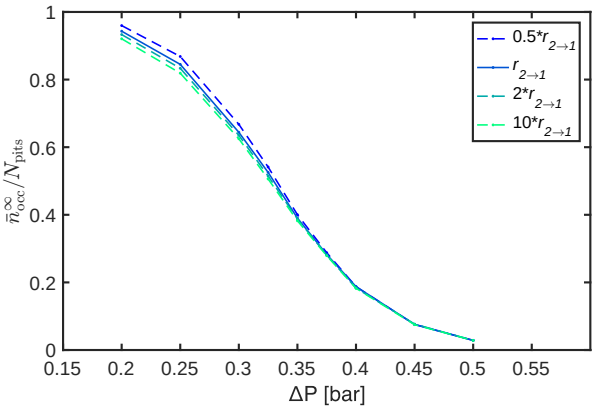
Figures S9 shows how changing r_{in}, $r_{1\to0}$ and $r_{2\to1}$ affect the fractional occupancy of a 1000 nm pit array. In an experiment, r_{in} can be changed without changing changing the applied pressure drop ΔP simply by changing the concentration of molecules in the buffer.



(a)



(b)



(c)

Figure S9: Simulation of the fractional occupancy at steady state for a 1000 nm pit array varying (a) the incoming rate r_{in}, (b) the rate $r_{1\to0}$, and (c) the rate $r_{2\to1}$. Full lines correspond to the values read off from Figure 4c in the main text and Figure S7. Dashed curves are for one parameter varied, while the others are kept fixed.

References

1. Reisner, W.; Larsen, N. B.; Flyvbjerg, H.; Tegenfeldt, J. O.; Kristensen, A. Directed self-organization of single DNA molecules in a nanoslit via embedded nanopit arrays. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 79–84.

2. Han, J.; Turner, S.; Craighead, H. Entropic trapping and escape of long DNA molecules at submicron size constriction. *Phys. Rev. Lett.* **1999**, *83*, 1688–1691.

3. Vestergaard, C. L.; Mikkelsen, M. B. L.; Reisner, W.; Kristensen, A.; Flyvbjerg, H. Transition state theory demonstrated at the micron scale with out-of-equilibrium transport in a confined environment. *Nat. Commun.* **2016**, *7*, 10227.

4. Gillespie, D. T. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* **1976**, *22*, 403–434.