Supplementary Information

The base pair scale diffusion of nucleosomes facilitates binding of transcription factors

Sergei Rudnizky^{1,#}, Hadeel Khamis^{1,2,#}, Omri Malik^{1,3,#}, Philippa Melamed^{1,3} and Ariel Kaplan $1,3,^*$

¹ Faculty of Biology, ² Faculty of Physics, and ³ Russell Berrie Nanotechnology Institute, Technion – Israel Institute of Technology, Haifa (32000), Israel

These authors contributed equally to this work

* To whom correspondence should be addressed. Email: akaplanz@technion.ac.il

MATERIALS AND METHODS

Proteins

Histones were purified as previously reported (1). The original plasmid for the DNA binding domain of Egr-1, from Dr. Scot Wolfe, was kindly provided by Dr. Amit Meller. The protein was expressed and purified as previously described (2, 3)

Molecular construct for single molecule experiments

DNA sequences corresponding to the ~200 bp of *Lhb* TSS nucleosome (1) were amplified by PCR from mouse genomic DNA with primers #1 and #2 listed in Table S1. The constructs were digested using *DraIII*-HF (New England Biolabs) overnight according to the manufacturer's instructions. A 10 bp hairpin (Sigma; #11) was ligated to the construct using T4 DNA ligase (New England Biolabs), in a reaction with 1:5 molar excess of the hairpin, at 16 °C. The construct was subsequently digested overnight with *BglI* (New England Biolabs) and mixed with recombinant histones to form mono-nucleosomes under conditions reported (1). Two ~2000-bp DNA handles were generated as previously reported (2), with following modifications: to generate double (instead of single) digoxigenin modification on one of the handles, a *BglI* restriction enzyme site was incorporated at the handle end. Two commercially purchased oligos with 3' and 5' digoxygenin terminal modifications were annealed and ligated to the handle after digestion with *BglI* (New England Biolabs; #7 and #8, Table S1). Reconstituted nucleosomes were ligated to the DNA handles using a rapid ligase system (Promega) in a 3:1 molar ratio, 30 min at room temperature. The full construct (i.e. handles + alignment segment + nucleosome, Table S2) was incubated for 15 min on ice with 0.8 μ m polystyrene beads (Spherotech) coated with anti-digoxygenin. The reaction was then diluted 500-fold in unzipping buffer (10 mM Tris·Cl (pH 7.4), 1 mM EDTA, 150 mM NaCl, 1.5 mM MgCl2, 1 mM DTT, 3% v/v glycerol and 0.01% BSA). Tether formation was performed *in situ* (inside the experimental chamber) by trapping an anti-dig bead (bound by nucleosomes) in one trap, trapping a 0.9 μ m streptavidin coated polystyrene beads in the second trap, and bringing the two beads into close proximity to allow binding of the biotin tag in the nucleosomal DNA to the streptavidin in the bead.

Optical Tweezers

Experiments were performed in a custom-made dual-trap optical tweezers apparatus as previously reported (1, 4). Experiments were conducted using a laminar flow cell (u-Flux, Lumicks).

Data analysis

Data were digitized and processed to obtain *j*, the number of unzipped bp, as previously reported (1). To improve the accuracy of the experiments, the 345 bp naked DNA alignment segment was used to perform a correlation-based alignment of all traces in an experiment, allowing shifting of the traces (i.e. redefining the position of zero extension) and stretching of up to 2%.

Nucleosome position and step size. The DNA construct was repetitively unzipped, with a total cycle time of 8 s (unless specified otherwise), starting from 2 pN, to the point where the nucleosome's region 1 is identified (but not disrupted) by detecting a force elevation of more than 3 pN from the naked DNA. The 2.5 kHz data was lowpass filtered to an effective 35 Hz bandwidth with a 70 points median filter, and the position of region 1 was defined as the center of the *j* distribution in the interaction region. The difference in region 1 position between successive cycles is interpreted as movement of the nucleosome. To generate Fig. 1f ,and Fig. 2c, the experimental traces were sliced into 5 min trajectories, the mean squared displacement (*MSD*) for each one as a function of time was calculated, and then averaged over the ensemble of trajectories. The nucleosome "step" was calculated by filtering the positions trajectories (running average with a 5 cycles window) and then calculating the difference in position between time points separated by 5 cycles.

The nucleosome step probability density function was calculated by counting the number of steps in the dataset that had a length in a certain length interval, and dividing by the total number of steps and the interval size. The obtained distribution was then fitted to a single exponential function (for H2A) or two separate exponential functions, for steps shorter or longer than 5 bp, respectively (for H2A.Z). The inset in Fig. 2d was generated by calculating the MSD of each experiments in its full length, and averaging over the ensemble of experiments. The same calculation was then repeated when the type 2 events were computationally eliminated from the traces.

S3

Potential of mean force. For H2A nucleosomes, which reach a steady state in their spatial distribution after a short time, we calculated the potential of mean force (PMF) as $PMF = -k_BT \log P(x)$, where $P(x)$ is the relative occupancy of position x by the nucleosomes in the ensemble.

Probability of exposure. In order to calculate the probability of exposure, we first converted the "construct coordinates" (bp unzipped) to "gene coordinates", where base pair number 346 corresponds to position -157 in the *Lhb* gene. Then, since region 1 is \sim 35 bps inside the total length of the wrapped DNA (1, 5), we subtract 35 bps from its experimentally determined position to obtain the nucleosome's 5' edge position (Fig. S10). As the binding site for Egr-1 is located between -104/-112, the probability of exposure was calculated for each trace by counting the number of cycles where the estimated nucleosome 5' edge position was downstream of -114, out of the total number of cycles in the experiment.

Egr-1 binding detection. Binding events on naked DNA were identified by detecting an increase in force of more than 0.5 pN as compared to the median force at the same position (18.7 pN). Applying the same criteria on an experiment without Egr-1 resulted in no detected binding events. In the case of nucleosomal DNA, Egr-1 binding could be detected only when the nucleosome 5' edge position was downstream of -120, leaving the Egr-1 site exposed.

Nucleosome movement as a function of Egr-1 binding. To check the immediate effect of Egr-1 binding on the movement of the nucleosome, we calculated the relative (upstream *vs*. downstream) conditional probability of type 2 events, given that Egr-1 is (or is not) bound to its site (Fig. 4c). First, we classified the set of all the type 2 events detected into those where Egr-1 was also detected, and those where it was not detected. Then, for each group, we counted the number of upstream (or downstream) events out of the total number of events in the group.

Dissociation time measurement. In order to measure the dissociation time of Egr-1 from its binding site, we used the method developed in our previous study (2). Briefly, we unzip the DNA until reaching the binding site and record the DNA breathing fluctuations. Binding of Egr-1 suppresses these fluctuations, and its typical dissociation time is longer than the time-scale of DNA breathing. Hence, the observed breathing kinetics, extracted with the HaMMy software (6), is used to characterize the

S4

binding kinetics: short-lived closed state (< 0.3 s) correspond spontaneous closing (i.e. DNA breathing), and long-lived closed states (> 0.3 s) are attributed to Egr-1 binding. The dissociation time is estimated as the mean of the long-lived closed events.

SUPPLEMENTARY DISCUSSION

Our measurements do not affect the spontaneous dynamics of nucleosomes

A number of control experiments were conducted to prove that the dynamics we observe is not induced by the experiment itself, i.e. that ours is a non-perturbative measurement.

- i. If the dynamics we observe were induced by the unzipping/rezipping, we would expect that the measured mobility will increase as we increase the frequency of probing. Hence, we characterized the MSD for 5 min experiments where the nucleosomes' position is probed every 8, 30 or 60 s, while keeping the force loading rate at the same value (12 pN/s). Fig. S1c shows that no significant differences are seen for either H2A or H2A.Z nucleosomes. Notably, the differences between H2A and H2A.Z nucleosomes are conserved for all the different probing frequencies.
- ii. To prove that the observed mobility is not specifically related to the introduction of structural defects during unzipping/rezipping, we compared the mobility of H2A.Z nucleosomes probed at two different unzipping/rezipping force loading rates. Introduction of structural defects, which implies that the structure of the nucleosome fails to reach its true energetic minima during rezipping, should depend on the rate of rezipping, as this determines how far from equilibrium the process takes place. Fig. S1d shows that no differences are observed in the 5 min MSD for experiments probed every 30 s, with two different loading rates: 12 pN/s (similar to the rest of the data) and 3 pN/s.
- iii. The possibility that structural defects are responsible for our observation is also ruled out by the symmetry of the step size distribution. Given the asymmetry of the experiment (we unzip/rezip only partially and from a specific side), one would expect that defects will be introduced only from this side. This means that the repositioning events are expected to be in a single direction, against the direction of unzipping, in contradiction with the observed symmetric distribution (Fig. S8).
- iv. To confirm that the integrity of the nucleosomes is conserved throughout the experiment (i.e. that some of the steps we observe are not due, for example, to

losing a dimer), we used the last cycle of the experiment, in which we perform a full unzipping curve. The expected signature with two regions of interaction (Fig. S9a), which is clearly distinguishable from the signature of hexasomes and tetrasomes, was observed in all of the nucleosomes we characterized. Moreover, quantification of the distance between region 1 and region 2, for both H2A and H2A.Z nucleosomes is similar to what was previously measured (Fig. S9b).

Our assay is sensitive to nucleosomal movements, not local dynamics

It is well known that the entry region of the nucleosome fluctuates, i.e. that contacts between the histone octamer and the DNA are spontaneously broken and reestablished. This "breathing" dynamics has been extensively studied (7–12), and shown to play a role in TF binding (7–9, 13) and RNAP elongation (14). In view of these previous results, it is important to consider how the breathing dynamics can affect our measurements. It may be argued that, every time a DNA molecule is unzipped, the position of the first interaction detected is different not because the nucleosome as a whole moves, as we claim, but because the specific interaction we detect undergoes breathing fluctuations. However, there are a number of points that help us rule out this possibility: First, we do not probe the position of the entry region in our experiments. We probe the position of "region 1" (1, 5, 15, 16), which correspond to DNA contacts with the H2A/H2B dimer at superhelical locations (SHL) − 3.5 to − 6.5 (or $+3.5$ and $+6.5$). These region are at \sim 35 bp on either side of the nucleosome's dyad (1, 5, 15, 16), which means that they are located \sim 35-40 bp from the "entry" region. Although, in principle, these regions can spontaneously unwrap from the DNA too, these events are associated with much higher energy barriers, and both experiments (7) and theoretical calculations (17) indicate that the equilibrium constant for such fluctuations is two orders of magnitude smaller than that for breathing of the entry region. Moreover, experimental evidence exists on the kinetics of such events (8, 18). It has been shown that while the re-wrapping time (on the miliseconds scale) is not significantly different between the entry region and regions further inside the nucleosome, the unwrapping rate decreases dramatically for the latter: For sites positioned 37 and 47 bp from the 5' end of the nucleosomal DNA, approximately the position of region 1, unwrapping times of 1 and 10 min, respectively, where measured (18). Hence, these breathing events are extremely rare and very short. In addition, the kinetics described above indicate that these events will not be detected in our experiments. The interaction of the unzipping fork with region 1 during an unzipping trace takes \sim 0.5 s, and the position of the interaction region is determined as the average position during the interaction time, after filtering the data down to 35 Hz. This means that the measurement is not sensitive to such short fluctuations, but only to the average position of region 1 during the interaction time. Since the unwrapping events, if they exist, are very rare, this average equals to the wrapped state. Finally, breathing is inherently an asymmetric process, as the energy involved is a function of the amount of DNA unwrapped. Assuming that what we measure is a breathing fluctuation, i.e. that we measure the position of the first wrapped DNA, we should see that the measured "position" of the nucleosome is distributed in time in a highly asymmetric manner, likely exponentially, since small breathing fluctuation are much more common than big ones. However, the distribution in the position measured for single nucleosomes (see for example the histograms in Fig. 3d) is symmetric. Taken together, these considerations rule out the interpretation of the observed dynamics as corresponding to breathing fluctuations.

Notably, although it is not possible to rule out completely that other conformational changes in the nucleosome occur during our experiments, our data rules out the possibility that they contribute significantly to our measurements and indicate that the dynamics we observe are essentially a "center-of-mass" movement. First, we use the last cycle of our experiments to completely unzip the nucleosome and measure the distance between region 1 and the dyad. If the observed dynamics were the result of local, large and stable conformational changes, we would expect to randomly observe that this distance is significantly smaller or larger than expected in many of our experiments, in contrast with our measurements. This suggests that the distance between region 1 and the dyad is conserved. In addition, our data shows that there is a very good correlation between the exposure of the TF binding site (as calculated from the measured position of region 1 by subtracting 35 bp) and actual binding of the TF (Figs. 2e and 4b). This suggests that the distance between region 1 and the 5'

S8

edge is conserved too. Taken together, this indicates that while the position of region 1 changes from cycle to cycle, its distance to the dyad and to the 5' edge do not, suggesting that the nucleosome moves as a whole.

Supplementary References

- 1. Rudnizky S, et al. (2016) H2A.Z controls the stability and mobility of nucleosomes to regulate expression of the LH genes. *Nat Commun* 7:12958.
- 2. Rudnizky S, et al. (2018) Single-molecule DNA unzipping reveals asymmetric modulation of a transcription factor by its binding site sequence and context. *Nucleic Acids Res* 46:1513.
- 3. Squires A, Atas E, Meller A (2015) Nanopore sensing of individual transcription factors bound to DNA. *Sci Rep* 5(1):11643.
- 4. Malik O, Khamis H, Rudnizky S, Marx A, Kaplan A (2017) Pausing kinetics dominates strand-displacement polymerization by reverse transcriptase. *Nucleic Acids Res* 45(17):10190–10205.
- 5. Hall MA, et al. (2009) High-resolution dynamic mapping of histone-DNA interactions in a nucleosome. *Nat Struct Mol Biol* 16(2):124–9.
- 6. McKinney SA, Joo C, Ha T (2006) Analysis of Single-Molecule FRET Trajectories Using Hidden Markov Modeling. *Biophys J* 91(5):1941–1951.
- 7. Polach KJ, Widom J (1995) Mechanism of protein access to specific DNA sequences in chromatin: a dynamic equilibrium model for gene regulation. *J Mol Biol* 254(2):130– 49.
- 8. Li G, Levitus M, Bustamante C, Widom J (2005) Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struct Mol Biol* 12(1):46–53.
- 9. Li G, Widom J (2004) Nucleosomes facilitate their own invasion. *Nat Struct Mol Biol* 11(8):763–769.
- 10. Wei S, Falk SJ, Black BE, Lee T-H (2015) A novel hybrid single molecule approach reveals spontaneous DNA motion in the nucleosome. *Nucleic Acids Res* 43(17):e111–e111.
- 11. Chen Y, et al. (2014) Revealing transient structures of nucleosomes as DNA unwinds. *Nucleic Acids Res* 42(13):8767–8776.
- 12. Miyagi A, Ando T, Lyubchenko YL (2011) Dynamics of Nucleosomes Assessed with Time-Lapse High-Speed Atomic Force Microscopy. *Biochemistry* 50(37):7901–7908.
- 13. Tims HS, Gurunathan K, Levitus M, Widom J (2011) Dynamics of Nucleosome Invasion by DNA Binding Proteins. *J Mol Biol* 411(2):430–448.
- 14. Hodges C, Bintu L, Lubkowska L, Kashlev M, Bustamante C (2009) Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* 325(5940):626–8.
- 15. Shundrovsky A, Smith CL, Lis JT, Peterson CL, Wang MD (2006) Probing SWI / SNF remodeling of the nucleosome by unzipping single DNA molecules. *Nat Struct Mol Biol* 13(6):549–554.
- 16. Dechassa ML, et al. (2011) Structure and Scm3-mediated assembly of budding yeast centromeric nucleosomes. *Nat Commun* 2:313.
- 17. Culkin J, de Bruin L, Tompitak M, Phillips R, Schiessel H (2017) The role of DNA sequence in nucleosome breathing. *Eur Phys J E* 40:106.
- 18. Tims HS, Gurunathan K, Levitus M, Widom J (2011) Dynamics of Nucleosome Invasion by DNA Binding Proteins. *J Mol Biol* 411(2):430–448.

SUPPLEMENTARY FIGURES

Figure S1: Our assay probes the spontaneous movements of nucleosomes with high sensitivity and without perturbing them. (a) Cumulative probability for the standard deviation in the position of nucleosomes reconstituted on the Widom 601 positioning sequence using canonical histones (blue) and nucleosomes reconstituted on a DNA sequence containing the –157/+43 segment of *Lhb* with canonical histone octamers (green) or histone octamers containing H2A.Z (red). The cumulative SD for the corresponding alignment sequence for each type of nucleosome is shown as a dashed line of the respective color, and is $\leq 2 \pm 1$ bp for the majority of the experiments (median \pm s.e.), setting the resolution of our method. (b) Ensemble-averaged mean squared displacement (MSD) for nucleosomes reconstituted on *Lhb* TSS (green) or the 601 nucleosome positioning sequence (blue), as a function of time. Data is analyzed and shown as in Fig. 1f. (c) The frequency of probing does not affect nucleosome mobility. Ensemble-averaged MSD after 5 min for different cycle times, measured with the same loading rate (12 pN/s), for H2A (green) and H2A.Z (red) nucleosomes reconstituted on *Lhb* DNA. Data shown as mean ± s.e.m; *P<0.05, ****P<0.0001, student's ttest. (d) The force loading/relaxation rate does not affect nucleosome mobility. Ensembleaveraged MSD after 5 min for experiments with a 30 s cycle time, performed with two different loading rates (12 pN/s and 3 pN/s), for H2A.Z nucleosomes. The number of experiments for each case is shown in Table S3.

Figure S2: Histograms for the total length of the experiments. (a) H2A nucleosomes (b) H2A.Z nucleosomes (c) H2A nucleosomes + Egr-1 (d) H2A.Z nucleosomes + Egr-1. The total number of experiments corresponding to each case is shown in Table S3.

Figure S3: Additional experimental traces, presented as in Fig. 1d. (a) H2A nucleosomes (b) H2A.Z nucleosomes (c) H2A nucleosomes + Egr-1 (d) H2A.Z nucleosomes + Egr-1. The number of experiments corresponding to each is shown in Table S3.

Figure S4: Potential of mean force for the nucleosomes' movement, calculated from the individual (a) *Lhb* H2A and (b) 601 H2A traces that are longer than 7 min. The calculated values where filtered with a 3 points window running average. $n_{Lhb}=15$, $n_{601}=14$. The standard deviation of the mean position, which is an indication of the degree of localization during the reconstitution, is 1.9 ± 0.6 bp for 601 and 32 ± 7.2 bp for *Lhb.*

Figure S5: Two representative traces for the coverage of the TF binding site due to H2A.Z nucleosome mobility, presented as in Fig. 3d.

Figure S6: (a) Ensemble averaged MSD for *Lhb* H2A.Z nucleosomes in the presence (purple) and absence (red) of Egr-1, calculated for traces that are longer than 10 min. Data is analyzed and shown as in Fig. 2c. The number of experiments corresponding to each case is shown in Table S3. (b) Probability distribution function of the step size for experiments with H2A.Zcontaining nucleosomes, whose 5' end is initially positioned between -160/-140, in the absence (yellow) or presence (purple) of Egr-1. Data is analyzed and shown as in Fig. 4b. Note, that two possibilities can explain the absence of large (>20 bp) steps, as shown in Fig. 4b: First, the majority of nucleosomes are reconstituted on the center of the *Lhb* TSS sequence (Fig. 4a), not in the region proximal to the 5' end of the construct. As >20 bp steps are a rare event (Fig. 2d), it is possible that the representation of these events is limited in Fig. S6b because of the smaller size of the data set. The second, and perhaps more intriguing scenario, is that nucleosome mobility may be affected by the properties of the local DNA sequence, which can affect the distribution of step size and restrict the formation of steps larger than 20 bp.

Figure S7: Kinetics of Egr-1 binding. (a) A DNA construct containing the *Lhb* TSS DNA or a DNA segment which includes the consensus binding sequence for Egr-1 was unzipped until reaching the extension corresponding to the Egr-1 binding site. The construct was then held under tension, letting the DNA fluctuate between locally 'open' and 'closed' states. Binding of Egr-1 to the DNA stabilizes the closed conformation, leading to a sudden repression of the fluctuations, which lasts until Egr-1 dissociation. The time difference between binding and dissociation (t_{bound}) at [Egr-1] = 500 nM was calculated as previously described (2). (b) An example of an experimental trace showing breathing fluctuations of the DNA in the absence (upper trace) or presence (lower trace) of Egr-1. In presence of Egr-1, breathing fluctuations are interrupted by sudden transitions to a 'closed' state, indicating binding of a TF to DNA. (c) The dissociation constant (k_{off}) was calculated for *Lhb* TSS and the consensus binding site. Data are shown as mean \pm s.e.m., n_{consensus} = 25 and n_{Lhb} =399. ****P $<$ 0.0001, two-sample student's *t*-test.

Figure S8: Step size probability distribution function for nucleosomes reconstituted on a Widom 601 positioning sequence using canonical histones (a, blue) and a DNA sequence containing the –157/+43 *Lhb* segment reconstituted with canonical histone octamers (b, green) or histone octamers containing H2A.Z (c, red). The probability distribution for the corresponding alignment sequence (grey) is shown for a reference. Data is analyzed as in Fig. 1e and shown on a logarithmic scale. The number of experiments corresponding to each case is shown in Table S3.

Figure S9: The integrity of the nucleosomes is conserved throughout the mobility assay. (a) Typical unzipping signatures for H2A (green) and H2A.Z (red) nucleosomes obtained from the last experimental cycle, at which histone-DNA interactions are fully and irreversibly disrupted. Δx is defined as the distance between region 1 (corresponding to interactions of the H2A/H2B dimer with DNA) and region 2 (which corresponds to interactions of H3/H4 tetramer at the dyad). The unzipping curve for naked DNA (black curve) is shown for a reference. (b) The measured Δx for H2A (38.5 ± 2.5 bp) and H2A.Z (39.2 ± 2.1 bp) is similar, and also consistent with the previously published \sim 40 bp value (5). Data shown as mean \pm s.e.m; n_{H2A}=21 and $n_{H2A.Z}=16$.

Figure S10: Estimation of the nucleosome's 5' edge position. Once the position of region 1 is measured, we subtract 35 bp to obtain position of the nucleosome's starting position (e.g. its 5' edge). The position of the interaction regions of the different histones with the DNA is shown in the inset. The experimental trace shown here is identical to the one shown in Fig. 1c.

Figure S11: Force-extension curves for repetitive unzipping experiments of nucleosomal (green) and non-nucleosomal (blue) DNA in presence of Egr-1. Unzipping curves for nucleosomal DNA without Egr-1 (orange) are shown for comparison. The black curves correspond to naked DNA.

SUPPLEMENTARY TABLES

Table S1: Primers used for PCR amplification of DNA sequences

Table S2: DNA sequence of the full unzipping construct. Grey: Alignment sequence. Green: Nucleosome position. Yellow: Egr-1 binding site.

