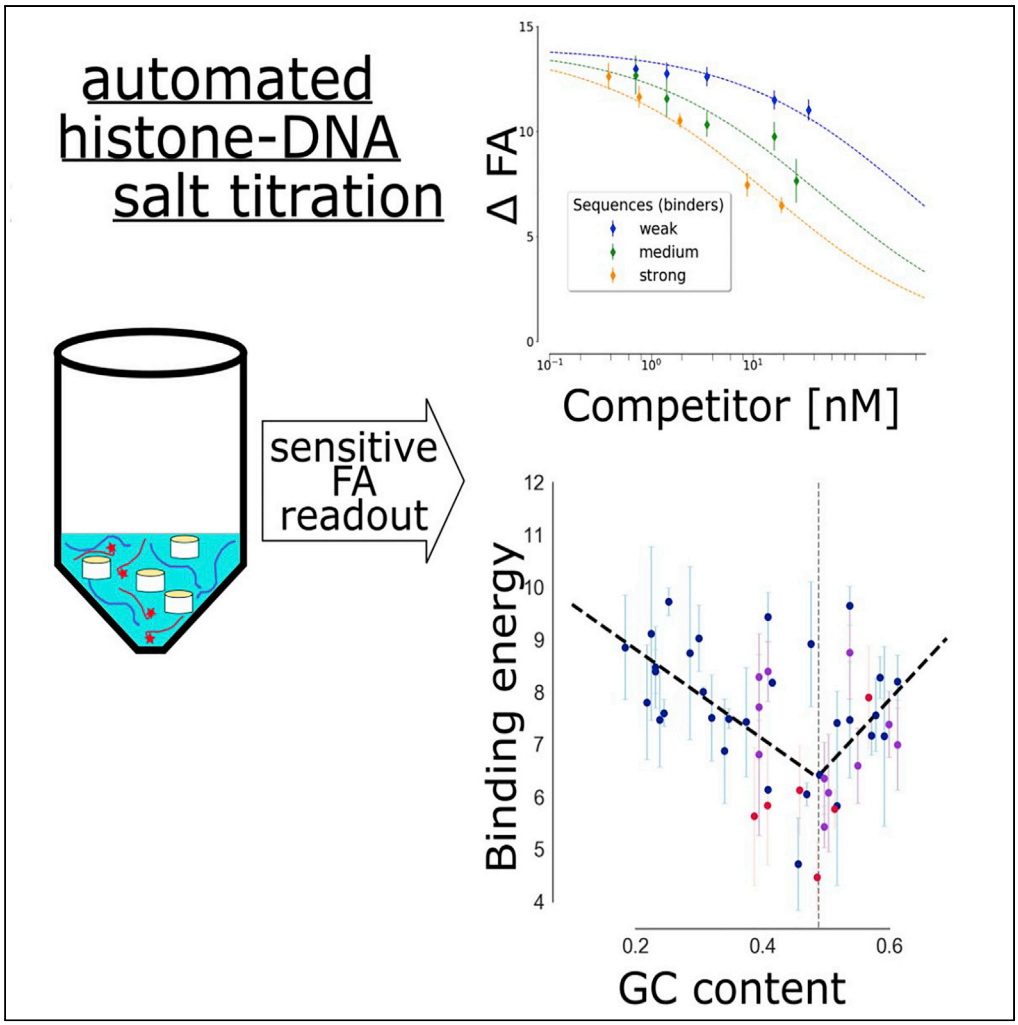
iScience

CellPress

Article

Sensitive Automated Measurement of Histone-DNA Affinities in Nucleosomes



Max Schnepf, Claudia Ludwig, Peter Bandilla, Stefano Ceolin, Ulrich Unnerstall, Christophe Jung, Ulrike Gaul

jung@genzentrum.lmu.de

HIGHLIGHTS

Robotics permits full titration series to measure histone-DNA binding affinities

Fluorescence anisotropy used as a fast, sensitive readout of bound/unbound DNA

Free energies span three orders of magnitude, less for naturally occurring sequences

GC content is a major determinant of measured binding free energies

Schnepf et al., iScience *23*, 100824
February 21, 2020
https://doi.org/10.1016/j.isci.2020.100824

iScience

CellPress

Article

Sensitive Automated Measurement of Histone-DNA Affinities in Nucleosomes

Max Schnepf,[1] Claudia Ludwig,[1] Peter Bandilla,[1] Stefano Ceolin,[1] Ulrich Unnerstall,[1] Christophe Jung,[1,2,*] and Ulrike Gaul[1]

SUMMARY

The DNA of eukaryotes is wrapped around histone octamers to form nucleosomes. Although it is well established that the DNA sequence significantly influences nucleosome formation, its precise contribution has remained controversial, partially owing to the lack of quantitative affinity data. Here, we present a method to measure DNA-histone binding free energies at medium throughput and with high sensitivity. Competitive nucleosome formation is achieved through automation, and a modified epi-fluorescence microscope is used to rapidly and accurately measure the fractions of bound/unbound DNA based on fluorescence anisotropy. The procedure allows us to obtain full titration curves with high reproducibility. We applied this technique to measure the histone-DNA affinities for 47 DNA sequences and analyzed how the affinities correlate with relevant DNA sequence features. We found that the GC content has a significant impact on nucleosome-forming preferences, but 10 bp dinucleotide periodicities and the presence of poly(dA:dT) stretches do not.

INTRODUCTION

Eukaryotes organize their genomes by wrapping their DNA around a complex of basic proteins called histones. Approximately 147 base pairs of DNA are wrapped 1.7 times around a histone octamer to form a nucleosome, which constitutes the basic unit of chromatin (Khorasanizadeh, 2004). Nucleosomes show a clear organization with respect to the DNA sequence especially around promoters, where a nucleosome-depleted region at the transcription start site is typically flanked by well-positioned nucleosomes on either side, referred to as the −1 and +1 nucleosomes, and by an array of regularly spaced nucleosomes downstream over the gene body (Jiang and Pugh, 2009; Lai and Pugh, 2017). It is commonly accepted that the *in vivo* positioning of nucleosomes is the result of the interplay of multiple determinants, which include the DNA sequence, but also the action of nucleosome remodelers, transcription factors, and other DNA-binding proteins/complexes such as the RNA polymerase II transcription machinery (Klemm et al., 2019). Among sequence parameters influencing nucleosome formation, the best studied are the GC content (Fenouil et al., 2012; Tillo and Hughes, 2009) and the base pair (bp) periodicity of flexible dinucleotides forming the contact sites of the nucleosomal DNA with the histone octamer (Drew and Calladine, 1987; Jin et al., 2016; Klug and Lutter, 1981; Shrader and Crothers, 1990; van der Heijden et al., 2012). Owing to the double helical nature of DNA, the same face contacts the histone complex every ten to eleven base pairs; nucleosome formation is thus facilitated by the periodical occurrence of alternatingly flexible (like AT, AA, or TT) and stiff (like GC) dinucleotide steps (kink-and-slide model) (Vasudevan et al., 2010). In addition, it is believed that the presence of short homopolymeric stretches of deoxyadenosine nucleotides referred to as poly(dA:dT) or dA:dT tracks, which are intrinsically stiff and are frequently found in nucleosome-depleted regions, impedes nucleosome formation (Jin et al., 2018; Raveh-Sadka et al., 2012; Segal and Widom, 2009). Overall, much effort has been devoted in recent years to characterize nucleosome sequence preferences *in vitro* (Krietenstein et al., 2012; Segal et al., 2006) and *in vivo* (Jin et al., 2018; Kaplan et al., 2010) and to predict nucleosome positions in the genome on this basis (Segal et al., 2006; Tillo et al., 2010). However, the precise degree to which the underlying sequence directs nucleosome positioning, and which sequence features are most important, is still a matter of debate (Jin et al., 2018; Kaplan et al., 2010; Struhl and Segal, 2013; Zhang et al., 2009). This is in part due to the fact that most studies have relied on MNase digestion of genomic DNA followed by deep sequencing and fragment counting to derive nucleosome occupancy values. This approach is subject to confounding effects of biological activity and/or the inherent biases of the MNase enzyme. It would thus be immensely useful to measure histone-DNA binding affinities directly, without other intervening features. However, few such data exist, mainly because of the lack of efficient experimental techniques.

[1]Gene Center and Department of Biochemistry, Center for Protein Science Munich (CIPSM), Ludwig-Maximilians-Universität München, Feodor-Lynen-Strasse 25, 81377 München, Germany

[2]Lead Contact

*Correspondence: jung@genzentrum.lmu.de

https://doi.org/10.1016/j.isci.2020.100824

The most widely used method for determining histone-DNA binding free energies *in vitro* was pioneered by Schrader, Crothers, and Widom (Lowary and Widom, 1998; Shrader and Crothers, 1990). In this approach, nucleosomes are typically reconstituted from purified core histones and DNA of mononucleosomal length by dialysis, or alternatively by stepwise dilution. A competition experiment is conducted using a mixture of the DNA of interest and low amounts of (usually radio- or fluorescently) labeled DNA, which serves as a reference to compare the nucleosome-forming capacity of different DNA sequences. Reconstituted samples are analyzed on polyacrylamide or agarose gels by an electrophoretic mobility shift assay (EMSA) (Thastrom et al., 1999) to calculate the fraction of reference DNA that reconstitutes into nucleosomes in a given DNA mixture. In this fashion, *relative* affinities (free energies) of histone octamers to differing DNA fragments can be determined. The assay works reliably but suffers from significant limitations: both the nucleosome reconstitution and the EMSA readout steps are time consuming and difficult to parallelize. This entails that histone-DNA free energies are usually determined using only a single concentration per sequence, raising issues regarding accuracy and reproducibility. As a result, affinity data are currently only available for a relatively limited number of sequences. Moreover, most studies have focused on artificially designed nucleosomal DNA sequences with strong, non-physiological binding properties (Eslami-Mossallam et al., 2016), whereas native genomic sequences have not been investigated extensively. Thus, there is a clear need for more accurate and comprehensive measurements of histone-DNA free energies.

In the current study, we developed a method to measure histone-DNA binding free energies in nucleosomes with high reproducibility and at medium throughput. Our technique is based on the classical approach but offers substantial improvements: (1) we carry out the competitive reconstitution of nucleosomes by small dilution steps using an automated liquid handling system and (2) we determine the fraction of bound DNA by measuring fluorescence anisotropy (FA) with an adapted epifluorescence microscope, following the approach we recently reported for measuring transcription factor-DNA binding energies (Jung, 2018, 2019). The high parallelization of the nucleosome reconstitution and the fast and sensitive fluorescence readout allowed us to obtain full titration curves for each individual histone-DNA interaction instead of single concentration measurements, resulting in a more accurate determination of binding free energies. We demonstrate the utility of this approach by measuring histone-DNA binding free energies for 47 different DNA sequences, including *Drosophila melanogaster* (*D. mel*) genomic nucleosomal sequences, synthetic DNA sequences derived from *D. mel* enhancers, and additional nucleosomal DNA sequences tested in previous studies (Cao et al., 1998; Filesi et al., 2000; Shrader and Crothers, 1990; Thastrom et al., 1999). We show that the free energies of nucleosome formation can be measured accurately and cover a wide dynamical range. Furthermore, we explored how the free energies correlate with DNA features such as the GC content, the 10 bp periodicity of flexible and stiff dinuleotides, and the number of short poly(dA:dT) stretches. We found GC content to be the most predictive feature in our data, explaining ~30% of the variation of the free energies.

RESULTS

Automated Assay to Determine Free Energies of Nucleosome Formation

We determined the free energies of nucleosome formation by titration of unlabeled DNA sequences (147–300 bp in length) competing for nucleosome reconstitution with a fluorescently labeled DNA reference in low amount (Figure 1A). Histones and competing DNAs are initially mixed at high salt concentration; in a slow dilution process, buffer is added in small steps, thus gradually increasing the interaction strength. The use of an automated liquid handling system permits carrying out this nucleosome reconstitution over a long period of time (12 h, typically overnight), helping to approximate thermodynamic equilibrium, and greatly improves the reproducibility and throughput of the assay. To ensure reproducibility of the reconstitution reaction, we had to limit evaporation (which typically occurs at borders and edges of well plate containers) and provide for a stable temperature. To this end, we designed and fabricated a metal block (Figure S5) accommodating up to 42 individual low protein binding tubes with a heated lid (Figure 1B and Transparent Methods). The metal block improves temperature stability and uniformity during the nucleosome formation process, while the heated lid reduces evaporation by preventing condensation at the lid.

The readout of the fraction of bound versus unbound fluorescently labeled reference DNA was carried out using FA (Roehrl et al., 2004) (Figure 1C) instead of the typical EMSA, thus offering the advantages of a fast and sensitive fluorescence readout. In brief, FA measures the rotational speed of a fluorescently labeled

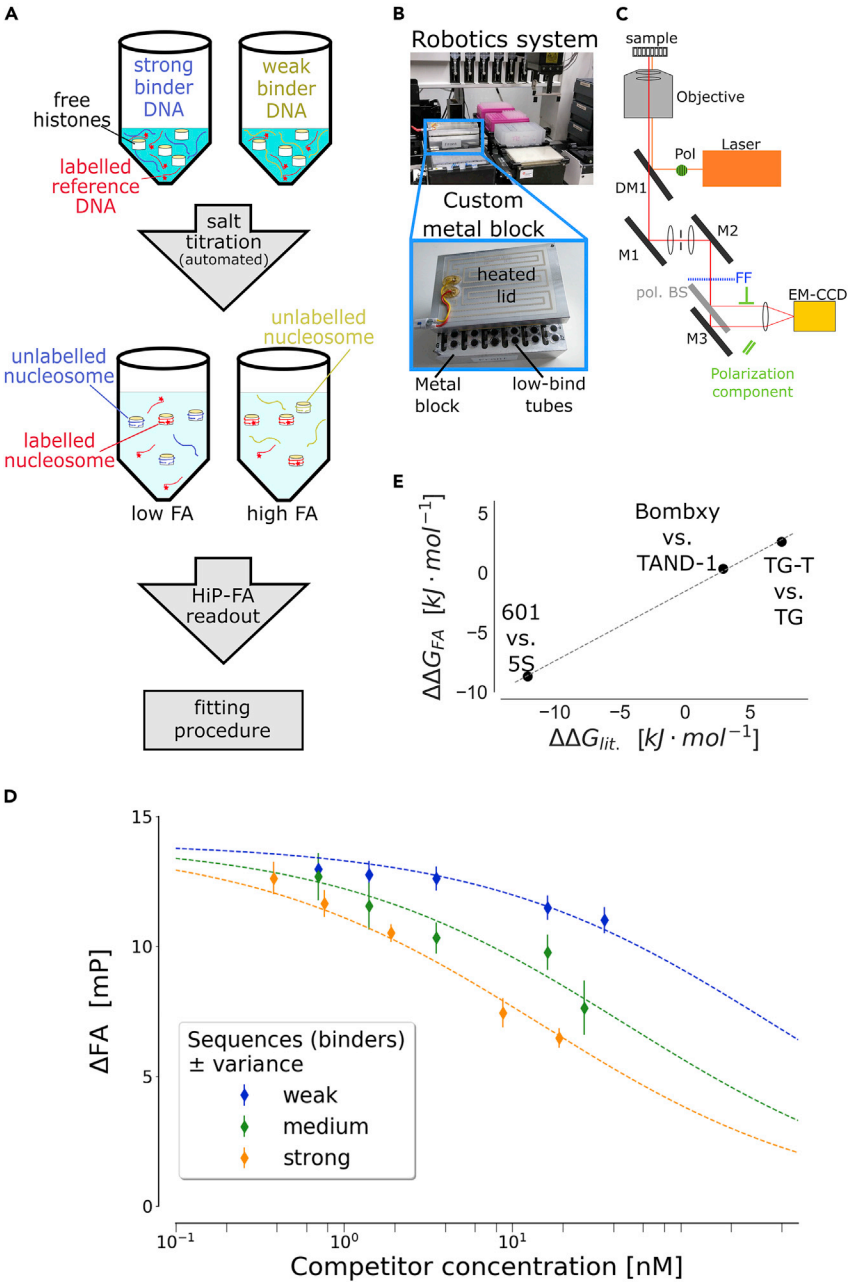


Figure 1. Competitive Nucleosome Reconstitution and Binding Assay

(A) Schematic representation of the assay workflow.

(B) Image of the robotic system used, with enlarged image of the custom metal block (see also Figure S5) with its heated lid.

(C) Schematic depiction of the fluorescence microscopy setup used for the FA readout.

(D) Histone-DNA affinity single titration curves for three different competitor sequences, together with their corresponding fits (dashed lines), for a weak, medium, and strong binder as indicated. Error bars refer to the standard deviation of the FA measurements, which were used to weight the individual points in the fitting procedure (Transparent Methods).

(E) Assay validation, using DNA sequences measured in previous studies (Cao et al., 1998; Filesi et al., 2000; Shrader and Crothers, 1990; Thastrom et al., 1999). The relative free energies determined in previous studies are plotted against the corresponding values obtained in this study; dotted line shows linear regression; Pearson correlation coefficient R = 0.99.

molecule: high FA indicates the presence of larger, and therefore slowly rotating, molecular complexes in the solution, in this case nucleosomes with incorporated fluorescently labeled reference DNA. If an unlabeled competitor DNA outcompetes this labeled reference for histone binding, the FA will decrease, as a higher proportion of the fluorescently labeled reference DNA molecules are unbound and thus rotating faster. Different FA levels (Figure S1) can therefore be used to calculate the fractions of bound versus unbound labeled reference DNA.

After nucleosome reconstitution, the samples are transferred to 96-well microscopy plates and FA is measured in each well using the microscopy setup described for HiP-FA in Jung, 2018, 2019 (Figure 1C). By performing the reconstitution with different unlabeled competitor concentrations, we obtain a full titration curve for each DNA sequence. The data can be fitted using the Hill equation, as shown for a weak ($\Delta\Delta G$ = 9.7 kJ·mol^{-1}; Dmel08), a medium ($\Delta\Delta G$ = 7.2 kJ·mol^{-1}; Dmel28), and a strong ($\Delta\Delta G$ = −2.2 kJ·mol^{-1}; 601) competitor sequence (Figure 1D).

To validate our assay, we measured seven nucleosomal sequences that had been tested in other studies (601, TGGA-2, TAND-1, TG, Bombyx, 5S, and TG-T [Cao et al., 1998; Filesi et al., 2000; Shrader and Crothers, 1990; Thastrom et al., 1999]). As shown in Figure 1E, our results are in excellent agreement with the previously measured free energies, calculated relative to their respective reference sequences. To evaluate our FA readout, we also measured affinities for selected sequences by EMSA. EMSA with fluorescent labels is prone to quenching effects; nevertheless, we find a reasonable agreement between these measurements and the affinities obtained by FA (Figure S2).

Applying the Method to Genomic and Synthetic Nucleosomal DNA Sequences

Most studies measuring free energies of nucleosome formation *in vitro* focused on DNA sequences that were selected or designed to cover a large range of affinities. In fact, the strongest known binders, like the well-known 601 sequence, are the result of heavy selection and are not found in native genomic DNA. A few naturally occurring sequences have been tested, showing lower affinities and a smaller dynamical range than synthetic sequences (Thastrom et al., 1999), but to date no comprehensive direct measurement of histone affinities of genomic nucleosomal sequences has been conducted. Taking advantage of the throughput and accuracy of our technique, we therefore decided to test 29 endogenous nucleosomal DNA sequences from *D. mel* (denoted Dmel01 - Dmel29); the sequences were selected from −1 nucleosomes as determined by MNase-Seq (unpublished data), which are well positioned but typically contain fewer cis-regulatory sequence elements than the +1 nucleosomes and whose positioning is presumably less affected by biological activity (Mavrich et al., 2008a, 2008b). The sequences were chosen randomly, but such that a range of GC contents from ∼20% to 60% was represented. A second group of tested sequences (denoted Synt01 - Synt11) was derived from synthetic enhancers driving expression in *D. mel* embryos.

In total, we obtained, after quality control (Transparent Methods), 147 titrations of 47 different DNA sequences (three measurements per sequence on average; Figure 2). All free energies ($\Delta\Delta G$) were determined relative to our reference sequence, a weakened version of the 601 sequence (601_dpl), which has the same GC content as the original 601 but shows less pronounced 10 bp dinucleotide periodicity patterns and is thus well suited to measure weaker competitor sequences (Transparent Methods for details). Overall, our measured free energies range over 12 kJmol·mol^{-1} (Figure 2), which is similar to the dynamical range reported by Thastrom et al. (1999); reproducibility is high, with a mean coefficient of variation (CV) of 24%. Interestingly, both the 601 sequence and its weakened derivative 601_dpl are strong outliers, and the free energies of all the other sequences are distributed over a much smaller range of 5.3 kJ·mol^{-1}.

Dependency on GC Content and Other DNA Features

To gain insight into the parameters contributing to histone-DNA affinities, we correlated our data with several known sequence determinants of nucleosome formation, namely, the GC content and sequence features affecting bending: the 10 bp periodicities of flexible (WW where W is A or T) and stiff (SS where S is G or C) dinucleotides and the presence of homopolymeric sequences poly(dA:dT).

We first plotted the free energies of nucleosome formation as a function of the GC content for all investigated sequences (Figure 3). We observe a decrease of $\Delta\Delta G$ (i.e., increased affinity) with GC contents for GC content values <∼0.5 and an inverted trend with increasing $\Delta\Delta G$s at higher GC content values. Thus, our

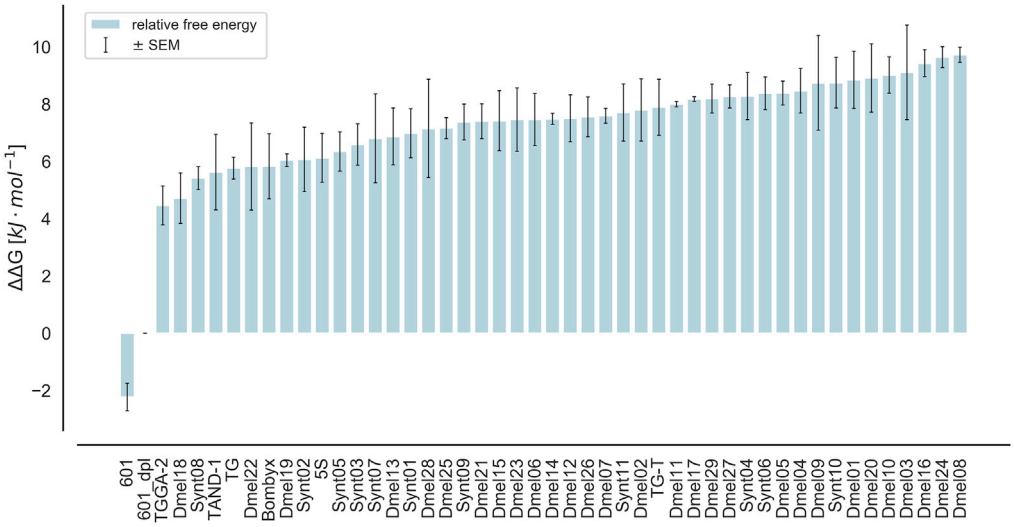


Figure 2. Overview of All Measured Histone-DNA Affinities

All affinities are shown as free energy of nucleosome formation relative to the 601_dpl reference sequence; mean ± SEM over, on average, three replicates. Names are taken from the original publications or indicate the different groups of sequences: Dmelxx: selected −1 nucleosomes from *D. mel*, Syntxx: synthetic enhancer constructs driving expression during *D. mel* embryo development.

data indicate that binding free energies are strongly influenced by the GC content, but in a non-monotonous fashion: there appears to be an optimal GC content value of ∼0.5. Interestingly, sequences from all three groups (Figure 3) follow this falling and rising pattern. The simplest model to describe the behavior of ΔΔG with respect to the GC content is given by a segmented linear regression with two segments, intersecting at the optimal GC content. We fitted the data using this model and found a relatively good correlation with a Pearson correlation coefficient (R) of 0.54 ($p = 0.0001$) and an optimal GC content value of 0.49, corresponding to a minimum ΔΔG value of 6.4 kJ·mol^{-1}. Thus, the GC content alone explains ∼30% (R^2) of the variance in the data. Note that for this analysis the extreme values of the two 601 variants were excluded, although both have a near-optimal GC content.

Another sequence feature that is thought to impact histone-DNA affinities is the periodic occurrence of flexible and rigid dinucleotides, aligning with the periodic changes in orientation of the DNA double strand facing the histones. In fact, the 601 sequence, which lies outside the general trend we observed with respect of GC content (Figure 3), contains very strong dinucleotide periodicities (Lowary and Widom, 1998), as do other specifically engineered sequences. Thus, we sought to determine whether dinucleotide periodicities influence the free energies among the pool of our measured sequences (owing to their extreme values, the 601 variants were again excluded from this analysis). We started by computing the 10 bp periodicity of WW (A or T) and SS (C or G) dinucleotides for all sequences using an autocorrelation function (Figures 4 and S3A and Transparent Methods) (Cui and Zhurkin, 2010). For both dinucleotide groups, we found relatively low 10 bp autocorrelations; the same result was obtained using Fourier transform, the alternative commonly used method (data not shown). A closer inspection of the individual autocorrelation values (Figure 4) revealed that the small average 10 bp autocorrelation observed is driven by a few literature sequences that in fact had been selected for their strong 10 bp dinucleotide periodicities (besides 601: Bombyx, TG, or TG-T). By contrast, most of the genomic −1 nucleosome and synthetic enhancer sequences exhibit very low 10 bp autocorrelation values. Note that this is not simply due to our sequence selection: the dinucleotide autocorrelations for all −1 nucleosomes and for all nucleosomes in *D. mel* show the same distribution of values (Figures S3B and S3C). When we then analyzed the influence of this sequence feature on the free energy of nucleosome formation by plotting the autocorrelation values versus ΔΔG (Figure 4), we found only a very weak correlation ($R = -0.15$ and -0.16 for WW and SS, respectively), which is driven by the artificial (=literature) sequences (without them $R = 0.06$ [SS] and 0.07 [WW]). This suggests that, at least among native nucleosomal sequences, dinucleotide periodicities are not a prominent feature and have no discernible impact on histone affinities.

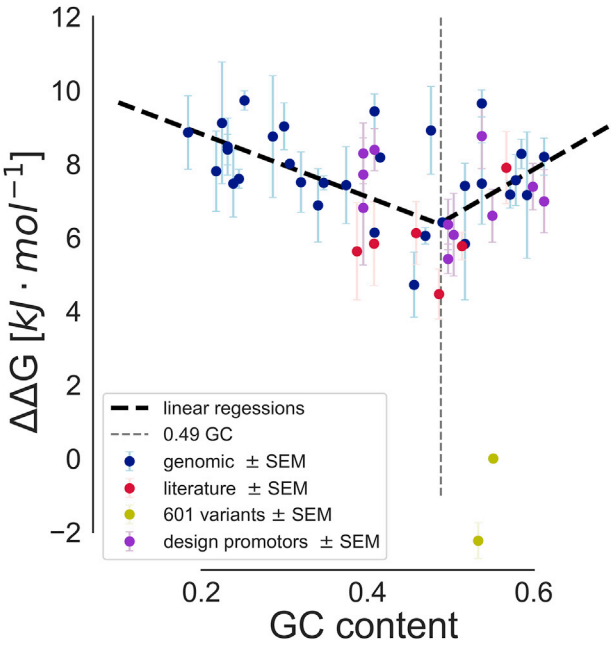


Figure 3. Correlation between Free Energy of Nucleosome Formation and GC Content

Scatterplot showing the relative free energies against the GC content for all investigated nucleosomal sequences. Data points are color coded according to their provenance; note the extreme values of the two 601 variants (yellow). Linear regression was performed over two sections for sequences with GC contents < or >0.49, respectively, resulting in a Pearson correlation coefficient between predicted and observed free energies of 0.54.

Finally, we evaluated the influence of poly (dA:dT) tracks, which are thought to disfavor nucleosome formation, on our measured histone binding energies (Figure S4). We do not find any significant effect, perhaps due to the low number of sequences containing long tracks (Figures S4A and S4B). However, for some of the sequences from the group of synthetic enhancers, in which this feature was systematically varied, we observe a weak effect of very short dA:dT tracks (n = 2 and 3) on ΔΔG, with changes of maximally 2 kJ·mol^{-1} (Figure S4C).

DISCUSSION

We developed an assay to determine histone-DNA affinities using salt titration and fluorescence anisotropy microscopy. Aided by automated liquid handling, we achieve slow gradual changes in the salt concentration by dilution in small steps over a long period of time and thus ensure that nucleosome formation takes place at conditions close to equilibrium. The readout is performed in an automated fashion using fluorescence anisotropy and allows the reproducible measurement of relatively small effects. This medium-throughput pipeline permits carrying out a full titration series, which is typically not possible in other competitive nucleosome formation assays. Since the analysis thus relies on a fitting procedure over a larger dataset, the determination of the free energies is more robust and it allows introducing a quality control step for the detection of experimental errors (see Transparent Methods).

Most recent efforts to systematically determine nucleosome sequence preferences have relied on genome-scale assays, where genomic DNA is digested by MNase into (typically mono-nucleosomal) fragments, which are then analyzed by high-throughput sequencing. For any given DNA region, the number of fragments covering it provides a measure of nucleosome occupancy at that position; these occupancy values in turn can then be used to derive nucleosome sequence preferences. However, when conducted using *in vivo* chromatin, the results of this procedure will reflect not only the inherent sequence preferences of the histone octamer, but also the action of the plethora of other biological factors influencing both nucleosome occupancy and positioning. Even if conducted on nucleosomes reconstituted *in vitro*, the imprecision and biases in cutting by the MNase will have a significant impact (Jin et al., 2018). Moreover, the result for a given stretch of genomic sequence will always also be influenced by the nucleosome binding properties of the surrounding sequence and corresponding steric hindrance effects.

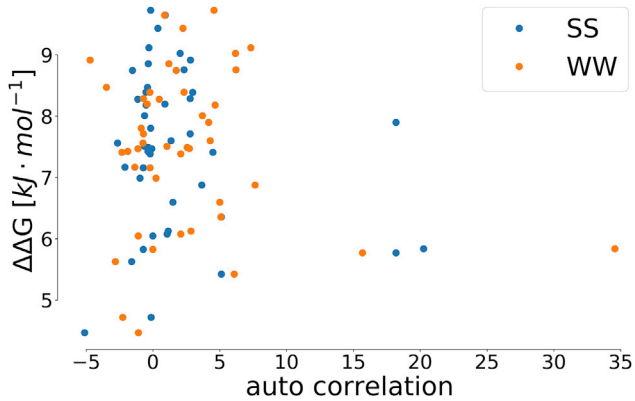


Figure 4. Relationship between the Free Energy of Nucleosome Formation and the Corresponding Autocorrelation at a Shift of 10 bp

No significant correlation can be observed for our sequences (R = −0.15 and −0.16 for WW and SS, respectively). The extreme points (autocorrelation>15) originate from literature-derived sequences selected for their strong periodicities (Bombyx, TG, TG-T).

By contrast, our assay can measure the "pure" histone binding affinity of any given piece of DNA precisely, without influence of other biological factors, chromatin features, or neighboring sequences, and it can do so with a throughput that is sufficient to comprehensively investigate the sequence determinants of nucleosome formation. Our approach cannot match the scale of the genomic sequencing-based assays and does rely on prior knowledge to select the sequences to be tested, but in permitting to isolate and measure the purely sequence-dependent aspect of nucleosome formation, it provides a valuable complement - similar to the way *in vitro* measurements of transcription factor-DNA affinities complement the *in vivo* tracking of transcription factor occupancies using ChIP-Seq.

Although our assay can capture histone-DNA affinities ranging over more than 2.5 orders of magnitude (in a non-log scale), we obtain a more limited dynamical range of only one order of magnitude for the histone binding free energies of the native genomic and synthetic enhancer sequences, consistent with observations in previous studies (Thastrom et al., 1999). Although we cannot be certain that our selected sequences represent the entire range of natively occurring nucleosomes, this suggests that the biological impact of sequence preferences on nucleosome positioning *in vivo* (Kaplan et al., 2010) is caused by relatively small variations in absolute affinities.

We found the GC content to be the one feature contributing significantly to the free energies of nucleosome formation for the investigated DNA sequences, explaining ~30% of the variance in the data (Figure 3). This finding is in agreement with studies of genome-wide nucleosome occupancy based on MNase digestion (Fenouil et al., 2012; Tillo and Hughes, 2009). Interestingly, we also observe that the relationship is non-monotonous, with an optimum GC content of ~0.5. Although some genomic studies found a monotonous dependency (Tillo et al., 2010), others do support the idea of an optimum GC content (Fenouil et al., 2012; Wang et al., 2014): Fenouil et al. (2012) examined GC content preferences of +1 nucleosomes and found an occupancy maximum at GC contents of 0.44 (*in vivo*) and 0.58 (*in vitro*), respectively; Wang et al. (Wang et al., 2014) found the occupancy of nucleosomes in exons to peak around a GC content of 0.5. However, the GC dependency of nucleosome occupancy observed in these genomic studies might be attributable to either biological activity (Tillo et al., 2010) or to the cutting biases of the MNase (Jin et al., 2018). Our data now show, with an independent method that captures the pure sequence preference of the histone-octamer, that GC content is indeed an important driving force of nucleosome formation.

Numerous studies have focused on 10 bp dinucleotide periodicity, which affects the bending properties of the DNA, as a major determinant of nucleosome formation. Although this feature certainly appears dominant in many designed sequences, with the 601 clone the strongest known case, we find neither strong intrinsic periodicities in our random selection of genomic sequences and synthetic enhancers nor any significant correlation between the autocorrelation values for these sequences and our ΔΔG measurements (Figure 4). This suggests that, in naturally occurring (euchromatic) DNA, dinucleotide periodicity is not

an important parameter driving differential histone-octamer binding. Finally, we investigated the influence on nucleosome formation of the occurrence of poly (dA:dT) stretches, which are thought to inhibit histone binding. Although globally we did not find an effect, perhaps due to the paucity of longer poly (dA:dT) tracks in our sample, we did observe, for a sub-group of synthetic enhancer sequences that are highly similar and expected to bind with the same strength based on GC content, a slight but detectable decrease of histone-DNA affinity with growing numbers of nucleotides in short dA:dT tracks (Figure S4C). Although more detailed investigation is needed, this suggests that even short dA:dT tracks can lead to a weakening of histone-DNA interactions. Interestingly, results from a genome-wide reconstitution study in yeast suggest that rather than directly affecting the histone-DNA binding interaction, poly(dA:dT) tracks might have a role in facilitating nucleosome displacement (Krietenstein et al., 2016), which our assay by design cannot track.

In conclusion, we have developed a method to measure the free energies of nucleosome formation reliably and at medium throughput, and we pinpoint the GC content of the DNA sequence as an important determinant of the histone-DNA interaction. We believe that our method will be highly valuable in investigating and comprehensively characterizing the sequence preferences of histone octamers. It can potentially be extended to investigate more complex interactions like the interplay between pioneer transcription factors and nucleosomes.

Limitations of the Study

Although the approach described here is much higher in throughput than existing *in vitro* methods, it cannot match the scale of genome-wide sequencing-based assays and thus relies on prior knowledge to select the sequences to be tested. To further substantiate our findings regarding the sequence determinants of histone-DNA affinities, more extensive studies with a larger selection of genomic and synthetic sequences will be needed.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY

All data sequences investigated in this study are listed with their corresponding binding energies in Data S1. All *Python* codes are available upon request.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.100824.

ACKNOWLEDGMENTS

This work was supported by the DFG (large equipment grant #INST 86/1312-1 FUGG), the Collaborative Research Centers SFB646 (Regulatory Networks in Genome Maintenance and Development) and SFB1064 (Chromatin Dynamics), the Center for Integrated Protein Science Munich (CIPSM), and the Graduate School of Quantitative Biosciences Munich (QBM). U.G. acknowledges support by the Alexander von Humboldt Foundation (Alexander von Humboldt-Professorship).

The authors want to thank the group of Peter Becker (especially Sandro Baldi) for providing help and equipment for the embryo collections and the group of Philipp Korber for providing the histone purification protocol, guidance, as well as purified histones for the final set of experiments. We thank Michael Till for crafting the metal block used for the nucleosome formations. Finally, we thank Sabine Bergelt for her comments on the manuscript.

AUTHOR CONTRIBUTIONS

C.J. and U.G. developed the project; M.S., C.L., and C.J. designed the experiments; M.S. and C.L. performed the experiments; M.S., C.L., and P.B. implemented the protocol on the robotics system; S.C. designed and produced the synthetic enhancer sequences; M.S., U.U., and C.J. performed data analysis and wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no conflict of interests.

Received: October 3, 2019
Revised: December 12, 2019
Accepted: January 6, 2020
Published: February 21, 2020

REFERENCES

Cao, H., Widlund, H.R., Simonsson, T., and Kubista, M. (1998). TGGA repeats impair nucleosome formation. J. Mol. Biol. *281*, 253–260.

Cui, F., and Zhurkin, V.B. (2010). Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro. J. Biomol. Struct. Dyn. *27*, 821–841.

Drew, H.R., and Calladine, C.R. (1987). Sequence-specific positioning of core histones on an 860 base-pair DNA. Experiment and theory. J. Mol. Biol. *195*, 143–173.

Eslami-Mossallam, B., Schiessel, H., and van Noort, J. (2016). Nucleosome dynamics: sequence matters. Adv. Colloid Interface Sci. *232*, 101–113.

Fenouil, R., Cauchy, P., Koch, F., Descostes, N., Cabeza, J.Z., Innocenti, C., Ferrier, P., Spicuglia, S., Gut, M., Gut, I., et al. (2012). CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. Genome Res. *22*, 2399–2408.

Filesi, I., Cacchione, S., De Santis, P., Rossetti, L., and Savino, M. (2000). The main role of the sequence-dependent DNA elasticity in determining the free energy of nucleosome formation on telomeric DNAs. Biophys. Chem. *83*, 223–237.

Jiang, C., and Pugh, B.F. (2009). Nucleosome positioning and gene regulation: advances through genomics. Nat. Rev. Genet. *10*, 161–172.

Jin, H., Rube, H.T., and Song, J.S. (2016). Categorical spectral analysis of periodicity in nucleosomal DNA. Nucleic Acids Res. *44*, 2047–2057.

Jin, H., Finnegan, A.I., and Song, J.S. (2018). A unified computational framework for modeling genome-wide nucleosome landscape. Phys. Biol. *15*, 066011.

Jung, C., et al. (2018). True equilibrium measurement of transcription factor-DNA binding affinities using automated polarization microscopy. Nature Communications *9*, 1605.

Jung, C., et al. (2019). High Sensitivity Measurement of Transcription Factor-DNA Binding Affinities by Competitive Titration Using Fluorescence Microscopy. Journal of Visible Experiments *144*.

Kaplan, N., Moore, I., Fondufe-Mittendorf, Y., Gossett, A.J., Tillo, D., Field, Y., Hughes, T.R., Lieb, J.D., Widom, J., and Segal, E. (2010). Nucleosome sequence preferences influence in vivo nucleosome organization. Nat. Struct. Mol. Biol. *17*, 918–920.

Khorasanizadeh, S. (2004). The nucleosome: from genomic organization to genomic regulation. Cell *116*, 259–272.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nat. Rev. Genet. *20*, 207–220.

Klug, A., and Lutter, L.C. (1981). The helical periodicity of DNA on the nucleosome. Nucleic Acids Res. *9*, 4267–4283.

Krietenstein, N., Wippo, C.J., Lieleg, C., and Korber, P. (2012). Genome-wide in vitro reconstitution of yeast chromatin with in vivo-like nucleosome positioning. Methods Enzymol. *513*, 205–232.

Krietenstein, N., Wal, M., Watanabe, S., Park, B., Peterson, C.L., Pugh, B.F., and Korber, P. (2016). Genomic nucleosome organization reconstituted with pure proteins. Cell *167*, 709–721.e12.

Lai, W.K.M., and Pugh, B.F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. Nat. Rev. Mol. Cell Biol. *18*, 548–562.

Lowary, P.T., and Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. J. Mol. Biol. *276*, 19–42.

Mavrich, T.N., Ioshikhes, I.P., Venters, B.J., Jiang, C., Tomsho, L.P., Qi, J., Schuster, S.C., Albert, I., and Pugh, B.F. (2008a). A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. Genome Res. *18*, 1073–1083.

Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venters, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R.L., Schuster, S.C., et al. (2008b). Nucleosome organization in the Drosophila genome. Nature *453*, 358–362.

Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012). Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat. Genet. *44*, 743–750.

Roehrl, M.H., Wang, J.Y., and Wagner, G. (2004). A general framework for development and data analysis of competitive high-throughput screens for small-molecule inhibitors of protein-protein interactions by fluorescence polarization. Biochemistry *43*, 16056–16066.

Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. Curr. Opin. Struct. Biol. *19*, 65–71.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I.K., Wang, J.P., and Widom, J. (2006). A genomic code for nucleosome positioning. Nature *442*, 772–778.

Shrader, T.E., and Crothers, D.M. (1990). Effects of DNA sequence and histone-histone interactions on nucleosome placement. J. Mol. Biol. *216*, 69–84.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267–273.

Thastrom, A., Lowary, P.T., Widlund, H.R., Cao, H., Kubista, M., and Widom, J. (1999). Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. J. Mol. Biol. *288*, 213–229.

Tillo, D., and Hughes, T.R. (2009). G+C content dominates intrinsic nucleosome occupancy. BMC Bioinformatics *10*, 442.

Tillo, D., Kaplan, N., Moore, I.K., Fondufe-Mittendorf, Y., Gossett, A.J., Field, Y., Lieb, J.D., Widom, J., Segal, E., and Hughes, T.R. (2010). High nucleosome occupancy is encoded at human regulatory sequences. PLoS One *5*, e9129.

van der Heijden, T., van Vugt, J.J., Logie, C., and van Noort, J. (2012). Sequence-based prediction of single nucleosome positioning and genome-wide nucleosome occupancy. Proc. Natl. Acad. Sci. U S A *109*, E2514–E2522.

Vasudevan, D., Chua, E.Y., and Davey, C.A. (2010). Crystal structures of nucleosome core particles containing the '601' strong positioning sequence. J. Mol. Biol. *403*, 1–10.

Wang, L., Stein, L., and Ware, D. (2014). The relationships among GC content, nucleosome occupancy, and exon size. arXiv. https://arxiv.org/abs/1404.2487.

Zhang, Y., Moqtaderi, Z., Rattner, B.P., Euskirchen, G., Snyder, M., Kadonaga, J.T., Liu, X.S., and Struhl, K. (2009). Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. Nat. Struct. Mol. Biol. *16*, 847–852.

iScience, Volume *23*

Supplemental Information

Sensitive Automated Measurement of Histone-DNA Affinities in Nucleosomes

Max Schnepf, Claudia Ludwig, Peter Bandilla, Stefano Ceolin, Ulrich Unnerstall, Christophe Jung, and Ulrike Gaul

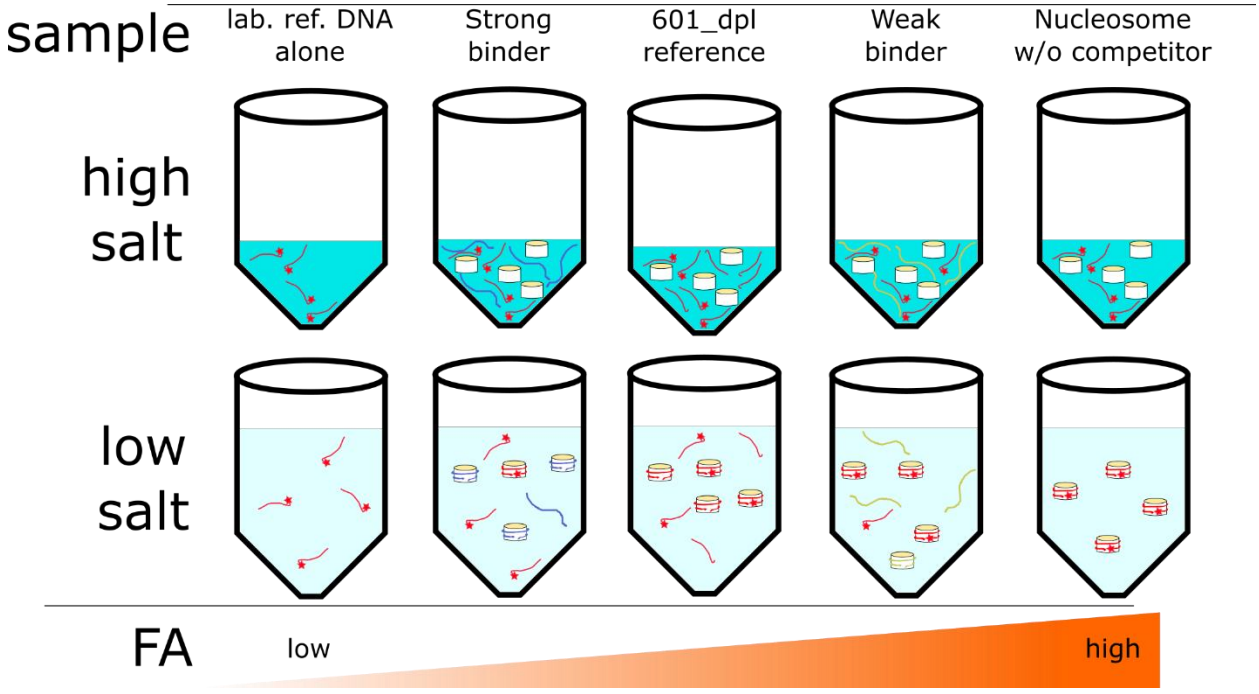
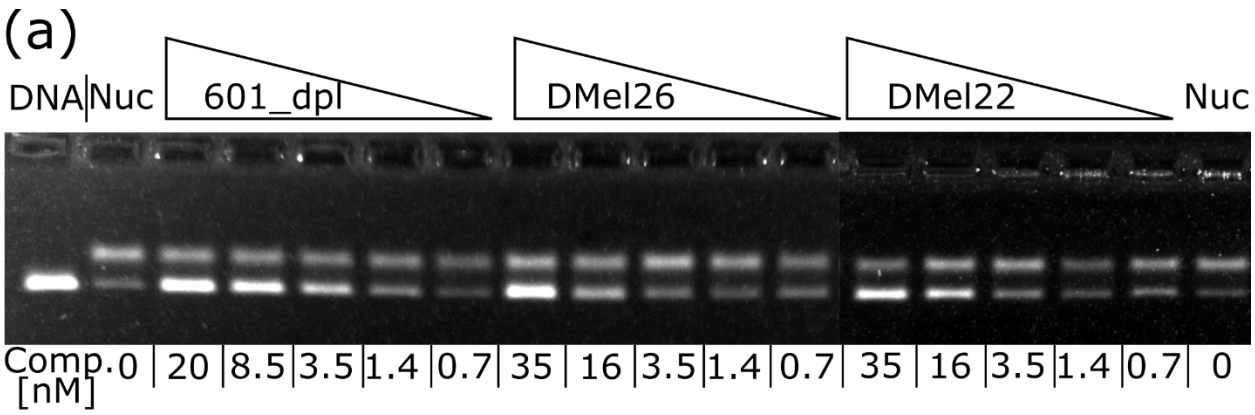


Figure S1. Schematic representation of different sample types and their corresponding fluorescence anisotropy levels, related to Figure 1. The schema depicts the composition of different samples mixed together with the fluorescently labeled reference DNA sequence and different competitor DNA sequences. The first sample (left) contains only the fluorescently labeled reference DNA, while all other samples also contain histone octamers. The 3 subsequent samples contain in addition (from left to right): a strong competitor sequence, the unlabeled reference sequence, and a weak competitor sequence, respectively. The last sample (right) contains nucleosome and labelled reference DNA without competitor DNA sequence. The samples are ordered by their endpoint FA range. Note that we had to use as labelled-DNA reference a sequence (601dpl; a weaker derivative of the 601 sequence) that binds relatively strongly to the histones. However, since most of our genomic sequences bind relatively weakly to histones, the use of a weaker binder as a reference would have been more suitable.

Unfortunately, this was technically not possible since the histones aggregate at the high concentration required to bind quantitatively to the reference (a known phenomenon for histones). Hence, due to this relatively strong reference sequence, the competitor does not fully compete with the labelled-DNA reference, even at high concentration, for DNA competitor with weak affinity to the histones.



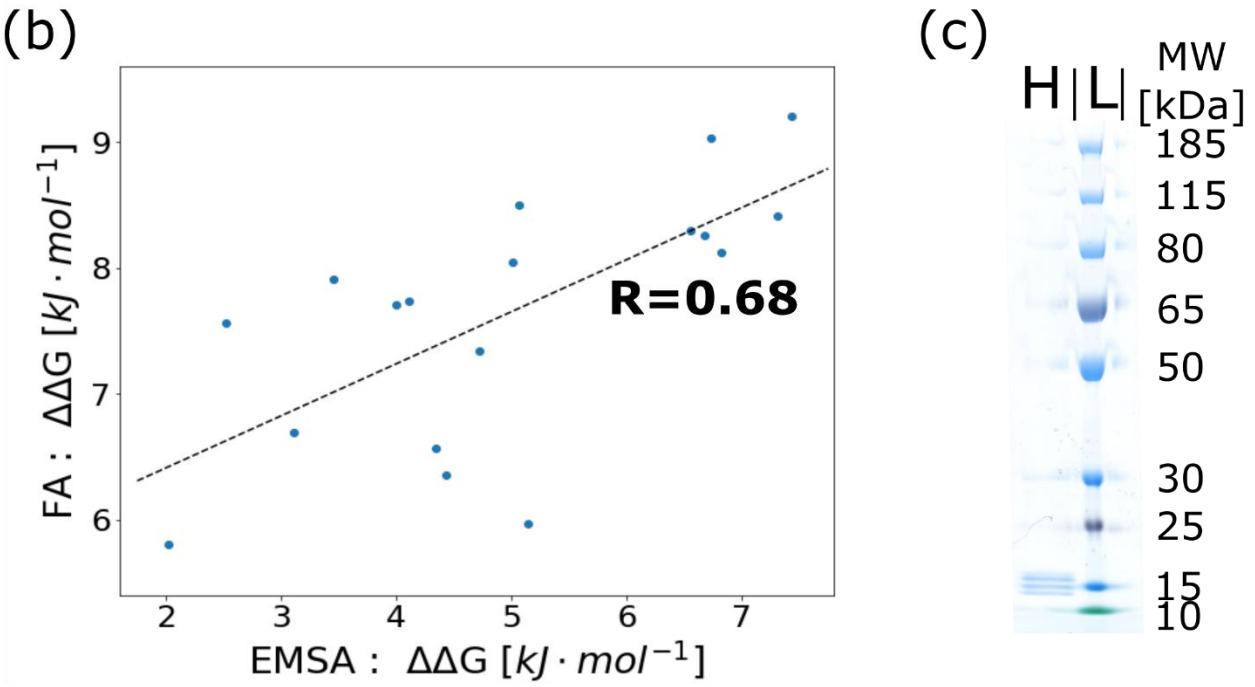


Figure S2. Competitive titrations using EMSA as readout and histone purification, related to Figure 1. (a) Three competitor titration curves analyzed using EMSA. The lanes are labeled with the sample names, “DNA” being the fluorescently labeled reference

sequence 601_dpl DNA, “Nuc” being the reference sequence with histone octamer, but without competitor DNA sequence. “601_dpl”, “Dmel26” and “Dmel22” denote the competitor sequences used for the titration. The triangles symbolize a decreasing concentration of competitor DNA, the concentration in nM is indicated at bottom. (b) Comparison of the nucleosome formation free energies obtained by EMSA and FA. The free energies obtained from 18 titrations with different competitor DNA sequences using FA as readout are plotted against the free energies obtained with the same samples, but using EMSA as readout. The dashed line shows a linear regression, R= 0.68. (c) SDS-PAGE gel of purified nucleosome. “H” denotes histone lane, “L” denotes the ladder, its molecular weights indicated on the right. The three histone bands (H3 -15.4 kDa- and H2A+H2B double band -13.4 kDa/13.7 kDa- and H4 -11.4 kDa) are visible in the lower part of the gel at the expected heights.

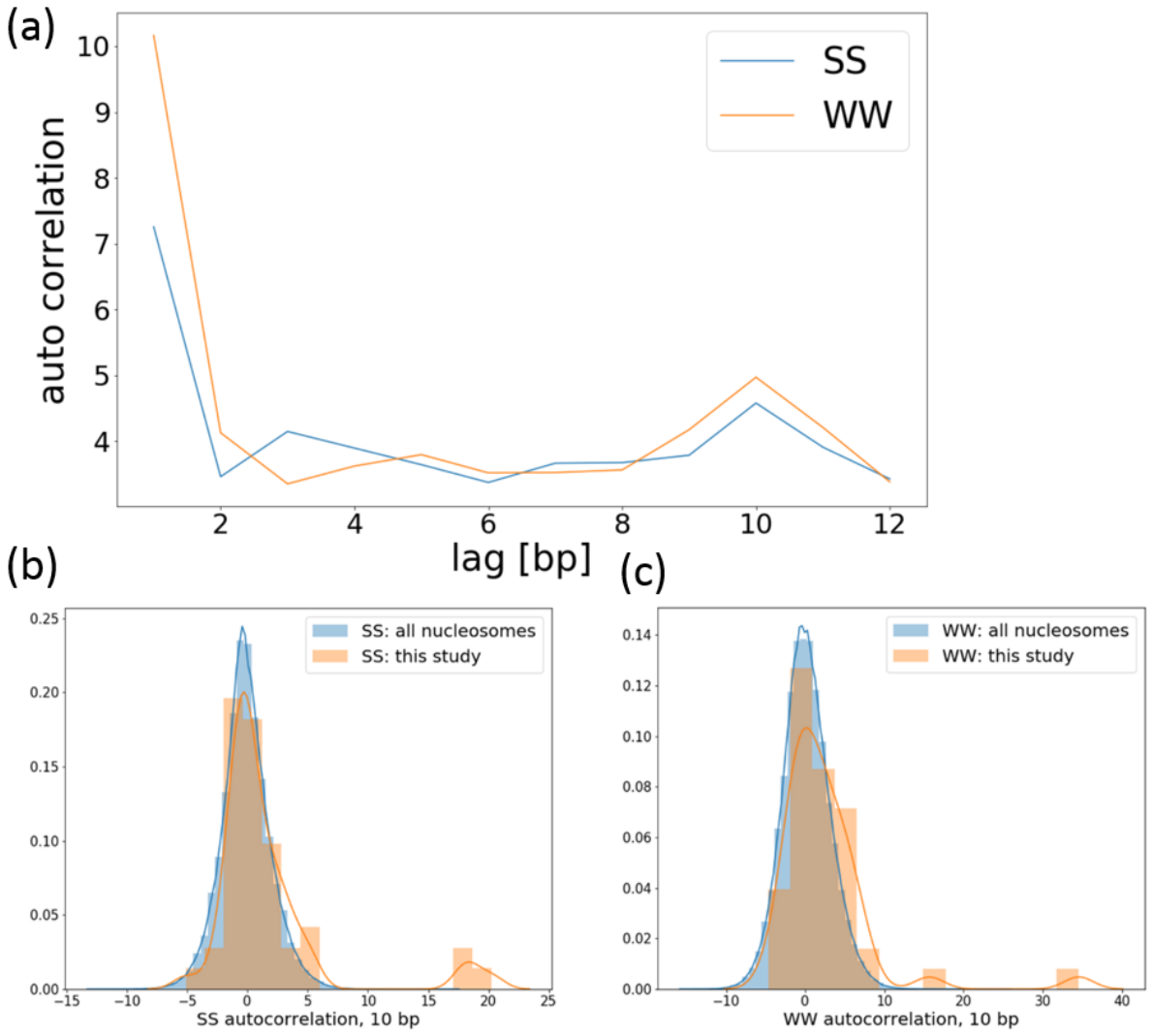


Figure S3. (a) Autocorrelation analysis of dinucleotides, related to Figures 3 and 4. Depicted are the average autocorrelation of SS (S= G or C) or WW (W=A or T) dinucleotides by their lag (in bp). Both curves show a weak peak at 10 bp. (b) and (c) Comparison of distributions of autocorrelation with a lag of 10 bp for sequences used in this study and nucleosomal sequences genome wide. (b) shows the distributions of SS (S= G or C) dinucleotides, (c) shows the distributions of WW (W= A or T) dinucleotides. The histogram for sequences used in this study is in orange, the genome wide distribution

is depicted in blue. Overall, the distribution of sequences used in this study reflects the general genomic distribution, except for a couple of extreme values stemming from literature sequences that were designed to have strong periodicities.

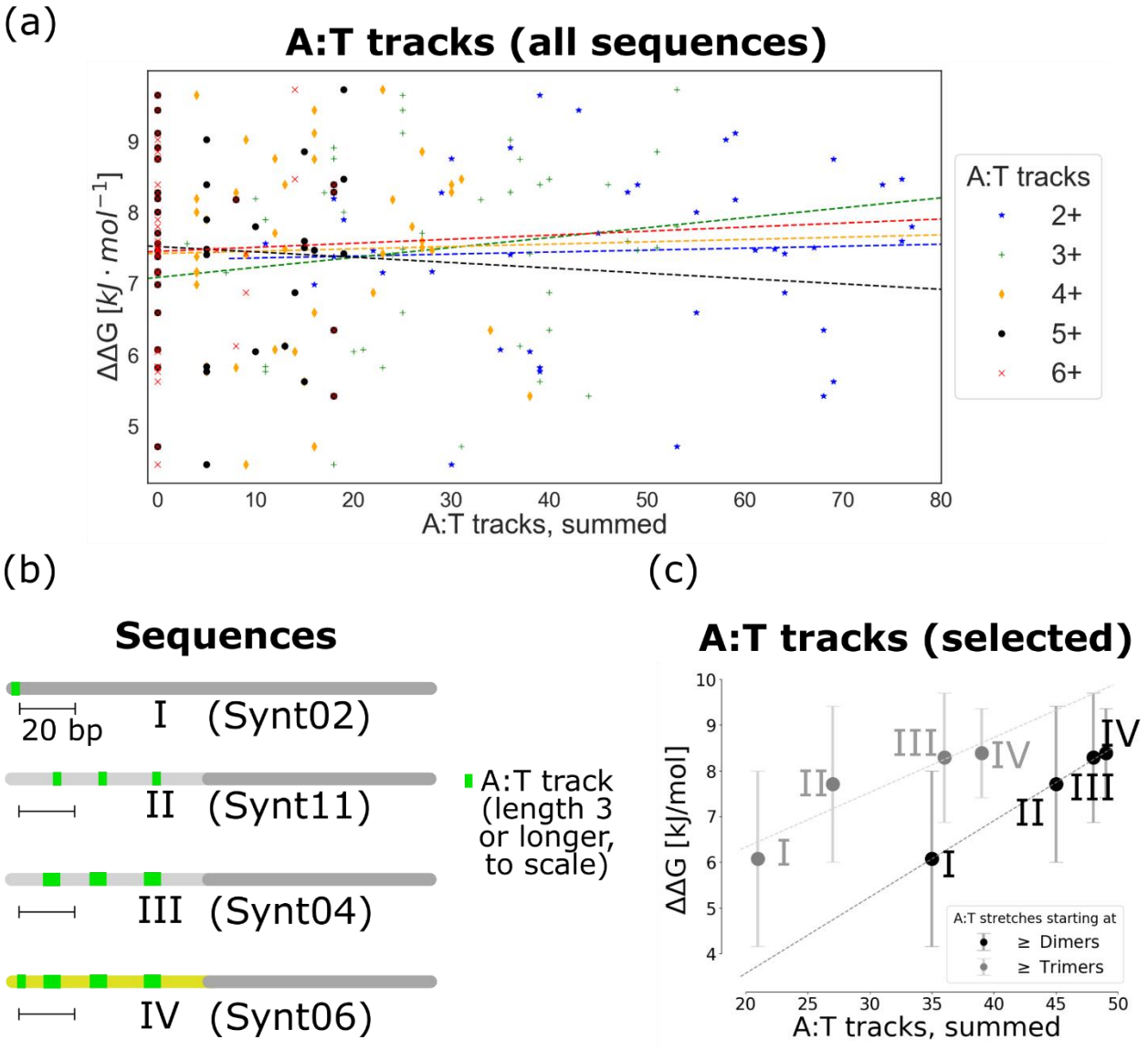


Figure S4. (a) Analysis of the influence of dA:dT tracks on free energy of nucleosome formation, related to Figures 3 and 4. ΔΔG values are plotted against the summed dA:dT track lengths, with varying minimal length (from 2bp to 6 bp as indicated). The dashed lines represent linear regressions with non-significant Pearson correlation coefficients ranging from R =-0.04 to R=0.17. (b) Schematic representation of four synthetic enhancer sequences (Synt02, Synt11, Synt04 and Synt06) with systematically varied dA:dT tracks (dA:dT tracks differing between the sequences highlighted in green, length to scale).

Yellow denotes a different background sequence. Sequences are denoted I to IV for simplicity. c) Free energy of nucleosome formation versus the total number of nucleotides in dA:dT tracks in selected synthetic enhancers (see a) for tracks of minimal length 2 (black) and minimal length 3 (grey); linear regression with dashed lines. All sequences would be expected to have very similar ΔΔG values based on their GC content.

Note regarding influence of dA:dT tracks on free energies of nucleosome formation: Poly(dA:dT) tracks are generally believed to impede nucleosome formation (Raveh-Sadka et al., 2012; Segal and Widom, 2009). The sequences investigated in this study show too few occurrences of sufficiently long dA:dT tracks (length>4; only 53% of the sequences contain at least one 5mer, 23% contain 6mers) to make a reliable statement about their influence on nucleosome binding energy (**Figure S4a**). When varying the minimal track length (2-6), we find no significant correlations between the total number of nucleotides contained in the dA:dT tracks with measured ΔΔG values (**Figure S4a**). However, four sequences (Synt02, Synt11, Synt04, and Synt06, in the following termed I, II, II and IV, respectively, for simplicity) from the group of synthetic enhancers constitute an important exception (**Figure S4b** and **S4c**). The synthetic enhancer sequences were specifically designed to investigate the influence of certain configurations of transcription factor binding sites in the *D. mel*. segmentation network, and, despite sharing most of their sequence, lead to strong differences in expression. One of the transcription factors of interest (Hunchback, Hb) possesses the consensus binding site AAAAAA (**Figure S4c**) (Štanojević et al., 1989). Whereas enhancer I contains no Hb binding sites, III contains three consensus sequences (in green in **Figure S4b**), and IV contains the same three sites, but with a different background sequence (yellow) and an additional AAA track at

the beginning. Finally, II differs from III only by single point mutations in the 3 Hb binding sites (AAAAAA > AAATAA). For these 4 systematically mutated sequences, the free energy of nucleosome formation increases with the total number of nucleotides contained in the dA:dT tracks (**Figure S4c)**; R= 0.999 and R= 0.92 for minimal track lengths of 2 and 3, respectively (there were too few longer tracks for statistics). Although measurable, the influence of these features on ΔΔG is rather weak (total range of 2 $kJ \cdot mol^{-1}$), and it might be completely obscured in cases where there are larger sequence differences, as in the majority of our investigated sequences (**Figure S4a**).

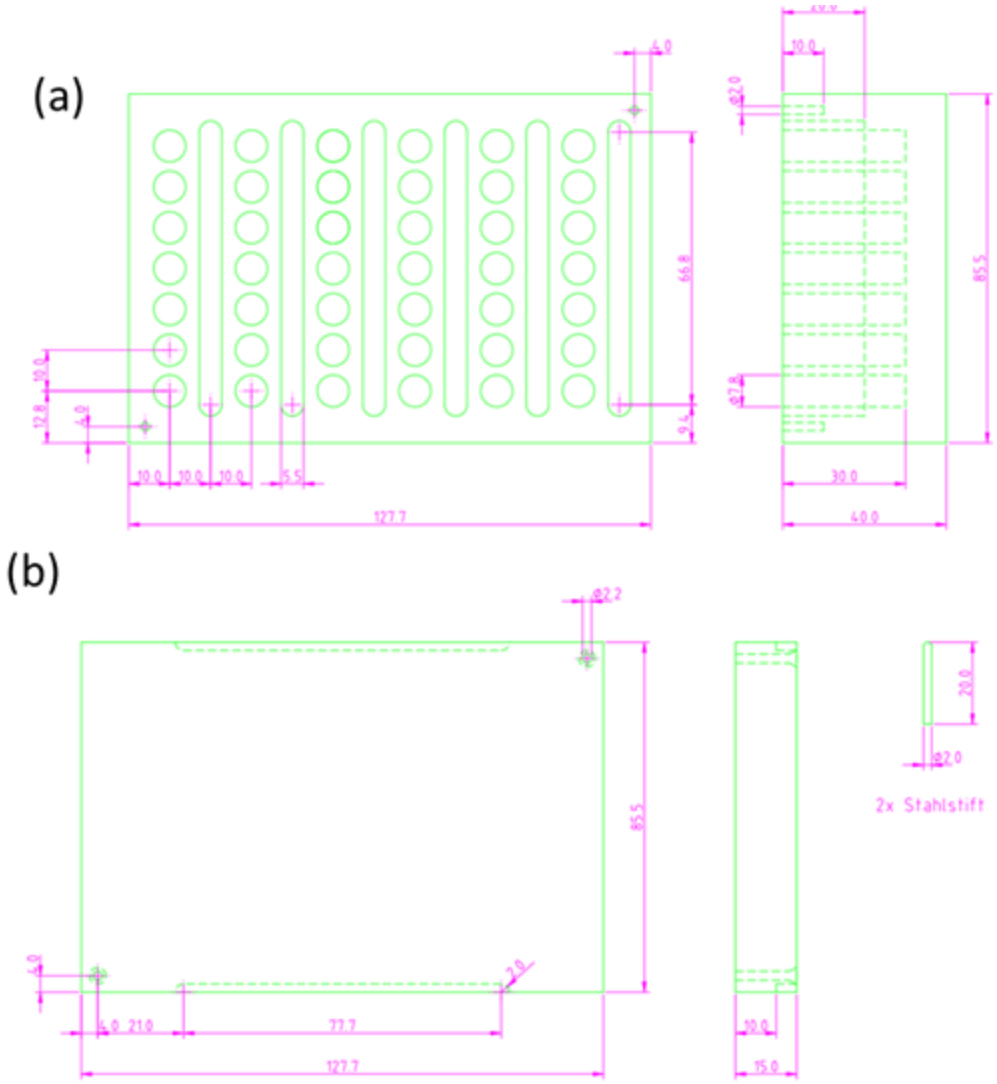


Figure S5. Technical drawings for the custom metal block made from aluminum, related to Figure 1. Tolerances according to ISO 8015 and ISO 2768-m, workpiece edge DIN 6784. (a) metal block. (b) the lid, “Stahlstift” = metal pin.

TRANSPARENT METHODS

PCR templates

The DNA sequences used in this study were amplified using PCR. The templates differ based on the type of sequence: the genomic -1 nucleosomes were amplified from genomic fly DNA, the template for the 601_dpl reference sequence was a ligation product, the literature-derived sequences were on gene synthesis plasmids, the synthetic enhancers on plasmids generated by restriction cloning.

Annealing and ligation

The 601_dpl reference sequence was assembled using six different single stranded oligomers of 34 -67 bases length (Eurofins, Germany), with three constituting the forward strand and three constituting the reverse strand. To ensure sufficient annealing, the DNA sequences contained overhangs of 14 bp complementary to the opposite strands. For annealing, 2 µl of each piece were mixed, the solution was heated to 70 °C in a standard PCR machine. The temperature was decreased to 25 °C at a rate of 0.1 K/s. The resulting annealed DNA was purified using a PCR purification kit (Qiagen) following the manufacturer's instructions. The cleaned DNA was eluted from the column using 30 µl of dH_2O. 10 µl of this eluate were ligated in 150 µl T4 DNA Ligase Buffer (NEB) using the T4 DNA ligase (NEB) in tenfold excess. The ligation was incubated for 30 min at RT. The reaction was cleaned up using the PCR purification kit (Qiagen) and eluting in 30 µl of H2O.

Fluorescence labeling via PCR

In order to generate the fluorescent labeled reference DNA, the aforementioned ligation product was amplified via a PCR reaction in which one of the primers was 5' fluorescently labeled with Cy5 (Eurofins, Germany). The product was again cleaned up using the PCR purification kit, and then eluted using 30 µl of elution buffer (Qiagen).

PCR amplification of competitor sequences

The unlabeled competitor sequences were amplified using touchdown PCR. As a template, genomic fly DNA (provided by M. Bozek (Bozek et al., 2019)) or a plasmid from gene synthesis (Synbio) was used. The resulting PCR products were cleaned up using a PCR clean-up kit. To ensure sufficient concentrations, up to eight PCR reactions of 50 µl each were pooled and eluted in 30 µl elution buffer.

Histone octamer purification

Embryo collection

The histones used in this experiment were extracted from *D. mel*, OregonR, wild type flies. Approximately $2x10^4$ flies per fly cage were supplied with one apple juice agar plate per cage. The plates were changed every 12 h. Starting after 36 h, fly embryos were collected. The embryos were rinsed off using water pressure and separated from other fly material over three consecutively smaller sieves. The embryos were then washed in cold embryo washing buffer (120 mM NaCl, 0.04 % Triton-X-100). The chorion was removed by stirring in 200 ml embryo washing buffer with 60 ml hypochlorite solution for 3 min. The hypochlorite was removed by rinsing the dechorionated embryos with water in the smallest sieve for 5 min. The embryos were snap frozen in liquid nitrogen and stored at -80 °C.

Nucleosome separation

The nucleosome were isolated from the collected *D.mel.* embryos using the protocol provided in Krietenstein et al. (Krietenstein et al., 2012).

Histone octamer isolation

The filtrate of histone octamers was loaded onto the ÄKTA system pre-equilibrated with embryo running buffer (630 mM KCl, 100 mM sodium phosphate buffer, pH 7.2). The histone octamers were loaded to a column consisting of 30 ml hydroxylapatite. After washing with 2 column volumes, the elution was performed with embryo elution buffer (2 M KCl, 100 mM sodium phosphate buffer, pH 7.2). The eluate was collected in fractions of 1 ml, the two fractions with the

highest absorbance were concentrated using centrifugal filters with 10 000 MWCO (Amicon-Ultra, Merck). The centrifugal filters were pre-rinsed with embryo elution buffer, and the concentration was carried out at 12 000 g, 4 °C. The resulting concentrate was mixed with glycerol to contain 50 % glycerol and stored at -20 °C. The purity was analyzed via SDS-PAGE (**Figure S2c**).

Nucleosome reconstitution

Determination of histone to DNA ratio

Before starting with titrations, the optimal ratio of histone octamers to fluorescently labeled DNA was determined. Each titration sample contained 2 µl of 20 ng/µl Cy5 labeled reference and 18 µl of nucleosome octamer dilution of various concentrations. A dilution series of the histone octamer stock solution in titration high salt buffer (1 M NaCl, 10 mM Tris, pH 7.8) was performed. The optimal dilution can vary significantly depending on the histone octamer preparation, and is thus best determined empirically. The optimization should be started at equimolar amounts of octamer and DNA.

When determining the affinity of a competitor sequence, in addition to labeled reference DNA and the histone octamer dilution at the optimal ratio, 2 µl of unlabeled competitor DNA of different concentrations were added. We used five concentrations per sequence, approximately logarithmically equidistant from each other with concentrations up to 35 nM final concentration. For a small number of sequences the highest measured concentration in the titration series was lower than 35nM, since it was difficult to obtain PCR products of that concentration.

Automated salt titration

The samples for a titration series were mixed in a 500 µl protein low binding reaction tube (Sarstedt) using low retention tips (Starlabs). These tubes were housed in a custom-made metal block with a heated lid to keep the temperature as constant as possible and minimize condensation at the lid (see Fig. S5). During the automated titration, the samples (22 µl) were

kept at 30 °C, the lid was heated to approximately 45 °C (see below). The samples were diluted in 16 steps over a time course of 12 hours using titration low salt buffer (1 mM EDTA,10 mM Tris, pH 7.8); for each step, a volume with 1 µl more than in the previous step was added after time intervals of 40 min, resulting in a final volume of 158 µl. After a final incubation of 60 min the samples were measured using fluorescence anisotropy or EMSA.

Metal block with heated lid

The metal block and the corresponding lid were milled from aluminum (technical drawing in **Figure S5**). To ensure a tight closing of the lid and the reaction cups, a microseal 'P' pad (BioRad) was applied to the inside of the lid. The lid was heated by applying a heater polyester foil (77 mm x 110 mm, 12 V, 12 W, thermo technologies) to the lid from the outside. The heater was operated at 9 V to achieve a surface temperature of approximately 45-50 °C.

EMSA

To validate our results with a different readout method, we measured the ratios bound/unbound DNA using EMSA for 18 individual titrations. We split the samples after salt titrations into two parts, mixing 20 µl of reconstituted sample with 25 µl of EMSA Buffer (1x Tris-Glycine Native Running Buffer). The diluted samples were run on a 0.5 % agarose gel (Tris Glycine Running Buffer) at 4 °C for 60 min. The gels were imaged using a Biorad imager.

Microscopy

FA of the reconstituted nucleosome samples was measured using a customized fluorescence microscope setup as described in Jung et al. (Jung et al., 2018) and (Jung et al., 2019). A schematic representation of the setup is shown in **Figure 1c**. To determine the FA signal, 100 µl of each sample were mixed with 100 µl of FA buffer (128 mM NaCl, 10 mM Tris, pH 7.8, 0.85 mM EDTA) using low retention tips in glass bottom SensoPlates (Greiner). Each sample was measured at twelve different z-planes; to reduce the effect of potential fluorescing protein aggregates that could lead to erroneous FA values, these measurements were repeated three times. The FA values were calculated using **Equation 1**, requiring the parallel ($I^{=}$) and

perpendicular (I^{+}) fluorescence intensities and the instrumental G-factor (G=1.15 for our setup, as determined by measuring free dye in solution). Regions of interest (ROI) were defined as equal-sized regions around the maxima in the respective channels.

Equation 1 $$FA(z,t) = \frac{I^{=}(z,t) - G*I^{+}(z,t)}{I^{=}(z,t)+2*I^{+}(z,t)}$$

Data processing

Outlier detection

In order to remove images containing potential aggregates or empty wells, all images from a plate were pooled and outliers were defined as those images showing average pixel values for the respective ROIs with a z-score (based on median absolute deviation) >2.0 (Iglewicz and Hoaglin, 1993). This threshold may be a setup specific factor and should be adjusted if necessary.

Data fitting and outliers removal

The data were fitted using the Hill equation (**Equation 2**), with n being fixed at 0.5 (we found this value to lead to the most robust results), x as the concentration of the competitor in nM. Start and end correspond to the asymptotes which were fixed per set of experiments (plate) The technical replicates (repeated FA measurements in the same well) were averaged for each data point and the inverses of their variances were used as weights in the fit. The fitting was performed using a weighted least square algorithm implemented in Python's Scipy. optimize package.

Equation 2 $$\mathrm{f(x)} = \mathrm{start} + \frac{(\mathrm{end}-\mathrm{start})*\mathrm{x^n}}{\mathrm{kd^n}+\mathrm{x^n}}$$

After the fitting of individual titration curves, the resulting Kds were calculated using the fol-lowing equation:

Equation 3 $$\Delta G = -RT \ln(Kd)$$

where R is the universal gas constant and T the temperature at which the titration was per-formed (303 K). The resulting ΔΔG values were subsequently averaged.

We only used the titration curves for which after the image outlier removal at least three of the five data points remained. The fitting parameters start and end were fixed per plate. Start was determined as a median of all start values when performing individual, free fits per curve (leaving start, end, and kd as free parameters). End was either set to the value of the DNA (different plates and dates). This can only be determined after measuring several plates.

Due to problems in titration or pipetting errors some curves needed to be removed completely from the analysis. This removal was performed based on the cost factor determined by the weighted least square fitting. We found this weighted cost to be a reliable indicator of problematic curves (its value was 80 mM^2 in our data set).

Data analysis

All data analysis was performed using Python 2.7. The packages relevant for the analysis are Scipy (1.0.0) and Numpy (1.15.2).

GC content analysis

The relationship between ΔΔG and GC content was analyzed using a segmented linear regression with the following partially evaluated functions:

Equation 3.1

for x< gc_optimum: $0 = x \cdot m_1 + c_1 - y$

Equation 3.2

for x>gc_optimum: $0 = x \cdot m_2 + m1 \cdot gc_{optimum} + c_1 - m_2 \cdot gc_{optimum} - y$

The regression was performed with a least squares fit, mimimizing equation 3 with x as the GC content and y as the binding free energy ΔΔG. m_1, m_2, c_1 and $gc_{optimum}$ were free parameters of

the fit. To determine the p-value, we took as null hypothesis that the slope of the linear regression of predicted versus measured values is not different from zero.

Note that the limited range of binding free energies for natural sequence makes the analysis challenging, as their respective values are very close to each other. However, whereas this narrow distribution doesn't allow to significantly distinguish from each other individual data points with close values, the statistical analysis of the entire distribution of the data points provides a way of testing the significance of the correlation. Although individually not determined with high precision, the data points contain collectively more information, which justifies their use in a global analysis

Autocorrelation analysis

For the autocorrelation analysis, the sequences were translated to a binary sequence, by evaluating for each pair of dinucleotides if it belongs to the group of interest (WW or SS, respectively). At each position in the sequence, the mean of the sequence was subtracted, and an autocorrelation was performed using the correlation function from numpy.

dA:dT Track analysis

The total number of bases occurring in consecutive stretches of A or T, respectively, with a given minimal length was calculated. Linear regressions were performed between the summed bases and the binding free energy.

EMSA evaluation

We used ImageJ to quantify titrations curves determined with EMSA. As the resulting total intensities per lane changed with concentrations, we normalized each intensity by the mean intensity of other points of the same concentration. We fitted the resulting curves with Equation 2 directly."

Supplemental References

Bozek, M., Cortini, R., Storti, A.E., Unnerstall, U., Gaul, U., and Gompel, N. (2019). ATAC-seq reveals regional differences in enhancer accessibility during the establishment of spatial coordinates in the Drosophila blastoderm. Genome research *29*, 771-783.
Iglewicz, B., and Hoaglin, D.C. (1993). How to Detect and Handle Outliers (ASQC Quality Press).
Jung, C., Bandilla, P., von Reutern, M., Schnepf, M., Rieder, S., Unnerstall, U., and Gaul, U. (2018). True equilibrium measurement of transcription factor-DNA binding affinities using automated polarization microscopy. Nat. Commun. *9*, 1605.
Jung, C., Schnepf, M., Bandilla, P., Unnerstall, U., and Gaul, U. (2019). High Sensitivity Measurement of Transcription Factor-DNA Binding Affinities by Competitive Titration Using Fluorescence Microscopy. JoVE, e58763.
Krietenstein, N., Wippo, C.J., Lieleg, C., and Korber, P. (2012). Genome-wide in vitro reconstitution of yeast chromatin with in vivo-like nucleosome positioning. Methods in enzymology *513*, 205-232.
Raveh-Sadka, T., Levo, M., Shabi, U., Shany, B., Keren, L., Lotan-Pompan, M., Zeevi, D., Sharon, E., Weinberger, A., and Segal, E. (2012). Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. Nat Genet *44*, 743-750.
Segal, E., and Widom, J. (2009). Poly(dA:dT) tracts: major determinants of nucleosome organization. Current opinion in structural biology *19*, 65-71.
Štanojević, D., Hoey, T., and Levine, M. (1989). Sequence-specific DNA-binding activities of the gap proteins encoded by hunchback and Krüppel in Drosophila. Nature *341*, 331-335.