Multi-exponential DNA Residence Behaviors of Transcription Factors Under The Discrete Affinity Model

Zafer Kosar [1,✉] **and Aykut Erbaş** [1]

[1]UNAM-National Nanotechnology Research Center and Institute of Materials Science & Nanotechnology, Bilkent University, Ankara, Türkiye

The transcription process is regulated by temporal interactions of transcription factors with DNA. In the last decade, computational and experimental studies revealed the residence times of transcription factors on DNA correlate with transcriptional output. Biochemical studies suggest that transcription factors exhibit bi-exponential dynamics, attributed to the binary affinity model composed of nonspecific and specific protein-DNA bindings. Recently, transcription factor residence times were shown to display a power-law pattern implicating protein-DNA affinity levels are rather continuous. Elucidating the underlying mechanisms of transcription factor residence distributions, beyond protein-DNA interaction strength, is crucial to construct a more complete understanding of transcriptional regulation. Here, by using molecular dynamics simulations of DNA and dimeric proteins, we demonstrate residence time behaviors of generic homodimeric transcription factors follow a multi-exponential pattern even with single and binary affinity levels between DNA and proteins, indicating the existence of emergent behavior. Our simulations reveal that DNA-protein clusters of various sizes contribute to this multi-exponential behavior. These findings add another layer to transcriptional regulation and, consequently, to gene expression by connecting protein concentration, DNA-protein clusters, and DNA residence times of transcription factors.

Multi-exponential Residence Times | Transcription Factor Dynamics | Transcriptional Regulation

Correspondence: ***zafer.kosar@bilkent.edu.tr***

Introduction

Gene expression dictates many aspects of cellular behavior, including response to extracellular factors and cellular identity. Among the many regulators of gene expression, transcription factors (TFs) play the most prominent role in inhibiting or activating the transcription of their target genes (Browning, 2004; Seshasayee et al., 2011; Bintu et al., 2005). Regulation of a gene by a TF protein starts at the binding site once the TF binds to its specific DNA sequence and forms protein-DNA complexes. Nevertheless, gene expression is a highly dynamic and regulated process. Thus, TF's unbinding from the DNA binding site might not be a fully stochastic process (de Jonge et al., 2022; Haberle and Stark, 2018). In accord, recent studies in TF dynamics revealed that the residence time (RT) of a TF on its DNA binding site (i.e., its unbinding rate) is intertwined with transcriptional output (Clauß et al., 2017; Lickwar et al., 2012; de Jonge et al., 2020). Thus, TF residence times on DNA directly contribute to the regulation of gene expression. Therefore, understanding residence time patterns becomes a key to the accurate prediction of gene expression behavior.

The duration of a TF on its DNA binding site can depend on the TF protein's affinity to the DNA, temperature, and 3D DNA structure (Kim and Shendure, 2019; Inukai et al., 2017). Kinetic studies focusing on unbinding (or dissociation) kinetics of TFs from their single-binding sites have demonstrated the concentration dependency of TF unbinding rates (Graham et al., 2011; Joshi et al., 2012). In scarce concentrations, a TF tends to stay bound to its target site for more extended periods of time. In contrast, the abundance of unbound TFs in solution leads to a competition for the binding sites on DNA, resulting in much higher unbinding rates (i.e., shorter residence times) via the process referred to as Facilitated Dissociation (FD) (Kamar et al., 2017; Koşar et al., 2022).

The impact of TF concentration is not limited to protein unbinding dynamics via FD. TFs can also contribute to 3D genome organization and form DNA-protein clusters in a concentration-dependent manner (Kim and Shendure, 2019; Noort et al., 2004; Skoko et al., 2006, 2004; Remesh et al., 2020; Koşar and Erbaş, 2022; Arold et al., 2010; Dame et al., 2000; Winardhi et al., 2015). Such structural effects are more pronounced with bacterial Nucleoid-Associated Proteins (NAPs), a class of DNA-binding proteins often with dual functionality. NAPs are involved in chromosome organization, and a number of them also function as transcription factors (Dillon and Dorman, 2010; Wang et al., 2011; Dorman, 2014). Due to their common multivalent nature Lee (1992), DNA-binding proteins can drive the bridging of DNA segments and thus the formation of DNA-protein clusters of various shapes and sizes and other forms of chromosome architectural effects (Skoko et al., 2004; Dillon and Dorman, 2010; Wang et al., 2011; Hammel et al., 2016; Dame, 2005; Verma et al., 2019). Molecular Dynamics (MD) simulations of model bacterial systems suggest that these events are highly dependent

on protein concentration and U_{ns}, nonspecific interaction potential (i.e., the affinity between the DNA-binding protein and nonspecific DNA).

The affinity of a TF to its specific binding site is often stronger compared to its binding affinity towards a nonspecific DNA sequence. However, nonspecific binding can contribute to the global chromosome organization. Contrarily, specific protein-DNA interactions can govern the formation of small local DNA-protein complexes (i.e., clusters) around the specific binding sites (Koşar et al., 2022; Lin et al., 2012; Brackley et al., 2013; Agback et al., 1998). DNA-protein complexes could significantly affect transcriptional regulation. Transcription factories are multiprotein complexes formed by TFs, RNA polymerase, coactivators, etc., within the eukaryotic nuclei. These factories not only carry out transcription but also organize nuclear architecture (Iborra et al., 1996; Cook, 2010; Melnik et al., 2011; Mitchell and Fraser, 2008). Therefore, multiprotein complexes or clusters are crucial regulators of transcriptional output.

Moreover, Eukaryotic euchromatin regions, which are not densely packed as heterochromatin regions, are more accessible by transcription factors thus, more transcriptionally active (Amemiya et al., 2022; Elgin and Grewal, 2003; Penagos-Puig and Furlan-Magaril, 2020). This notion emphasizes the significance of local genome architectures in residence times. In other words, the contribution of DNA-binding proteins to gene regulation is not restricted to their functions as TFs since their activity in domain-specific and global genome organization is also crucial.

A recent experimental study employing single molecule tracking (SMT) demonstrated residence durations of several TFs and chromatin-associated proteins dynamics follow a power-law pattern (Garcia et al., 2021). These findings support a continuum model for TF dynamics and TF-DNA interaction affinities rather than a bi-exponential model attributed to binary specific and nonspecific affinities.

In the bi-exponential model, TFs were considered to have longer residence times on their specific target DNA sequence and shorter residence times on nonspecific DNA, generating a bi-exponential distribution of residence durations (Chen et al., 2014; Ball et al., 2016; Morisaki et al., 2014). Contrary to this suggestion, even with a single affinity ($U_{ns} = U_{sp}$) and binary affinity model ($U_{ns} < U_{sp}$), we observed an apparent multi-exponential pattern in our model, indicating an emergent behavior. In parallel with our findings, multi-exponential was used to interpret the residence behaviors of several SMT studies (Hipp et al., 2019; Reisser et al., 2020; Agarwal et al., 2017).

Consequently, we explored the driving factors of multi-exponential residence behaviors of TFs. In particular, high nonspecific affinity cases, which enabled the formation of much larger clusters, exhibited distinct RT distributions and required more exponents to match with decay curves. Investigation of TF-DNA cluster formations of different sizes revealed that cluster dissipation mean lifetimes are coupled to their sizes. Therefore, concentration and affinity-dependent cluster formations could be the driving factors of the observed multi-exponential patterns.

Finally, we explored the distributed affinity model to check whether the additional complexity would lead to power-law behavior. However, it is clear that limited discrete affinity distribution again forms a multi-exponential decay pattern and does not provide sufficient complexity to generate a power-law behavior.

Here, we demonstrate that TF dynamics follow multi-exponential patterns without the need for multiple TF-DNA affinity levels. We also show how these behaviors are impacted by cluster formations. Our model predicts the power-law behavior might be plausible even with discrete affinities with some additional complexities. These findings may help establish a more advanced understanding of transcriptional regulation, thus of regulation of gene expression.

Results

Residence time behaviors are multifactorial

We employed multiple cases for nonspecific interaction affinities ranging from a very weak $1kT$ to a strong $4kT$ (i.e., $U_{ns} = U_{sp}$) per bead, where $U_{sp} = 4kT$ per bead in all cases. For each of these affinity levels, four different physiologically relevant (Verma et al., 2019; Azam et al., 1999; Ball et al., 1992) concentrations ranging from $10 - 60\mu M$ of TFs were employed in the simulations. We obtained residence time patterns, as shown in Figure 1 and as described in methods, in the form of occurrence versus duration, where occurrence is the number of times a duration was achieved or observed. We then tried fitting several equations to define TF residence patterns (Figure 1C). Fit equations included exponential decay with up to five exponents as well as a power-law equation (Figure 2).

The projection of RT patterns in the log-log scale exhibited an apparent arching pattern, eliminating the possibility of a good power-law fit, which requires a straight pattern in such a scale. The single exponential decay (ED) equation can only describe short-duration ($< 10a.u.$) distributions regardless of affinity and concentration (Figure S2A, Figure S1), implying the full-spectrum residence time behavior is not dependent on a single parameter. This behavior was indeed expected due to distinct NS and SP levels leading to at least a bi-exponential behavior.

For most of the cases, our simulations utilized binary values where, for each simulation, there was only one nonspecific and one specific affinity where $U_{ns} < U_{sp}$. If there was no emergent factor affecting residence times, a double exponential decay (DED) would be sufficient to interpret the residence time patterns obtained from the simulations. Thus, we used a double exponential decay equation for characterizing RT patterns. DED provided

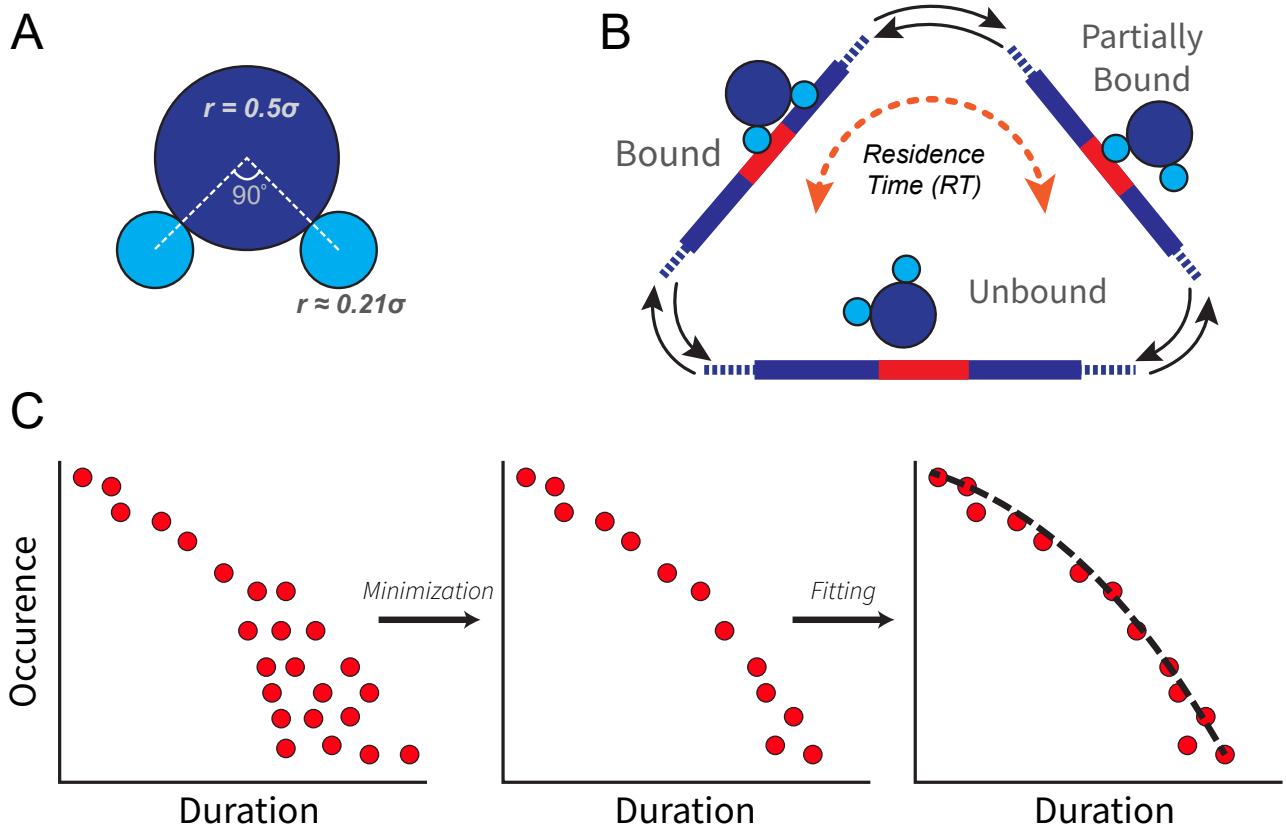

Figure 1. Graphical abstract of the study
(**A**) The coarse-grained model of a generic homodimeric transcription factor. The light blue parts represent the binding domains and dark blue part is the hinge (non-binding) domain. Radiuses of the coarse-grained bead are given in Lennard-Jones (LJ) units and the angle between binding domains are shown in degrees. (**B**) Possible binding states of a TF and calculation of residence times (RT). A TF could be fully or partially bound to DNA, or it could be in an unbound state in which it does not interact with the DNA. Note that red and dark blue regions represent specific binding sites and nonspecific binding sites of the DNA, respectively. Nevertheless, binding to the either site are considered equivalent. (**C**) Collection, minimization, and analysis of TF residence patterns. First, bound durations of each TF throughout the simulations are collected which is then minimized to their medians. Resulting patterns are analyzed fitting several exponential equations.

much better fits compared to the ED equation but notably failed to generalize well for all durations (Figure 2). That is certainly noteworthy, considering simulations utilized a single type of TF and two distinct types of DNA sites.

The employed strategy here was to increase the number of exponents to characterize TF residence behaviors better. We gradually increased the number of exponents for the decay equation to up to five exponents. On top of the visual inspection of the fits, we quantitatively analyzed their accuracy via normalized Residual Sum of Squares (RSS), where the lower RSS indicates a better fit. The increase from single to double as well as from double to triple exponential decay resulted in the order of magnitudes lower RSS (Figure 2B). Additional increments in the number of decay exponents also reduced RSS for the fits, but the changes were not as drastic. These analyses demonstrate that RTs follow a multi-exponential decay pattern, suggesting RT patterns are shaped by additional factors besides binary DNA-protein binding affinities.

Residence time behaviors are dependent on the emergent behavior of TF concentration and binding energies

Residence time behaviors (Figure 2A, Figure S1) suggest a pattern beyond a bi-exponential system that cannot be simply explained nor attributed to the dual model of short-lived TF-DNA interactions on nonspecific sites and longer-lived, more stable interactions on specific binding sites. Therefore, an emergent behavior is required to explain such multi-exponential patterns. Specifically, patterns obtained from elevated concentration and high-affinity kT cases require more exponents, indicating such behaviors emerge as the results of concentration and energy levels.

The relatively high binding energies and concentration of DNA-binding proteins were shown to lead to cluster formations as well as local and global condensations of the chromosome (Koşar et al., 2022; Lin et al., 2012; Brackley et al., 2013; Agback et al., 1998). Unsurprisingly, higher concentrations yield larger clusters (Figure 3C). The remarkable part is that global compaction of the chromosome, and cluster sizes, even conformations are mainly regulated by nonspecific interactions, and specific interactions have little effect. This behavior can be attributed to the abundance of nonspecific

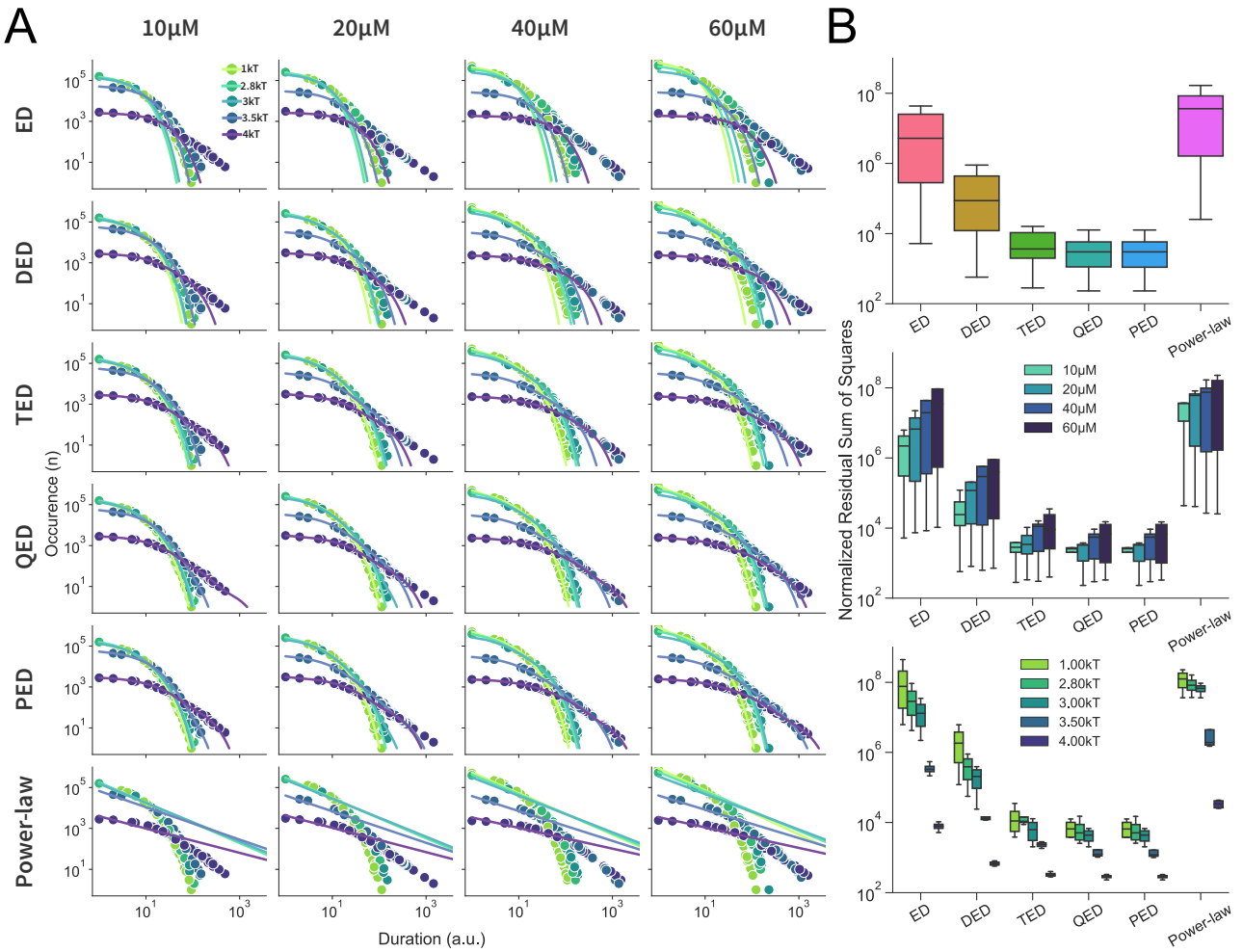

Figure 2. Analysis of TF residence time patterns
(**A**) Distribution of TF residence times in arbitrary unit (a.u.) at various concentrations for distinct U_{ns} levels and fitting Exponential Decay (ED), Double Exponential Decay (DED), Triple Exponential Decay (TED), Quadruple Exponential Decay (QED), Pentuple Exponential Decay (PED), and Power-law equations to residence time data. (**B**) Normalized Residuals Sum of Squares (RSS) of the fits depicting the difference between observation and the fits. For all equations, overall RSS, RSS by concentrations, and by binding affinities are shown, respectively. Note that specific binding energies are $4kT$ for all the systems.

DNA sites over specific sites. In this work, we reduced the impact of global chromosome compaction via miniaturized TF binding domains (Figure 1A), minimizing the bridging of multiple DNA segments. Moreover, the use of adequate binding energies eliminated the possibility of high chromosomal compaction even at high concentrations (Figure 3A). This strategy allowed TFs to roam relatively freely within the cellular confinement.

Our model depicts TFs residing for increasingly higher durations on DNA instead of freely roaming around with the increasing NS potential with fixed specific binding energy (Figure 3A-B). Consistent with the previous findings, at a very low NS potential ($U_{ns} = 1kT$), bound proteins are sparsely distributed around the DNA polymer, mainly near binding sites, but do not form DNA-polymer complexes (i.e., clusters) (Figure 3). Relatively higher NS potentials ($U_{ns} = 2.8kT$ and $U_{ns} = 3kT$) led to primarily small and globular clusters. Whereas stronger NS affinities ($U_{ns} = 3.5kT$ and $U_{ns} = 4kT$) enabled the formation of much bigger clusters with filamentous conformation (Figure 3C). Therefore, we hypothesized the cluster sizes and conformations to play significant roles in TF residence patterns.

Cluster formations drive multi-exponential residence time patterns

To unravel the relationship between DNA-protein clusters and multi-exponent patterns, we initially considered tracking and collecting residence times of TFs for each individual cluster, as described in Figure 1. However, clusters are dynamic formations, and it is not rational to follow TFs that were initially part of a cluster because after they dissipate, they are free to bind anywhere on DNA. Therefore, we rather investigated the dissipation mean lifetimes of individual clusters (see methods). As opposed to RT pattern acquisitions, partial unbinding events were also counted. This modification was needed simply because at high nonspecific affinities, the time needed for obtaining cluster decay rates would exceed simulation lifetimes when only full unbinding events were counted.

At relatively low nonspecific energies ($U_{ns} = 2.8kT$ and $U_{ns} = 3kT$), correlation analysis of cluster size and their dissipation durations did not reveal any relation. Con-

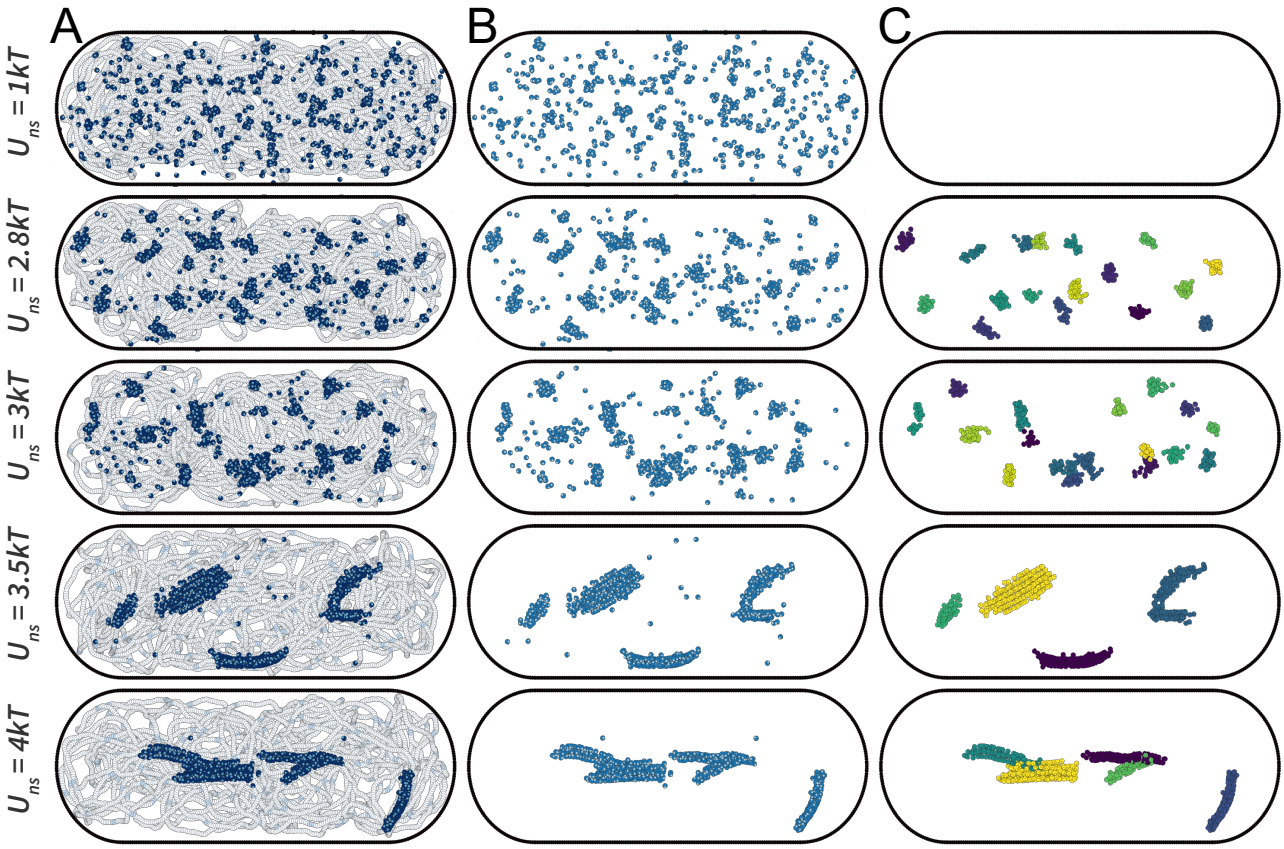

Figure 3. Visualizations of the system, protein distributions, and cluster formations of the coarse-grained bacterium model at various nonspecific binding energies
(**A**) Overview of the system DNA (white-light blue), and TFs. (**B**) TF distributions within the confinement. (**C**) Cluster formations of TFs. Here, the coloring only serves to distinguish distinct protein clusters. Snapshots were obtained from systems with TF concentrations of $60\mu M$. Specific binding energies are $4kT$ for all the systems.

trarily, higher nonspecific interaction potentials ($U_{ns} = 3.5kT$ and $Uns = 4kT$) led to correlations with coefficient of $r \geq 0.5$ (Figure 4B), where larger clusters had longer mean lifetimes. These imply differently sized clusters could dissipate at diverse rates, contributing to multi-exponential behavior that is more apparent for high nonspecific affinity cases (Figure 2A).
It should also be noted that at $2.8kT$ and $3kT$ nonspecific affinities, cluster sizes ranged from 20 to 45 and 20 to 70 TFs, respectively (Figure 4B). The size ranges were significantly improved with higher nonspecific interactions. The cases of $3.5kT$ and $4kT$ nonspecific affinities allowed the formation of clusters of size ranging from 50 to 250 TFs (Figure 4B). This increase, of course, provides advantages for the analysis of the relation between lifetimes and cluster sizes.

TF residence times depend on their position in clusters

Upon establishing the relationship between cluster sizes and their lifetimes, we considered analyzing the residence times of TFs that are located at different regions of the clusters. Cluster-associated TFs can be classified as surface TFs and core TFs. While the core TFs are rather trapped within the cluster, surface TFs are more exposed and could be expected to have shorter residence durations as they are free to unbind. Additionally, TFs that are not part of any cluster are classified as free TFs, which are useful as references against cluster-associated TFs (Figure S2B).

Visual investigation of individual proteins located at different positions with respect to the TF-DNA clusters did not show apparent distinction in their residence times (Figure S2). However, in an analytic (i.e., t-test) and more comprehensive approach utilizing multiple time steps, core TFs exhibited significantly longer residence times compared to surface and free TFs (Figure 5A). However, this behavior was limited to the high nonspecific affinity cases and not observed at lower nonspecific potentials, which may add another layer of complexity to residence behaviors, therefore, could explain the more prominent multi-exponential residence patterns at such energy levels. At $U_{ns} = 3.5kT$ case, surface TFs also had significantly higher residence times in comparison to free TFs (Figure 5A). This disparity indeed reveals that even located at the exposed regions of clusters, TFs could behave differently compared to freely roaming TFs. Overall, the variations among core, surface, and free TFs residence times could shape the overall residence distributions and further contribute to multi-exponential DNA-residence behaviors of TFs at single and binary affinity models.

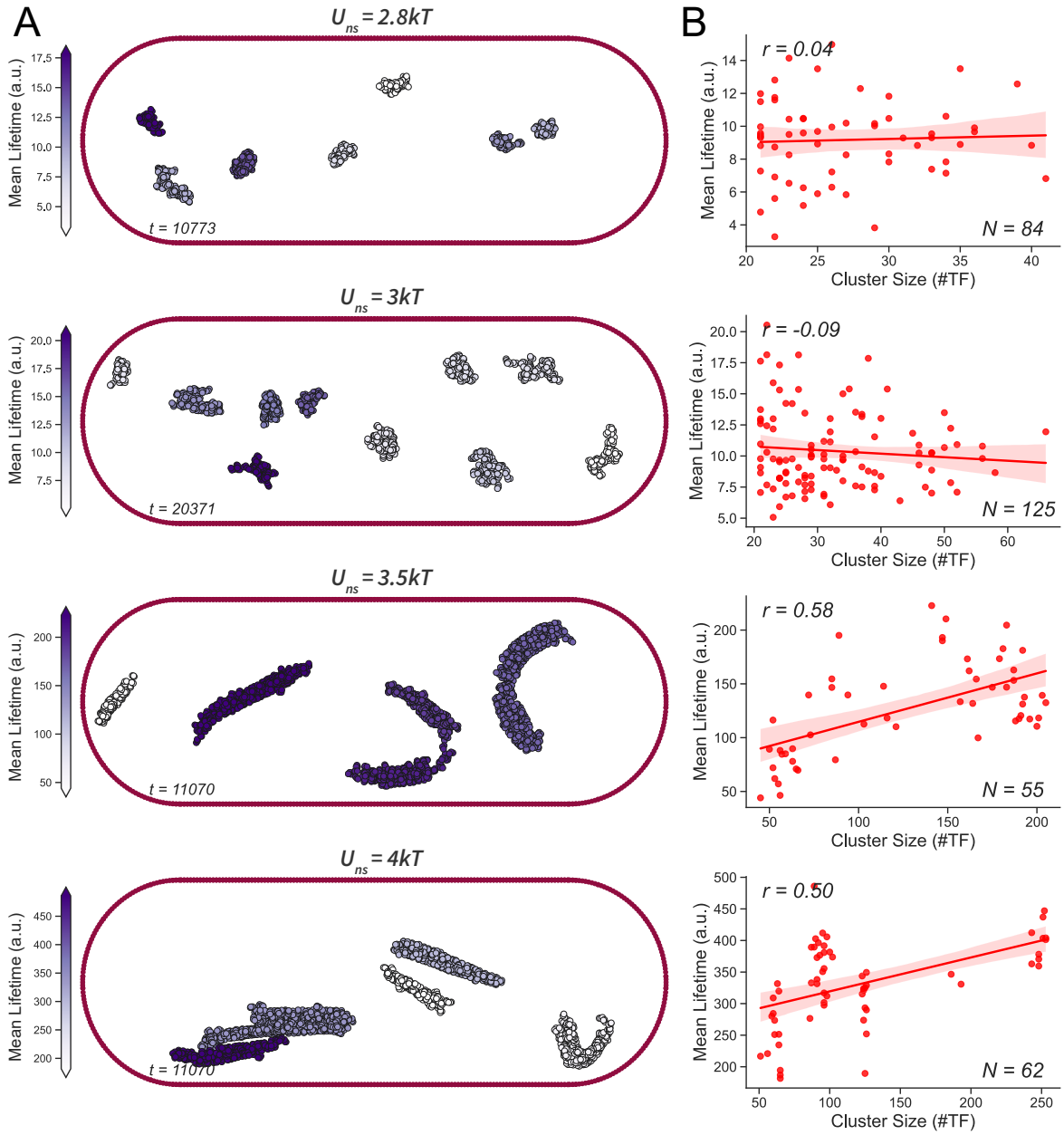

Figure 4. Correlation analysis of cluster sizes and dissipation times
(**A**) Snapshots of the TF clusters at given nonspecific energies from random timeteps. Darker colors implicate higher mean lifetime as indicated by the colorbars on the left. (**B**) Mean lifetimes in arbitrary units versus cluster sizes in number of TF they are composed of. Regression lines are used to determine correlations. Pearson correlation coefficients (r) are given on the upper-left side and number of the clusters in the analysis are given on the bottom-right side. TF concentrations were $60\mu M$ for all cases. The threshold was set as minimum 20 TFs for cluster size for decay rate analysis due to statistical reasons. 12 timesteps were used for sampling and the fixed specific binding energy was $4kT$ for all simulations.

Another factor that should be considered is the differing number of TFs of each type for distinct cases. Even though the concentration ($60\mu M$) and specific binding potentials ($U_{sp} = 4kT$) are fixed among the cases, the higher nonspecific energies drive clustering rather than scattering (Figure 3B), also leading to the formation of much bigger clusters (Figure 3C, Figure 4B). Thus, more of the TFs are located within the clusters for high nonspecific affinity cases (Figure 5B). Moreover, the larger the clusters, the higher the core TF percentages were (Figure 5B), as expected due to the decrease in the surface-to-volume ratio.

Distributed affinity models do not lead to power-law behavior

Our MD simulations revealed that even single and binary affinity models could lead to multi-exponential residence time patterns. The next step was to mimic the continuum affinity model with a distributed affinity model, where we employed 13 distinct DNA site types and assigned them diverse binding affinities with DNA. The affinities inclusively ranged between $1 - 4kT$ and $2 - 5kT$ with an increment of $0.25kT$ and the number of DNA sites for the given affinity assigned to provide either a normal (i.e., Gaussian) or uniform distribution. Contrary to our expectations, distributed affinity mod-

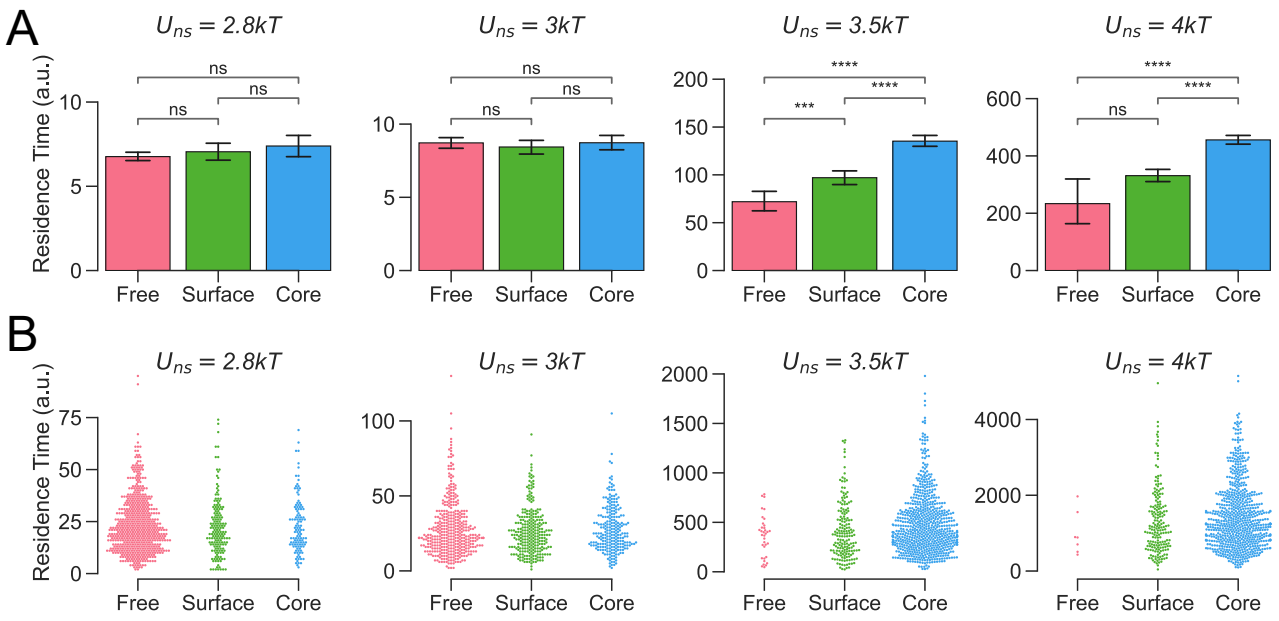

Figure 5. Residence time analysis based on the TF localization with respect to clusters
(**A**) Differences among of TF residence times for free, surface, and core TFs at given nonspecific affinities. Welch's t-test was used for significance analysis. (**B**) Distributions of TF residence times for the given TF types. 12 timesteps were used for sampling and the fixed specific binding energy of $4kT$ was used for all simulations. Representative 12% of the data are shown to prevent over-crowding. Data was obtained from $60\mu M$ cases. Free TFs do not belong in any clusters. TFs on the Surface and in the Core of clusters were named accordingly.

els did not produce power-law behaviors . Instead, the triple-exponential fit was the better predictor of the models (Figure 6). Further power-law fit attempts via adding weights for the power-law equation provided better fits partly for higher durations, yet it failed to explain overall residence time patterns (Figure 6). Therefore, we might speculate that discrete distribution models may not be enough for power-law behavior, and such behavior would require much higher complexity than our MD simulations could provide to behave completely as a continuum model.

Discussion

Our findings reveal an apparent multi-exponential behavior for the discrete affinity model of TF-DNA bindings. Single and binary affinities leading to multi-exponential distributions were beyond the prior suggestion of bi-exponential behavior due to specific and nonspecific binding sites. Here, we demonstrate how this multi-exponential behavior is shaped by TF-DNA complexes, the dependency of these formations, and the residence times behavior to nonspecific binding affinities.

One of the prominent features of our previous coarse-grained model was the global chromosome organization. Although that system allowed the modeling of the role of nonspecific affinities, selected bead size for TF binding domains led to multiple interactions facilitating the chromosomal collapse. Extreme compaction of the chromosome and resulting residence times of TFs would exceed simulation durations, making it unlikely to obtain RT patterns. By lowering the bead size of TFs, we minimized multiple binding and over-compaction. That also enabled the decoupling of 3D genome organization from RT patterns, increasing the control over other variables.

Similar to any MD study, this work has considerable limitations. Most prominently, the simplification of the cell for coarse-grained MD simulations removes most of the complexity possessed by the actual cellular systems. Additionally, we used a single type of TF for all MD simulations with discrete binding affinities. In combination, reduced complexity may explain the lack of observation of power-law behavior, which was attributed to continuous TF-DNA affinities in experimental studies. Of course, the cellular complexity is persistent in the experimental setup. Therefore, other factors, such as fluctuating TF levels and dynamic chromosomal landscape, should not be neglected. However, our simulations, even with discrete affinities, exhibited multi-exponential behaviors, suggesting a power-law pattern is highly possible with continuous affinities, which is beyond the scope of this study.

As their primary roles, TFs can inhibit and activate transcription. The formation of stable protein-DNA complexes extends the duration of a TF's residence on its target DNA site. That, in turn, could significantly enhance their inhibitory or activatory effects. We may also speculate that larger and more compact clusters may decrease the accessibility by RNA polymerase, such as in heterochromatin-like domains (Amemiya et al., 2022), effectively reducing the transcriptional output. Contrarily, similar protein-DNA complexes, but in the form of transcriptional machinery, could initiate or enhance transcription.

Even though we demonstrated cluster size is an important factor for residence times, we also wanted to reveal any possible relationship between cluster shape and dissipation rates. However, investigation of the conformations of the cluster formations on residence times was inconclusive due to cluster sizes being highly dependent on nonspecific affinities. Thus, establishing a clear relation between cluster shape and RT requires eliminating affinity as a variable. In turn, that would require a separate MD model. Furthermore, clusters are not fixed formations; they are highly dynamic. Their sizes, conformations, orientations, and even the DNA segments they interact with change over time in the system. Moreover, clusters can merge and split, making it rather complicated to track individual clusters and make accurate analyses.

Additionally, the utilization of dimeric TF models may have contributed to the multi-exponential behavior in two ways. First, our definition of unbound TF requires both binding domains not to be in touch with any DNA molecule. In other words, partial bindings to DNA (i.e., bindings with only one binding domain) are equivalent to full bindings in terms of being considered bound. The difference in stability of partial binding and full binding may indeed lead to distinct residence behaviors contributing to multi-exponential behavior. Moreover, divalent interactions are significantly more prone to facilitated dissociation (Kamar et al., 2017; Chen et al., 2018). Since FD is pronounced at high concentrations (Koşar et al., 2022), multi-exponential behavior could be partially attributed to FD as it can facilitate the unbinding of the exposed TFs (i.e., free or surface) more than the unexposed TFs (i.e., core).

Noise in gene expression is an important factor driving cellular heterogeneity (Liu et al., 2019). Gene expression noise could lead different cells in a homogeneous population to distinct phenotypes even under the same environmental conditions (Raser and O'Shea, 2005). This resulting cellular heterogeneity may provide evolutionary advantages to cells. The primary driving event of gene expression noise is considered to be TF binding (Parab et al., 2022). Also, infrequent or rare biochemical processes contribute to noise in gene expression (Raser and O'Shea, 2005). Such infrequency or lower occurrence could be seen for the higher duration residences in our systems, which are more apparent at high nonspecific affinity and high protein concentration cases. Therefore, such cases could lead to higher gene expression noise and yield a more heterogeneous population of cells. It should also be noted that longer residence times are more likely to cause transcriptional bursts (Raser and O'Shea, 2005), which is another notion that contributes to noise in gene expression. On the other hand, shorter DNA residence times are suggested to lower gene expression noise (Azpeitia and Wagner, 2020). Our study sheds light on DNA residence distributions and may explain the diverse noise in gene expression responses.

In this study, we demonstrate that TFs follow multi-exponential patterns with discrete affinities in our MD model. We discuss how binding affinity and concentration of TFs dictate these behaviors. There are indeed some implications of this behavior on biological systems. Isolated from cellular complexity and continuum model, model homodimeric TFs exhibit multi-exponential patterns even with single and binary affinities. This type of behavior might be one of the underlying reasons for gene expression noise and consequent cellular heterogeneity and contribute to cellular differentiation. Moreover, the distributions of TF-DNA residence times may help explain discrete transcriptional bursts. Lastly, DNA-protein clusters in bacterial chromosomes could drive Topologically Associated Domain-like domain formations and further affect the regulation of gene expression. Overall, our findings might contribute to a more comprehensive understanding of gene expression and regulation.

Methods

Modelling the system

We used a modified version of our previous coarse-grained model bacterial system mimicking an *E. coli* bacterium. The model system includes a fairly relaxed chromosome with uniformly distributed binding sites confined within cell wall-like boundaries resembling that of a rod-shaped bacterium. Generic homodimeric transcription factors with various concentrations $(10 - 60\mu M)$ were placed randomly in the volume created by confinement.

Modelling the DNA polymer

For coarse-grained modeling, we used 10bp $\approx$ 1 bead bead "Kremer-Grest" (KG) model for DNA where 1 bead has a diameter of 1σ in Lennard-Jones (LJ) units corresponding to $\sim 3.4nm$. We set the persistent length to 15 beads corresponding to $50nm$ consistent with double helix DNA molecule. We established an $N = 12000$ KG bead DNA model where we maximized the number of binding sites and left sufficient spacing for DNA segmental flexibility to minimize the impacts by bridging the binding sites. The initial circular DNA structure was compacted using self-attractive forces to reduce its size to fit within the available volume, similar to our previous work (Koşar et al., 2022). The binding sites are composed of three beads (as opposed to two beads in our previous work), increasing the likelihood of maximum interaction with the proteins. 150 binding sites are placed uniformly along DNA with 80-bead spacing.

Constructing the confinement

To mimic the rod-like shape of the bacterial cell wall, we first built an open-ended cylinder (or simply a pipe) with a diameter of $R = 2r$ and a height of $2R$. Then, two semi-spheres with radiuses of r are used as caps to close the endings of the cylinder. The overall structure

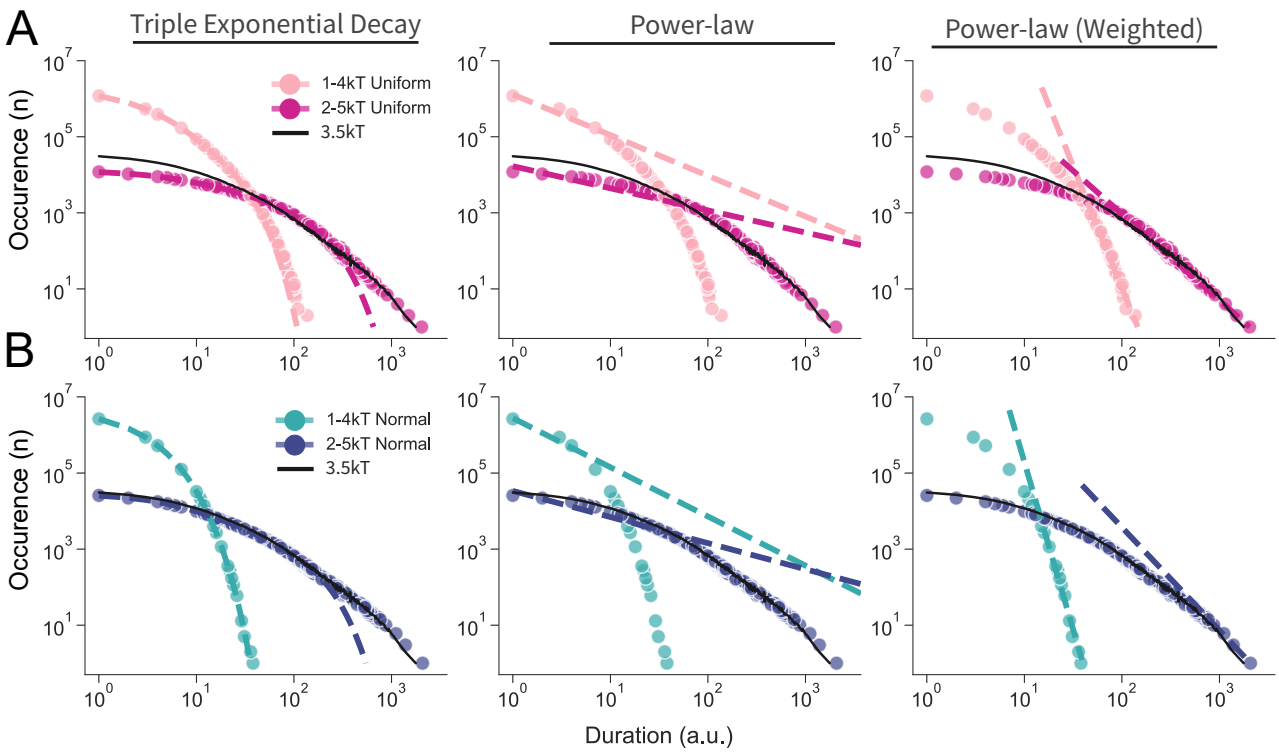

Figure 6. TF residence time behaviors at distributed affinities
(**A**) Uniform distribution and (**B**) Normal or Gaussian distribution affinity models. Pattern analysis via fitting of triple exponential, power-law, and weighted power-law equations from left to right respectively. TF concentrations were $60\mu M$ for all cases. Black lines are shown as reference from fixed nonspecific affinity of $3.5kT$. Specific binding affinities are $4kT$ for all the systems.

has a volume of $2/3 \times \pi \times R^3$. The radius ($r$) was set to provide a 1% DNA volume fraction to match that of *E. coli*. Therefore, r corresponds to $\sim 30\sigma$ in LJ units ($\sim 100nm$) for a DNA polymer length of $N = 12000$. The fixed beads of the cell wall are placed dense enough to provide effective boundaries for DNA and proteins. Nevertheless, there were an insignificant number of TF leaks ($< 2\%$) for extended simulations.

Modelling of the transcription factors

We employed a generic model of homodimeric TFs in which a TF has two identical binding domains and a hinge domain with no affinity to DNA. Binding domains were placed around the hinge domains with $90°$ angles in a semi-flexible fashion with 12 beads of persistent length. Radiuses were set to 0.5σ and 0.21σ for hinge and binding domains, respectively (Figure 1A). We used relatively small binding domains to minimize multiple interactions by single binding domains and prevent over-condensation of the nucleoid. The exact sizes and angles were used to ensure the well-fittings of dimeric TFs into three-bead binding sites. We employed four different concentration levels (i.e.,$10\mu M$, $20\mu M$, $40\mu M$, $60\mu M$). The corresponding number of TFs for the given concentrations was calculated using the volume provided by the confinement. TFs were distributed into the volume at random coordinates.

Modelling the TF-DNA affinities

We used a fixed specific interaction potential of $U_{sp} = 4kT$ per bead, energy high enough for robust binding also low enough to allow unbinding in the timeframe of our simulations, enabling us to extract residence time patterns. Varying nonspecific interaction potentials in the range of $1-4kT$ allowed tracking of residence times at distinct local compaction levels and diverging cluster sizes, as well as differing protein distribution over DNA. For the distributed affinity cases, there was no specific binding potential. Instead, affinities inclusively followed the given ranges of $1-4kT$ and $2-5kT$ with the increment of $0.25kT$, resulting in 13 distinct affinity levels throughout DNA to ensure either Gaussian or uniform distributions.

Calculation of residence durations

Transcription factors are considered bound under the condition that at least one of two binding sites is in direct contact with the DNA polymer. For each timestep, each protein is marked either bound or unbound. Then, the duration of each uninterrupted bound state was calculated (Figure 1B). Each time a particular residence duration was encountered, the corresponding occurrence was incremented by 1. We then utilized pooled occurrences, containing the number of occurrences for each possible duration ($1 - tmax$), for analyzing residence patterns of transcription factors from simulations with distinct parameters. Obtained data is then minimized to ensure one occurrence value has only one correspond-

ing duration by taking the median of the durations. This step was necessary for equation fits and visualizations.

Fitting equations to the residence time distributions

We used several equations to interpret the behavior of transcription factors. Initially, we used single exponential decay $c \times \exp(-t \times k)$ and power-law $c \times t^{-k}$ where c stands for coefficient and k for decay rate or τ^{-1}. Due to suboptimal fitting with these equations, we included double, triple, quadruple, and pentuple exponential decays in the form of $c_1 \times \exp(-t \times k_1) + ... + c_n \times \exp(-t \times k_n)$. Resulting fits were graphed and used for calculating the normalized Residual Sum of Squares (RSS). Normalization was achieved by dividing RSS by the number of durations with non-zero occurrence points to prevent more observations leading to higher total RSS.

Clustering the transcription factors

Transcription factors within the threshold distance of each other are accepted to be part of the same cluster. We set the threshold distance to 2.1σ in LJ units in agreement with visual inspections. For the overview of the system, the minimum number of transcription factors to form a cluster was 12, and for decay analysis, it was 20, ensuring a more accurate estimation of mean lifetimes.

Cluster analysis

Cluster analysis included size, conformation, surface analysis, and decay rates. The size of a cluster is simply the number of transcription factors within that particular cluster. To classify the conformation (or shape) of a cluster, we first determined three possible formations, namely filamentous, globular, and semi-filamentous. We formed Rg tensor for each cluster and evaluated their eigenvalues. To distinguish surface and core cluster proteins, we employed the Convex-Hull algorithm. Decay rates (reverse of mean lifetimes) for clusters were calculated by fitting a single exponential decay equation to $N(t)/N_0$. Decay rates require statistically meaningful numbers for reliable analysis of lifetimes, which led us to select 20 as the threshold for minimum cluster size. We then used linear regression for correlation analysis between cluster dissipation mean lifetimes and their sizes.

Code Availability

The codes necessary to reproduce the analyses and the graphs are available at https://github.com/Zaf4/residence2.

Acknowledgment

ZK thanks Alper Genceroglu, Irem Niran Çagıl, Leyla Yalçınkaya, Tutku Muratoglu, and Fatma Chafra for their review of the initial manuscript.

References

Agarwal, H., Reisser, M., Wortmann, C., and Gebhardt, J. C. M. (2017). Direct observation of cell-cycle-dependent interactions between ctcf and chromatin. *Biophysical Journal*, 112. doi: 10.1016/j.bpj.2017.04.018.

Agback, P., Baumann, H., Knapp, S., Ladenstein, R., and Härd, T. (1998). Architecture of nonspecific protein–dna interactions in the sso7d–dna complex. *Nature Structural Biology 1998 5:7*, 5:579–584. doi: 10.1038/836.

Amemiya, H. M., Goss, T. J., Nye, T. M., Hurto, R. L., Simmons, L. A., and Freddolino, P. L. (2022). Distinct heterochromatin-like domains promote transcriptional memory and silence parasitic genetic elements in bacteria. *The EMBO Journal*, 41:e108708. doi: 10.15252/EMBJ.2021108708.

Arold, S., Parkinson, G., Ladbury, J., and Leonard, P. (2010). H-ns forms a superhelical protein scaffold for dna condensation. *Proc. Natl Acad. Sci. USA*, 107:15728–15732. doi: 10.1073/pnas.1006966107.

Azam, T. A., Iwata, A., Nishimura, A., Ueda, S., and Ishihama, A. (1999). Growth phase-dependent variation in protein composition of the escherichia coli nucleoid. *Journal of Bacteriology*, 181:6361–6370. doi: 10.1128/jb.181.20.6361-6370.1999.

Azpeitia, E. and Wagner, A. (2020). Short residence times of dna-bound transcription factors can reduce gene expression noise and increase the transmission of information in a gene regulation system. *Frontiers in Molecular Biosciences*, 0:67. doi: 10.3389/FMOLB.2020.00067.

Ball, C. A., Osuna, R., Ferguson, K. C., and Johnson, R. C. (1992). Dramatic changes in fis levels upon nutrient upshift in escherichia coli. *Journal of Bacteriology*, 174:8043–8056. doi: 10.1128/JB.174.24.8043-8056.1992.

Ball, D. A., Mehta, G. D., Salomon-Kent, R., Mazza, D., Morisaki, T., Mueller, F., McNally, J. G., and Karpova, T. S. (2016). Single molecule tracking of ace1p in saccharomyces cerevisiae defines a characteristic residence time for non-specific interactions of transcription factors with chromatin. *Nucleic Acids Research*, 44. doi: 10.1093/nar/gkw744.

Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. Transcriptional regulation by the numbers: Models, (2005). ISSN 0959437X.

Brackley, C. A., Taylor, S., Papantonis, A., Cook, P. R., and Marenduzzo, D. (2013). Non-specific bridging-induced attraction drives clustering of dna-binding proteins and genome organization. *Proceedings of the National Academy of Sciences of the United States of America*, 110. doi: 10.1073/PNAS.1302950110.

Browning, S. B. D. (2004). The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol.*, 2:57–65.

Chen, J., Zhang, Z., Li, L., Chen, B. C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., Tjian, R., and Liu, Z. (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*, 156. doi: 10.1016/j.cell.2014.01.062.

Chen, T.-Y., Cheng, Y.-S., Huang, P.-S., and Chen, P. (2018). Facilitated unbinding via multivalency-enabled ternary complexes: New paradigm for proteindna interactions. doi: 10.1021/acs.accounts.7b00541.

Clauß, K., Popp, A. P., Schulze, L., Hettich, J., Reisser, M., Torres, L. E., Uhlenhaut, N. H., and Gebhardt, J. C. M. (2017). Dna residence time is a regulatory factor of transcription repression. *Nucleic Acids Research*, 45:11121–11130. doi: 10.1093/NAR/GKX728.

Cook, P. R. A model for all genomes: The role of transcription factories, (2010). ISSN 00222836.

Dame, R., Wyman, C., and Goosen, N. (2000). H-ns mediated compaction of dna visualised by atomic force microscopy. *Nucleic Acids Res.*, 28:3504–3510. doi: 10.1093/nar/28.18.3504.

Dame, R. T. (2005). The role of nucleoid-associated proteins in the organization and compaction of bacterial chromatin. *Molecular Microbiology*, 56:858–870. doi: 10.1111/J.1365-2958.2005.04598.X.

de Jonge, W. J., Brok, M., Lijnzaad, P., Kemmeren, P., and Holstege, F. C. (2020). Genome-wide off-rates reveal how dna binding dynamics shape transcription factor function. *Molecular systems biology*, 16. doi: 10.15252/MSB.20209885.

de Jonge, W. J., Patel, H. P., Meeussen, J. V., and Lenstra, T. L. Following the tracks: How transcription factor binding dynamics control transcription, (2022). ISSN 15420086.

Dillon, S. C. and Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology 2010 8:3*, 8:185–195. doi: 10.1038/nrmicro2261.

Dorman, C. J. (2014). Function of nucleoid-associated proteins in chromosome structuring and transcriptional regulation. *Microbial Physiology*, 24:316–331. doi: 10.1159/000368850.

Elgin, S. C. and Grewal, S. I. (2003). Heterochromatin: silence is golden. *Current Biology*, 13:R895–R898. doi: 10.1016/J.CUB.2003.11.006.

Garcia, D. A., Fettweis, G., Presman, D. M., Paakinaho, V., Jarzynski, C., Upadhyaya, A., and Hager, G. (2021). Power-law behavior of transcription factor dynamics at the single-molecule level implies a continuum affinity model. *Nucleic Acids Research*, 49:6605–6620. doi: 10.1093/nar/gkab072.

Graham, J. S., Johnson, R. C., and Marko, J. F. (2011). Concentration-dependent exchange accelerates turnover of proteins bound to double-stranded dna. *Nucleic Acids Research*, 39:2249–2259. doi: 10.1093/NAR/GKQ1140.

Haberle, V. and Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation, (2018). ISSN 14710080.

Hammel, M., Amlanjyoti, D., Reyes, F. E., Chen, J. H., Parpana, R., Tang, H. Y., Larabell, C. A., Tainer, J. A., and Adhya, S. (2016). Hu multimerization shift controls nucleoid compaction. *Science Advances*, 2. doi: 10.1126/SCIADV.1600650.

Hipp, L., Beer, J., Kuchler, O., Reisser, M., Sinske, D., Michaelis, J., Gebhardt, J. C. M., and Knöll, B. (2019). Single-molecule imaging of the transcription factor srf reveals prolonged chromatin-binding kinetics upon cell stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, 116. doi: 10.1073/pnas.1812734116.

Iborra, F. J., Pombo, A., Jackson, D. A., and Cook, P. R. (1996). Active rna polymerases are localized within discrete transcription 'factories' in human nuclei. *Journal of Cell Science*, 109. doi: 10.1242/jcs.109.6.1427.

Inukai, S., Kock, K. H., and Bulyk, M. L. Transcription factor–dna binding: beyond binding site

motifs, (2017). ISSN 18790380.

Joshi, C. P., Panda, D., Martella, D. J., Andoy, N. M., Chen, T. Y., Gaballa, A., Helmann, J. D., and Chen, P. (2012). Direct substitution and assisted dissociation pathways for turning off transcription by a merr-family metalloregulator. *Proceedings of the National Academy of Sciences of the United States of America*, 109:15121–15126. doi: 10.1073/PNAS.1208508109.

Kamar, R. I., Banigan, E. J., Erbas, A., Giuntoli, R. D., Cruz, M. O. D. L., Johnson, R. C., and Marko, J. F. (2017). Facilitated dissociation of transcription factors from single dna binding sites. *Proceedings of the National Academy of Sciences of the United States of America*, 114:E3251–E3257. doi: 10.1073/PNAS.1701884114.

Kim, S. and Shendure, J. (2019). Mechanisms of interplay between transcription factors and the 3d genome. *Molecular Cell*, 76:306–319. doi: 10.1016/J.MOLCEL.2019.08.010.

Koşar, Z. and Erbaş, A. (2022). Can the concentration of a transcription factor affect gene expression? *Frontiers in Soft Matter*, 0:6. doi: 10.3389/FRSFM.2022.914494.

Koşar, Z., Attar, A. G., and Erbaş, A. (2022). Facilitated dissociation of nucleoid-associated proteins from dna in the bacterial confinement. *Biophysical journal*, 121:1119–1133. doi: 10.1016/J.BPJ.2022.03.002.

Lee, K. A. Dimeric transcription factor families: It takes two to tango but who decides on partners and the venue?, (1992). ISSN 00219533.

Lickwar, C. R., Mueller, F., Hanlon, S. E., McNally, J. G., and Lieb, J. D. (2012). Genome-wide protein–dna binding dynamics suggest a molecular clutch for transcription factor function. *Nature 2012 484:7393*, 484:251–255. doi: 10.1038/nature10985.

Lin, J., Chen, H., Dröge, P., and Yan, J. (2012). Physical organization of dna by multiple non-specific dna-binding modes of integration host factor (ihf). *PLOS ONE*, 7:e49885. doi: 10.1371/JOURNAL.PONE.0049885.

Liu, J., François, J.-M., and Capp, J.-P. (2019). Gene expression noise produces cell-to-cell heterogeneity in eukaryotic homologous recombination rate. *Frontiers in Genetics*, 10. doi: 10.3389/fgene.2019.00475.

Melnik, S., Deng, B., Papantonis, A., Baboo, S., Carr, I. M., and Cook, P. R. (2011). The proteomes of transcription factories containing rna polymerases i, ii or iii. *Nature Methods*, 8. doi: 10.1038/nmeth.1705.

Mitchell, J. A. and Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes and Development*, 22. doi: 10.1101/gad.454008.

Morisaki, T., Müller, W. G., Golob, N., Mazza, D., and McNally, J. G. (2014). Single-molecule analysis of transcription factor binding at transcription sites in live cells. *Nature Communications*, 5. doi: 10.1038/ncomms5456.

Noort, J. V., Verbrugge, S., Goosen, N., Dekker, C., and Dame, R. T. (2004). Dual architectural roles of hu: Formation of flexible hinges and rigid filaments. *Proceedings of the National Academy of Sciences of the United States of America*, 101:6969–6974. doi: 10.1073/PNAS.0308230101.

Parab, L., Pal, S., and Dhar, R. (2022). Transcription factor binding process is the primary driver of noise in gene expression. *PLoS genetics*, 18:e1010535. doi: 10.1371/JOURNAL.PGEN.1010535.

Penagos-Puig, A. and Furlan-Magaril, M. Heterochromatin as an important driver of genome organization, (2020). ISSN 2296634X.

Raser, J. M. and O'Shea, E. K. (2005). Noise in gene expression: Origins, consequences, and control. *Science*, 309:2010–2013. doi: 10.1126/science.1105891. Original paper for "gene expression NOISE".

Reisser, M., Hettich, J., Kuhn, T., Popp, A. P., Große-Berkenbusch, A., and Gebhardt, J. C. M. (2020). Inferring quantity and qualities of superimposed reaction rates from single molecule survival time distributions. *Scientific Reports*, 10. doi: 10.1038/s41598-020-58634-y.

Remesh, S. G., Verma, S. C., Chen, J. H., Ekman, A. A., Larabell, C. A., Adhya, S., and Hammel, M. (2020). Nucleoid remodeling during environmental adaptation is regulated by hu-dependent dna bundling. *Nature Communications 2020 11:1*, 11:1–12. doi: 10.1038/s41467-020-16724-5.

Seshasayee, A. S. N., Sivaraman, K., and Luscombe, N. M. (2011). An overview of prokaryotic transcription factors: A summary of function and occurrence in bacterial genomes. *Sub-Cellular Biochemistry*, 52:7–23. doi: 10.1007/978-90-481-9069-0_2/FIGURES/2_2_209873_1_EN.

Skoko, D., Wong, B., Johnson, R. C., and Marko*, J. F. (2004). Micromechanical analysis of the binding of dna-bending proteins hmgb1, nhp6a, and hu reveals their ability to form highly stable dnaprotein complexes†. *Biochemistry*, 43:13867–13874. doi: 10.1021/BI048428O.

Skoko, D., Yoo, D., Bai, H., Schnurr, B., Yan, J., McLeod, S. M., Marko, J. F., and Johnson, R. C. (2006). Mechanism of chromosome compaction and looping by the escherichia coli nucleoid protein fis. *Journal of Molecular Biology*, 364:777–798. doi: 10.1016/J.JMB.2006.09.043.

Verma, S. C., Qian, Z., and Adhya, S. L. (2019). Architecture of the escherichia coli nucleoid. *PLOS Genetics*, 15:e1008456. doi: 10.1371/JOURNAL.PGEN.1008456.

Wang, W., Li, G. W., Chen, C., Xie, X. S., and Zhuang, X. (2011). Chromosome organization by a nucleoid-associated protein in live bacteria. *Science*, 333:1445–1449. doi: 10.1126/SCIENCE.1204697.

Winardhi, R. S., Yan, J., and Kenney, L. J. (2015). H-ns regulates gene expression and compacts the nucleoid: Insights from single-molecule experiments. *Biophysical Journal*, 109:1321–1329. doi: 10.1016/J.BPJ.2015.08.016.