Protein search processes mediated by chromatin topology

Shuvadip Dutta[1], Adarshkrishnan R.[1], Ranjith Padinhateeri[2],[*] and Mithun K. Mitra[1][†]

[1]*Department of Physics, Indian Institute of Technology Bombay, Mumbai 400076, India*

[2]*Department of Biosciences and Bioengineering,*

Indian Institute of Technology Bombay, Mumbai 400076, India

Abstract

We investigate the role of compaction of chromatin domains in modulating search kinetics of proteins. Collapsed conformations of chromatin, characterised by long loops which bring distant regions of the genome into contact, and manifested structurally as Topologically Associated Domains (TADs) affect search kinetics of DNA associated transcription factors and other proteins. In this study, we investigate the role of the compactness of chromatin on the dynamics of proteins using a minimal model. Using analytical theory and simulations, we show that an optimal compaction exists for which the residence time of proteins on a chromatin-like polymer backbone is minimum. We show that while bulk diffusion is an advantageous search strategy for extended polymers, for highly folded polymer domains, intersegmental transfers allow optimal search. We extend these results to more detailed polymer models - using the Freely Rotating Chain model, a Lennard-Jones bead-spring polymer model, which approximates chromatin behavior. We show that our results continue to hold for these polymer models, with a minimum residence time at an optimum polymer compaction. Finally, we also analyse the dynamics of proteins on networks generated using experimental chromatin conformation data from 8355 TADs extracted from human chromosomes. Our analysis suggests that TADs exist near this zone of optimality, indicating that chromatin conformations can play a crucial role in modulating protein search strategies.

[*] ranjithp@iitb.ac.in

[†] mkmitra@iitb.ac.in

I. INTRODUCTION

The mechanism of target search by DNA-associated proteins such as Transcription Factors (TFs) on DNA backbones remains an open question [1–4]. Simple three-dimensional diffusion to locate specific binding sites cannot account for search timescales, necessitating proposals of facilitated diffusion processes [5–8]. Four distinct search mechanisms have been proposed to explain the dynamic motion of proteins along DNA: 1D sliding, hopping or jumping, 3D diffusion, and intersegmental transfer [9]. A combination of 3D and 1D diffusion has been shown to lead to optimal search times compared to a pure 1D or a pure 3D search [5]. However, experimental evidence points to a preponderance of sliding modes of transport, with only very small time spent in the bulk (SI Table S1) [10, 11]. In contrast, intersegmental hops or "direct transfers" have been observed experimentally [12], especially in densely packed nucleoids. This assumes particular significance for protein motion in mammalian cells, where highly packed chromatin regions with high density and viscosity hinder free diffusion in the solution phase [13], and multivalent proteins, due to nonspecific interactions, slow the mobility of other searching proteins, limiting their capacity for free 3D diffusion [14].

In this crowded milieu, intersegmental transfer from one segment of DNA to another via long loops assumes particular significance as a mechanism for effective search, which does not include free diffusion in the bulk [9]. While most studies assume that intersegmental transfer moves the protein to an uncorrelated segment, and thus has the same effect as 3D diffusion [2], this is not true for densely packed chromatin, where backbone correlations persists over long length scales, and impact the intersegmental transfer dynamics. Numerous theoretical and experimental studies have delved into the hierarchical organization of the genome across various length scales [15–21]. In mammals, the genome is organised into compact regions, such as Topologically Associating Domains (TADs) [22–24]. TADs are structural and functional units of the genome characterised by large squares of increased contact frequency tiling the contact matrices' diagonal [23] (Fig. 1a). Concurrent with this high intra-domain contacts inside TADs, is a relative insulation among adjacent domains through TAD boundaries [25, 26]. TAD boundaries, marked by the enrichment of H3K4Me3 and the presence of abundant Pol2 [27], serve as hubs where the promoter regions of numerous transcriptional units are strategically located. These enhanced contacts within TADs

result in a higher occurrence of long intersegmental hops, consequently influencing the kinetics of the search processes of proteins on chromatin [28]. Indeed, search processes on polymer backbones have shown surprising results for protein kinetics [8], both in presence and absence of polymer fluctuations [29, 30]. DNA conformations and looped structures such as rosettes have been predicted to affect search kinetics [31–33].

In this Letter, we investigate the role of 3D chromatin organization, in particular within folded TAD-like regions, in the target search process by proteins. We show how the search times of proteins crucially depend on the topology of polymer domains characterised by long-range intersegmental contacts. In biologically relevant regimes, we show that 3D diffusion is not a efficient search strategy compared to search mediated through intersegmental jumps. Finally, we analyse available experimental data to show that conformations of human TADs are optimised for efficient search.

II. A NETWORK REPRESENTATION OF CHROMATIN

We first consider a contact space network representation of a polymer domain (SI Sec. 1.1, Fig. S1). Two beads i and j ($1 \leq i, j \leq L-1$) are assumed to be in contact with a probability $P_c(s) = cs^{-\gamma}$, where $s = |i-j| > 1$. A protein performs a random walk (SI Sec. 1.2) along the contact space, and can also dissociate to the bulk from any site with a probability p_{off} (Fig. 1b,c). The target sites are located just outside of this polymer domain, at $i = 0, L$, motivated by the presence of transcriptional hubs at the TAD boundaries (Fig. 1a). A protein in the bulk can attach randomly to any site of the polymer, including the target sites, with a mean exploration time of τ_f in the solution phase before it reattaches to the polymer.

We compute the configuration-averaged mean search times $\langle T_{\text{search}} \rangle$, defined as the time a protein takes to reach either boundaries of the polymer for the first time (SI Sec. 1.3). The mean search time can be computed as

$$\mathbb{T} = \mathbb{A}^{-1}\mathbb{B}$$

where $\mathbb{T}_i$ is the time taken to reach the target starting from the ith site. The matrices $\mathbb{A}$

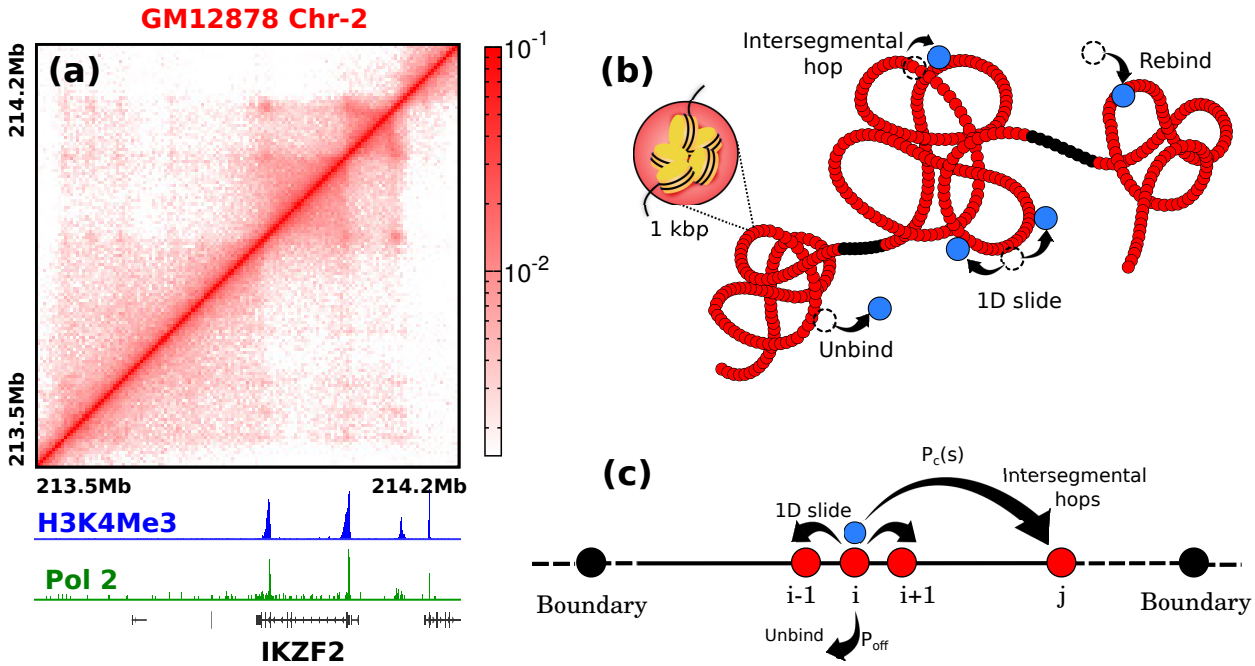

FIG. 1. **(a)** A representative contact domain shows TADs as square blocks of increased interaction frequencies [34]. The appearance of a TAD is often accompanied by H3K4Me3 and Pol-2 enrichment peaks at the boundaries, as well as the activation of a gene whose promoter lies at one of the peak loci. **(b)** Coarse-grained (1 kbp) polymer representation of chromatin with compact TAD regions (red) separated by relatively extended boundary regions (black). **(c)** A contact space representation of TAD-like polymer domains. Each bead is connected via bonds to its neighbor along the backbone. In addition, bonds between genomically distant but spatially neighbouring beads allow for intersegmental jumps. Black beads denote the boundaries of the TAD.

and $\mathbb{B}$ are defined by,

$$\mathbb{A}_{ij} = \delta_{i,j} - \frac{1-p_{\text{off}}}{N_i} C_{ij} - p_{\text{off}}\delta_{j,f} \;\; \forall i \neq 0, L, f$$
$$\mathbb{A}_{fj} = \delta_{f,j} - \frac{1-\delta_{f,j}}{L+1} \qquad \mathbb{A}_{ij} = \delta_{i,j} \text{ for } i = 0, L$$
$$\mathbb{B}_i = \tau + (\tau_f - \tau)\delta_{i,f} - \tau(\delta_{i,0} + \delta_{i,L})$$

where, τ is the time taken for a single step, C_{ij} denotes the connectivity matrix, and N_i denotes the total number of bonds for the ith site. Note that the polymer backbone connectivity implies that $C_{i,i-1} = C_{i,i+1} = 1$ necessarily. The site $i = f$ denotes the freely diffusing state in solution. For simplicity, we assume the protein starts from the middle of the polymer domain, $\langle T_{\text{search}} \rangle = \langle \mathbb{T}_{L/2} \rangle$.

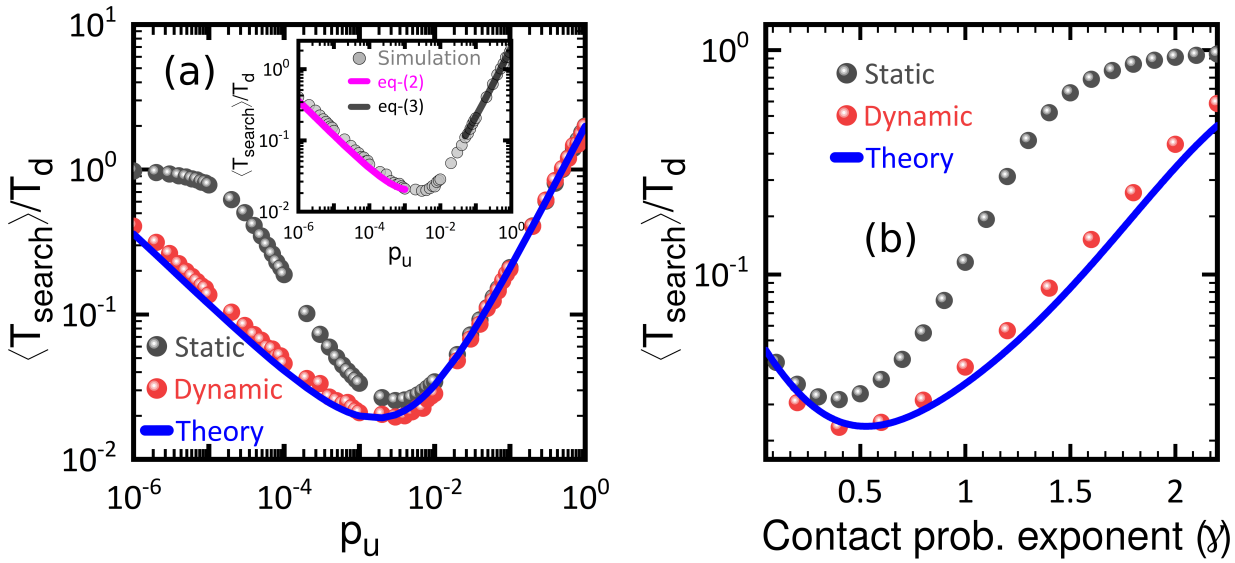

FIG. 2. **(a)** Scaled mean search time ($\frac{\langle T_{\text{search}}\rangle}{T_d}$) for static (black) and dynamic (red) conformations in a uniformly connected network domain vs probability of contact (p_u) between any two non-neighbouring beads for $L = 500$. The solid blue line is the analytical solution of Eq. 1. The inset shows the low-p_u and high-p_u approximations from Eqs. 3 and 4. **(b)** ($\frac{\langle T_{\text{search}}\rangle}{T_d}$) in power-law connected network domain as a function of contact probability exponent(γ) for $L = 500$ also shows non-monotonic behaviour for both static (black) and dynamic (red) configurations.

Motivated by experimental results of very high estimates of time a protein spends non-specifically bound to the DNA [10, 11], and consistent with high volume fractions inside TADs [35] (SI Sec. 5,Fig. S12), we first consider the limiting case $p_{\text{off}} = 0$, so that target search happens only along the polymer backbone, via sliding and intersegmental jumps. For simplicity, we first assume $\gamma = 0$, so that all beads have a uniform connection probability $P_c(s) = c \equiv p_u$. We observe a non-monotonic relation between $\langle T_{\text{search}}\rangle$ and p_u (Fig. 2a). When $p_u = 0$, the protein performs a 1D random walk, giving $\langle T_{\text{search}}\rangle = \frac{L^2}{8D} \equiv T_d$ where $D = \frac{1}{2\tau}$ [36]. As p_u increases, more paths to reach the boundaries emerge, reducing $\langle T_{\text{search}}\rangle$. However, beyond a threshold connectivity, $\langle T_{\text{search}}\rangle$ increases due to excess intersegmental bonds diverting the walker. This non-monotonic behavior of $\langle T_{\text{search}}\rangle$ remains robust even when the polymer network undergoes dynamic re-configurations (SI Sec. 2, Fig. S2) at a fixed connection probability (Fig. 2a red symbols, and SI Fig. S3a)). Dynamic rewiring of network connections reduces $\langle T_{\text{search}}\rangle$, with faster configuration changes yielding lower $\langle T_{\text{search}}\rangle$ at low p_u (SI Fig. S3a). However, at high p_u values, network dynamicity offers

little advantage, as most nodes already have numerous long-range connections. The optimal behavior of $\langle T_{\text{search}} \rangle$ is consistent for all polymer length (SI Fig. S4), differential rates for sliding and intersegmental jumps (SI Fig. S3b), and differential dynamic rewiring rates (SI Fig. S3a).

The non-monotonic behaviour persists when we use a power law connectivity with $\gamma \neq 0$. For a self-avoiding polymer, $\gamma = 2.2$, which reduces to $\gamma = 1.5$ for a RW polymer. For chromosomes, on an average, $\gamma \approx 1.08$ [37], while within highly folded TADs, experiments report even smaller values for γ [38]. As a polymer undergoes collapse, contacts become more frequent, reducing the contact probability exponent γ. We observe a non-monotonic behaviour of the search times for both and static and dynamic realisations of such power-law networks, as shown in Fig. 2b. Again, when the polymer is relatively open (larger γ), dynamic rewiring of the connectivities helps reduce search times, similar to the observation from the simpler uniform connectivity ($P_c(s) = p_u$) case.

For the case of the dynamic polymer, where the polymer undergoes rewiring at a fixed connection probability p_u, we can solve for the mean search times analytically. The mean search times ($T_k(p_u)$) for a protein starting from the kth site obeys,

$$\mathbb{T}_k = \tau + r_s(p_u)\left(\mathbb{T}_{k+1} + \mathbb{T}_{k-1}\right) + \sum_{j \neq nn} r_h(p_u)\mathbb{T}_j. \tag{1}$$

The total number of bonds for any site i is $N_i = 2 + p_u(L-3)$, and all hops are equally likely. The probability for sliding to any of the two neighbouring sites is then $r_s(p_u) = 1/N_i$ and the probability of hopping to a non-neighbouring site via a intersegmental hop is $r_h(p_u) = p_u/N_i$ on average (SI Sec. 3). This can be solved exactly, subject to the absorbing boundary conditions at the two ends to yield a closed form solution (SI Sec. 3.2).

$$\langle T_{\text{search}} \rangle = \frac{\beta}{\alpha} + 2e^{\frac{L}{2}\sqrt{\alpha}} \left(\frac{\phi_1 \tau + \frac{\phi_2 \beta}{\alpha}}{\phi_3} \right) \tag{2}$$

where

$$\alpha = \frac{1 + 3r_h - 2r_s}{r_s - r_h}$$

$$\beta = \frac{\tau \left(1 + 2r_h(1 + r_h) + \frac{2r_h \phi_1 \psi(1+2r_h)}{\phi_3} + 2r_h(r_s - r_h)\frac{\phi_1 \phi_4}{\phi_3}\right)}{r_s - r_h - \frac{r_h(L-3)(1+2r_h)}{\alpha} - \frac{2r_h \phi_2 \psi(1+2r_h)}{\alpha \phi_3} - \frac{2r_h(r_s - r_h)}{\alpha}\left(1 + \frac{\phi_2 \phi_4}{\phi_3}\right)}$$

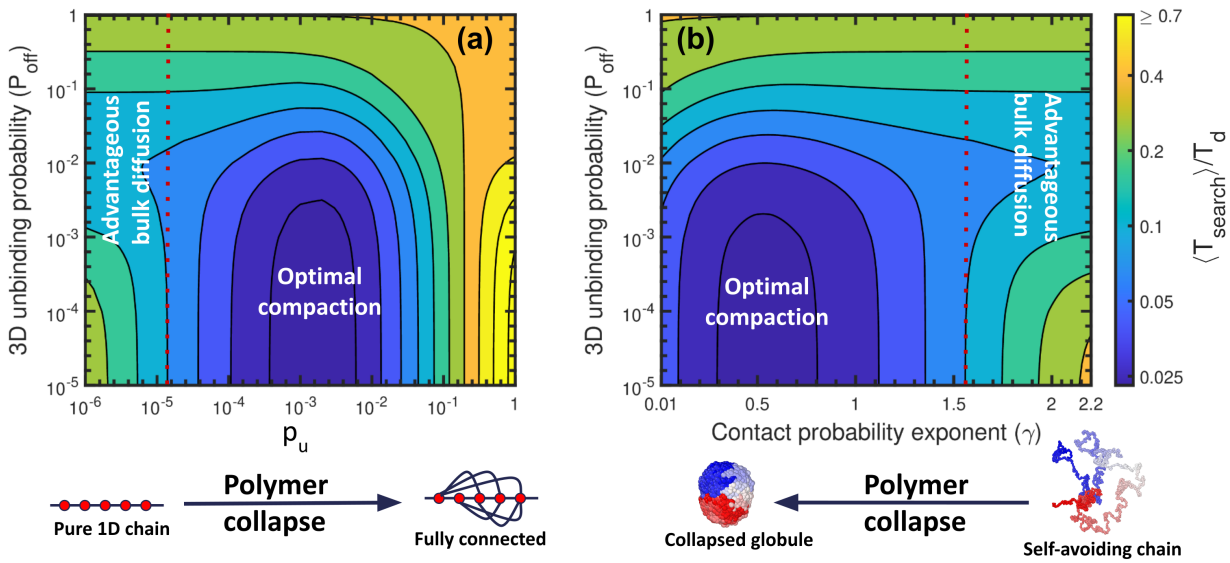

FIG. 3. Contour plot of $\langle T_{\text{search}}\rangle/T_d$ as a function of 3D unbinding probability (p_{off}) and polymer compaction for **(a)** a uniformly connected domain and **(b)** a power-law connected domain of $L = 500$. In a relatively open chain (see left side in **(a)** and right side in **(b)**, there is an optimal p_{off} for which $\langle T_{\text{search}}\rangle$ is minimal suggesting advantage of 3D excursion. However, as polymer collapses, $\langle T_{\text{search}}\rangle$ is minimum for the smallest p_{off} suggesting that 3D excursions are not advantageous. For biological p_{off} regimes, there always exists an optimal polymer compaction where $\langle T_{\text{search}}\rangle$ is minimum. The fact that the global minimum of $\langle T_{\text{search}}\rangle$ occurs for highly folded conformations and in regions where unbinding is low suggests the role of 3D structure in search process (see text).

$$\phi_1 = 1 + (r_s + r_h)(1 + r_h)$$

$$\phi_2 = (1 + r_s + r_h)r_h(L - 3) + r_s^2 + r_s - (r_h + 1)^2$$

$$\phi_3 = \left[1 + r_h - r_s^2 + r_h^2\right]\left(e^{2\sqrt{\alpha}} + e^{(L-2)\sqrt{\alpha}}\right) - 2\psi r_h(1 + r_s + r_h) - (r_s - r_h)\left(e^{3\sqrt{\alpha}} + e^{(L-3)\sqrt{\alpha}}\right)$$

$$\phi_4 = e^{2\sqrt{\alpha}} + e^{(L-2)\sqrt{\alpha}}$$

$$\psi = \left[e^{2\sqrt{\alpha}}\left(\frac{e^{(L-3)\sqrt{\alpha}} - 1}{e^{\sqrt{\alpha}} - 1}\right)\right]$$

The analytic matches expression exactly with the simulation results (blue solid line in Fig 2a). For low values of p_u, the governing equation can be simplified to obtain a solution

(SI Sec. 3.2.1),

$$\frac{\langle T_{\text{search}}\rangle}{T_d} \overset{p_u\to 0}{=} \frac{4\tau}{L^2(\alpha(r_s-r_h)-r_h\rho)}\left[1-\frac{2e^{\frac{L}{2}\sqrt{\alpha}}}{1+e^{L\sqrt{\alpha}}}\right] \tag{3}$$

where, $\rho = (L-1) - \frac{2e^{\sqrt{\alpha}}(e^{\sqrt{\alpha}(L-1)}-1)}{(1+e^{L\sqrt{\alpha}})(e^{\sqrt{\alpha}}-1)}$ and $\alpha = \frac{1+3r_h-2r_s}{r_s-r_h}$. To leading order, then, we have,

$$\frac{\langle T_{\text{search}}\rangle}{T_d} \overset{p_u\to 0}{\approx} 1 - \frac{(L^3-5L^2-20L+48)}{48}p_u \quad \text{for } p_u \ll \frac{1}{L^2}$$

a decreasing function of p_u. At high connection probability, as $p_u \to 1$, all sites become equivalent, and a simplified governing equation can be solved to yield (SI Sec. 3.2.2),

$$\langle T_{\text{search}}\rangle \overset{p_u\to 1}{=} 2 + (L-3)p_u + \frac{(2+(L-4)p_u)^2}{2p_u} \tag{4}$$

which implies a linear increase with p_u as $p_u \to 1$, also consistent with a solution using the effective medium approach [39] (SI Sec. 3.2.3 and SI Fig. S5,S6). As can be seen from Fig 2a inset, the analytic expression at both small and high p_u limits matches the simulation results.

How does this search strategy mediated by intersegmental jumps compare with the facilitated diffusion in the presence of 3D diffusion ($p_{\text{off}} \neq 0$)? We first look at the case of $\gamma = 0$ (Fig. 4a), so that $P_c(s) = p_u$. At $p_u = 0$, we recover the classical facilitated diffusion result of an optimum search at an intermediate value of p_{off}. As p_u increases, the advantage obtained through bulk diffusion reduces, until beyond a certain threshold, unbinding from the backbone offers no advantage to the target search process. This general result holds true when we look at power-law connectivities (Fig. 4b) as well. For the self-avoiding polymer ($\gamma = 2.2$, 3D diffusion offers an advantage to target search. As the polymer collapses, and hence γ decreases, the relative speed-up through bulk excursions decreases. Beyond the RW polymer ($\gamma \sim 1.5$), unbinding from the backbone offers no advantage to the target search process, as intersegmental hops allow a more efficient mechanism to explore distant sites. Indeed for collapsed polymers (low γ or large p_u), and in the vicinity of the optimal compaction state, remarkably, 3D diffusion is not helpful for target search, with a monotonic increase in search times with increasing unbinding probability (SI Sec. 3.3, Fig. S7).

III. POLYMER REPRESENTATIONS OF CHROMATIN

We now examine whether these effects persist when canonical polymer models are used to model chromatin structure. Thus the allowed intersegmental jumps arise directly from

polymer topology, and hence random walks on these backbones have direct relevance to protein search processes.

A. Freely Rotating Chain Model

We first study the Freely Rotating Chain (FRC) model which has been used to investigate chromatin behaviour [40–43]. Neighbouring bonds in an FRC polymer are constrained to lie on a cone subtending an angle θ [44]. The angle θ controls the compaction of the polymer, with large θ leading to a collapsed state (see SI Fig. S8A). As the polymer collapses, the number of non-neighbouring bonds N_b increases monotonically (see SI Fig. S8B), indicating a greater probability of intersegmental jumps. We simulated search on these FRC backbones (see SI Sec. 4.1), and the mean residence times $\langle T_{\text{search}} \rangle$ shows a non-monotonic behaviour with increasing bond angle (Fig. 4A). This can be also interpreted as a function of polymer compactness or collapse. Thus in line with the expectations from the network model of polymer, there exists an optimal compaction where residence times are minimum.

B. Lennard-Jones polymer model

We next use a bead-spring polymer model of chromatin with attractive Lennard-Jones (LJ) potential [45–47]. This introduces self-avoidance effects, and bond length fluctuations absent in the FRC model. Any pair of beads of diameter σ, separated by distance r, interacts via the potential,

$$E_{LJ}(r) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] & r < 2.5\sigma \\ 0 & r \geq 2.5\sigma \end{cases} \tag{5}$$

where ϵ is the LJ interaction strength (see SI Sec. 4.2). A LJ polymer undergoes a second-order phase transition from a coiled to a compact state (See SI Fig. S9A), accompanied by an increase in the number of intersegmental bonds (see SI Fig. S9B), as ϵ is increased. The residence times $\langle T_{\text{search}} \rangle$ on these LJ polymer backbones exhibited a non-monotonic pattern as the LJ interaction strength is increased (Fig. 4B), or equivalently as the polymer compaction increases. There was a distinct minimum observed in the residence times, occurring close to the collapse transition. Thus a LJ polymer qualitatively mirrors the results obtained from the uniform network and FRC polymer models.

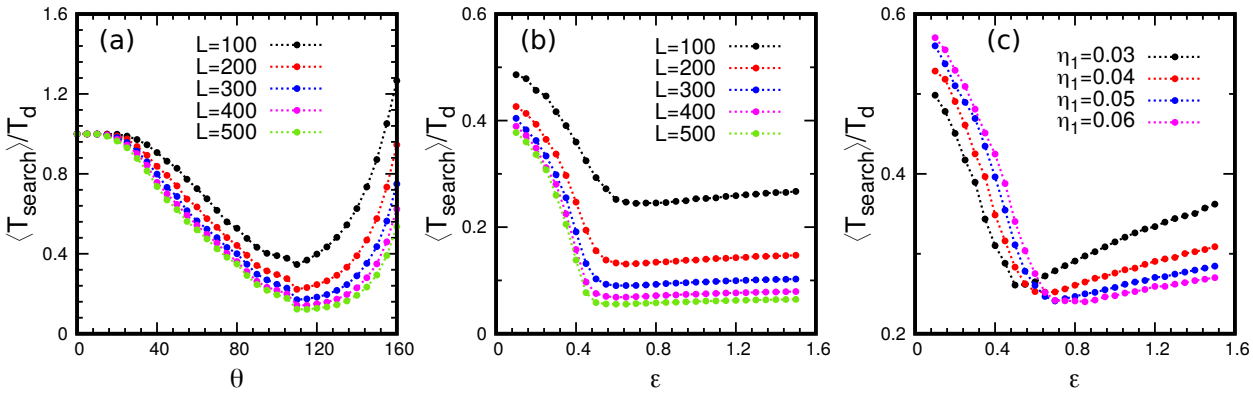

FIG. 4. **(a)** Scaled mean residence time ($\frac{\langle T_{\text{search}}\rangle}{T_d}$) in FRC polymer domain as a function of angle (θ) between any two bond vectors for $L = 100 - 500$. **(b)** $\frac{\langle T_{\text{search}}\rangle}{T_d}$ in LJ polymer domain as a function of LJ interaction strength (ϵ) between any two non-bonded beads for $L = 100 - 500$. **(c)** $\frac{\langle T_{\text{search}}\rangle}{T_d}$ in soft LJ polymer domain as a function of soft LJ interaction strength (ϵ) between any two non-bonded beads for $L = 100$ with varying softness characterised by η_1.

However, unlike the significant rise seen in the FRC and the network model, the extent of $\langle T_{\text{search}}\rangle$ increase beyond a critical ϵ is less pronounced in this case. This limited increase is due to the inherent hard-core repulsion in the LJ potential, which imposes a lower limit on polymer compaction, and on the number of intersegmental bonds. This interplay between LJ interactions and polymer structure thus introduces a unique feature that distinguishes it from both FRC and the network model.

C. Soft Lennard-Jones polymer model

Is this weak non-monotonic rise in $\langle T_{\text{search}}\rangle$ then the correct physical description of residence times? The answer hinges on whether hardcore LJ potential best represents coarse-grained chromatin conformations. Experimental and theoretical studies of chromatin configurations and 3D distances suggest that a soft inter-bead potential that allows for overlap is a more realistic description of coarse-grained chromatin [20, 48]. We study a polymer model with following soft inter-bead potential,

$$E_{soft}(r) = \begin{cases} V_0[1 - (\frac{r}{r_m})^{\eta_1}]^{\eta_2} - \epsilon & 0 \leq r < r_m \\ 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] & r_m \leq r < 2.5\sigma \\ 0 & r \geq 2.5\sigma \end{cases} \tag{6}$$

where $V_0 - \epsilon$ is the energy penalty at complete overlap, and η_1 and η_2 tune the softness of the potential (see SI Sec. 4.3 and SI Fig. S10). We investigate $\langle T_{\text{search}} \rangle$ in polymers with varying degrees of softness by tuning η_1. A lower value of η_1 indicates a softer polymer, whereas high η_1 resembles the classical LJ potential. For a very soft polymer ($\eta_1 = 0.03$), $\langle T_{\text{search}} \rangle$ shows a strong non-monotonicity, analogous to the FRC and the network results. This is because a soft polymer core cannot effectively enforce volume exclusion, resulting in a more tightly packed polymer compared to the conventional LJ model (see SI Fig. S11A). This then leads to a higher number of non-neighboring contacts within the polymer domain (see SI Fig. S11B). As we decrease the softness (increasing η_1), the degree of non-monotonicity in the residence times decreases and approaches the LJ-like behaviour. Thus non-monotonic behaviour of residence times remains a crucial feature of protein diffusion on coarse-grained chromatin-like soft polymer topologies. The degree of non-monotonicity $\langle T_{\text{search}} \rangle$ is controlled by the softness of the inter-bead potential.

IV. POLYMERS FROM TAD DATA

Is this non-monotonic behaviour of search times relevant for real TADs? Where do TAD conformations lie with respect to this optimal compaction state? To answer these, we utilized publicly accessible experimental Hi-C contact probability data [34] and performed the following theoretical study. We analysed Hi-C data for 8355 TADs at 1kbp resolution from GM12878 cell line Fig. 5a. For each TAD of length L, the contact probability curve was fitted to the equation $P_c(s) = cs^{-\gamma}$ in the region $s = 10$ to $(L - 100)$ kbp Fig. 5b. The distributions of the best fitting values of c and γ are shown in 5c,d and yield $\langle c \rangle = 0.0214$ and $\langle \gamma \rangle = 0.71$ (most probable gamma values are in the range $\approx (0.66, 0.75)$) (SI Sec. 6, Fig. S13). We computed protein target search times for wide range of L and γ on a power-law network with $c = \langle c \rangle$ and is shown as contour map in 5e. Strikingly, when we plot the best fitting γ values along with the length L for each TAD in the same Fig. 5e we find that they

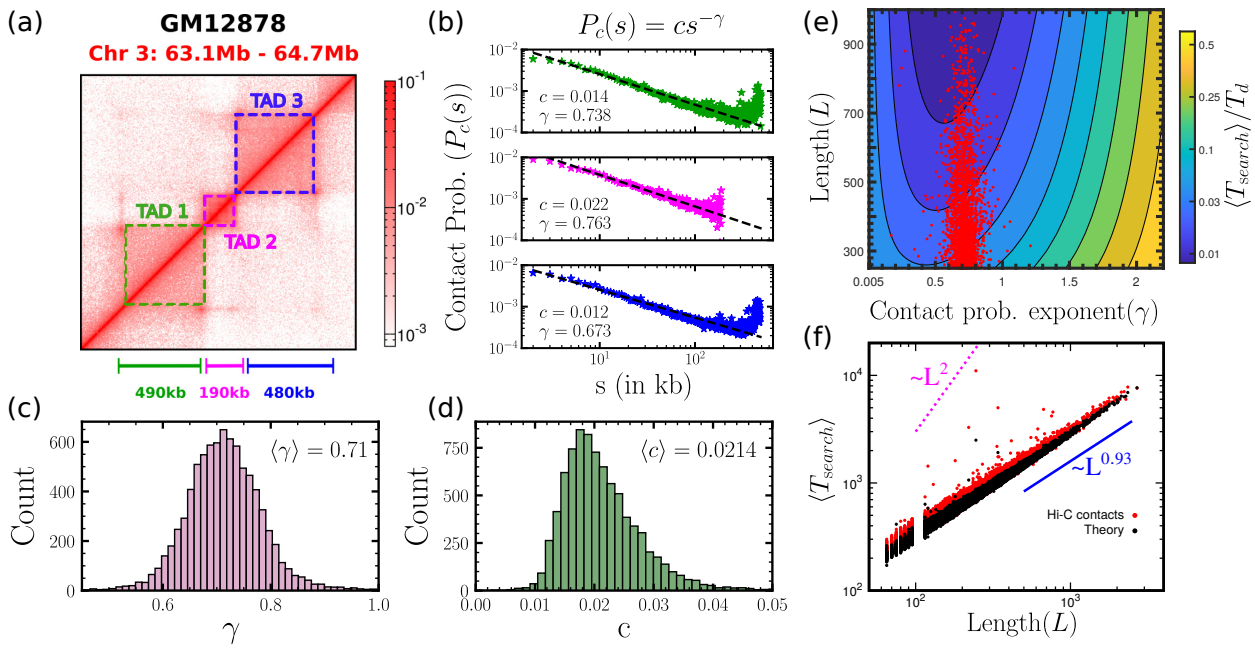

FIG. 5. **(a)** An example of 3 distinct TAD regions in Hi-C contact map from [34]. **(b)** Contact probability ($P_c(s)$) as a function of genomic distance (s), fitted using $P_c(s) = cs^{-\gamma}$ for the three TADs in (a). **(c-d)** Distribution of best fit parameters c and γ for all 8355 collected TADs from GM12878 data [34]. **(e)** Contour plot of mean search times (scaled) as a function of γ and L. The fact that experimentally observed [34] L and γ values (red dots) fall very close to the minimum is yet another evidence that 3D chromatin structure is relevant in the search processes. **(f)** Search times with TAD lengths show a very different scaling compared to purely 1D diffusive search. Search time computed from simulating walks on network with Hi-C contact probability (red dots) and network with observed power law properties (black) show $\langle T_{\text{search}}\rangle \sim L^{0.93}$.

fall near the optimal search times (see red dots). This suggests that TAD structures for all human TADs are optimised for efficient search.

To investigate this further we generated chromatin network by following two different procedures: (i) power-law networks for each TAD of length L by taking the best fitting c and γ values from above, and (ii) direct Hi-C networks for all TADs by inserting connections directly with experimentally measured contact probabilities P_{ij}. While the former gives accurate average contact probability for each $s = |i-j|$, the latter gives us accurate sequence-specific contact probability for every pair (i, j). We performed the search simulations on both networks and the results are shown in Fig. 5f. Note that, both the results (red,black) give

comparable search times and the scaling with TAD lengths is close to ballistic ($\langle T_{\text{search}} \rangle \sim L^{0.93}$). This implies that the TADs represented by red dots in Fig. 5e not only have the appropriate (L, γ) but also have the optimal search times. This suggests that random walks on these TADs can lead to large speed-ups compared to the diffusive timescales ($\sim L^2$) purely due to the underlying TAD structure (Fig. 5f). Further, despite sequence-specific effects, search times on these Hi-C networks is accurately captured by power-law network models, and they offer a true estimate of times expected for search processes on chromatin backbones.

V. DISCUSSION

In conclusion, we propose that facilitated diffusion which takes into account correlated intersegmental jumps can leverage the compaction state of chromatin, characterized by the presence of TADs and looped polymer conformations, to access distant genomic regions for more efficient search. This is in line with direct evidence from single-molecule experiments which have reported that search process by a site-specific restriction enzyme on DNA becomes almost doubly efficient when the DNA configuration is collapsed instead of a fully extended configuration [49]. Remarkably, contrary to the prevailing picture of bulk diffusion mediated search, we show that for highly folded polymer conformations, such as in chromatin, bulk diffusion hinders the search process, possibly offering a reconciliation with experimental estimated of very large times a protein spends non-specifically bound to DNA. An analysis of connection probabilities from experimentally determined Hi-C data suggests that TADs reside near this zone of optimal compaction where search times are minimised. The structural organization of the genome within TADs may thus have evolved to facilitate rapid and precise dynamics of DNA-associated proteins, which is crucial for gene regulation and cellular function.

VI. ACKNOWLEDGEMENT

SD acknowledges fellowship support from the PMRF, MoE, India. MKM and RP acknowledge funding from Core Research Grant by Science and Engineering Research Board,

India (Grant number: CRG/2022/008142).

---

[1] A. D. Riggs, S. Bourgeois, and M. Cohn, Journal of molecular biology **53**, 401 (1970).

[2] O. G. Berg, R. B. Winter, and P. H. Von Hippel, Biochemistry **20**, 6929 (1981).

[3] R. B. Winter, O. G. Berg, and P. H. Von Hippel, Biochemistry **20**, 6961 (1981).

[4] S. E. Halford and J. F. Marko, Nucleic acids research **32**, 3040 (2004).

[5] O. G. Berg and P. H. von Hippel, Annual review of biophysics and biophysical chemistry **14**, 131 (1985).

[6] A. Veksler and A. B. Kolomeisky, The journal of physical chemistry B **117**, 12695 (2013).

[7] L. Mirny, M. Slutsky, Z. Wunderlich, A. Tafvizi, J. Leith, and A. Kosmrlj, Journal of Physics A: Mathematical and Theoretical **42**, 434013 (2009).

[8] O. Bénichou, C. Chevalier, B. Meyer, and R. Voituriez, Physical review letters **106**, 038102 (2011).

[9] J. Gorman and E. C. Greene, Nature structural & molecular biology **15**, 768 (2008).

[10] J. Elf, G.-W. Li, and X. S. Xie, Science **316**, 1191 (2007).

[11] K. Bettridge, F. E. Harris, N. Yehya, and J. Xiao, The Journal of Physical Chemistry B **127**, 3816 (2023).

[12] R. D. Giuntoli, N. B. Linzer, E. J. Banigan, C. E. Sing, M. O. de la Cruz, J. S. Graham, R. C. Johnson, and J. F. Marko, Journal of Molecular Biology **427**, 3123 (2015).

[13] C. M. Caragine, S. C. Haley, and A. Zidovska, Phys. Rev. Lett. **121**, 148101 (2018).

[14] K. Kamagata, R. Kusano, S. Kanbayashi, T. Banerjee, and H. Takahashi, Nucleic Acids Research **51**, 6654 (2023), https://academic.oup.com/nar/article-pdf/51/13/6654/50903029/gkad471.pdf.

[15] Y. Qi and B. Zhang, PLoS computational biology **15**, e1007024 (2019).

[16] G. Shi and D. Thirumalai, Physical Review X **11**, 011051 (2021).

[17] M. Di Pierro, B. Zhang, E. L. Aiden, P. G. Wolynes, and J. N. Onuchic, Proceedings of the National Academy of Sciences **113**, 12168 (2016).

[18] T.-H. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, O. J. Rando, R. Tjian, and X. Darzacq, Molecular cell **78**, 539 (2020).

[19] T. Misteli, Cell **183**, 28 (2020).

[20] S. Kadam, K. Kumari, V. Manivannan, S. Dutta, M. K. Mitra, and R. Padinhateeri, Nature Communications **14**, 4108 (2023).

[21] S. Guha and M. K. Mitra, Soft Matter **19**, 153 (2023).

[22] E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. Van Berkum, J. Meisig, J. Sedat, *et al.*, Nature **485**, 381 (2012).

[23] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, Nature **485**, 376 (2012).

[24] Q. Szabo, F. Bantignies, and G. Cavalli, Science advances **5**, eaaw1668 (2019).

[25] A. L. Sanborn, S. S. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, *et al.*, Proceedings of the National Academy of Sciences **112**, E6456 (2015).

[26] D. Jost, P. Carrivain, G. Cavalli, and C. Vaillant, Nucleic acids research **42**, 9553 (2014).

[27] T.-H. Hsieh, A. Weiner, B. Lajoie, J. Dekker, N. Friedman, and O. Rando, Cell **162**, 108 (2015).

[28] M. Mazzocca, A. Loffreda, E. Colombo, T. Fillot, D. Gnani, P. Falletta, E. Monteleone, S. Capozi, E. Bertrand, G. Legube, *et al.*, Nature Communications **14**, 6433 (2023).

[29] I. M. Sokolov, J. Mai, and A. Blumen, Phys. Rev. Lett. **79**, 857 (1997).

[30] T. Schötz, R. A. Neher, and U. Gerland, Phys. Rev. E **84**, 051911 (2011).

[31] C. A. Brackley, M. E. Cates, and D. Marenduzzo, Physical review letters **109**, 168103 (2012).

[32] A. Mondal and A. Bhattacherjee, The Journal of Physical Chemistry B **121**, 9372 (2017).

[33] K. Mondal and S. Chaudhury, The Journal of Physical Chemistry B **124**, 3518 (2020).

[34] S. S. Rao, M. H. Huntley, N. C. Durand, E. K. Stamenova, I. D. Bochkov, J. T. Robinson, A. L. Sanborn, I. Machol, A. D. Omer, E. S. Lander, *et al.*, Cell **159**, 1665 (2014).

[35] H. D. Ou, S. Phan, T. J. Deerinck, A. Thor, M. H. Ellisman, and C. C. O'shea, Science **357**, eaag0025 (2017).

[36] S. Redner, *A Guide to First-Passage Processes*, A Guide to First-passage Processes (Cambridge University Press, 2001).

[37] E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, and J. Dekker, Science **326**, 289 (2009), https://www.science.org/doi/pdf/10.1126/science.1181369.

[38] B. Chan and M. Rubinstein, Proceedings of the National Academy of Sciences **120**, e2222078120 (2023).

[39] V. Sood, S. Redner, and D. ben Avraham, Journal of Physics A: Mathematical and General **38**, 109 (2004).

[40] H. Schiessel, W. M. Gelbart, and R. Bruinsma, Biophysical Journal **80**, 1940 (2001).

[41] G. Bajpai, I. Jain, M. M. Inamdar, D. Das, and R. Padinhateeri, PLoS computational biology **13**, e1005365 (2017).

[42] G. Bajpai and R. Padinhateeri, Biophysical Journal **118**, 207 (2020).

[43] A. J. Spakowitz, The Journal of Chemical Physics **151** (2019).

[44] M. Rubinsten, *Polymer physics* (United States of America, 2003).

[45] C. A. Brackley, J. Johnson, S. Kelly, P. R. Cook, and D. Marenduzzo, Nucleic acids research **44**, 3503 (2016).

[46] A. Buckle, C. A. Brackley, S. Boyle, D. Marenduzzo, and N. Gilbert, Molecular cell **72**, 786 (2018).

[47] S. Shin, G. Shi, and D. Thirumalai, PRX Life **1**, 013010 (2023).

[48] S. Fujishiro and M. Sasai, Proceedings of the National Academy of Sciences **119**, e2109838119 (2022).

[49] B. van den Broek, M. A. Lomholt, S.-M. Kalisch, R. Metzler, and G. J. Wuite, Proceedings of the National Academy of Sciences **105**, 15738 (2008).