



The effect of configurational complexity in hetero-polymers on the coil-globule phase transition

Fabrizio Tafuri, Andrea M. Chiariello^a

Dipartimento di Fisica, Università di Napoli Federico II, and INFN Napoli, Complesso Universitario di Monte Sant'Angelo, 80126 Naples, Italy

Received: 2 November 2022 / Accepted: 5 February 2023
© The Author(s) 2023

Abstract The coil-globule transition of hetero-polymer chains is studied here. By means of extensive Molecular Dynamics simulations, we show that the transition is directly linked to the complexity of the chain, which depends on the number of chemical species defined in the environment and the location of the binding sites along the polymer. In addition, when the number of species increases, we find that the distribution of binding sites plays an important role in triggering the transition, beyond the standard control parameters of the polymer model, i.e. binders concentration and binding affinity. Overall, our results show that by increasing the system complexity new organizational layers emerge, thus allowing a more structured control on the polymer thermodynamic state. This can be potentially applied to the study of chromatin architecture, as such polymer models have been broadly used to understand the molecular mechanisms of genome folding.

1 Introduction

Understanding how genome folds in the nucleus of cells is one of the most important questions in Molecular Biology. Indeed, the development of powerful experimental technologies [1] allowed to show that chromosomes exhibit a complex architecture, occurring at different genomic length scales in the cell nucleus [2–5]. Specifically, chromatin is folded into Mb-sized domains called topologically associating domains (TADs, [6, 7]); TADs have internal structure (sub-TADs) and are included in larger domains, named A and B compartments [8], associated with open and active chromatin and more compact and inactive chromatin, respectively. Also, TADs form a non-random hierarchy known as metaTAD structure [9]. Such spatial structure of chromatin is linked to cellular activities. For example, distant regulatory elements (enhancers) located along DNA sequences trigger the functions of target genes through physical contact. Anyway, the mechanisms governing such complex and precise organization are not fully understood.

To tackle the organization problem at different length scales, polymer models have been employed. Some models consider chromatin organization in a non-equilibrium state, e.g. the fractal globule (FG) model [10, 11], a transient state occurring during the collapse of the polymer chain, or emerging from non-equilibrium, active processes in cell nuclei [12, 13]. Analogously, loop extrusion (LE) model [14, 15] considers the formation of chromatin loops through the action of active motors (typically cohesin or condensin) that slides along the chromatin filament and blocking at anchor points (converging CTCF binding sites). Other models are based instead on equilibrium mechanisms, such as micro-phase separation of chromatin [16], chromatin interacting with proteins [17, 18] or chromatin interacting with phase-separated protein clusters [19, 20]. In any case, polymer physics models provide quantities of interest useful to understand the complexity of such spatial organization as they can be directly compared with experiments. For example, by means of microscopy-based experiments (e.g. by FISH or STORM methods [21]) the spatial distances between several loci can be simultaneously measured and compared with model predictions [19, 22]. In addition, the Hi-C technique [8] allows to measure the frequency with which two DNA fragments physically co-localize in space. Again, from the Hi-C contact matrix, we can calculate quantities (e.g. the mean contact probability of two regions as a function of their linear distance) that can be compared with the model and then reveal information about the physical properties of the chromatin in a specific genomic region.

In this work, we focus on the String and Binders Switch (SBS) polymer model [18], which has been successfully employed to describe several features of genome folding, as contact probability scaling [23], spatial organization of real loci [24] or multiple contacts [19]. In the SBS scenario, the contacts between distant regions are mediated by diffusive molecules (named binders) that can attractively interact with binding sites located along the polymer. The position and the number of types of binding sites define the complexity of the polymer and can be conveniently set to describe different contexts, as simple homo-polymers (a uniform polymer with just one type of binding sites), block co-polymers (two or more types of binding types arranged in linear segments along the

^a e-mail: chiariello@na.infn.it (corresponding author)

polymer) or hetero-polymers with multiple binding sites located along the polymer. In this regard, computational methods [25–27] have been designed to customize polymer models aiming to accurately describe genome architecture. Using machine learning approaches fed with experimental data, e.g. Hi-C contact matrices, it is possible to infer a polymer model that best reproduces the degree of compaction of a specific genomic region of interest, i.e. the minimum number of different types of binding sites and their spatial distribution along the hetero-polymer chain [28].

In this work, we use massive parallel Molecular Dynamics (MD) simulations to explore, in a highly controlled polymer system, the role of configurational complexity of hetero-polymers in the coil-globule phase transition. We start from a simple, uniform homo-polymer and gradually increase the complexity of the chain by introducing different types of binding sites. Then, we study the physical properties of the system in each condition at equilibrium, focusing on the thermodynamic state of the polymer (coil or globule). We find that the folding state of an alternating hetero-polymer is affected by the number of types of binding sites, for fixed binding affinity and binder concentration. Indeed, more complex alternating hetero-polymers require higher interaction affinities to fold into the globular state. Furthermore, we find that polymers exhibit different equilibrium states when different arrangements of the binding sites are considered. Hetero-polymers with random, less complex, arrangements of binding sites can transit into the globular state, whereas hetero-polymers with regular alternating sequences of different types do not, being the other control parameters unchanged. This implies that the folding state of complex polymers may be controlled, beyond the binder concentration and binding affinity, through the configuration of the binding sites along the polymer chain. In a biological framework, different binding types are typically mapped into different molecular signatures. Therefore, our study suggests that their abundance found in cell nuclei could be another regulatory mechanism to control genome folding and gene activity.

2 Results

2.1 The system

To investigate the general properties of the coil-globule phase transition in hetero-polymers, we consider a polymer model equipped with binding sites able to attractively interact with diffusive molecules (binders) which populate the surrounding environment. Biologically, the polymer represents a chromatin filament, the binding sites represent genomic regulatory elements, as enhancer, promoters or binding sites associated with protein binding motifs and the binders represent the molecular factors that constantly bind to the chromatin in the cell nucleus (such as transcription factors) [17]. In a very simple version of this model (generally known as the SBS model, [18]), the polymer consists of one binding sites type and inert, non-interacting beads (schematically represented as red and grey bead, respectively, in Fig. 1a). Nevertheless, as several different proteins exist in cell nuclei, the model can be naturally generalized to envisage the existence of more binding site types, associated with an equal number of cognate types of binders. In any case, the system control parameters are the binder concentration c and binding affinity E_{int} . If both c and E_{int} are above threshold, the system undergoes the coil-globule phase transition [18, 23] and the polymer collapses in a stable, globular phase. In the following, we will explore such thermodynamic behaviour by varying the polymer complexity. Specifically, when the number of types of binding sites is increased (schematically, the number of colours), the polymer is considered more complex. If the number of types is fixed (i.e. the number of colours), then their arrangement along the sequence is important. In this case, sequences of binding sites regularly distributed according to some ordering scheme (e.g. alternating, see Sect. 2.2) are considered more complex than polymers with randomly distributed binding sites (see Sect. 2.3). Details about parameters of the simulations can be found in the Methods section.

2.2 Phase transitions of alternating hetero-polymer chains

We first study the phases of hetero-polymers as a function of the number types of binding sites, with fixed binder concentrations and binding affinities. For this purpose, we simulate the time dynamics of alternating hetero-polymer chains, where an ordered subsequence of differently coloured beads is repeated along the polymer (Methods). Additional effects of binding sites spatial distributions along the polymer will be considered in the following sections. The simplest model of alternating SBS hetero-polymer is the uniform homo-polymer, where the non-interacting bead (grey) and binding site (red) alternate along the string (Fig. 1a). We simulated the MD dynamics of such system by varying the binding affinity, for fixed binder concentration. The goal was to identify the affinity E_{int} values for which the homo-polymer chain can undergo the coil-globule transition. As usual, we used the gyration radius ([23], Methods) to monitor the transition during time. Indeed, when the transition occurs, the polymer size is reduced, and the gyration radius sharply drops (Fig. 1, b). In agreement with the predictions of Polymer Physics ([29, 30]), three thermodynamic states are found, with the transition threshold (Θ -point) identified around $E_{\text{int}} = 4.3k_{\text{B}}T$. Consistently, in each phase, the scaling exponents for mean contact probability and mean square distance (Methods) are compatible with the theoretical value (Fig. 1c, d). Then, we gradually increase the complexity of the chains by adding more types (colours) of binding sites to the sequence. As first step, we use an alternating chain in which there are two types of binding sites (red and green) and use the same binder concentration for each type. To compare the behaviour at equilibrium of different hetero-polymer chains in the same condition, we set the binding affinity $E_{\text{int}} = 4.8 k_{\text{B}}T$ which gave the transition for the chain with a single type of binding site. As before, we observe that with this affinity

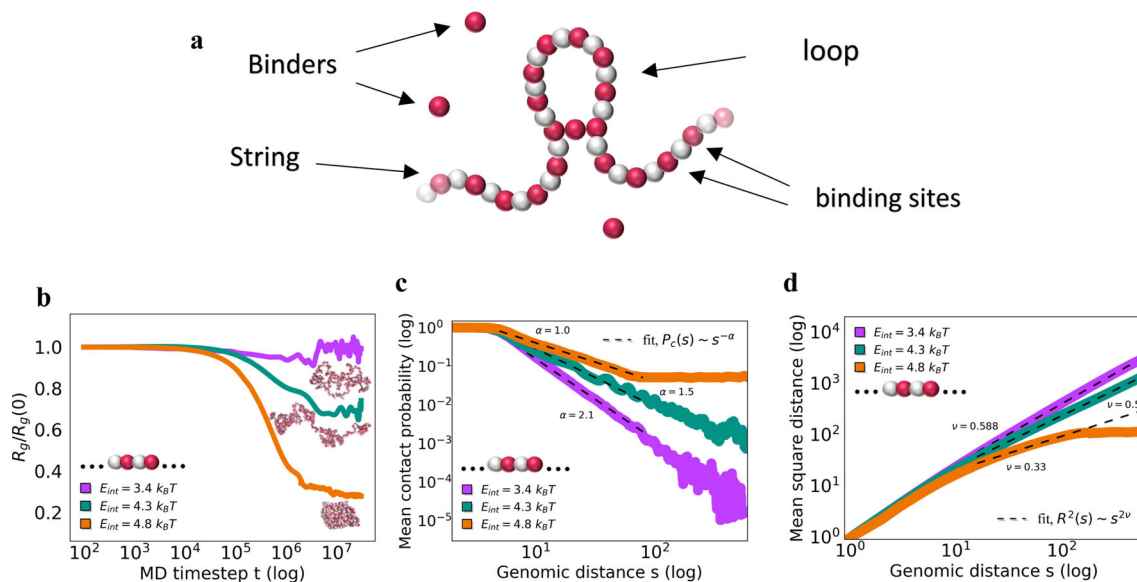


Fig. 1 The coil-globule phase transition of a hetero-polymer SBS model. **a** A schematic cartoon of an SBS hetero-polymer model is shown. In this case, one type of binding site, coloured in red, is present. **b, c, d** Thermodynamic phases of an alternating SBS polymer with one type of binding site, as function of the binding affinity E_{int} . **b** The time dynamics of normalized gyration radius shows that the transition occurs as the binding affinity increases. Examples of 3D snapshots taken from MD simulations are reported. **c** Scaling of mean contact probability $P_c(s) \sim s^{-\alpha}$ in the three different thermodynamic equilibrium phases. **d** Scaling of the mean square distance $R^2(s) \sim s^{2\nu}$. In the open SAW conformation ($E_{int} = 3.4 \text{ k}_B\text{T}$) the scaling exponents result $\alpha = 2.1$, $\nu = 0.588$, in the compact globule state ($E_{int} = 4.8 \text{ k}_B\text{T}$) $\alpha = 1.0$, $\nu = 0.33$ and in Θ -point thermodynamic phase ($E_{int} = 4.3 \text{ k}_B\text{T}$), $\alpha = 1.5$, $\nu = 0.5$. Binder concentration is fixed for all cases

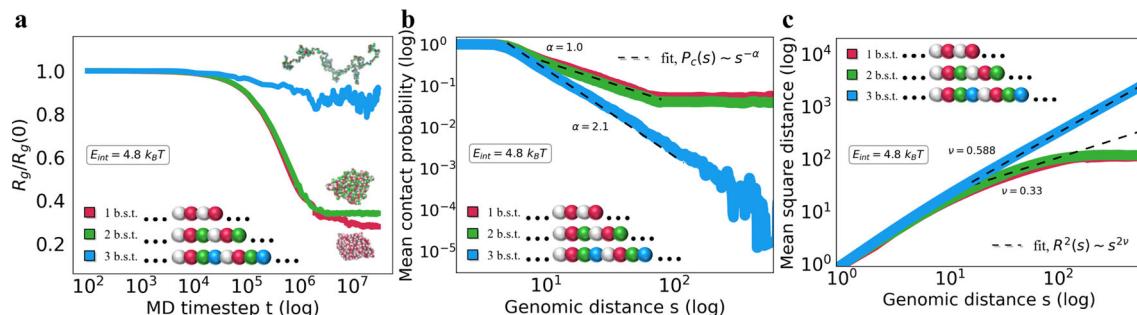


Fig. 2 Number of binding site types is important in the phase transition of an alternating hetero-polymer. The folding state of an alternating hetero-polymer chain depends on the number of binding site types (b.s.t) along the polymer. **a** $R_g/R_g(\text{SAW})$ dynamics is very similar for the 1 and 2 b.s.t. hetero-polymers (red and green curves) where the transition occurs. Conversely, in the 3 b.s.t. alternating polymer (cyan curve) the transition does not occur. Examples of 3D equilibrium structures are also shown. **b** and **c** Equilibrium mean contact probability and mean square distance in the different hetero-polymers

the chain with two types can transit into the globular phase (Fig. 2a, green curve), as confirmed by the mean contact probability and mean square distance (Fig. 2b and c, green curve). By repeating the process, we increase again the complexity and consider a chain with three binding site types (red, green and blue), following the same alternating scheme, at the same concentration c and affinity E_{int} . Unlike the previously considered hetero-polymer chains, in this case the transition does not occur, as no substantial change in the gyration radius is detected from the initial SAW configuration (Fig. 2a, cyan curve). Analogously, the equilibrium $P_c(s)$ and $R^2(s)$ clearly show that the polymer stays in the thermodynamic SAW phase (Fig. 2b and c, cyan curves). Importantly, we verified that those results do not depend on the simulation time, as longer simulations returned analogous results (see Methods).

Overall, our results show that the thermodynamic equilibrium phases of alternating hetero-polymer chains are importantly affected by the number of types of binding site, that is by its complexity. At the binding affinity $E_{int} = 4.8 \text{ k}_B\text{T}$, the alternating hetero-polymer chain undergoes a phase transition to the compact state for only one and two types of binding site; the addition of a third (or obviously more) colours keeps the polymer in the open conformation and higher interaction affinities are therefore required to trigger the transition.

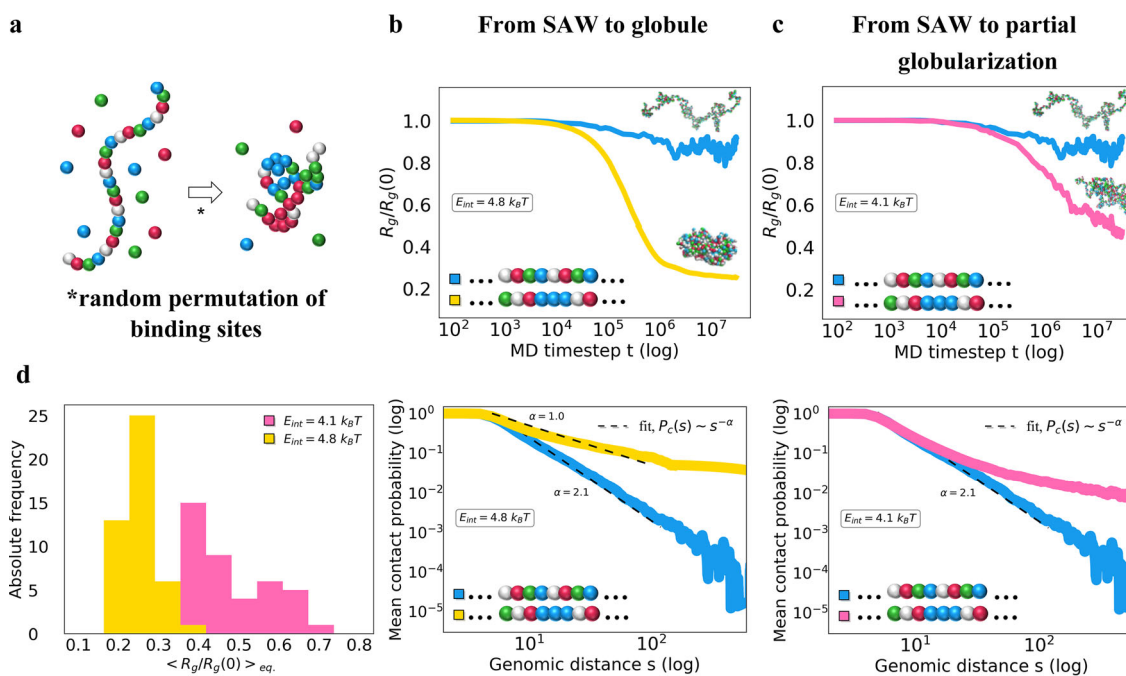


Fig. 3 Rearrangement of binding site positions of regular alternating hetero-polymer induces coil-globule phase transition. **a** Schematic cartoons of alternating and randomly permuted SBS polymers. Alternating hetero-polymers do not undergo the coil-globule transition (left), whereas randomly rearranged hetero-polymers can do (right). **b** Normalized gyration radius dynamics (top) and equilibrium mean contact probability (bottom) of alternating (cyan curve) and random hetero-polymers (yellow curve) at binding affinity $E_{int} = 4.8 k_B T$. Examples of 3D snapshots taken from MD simulation at equilibrium are reported. **c** as in **b**, for $E_{int} = 4.1 k_B T$. Curves are an average over different independent random permutations. **d** Distribution of the normalized gyration radius at equilibrium from independent randomly permuted hetero-polymers with 3 binding site types at affinities $E_{int} = 4.1 k_B T$ (pink) and $E_{int} = 4.8 k_B T$ (yellow)

2.3 Phase transitions of random hetero-polymer chains

We have shown that the number of types along the chain plays an important role in triggering the phase transition. Next, we investigated the potential role in the arrangement of the binding sites in driving the phase transition. To this aim, we focused on hetero-polymer chains with the three afore-mentioned binding site types and rearranged the positions of the beads. As first rearrangement scheme, we considered random permutations of the binding sites, where the bead positions (including the grey ones) are randomly rearranged along the string (i.e. polymers are less complex, Fig. 3a). MD simulations were then performed by keeping binding affinity $E_{int} = 4.8 k_B T$ and binder concentration unchanged with respect to the previous case. Interestingly, the results of the simulations revealed that the rearranged polymer collapses into a stable globular shape, although the transition was not observed for the regular alternating chain. To check the robustness of the results, we considered several (approximately 50) independent random permutations, and the transition was observed in all cases (Fig. 3b, d, Methods). Next, we verified that such effect is not strictly dependent on the specific binding affinity. Indeed, simulations with lower binding affinity (Fig. 3c, $E_{int} = 4.1 k_B T$) exhibit a similar behaviour, although the transition is not complete, and the polymer is only locally compacted (Fig. 3c, Methods). These results point towards a scenario in which the configuration of the binding sites along the chain is another possible, general control mechanism to induce the phase transition (see next section) in this kind of systems. Again, such results have interesting implications for the chromatin architecture, as the regulation of some genomic regions can be orchestrated not only by the mere presence of the regulatory elements, but it is also encoded in their specific position along the genome. The complexity of the polymer therefore allows a more precise control on gene activity. In the next section, we investigate the microscopic origin of this process and give a mechanistic insight to it.

2.4 Microscopic origins of the phase transition

We then investigated which microscopic mechanism originates the different behaviours observed in the alternating and the randomly permuted polymer, although they carry the same number of binding sites for each type. We hypothesized that in the randomly permuted configurations there exist regions with consecutive beads of the same type (not present in the alternating polymer) which likely form more stable interactions with the binders and allow more stable contacts with distant regions, inducing in this way the phase transition of the entire polymer. Therefore, to quantitatively evaluate this hypothesis, we considered the following strategy: in brief, we start from alternating hetero-polymer chains and gradually create pairs of consecutive binding sites, by swapping the position of selected beads (Fig. 4a, see Method for the details of the developed algorithm). Then, we studied how the polymer folding state

Fig. 4 Frequency of homolog pairs favours the phase transition. **a** Swapping scheme used to make hetero-polymers with a controlled number of homolog bead pairs (see Methods). **b** Normalized gyration radius calculated at equilibrium for different frequency of pairs of same type (see Methods). Error bars represent the standard error computed from the ensemble of independent simulations. Examples of 3D snapshots are also shown

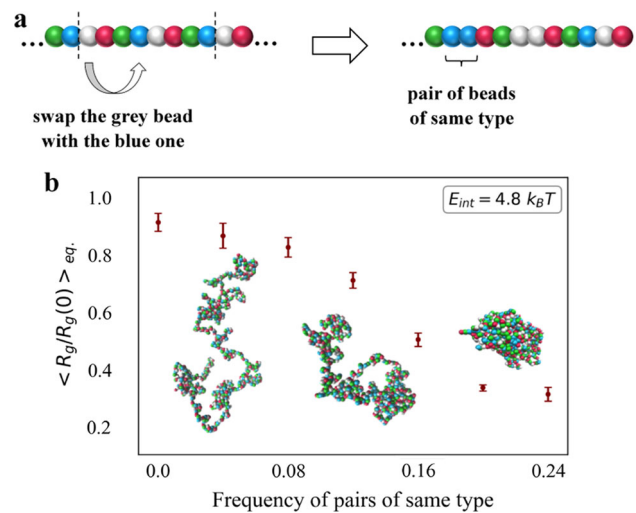
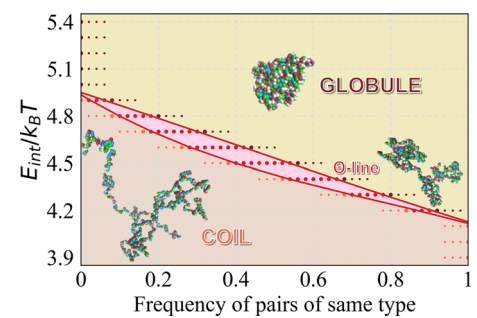


Fig. 5 Phase diagram of the polymer equilibrium state for different binding affinities and pair frequency. At low binding affinities or pair frequencies, the polymer is in the open conformation (thermodynamic coil phase). Conversely, for values of these parameters above the threshold, the hetero-polymer is able to reach the globular state. Examples of 3D snapshot for each equilibrium state is also reported



varies as a function of the number of pairs of binding sites of the same (any) colour, at a fixed binding affinity E_{int} and binder concentration c . Since in the previous section we studied alternating and random chains composed of three binding site types and an inert (grey) species, this strategy allowed to consider intermediate, highly controlled configurations. Interestingly, we find that the polymer starts to be able to transit into the globular state after a threshold number of pairs (estimated around 100, i.e. 50 bead swaps) are formed, as shown in Fig. 4b (here $E_{int} = 4.8 k_B T$). It is also worth to note that with this affinity a few pairing events (fraction of 0.1–0.2, see Methods) are required to trigger the transition (see next section), highlighting that macroscopic effects can be induced by microscopic changes of the system [31]. Overall, these results suggest that the presence of regions with consecutive beads of the same type enhances at the microscopic scale the avidity of the polymer to attract binders, which in turn can form more stable loops and drive the polymer in the globular state.

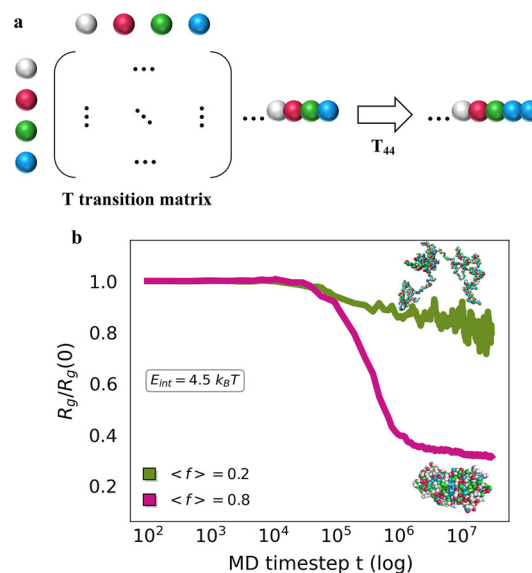
2.5 Binding sites configurations are another control parameter: phase diagram

Next, we systematically studied the equilibrium thermodynamic phases of hetero-polymer chains in several conditions, by simultaneously varying binding affinity E_{int} and the frequency of bead pairs of same type along the polymer (see Methods). In this way, we obtain the phase diagram of the system (Fig. 5). Interestingly, we identify a range of affinities E_{int} where the previously discussed configurational effect is observed. Our results show that lower binding affinities require higher pair frequency to allow the phase transition; conversely, if we increase the affinity, the globular state is achieved at lower pair frequencies (Fig. 5). In agreement with the SBS model [18] and general Polymer Physics [29, 30], we observe that for sufficiently high values of the binding affinity (here $E_{int} \geq 5.1 k_B T$) the polymer undergoes the phase transition, regardless the pair frequency. On the other hand, at low binding affinities, the loops are not stable, the transition does not occur (here $E_{int} \leq 4.1 k_B T$), and therefore, the hetero-polymer is an open SAW conformation, for any pair frequency. In addition, the thermodynamic phase of the alternating chain (i.e. pair frequency 0) changes from coil to globule in the binding affinity range 4.8–5.1 $k_B T$ (left part of Fig. 5). As shown in the previous sections, the number of types of binding sites is a control parameter of the phase transition. At fixed binder concentration, when the complexity of the chain increases, a greater binding affinity is required for the globular state to be achieved.

It is important to stress that we considered rearrangements only involving pairs of consecutive beads of the same type. Of course, other rearrangement schemes (involving, e.g. consecutive triplets, quadruplets and so on) can be implemented and likely expand the phase diagram with more articulated parameters settings. Overall, our results suggest that the spatial arrangement of binding site types along the polymer is an important feature to control the coil-globule phase transition.

Fig. 6 Markov chain approach to model stochastic hetero-polymers.

a To generate stochastic hetero-polymers, a transition matrix T is defined. For example, a blue bead is observed after another blue bead with a probability T_{44} . By setting the entries T_{ij} , a specific average pair frequency $\langle f \rangle$ can be obtained (see Methods). **b** Time dynamics of normalized gyration radius in stochastic hetero-polymers with average pair frequency $\langle f \rangle = 0.2$ (green) and $\langle f \rangle = 0.8$ (purple), interaction affinity $E_{\text{int}} = 4.5 \text{ k}_B T$. Consistently with the phase diagram, in the polymer with $\langle f \rangle = 0.2$ coil-globule transition is not observed, whereas it occurs in polymers with $\langle f \rangle = 0.8$



2.6 Markov approach of stochastic hetero-polymer chains

The above considered hetero-polymer chains have been generated deterministically. To verify the robustness of the results with an alternative approach, we simulated stochastic hetero-polymer chains, in which the spatial distribution of bead types is described by a Markovian process (see Methods). In brief, the type of each bead along the string is given according to a specific probability, described by the transition matrix associated with the hetero-polymer chain and depends only on the previous bead type (Fig. 6a). By setting the entries of the transition matrix, a stochastic polymer with a specific average pair frequency $\langle f \rangle$ can be generated (see Methods). Taking advantage of the phase diagram in Fig. 5, we first applied this approach using $\langle f \rangle = 0.2$ (few pairs) and $E_{\text{int}} = 4.5 \text{ k}_B T$. In this case, the loops are not stable, and the transition does not occur. Conversely, in stochastic polymers with $\langle f \rangle = 0.8$ (abundant pairs) and $E_{\text{int}} = 4.5 \text{ k}_B T$, the phase transition occurs, and a thermodynamic globular phase is observed (Fig. 6b). Those results are fully consistent with the phase diagram (Fig. 5) and, more generally, support the scenario where the pair frequency is an additional control parameter of the coil-globule phase transition.

3 Discussion

In this work, we investigated the coil-globule phase transition of polymer models typically used to describe chromatin architecture [32]. In particular, we focused on the relationship between the complexity of a hetero-polymer chain and its ability to transit in a compact, globular structure. We found that by increasing the number of molecular types composing an alternating hetero-polymer, the system can escape from its stable, globular thermodynamic phase. In addition, we found that the phase transition of a hetero-polymer depends on the arrangement of the binding sites. If the affinity and binder concentration is high enough, random hetero-polymers undergo the coil-globule phase transition, whereas more complex, regular alternating hetero-polymers do not, being the other parameters unchanged. We have shown that the key feature controlling this behaviour is the presence of pairs of same type along the sequence of binding sites, which likely form more stable contacts and, if such pairs are enough frequent, induce the coil-globule phase transition. Using different approaches, we explored several polymer configurations, with different pair frequencies and binding affinities. Therefore, we were able to identify an entire range of values where such effect is observed and obtained the phase diagram of the system, suggesting a scenario in which the above discussed configurational complexity is another feature to control the thermodynamic phase of these polymer systems. Importantly, the phase behaviours described in our model are typical of many other complex systems of physics [33–37].

It is worth to note that our model is based only on an equilibrium process, i.e. phase separation [23] driven by the interaction of the polymer with the binders. In real cell nuclei, other mechanisms, based on non-equilibrium processes, are known to play important roles in organizing chromatin structure and gene regulation, as e.g. loop extrusion [14, 15] or general active processes driven by ATP consumption [13]. Therefore, it will be interesting to investigate the role of the above-discussed configurational complexity in the coil-globule phase transition with the presence of these active mechanisms. Biologically, the different types of binding sites can be linked to the different proteins that typically populate the nuclear environment and that bind to the chromatin filament [38] to drive chromatin architecture and genome function. The presence of multiple, different molecular factors that bind to certain genomic

regions along with the location of the regulatory elements can be therefore explained as a strategy adopted by the cell to allow a more articulated and precise control of gene activity.

4 Methods

4.1 The SBS model

According to the SBS model [18], the polymer chain is modelled as a sequence of N beads and the interactions between distant sites are mediated by binders, modelled as simple, spherical particles. Beads that interact with the binders are named binding sites. At the beginning of the simulation, the polymer is initialized as a SAW (self-avoiding walk) configuration and the binders are randomly placed in the simulation box. Then, interaction between the binding sites and cognate binders can induce the formation of loops. The stability of these loops depends on the control parameters of the model, i.e. the binder concentration (c) and the binding affinity (E_{int}). If the concentration or energy is above threshold, the loops become stable, and the polymer transits from an open conformation to a globular, compact phase. The thermodynamic phase intermediate to the coil and globule states is known as Θ -point, and the corresponding folding state is described by the random walk (RW) model. Binding sites can be specific, i.e. they can interact only with associated cognate binders. Schematically, different types can be identified with different colours. In general, each string is composed of several types of binding sites. For the sake of simplicity, in this work, we consider the same binding affinity for each colour. We consider also an inert, non-interacting type (coloured in grey) of bead, which therefore has no cognate binders.

4.2 Details of the Molecular Dynamics (MD) simulations

Beads and binders are subject to Brownian motion, so each particle obeys a Langevin equation [39], solved numerically with a standard velocity Verlet's algorithm using the freely available LAMMPS package [40]. Interactions are modelled as previously described [23]. Briefly, all particles experience a repulsive interaction modelled as a truncated, repulsive Lennard–Jones (LJ) potential [41], with diameter $\sigma = 1$, mass $m = 1$, energy scale $\varepsilon = 1 \text{ k}_B\text{T}$ (k_B is the Boltzmann constant, and T is the temperature of the system) and distance cut-off $r_{\text{cut}} = 1.12\sigma$ (we adopted dimensionless units and used the notation given in [41]). Interactions between bead and binders are modelled as a truncated, attractive LJ potential [23] with distance cut-off $r_{\text{cut}} = 1.3\sigma$ and ε depending on the specific parameter setting. Consecutive beads along the polymer are connected by standard FENE springs [41], with distance cut-off $R_{\text{int}} = 1.6\sigma$ and constant $K = 30 \text{ k}_B\text{T}/\sigma^2$.

In this work, all hetero-polymer chains are composed of $N = 1000$ beads and the system is confined to a cubic box with periodic boundary conditions and linear size $D = 70$. We explored an affinity (given by the minimum of the LJ attractive potential [23]) range in the weak biochemical energy scale, in the approximate range $E_{\text{int}} = 3 \div 5 \text{ k}_B\text{T}$. For the sake of simplicity, we used the same binder concentrations for all the types and considered 500 binders for each colour. This value is the same for all the simulations performed. We let the system evolve until stationarity is reached, as shown by the gyration radius as a function of MD time. We consider an integration timestep $\Delta t = 0.012$ ([39, 40]) and used $3 \cdot 10^7$ timesteps for each MD simulation. To check the robustness of the results, we performed also longer MD simulations (up to $3 \cdot 10^8$) of regularly alternating hetero-polymers with 3 binding site types at $E_{\text{int}} = 4.8 \text{ k}_B\text{T}$ (see Fig. 2a) and found analogous results. For each parameter setting, we simulated up to 30 independent runs and the physical quantities of interest ($R_g/R_g(\text{SAW})$, $P_c(s)$ and $R^2(s)$, see next sections) were calculated as ensemble averages.

4.3 Types of hetero-polymer chains

4.3.1 Alternating and random hetero-polymers

Alternating hetero-polymer chains refer to those hetero-polymers in which a sub-sequence of monomers of any type repeats along the string. If the subsequence of a polymer of length N contains monomers of m different types, we design an alternating hetero-polymer in the following way. We take m beads of different types and insert them into the first m sites of the string. From the $(m+1)$ -th position we insert other m beads in the same order as the previous sub-sequence and iterate this mechanism until position N is occupied. Random hetero-polymer chains are generated from alternating ones by randomly permuting the positions of each bead along the string.

4.3.2 Beads swap and frequency of pairs of same type beads along the polymer

To consider intermediate configurations between the alternate and the randomly arranged polymer previously defined, we designed a procedure that transforms alternating hetero-polymer chains into polymers where the frequency of pairs of similar beads is fixed. The procedure was applied on the polymer having the inert type (grey) and three types of binding sites (red-green-blue in the figures). Specifically, the algorithm takes as input a hetero-polymer chain, where the main sub-sequence is the quadruple containing the colours grey, red, green and blue, grey being the non-interacting particle. To make consecutive pairs of beads of the same type

along the polymer, we first select the positions of two differently coloured beads and then swap their position. We start with the grey and blue beads, as schematically shown in Fig. 4a. Because of such a substitution, there are now two pairs of same type beads, a blue one and a grey one. We define n_{14} the number of such substitutions. The maximum number of substitutions is 125, since the polymer contains 250 beads of each of the four types. To make polymers with symmetric arrangements of pairs, for a fixed number of substitutions $n_{14} \in \{1, \dots, 125\}$, we divide the length of the polymer into n_{14} blocks, consider the first quadruple of each block, and swap the grey bead with the blue one within each quadruple. Once made all possible pairs of grey and blue beads, we start again the procedure and swap the beads with unchanged spatial distribution, i.e. green and red types. The number of substitutions is defined n_{23} and again $n_{23} = 125$ are possible. We define the general parameter $n = n_{14} + n_{23}$ as the number of substitutions of the beads, regardless of their type. By definition $n \in \{0, \dots, 250\}$, where $n = 0$ is the alternating polymer with four types of beads. By normalizing this number in the range $[0, 1]$, we obtain the pair frequency which is used as control variable in Figs. 4 and 5.

4.4 Markov transition matrices for stochastic hetero-polymer chains

To generate hetero-polymers with a fixed number of pairs of same type we employed Markov chains described by a transition matrix \mathbf{T} (Fig. 6a, left), whose entries T_{ij} with $i, j \in [1, m]$, m number of types, represent the probability that a bead of type i (i.e. colour i) is followed by a bead of type j (i.e. colour j). We have, for $m = 4$ (i.e. 4 colours, schematically represented by grey, red, green and blue in Fig. 3a):

- random hetero-polymers. The transition probabilities do not depend on the type of beads, i.e. $T_{ij} = 0.25, \forall i, j$;
- regular alternating hetero-polymers. The subsequence of beads is grey-red-green-blue, that is 1-2-3-4 and is regularly repeated along the polymer. The transition probability is $T_{12} = T_{23} = T_{34} = T_{41} = 1, 0$ otherwise;
- alternating hetero-polymers with an average fraction $\langle f \rangle$ of pairs. Without loss of generality, we first consider only grey or blue pairs. In this case, we have the transition matrix is:

$$\mathbf{T} = \begin{pmatrix} \langle f \rangle & 1 - \langle f \rangle & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \langle f \rangle & 0 & 0 & 1 - \langle f \rangle \\ 1 - 2\langle f \rangle & \langle f \rangle & 0 & \langle f \rangle \end{pmatrix}$$

where $\langle f \rangle \in [0-0.5]$. When all swaps of grey species with blue ones have been made, red and green sites can be swapped and the transition matrix becomes:

$$\mathbf{T} = \begin{pmatrix} 0.5 & 0.5 - \langle f \rangle & \langle f \rangle & 0 \\ \langle f \rangle & \langle f \rangle & 1 - 2\langle f \rangle & 0 \\ 0.5 - \langle f \rangle & 0 & \langle f \rangle & 0.5 \\ 0 & 0.5 & 0 & 0.5 \end{pmatrix}$$

where again $\langle f \rangle \in [0-0.5]$ and the resulting frequency is $0.5 + \langle f \rangle$.

4.5 Quantities of interest in polymer physics

4.5.1 $R_g/R_g(\text{SAW})$ and 3D graphic visualization

The gyration radius R_g measures the size of the sphere containing the polymer beads and has been calculated according to its definition: $R_g^2 = \frac{1}{2N^2} \sum_{i,j}^N |\mathbf{R}_i - \mathbf{R}_j|^2$, where \mathbf{R}_i is the position of i -th bead. To monitor the phase transition, we used the dimensionless quantity $R_g/R_g(\text{SAW})$, i.e. the gyration radius normalized with respect to its initial value in the SAW state. 3D structures were generated using the software Visual Molecular Dynamics [42].

4.5.2 Mean contact probability $P_c(s)$ and mean square distances $R^2(s)$

The contact probability $P_c(s)$ and quadratic distance $R^2(s)$ were calculated as follows. We define the genomic distance s between beads of sites i and j of the polymer as the number of beads separating them, i.e. $s = |i - j|$. For each s , the mean square distance $R^2(s)$ is computed as the arithmetic average of all possible bead pairs having the same genomic distance s . We then define a distance threshold $\lambda\sigma$ (we set $\lambda = 3.5$). If two beads have a distance lower than the threshold they are considered in spatial contact. The mean contact probability is the average frequency of finding two beads at distance s in contact.

Acknowledgements AMC acknowledges financial support from “Programma per il Finanziamento della Ricerca di Ateneo Linea B” (FRA) 2020, University of Naples Federico II. A.M.C. acknowledges support from the CINECA ISCRA grant IDs HP10CCZ4KN and HP10C8JWU7. The authors acknowledge computer resources from INFN, ENEA CRESCO/ENEAGRID [66] and Scope/ReCAS/Ibisco at the University of Naples.

Author contributions AMC conceived and supervised the project. FT performed and analysed numerical simulations. AMC and FT wrote the manuscript.

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

Data Availability Statement This manuscript has associated data in a data repository. [Authors' comment: Data will be made available upon reasonable request.]

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- R. Kempfer, A. Pombo, Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020)
- C.L. Woodcock, R.P. Ghosh, Chromatin higher-order structure and dynamics. *Cold Spring Harb Perspect Biol* **2**, a000596 (2010)
- W.A. Bickmore, B. Van Steensel, Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270–1284 (2013)
- T. Misteli, Beyond the sequence: cellular organization of genome function. *Cell* **128**, 787800 (2007)
- J.R. Dixon, D.U. Gorkin, B. Ren, Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016)
- J.R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012)
- 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., Spatial partitioning of the regulatory landscape of the X inactivation centre. *Nature* **485**, 381–385 (2012)
- 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 et al., Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009)
- J. Fraser, C. Ferrai, A.M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B.L. Moore, D.C. Kraemer, S. Aitken et al., Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* **11**, 852 (2015)
- A.Y. Grosberg, S.K. Nechaev, E.I. Shakhnovich, The role of topological constraints in the kinetics of collapse of macromolecules. *J. Phys.* **49**, 2095–2100 (1988)
- A. Grosberg, Y. Rabin, S. Havlin, A. Neer, Crumpled globule model of the three-dimensional structure of DNA EPL. *Europhys. Lett.* **23**, 373 (1993)
- A. Agrawal, N. Ganai, S. Sengupta, G.I. Menon, Nonequilibrium biophysical processes influence the large-scale architecture of the cell nucleus. *Biophys. J.* **118**(9), 2229–2244 (2020)
- N. Ganai, S. Sengupta, G.I. Menon, Chromosome positioning from activity-based segregation. *Nucleic Acids Res.* **42**(7), 4145–4159 (2014)
- G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, L.A. Mirny, Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038 (2016)
- A.L. Sanborn, S.S.P. Rao, S.-C. Huang, N.C. Durand et al., Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci.* **112**, E6456 (2015)
- D. Jost, P. Carrivain, G. Cavalli, C. Vaillant, Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acid Res.* **42**(15), 9553–9561 (2014)
- C.A. Brackley, S. Taylor, A. Papantonis, R. Peter, D. Marenduzzo, Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization. *Proc. Natl. Acad. Sci. USA* **110**(38), E3605–E3611 (2013)
- M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, M. Nicodemi, Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. USA* **109**, 16173–16178 (2012)
- A.M. Chiariello, S. Bianco, A.M. Oudelaar, A. Esposito, C. Annunziatella, L. Fiorillo, M. Conte, A. Corrado, A. Prisco, M.S.C. Larke, J.M. Telenius, R. Sciarretta, F. Musella, V.J. Buckle, D.R. Higgs, J.R. Hughes, M. Nicodemi, A dynamic folded hairpin conformation is associated with α -globin activation in erythroid cells. *Cell Rep.* **30**, 2125–2135 (2020)
- F. Rossi, R. Buschow, L.V. Glaser, T. Schubert, H. Staeger, A. Grimme, H. Will, T. Milke, M. Vingron, A.M. Chiariello, S. Kinkley, Connecting the dots: PHF13 and cohesin promote polymer-polymer phase separation of chromatin into chromosomes. *BioRxiv* (2022). <https://doi.org/10.1101/2022.03.04.482956>
- B. Bintu, L.J. Mateo, J. Su, N.A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A.N. Boettiger, X. Zhuang, Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* (2018). <https://doi.org/10.1126/science.aau1783>
- M. Conte, L. Fiorillo, S. Bianco, A.M. Chiariello, A. Esposito, M. Nicodemi, Polymer physics indicates chromatin folding variability across single-cells results from state degeneracy in phase separation. *Nat. Commun.* **11**, 3289 (2020)
- A.M. Chiariello, C. Annunziatella, S. Bianco, A. Esposito, M. Nicodemi, Polymer physics of chromosome large-scale 3D organisation. *Sci. Rep.* **6**, 29775 (2016)
- M. Barbieri, S.Q. Xie, E.T. Triglia, D.S. Inês, M.R. Branco, D. Rueda, M. Nicodemi, A. Pombo, Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **24**, 515–524 (2017)
- Q. Li, H. Tjong, X. Li, K. Gong, X.J. Zhou, I. Chiolo, F. Alber, The three-dimensional genome organization of *Drosophila melanogaster* through data integration. *Genome Biol.* **18**, 145 (2017)
- F. Serra, D. Bau, M. Goodstadt, D. Castillo, G.J. Filion, M.A. Marti-Renom, Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors *PLOS Comput. Biol.* **13**, e1005665 (2017)
- G. Nir, I. Farabella, C. Perez Estrada, C.G. Ebeling, B.J. Beliveau, H.M. Sasaki, S.H. Lee, S.C. Nguyen, R.B. McCole, S. Chatteraj et al., Walking along chromosomes with super-resolution imaging, contact maps, and integrative modelling. *PLoS Genet.* **14**, e1007872 (2018)
- S. Bianco, D.G. Lupiáñez, A.M. Chiariello, C. Annunziatella, K. Kraft, R. Schopflin, L. Wittler, G. Andrey, M. Vingron, A. Pombo et al., Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat. Genet.* **50**, 662–667 (2018)
- P. De Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, 1970)
- Doi, M., Edwards, S., F. (1988). *The Theory of Polymer Dynamics*. *Oxford science publications*

31. A.M. Chiariello, F. Corberi, M. Salerno, The interplay between phase separation and gene-enhancer communication: a theoretical study. *Biophys. J.* **119**(4), 873–883 (2020)
32. A. Esposito, A.M. Chiariello, M. Conte, L. Fiorillo, F. Musella, R. Sciarretta, S. Bianco, Higher-order chromosome structures investigated by polymer physics in cellular morphogenesis and differentiation. *J. Mol. Biol.* **432**(3), 701–711 (2020)
33. M. Nicodemi, A. Coniglio, H.J. Herrmann, Compaction and force propagation in granular packings. *Physica A* **240**, 405 (1997)
34. M. Nicodemi, H.J. Jensen, Creep of superconducting vortices in the limit of vanishing temperature: a fingerprint of off-equilibrium dynamics. *Phys. Rev. Lett.* **86**, 4378 (2001)
35. A. Coniglio, M. Nicodemi, Jamming transition of granular media. *J. Phys. Condens. Matter* **12**, 6601 (2000)
36. M. Nicodemi, Force correlations and arches formation in granular assemblies. *Phys. Rev. Lett.* **80**, 1340 (1998)
37. D.S. Grebenkov, M. Pica Ciamarra, M. Nicodemi, A. Coniglio, Flow, ordering, and jamming of sheared granular suspensions. *Phys. Rev. Lett.* **100**, 078001 (2008)
38. L. Fiorillo, S. Bianco, A. Esposito, M. Conte, R. Sciarretta, F. Musella, A.M. Chiariello, A modern challenge of polymer physics: novel ways to study, interpret, and reconstruct chromatin structure. *WIREs* **10**(4), e1414 (2020)
39. M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, 1987)
40. S. Plimpton, Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995)
41. K. Kremer, G.S. Grest, Dynamics of entangled linear polymer melts: a molecular-dynamics simulation. *J. Chem. Phys.* **92**, 5057–5150 (1990)
42. Visual Molecular Dynamics. VMD documentation: <https://www.ks.uiuc.edu/Research/vmd/current/ug/>