Supporting Information Kinetic mechanisms of crumpled globule formation

January 9, 2020

1 Calculation of P(s) and R(s)

To calculate P(s) we counted N_i - the number of beads, which have less distance to the *i*-th bead, than cutoff radius = 1.5, and then averaged N_i over all measurements in a chain:

$$P(s) = \frac{1}{N-s} \sum_{i=1}^{N-s} N_i(t_0)$$
 (Eq. S1)

In R(s) we calculated average spatial distance between beads, separated by s beads along the chain.

$$R(s) = \frac{1}{N-s} \sum_{i=1}^{N-s} ((x_i(t_0) - x_{i+s}(t_0))^2 + (y_i(t_0) - y_{i+s}(t_0))^2 + (z_i(t_0) - z_{i+s}(t_0))^2)^{1/2}$$
(Eq. S2)

2 Procedure of MSD(t) **calculation**

We calculated MSD(t) according to the following procedure. Using the annealed structure as the initial structure, we continued the simulations, and structures were generated every Δt DPD time steps. However, we chose two different time steps: $\Delta t_1 = 10$ DPD time steps and $\Delta t_2 = 2 \times 10^4$ DPD time steps. Simulations, where structures were generated every Δt_1 DPD time steps, were performed during $T_1 = 2 \times 10^4$ DPD time steps, and, where structures were generated every Δt_2 DPD time steps, were performed during $T_2 = 4 \times 10^7$ DPD time steps. This was done for computational efficiency: we needed to measure MSD(t) in a very broad time range $t \in [10, 4 \times 10^7]$, and if we give out structures with one time step $\Delta t = 10$, it would not be possible to measure MSD(t) in a reasonable time due to extraordinarily large number of structures to analyze. Therefore, we "split" the time range into two parts, and gave out structures with two different time steps. We ensured the correctness of such approach by calculating $MSD(t = 2 \times 10^4)$ and verifying, that both dependencies give the same value and the standard deviation of mean squared displacement.

The program analyzed n-th pair of structures, separated by $t = n\Delta t_k$, (k = 1, 2) DPD time steps from each other, counting $MSD_n(t)$: the average squared displacement of beads in the given pair of structures. We also distracted displacement of the center of mass of the chain (COM) from displacements of beads. Then program analyzed next pair of structures, separated by t DPD time steps from each other, and calculated $MSD_{n+1}(t)$ etc. MSD(t) was calculated as the average over all $MSD_n(t)$ values. Standard deviation of MSD(t) was calculated as the standard deviation in the set of $MSD_n(t)$ values. Exact equation for MSD(t) calculation is given in Equation Eq. S3.

$$\begin{split} MSD(t,t < T_k - t_0) &= \frac{1}{N} \frac{\Delta t_k}{T_k - (t + t_0)} \sum_{n=0}^{(T_k - (t + t_0))/\Delta t_k} \sum_{i=1}^N ((x_i(t + t_0 + n\Delta t_k) - x_{COM}(t + t_0 + n\Delta t_k)) - (x_i(t_0 + n\Delta t_k) - x_{COM}(t_0 + n\Delta t_k))^2 + ((y_i(t + t_0 + n\Delta t_k) - y_{COM}(t + t_0 + n\Delta t_k)) - (y_i(t_0 + n\Delta t_k) - y_{COM}(t_0 + n\Delta t_k))^2 + ((z_i(t + t_0 + n\Delta t_k) - z_{COM}(t + t_0 + n\Delta t_k)) - (z_i(t_0 + n\Delta t_k) - z_{COM}(t + n\Delta t_k)) - (z_i(t_0 + n\Delta t_k) - z_{COM}(t + n\Delta t_k)) - (z_i(t_0 + n\Delta t_k) - z_{COM}(t_0 + n\Delta t_k))^2, k = 1,2 \end{split}$$
(Eq. S3)

3 Bond formation/breaking algorithm

In this section we describe in detail, how formation and breaking of pairwise reversible bonds is realized in our simulations. Every $N_{stp} = 200$ DPD time steps program chooses an i-th bead, and forms a list of beads, which are spatially closer to the chosen i-th bead, than cutoff radius $R_c = 1.0$. Then this list is sorted by spatial distance to the i-th bead (beads, which are the closest to the chosen bead, are on the top of the list). Then, starting from the top of the list, we pick a j-th bead and it can form a bond with the i-th bead with probability of bond formation, which is set constant in the simulation (if the bead on the k-th position in the list does not form a bond, the bead on the k+1-st position is checked, etc, until the end of the list). Therefore, our algorithm of bond formation favors bond formation of the spatially closest beads to the chosen one. Formed bond is identical to the backbone bonds of the chain. Then program

iterates over all i-th beads until the end of the chain. If the chosen i-th bead already forms a bond, it can break the bond with probability of breaking a bond β , which is set constant during the simulation as well.

To study non-equilibrium properties of the chain, we needed to forbid self-intersections of the chain. To do so, we set large bond stiffness constant K = 150.0, repulsion parameters, and small equilibrium bond length, which was equal to 0.6 in all our simulations (see Fig. S1). By choosing these parameters we guaranteed the absence of self-intersections¹, but broke detailed balance (Fig. S1). Therefore, there is unequal probability distributions of finding two beads at a distance r_{ij} at the moment of bond formation/breaking (Fig. S1). The broken detailed balance leads to the difference between absolute values of gained and dissipated energy after every bond formation and breaking event. To judge whether this effect can affect the system, first we determined the average value of potential energy change at the moment of bond formation and breaking. This can be done analytically in our system: the potential energy for bonds in our system is written as $U_{ij}^b = \frac{1}{2}K(r_{ij} - r_0)^2$, $r_0 = 0.5$ DPD units in all our simulations. The average distance between beads at the moment of bond formation is approximately 0.84 DPD units (see Fig. S1). Hence, the average potential energy change at the moment of bond formation +8.67 energy units, in equilibrium $k_BT = 1$ energy unit by default. Due to repulsion forces, equilibrium bond length is greater than r_0 and equals to 0.6. Therefore, the loss of potential energy at the moment of bond breaking is -0.75 energy units.

These estimates provide us the opportunity to evaluate how broken detailed balance affects our system. The possible effect could be incorrect local non-equilibrium dynamical behavior of beads after bond formation and breaking events. Dissipative and random forces acting between beads act as a thermostat in DPD². Hence, if simulation parameters are chosen correctly, local gain and dissipation of energy after bond formation/breaking can be compensated by dissipative and random forces from surrounding beads. To ensure this compensation, we have chosen integration time step $\delta t = 0.02$, noise level $\sigma = 3$ and DPD density $\rho = 3$. If these parameters are chosen, the relaxation time of exponential decrease of temperature from $k_B T = 10$ energy units to $k_B T = 1$ is around 10 DPD time steps (Ref.², Section iii). Therefore, characteristic time for non-equilibrium state existence with $k_B T \approx 8.67$ energy units is approximately 10 DPD time steps. Moreover, due to repulsion forces acting on the beads after bond formation, not all excessive potential energy 8.67 energy units will heat up the system locally, therefore, the non-equilibrium state will have even less temperature after bond formation. Hence, beads, which formed a bond, become equilibrated with environment after maximal time of approximately 10 DPD time steps. We studied dynamical behavior (*MSD*(*t*)) on the time scales larger than 10 DPD time steps, and bond lifetime values used in the work were 5×10^3 , 5×10^4 and 5×10^5 DPD time steps. Hence, all quantities measured in our work are not affected by the energy gain and dissipation due to broken detailed balance.

4 Procedure of globule equilibration in SPS

In our work, we studied conformational properties of equilibrium globule as a reference state. As we mentioned in the paper, we obtained stable transient state, similar to crumpled globule, during both types of coil-globule transition: in a poor solvent and induced by pairwise reversible bonds. To obtain equilibrium globule, we used crumpled globule state, formed in a weak poor solvent ($\chi \approx 1.5$, SPS), after $t = 8 \times 10^7$ DPD time steps, as initial structure. The crumpled globule state is stabilized by topological interactions, therefore, we needed to facilitate self-intersections of the chain to obtain Gaussian globule. To do so, we set monomer-monomer and solvent-solvent repulsion parameters a_{ii} (i = 1,2) equal to 25.0, and monomer-solvent repulsion was equal to $a_{12} = a_{21} = 29.5$. Since χ value is determined from difference between a_{ij} and a_{ii} (if $a_{ii} = a_{jj}$ and, $a_{ij} = a_{ji}$)², we conserved the solvent quality, but significantly reduced overall repulsion between polymer beads, therefore, facilitating self-intersections of the chain. We also reduced bond stiffness parameter from K = 150.0 to K = 4.0, and increased integration time step to $\delta t = 0.04$. Therefore, as previously described in Ref.³, we allowed chain to self-intersect, and conserved temperature $k_BT = 1$ in the NVT-ensemble. We performed simulation with these parameters for $t = 5 \times 10^7$ DPD time steps. Then we changed simulation parameters to the initial values, described in Methods for SPS, and performed simulations for $t = 2 \times 10^7$ DPD time steps.

Notes and references

- [1] P. Nikunen, I. Vattulainen and M. Karttunen, Physical Review E, 2007, 75, 036713.
- [2] R. D. Groot and P. B. Warren, The Journal of chemical physics, 1997, 107, 4423-4435.
- [3] M. V. Tamm, L. I. Nazarov, A. A. Gavrilov and A. V. Chertovich, *Physical review letters*, 2015, 114, 178102.

5 Additional plots

In this section, we provide additional dependencies, supporting the information in the paper.



Figure S1 Probability distributions of distance between centers of beads in the moment of bond breaking (red) and bond formation (blue). Samples consist of 96514 distances for each histogram, data is normalized.





(b)

(c)

Figure S2 α_L and α_G values (exponents of R(s) dependencies on the "local" and "global" scale shown in black and red, respectively) depending on the time passed since simulation started. Figure (a) shows dependencies for the System in a Poor Solvent (SPS) in a weak poor solvent ($\chi \approx 1.5$), and for the System with Pairwise Bonds (SPB) with fraction of bonds $f \approx 0.9$, bond lifetime is $\tau = 5 \times 10^3$ steps. Figure (b) shows dependencies for System with Pairwise Bonds (SPB) with maximal fraction of bonds $f \approx 1$, bond lifetime is $\tau = 5 \times 10^5$ steps, and Figure (c) for SPB with fraction of bonds $f \approx 0.55$. It should be taken into account, however, that fraction of bonds $f \approx 0.55$ is reached only after $t = 2 \times 10^6$ DPD time steps (Figure c), on the shorter time scale fraction of bonds was constantly increasing (Figure S4b).



Figure S3 Dependency of the radius of gyration of the chain ($N = 2 \times 10^4$) on the fraction of bonds with lifetime equal to $\tau = 5 \times 10^4$. Coil-globule transition occurs at $f \approx 0.3$. Compact conformations are obtained if $f \approx 0.5$ and larger.



Figure S4 Time dependencies of fraction of bonds in the SPB, bond lifetime is equal to $\tau = 5 \times 10^4$ steps. Bond formation started after $N_{stp} = 200$ DPD time steps passed in every simulation. a) Probability of reversible bond formation is 1.0. As a result, fraction of bonds quickly stabilizes at value $f \approx 1$. b) Probability of reversible bond formation is 0.001, therefore, fraction of bonds reaches the stable value $f \approx 0.55$ relatively slowly.



Figure S5 The figure shows the experimental contact probability dependencies for different species and cell lines. Blue dashed line corresponds to the equilibrium Gaussian globule and red dashed line corresponds to the fractal globule. Well-known dependence of the contact probability on genomic distance $P(s) \propto s^{-1}$ for Homo sapiens actually is not strictly enforced. For example, contact probability depends on considering scales and it works for another biological species as well. Moreover, scalings for different cell lines of the same species could significantly differ from each other on the same scale. Apparently minor protocol changes in Hi-C experiment, different equipment and different data processing have a significant impact on the final results. To plot this figure we used publicly available Hi-C datasets re-analyzed with distiller software with standard parameters (https://github.com/mirnylab/distiller-nf). All replicates, if available, were merged together. The data for Homo sapiens was retrieved for: K562 cell line (GEO ID GSE63525, https://www.ncbi.nlm.nih.gov/pubmed/25497547), A549 GSE105600, https://www.ncbi.nlm.nih.gov/pubmed/22955616; Drosophila melanogaster cell lines Kc167, Dm3, OSC, S2: GSE69013 https://www.ncbi.nlm.nih.gov/pubmed/22955616; Drosophila https://www.ncbi.nlm.nih.gov/pubmed/29972771; Mus musculus data from GSE96611 https://www.ncbi.nlm.nih.gov/pubmed/29335546 and from GSE96107 https://www.ncbi.nlm.nih.gov/pubmed/29335



Figure S6 "Local" (α_L) and "global" (α_G) *R*(*s*) scaling exponents for initial structures depending on the time passed since equilibration in athermal solvent started, chain length $N = 2 \times 10^4$ beads, polymer concentration $n \approx 6\%$. Magenta and green dots represent equilibrium values of the exponents for the swollen and Gaussian chain, respectively.



Figure S7 "Local" (α_L) and "global" (α_G) R(s) scaling exponents for chain length $N = 2 \times 10^4$ beads, polymer concentration $n \approx 1\%$, $f \approx 1$, bond lifetime $\tau = 5 \times 10^4$ DPD steps, depending on the time passed since simulation started.





Figure S8 Particle density profile in the simulation box across x-coordinate (Fig. a), y-coordinate (Fig. b) and z-coordinate (Fig. c).