Supplemental Information

Scaling arguments for $P_c(s)$ in fractal and equilibrium globules, and random walks

Recall that the contact probability $P_c(s)$ is the probability that two monomers separated by the distance s along the chain come close enough in the space to make a contact. The probability of forming a contact in a globular state with a constant density can be roughly estimated as an inverse of a volume $V(s)$ in which two ends of a subchain of legnth s reside: $P_c(s) \sim 1/V(s) \sim 1/R_G^D(s)$. Here, $R_G(s)$ is the gyration radius of the chain fragment of s monomers. For a curve with dense space-filling subchains, the volume of a subchain is proportional to the number of monomers $V(s) \sim s$, yielding $R_G(s) \sim s^{1/3}$ and $P_c(s) \sim s^{-1}$. This decay has been discussed in the literature [15], so the above naive derivation of the s^{-1}–law pursues mainly illustrative aims.

For a random walk, subchains are gaussian, and therefore have average size $R_G(s) \sim s^{1/2}$. Therefore, two ends of a subchain are located within the volume $V(s) \sim R_G^3 \sim s^{3/2}$. Probability that the two ends of the subchain of length s are in contact is approximately inversely proportional to the volume in which the subchain resides $P_c(s) \sim \frac{1}{V(s)} \sim s^{-3/2}$. In the equilibrium globule, subchains of sizes $s < N^{2/3}$ behave as random walks, and thus have the contact probability $P_c(s) \sim \frac{1}{V(s)} \sim s^{-3/2}$. For subchains longer than $s = N^{2/3}$, subchains start reflecting from the boundaries of the confinement, positions of the two ends become uncorrelated in the volume of the entire chain, and thus $P_c(s)$ exhibits a plateau.

We note that $P_c(s)$ and $R_G(s)$ for the equilibrium globule are qualitatively similar with $P_c(s)$ and $R_G(s)$ for a phantom chain confined to the same box; plots for $Rg(s)$ are shown in the figure S1. However, for a phantom chain confined to the same volume, the plateau in $P_c(s)$ and $R_G(s)$ starts earlier, because the chain is repelled by the walls and thus has a higher density in the center and a lower density next to the walls, which corresponds to an effectively smaller confining volume. Repulsion of a phantom chain by confining walls is well-studied, and analytic expression for spatial density can be obtained; for details see [37].

Relation between $P_c(s)$ and $R_G(s)$

Contact probability, $P_c(s)$, it not a dimensionless quantity, and contains information both about the size of a subchain, ($R_G(s)$), and the internal structure of the subchain. To distinguish between the two effects on $P_c(s)$, we multiplied the $P_c(s)$ plot by the approximate subchain volume, $R_G(s)^3$, thus accounting for the effect of subchain size. However, we found that the results were drastically different dependent on whether we used gyration radius $R_g(s)$, or average distance between monomers separated by s ("End-to-end distance", ETE), as a measure of subchain size. Specifically, $Pc(s)$ renormalized using gyration radius is monotonically increasing with chain size, while $Pc(s)$ renormalized using end-to-end distance has a maximum for intermediate-size subchains (see Fig. S2). This discrepancy is likely explained by the lack of well-defined scaling relations in our system. These results highlight that extra care should be taken when interpreting slopes of $P_c(s)$ and $R_G(s)$ curves as indicators of scaling relations. However, we emphasize that the contact probability is relevant for studies of chromosomes, as it is related to the frequency of interactions between genomic elements, and thus can be measured directly using Hi-C.

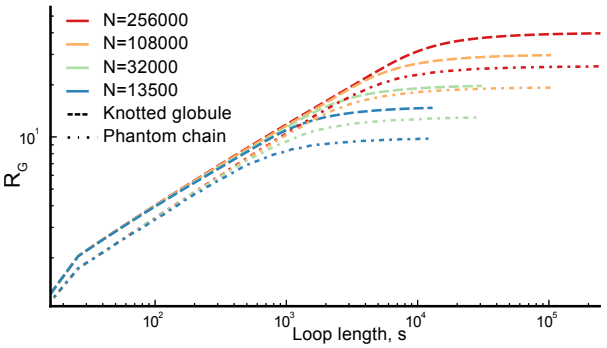

Fig. S1 Gyration radius for subchains of knotted globules and confined phantom chains; similar to Fig. 3a.

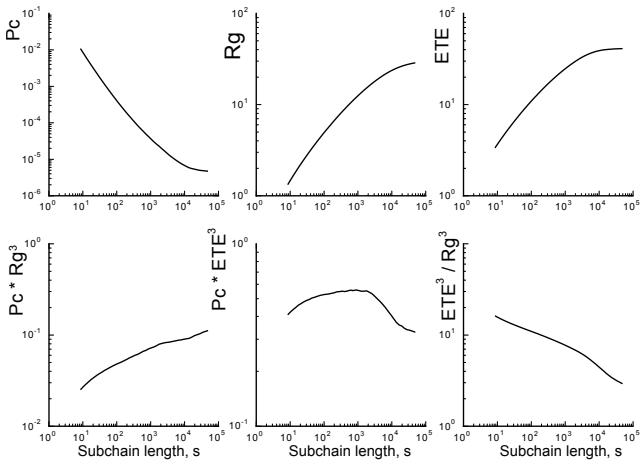

Fig. S2 (top) Average contact probability, gyration radius of subchains, and end-to-end distance for 256 000 long unknotted globules. (bottom) $P_c(s)$, renormalized by gyration radius (left), end-to-end distance(middle). Cube of the ratio of R_g to the end-to-end distance. (right)

Effects of confinement

Most of the larger loops in our system are in contact with the confining boundary. However, if we only focus on the loops which are not touching the boundary, we can still see the difference between the knotted and the unknotted globules. Fig. S4 shows the $R_G(s)$ plots for different chains, only focusing on the loops which are not in contact with the boundary. Note that $R_G(s)$ plots terminate at $s \leq N$ because most of the large loops have at least one of the monomers in contact with the boundary.

We note that gyration radius, R_G is characterizing the second moment of the spatial distribution of monomers in a subchain: $R_G = < (x-\bar{x})^2 > + < (y-\bar{y})^2 > + < (z-\bar{z})^2 > = < (r - c.o.m.)^2 >$, where c.o.m. denotes center of mass. Here we study if higher moments of the distribution to the distances to the center of mass are different in subchains of the knotted and unknotted globules. However, we can only compare subchains of the same average size R_G, as the effect of confinement on the distribution of distances to the center of mass is dependent on the R_G. For subchains in a confined volume, spatial distances exceeding the confining box size are not possible; this cuts off tails of the distribution of distances between monomers and makes the distribution non-gaussian, especially when R_G reaches the size of confining box.

To test whether knotted and unknotted globules have different distribution of distances to the center of mass, we plotted normalized moment ratios this distribution. Gyration radius is a second moment, $< (r-\bar{r})^2 >$. We plotted normalized second moment, $< r^2 > / < r >^2$ as a function of $< r >$, and normalized fourth moment as a function of the second. The latter is analogous to the kurtosis of the distribution. Since subchains of the same R_G or $< r >$ are expected to experience the same boundary effects, then any differences in the plots would be indicative of a non-trivial structure of subchains in unknotted globules, as compared to knotted. Surprisingly, we find no difference in how the normalized higher moments of the distribution depend on the mean. This is indicative of the fact that subchains of the same spatial size in knotted and unknotted globules have comparable spatial structure (Fig. S3).

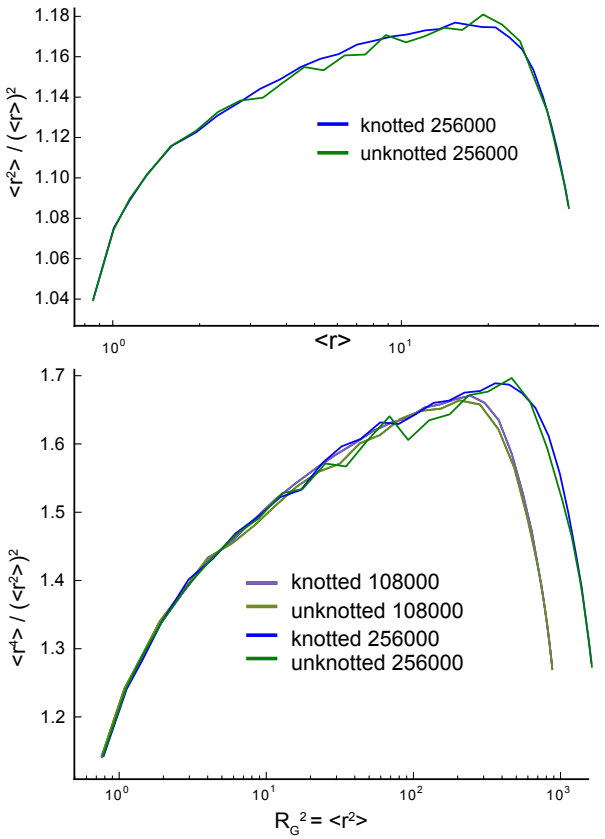

Fig. S3 (top)Dependence of the normalized second moment of spatial distribution, $< r^2 > /(< r >)^2$ on the first moment,$< r >$; r is distance to the center of mass of a subchain. (bottom) Dependence of the normalized fourth moment of spatial distribution, $< r^4 > /(< r^2 >)^2$ on the second moment,$< r^2 > = R_G^2$.

Topological properties of subchains

To obtain knotting diagrams, we evaluated $\varkappa$ for all loops in a given globule. We then divided all loop start positions and loop length in the equal number of bins (500 monomer bins for 32 000 globules, 1 000 for 108 000). For each bin of loop lengths and loop start, we evaluated average $\varkappa$ for all loops contributing to a given point, and record the average value of the knotting complexity. We then display average loop complexities on a map, using "jet" colormap, and showing $\sqrt{\varkappa}$ to make unknots easily distinguishable. White regions of the map correspond to the pairs of regions of the globule which formed no loops at all.

Figure S5 shows matrices of knot complexity for 32000-long knotted and unknotted globules. As we can see, knotting of loops in the unknotted globule is highly variable, and for any loop length it ranges from completely unknotted to highly knotted. On the contrary, for loops in the knotted globule, knot complexity depends strongly on loop length, and is not very variable for a given loop length (color forms uniform vertical stripes).

We note that in both types of globules, loops longer than $N/2$ are more knotted than shorter loops. Each contact in the unknotted globule separates the polymer into two loops: one loop that is larger than $N/2$ and the other that is smaller. If the two loops are unlinked, they would both have to be unknotted. However, if they are linked, then their knotting complexities are not necessarily equal; in fact, it often happens that the shorter loop would be unknotted and the longer loop would

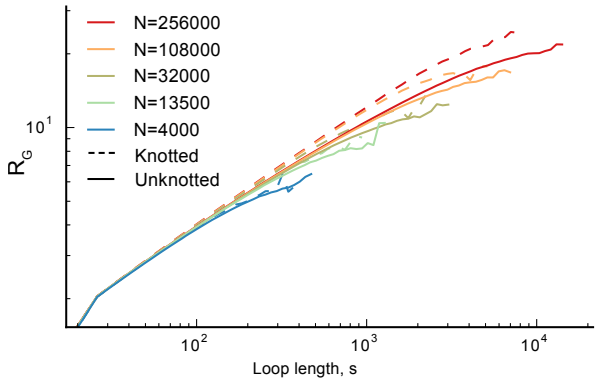

Fig. S4 The same plot as in Fig. 3a, but evaluated for loops which are not in contact with the boundary. A loop was defined to be in contact with the boundary if at least one monomer of the loop occupied the site directly adjacent to the box.

be knotted. For example, in a 108 000-long globule, 98.8% loops of length 500 to 1000 are unknotted, while 58% of the matching loops of length 107 000 to 107 500 are knotted with the average knot complexity $\varkappa = 14$.

Polymer Simulations and Analyses

Initial conformation

We note that since simulations are preformed at equilibrium, initial conformation is irrelevant to the conclusions made in the paper, and is presented for consistency and responsibility only. We initialize polymer simulations from two different starting conformations: growing polymer ring, and polymer packed in a grid. A growing polymer ring on a cubic lattice was created from a small 4-monomer ring by incrementally adding two monomers to it. At a first step, a 4-monomer ring was placed in the center of the box used for simulations. Then, one bond was chosen randomly. We then tried to extend the polymer at the chosen bond by two monomers, by making the bond into a kink. To do this, we considered another bond, obtained by shifting this bond by distance 1 in a random direction perpendicular to the bond (chosen out of 4 possible directions). If both locations of the shifted bond were free, the polymer was extended to incorporate this bond. For example, if the bond was going in $+z$ direction: $... \to (0,0,0) \to (0,0,1) \to ...$, and we attempted to grow it in the $-y$ direction (chosen randomly out of $+x,-x,+y,-y$), we would check positions $(0,-1,0)$ and $(0,-1,1)$. If both of them were free, the polymer sequence would be changed to $... \to (0,0,0) \to (0,-1,0) \to (0,-1,1) \to (0,0,1) \to ....$ If at least one of them were occupied, selection of a random bond would be repeated. The process was repeated till the polymer grew to the desired length.

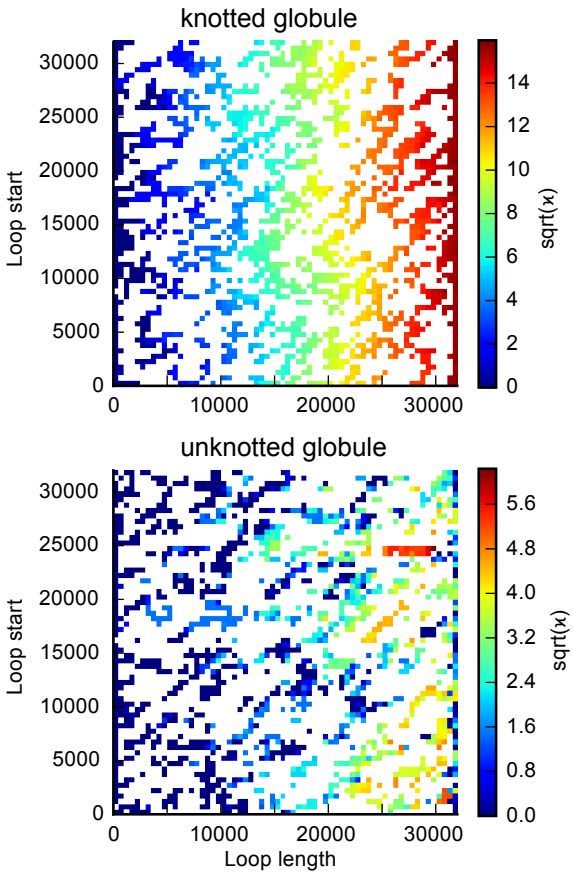

Fig. S5 Matrices of loop knot complexity for a single 32 000-long knotted globule (top), and unknotted globule (bottom). Color shows $\sqrt{\varkappa}$ to better highlight unknots. Note that the color scale is different for the two images.

Polymer densely packed in a grid was created as described below. Here $\{\}$ indicates the largest unit, $[\,]$ indicates a smaller unit, and $(\,)$ indicates a monomer

$$\begin{aligned}
&\{[(1,1,1) \to (2,1,1) \to ... \to (N-2,1,1) \to \\
&\quad (N-2,2,1) \to ... \to (1,2,1)] \to \\
&\quad [(1,3,1) \to ...(1,4,1)] \to ... \to \\
&\quad [(1,N-3,1) \to ... \to (1,N-2,1)] \to \\
&\quad [(1,N-2,2) \to ... \to (1,N-3,2)] \to ... \to \\
&\quad [(1,2,2) \to ... \to (1,1,2)]\} \to \\
&\{[(1,1,3) \to ... \to (1,1,4)]\} \to ... \to ((a,b,c))
\end{aligned}$$

after which it was crawling back to $(1,1,1)$ along the (x,y) plane, and then through the $(0,y,z)$ plane) when the distance to the end was barely enough to make the return path.

All simulations in the paper were initialized from a grow-

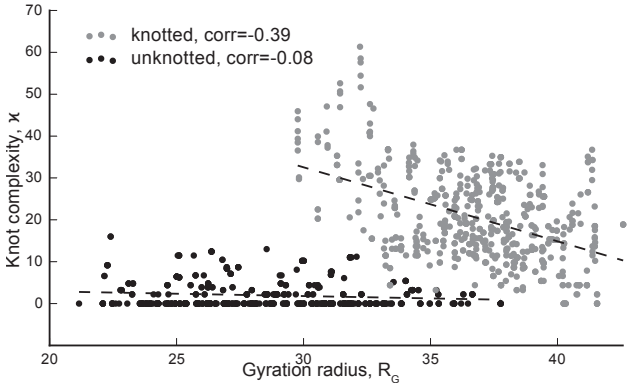

Fig. S6 A scatter plot of $R_G(s)$ and knot complexity for $s = 20\,000 - 22\,000$ loops in $N = 256\,000$ knotted and unknotted globules.

ing polymer ring; a control initialized from a polymer packed in a grid was performed for $N = 108\,000$, and showed $R_G(s)$ and $P_c(s)$ within 1% of that for a system initialized from the growing ring.

Polymer systems

We performed polymer simulations as described in [11], considering 6 system sizes: 256 000 monomers in a $80 \times 80 \times 80$ box, 108 000 monomers in a $60 \times 60 \times 60$ box, 32 000 monomers in a $40 \times 40 \times 40$ box, 13 500 monomers in a $30 \times 30 \times 30$ box, 4000 in a $20 \times 20 \times 20$ box and 2048 in a $16 \times 16 \times 16$ box. Simulation times, measured in $\log t / \log N$ are shown in Fig. S10. We performed 1000 replicas for unknotted globules up to 13 500 monomers, 400 replicas for 32 000 monomers, 30 replicas for 108 000 monomers and 20 replicas for the largest system.

We simulated topologically relaxed system by allowing co-occupation of the same lattice site by two monomers with a probability of 0.002. To test whether this procedure introduces any change in spatial properties of the polymer, we turned on full strength excluded volume interaction in $N = 108\,000$ equilibrated topologically relaxed globules, and simulated them for additional $1 \cdot 10^7$ MC steps per monomer. This process resolved all co-occupation events after $1 \cdot 10^3$ MC steps per monomer, and simulated the globule for 10^4 times longer. We then compared gyration radii R_G of the whole chain with co-occupation and without co-occupation, and found no noticeable difference in the full-chain R_G (Fig. S7 top). Note that if the chain was made completely phantom (i.e. 100% co-occupation), resulting shift of R_G would be much more dramatic, and R_G would decrease much more (Fig. S7, S1), producing visually different chains.

The knot complexity $\varkappa$ of the whole chain for topologically relaxed polymer system is increasing sharply with chain size.

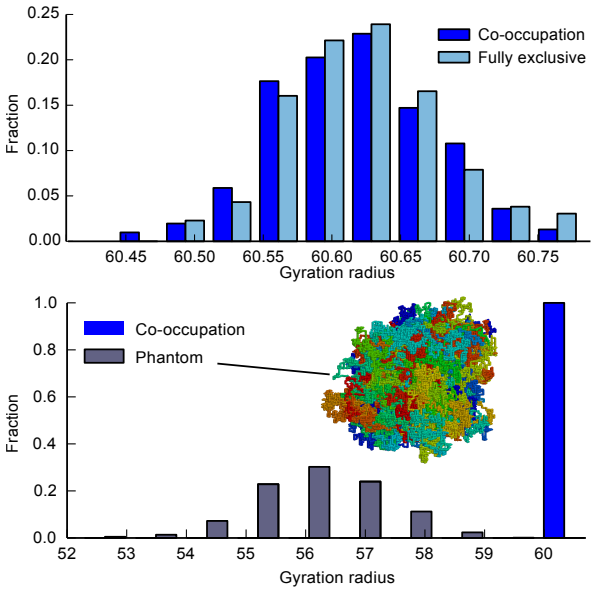

Fig. S7 (top) R_G of the whole knotted globule with partial co-occupation of lattice cites (i.e. how the knotted globule was simulated; shown in blue) is compared to the R_G of the knotted globule for which full strength excluded volume interactions were introduced (light blue). (bottom) R_G of the knotted globule with co-occupation (blue) is compared to the R_G of the phantom chain (gray). Note that the X scale is different compared to the above plot. Insert shows a sample conformation of the phantom chain. Note that it does not have a sharp boundary, unlike knotted or unknotted globules (Fig. 1)

For $N = 256\,000$, calculation of Alexanders polynomial for the whole chain was computationally unfeasible; For $N = 2048 - 108\,000$, average $\varkappa$ in a knotted globule is shown in Fig. S8.

Equilibration

To quantify equilibration of the system, we evaluate two observables: $P_c(s)$ and $R_G(s)$ for subchains of length $s = N^\alpha$, where α was equal $0.2, 0.35, 0.5, 0.65, 0.8$. We performed simulations of small systems ($N \leq 32\,000$) for long time (Fig. S10) with $400 - 1000$ replicas for each system size. We then quantitatively defined equilibration as a point where all 10 $R_G(t)$ and $P_c(t)$ curves defined above fall within 1% deviation from the final $P_c(s)$ or $R_G(s)$ (Fig. S9). The final value for the plot was defined as an average over the last half of the trajectory. We note that equilibration time have always occurred within the first 10% of the trajectory (20% for $N = 108\,000$), allowing comparison to the last half.

We then plot $log(t_{eq})$ as a function of s, and obtain that the system equilibrates for $t < N^{1.6}$ steps per monomer for all systems up to 32 000 monomers. Specific values of the

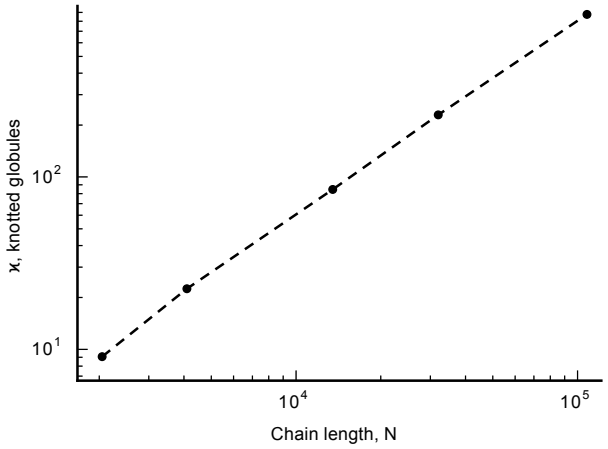

Fig. S8 Average $\varkappa$ for knotted globules of sizes 2048, 4000, 13 500, 32 000, and 108 000.

power $\alpha = \log(t)/\log(N)$ were: 1.583, 1.558, 1.543, 1.552 for 2048, 4096, 13 500, 32 000 monomers respectively (Fig. S10). We then extrapolate the $N^{1.6}$ time to longer chain lengths, $N = 128\,000$, 256 000, as we cannot obtain enough statistics to estimate $P_c(s)$ to less than 1% at that chain lengths. A longer, 128 000-monomer system was simulated for $6N^{1.6}$, while 256 000 system was simulated for $2.2N^{1.6}$. Based on the $P_c(s)$ and $R_G(s)$ plots, one can see that the 108 000-monomer system is equilibrated according to our criteria; however, we cannot estimate equilibration time precisely because of fluctuation in in $R_G(s)$ and $P_c(s)$ plots for different time points (Fig. S9) due to a small number of replicas. Consequently, we cannot use this system size for extrapolation. Thus we conclude that 108 000-monomer system is equilibrated, while 256 000-monomer system is equilibrated based on the extrapolation.

Contact probability and gyration radius analyses

We define two monomers to be in contact if they occupy two adjacent sites on a cubic lattice. To plot $P_c(s)$, we first divide all inter-monomer separations in bins of logarithmically increasing sizes, starting at 6 and going with a step of 1.1, rounding to the nearest even integer and removing duplicates. This yielded a sequence of $6, 8, 10, ..., x, 1.1x, 1.1^2x, ..., N$.

To define $R_G(s)$ for loops, we divided all distances to similar logarithmically space bins. We then considered all contacts which occur at that distance, and sampled a fixed number of loops from that set. If no contacts were present at a given bin of separations, we ignored this conformation; however, this happened in a fraction of a percent of cases. Averages were performed over replicas and over time. To obtain plots in the main text, these quantities were averaged over the last half of the trajectory for each run; for time-dependence plots they were averaged over logarithmically-spaced time bins with a

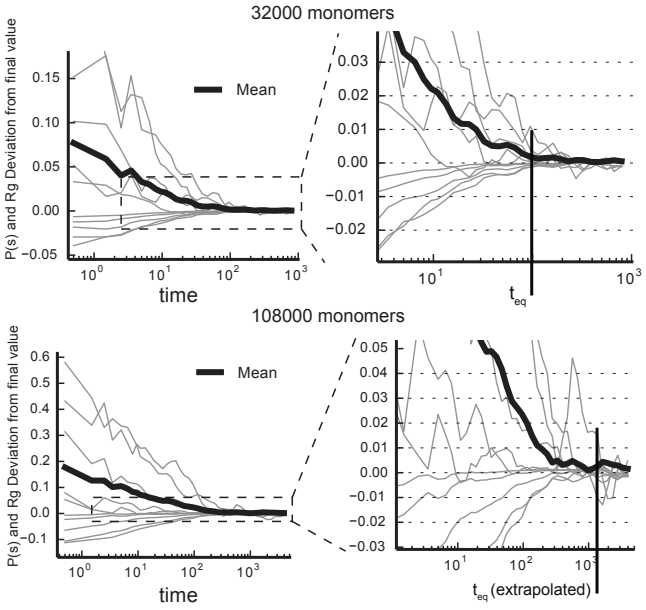

Fig. S9 Contact probability P_c and gyration radius R_G shown as a function of time for five distinct values of subchain length s for a 32 000-long globules (top) and 108 000-monomer globules (bottom). Thick black line labelled as "mean" denotes average absolute deviation of 10 curves from zero. Plots were divided by the mean value of the last 50% of the trajectory, and then 1 was subtracted to show relative deviation from the final value. Equilibration time for $N = 32\,000$ monomers was estimated as described in the text and equals $N^{1.55}$. Extrapolated time for $N = 108\,000$ monomers is calculated from a more permissive $N^{1.6}$ dependence.

coefficient of 1.2.

Calculation of Alexander polynomial

Finding Alexander polynomial for large rings could be a computationally intensive task. Here, we used the following strategy to do it.

First, we attempted to simplify the polymer ring using the previously used algorithm. We chose a pair of neighboring polymer bonds and attempted to remove a particle between them, replacing two bonds with one. We then checked if any of the other bonds are crossing the triangle formed by the two original bonds and the new bond. If none of the bonds were crossing this triangle, we accepted the replacement. We then continued choosing bonds until no further simplification could be made. This part was thoroughly tested, and we detected no differences in Alexander polynomial of the original and simplified polymer.

Second, we used the code adjusted from[42] to calculate the value of Alexander polynomial at -1.1. The current code is at http://bitbucket.org/mirnylab/openmm-polymer in a knots

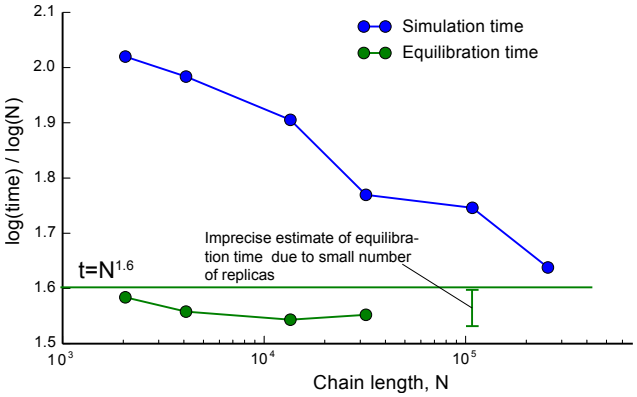

Fig. S10 Time of simulations and time of equilibration shown for unknotted globules of various sizes. Time is measured in logarithm base N, where N is number of monomers. Equilibration time for $N = 108\,000$ cannot be precisely estimated because fluctuations in the average $P_c(s)$ and $R_G(s)$ exceeds 1% due to small sample size . Upper and lower values of the range shown for $N = 108\,000$ denote the time when all 10, or 9 out of 10 observables fall below 1% relative deviation. Note that in Fig. S9 time points are logarithmically spaced, and hence time points for longer times aggregate information over longer time spawns. Therefore fluctuations due to the small number of replicas decrease with time, and allow us to put an upper boundary on the equilibration time for $N = 108\,000$.

folder.

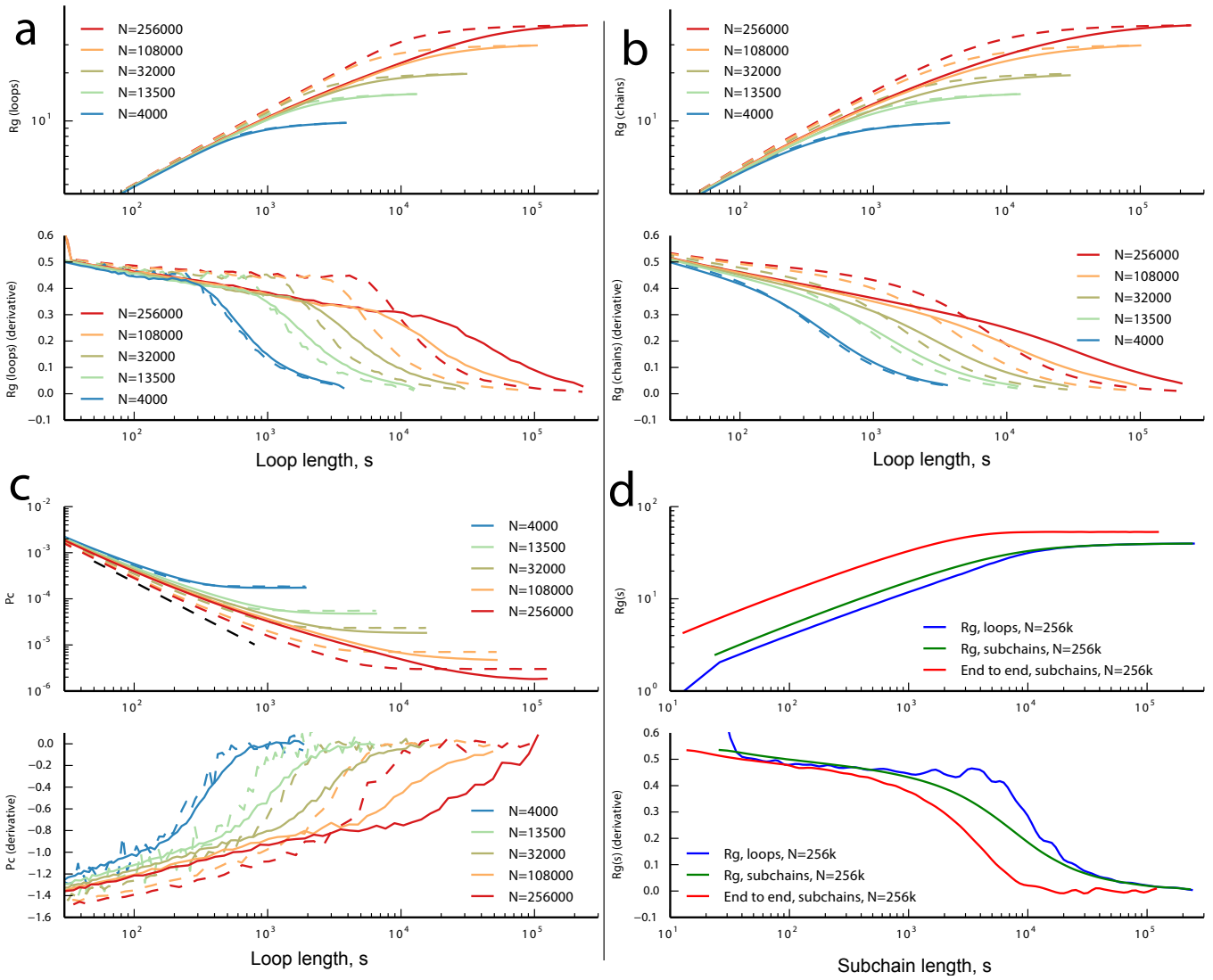

Fig. S11 In this figure we explore whether scaling regimes for $R_G(s)$ and $P_c(s)$ can be established. We compute a slope in the log-log coordinates as $d\log R_G(s)/d\log(s)$ and plot it as a function of $\log(s)$. As in the main text, solid line corresponds to the knotted globule, and dashed line - to the unknotted. Each of the four panels has two plots: the top plot is the quantity of interest ($R(s)$ or $P_c(s)$), and the bottom plot is its slope. **(a)** $R_G(s)$ for loops, similar to Fig. 3. **(b)** $R_G(s)$ evaluated for subchains. Note that both for the knotted and unknotted globule, all transitions between scaling regimes are more smooth. **(c)** The contact probability $P_c(s)$. The effect of confinement is more evident here affecting the scaling of P_c even for $s = 100$ for the knotted globule of $N = 256\,000$. Moreover, $P_c(s)$ has a plateau at $s \sim N$ for the knotted globule, which is not observed for corresponding $R_G(s)$. **(d)** Three quantities characterizing the spatial size of subchains in the knotted globule: R_G for loops, R_G for subchains, and the end-to-end distance R. For the same subchain length s, $R_G(s)$ for loops is smaller, and thus it starts to feel the effects of confinement at larger s, showing a clear $R_G(s) \sim s^{0.5}$ scaling regime up to several thousand monomers. $R_G(s)$ evaluated for subchains is larger, and the effect of constraints starts to manifest itself for smaller values of s. Finally, the end-to-end distance is the largest of the three, which makes it feel the confinement at even smaller s. The end-to-end distance shows a distinct plateau at $s \gtrsim 10\,000$. A similar plateau is seen in the $P_c(s)$ plots.