Supplementary Information: TOPOLOGY SORTING AND CHARACTERIZATION OF FOLDED POLYMERS USING NANO-PORES

Narges Nikoofard[a] and Alireza Mashaghi*[b]

1 Translocation of 2-contact chain in folded configuration

Consider translocation of a 2-contact chain through a nano-pore. Figs. S1(a) and S1(b) show the average number of passed monomers versus time for a polymer with 64 monomers ($l = 12$). The chain is pulled into a nano-pore with the internal diameter $d = 3b$, using two different forces $F = 10\frac{k_BT}{b}$ and $F = 6\frac{k_BT}{b}$. The contacts are supposed to be permanent. The shoulders in these figures correspond to pauses in the passage of the polymer due to the connected monomers. In Fig. S1(a), the first pause occurs at $N_{passed} = 13$ for all topologies, which is the position of the first contact site. A second shoulder is observed for series topology around $N_{passed} = 39$, which is the position of the third contact site.

In Fig. S1(b), the pause at $N_{passed} = 13$ is still observed. Here, we have two extra shoulders for the parallel topology. The extra shoulders correspond to: (1) when the first larger loop and the chain end are inside the nano-pore, around $N_{passed} = 39$; and (2) when the final loop is only remained to pass the pore, at $N_{passed} = 52$. The series and the cross topologies also have shoulders around $N_{passed} = 52$, when the following tail of the chain is going to pass the nano-pore.

The average position of the first monomer in the passage of the 2-contact chain with 64 monomers is shown in Figs. S2(a) and S2(b). In Fig. S2(a), two shoulders are seen in the case of the series topology under the strong force. Two shoulders are seen in the case of the parallel topology under the weak force, in Fig. S2(b). The second pause of the parallel topology in Fig. S2(b) occurs at the same time with the third pause in Fig. S1(b) corresponding to the passage of the final loop.

To generalize the obtained results, passage of a 2-contact chain with double spacing $l = 24$ between the contact sites is investigated using weak and strong pulling forces (Figs. S3(a) and S3(b)). All topologies have a shoulder at the position of the first contact site $N_{passed} = 25$. The series topology has an extra shoulder under the strong force around the position of the third contact site $N_{passed} = 75$ (Fig. S3(a)). The parallel topology has two extra shoulders in the weak force, at the middle of its larger loop $N_{passed} = 75$ and upon entry of the smaller loop $N_{passed} = 100$ (Fig. S3(b)). The inset plots show the average number of monomers inside the nano-pore versus time. The peaks in the insets occur at the same time with the shoulders in the main plots.

Additional simulations are performed to study the crossover between weak and strong forces, where series and topologies have extra pauses. Different pulling forces are examined $F = 10 - 5\frac{k_BT}{b}$ and the results are presented in Figs. S4(a)-S4(f). The transition seems to happen at $F = 7\frac{k_BT}{b}$, where the extra shoulder of the series topology disappears and one extra shoulder starts to appear for the parallel topology. The entropic fluctuations at the force $F = 5\frac{k_BT}{b}$ are so strong that no shoulders or peaks can be tracked in this pulling force.

[a] *Institute of Nanoscience and Nanotechnology, University of Kashan, Kashan 51167-87317, Iran.*
[b] *Harvard Medical School, Harvard University, Boston, MA 02115, USA. E-mail: Alireza_Mashaghi-Tabari@meei.harvard.edu*

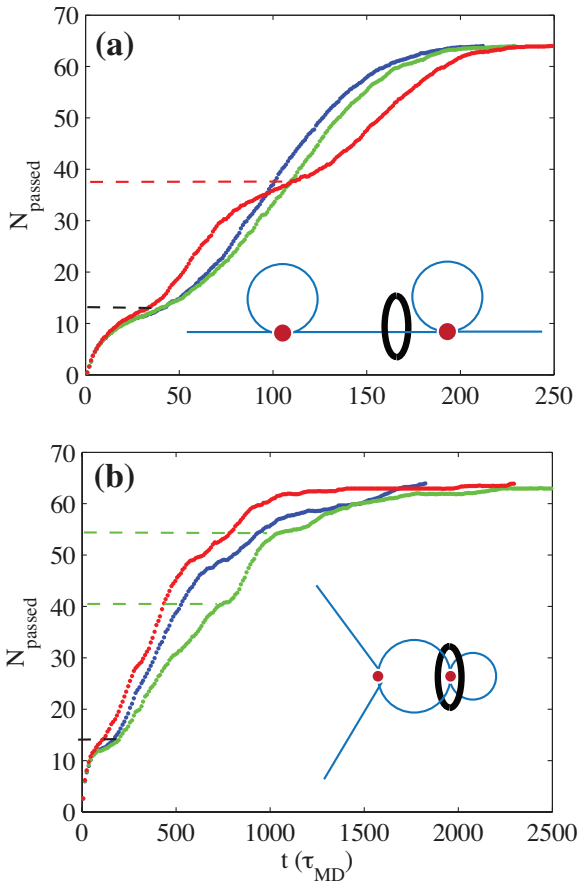

Figure S1: Average number of passed monomers versus time, for a 2-contact chain with 64 monomers. (a) The pulling force is $F = 10\frac{k_B T}{b}$. Schematics show configuration of the series topology during the second pause. (b) $F = 6\frac{k_B T}{b}$. Schematics show configuration of the parallel topology during the third pause.

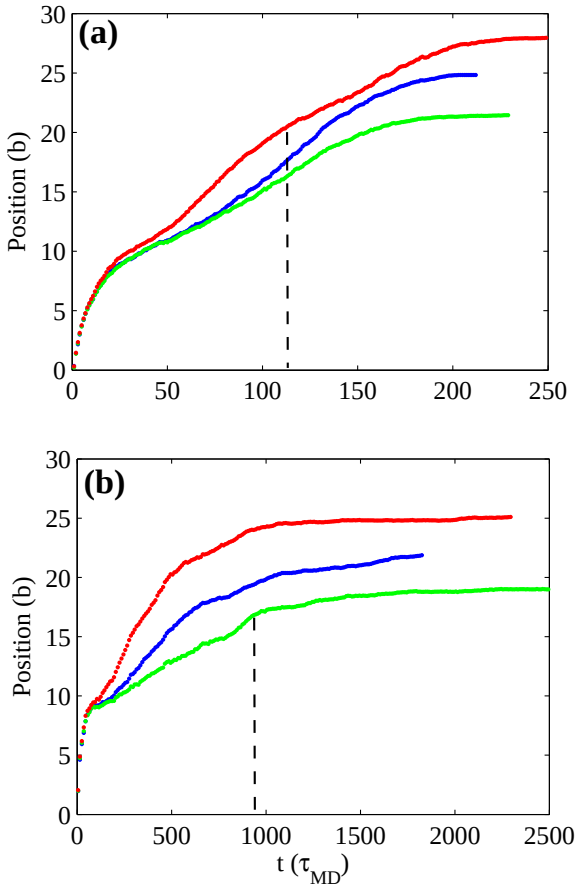

Figure S2: Average position of the first monomer versus time in the translocation of a 2-contact chain with 64 monomers. (a) The pulling force is equal to $F = 10\frac{k_B T}{b}$. Two shoulders are observed for the series topology. (b) $F = 6\frac{k_B T}{b}$. Two shoulders are observed for the parallel topology.

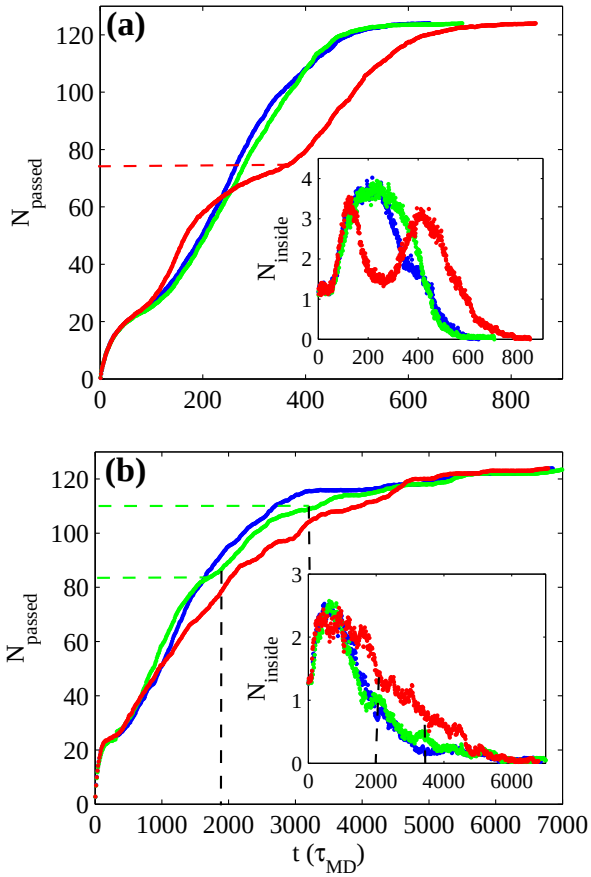

Figure S3: Average number of passed monomers versus time for a 2-contact chain with 124 monomers, for (a) $F = 10\frac{k_B T}{b}$ and (b) $F = 6\frac{k_B T}{b}$. The series topology has two shoulders in the strong force and the parallel topology has three shoulders in the weak force. Insets: Average number of monomers inside the nano-pore versus time. The peaks in the insets occur at the same time with the shoulders in the main plots.

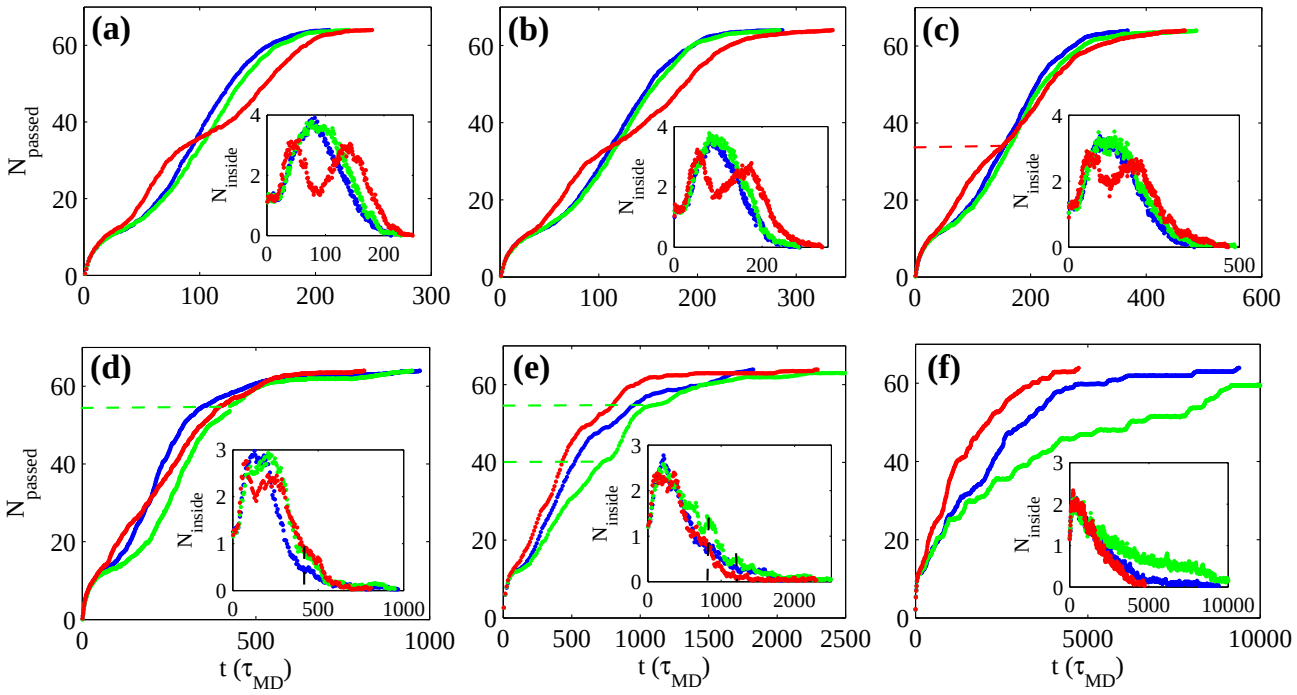

Figure S4: Crossover between weak and strong forces. Average number of passed monomers versus time, for (a) $F = 10\frac{k_BT}{b}$, (b) $F = 9\frac{k_BT}{b}$, (c) $F = 8\frac{k_BT}{b}$, (d) $F = 7\frac{k_BT}{b}$, (e) $F = 6\frac{k_BT}{b}$ and (f) $F = 5\frac{k_BT}{b}$. The extra shoulder seen at strong forces for the series configuration disappears at $F = 7\frac{k_BT}{b}$. One extra shoulder appears for parallel configuration in this pulling force. No information about the chain topology is obtainable under the weakest pulling force $F = 5\frac{k_BT}{b}$, where entropic contributions dominate.

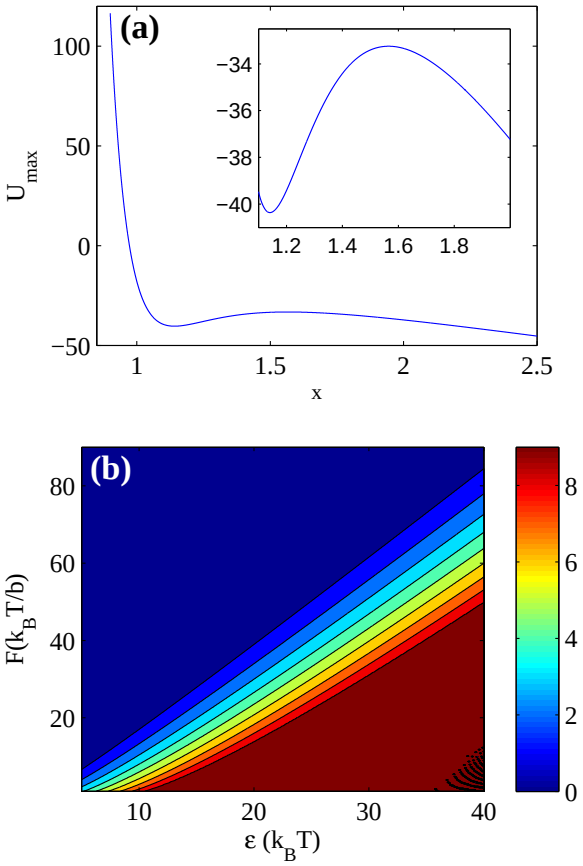

Figure S5: (a) A representative of the potential energy versus the distance between the contact sites in a chain pulled through a nano-pore. The bond strength and the pulling force are equal to $20k_BT$ and $18\frac{k_BT}{b}$, respectively. Inset: Plot of the potential energy focused in the barrier region. (b) Contour plot of the isoclines of the barrier height for various bond strengths and pulling forces, obtained from numerical calculations. The barrier height is only shown in the accessible region to simulation, $U_{max} < 9k_BT$. The irregularities in the lower left corner of the plot are the result of the limitations of the numerical methods for rapidly changing functions. These can be easily removed at the cost of higher computations.

2 Theoretical description of the contact rupture during the chain passage

Here, we develop a theoretical model to describe dependence of the pulling force on the bond strength. There are two parameters that determine the time required to rupture a bond during the chain passage; the bond strength and the pulling force. Our test simulations show that the pulling force needed to rupture the contacts in a reasonable time has a non-linear dependence on the bond strength.

As explained in the main manuscript, the contact sites attract each other through the Lennard-Jones potential. Thus, the potential energy landscape between the connected monomers is $U(x) = 4\epsilon_0 \left[\left(\frac{1}{x}\right)^{12} - \left(\frac{1}{x}\right)^{6} + \frac{1}{4}\right] - Fx$. Here, x is the distance between the monomers scaled to the monomer size. ϵ_0 and F are the bond strength and the pulling force, respectively. This potential function is plotted in Fig. S5(a) for $\epsilon_0 = 20k_BT$ and $F = 18\frac{k_BT}{b}$. It is seen that the contact sites should overcome a barrier to get fully detached. The time required to escape the barrier is proportional to the exponential of the maximum height of the potential, $\exp\left(\frac{U_{max}}{k_BT}\right)$. The maximum height of the potential (or

the barrier height) for $\epsilon_0 = 20k_BT$ and $F = 18\frac{k_BT}{b}$ is around $7k_BT$ (inset of Fig. S5(a)).

To find the barrier height for various pulling forces and bond strengths, we calculate the maximum and the minimum of the potential energy numerically. Positions of the minimum and the maximum of the potential energy are obtained from solutions of the equation $\frac{\partial U(x)}{\partial x} = 0$. Then, the barrier height U_{max} is equal to the difference between the potential energy at the maximum and the minimum of the potential.

As an example, we consider thermodynamic stability of polynucleotide hairpins which changes between $3.5 - 42.5k_BT$ in experiments [1]. Accordingly, we vary the bond strength in the numerical solutions between $5 - 40k_BT$. The pulling force is changed in the interval $1 - 90k_BT/b$. Only barrier heights smaller than $9k_BT$ are accessible to simulations. The contour plot of isoclines of the barrier height in the accessible region of the phase space is shown in Fig. S5(b).

In Fig. S5(b), there is a region in which the barrier height is equal to zero. A zero height of the barrier eliminates the effect of the entropic fluctuations on the passage process. Thus, to find the position of the contact sites using the trend of the position of the first monomer (Fig. 4 of the main manuscript), the pulling force in simulations should be chosen such that the barrier height is zero. However, to discriminate between the three topologies by using the passage time (Fig. 5 of the main manuscript), the simulations should be performed in the region that the height of the barrier is nonzero.

3 Translocation of 2-contact chain coupled to unfolding

Consider translocation of a 2-contact chain through a nano-pore with internal diameter $d = 1.4b$. The chain has to unfold to pass through the nano-pore. The strength of the bonds determines the required pulling force to unfold the chain. Shoulders are observed in the time profile of the average number of passed monomers using suitable pulling forces (Figs. S6(a) and S6(b)). These pauses are due to the time required to rupture the contacts. The first pause for all topologies occurs at the position of the first contact site, $N_{passed} = 12$. The second pause happens at $N_{passed} = 24$ for parallel and cross topologies, and at $N_{passed} = 36$ for series topology. The suitable force should be strong enough to overcome the bond strength easily. As shown in the inset plots of Figs. S6(a) and S6(b), fluctuations in weak pulling forces hide the pauses caused by topology. Too strong forces are not appropriate either because they eliminate the effect of topology completely (results not shown here).

In the strong pulling forces, the leading end of the polymer becomes completely stretched just before breakage of each of the contacts. So, the average position of the first monomer can be also used to read the position of the contact sites along the chain (Figs. 4(a) and 4(b) in the main manuscript).

The average translocation time differs for 2-contact chains with different topologies, in each bond energy and pulling force. Topologies that have the minimum and the maximum of the average passage times for 50 data points are shown in Figs. S7(a) and S7(b). Topologies corresponding to the minimum and the maximum of the average passage times for 25 data points are shown in Figs. S7(c) and S7(d). It is observed that generally the orders of the average passage times in each bond energy and pulling force do not depend on the data sets used for averaging.

The values of the average passage times for the passage of the 2-contacts chain are shown in Table S1. It is seen that the average passage times in the bond energy $30k_BT$ and the pulling force $35\frac{k_BT}{b}$ are larger than the other times given in the table. Thus, the pulling force $35\frac{k_BT}{b}$ is not strong enough to be able to break the contacts with the energy $30k_BT$. This causes the entropic effect to be dominant and the dependence of the order on the data set (observed in Fig. S7). The other difference in the order in Fig. S7 occurs at the bond energy $40k_BT$ and the pulling force $50\frac{k_BT}{b}$. The passage times in these values of the bond energy and pulling force are relatively small. Thus, the dependence on the data set is caused by a strong force which decreases the effect of topology on the passage time.

4 Translocation of 5-contact chain coupled to unfolding

The effect of topology of a 5-contact chain on the average passage time through a nano-pore is studied in three sets of simulations. The simulations for each bond energy and pulling force are repeated for 150 random chains fulfilling one of the following conditions: $N_c = 8$, $N_p = 8$ or $N_s = 8$. The average passage time changes with the dominant topology in the configurations of the chains. The minimum and the maximum passage times are shown in Figs. 5(c) and 5(d) in the main manuscript. To show that the obtained order for the passage times is meaningful, we calculate

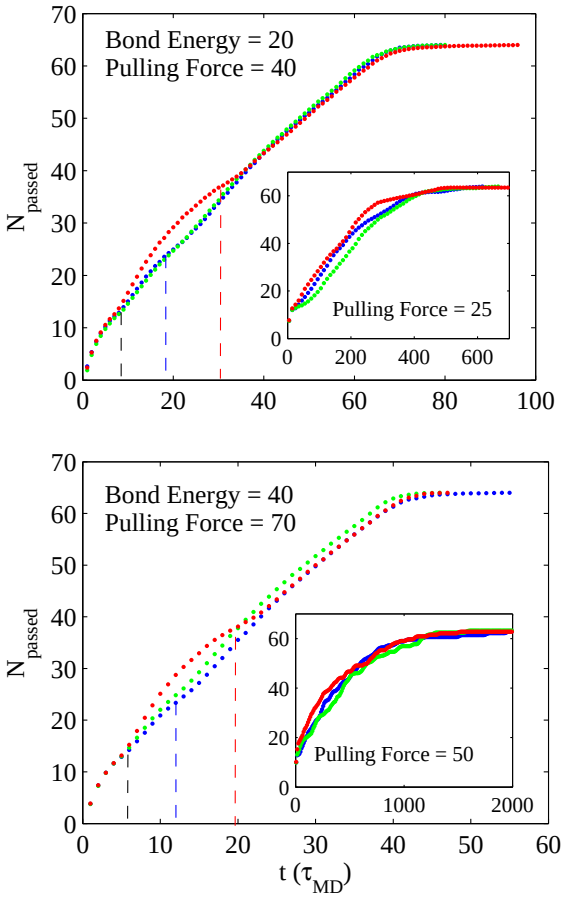

Figure S6: Average number of passed monomers versus time for two different bond strengths. All of the three topologies show two pauses in their passage process. This is because of the time needed to rupture the two contacts in their configuration. Insets: It is not possible to track the pauses in weaker pulling forces in the same bond energies.

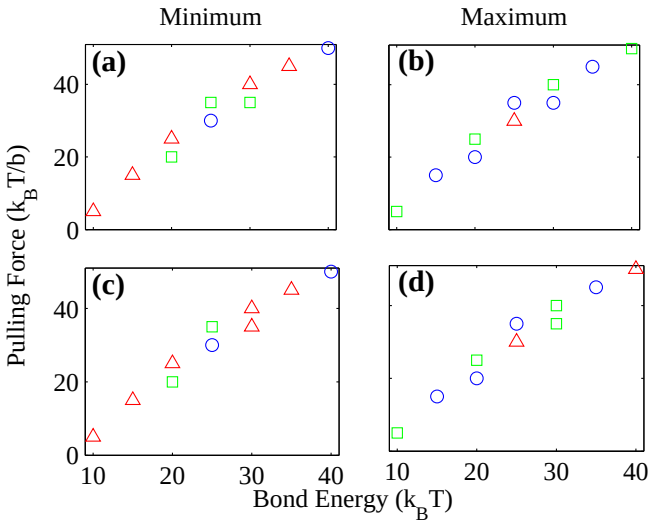

Figure S7: Topologies with the minimum and the maximum passage times in the translocation of a 2-contact chain. In panels (a) and (b), the passage times are averaged over 50 realizations. In panels (c) and (d), the passage times are averaged over 25 realizations. It is observed that only one symbol is different between the panels (a) and (c). Two symbols are different between the panels (b) and (d).

the average passage time over 50, 100 and 150 data points. The orders of the average passage times for different bond energies and pulling forces do not generally depend on the data set used for averaging (Table S2). This is true while the data set contains more than 100 data points.

The fourth set of simulations is performed for 150 completely random chains. The order of the random set changes among the other three sets, with the number of data points used for averaging (Table S2). This dependence on the data set shows that the 150 randomly chosen chains cannot represent the whole phase space containing 1500 chains.

References

[1] Schink S.; Renner S.; Alim K.; Arnaut V.; Simmel F. C.; Gerland U. Quantitative analysis of the nano-pore translocation dynamics of simple structured polynucleotides. *Biophys. J.* **2012**, 102, 85-95.

2-contact chain	Topology	50 data points	Order	25 data points	Order
Bond Energy $= 10k_BT$ Pulling Force $= 5\frac{k_BT}{b}$	Cross	839	2	824	2
	Parallel	861	3	927	3
	Series	775	1	753	1
Bond Energy $= 15k_BT$ Pulling Force $= 15\frac{k_BT}{b}$	Cross	504	3	518	3
	Parallel	454	2	432	2
	Series	438	1	414	1
Bond Energy $= 20k_BT$ Pulling Force $= 20\frac{k_BT}{b}$	Cross	1286	3	1201	3
	Parallel	789	1	739	1
	Series	814	2	855	2
Bond Energy $= 20k_BT$ Pulling Force $= 25\frac{k_BT}{b}$	Cross	269	2	270	2
	Parallel	293	3	276	3
	Series	268	1	240	1
Bond Energy $= 25k_BT$ Pulling Force $= 30\frac{k_BT}{b}$	Cross	547	1	505	1
	Parallel	589	2	538	2
	Series	613	3	607	3
Bond Energy $= 25k_BT$ Pulling Force $= 35\frac{k_BT}{b}$	Cross	173	3	159	3
	Parallel	142	1	139	1
	Series	166	2	156	2
Bond Energy $= 30k_BT$ Pulling Force $= 35\frac{k_BT}{b}$	Cross	1315	3	1132	2
	Parallel	1058	1	1180	3
	Series	1277	2	952	1
Bond Energy $= 30k_BT$ Pulling Force $= 40\frac{k_BT}{b}$	Cross	209	2	161	2
	Parallel	236	3	209	3
	Series	139	1	149	1
Bond Energy $= 35k_BT$ Pulling Force $= 45\frac{k_BT}{b}$	Cross	434	3	480	3
	Parallel	415	2	382	2
	Series	353	1	305	1
Bond Energy $= 40k_BT$ Pulling Force $= 50\frac{k_BT}{b}$	Cross	584	1	478	1
	Parallel	603	3	610	2
	Series	587	2	636	3

Table S1: Average translocation time of 2-contact chains with three different topologies. The averages are taken over different sets of data, containing different number of realizations. The order of the average passage times of the three topologies does not change with the data set in most cases.

5-contact chain	Topology	50 data points	Order	100 data points	Order	150 data points	Order
Bond Energy $= 10k_BT$ Pulling Force $= 5\frac{k_BT}{b}$	$N_c = 8$	2635	3	2691	3	2608	3
	$N_p = 8$	2478	1	2546	2	2568	2
	$N_s = 8$	2505	2	2523	1	2496	1
	Random	2610		2637		2649	max
Bond Energy $= 15k_BT$ Pulling Force $= 15\frac{k_BT}{b}$	$N_c = 8$	1489	2	1466	2	1460	2
	$N_p = 8$	1426	1	1435	1	1437	1
	$N_s = 8$	1523	3	1506	3	1536	3
	Random	1502		1569		1544	max
Bond Energy $= 20k_BT$ Pulling Force $= 20\frac{k_BT}{b}$	$N_c = 8$	3058	2	3104	2	3173	2
	$N_p = 8$	2965	1	3102	1	3132	1
	$N_s = 8$	3394	3	3412	3	3507	3
	Random	3208		3376		3274	
Bond Energy $= 20k_BT$ Pulling Force $= 25\frac{k_BT}{b}$	$N_c = 8$	922	2	952	3	923	1
	$N_p = 8$	953	3	943	2	944	2
	$N_s = 8$	898	1	916	1	954	3
	Random	943		930		943	
Bond Energy $= 25k_BT$ Pulling Force $= 30\frac{k_BT}{b}$	$N_c = 8$	1731	2	1801	3	1750	2
	$N_p = 8$	1688	1	1660	1	1637	1
	$N_s = 8$	1804	3	1757	2	1853	3
	Random	1693		1726		1741	
Bond Energy $= 25k_BT$ Pulling Force $= 35\frac{k_BT}{b}$	$N_c = 8$	620	2	607	2	613	2
	$N_p = 8$	590	1	603	1	609	1
	$N_s = 8$	722	3	711	3	692	3
	Random	605		616		639	
Bond Energy $= 30k_BT$ Pulling Force $= 35\frac{k_BT}{b}$	$N_c = 8$	3148	1	3254	1	3376	1
	$N_p = 8$	3588	3	3675	3	3698	3
	$N_s = 8$	3512	2	3294	2	3455	2
	Random	3373		3396		3500	
Bond Energy $= 30k_BT$ Pulling Force $= 40\frac{k_BT}{b}$	$N_c = 8$	912	3	920	3	948	3
	$N_p = 8$	840	1	863	1	874	1
	$N_s = 8$	893	2	902	2	903	2
	Random	917		958		946	
Bond Energy $= 35k_BT$ Pulling Force $= 45\frac{k_BT}{b}$	$N_c = 8$	1476	3	1466	3	1428	3
	$N_p = 8$	1443	2	1402	2	1419	2
	$N_s = 8$	1427	1	1397	1	1363	1
	Random	1405		1473		1488	max
Bond Energy $= 40k_BT$ Pulling Force $= 50\frac{k_BT}{b}$	$N_c = 8$	1988	2	1994	3	1982	3
	$N_p = 8$	1998	3	1936	2	1943	2
	$N_s = 8$	1867	1	1845	1	1785	1
	Random	2090		1993		2007	max

Table S2: Average translocation time of random 5-contact chains. In three sets of simulations, the random chains are chosen such that one of the three topological arrangements is dominant in their configuration. In the fourth set of simulation, the chains are chosen completely at random. The averages are taken over different number of realizations. The order of the three sets with dominant topologies does not change with the data set while the data set contains more than 100 realizations in most cases. However, the completely random set changes order with the number of realizations used for averaging.