Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.
This journal is © the Owner Societies 2017

*Supplementary Materials for*

# Topology of internally confined polymer chain

Maziar Heidari[a,b†], Vahid Satarifard[a,c†], Sander J. Tans[d], Mohammad Reza Ejtehadi[e], Samaneh Mashaghi[f], Alireza Mashaghi[a,*]
[a] *Leiden Academic Centre for Drug Research, Faculty of Mathematics and Natural Sciences, Leiden University, Leiden, The Netherlands*
[b] *Max Planck Institute for Polymer Research, Mainz, Germany*
[c] *Max Planck Institute of Colloids and Interfaces, Potsdam, Germany*
[d] *FOM institute AMOLF, Amsterdam, The Netherlands*
[e] *Department of Physics, Sharif University of Technology, Tehran, Iran*
[f] *School of Engineering and Applied Sciences and Department of Physics, Harvard University, Cambridge, Massachusetts, USA*
∗*Correspondence: a.mashaghi.tabari@lacdr.leidenuniv.nl and*
†*Equal contribution*
(Dated: May 30, 2017)

## COMPARISON OF FOLDING TIME LANDSCAPE

The folding time for a polymer chain with length $L = 50l$ is shown in Fig. 1. The chain has four binding sites (i.e. two contacts) which are randomly distributed over the chain. Each curve is representing the case explained in the main text.

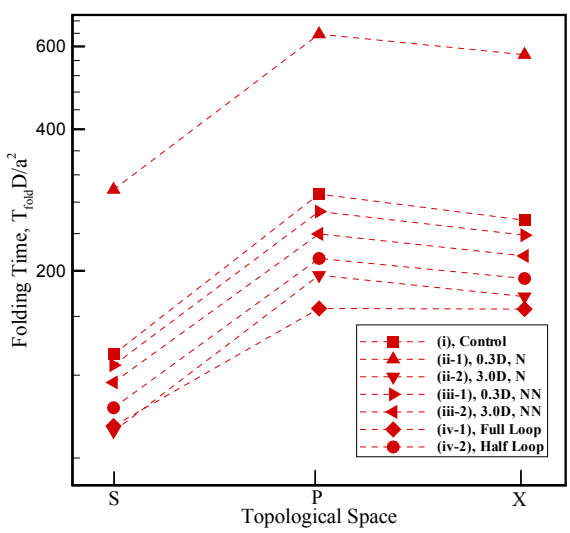


FIG. 1. Rescaled folding time ($T_{fold}D/a^2$) of polymer with two contacts in different topological spaces, series (S), parallel (P) and cross (X). The standard errors are smaller than the symbols' size. The polymer length is $L = 50l$ and the names of each curve is given in the legend according to the definitions described in the main text.

Fig. 2 shows the folding time landscape of the chain, when the external molecule (Chaperone) sticks to either the chain's binding sites(Internal loop) or adds another loop to the chain native state (External loop). In both cases, the diffusion constant of the chaperone could be higher ($3.0D$) or lower ($0.3D$) than the diffusion constant of chain monomers ($D$). This figure is similar to Fig. 3 in the main text but the lower and upper bounds in the color bars are set differently.

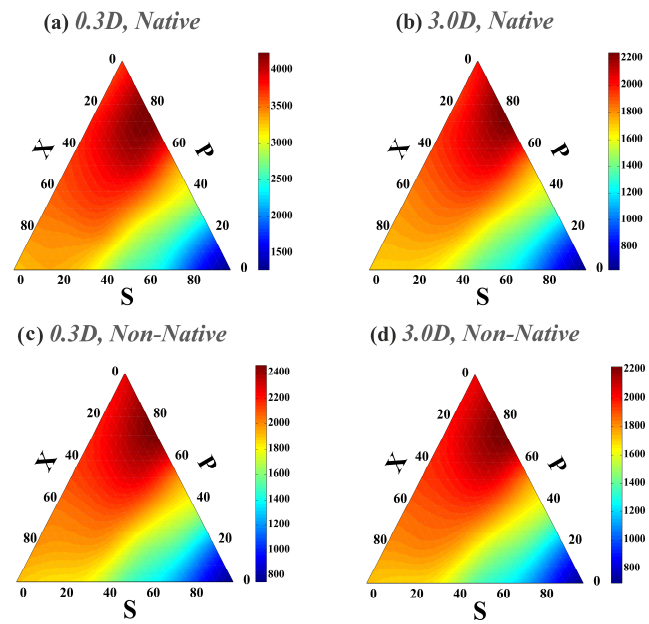


FIG. 2. Ternary plots of rescaled folding time ($T_{fold}D/a^2$) for a polymer with four contacts in its native state when one internal or external contact has different diffusion constant with respect to the other polymer contacts. Plot (a)-(d) respectively corresponds to the cases ($ii$-1)-($iii$-2) which are explained in the main text.

## FOLDING RATES ESTIMATION

### Approximate solution

Based on the shortest distance in the final folded state of the chain, it is possible to find simple expressions that roughly estimate the folding rates of the different topo-

logical states [1, 2]. For a chain whose length is $L$ and its binding sites are separated by contour lengths of $l_1$, $l_2$ and $l_3$ (see Fig. 3), the folding rates of S-, P- and X-loops can be estimated as $r_s = 1/(l_1^{3/2} + l_3^{3/2})$, $r_p = 1/(l_2^{3/2} + (l_1+l_3)^{3/2})$ and $r_x = 1/((l_1+l_2)^{3/2} + (l_2+l_3)^{3/2})$ respectively. The averaged values of rates can be obtained as

$$\langle r_y(l_2)\rangle = \frac{\sum_{l_1,l_3=1}^{L} r_y \theta(L-(l_1+l_2+l_3))}{\sum_{l_1,l_3=1}^{L} \theta(L-(l_1+l_2+l_3))} \qquad (1)$$

Where $\theta(x) = 1$ for $x \geq 0$ and $\theta(x) = 0$ otherwise.

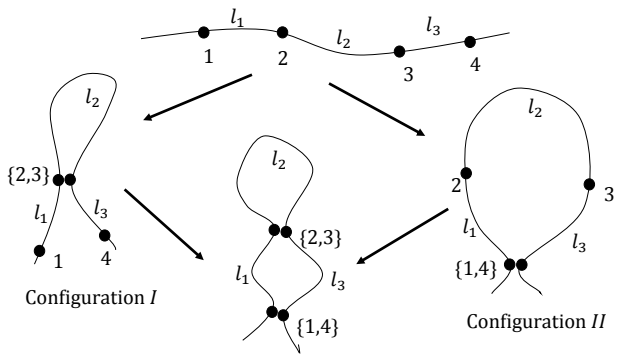


FIG. 3. A chain with four binding sites. The binding sites are separated by contour segments of $l_1$, $l_2$ and $l_3$ along the chain. There are two possible pathways, configuration I and II, to close the binding sites of the chain into the parallel state.

The resulting averages are shown in Fig. 4. The results are obtained from a chain of length $L = 200l$, where $l$ is a chain segment, and the average is taken over the ensembles of loops in which the length $l_1$ and $l_3$ are varied in the range of $l$ to $80l$. For the whole range of inter-loop distances ($l_2$), the average folding rate of the S-loops is higher than the rate of the other loop sets in the parallel and cross topologies. Additionally, $\langle r_s\rangle$ remains constant since the folding rate of the S-loops is independent of the inter-loop distance. For the short inter-loop distances ($l_2 < 4$) X-loops fold faster than P-loops, however, as the inter-loop distances grow, the folding rates of the loops in both topological sets are reduced and the P-loops starts to fold faster than X-loops. The result of averaging all folding rates over the inter-loop distances is ($\langle r_s\rangle_{l_2}$,$\langle r_p\rangle_{l_2}$,$\langle r_x\rangle_{l_2}$ $=10^{-3}\,(3.5, 1.9, 1.5)$). The S-loops have the fastest folding kinetics while the P-loops fold and X-loops have approximately the same overall folding rate. There are two main differences between the folding time obtained from the current relation and the KMC method. In the relation proposed for folding rates of the P-loops, it is assumed that the zipping mechanism is always effective. Thus the sequence in the folding process is deterministic and the shorter loop folds initially. However, in the KMC, the sequence of the folding process occurs probabilistically and the probability of the event that prior to the smaller loop, the larger loop is closed is not zero. The second difference is that the parameter space in KMC method is not fully explored. As it is shown in the inset of Fig. 4, the folding rates of the X-loops exceed the P-loops for the relatively large loops.

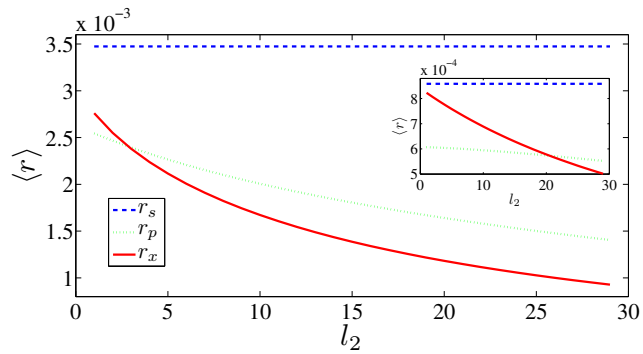


FIG. 4. Rescaled folding rates ($\langle r\rangle$) of a chain having the length of $L = 200l$ and four binding sites obtained from equation 1 against the inter-loop distance $l_2$. The binding sites are randomly distributed along the chain with the following contour lengths, $l_1$, $l_2$ and $l_3$. The ensemble average is taken over the lengths $l_1, l_3 < 80l$. Inset shows the rescaled folding rate which is averaged only over the large loops, $60l < l_1, l_3 < 80l$.

### Zipping mechanism in ideal chain

Depending on which loop is closed faster, there are two possible events in the folding process of parallel loops. As illustrated in Fig. 3, the linear Gaussian chain has four binding sites which are separated by contour segments of $l_1$, $l_2$, $l_3$. Each segment of the chain is modeled by a spring whose elastic stiffness is $\gamma$. As explained in KMC section, the association rate between the binding sites is scaled by the effective stiffness between each pair of the binding sites, $k_a\ \gamma_{eff}^{3/2}$. For the configuration I, the inner loop is closed firstly, and the stiffness matrix of the configuration becomes

$$K_I = \gamma \begin{pmatrix} \frac{1}{l_1} & -\frac{1}{l_1} & 0 & 0 \\ -\frac{1}{l_1} & \frac{1}{l_1}+\frac{2}{l_2} & -\frac{2}{l_1} & 0 \\ 0 & -\frac{2}{l_2} & \frac{2}{l_2}+\frac{1}{l_3} & -\frac{1}{l_3} \\ 0 & 0 & -\frac{1}{l_3} & \frac{1}{l_3} \end{pmatrix} \qquad (2)$$

The stiffness matrix is obtained based on finite element theory of the linear spring network, for more information see ref. [3, 4]. Alternatively, for the event in which the outer loop is closed initially, as shown in Fig. 3, the

stiffness matrix becomes:

$$K_{II} = \gamma \begin{pmatrix} \frac{1}{l_1} + \frac{1}{L} & -\frac{1}{l_1} & 0 & -\frac{1}{L} \\ -\frac{1}{l_1} & \frac{1}{l_1} + \frac{1}{l_2} & -\frac{1}{l_2} & 0 \\ 0 & -\frac{1}{l_2} & \frac{1}{l_2} + \frac{1}{l_3} & -\frac{1}{l_3} \\ -\frac{1}{L} & 0 & -\frac{1}{l_3} & \frac{1}{l_3} + \frac{1}{L} \end{pmatrix} \quad (3)$$

To calculate the effective association rate of the folded structures for both configurations, it is required to solve the following equation,

$$X = K^{-1} F \quad (4)$$

Where for the configurations I and II the corresponding vectors are $F_I = [1\ 0\ 0\ \ -1]^T$ and $F_{II} = [0\ 1\ \ -1\ 0]^T$. The resulting effective stiffness for two configurations are obtained as

$$\gamma_{e,I} = \frac{2\gamma}{2l_1 + l_2 + 2l_3} \quad (5)$$

$$\gamma_{e,II} = \frac{2\gamma}{l_2} \frac{l_1 + l_2 + l_3}{2l_1 + l_2 + 2l_3} \quad = \left(1 + \frac{(l_1 + l_3)}{l_2}\right) \gamma_{e,I} \quad (6)$$

It is clear that $\gamma_{e,I} < \gamma_{e,II}$ however, the difference becomes smaller if size of the inner loop grows to the size of the outer loop $l_2 \to l_1 + l_3$. The corresponding folding time is calculated as

$$T_I^{fold} = T_1 + A\gamma_{e,I}^{-3/2} \quad (7)$$

$$T_{II}^{fold} = T_1 + A\gamma_{e,II}^{-3/2} \quad (8)$$

Where $A$ is a constant prefactor. An equal folding time ($T_1$) is assigned for both configurations since the total association rates of the binding pairs in the both configurations ($k_{tot}$) is the same (see relation 3 in the main text). This means that the escape time of unfolded state of the chain into the partially folded state (either configuration I or II) is the same. Hence, since $T_I^{fold} > T_{II}^{fold}$, the zipping mechanism is not an effective pathway in the KMC folding of the Gaussian chain model. The reason can be explained by the constrains that both configurations have. For the configuration II, since the chain is ideal, we can neglect the other two tails of the chain and assume it as a ring. In this case, since the starting and finishing points of the chain are constraint to a point, the chain has less configuration entropy compared to when a point on the chain is fixed (Configuration I).

### Comparing approximate solution with KMC

In this section, we assume a Gaussian chain having $L = 200l$ segments and the binding site along the chains are distributed such that the two lengths $l_1 = 80l$ and $l_3 = 80l$ are fixed and $l_2$ varies between $5l$, $10l$ and $15l$. The folding rates obtained from both equation (1) and KMC method is shown in Fig. 5. All folding rates are normalized by the constant $\langle r_s \rangle$. As expected, the folding rate of S-loops obtained from the both methods does not depend on the inter-loop distances $l_2$. Additionally, there exists similar trends and scaling for the averaged folding rates of P-loops and X-loops which confirms the KMC model approximately agrees well with equation 1. However, large error bars are observed for the folding rates of P-loops with respect to the other topological states. This is originated from the bi-modal distribution of P-loops folding pathways which is already explained in the previous section. In this regard, the ensemble of P-loops is separated into two independent configurations I and II ensembles and then the average is taken over each ensemble, i.e. $\langle r_p \rangle_{confI}$ and $\langle r_p \rangle_{confII}$. As this is shown in Fig. 5(b), while the error bars of rates in each ensemble have been reduced dramatically, the folding rate of P-loops in the second configuration exceed the rates of S-loops. This is somehow expected because in this case, the size of the loops is much larger than the inter-loop distance $l_1,\ l_3 > l_2$.

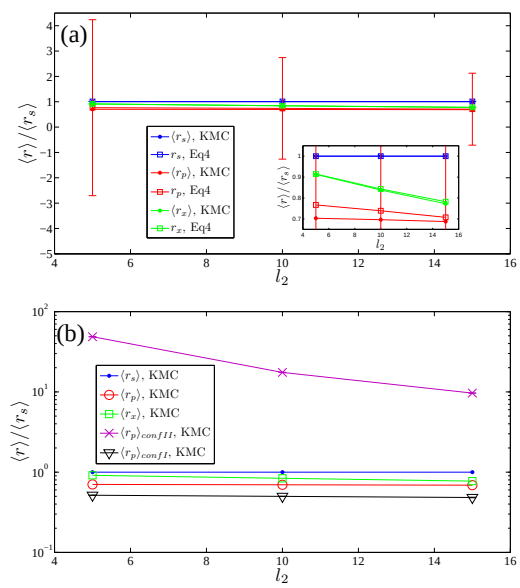


FIG. 5. (a) Rescaled folding rates ($\langle r \rangle / \langle r_s \rangle$) of a chain having the length of $L = 200l$ and four binding sites. The binding sites are distributed along the chain with $l_1 = l_3 = 80l$ and the length $l_2$ varies in $l_2 = 5l$, $10l$ and $15l$. Inset shows the rescaled similar figure. (b) Folding rate of the same system as explained in (a) but with extra ensemble averages $\langle r_p \rangle_{confI}$ and $\langle r_p \rangle_{confII}$.

## MOLECULAR DYNAMICS SIMULATIONS

The coarse-grained model of polymer is simulated by beads and springs. The total potential of the chain is given by $U = U_{bond} + U_{bend} + U_{excl}$[5, 6]. The connectivity of the monomers is controlled by bonding energy

$U_{bond} = k_s/2 \sum (r_{i,i+1} - d_0)^2$, where $k_s$ and $d_0$ denote the spring stiffness and equilibrium distance respectively. The bending energy is given by $U_{bend} = k_\theta/2 \sum (\theta_{i,i+1} - \theta_0)^2$. The bending stiffness, angle between two neighboring bonds and equilibrium angle are denoted by $k_\theta$, $\theta_{i,i+1}$ and $\theta_0$ respectively. The self-avoidance is controlled by repulsive part of the Lennard-Jones potential $U_{excl} = 4\epsilon \sum \left[ (\sigma/r_{i,j})^{12} - (\sigma/r_{i,j})^6 \right]$, where $\epsilon$ and $\sigma$ denote the well depth and monomer diameters. The simulations were carried out in canonical (NVT) ensemble in temperature $k_B T = \epsilon = 1$, using the Nose-Hoover thermostat via LAMMPS package [7].


[1] T. Guérin, O. Bénichou, and R. Voituriez. Non-markovian polymer reaction kinetics. *Nature Chemistry*, 4(7):568–573, 2012.

[2] A. Szabo, K. Schulten, and Z. Schulten. First passage time approach to diffusion controlled reactions. *J. Chem. Phys.*, 72(8):4350–4357, 1980.

[3] D. E. Makarov and H. Metiu. A model for the kinetics of protein folding: kinetic monte carlo simulations and analytical results. *J. Chem. Phys.*, 116(12):5205–5216, 2002.

[4] D. E. Makarov and G. J. Rodin. Configurational entropy and mechanical properties of cross-linked polymer chains: implications for protein and rna folding. *Phys. Rev. E.*, 66(1):011908, 2002.

[5] O. Mashinchian, S. Bonakdar, H. Taghinejad, V. Satarifard, M. Heidari, M. Majidi, S. Sharifi, A. Peirovi, S. Saffar, M. Taghinejad, et al. Cell-imprinted substrates act as an artificial niche for skin regeneration. *ACS Appl. Mater. Interfaces*, 6(15):13280–13292, 2014.

[6] D. Meluzzi and G. Arya. Recovering ensembles of chromatin conformations from contact probabilities. *Nucleic Acids Res*, page gks1029, 2012.

[7] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.*, 117(1):1–19, 1995.