Supplementary Information for "Circuit topology of self-interacting chains: implications for folding and unfolding dynamics"

Andrew Mugler[1,2,3], Sander J. Tans[3], and Alireza Mashaghi[4,*]

[1]Department of Physics, Purdue University, 525 Northwestern Avenue, West Lafayette, IN 47907, USA

[2]Department of Physics, Emory University, 400 Dowman Drive, Atlanta, GA 30322, USA

[3]FOM Institute AMOLF, Science Park 104, 1098 XG Amsterdam, The Netherlands

[4]Harvard Medical School, Harvard University, 25 Shattuck St, Boston, MA 02115, USA

1 Topology matrix during the unfolding process

The dynamics of folding or unfolding can be illustrated using the matrix representation of topology. Fig. S1 shows an example corresponding to a particular unfolding pathway in Fig. 2B of the main text.

2 Types of forbidden unfolding transitions

As noted in the main text, forbidden transitions in an unfolding tree may arise in several distinct ways: (i) the contact might not lie along the shortest path(s) from one end of the chain to another, (ii) the transition might originate from an inaccessible state, or (iii) the transition might involve breaking two contacts and reforming another. These three types of forbidden transition are illustrated for an example chain in Fig. S2A, and the statistical effects of topology on these three types across many chains are shown in Fig. S2B. Several trends are apparent in Fig. S2B, for reasons similar to those underlying Fig. 2C in the main text, as we now describe.

The fractions of type i and type ii forbidden transitions correlate with topological features in similar ways (Fig. S2B, first and second rows): both generally increase with the fraction of parallel relations, decrease with the fraction of series relations, and have little dependence on the fraction of cross relations. These are the same trends as seen in Fig. 2C in the main text, and they arise for the same reason: a parallel relation forbids one of two possible unfolding transitions, whereas a series relation does not. The fact that both types show these correlations is because type ii transitions "amplify" type i transitions. That is, as seen in Fig. S2A, each type i transition gives rise to several type ii transitions further down the tree. As such, types i and ii transitions correlate similarly with topology, and the overall fraction of type ii is higher than that of type i, as seen in Fig. S2B.

*Alireza-Mashaghi-Tabari@meei.harvard.edu

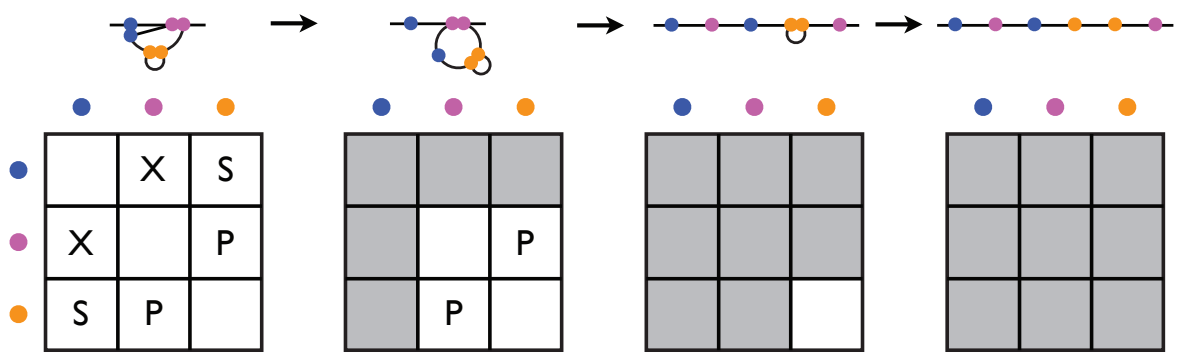


Figure S1: Matrix representation of an unfolding pathway shown in Fig. 2B of the main text. For contacts that are ripped apart, the corresponding column and row are shaded. For a given configuration, any two contacts are assigned a topological relation from the set {S, P, X}.

On the other hand, the fraction of type iii forbidden transitions shows the opposite trends as types i and ii (Fig. S2B, third row): the fraction of type iii forbidden transitions decreases with the fraction of parallel relations and increases with the fraction of series relations. The reason is that whereas type i and ii transitions tend to restrict the size of the unfolding tree by forbidding transitions "outside" of it, type iii transitions are generally found "inside" of folding trees (see Fig. S2A). Therefore, since parallel relations introduce more type i and ii forbidden transitions, which results in narrower trees, they also reduce the available space for type iii forbidden transitions, which generally reside within these trees. Conversely, since series relations allow for wider trees, they allow for more type iii forbidden transitions within these trees.

Fig. S2B shows that the majority of forbidden transitions are type ii. For this reason, type ii transitions, together with type i, dominate over type iii transitions, and set the direction of trends seen for all forbidden transitions in Fig. 2C in the main text.

3 Topology and folding rate, for stochastic folding

In the main text, we show how topology affects folding rates under the assumption that folding is deterministic (Fig. 3 in the main text). Here we show that these effects persist when folding is stochastic.

As in deterministic folding, the time for each contact to form is proportional to the shortest inter-contact distance along the partially folded chain, raised to the 3/2 power. Here, in stochastic folding, the probability of a contact forming is inversely proportional to its formation time. The folding time is then the total time to go from the fully unfolded state to the fully folded state, averaged over all paths with their associated probabilities, and the folding rate is the inverse of this time.

Fig. S3 shows results from the same protocol as Fig. 3C and D of the main text, but with stochastic folding instead of deterministic folding. We see that, just as in deterministic folding, the folding rate increases with the fraction of parallel relations at constant contact order; and, for proteins without parallel relations, the folding rate also increases with the fraction of cross relations at constant contact order.

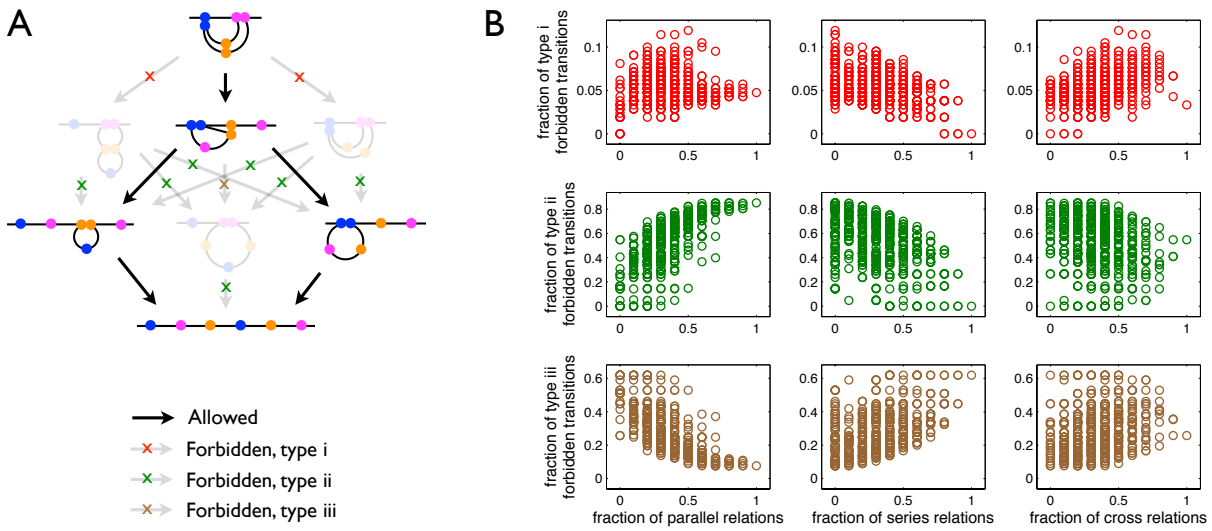


Figure S2: Topology affects all three types of forbidden unfolding transitions. (A) The three types are illustrated for an example chain with $C = N/2 = 3$ contacts: (i) the contact does not lie along the shortest path(s), (ii) the transition originates from an inaccessible state, or (iii) the transition involves breaking two contacts and reforming another. (B) The fractions of each type are plotted against the fractions of each topological binary relation for all 513 topologically unique chains with $C = N/2 = 5$.

4 Topology and intermediate folding states

In the main text, we show the topological composition of chains that exhibit certain numbers of intermediate states during unfolding (Fig. 4E of the main text). Here, we show results for folding. We find almost identical results (compare Fig. S4 to Fig. 4E of the main text). This demonstrates that the effects of topology on the presence of intermediate states is highly robust to the protocol that governs state-to-state transitions, persisting both for tension-based unfolding and stochastic folding.

5 Transition probabilities during folding and unfolding

When calculating the path-weighted degeneracy, path probabilities are set in the following way. For folding, in each step, the probability of a contact forming is inversely proportional to its formation time (stochastic folding). For unfolding, in each step, only contacts involved in the shortest path(s) between the chain ends are candidates to break. We set the break probability of each of these contacts to be proportional to the amount of force acting on the contact. The force is calculated mechanically, assuming equal tension on parallel chain segments when paths overlap.

Fig. S5 provides an example of how tension sets the probabilities for each contact to break during the unfolding process. In the top configuration in Fig. S5, the magenta contact must resist the entire pulling force F. The tension in each of the two parallel segments connecting the blue and magenta contacts is therefore $F/2$. The blue contact counteracts this tension only, such that this contact must resist only half the force, $F/2$. We therefore make the blue contact half as likely to break as the magenta contact, or $p = 1/3$ and $2/3$, respectively.

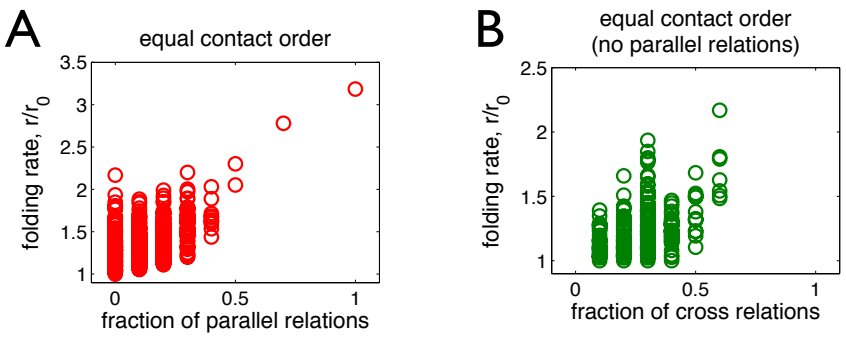


Figure S3: Topology affects the folding rate when folding is stochastic. (A) At constant contact order, folding rate increases with the fraction of parallel relations. (B) At constant contact order, for chains with no parallel relations, folding rate increases with the fraction of cross relations. Data are for 10^5 chains with $C = 5$ contacts in randomly sampled arrangements and N sufficiently large to accommodate the all-series arrangement, whose folding rate is r_0.

The above protocol is equivalent to making the break probability of each contact proportional to the number of shortest paths in which it is involved: here, the magenta contact is involved in both shortest paths, while the blue contact is involved in one (the bottom segment only).

6 Topology and energy

Because topology has a significant effect on folding and unfolding dynamics, as demonstrated in the main text, we advocate that it be considered alongside other standard predictors of folding and unfolding dynamics, such as energetic and geometric properties of linear chains. We illustrate this point in Fig. S6 with a simple example that considers both topology and energy.

In Fig. S6, two molecules are shown, one with two strong contacts in series (top) and one with two weak contacts in parallel (bottom). An observer only aware of their topologies would predict that the bottom molecule would take longer to unfold, because in the bottom molecule the contacts are in parallel, and thus the outer contact must break before the inner contact feels a force. In contrast, an observer only aware of their bond energies would predict that the top molecule would take longer to unfold because its bonds are stronger. Resolving these contradictory predictions requires a framework that takes both topology and energy into account. This framework would need to be developed and fine-tuned in conjunction with experiments.

Although the example in Fig. S6 is sufficiently simple to make the outlined effects obvious by inspection, it serves to illustrate that point that both topology and energy can have important effects on (un)folding dynamics, and the interplay between the two may be far more subtle in more complex chains.

7 Computational protocol

In the interest of clarity, as well as to facilitate extensions of our work, in Fig. S7 we summarize the computational protocol used to generate and analyze the data in Figs. 2, 3, and 4 in the main text.

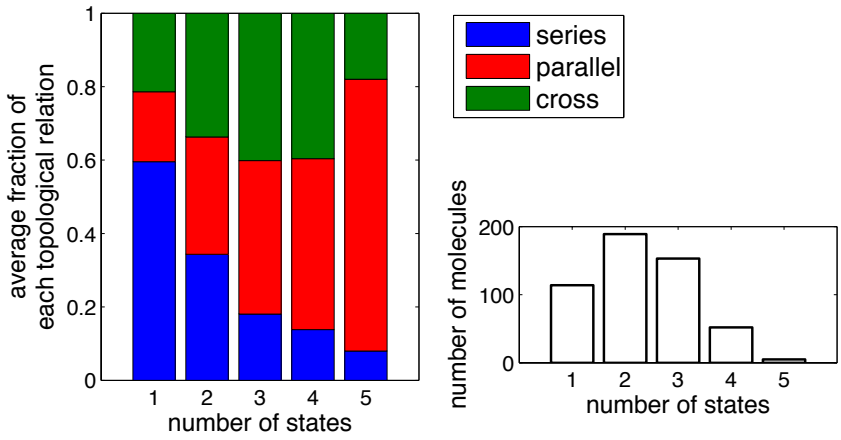


Figure S4: Topology influences the number of intermediate states visited during folding. For all chains with length $N = 10$ and $C = 5$ contacts that exhibit a given number of states, we plot the average fractions of each topological relation (left) and the number of such chains out of the total of 513 (right).

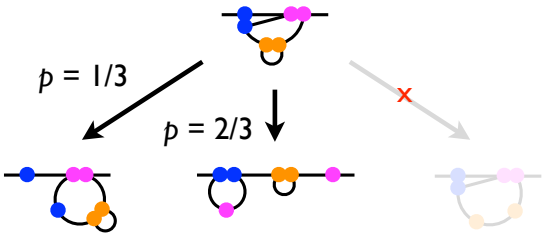
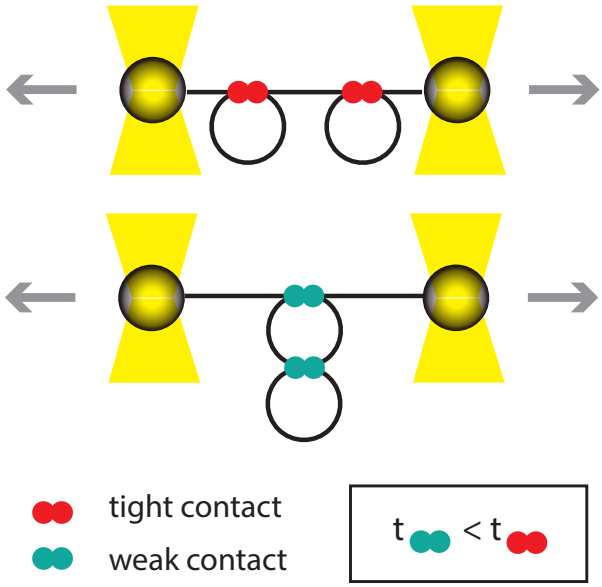


Figure S5: An example configuration and its unfolding probabilities.



Picture	Unfolding time
Topology	$t_S < t_P$
Energy	$t_P < t_S$

Figure S6: Simple example in which a prediction based on topology contrasts with a prediction based on bond energy, and thus the two frameworks must be considered in conjunction. Two molecules are pulled into unfolded states by tethering their ends to beads which are controlled using optical traps. The top molecule contains two strong contacts in series, while the the bottom molecule contains two weak contacts in parallel. An observer only aware of their topologies would predict that the bottom molecule would take longer to unfold, because in the bottom molecule the contacts are in parallel, and thus the outer contact must break before the inner contact even begins to feel a force. In contrast, an observer only aware of their bond energies would predict that the top molecule would take longer to unfold because its bonds are stronger and thus take longer to break.

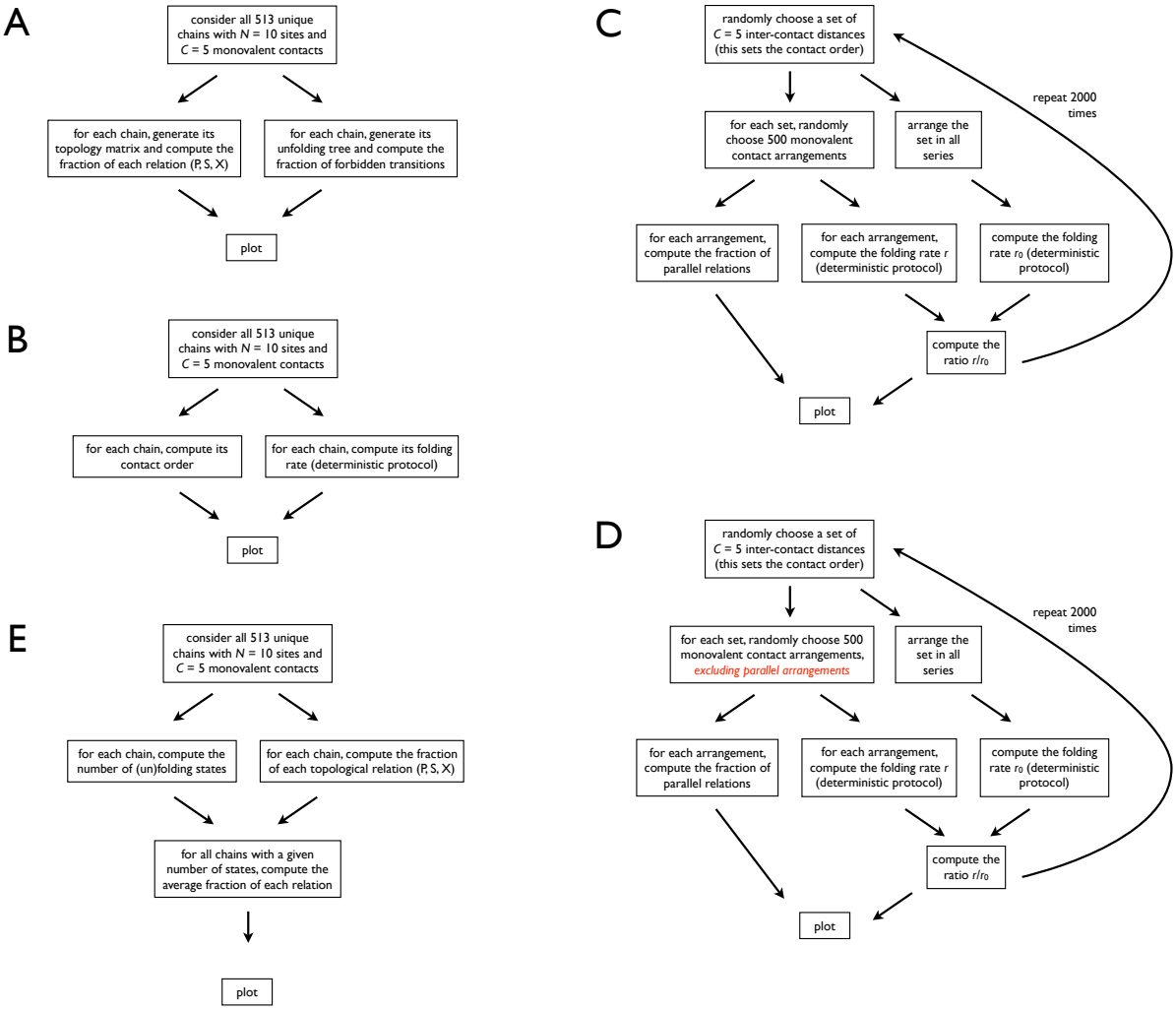


Figure S7: Flowcharts summarizing the computational protocol used to take and analyze the data for figures in the main text: (A) Fig. 2C, (B) Fig. 3B, (C) Fig. 3C, (D) Fig. 3D, (E) Fig. 4E and F.