Simplification of Topology by Type II Topoisomerases and Kinetic Proofreading model

Kalin Vetsigian

December 5, 2001

Topoisomerases are enzymes that have evolved to solve the topological problems associated with DNA manipulation in the cell, such as in transcription, replication, packing and unpacking of DNA. Since these enzymes are essential for the correct functioning of the cell they are important drug targets. People study topoisomerases from many different perspectives: protein structure, detailed atomic level mechanisms of how they work, single molecular experiments, effects of mutations on enzyme function, etc. However, the first question I have as a beginner is what topoisomerases do. I would also like to imagine at least one physically plausible mechanism capable to account for the observed functions. I will discuss a relatively recent experimental paper by Rybenkov et. al.(1997). [1], revealing unexpected abilities of various type II topoisomerases, and theoretical papers by Yan, et al. [2],[3] giving a possible explanation.

There are two types of topoisomerases (topos). Type I topos cut one of the two DNA strands, rotate it at 360 degrees around the other and reseal it. Type II topos cut both strands, pass another double stranded DNA segment through the gap and reseal the broken strands. Further very important distinction is between "passive" topoisomerases which simply catalyze passage of DNA strands through each other - helping the DNA molecules to relax toward topological thermal equilibrium, and "active" topoisomerases which fight to establish a different distribution of topological species. Topoisomerases that do not hydrolyze ATP necessarily belong to the "passive" class because otherwise they would play a role of Maxwell demons. From physical point of view these topos convert a real DNA chain into a phantom chain which can pass through itself. Gyrases (type II) use ATP hydrolysis to introduce negative supercoiling in DNA. Reverse gyrases (type I topos present in some thermophilic organisms) use ATP to positively supercoil. However, there are type II topoisomerases that use ATP but still do not supercoil. Instead they relax supercoiled DNA just as the "passive" topos do. Though there is no thermodynamic problem with that one, can ask why cells waste ATP for something that can be done passively (for ex. gyrases can relax supercoiling even in the absence of ATP)? The only available answer before Rybenkov, et al. paper (1997) was that ATP is used only to somehow accelerate the relaxation.

The finding of the paper is that some type II topoisomerases can drastically

reduce the probabilities for knots (up to 90 times) and catenanes (up to 16 times) relative to the thermal equilibrium values in circular DNA ensembles . In addition the equilibrium distribution of linking numbers in circular DNA, though not shifted (no supercoiling is induced) is narrowed about 1.4 times. (The exact numbers vary with the experimental conditions and topoisomerases being used.) These results are surprising because they imply that the small typo II topos (less than 10nm in diameter) are able to somehow sense the global topology of much larger DNA knots and catenanes (on length scales of thousands of bp). Random passage of DNA segments when they hit each other would just yield the thermal equilibrium. Due to their small size, topoisomerase must use local interactions with DNA to recognize DNA topology. However, local interactions provide only limited information while the observed effects look large. Another possible direction for explanation, not relying on the sufficiency of local interactions, is that DNA is effectively shortened due to cooperativity between several topos acting simultaneously- short DNA are less likely to form knots and catenanes, or deviate from the relaxed linking number. However cooperativity was ruled out by previous experiments (see references in [1]).

An interesting question is how we create " phantom chain" thermal equilibrium distributions in practice. The establishment of thermodynamic topological equilibrium distribution of linking numbers in a circular double stranded DNA is easy - we just act with "passive" type I topos that can relax both positive and negative supercoiling. (Notice that the fact that a topo does not use ATP does not by itself guarantee establishment of full topological thermal equilibrium from all initial conditions. For. example there are topos that relax only positive or negative supercoiling. But we believe that the topos used establish such equilibrium). The situation is more problematic for knots and catenanes because all type II topos are not passive, and besides they are subject to the experiment itself. So we need an independent way of obtaining equilibrium distributions. The method used involved cyclization of linear DNA with cohesive ends (Figure 1). If cyclization is sufficiently slow it is believed that it results in a topological equilibrium distribution. This is supported by Monte Carlo simulations, as well as agreement with linking number distributions generated as discussed above. In catenation experiments, P4 DNA (10kb long) was cyclized in the presence of excess amounts of pAB4 circular DNA (7 kb long), and the fraction of heterodimer catenanes was observed. For simplicity, both DNA-s were nicked to insure that they are not supercoiled. Knot experiments were performed with cyclization of both P4 DNA and pAB4 DNA. Because the DNA molecules used were short, almost all of the knots were of the simplest type - trefoils, and almost all of the catenanes singly linked. So in this experiments we can to a very good approximation consider only two topological states: knot - unknot, or unlink - single link. The concentrations of different topological species were resolved using a standard agarose gel electrophoresis technique.

Jan et al. noticed that the probability distributions are roughly squared in the presence of topo II. This was reminiscent to the squaring of error probabilities in proofreading models [4]. Correspondingly they proposed a mechanism that is a simple variation of the original proofreading scheme. However, one can

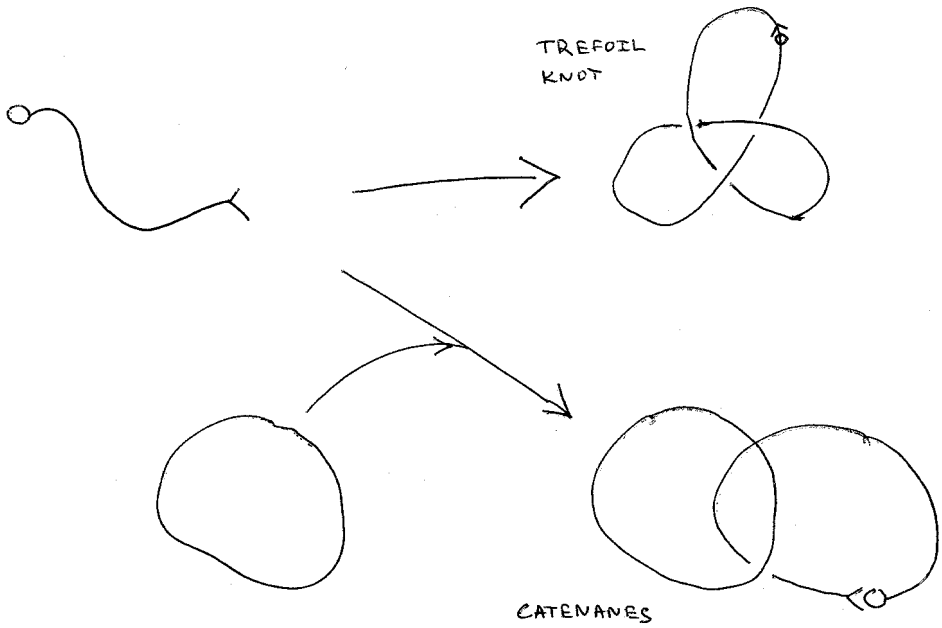

Figure 1:

understand the essence of their suggestion without reference to the proofreading model. The local property that the topos can use to sense topology is the random collision rate between DNA strands. It is clear that the collision rate in a knotted molecule will be somewhat higher than in an unknoted one. Same is true for catenanated molecules and supercoiled molecules (the higher the deviation from equilibrium linking number, the more supercoiled the molecule is, the closer the DNA strands are on average). Sensitivity to collision rates requires memory, and thus irreversibility coming from ATP hydrolysis. The minimal model is that two successive collisions between a topoisomerase bound segment and another segment must occur in a small time interval to achieve strand passage. More specific mechanism is shown on Figure 2. The topo sitting on a DNA (1) is activated after the first collision (2), then brought to a high energy state (1*) by an irreversible ATP hydrolysis. If another DNA strand hits the topo while it is in 1* state there is a fixed probability that this strand will pass to the other side. However, the high energy 1* state is unstable and decays with rate γ (the reverse rate is negligible due to the large free energy difference between 1* and 1). This gives an overall sensitivity to collision rates. It is reasonable to assume that k and v are proportional to the collisions rates in knotted and unknoted states. k' and v' are the corresponding rates when the topo is in 1* state. A reasonable assumption is $k/v = k'/v' =$ (collision rate when knotted)/(collision rate when unknotted).

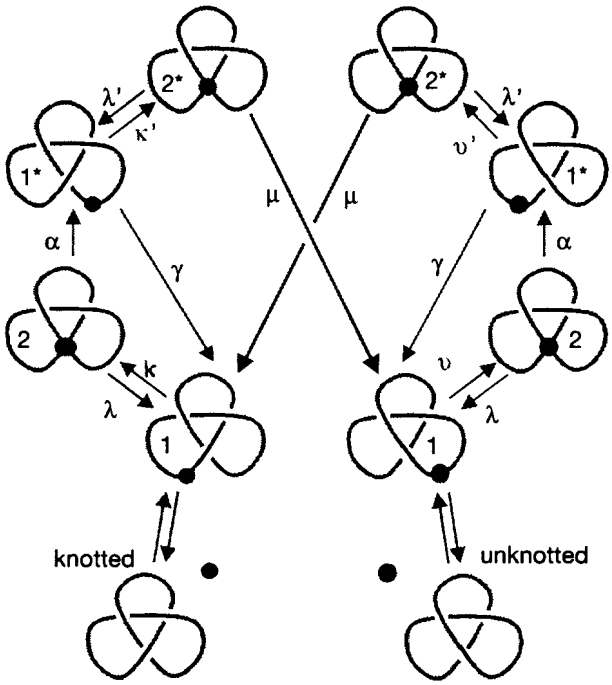

Figure 2: Proposed kinetic model for type II topos. The topo is represented by a filled circle.

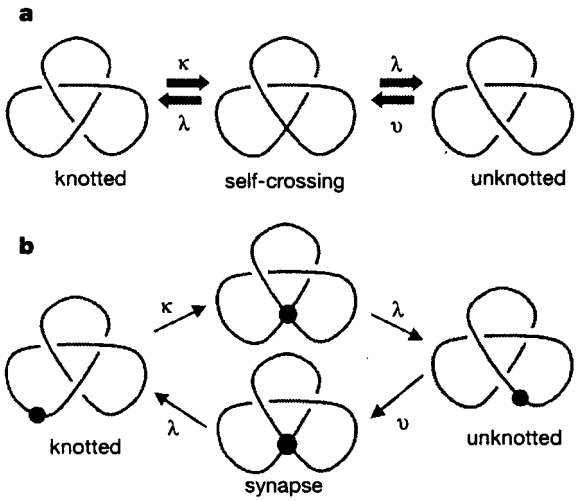

Figure 3: **a** "Phantom" DNA loop. **b** Simplest kinetic model of type II topos. The topo is represented by a filled circle.

Now we can look at a model that gives topological equilibrium such as the one presented on Figure 3. Simple calculation gives $P^{eq}_{knot}/P^{eq}_{unknot} = v/k$. The steady state for the diagram shown on Figure 2 is

$$\frac{P_{knot}}{P_{unknot}} = \frac{(\gamma[\lambda' + \mu] + k'\mu)vv'}{\gamma([\lambda' + \mu] + v'\mu)kk'} \ . \tag{1}$$

In the limit $\gamma[\lambda' + \mu] >> k'\mu$, $\gamma[\lambda' + \mu] >> v'\mu$, i.e. for sufficiently fast decay of the high energy topo state, we get an estimate with no adjustable parameters.

$$\frac{P_{knot}}{P_{unknot}} = (\frac{P^{eq}_{knot}}{P^{eq}_{unknot}})^2 \tag{2}$$

The above estimate implies that off-equilibrium steady state distributions generated by different topoisomerases are approximately the same. However this contradicts the experiments in [1]. E.coli IV reduces knots by 90 times while topos II from *Saccharomyces cerevisiae* reduces knots by only 5 times. Yan, et. al. do not mention this difficulty. Instead they relax the above assumptions to get $\frac{P_{knot}}{P_{unknot}} \leq (\frac{P^{eq}_{knot}}{P^{eq}_{unknot}})^2$ which seems consistent with the data. An analogous argument goes for catenanes.

In [3] the argument is extended to the linking number distributions. Since type II topos change the linking number by 2, odd and even linking number distributions achieve steady state independently (staring from the thermal equilibrium). Though in this case we have infinitely many states to consider, instead of two, the result is similar due to the fact that states of different linking numbers

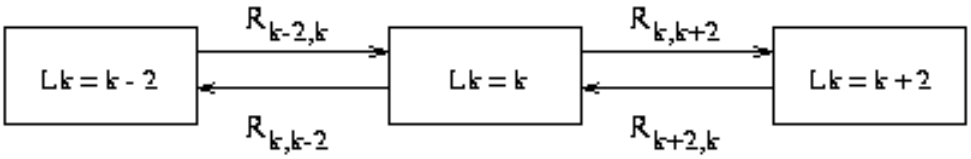

Figure 4:

are connected in a 1D chain (Figure 4). If $R_{k,k+2}$ is the overall rate for changing the linking number from k to $k+2$, then the condition that there is no net flow of probability along the chain gives

$$\frac{P_{k+2}}{P_k} = \frac{R_{k,k+2}}{R_{k+2,k}} \tag{3}$$

for a steady state. Under analogous assumptions for the rates we get $\frac{P_{k+2}}{P_k} = (\frac{P^{eq}_{k+2}}{P^{eq}_k})^2$ which upon separate normalization for odd and even states, and noticing that for the few kb DNA used $N_{even}/N_{odd} = 1.000001$ in thermal equilibrium we get:

$$P_k = A(P^{eq}_k)^2, \tag{4}$$

where A is determined from $\sum P_k = 1$, and there are no adjustable parameters. This prediction compares very well with the results for *E. coli* topoisomerase IV shown on Figure 5. Again, distributions created by other topos do not agree with the parameter free estimate. In all cases (knots, catenanes, linking number distribution) the agreement is good for the "most efficient" topo (*E. coli* IV) and becomes worse when we go to the "less efficient" ones. The authors indicate that this can be accounted for by including a probability that the topos are activated by thermal conformational fluctuations without being hit from a DNA strand. This omitted "leaky"effect is more important for the "less efficient" topos, and is characterized by a constant which depends on the topo. However, the authors do not perform any computations or comparison with experiment.

Since under reasonable assumptions for the rate constants the model from Figure 2 will have very few adjustable parameters after accounting for the "leaking", there are many tests that can be done by comparing it to both experimental data and data coming from simulations of polymer chain fluctuations. If the basic mechanism - sensitivity to collision rates - is correct, it might be possible to characterize each topo by just few constants which will open the road to theoretical predictions (using simulations) of what happens in complicated situations involving long DNA molecules. Of course, the topo parameters will depend on the properties of the solution, such as ion concentration etc. Then variations in ion concentrations, for example, will complicate quantitative predictions.

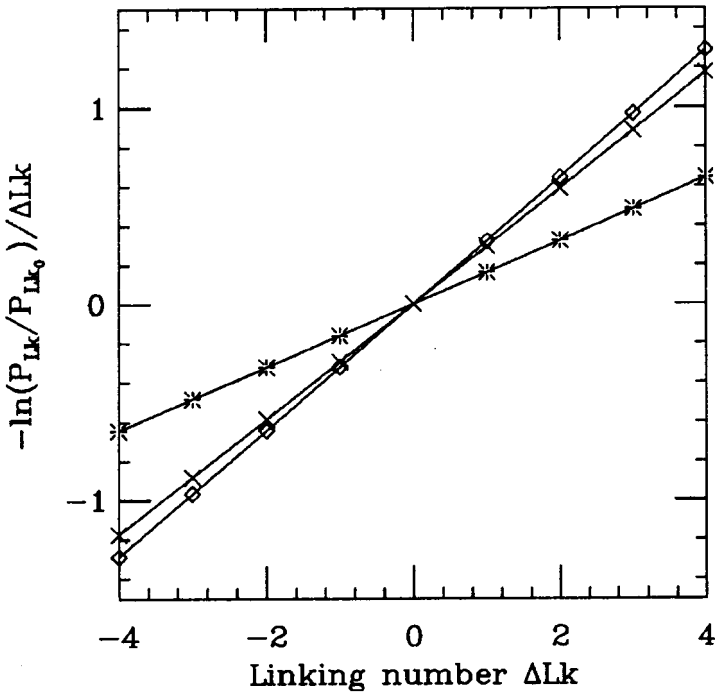

Figure 5: Comparison of the linking number distributions obtained in thermal equilibrium (stars), in the type II topoisomerase-driven steady state (crosses), and from the two-collision kinetic proofreading model (diamonds).

References

[1] Rybenkov, V.V., Ullspeger, C., Vologodskii, A. V., Cozzarelli, *Science* **277**, 690-693 (1997).

[2] Yan, J., Magnasco, M. O., Marko, J.F., *Nature* **401**, 932-935 (1999)

[3] Yan, J., Magnasco, M. O., Marko, J.F., *Phys. Rev. E* **63**, 031909 (2001)

[4] Hophield, J.J., *Proc. Natl. Acad. Sci. USA* **71**, 4135-4139 (1974)