Communicating Research to the General Public

At the March 5, 2010 UW-Madison Chemistry Department Colloquium, Prof. Bassam Z. Shakhashiri, the director of the Wisconsin Initiative for Science Literacy (WISL), encouraged all UW-Madison chemistry Ph.D. candidates to include a chapter in their Ph.D. thesis communicating their research to non-specialists. The goal is to explain the candidate's scholarly research and its significance to a wider audience that includes family members, friends, civic groups, newspaper reporters, program officers at appropriate funding agencies, state legislators, and members of the U.S. Congress.

Over 20 Ph.D. degree recipients have successfully completed their theses and included such a chapter.

WISL encourages the inclusion of such chapters in all Ph.D. theses everywhere through the cooperation of Ph.D. candidates and their mentors. WISL is now offering additional awards of $250 for UW-Madison chemistry Ph.D. candidates.

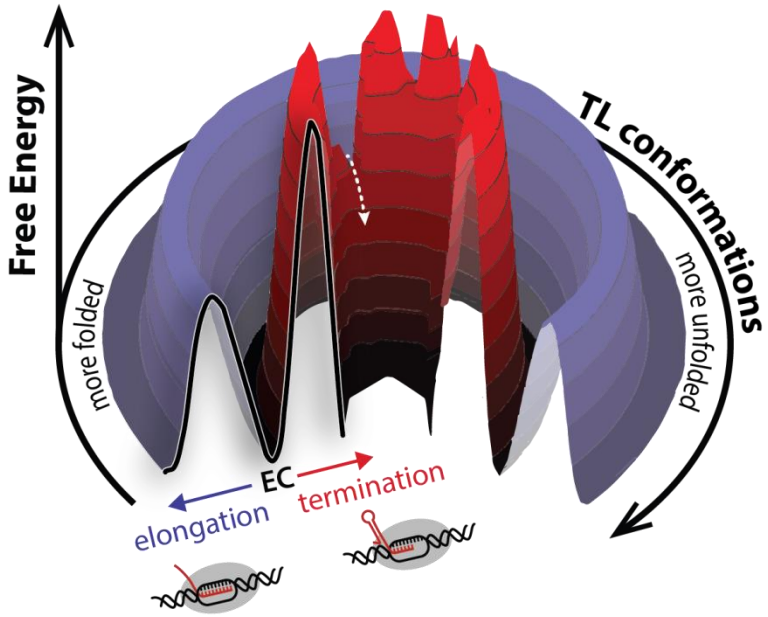
Wisconsin Initiative for Science Literacy

The dual mission of the Wisconsin Initiative for Science Literacy is to promote literacy in science, mathematics and technology among the general public and to attract future generations to careers in research, teaching and public service.

UW-Madison Department of Chemistry
1101 University Avenue
Madison, WI 53706-1396
Contact: Prof. Bassam Z. Shakhashiri
bassam@chem.wisc.edu
www.scifun.org

May 2014

Trigger loop dynamics aid intrinsic transcription termination by *Escherichia coli* RNA polymerase



Ananya Ray-Soni

In the laboratory of Professor Robert C. Landick

University of Wisconsin – Madison

Madison, WI

Have you ever considered how we get different organs and body parts if they *all contain the same DNA*?

In forensic science (think to your arsenal of knowledge from crime shows like CSI and NCIS), crime scenes are always combed for hairs or dry skin particles that may have been left by the perpetrator. From these hair or skin particles, forensic scientists can isolate DNA, which encodes the unique genetic information that makes a person who they are. In essence, it provides a DNA "fingerprint" of anyone that was at the crime scene. Once the police have a suspect, they can then take a cheek swab, isolate the DNA from this cheek swab, and compare it to the DNA from the hair/skin particles they found on the scene. If the two DNA samples match, it is likely that their suspect was in fact at the scene of the crime.

…But wait! Why would the DNA from such different parts of the body with completely different appearances, textures, and functions be comparable? It turns out they are not just comparable; they are in fact the *same*! Think about this for a second – we know that our DNA **genome** encodes the genes required for each biological "machine" to function. So if all of the different parts of a human body that are so different from each other contain the *same DNA*, how do they look and behave so differently? This is because even if all parts (and the **cells** that make up those parts) have the same genes and DNA, only some of the genes are "turned on" in each part. In other words, although all of the information needed for us to grow and survive is encoded in our DNA, our cells can only utilize this information if the necessary gene is turned on and is used to produce the biological "machine" that it encodes. And which machines are being produced in a particular cell determines its **cell-type**. By controlling which sets of genes are "on" or "off" in each cell, the same DNA can be used to produce all of the diverse cell-types that make up a human body.

Sidebar: What are **cells** and **cell-types**?

Cells (often called the "building blocks of life") are the smallest structural and functional units of all living organisms that are *self-sustaining* and have the potential to self-replicate. What do I mean by self-sustaining? All cells have the ability to produce the energy and biological resources to be able to live and grow without support from other biological structures. In fact the simplest organisms are single-cellular organisms (*e.g.* **bacteria**, see sidebar below), which can grow and replicate completely independently. Cells may have various different characteristics depending on the genes encoded in their DNA, as well as their **gene expression patterns** (*i.e.* which genes are "on" or "off").

In multi-cellular organisms (sometimes called "higher organisms," *e.g.* humans) all cells within the organism share the same DNA. However, these cells vary in their gene expression patterns, resulting in cells with distinct characteristics. We call cells with the same characteristics the same **cell-type**. Cells of the same or related cell-type often adhere together to form **tissues** (*e.g.* muscle tissue is constituted of muscle cells). And tissues of different types that work together to carry out one larger biological function (*e.g.* pumping blood to the body) form organs (*e.g.* a heart).

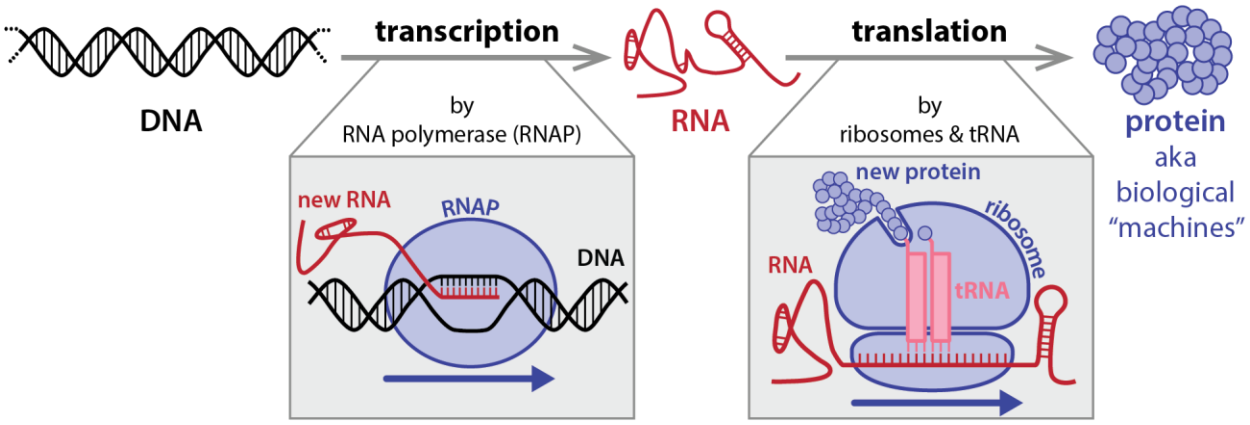
cells (*e.g.* nerve cells) → tissues (*e.g.* nerve tissue) → organs (*e.g.* brain)

How are some genes specifically "turned on", while others are "off"?

To think about how genes are turned on and off, we must first understand how the corresponding biological machines are made. In biology, the **central dogma** describes the process by which a gene is turned on and used to create the biological machine it encodes (**Figure 1**). In a process called **transcription**, the biological machine called **RNA polymerase (RNAP)** "reads" the information in the DNA gene and converts it to RNA. The RNA acts as a sort of temporary message, from which protein (*a.k.a.* the biological machines

we've been talking about) can be made in a process called **translation**. Once these proteins are made, they can carry out their functions in the cell.

Since the RNA needs to first be produced for the protein to be made from it, the process of transcription is the first step at which a gene can be turned on or off. Different environmental factors and cellular signals help ensure that RNAP transcribes the correct genes for each cell-type and condition. *I.e.*, we say that the process is **regulated** by different factors. In this work, I was studying how RNAP facilitates the process of transcription.



<u>Figure 1:</u> The central dogma of gene expression: how RNA and proteins are made from DNA.

RNA is produced from DNA by RNA polymerase (RNAP) in a process called transcription. Proteins are then produced from the RNA transcript by the ribosome and tRNAs in a process called translation. Both RNAP and ribosomes are also proteins, while tRNA is actually a functional RNA molecule (see also **Figure 3**)!

Why study transcription?

As I alluded to in the previous section, transcription is the first step at which gene expression (we say a gene is expressed when the gene is "on") can be regulated; translation of RNAs to proteins can act as an on/off switch too! However, both transcription and translation use valuable cellular resources to produce RNAs and proteins, and transcription of a gene *must* occur before translation. So if a gene needs to be off in certain conditions or cell-types, transcription is the less resource-"expensive" step to act as an on/off switch, since it is the first step of gene expression.

Transcription can largely be divided into 3 stages (**Figure 2**):

(1) **initiation** of transcription: in which RNAP binds to the DNA at the start of a gene that is turned on and begins **transcribing** the RNA **transcript** from the gene;

(2) **elongation** of the new RNA transcript: in which RNAP decodes the entire message that is encoded in the DNA gene and converts it to RNA;

(3) **termination** of transcription: which causes the transcribing **elongation complex (EC)** to be dissociated into its components – the new RNA transcript, the DNA genome, and RNAP.

In this work I was interested in understanding how RNAP carries out the last step of transcription – termination.

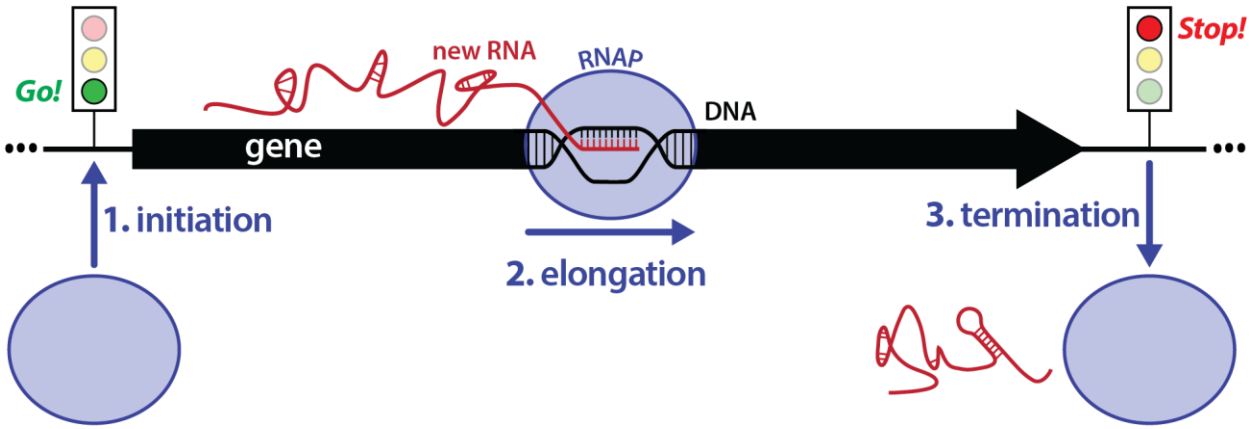


Figure 2: **The 3 stages of transcription.**

So initiation turns the genes on; why does it matter whether you terminate the process or not?

While initiation is very important for turning genes *on*, termination is essential for turning them *off* (**Figure 3**). It is also important for recycling RNAP for new initiation events, releasing the RNA, and it prevents collisions between RNAP and other proteins that travel along the DNA, as this could damage the genome.

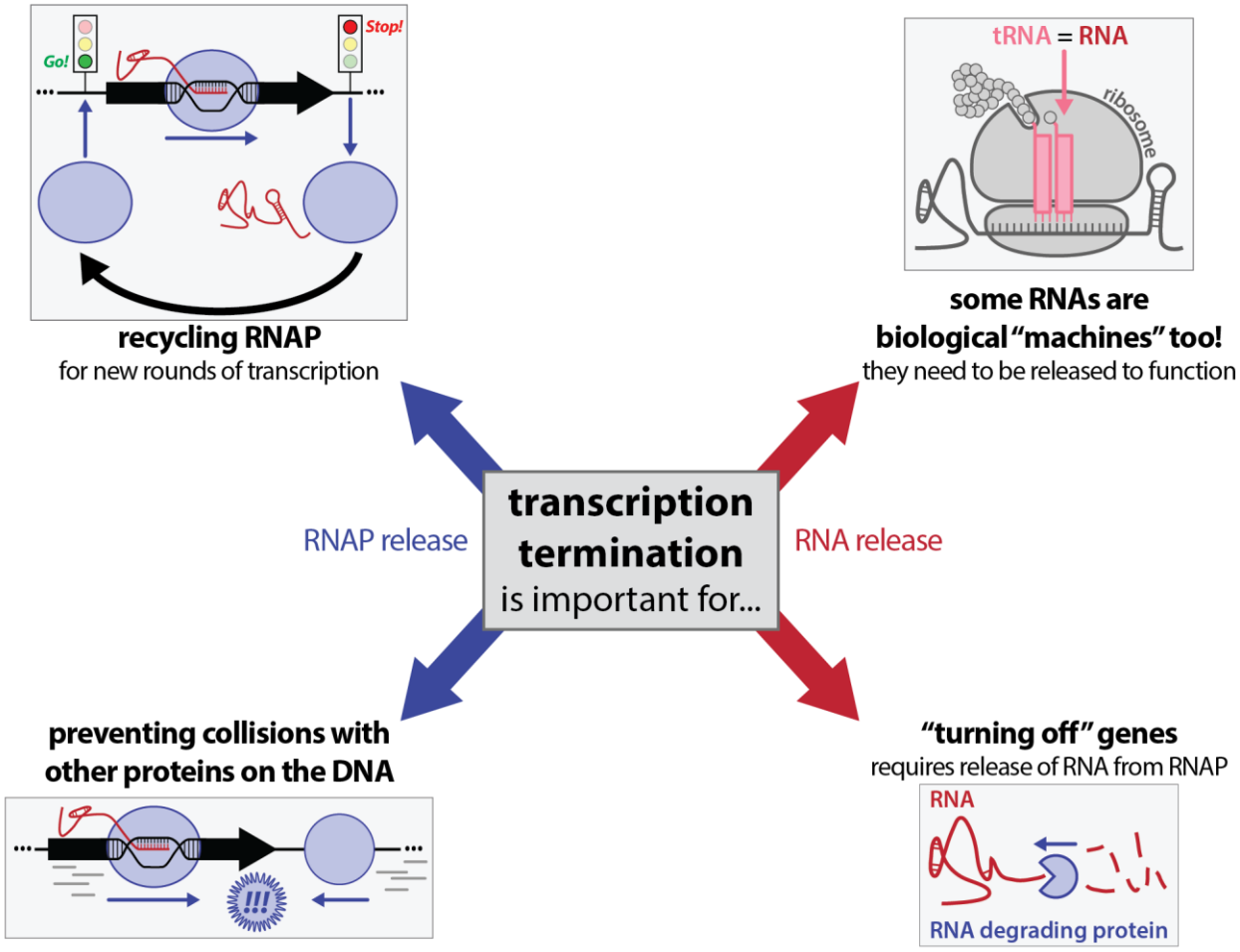


Figure 3: The importance of transcription termination is manifold.

Studying termination in bacteria

I am studying transcription termination in **bacteria**. Bacteria are diverse organisms present in and around everything we encounter, and understanding how they function has great potential for many different applications (see the sidebar below for discussion on the significance of bacteria)! In addition, the efficient minimal biological systems used by bacteria often provide a great way to understand how more complex versions of the analogous system may function in humans and other higher organisms.

Sidebar: Why are **bacteria** so important?

Bacteria are microscopic single-cell organisms that pre-date human existence on Earth. This means that they are able to carry out all of the functions needed to live and grow in just one cell! The elegant genetic makeup of these diverse organisms also allow them to exist in various temperate and extreme environments including the soil, hot springs, extreme cold or acidic environments.

We often think of bacteria as harmful, pathogenic organisms that make us sick. But this isn't true of *most* bacteria. While there are some bacteria that are parasitic, and are harmful to whatever host organisms they infect, the vast majority of bacteria is harmless and coexists with us peacefully. Did you know that every surface that we come in contact with is covered with bacteria? On average, one teaspoon of water contains ~5 million bacteria! In fact, the human body may be home to more bacterial cells than human cells! But don't be alarmed – most of these bacteria are completely harmless, and some of them actually *help us survive*. For instance, most bacterial cells associated with the human body inhabit our gut and help us unlock vital nutrients from our food.

Scientists in many industries have also found ways to utilize the unique abilities of different bacteria, *e.g.* for producing cheese and yogurt, cleaning up oil spills, mining metals, and producing fuel from food waste. And this is just the tip of the iceberg! As these examples hopefully illustrate, there is a lot of value and utility to understanding bacterial function, and likely a great deal of untapped potential to be uncovered from continued bacterial research!

What do we know about bacterial termination so far?

In bacteria, **intrinsic termination** is one of the 2 major ways by which transcription is terminated, and it is signaled by the formation of a structure in the new RNA transcript that is being made by RNAP. Interestingly, *not all ECs that encounter the RNA termination signal actually terminate*. The EC must make a *decision* between continuing elongation, and terminating. This decision involves a choice between 2 paths that require more or less energy to cross. Think of the EC as yourself riding a bike to either a friend's house or home after school. Assume your only motivation to go home or to your friend's house is based on the difficulty of crossing the path to get there. Under this assumption, if the hill to your friend's place is much steeper than the hill on the road home, you need more energy to go to your friend's place, and you are more likely to choose to go home (**Figure 4**).

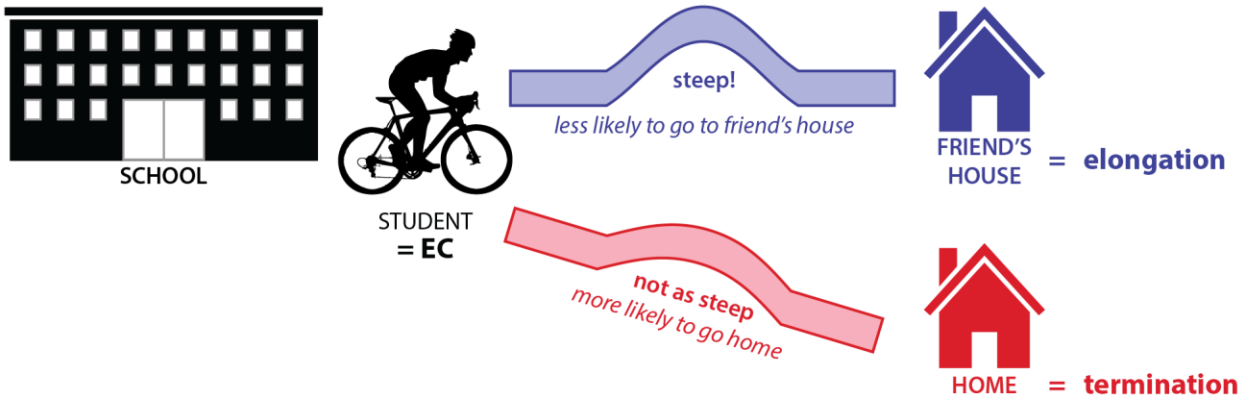
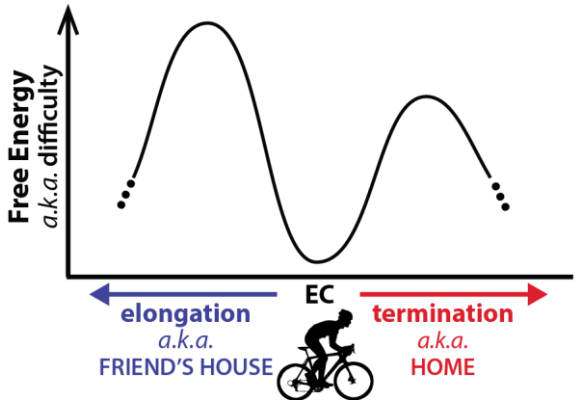


Figure 4: The difficulty of crossing hills determines which path you will take.

In this analogy, the student riding a bike represents the transcription elongation complex (EC). Here we assume that the probability that the student will go HOME (*i.e.* terminate) versus to his/her FRIEND'S HOUSE (*i.e.* elongate) is determined by the difficulty of crossing the hill on the path to either destination. *Biking cartoon adapted from http://www.veloscience.org/?p=4403.*

Similarly, the relative probability of carrying out 2 different **reactions** (*i.e.* crossing 2 different hills in our analogy) can be depicted as an **energy diagram** (**Figure 5**)**.** When an EC encounters 2 different **energy barriers** (hills) to elongation (your friend's house) versus termination (home), a higher energy barrier is harder to cross, making it the less probable choice (**Figure 5**).



<u>Figure 5:</u> An energy diagram depicts the relative energy barriers to the competing reactions of termination (right) versus elongation (left).

The height of the peak for either reaction indicates the amount of energy required to complete the reaction, where a higher peak requires more energy, and is therefore less probable. At termination sites, the energy barrier to termination is lower than the barrier to elongation, resulting in appreciable amounts of termination.

To complicate matters slightly, termination is a process with many steps, so the EC has to make the decision between elongation and termination many times before it finally successfully terminates. By our analogy, you must cross a few hills on your way home, and before each hill, you have the choice to cross a different hill and go to your friend's house instead (**Figure 6A**). Right now, the way termination is studied, researchers most commonly look at what percent of ECs terminated by the end of the experiment (called **termination efficiency, TE; Figure 6B, top**), *i.e.* they determine how many kids ultimately went home at the end of

the day. But with so many steps and decision points between school and home, a lot of information is lost about each of those decisions. I wanted to devise a method to "see" how many ECs choose to go over the hill towards home or the one towards their friend's house, *at every decision point*.

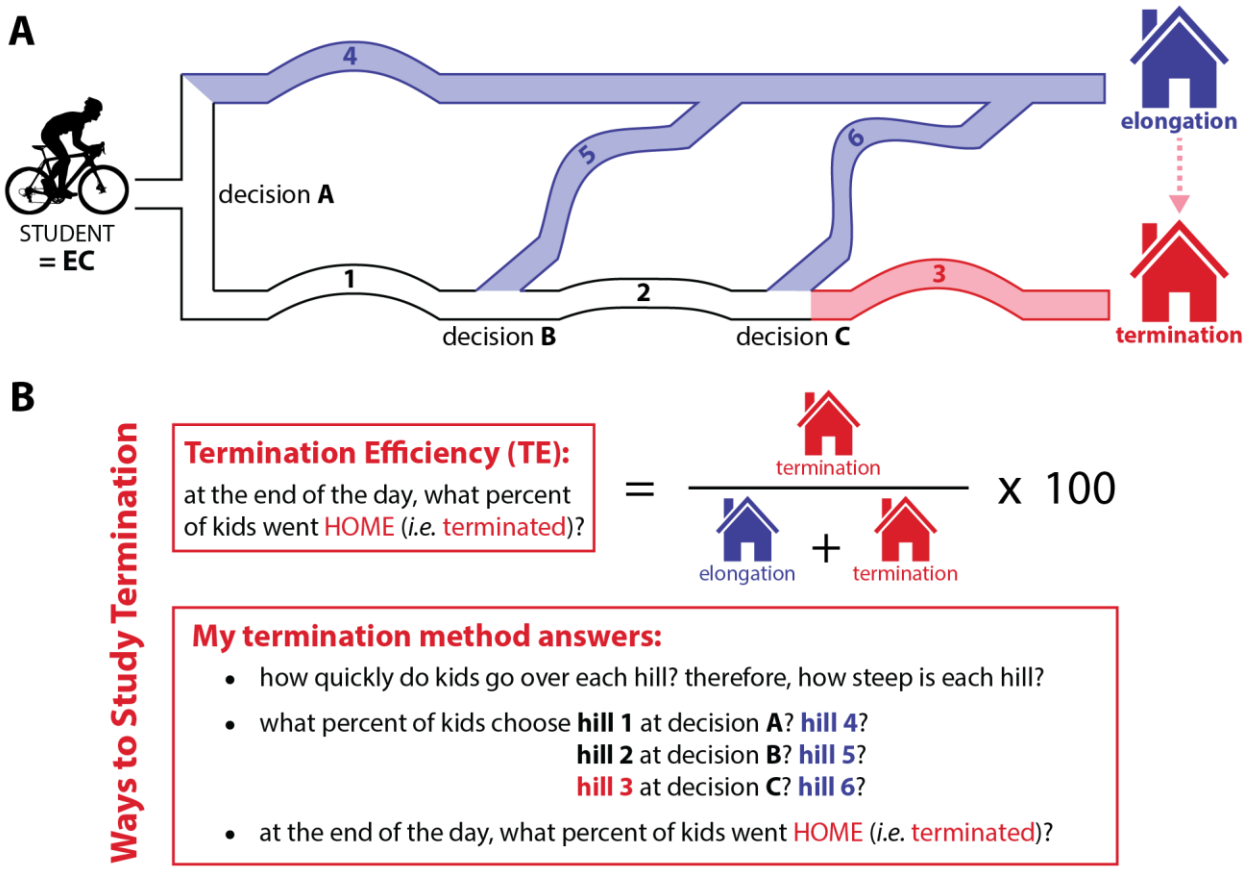


Figure 6: There are many steps and decision points on the path to termination.

(A) The multiple steps to termination of varying difficulty are depicted as numbered hills. Higher hills indicate greater difficulty, and thus lower probability of crossing that hill. The EC must cross **hills 1, 2 and 3** to successfully terminate. Paths that result in elongation and termination are *shaded blue* and *red,* respectively. The *pink dashed arrow* indicates that while not all ECs will terminate at the given termination site, they will eventually terminate at a later site on the DNA. **(B)** Termination efficiency (top) is the most common and easy-to-use method of studying termination. This method has been used to determine a great deal of information about the RNA termination signal. The method I developed in my graduate work (bottom), however, yields a lot more information about each of the steps of termination, and is thus a powerful new way to study the termination process. See explanation in the section below.

Problem solving time! How do I watch ALL the decision points?

When you are riding your bike over a hill, the height and steepness of the hill will not only affect your motivation to take that path, but also the speed at which you cross this hill. If the hill is higher, it will take you longer to cross the hill. Analogously, a population of ECs cross a higher energy barrier at a slower speed (or **rate**) than a lower energy barrier. Why? A higher energy barrier means that the reaction requires more energy to overcome the barrier, and is therefore less likely; a reaction that requires less energy is more likely. Another way of saying this is: if the reaction requires less energy, more ECs are likely to obtain the necessary energy to overcome the barrier. This results in a faster reaction rate, which we can measure in our experiments. Then, using the information about the rate at which ECs cross different energy barriers, I can figure out how high the barrier was. BUT, as we saw in **Figure 6A**, the termination process is complicated, with many hills and decision points.

When we encounter a complex problem in arithmetic, we break it down in to simpler parts, right? For *e.g.*, if I ask you to calculate: 22 + (8 x 4), PEMDAS (the order of operations) tells you to do each function one at a time. You would first compute the part in parentheses (8 x 4 = 32), simplifying the problem to: 22 + 32. Then the problem is simply addition, and you can easily determine that the solution is 54.

Similarly, to simplify the problem of too many different paths home and to the friend's house, I first measured the reaction rates with the road leading home (to termination) blocked. How did I do this? Recall how I told you that there was an RNA signal that causes termination. Well I changed the signal just enough, so it doesn't cause termination anymore. In our metaphor, I effectively made one of the hills leading home (**hill 2** in **Figure 6A**) so high that it was impossible for anyone to go over them. With **hill 2** blocked, no ECs could cross **hills 2, 3 or 6**. In this way I could simplify the problem and determine the reaction rates for crossing only **hills 1, 4 and 5** first.

With half of the reaction rates determined, I next wanted to find out how quickly ECs could cross the remaining hills on the path home. To do this, I gave the ECs the full termination signal, which made **hill 2** "crossable", and followed how quickly ECs terminated (went home) or elongated (went to a friend's house). With half of the information already in place – the reaction rates for crossing **hills 1, 4 and 5** – I only had to figure out the other half of the information from this experiment. This was MUCH more doable than trying to extract all the information from one experiment!

Using this method, I could glean a lot more information about the termination process, and each of the steps along the way, including the overall termination efficiency (**Figure 6B, bottom**). I was able to determine how quickly ECs crossed all the hills to either the friend's place or home (*i.e.* the reaction rates for crossing each of the energy barriers). By extension I could also calculate the relative heights of each of those hills (or energy barriers). And since we know that an EC is twice as likely to go over a hill that is half the height of another, I could also figure out how likely it was that an EC would take the path towards home at each decision point. Thus, the method I developed helped me identify which steps in the termination process are faster and slower, and which decision points are the most critical in determining if an EC ultimately terminates.

Cool, so I have a new method! …what else can I use it for?

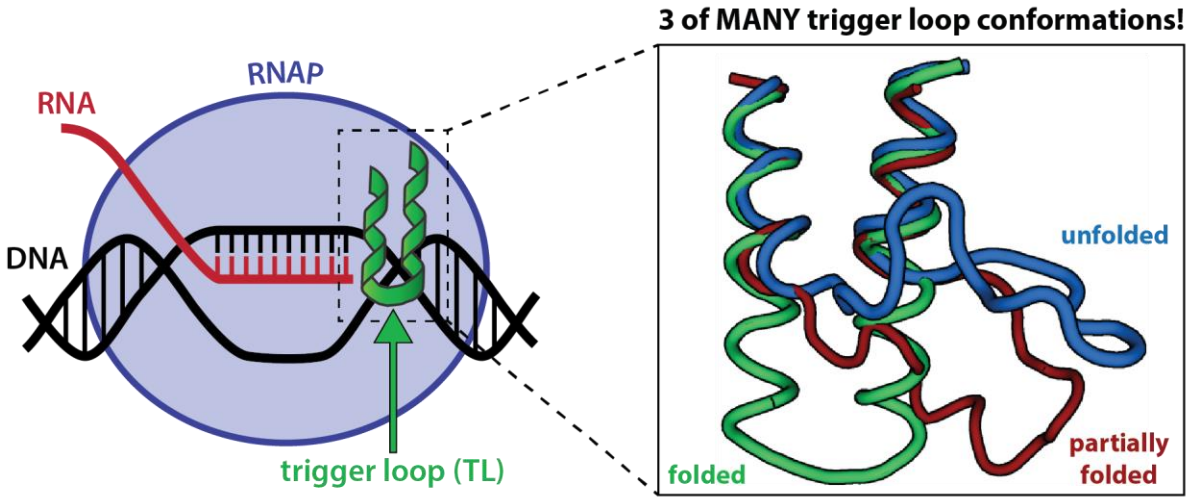
As I briefly described above, intrinsic termination is instigated by a sequence signature in the new RNA transcript, which forms an RNA structure within the EC. Through many carefully designed experiments, scientists before me have figured a lot out about how specific elements of this RNA sequence help signal termination of the EC, and favor the termination decision. However, the way that RNAP responds to this RNA termination signal and enables termination remained almost completely unknown. I wanted to use my method to start to understand how movements of RNAP facilitate termination.

Proteins are highly mobile structures that change shape (or **conformation**) to carry out their functions as biological machines. You can imagine that any machine has many moving parts (like levers and gears), and it's important to understand how these moving parts facilitate the main overall function(s) of the machine. Similarly the specific movements of different elements (or **modules**) of the protein structure – "cogs" in the machine – together facilitate the function of the protein. I was interested in determining the function and conformational changes of one such module of RNAP: the trigger loop (TL; **Figure 7**). The TL is known to be highly structurally dynamic, and requires this flexibility to enable RNAP to carry out many of its functions during transcription. I wanted to know whether the TL is important for any of the steps and decision points in the termination process.

How do we find out if parts of a molecular machine are important? In biochemistry, we either remove the module entirely (called deleting it), or change it in a specific, predictable way (called mutating it). We made RNAPs with such a TL deletion or TL mutations that behaved in ways that we had already determined by other experiments. I assembled ECs using these TL-deleted or TL-mutated RNAPs, and tested their abilities to terminate and elongate using my method. With this information I could determine:

(1) how the TL impacts the ability of the EC to cross each hill on the paths to termination and elongation, and

(2) how likely these ECs were to move towards termination at each decision point.

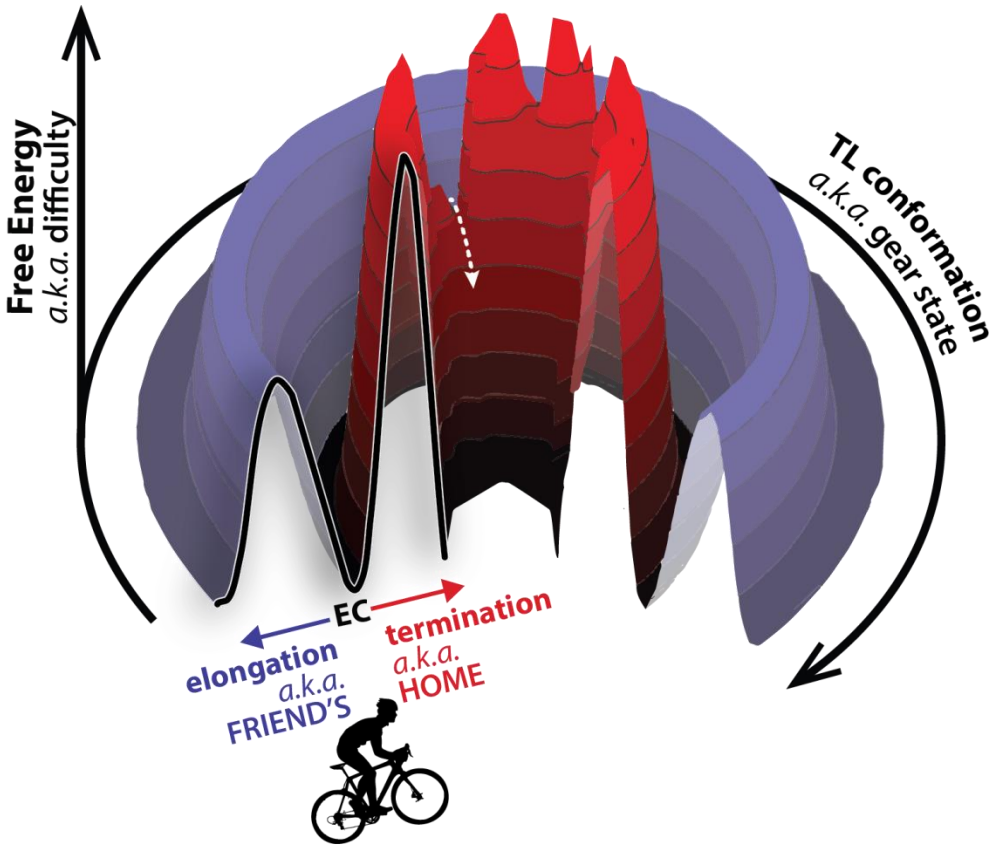


<u>Figure 7:</u> The structurally dynamic trigger loop (TL) module of RNA polymerase (RNAP) must adopt diverse conformations to enable the different functions of RNAP during transcription.

What did I find out? Was the TL involved in termination? Did it help send ECs down the termination path?

Spoiler alert: yes it did! From our experiments, we found out that the TL was in fact very important for a number of the decision points on the path to termination. I found that in the absence of the TL (*i.e.* with TL-deleted RNAP), the ability of the ECs to complete the final step of termination (**hill 3** in **Figure 6A**) was *severely decreased*! Strikingly, this impairment of TL-deleted ECs in termination was not detectable from measurements of termination efficiency. The termination method I developed was thus instrumental in identifying this important role of the TL in termination.

My results also indicated that mobility of the TL*, i.e.* its ability to take on many shapes or conformations, was very important for its function in termination. In our analogy, think of the bike you are riding as a multi-gear bike. The same hill may seem more or less daunting depending on the "conformation" of the gears, *i.e.* the gear-setting. When the bike is set to the ideal gear-setting, it seems easier to cross the hill. Conversely, in other gear-settings (too high or too low settings alike) it may be *extremely* tiring to cross the hill, making it seem like an insurmountable obstacle! In this scenario, the TL is the gear-shifter, allowing you to change the gear (or conformation) of the bike (EC). Now if the gear-shifter is absent from your bike, and you aren't able to switch gears, you are stuck to one gear-setting. Crossing the hill will always seem to be of the same level of difficulty. By having the gear-shifter, you have the option to try different gears, and find the setting that makes crossing the hill the least difficult. Analogously, the TL (gear-shifter) thus enables the EC (you, riding a bike) to sample various conformations (gear-settings) and find conformations that have the lowest energy barrier on the path to termination (gear- settings that make crossing the hill seem the least difficult; **Figure 8, white dashed arrow**).



<u>Figure 8:</u> Flexibility of the TL allows it to explore conformations for which the energy barrier to termination is relatively low.

A hypothetical 3D energy diagram (related to the energy diagram in **Figure 5**) depicts the relative energy barriers (vertical axis) to termination (*red*) versus elongation (*blue*) for different TL conformations (circular axis). Our results suggest that when the TL is flexible, it enables the EC to adopt states for which the energy barrier to termination is relatively low (depicted by the *white dashed arrow*). However, when TL dynamics are limited (in the TL-deleted or TL-mutated ECs), the EC is likely trapped in a state for which the energy barrier to termination is much higher, making it much less feasible. NOTE: The energy barrier to elongation is shown as a single height in this depiction; however, it likely also changes as a function of TL conformation.

Can this method be used by anyone else? What can they use it for?

Absolutely! In science, we publish our work in peer-reviewed journals that are publicly available. This means that we write up all of our results in a story, explaining exactly how we did each experiment, and submit the story to a journal that we think will get an interested audience. Then other experts studying similar problems will read our work and make sure all of the logic and experiment design checks out. If they approve it, our work gets published in the journal, and it's available for anyone to read!

The way we designed our assay, it's easily adaptable for asking other questions about the different steps in termination. Just like the TL, other cogs in the RNAP machine undoubtedly play a role in the process of termination. The key is to design smart and thoughtful experiments and RNAP mutants to address those questions. Someone could designs mutations in other parts of RNAP, or the termination signal itself, and determine the effects at each decision point. With more and more of these kinds of experiments, we can have a much clearer picture of how all the cogs in the RNAP machine must work together to facilitate a complete successful journey home.

Are there any human health-related implications of my work?

While the bacteria provide a model system for understanding the fundamentals of human biology, there are some key differences between bacterial RNAPs and human RNAPs that can help us target bacterial RNAPs with antibiotics. As you may be aware, antimicrobial resistance is a growing and serious problem worldwide. The Centers for Disease Control and the World Health Organization (WHO) have called for bacteriologists to identify more new antibiotics to help combat this problem.

Specifically, the WHO recently identified 12 bacterial organisms that are extremely dangerous and resistant to all or most available drugs. Some of these 12 pathogens contain a specific element within their TL domains that is absent in human RNAPs. My work as well as others' indicates that this element is very important for the activities of the TL, and RNAP, and could therefore serve as a potent and specific drug target. Moreover, the way to kill pathogens like these is to target their essential processes – processes that they need to have functioning for survival. Transcription termination is just one such process that is essential to bacteria; there are many other processes that need to be better understood so drugs can be designed against them. There are many, many more problems to be solved. Will you join the workforce?

Acknowledgements:

I thank Beth Boudreau for helpful discussion during preparation of the manuscript and development of the metaphor. This work was supported by funding from the National Institutes of Health (grant number GM38660 to Robert Landick), and the University of Wisconsin–Madison Biochemistry Department.