

A Primer on RNA Folding with Respect to the Ensemble and an Introduction to the Theory of the Mechanics of RNA

Author: Jennifer Pearl

Email: jenpearl5@gmail.com

Date: 5/3/2023

Rev: Rev 2.1

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Chapter 1

The ensemble and how RNA takes shape.

I have wanted to write this for a very long time, but the thought of writing a document or a book on how I understand RNA to behave has felt daunting to me. However at this point in time after now being published as a contributor and to have my work cited in reference to my research in RNA design methods in a peer review journal and seeing my many theories proven right through *in silico* and *in vivo* testing I feel it is time to put my thoughts into words.

To start with, I fully believe that RNA should be thought of as a mechanical system that lives in an energy world. It is thus bound by the rules of mechanics more than if not the same as the effects of the energies in the loops. What do you think the energies are actually DOING? I capitalize DOING because I want to emphasize that I am talking about the action of RNA. All RNA has action whether it be static or a riboswitch. When a switch is made there are shifts in energies and the hydrogen atoms that bind the nucleotides together are pulled apart and pushed back together somewhere else that they have a strong attraction too. The shape it eventually settles into is what is called the MFE or minimum free energy state and it is the one that a RNA design likes to hold most of the time. Now read that most of the time as “usually kinda but not really when you think of it”. The reason for this answer is because the MFE structure is only one data point in what is called the ensemble.

The ensemble consists of all the possible pairing configurations of the nucleotides in the primary structure and thus it consists of all the secondary structures that the single primary structure can take. This may sound a bit confusing at first if you are not experienced with some of these terms. Generally people don't learn about the ensemble until very late in their education. I think this does a disservice and actually ingrains a misunderstanding of RNA. RNA operates on probabilities and not on hard rules. It's very much akin to how scientists look at particles in a particle accelerator. You have to read the tea leaves so to speak. The problem is then that until you learn about the ensemble and how it works you are taught that RNA only does this one thing or that one thing. The ensemble is the RNA structure and not just a list of shapes it can take. Let's learn how RNA works so that we can better understand this concept.

For a moment let's take a step back and make sure we understand how RNA forms before continuing with the discussion about MFE. First, the primary structure is the technical term for the RNA sequence itself in letter form. RNA consists of 4 types of nucleotides, which are Adenine (A), Guanine (B), Uracil (U) and Cytosine (C). These are arranged together to form a

RNA strand such as AAUAGUACACUUUUUUACG. It is the attraction and repulsion of these nucleotides that causes the formation of a structure, and it is the 2D representation of the structure that is called the secondary structure. The 3D representation of the structure is called the tertiary structure, but it is not modeled very often. We generally deal with secondary structures while doing analysis as it is a good way to represent the RNA attractions in a straightforward manner, and we will be using them for this journey we are taking. Now, you're probably thinking, isn't it straight forward then that a specific shape would occur based on the sequence? Well, not exactly but sorta, as the nucleotides are not bound by absolutes... they are instead bound by probabilities. Much in the same way that quantum mechanics deals with probabilities in positions of components of molecules, so does RNA deal with arrangement and selection of secondary structures and this goes back to what I said at the start of the book in the first paragraph. We know where they are generally supposed to be, but we will not be right 100% of the time and we can not be. There are so many variables as well, just the changing of the temperature a fraction of a percentage can shift the energy equilibrium of a RNA molecule so that it then changes into a completely unrecognizable shape from before.

So we can take this concept of a structure, its nucleotides, and the many possibilities from all the variables and we call it the ensemble as discussed before. Now the structure that the RNA likes to take the most and requires the least amount of energy to maintain is called the "minimum free energy" (MFE) structure and all other structures in the ensemble are generally discussed in relation to that. The MFE is also by my definition the lowest energy state the RNA can take as all things move toward entropy in mechanics and physics, but the molecule itself is fighting the draw to entropy and is able to use what "energy it produces" to fight this draw. It is at this point that we have an equilibrium of forces which we will learn more about later. For now I think we have a good, very basic understanding of the MFE so let's break out of this side conversation and get back to the discussion at hand, which might seem like a rehash but covers important concepts.

Starting with the MFE we just discussed, as you move a RNA molecule further from its MFE state to a higher energy state, the number of possible secondary structures increases. For example, if a RNA molecule had a MFE state with -30kcal free energy there may only be 2 predicted secondary structures, but if that same molecule's energy was shifted to -20kcal it would have 10 unique secondary structures predicted. This is because the overall energy state of the design has changed in a positive direction, so the originally predicted pairing probabilities of the base pairs has shifted along with the calculations and now pairs that were strong are just too weak or those that were originally too weak are just now strong enough to pair up. This is why predicted secondary structure can be different by orders of magnitude or just slightly with a small change. Now with this in mind, the overall energy state of the RNA is calculated using temp as a variable at time of calculation. It is impossible to keep a perfect temperature and thus any fluctuations will cause energy shifts. There are many things that can tug and pull on that MFE state, in a constant fight to pull the MFE away from its cozy low energy home out into the busy world of higher energies.

Now with all that in mind and taken into consideration I hope you understand that RNA isn't always able to be explained or presented in absolutes, but instead in probabilities and likelihoods. When you think of this, if you focus all your energies on the MFE or just energies alone you may miss something. It is more of a kaleidoscope of what could happen, than a dial in type thing I hope you see.

You have to think of it in terms of how at any one time all the RNA nucleotides are tugging and pulling on every other molecule with fluctuating drivers of these forces. It does not matter how far away they are, as there is always a pull no matter how far it seems. These tiny forces are represented in the partition function as the statistical probability that two nucleotides will bond and that is what is referred to as pairing probabilities.

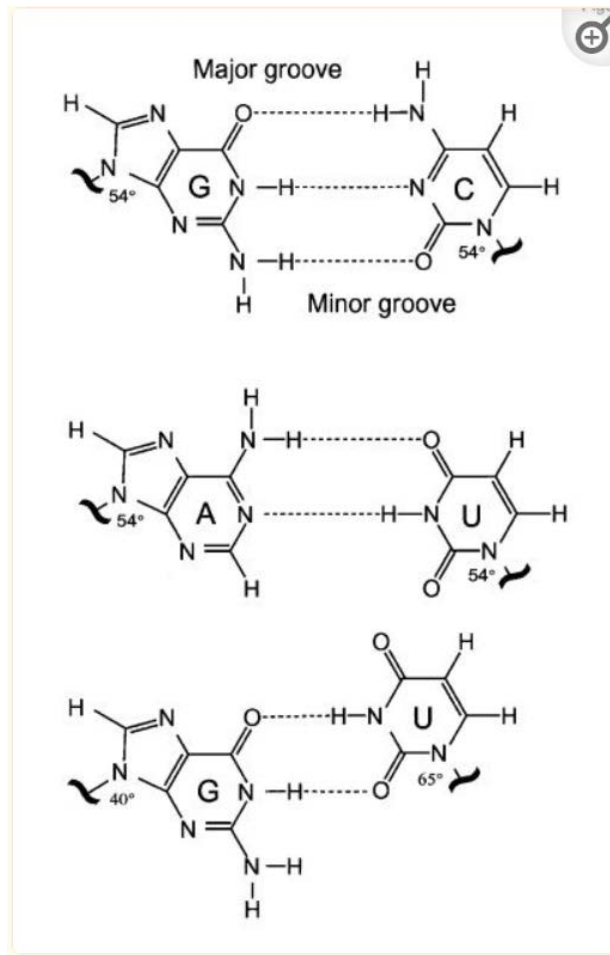
Pairing probability is a term used to refer to statistical calculations performed that represent the statistical probability that specific nucleotides will pair up and form a base pair. This does not care if a base pair is a Watson-Crick canonical base pair or whatever. I tend to think of it as the amount of attraction and pull each nucleotide has on each other in the entire primary structure.

These pairing probabilities then guide and really determine the secondary structure of the RNA based on the energies of the bonds whose forces are represented by the pairing probabilities. I'm not really sure at what point energies come into play in my mind, but in general energies are defined by the type of base pair involved as well as the base pairs and unbound nucleotides nearest it. This means that any nucleotides that have a potential to influence another nucleotide in the primary structure will have a pairing probability and that is what you see in the dot plots (we will cover dot plots further later). It is with certain geometries that you then get energy bonuses such as special loops, stacks, binding sites, etc., this includes boosting. In this thought process, boosting is adding a nucleotide in a location that causes it to have a pull on other RNA and thus increasing its strength of pairing probabilities and attractions at certain locations to help achieve a desired shape. This might be what is actually creating energy bonuses. It's the rigidity of the structure and the high pairing probability is a result of string energies. It is a bit circular, but only to a point.

I talk about pairing probabilities but I have not explained exactly how RNA bonds and thus how these probabilities are "generated". I will go into the probabilities a little bit more in the next chapter when we discuss the partition function. At this point I want to just tell you how RNA nucleotides bond to other nucleotides to give you a more complete understanding as we continue this discussion. I really feel like I should avoid not going into the smallest detail. We should strive in science, I think, to not assume everyone knows by heart everything they "should know" about a field. It often comes with lots of trial and a lot more error before we learn this all.

If you remember earlier I mentioned that RNA had shifts in hydrogen atoms that bind the RNA together, The RNA is actually bound by nucleotide pairs whose bonds themselves are generated by the location of the hydrogen binding sites on one nucleotide and the number of matching hydrogen atoms on the on-coming nucleotide at the same time. The amount of binding sites that line up on both nucleotides determines the strength of the bond. The binding sites are determined by the nucleotide type and as such, each nucleotide type has a specific

number and location. The Watson-Crick base pairs, which are GC and AU pairs, are the strongest. GC pairs have 3 hydrogen bonds and AU pairs have 2 hydrogen bonds. GU pairs which are not considered Watson-Crick pairs and also that which is weaker than an AU pair also have two functional hydrogen binding sites, but the major difference is that they are off-set. See this paper for reference and the following diagram from it. (reference <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1083677/>)



The application of analysis of pairing probabilities and structures within RNA is an interesting thing that has been the focus of the majority of my research for the entirety of the project. It was the analysis of the structures within RNA that were calculated by NUPACK and Vienna 2 that resulted in all the winning designs submitted by Sara during the DOE that the new PNAS paper is based on. DOE stands for Design of Experiment and is a scientific process where you literally perform experiments while recording specific values found during the investigation. In this experiment we were allowed to design any RNA sequence we wanted to make very simple hairpin switches with the explicit intent of marking if it is predicted that the design will do well or poorly, or even any other info that would assist with prediction of proper folding. \

The intent is that you can then later look over the data and in essence form a general understanding of the system you are investigating, as you are basically looking at a series of snapshots of the system as it progresses through various stages. In the manufacturing engineering world you would perform a DOE on a new product being built to understand what the actual performance range is of the device in order to educate everyone on how to make improvements. There are a few ways to look at that data and perform an analysis. The next chapter will go over some of the core concepts of the ensemble in relation to the partition function and how to start analyzing it to glean insight into how the RNA molecule will perform in the real world. Proof Read JMP 1-1-23

Chapter 2

Analysis of Partition Function: Pairing probabilities and the ensemble

My Eterna research has shown that the partition function is probably the most important aspect of an RNA design yet it is not a very well understood thing.

When I first started learning about RNA and the ensemble, I thought the partition function was the ensemble. This very important, yet tremendously small, aspect of RNA is very misunderstood by most. The partition function is just a single number, and by itself means really nothing. It is only after it is applied to the predicted structures in the ensemble that it means something and gives us vital information in the form of the pairing probabilities we discussed last chapter. So how does this all happen... Well, we have to do two sets of calculations on the RNA. First, we have to build the ensemble so to speak, and to do this we have to calculate all the predicted forces found in the arrangement of the nucleotides in the primary structure. This generates a list of all the possible structures but they are not sorted yet.

With this list of structures and their forces we apply an algorithm and bayesian probabilities which generates the partition function. Secondly, the partition function value is then used as a variable in a second algorithm where we apply it to the same list of structures and their energies to give us the probabilities that these nucleotide pairing will occur. It is with the probabilities that we are able to generate a list of pairing probabilities for each nucleotide in each structure and this is represented in the secondary structure as if the probability is high enough then it is considered to be bonded and if too low then it is predicted to be not bonded.

There are many ways to calculate the partition function, but currently we generally use McCasill's method developed in the 1990's as it is quicker and more accurate than its predecessors. A good explanation for it is found at the RNAlib-2.5.0 documentation website (https://www.tbi.univie.ac.at/RNA/ViennaRNA/doc/html/pf_algorithm.html). Here is the Abstract form McCaskill 1990 (<https://pubmed.ncbi.nlm.nih.gov/7685689/>) as I really like it and it helped me start to really understand things.

"A novel application of dynamic programming to the folding problem for RNA enables one to calculate the full equilibrium partition function for secondary structure and the probabilities of various substructures. In particular, both the partition function and the probabilities of all base pairs are computed by a recursive scheme of polynomial order N^3 in the sequence length N . The temperature dependence of the partition function gives information about melting behavior for the secondary structure. The pair binding probabilities, the computation of which depends on the partition function, are visually summarized in a "box matrix" display and this provides a useful tool for examining the full ensemble of probable alternative equilibrium structures. The calculation of this ensemble representation allows a proper application and assessment of the predictive power of the secondary structure method, and yields important information on alternatives and intermediates in addition to local information about base pair opening and

slippage. The results are illustrated for representative tRNA, 5S RNA, and self-replicating and self-splicing RNA molecules, and allow a direct comparison with enzymatic structure probes. The effect of changes in the thermodynamic parameters on the equilibrium ensemble provides a further sensitivity check to the predictions.”

Bonhoeffer S, McCaskill JS, Stadler PF, Schuster P. RNA multi-structure landscapes. A study based on temperature dependent partition functions. *Eur Biophys J.* 1993;22(1):13-24. doi: 10.1007/BF00205808. PMID: 7685689.

So now that we understand the partition function and how it generates the probability numbers we use, let's get back to the ensemble and the structures that are defined by it. I keep pushing the ensemble as it is the most important aspect I think, and to give you an idea why, the only standard metric that is currently part of Vienna2 and NUPACK that I have ever seen track with a design score is ensemble diversity or ensemble defect. Let that sink in for a moment....

In fact when I first noticed this back in 2016, I did a search into the old Eterna forum posts and found people pointing this out around 2011, but the post I found was largely ignored... I have never seen any other metric, such as MFE, centroid energy, whatever actually contains a signal that something is happening that is not a completely unique problem that changes from design goal to design goal. This is because all the RNA sequences are unique and there will never be ideal energy levels I believe, but there are ideal mechanical configurations and ensemble diversity's signal I believe is based on its relation to how stable the secondary structure is and how much diversity there is in the ensemble. The only problem with ED is that it is too generalist and applies to the ensemble. The problem with this is that I have found that only the first couple kcal of energy away from the MFE is needed and the ensemble contains MFE hundreds of steps away that are skewing the results. In statistics it is important to avoid a single data being more weighted than others, but that is what has actually happened with the ensemble diversity and defect metrics. As such I developed a new metric called Ensemble Variation that gives a instantaneous value at any energy level, verses single average metric with thousands of data points that go so far off into the distance that they are thus by accident given more weight than the the 1-4 kcal range that is most representative of the behavior of RNA in controlled conditions. Basically the ED gives a general sense for RNA ensemble diversity, but EV will tell you if it will be stable or not and maintain the MFE target structure based on this diversity, and also how much noise and diversity is there at the lower energy deltas or any energy point really.

Now why dedicate a portion of a chapter about pairing probabilities with an ensemble defect (ED) screed? The answer is that pairing probabilities essentially determine the the Ensemble Diversity/Defect value from Vienna and Nupack respectively. The ED really points to a RNA's average stability over the entire ensemble from the MFE all the way to +100Kcal if it was to go that way high. The point is that the full ensemble has 10's of thousands of alternate structures in it and ED points to how all off these compare directly to the MFE structure.. My research shows time and time again a clear slope in all the plots across all the labs that ED has a signal that clearly shifts as the Eterna score and fold change go up, regardless of the lab. The thing though is that each lab seems to have a bit of a different ED range that is good for each lab but they are

all kinda close. The only issue is that ED is way too wide and not very good for a fine tune understanding of stability. That is why I developed the Ensemble Variation metric that I will discuss later on.

Now to get back to pairing probabilities, one way I like to imagine them (and Eterna players will know what I mean when I talk about this), is I think of the dot plots you can look at in Eterna as those are a graphical representation of the pairing probabilities data and the entire ensemble. A dark spot is a strong predicted bond and a light spot is a weak pairing and a white spot is no pairing. The more designs that have a nuc pair the stronger that predicate bond and thus pairing probability. The data for this is generated by NUPACK and VIENNA2 as numerical values representing the statistical calculation of the chance for each to bond. It is normalized to the value of 1 with 1 being the strongest bond possible (i.e. 100% chance to bond) and 0 being no bond present (i.e. 0% chance to bond). In fact, the energy models will only list a pair if its probability to pair is > 0 by any number. The models output a list of every potential pair and probabilities for those pairs. That is what you are seeing when you look at the dot plots as well as the cool new tool arcplot, in that they are a graphical representation of those attractions and bonds.

To get back on point and how this all applies to the ensemble, the ensemble can be thought of as all the possible secondary structures that a design can take if you consider the entire range of possible kcal's and not just the MFE which is just the edge of the ensemble. Since the probabilities are just a statistical probability there are chances that the design will have slightly different results on occasion, and maybe even big sometimes, and this is enough to nudge the design and make it change shape often.

When we get the MFE secondary structure then, we are actually looking at what is predicted to be the most probable and common shape it will take at the lowest and thus strongest energy levels that can hold the RNA in place. You can however, peek at what things would be like if the energy levels were not perfect and that is the subopt function of NUPACK. With this you can get a list of secondary structures in the ensemble found at a specific energy range from the MFE. You can also get the energy delta of the energies that each unique alternate structure will appear at in relation to the MFE energy level. This function gives you the complete list of alternate structures and thus you can tell how many there are and what pairs are similar. Remember I said a few times that pairing probabilities define the secondary structure and we saw a glimpse of that when reviewing McCaskill's partition function algorithm, well you can see that in action now and understand what is happening. The pairing probabilities are related to the subopt alternate structures and vice versa and it all feeds into an understanding of how each nucleotide will pair. I feel like I am starting to kick a dead horse so I will close with this final all encompassing summary of what we just went over. This will ensure that we are ready for the next step, which is how all of this explains and feeds into an understanding that RNA and the shape it takes is ruled entirely by the laws of mechanics. The same laws that allow us to understand how the planets move through space and orbit the sun, as well as model those behaviors. We will in fact be modeling the behavior of RNA and using the dual body physics model to help teach the concepts after a review of the laws of Neutronian physics.

Let's walk through this process one last time now that we know what is going on in the steps.

A RNA sequence is the most stable at its Minimum free energy state (MFE) and naturally wants to stay there when at the temperature the MFE was calculated for. This is because as RNA gets colder the bonds become stronger and as it gets warmer the bonds begin to fall apart (this is what the melt point metric shows). Now it's impossible to keep a molecule at a specific temperature perfectly and there are always numerous environmental factors affecting RNA, so the actual energy levels fluctuate a little. When the RNA fluctuates its energy levels and thus the secondary structure vary depending on the delta from MFE, and the design will fluctuate between all the possible secondary structures in that kcal delta group of the ensemble. This means that if the energy levels go back to the MFE level, then it will be back at the MFE secondary structure, and if it goes to a different energy level it will take on that different shape.

Now some designs may have very few different alternate secondary structures and some may have a significant amount of alternate secondary structures in each unique group. Some of the groups will have very little change or shift from the MFE secondary structure and some will have a lot of change or shift between the secondary structures, regardless of how many secondary structures there are. The less variation and the less alternate secondary structures in each group of kcal delta's the more stable the design will be then. This leads into another question...why is it more stable when there is less variation? That answer is given I believe through looking at RNA through the lens of mechanics of materials for dynamic loads.

Chapter 3

What is mechanics and how does it explain RNA behavior?

Mechanics is roughly and to over-simplify is that study and application of the movement of bodies in motion and at rest. It is based on physics and allows us to understand the complex action on a mathematical level of when an object falls or is moved by some force. It covers statics and dynamics or understanding objects at rest and in motion respectively. We can accurately predict paths of movement and it is this same study that allows us to land people on the moon. Mechanics of materials then is the study of mechanics in complex systems combining what we know about static and dynamics and integrating it into a model of a complex system.

What this does for us then is that it explains how the spring action in a spring happens from a trigonometric, mathematica, and potential energy perspective. When an object moves on a fulcrum, or is pulled and pushed apart it behaves according to the laws of physics as we understand it through the laws of mechanics and ultimately by mechanics of materials as that is the study of mechanics in systems with statics and dynamics mixed. In mechanics you start learning about moments, which are really weird and the best I understand a moment is a force that resists movement. I usually hear it associated with rotating objects like with a moment of inertia, which is how much an object resists starting and stopping in motion. We will go into these in more depth later when we begin discussing how to link RNA to mechanics and how to perform the translation in you head.

So the reason I go into depth about mechanics is that I believe RNA should be thought of as a mechanical object and not only an energy object. As thus we should consider how the mechanical systems of the RNA design will perform when evaluating a design. This was part of the focus of my earlier writing on RNA and a large portion of this book are actually from my earlier work titled "Challenge Puzzle Solving Guide rev 2.0", however I will be adding a large amount of commentary and analysis to it. I think it has some good info that I want to incorporate into this book.

To start, I have long thought RNA should not be considered a single solid structure but thought of as a multitude of smaller segments that have become interconnected to each other. The smaller segments are the stacks between the loops and the loops are where they join together to form a single structure. They don't necessarily understand that they are a single strand (I'm talking more mechanical memory like bent metal), as once the design is folded it now functions as a complex mechanical system under a dynamic load trying to rip it apart and push it together and it still behaves in such a way that when pulled apart it forms a single large segment. An interesting fact is the the RNA backbone is what keeps a RNA chain of molecules together. They are linked to each other via a super strong hitch so to speak that requires a lot of energy to form and a similarly large amount of energy to separate. This is why a RNA design may fall apart and stop folding, but the RNA chain is left unbroken.

I believe that this approach will help alleviate difficulty in conveying the essence of what is happening behind the scenes when mutating specific nucleotides (nts) in such a way that it begins to pull itself in on itself as we mutate each nucleotide one at a time and then model that change in the mathematical models. The RNA will not start to fold right until the energies are just right and everything comes together as a critical mass of sorts is reached. This critical mass I think is the result of an equilibrium of forces within the design. The critical mass is reached when the design has a just low enough negative free energy to want to pull together and form a structure. It is fighting the energetic desire to form a straight line so to speak.

This is an important way to think about it because it helps drive home some things. The critical mass must be reached in order to break free from the straight line generated by the complete A pair stacks, and this straight line is formed by massive repulsive energies as A's repulse A's. This results in no strong pairing probabilities and only repulsion and thus no shape since there are no attractions. You can start to increase the attraction by changing or mutating the nucleotides one by one in order to start forming complementary base pairs. To severely oversimplify things you stabilize the design by mutating specific nucleotides to cause complementary base pairs to form which are GC, AU or GU base pairs. Now after this structure is formed you can still pull it apart as long as there is enough energy to counteract the lower free energy of the design. This is the complex mechanical system that is actually a single string and this is important to understand how mechanics of materials affects multi-state switch performance and even single state performance.

For a single state you need to design the RNA in a way to make sure that the hairpins and loops have relatively strong nucleotide bonds to ensure the loops form right and the stacks need to be extremely strong. The best single state structures we found back in the day were designs that had very little "noise" as we called it at the time in the dot plot. This we now know is the pairing probabilities of the design and if there is a lot of noise there are a lot of secondary structures in the ensemble at the closer kcal deltas and thus a less stable design. I used to talk about the sinkhole effect where if you mess up a RNA design it can suddenly change the entire structure drastically. This is a behavior of a design with low stability causing a drastic recalculation of the energies. As we move closer to a more stable design it can handle a bit more change as the ensemble gets smaller and there are less drastic changes. Now before I get into the weeds more with this, you may be asking yourself...as I was a few times, why am I going into so much detail about this and haven't even gotten to what is really happening...is this mindless dribble... I thought about that a moment myself and realized that though this part is quite mundane it is essential to properly understand something. I need to fully explain the process and that means going into some boring minute details. With that being said...

Since you are still here let's dig into and model the behavior of the RNA, if you will, and see what is actually happening as you theoretically make a design and multiple stacks and loops stable as that allows discussion into what is going on in a switch, it's a bit of a thought experiment if you will. In the next chapter we will connect mechanics to RNA behavior prediction through an understanding of the laws of mechanics.

Chapter 4

Introduction to application of mechanics to RNA including a short discussion on laws of mechanics

Single state RNA modeling thought experiment called “The hairpin: A simple structure to work through and observe pairing probabilities forming the ensemble and MFE structure connecting it to the mechanical study of statics”

The hairpin is a fun structure as it is not some theoretical stack with a loop devoid of external forces, for to explain stack energies and do so without the loop bit is very similar to teaching people everything about the classification of animals, but telling people to ignore the swifty bit at the end of most of them. The tail is an important thing and many animals use it for balance. I hope to demonstrate in this chapter and beyond what I really mean by that. I also hope to allow us to take a peek behind the curtains so to speak and understand why much of our practical knowledge works for how we understand energies in Eterna and thus allow us to progress to much more complex systems. Mechanics explains many things happening in simple single state designs, but it is only when you get into switches that they seem to matter the most.

Let's jump straight in then. Imagine that you have a long hairpin with alternating GC closing pairs, with alternating orientations of AU nucleotide pairs in the rest of the stack. This design would fold well in Eterna. Let's do a comparison of a 15 nucleotide long hairpin with the primary structure of “GAUAUCAAAGUAUC” that we will call the “alternating_AU_hairpin_sequence” and “GAAACAAAAGUUUUC” which we will call the “unidirectional_AU_hairpin_sequence” and they are both the same target secondary structure.



If you were to run this through Nupack4.0 using rna06 (rna06 is the name of a standard energy parameter set released in 2006 based on various recent research papers and is academically

recognized), you get the nucleotide pairing probabilities that are greater than non-zero, as well as those that have a greater than 0.0001 probability to pair. The 0.0001 filter is to filter out probabilities that are noise and would get in the way of the examples. Later on we will be removing this filter. Each pair has a base pair represented in i,j notation followed by a space and the pairing probability. Here is the nupack data for that first design.

alternating_AU_hairpin_sequence

MFE Secondary Structure = ((((((.....))))))

MFE Free Energy=-4.200000286102295

MFE Stack Energy=-4.200000286102295

Ensemble Defect (ED) = 0.01440257489139193

Kcal_Delta=1, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1

Ensemble kcal delta secondary structures and energy's

1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

Kcal_Delta=2, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1

Ensemble kcal delta secondary structures and energy's

1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

Kcal_Delta=3, EV_old=0.08333333333333333, EV_old_norm=41.666666666666664,

EV_new=133.33333333333331, structure_count=3

Ensemble kcal delta secondary structures and energy's

1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

2: .(((.....))). FreeEnergy: -2.44: Stack Energy: -1.80

3: (((.....))) FreeEnergy: -1.70: Stack Energy: -1.70

Kcal_Delta=4, EV_old=0.125, EV_old_norm=62.5, EV_new=200.0, structure_count=4

Ensemble kcal delta secondary structures and energy's

1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

2: .(((.....))). FreeEnergy: -2.44: Stack Energy: -1.80

3: (((.....))) FreeEnergy: -1.70: Stack Energy: -1.70

4: ..(((.....))).. FreeEnergy: -1.59: Stack Energy: -1.10

non-zero pair probs

0.0001

Pair_1: 1,16 0.9300543613

Pair_2: 2,15 0.9828239976

Pair_3: 3,14 0.9955238083

Pair_4: 4,13 0.9971616253

Pair_5: 5,12 0.9978022090

Pair_6: 6,11 0.9814199397

unidirectional_AU_hairpin_sequence

MFE Secondary Structure = ((((((.....))))))

MFE Free Energy=-3.200000286102295

MFE Stack Energy=-3.200000286102295
Ensemble Defect (ED) = 0.021009203060565033
Kcal_Delta=1, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1
Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20

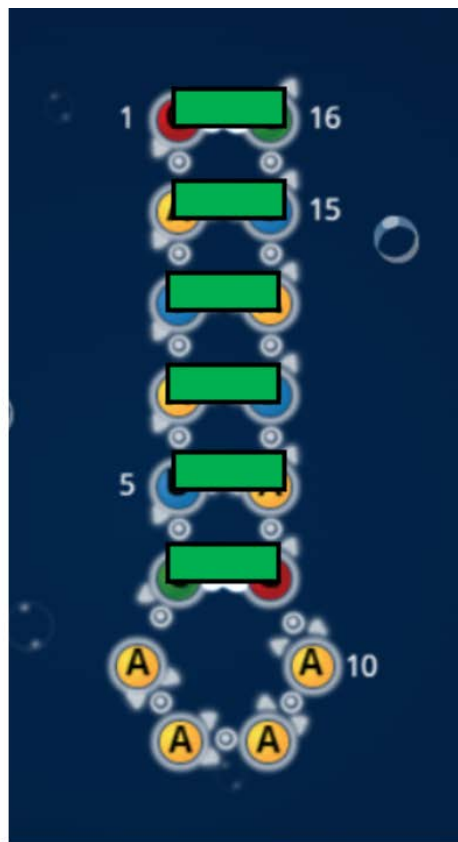
Kcal_Delta=2, EV_old=0.0625, EV_old_norm=31.25, EV_new=100.0, structure_count=2
Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
2: ((((((.....)))))) FreeEnergy: -1.20: Stack Energy: -1.20

Kcal_Delta=3, EV_old=0.08333333333333333, EV_old_norm=41.666666666666664,
EV_new=133.33333333333331, structure_count=3
Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
2: ((((((.....)))))) FreeEnergy: -1.20: Stack Energy: -1.20
3: .(((.....))). FreeEnergy: -1.44: Stack Energy: -0.80

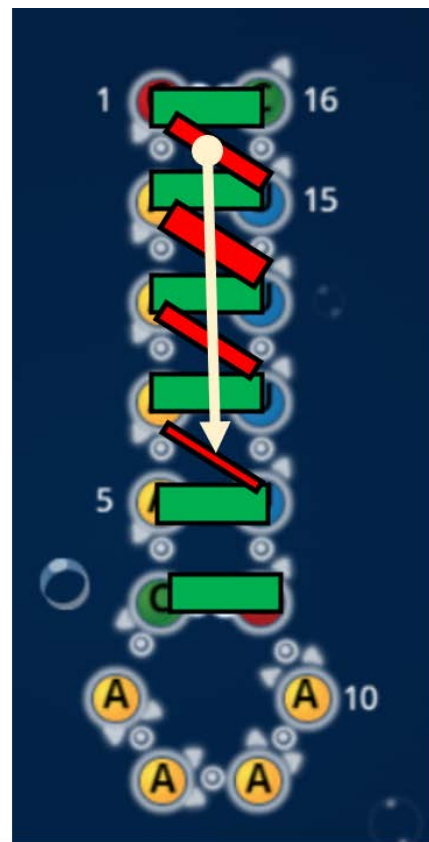
Kcal_Delta=4, EV_old=0.21428571428571422, EV_old_norm=107.1428571428571,
EV_new=342.8571428571428, structure_count=7
Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
2: ((((((.....)))))) FreeEnergy: -1.20: Stack Energy: -1.20
3: .(((.....))). FreeEnergy: -1.44: Stack Energy: -0.80
4: FreeEnergy: +0.00: Stack Energy: +0.00
5: ..(((.....))). FreeEnergy: -0.47: Stack Energy: +0.00
6: ((.(((.....))).) FreeEnergy: +0.30: Stack Energy: +0.00
7: (((.(((.....))).)) FreeEnergy: +0.30: Stack Energy: +0.00

non-zero pair probs
Pair_1: 1,15 0.0003830565
Pair_2: 1,16 0.9268202371
Pair_3: 2,14 0.0004217807
Pair_4: 2,15 0.9791552059
Pair_5: 3,13 0.0003397172
Pair_6: 3,14 0.9872231286
Pair_7: 4,12 0.0002270456
Pair_8: 4,13 0.9898084436
Pair_9: 5,12 0.9924448121
Pair_10: 6,11 0.9565416761

Now if you were to visualize this in Eterna and then overlay a diagram of the pairing probabilities you will get a good visual aid of what is going on with the internal forces of the single stack of the hairpin.



alternating AU hairpin sequence



unidirectional AU hairpin sequence

If you look at the two diagrams above you can see the forces for the intended complementary base pairs (those that are necessary to properly form the hairpin) represented as green squares, and these are the strongest pairings. The intensity of the predicted pairing probability and thus bonding strength is represented by the thickness of the bar. The weaker secondary forces are represented as red bars with the same rules.

In the `alternating_AU_hairpin_sequence` you can see that the pairing probs found (green bars) at the filtering level we have chosen only exist for nucleotide pairs in the stack that correspond to the target structure. This is the definition for this document when referring to preferred pairing. When you look at the `unidirectional_AU_hairpin_sequence` you can see that there are the same preferred pairing pairs, but now there are also extra predicted pairs of nucleotides. This secondary structure is shifted one down for nucleotides 1 through 4. These “shifts” in pairing probs are caused by internal forces and this is the point where mechanics of materials begins to help explain the scientifically accepted interpretations for the reason behind the behavior of RNA after the correct energy levels and equilibriums have been reached to establish the resulting pairing and thus alternate structures.

The first step in understanding this I think is to start with the concept of how a structure is built. It must have a foundation and structural support. What this means and looks like is dependent on each structure but it is what I think every structure boils down to in an almost oversimplified way. The first piece is the bolt that holds structures together. Think of each bar that represents a pairing probability as a bolt that is used to fasten a piece of metal to another piece of metal, or one structure to another. However, In this instance the empty-ish line down the middle is the weak point as that is where the hydrogen bonds are formed. These hydrogen bonds are what gives RNA their bonding strength by forming a connection or a pair and this then introduces a structural rigidity to the RNA by the nature of the pieces being secured together. This is because originally the RNA was just a bunch of nucleotides connected by the '3 and '5 locations and had an extremely strong strength, and when the RNA folds in on itself the hydrogen molecules in the RNA then start binding with the hydrogen receptor sites on the other molecules. Once all the nucleotides have been bolted down so to say, the entire stack is now a single physical structure with a mass from the nucleotides and the rigidity of the bindings. Where there is rigidity and a structure there are also internal forces desiring to reach an equilibrium, and we still have not considered the loop which we will add in later on in this conversation.

Think then of the undesired forces or undesired pairing probabilities in this design as unwanted internal forces that detract from the forces of the good pairing. This detraction exists because if you have a structure with focus pulling in opposite directions, the force of one direction negates the equivalent force in the other direction to put it roughly. With the examples given I hope you can see how the observed forces get the biggest in the 1 to 3 nucleotide range with the 2,14 pair being pretty strong. Since all the AU's have the same orientation the A's are free to pull towards the next U down from the intended U at will until it reaches the resistive forces of the loop and you can see this as the unwanted forces decrease and disappear at the loop.

This makes the design much less stable. You can see in the ensemble data, that for the alternating_AU_hairpin_sequence the group size for alternate secondary structures for the 1 and 2 kcal deltas is 1 alternate secondary structure for each group. That means that there is only the MFE structure found in the range of both 1 and 2 kcal deltas. As a sidenote those ranges I found are a really good quick guide for stability when dealing with Ensemble Variation. If you look at those same deltas for unidirectional_AU_hairpin_sequence you see that the 1 kcal delta has 1 secondary structure, but the 2 kcal has 2. The ensemble alternate structures are starting to increase now as the energy delta increases in the nupack data. This is a result of the increased number of observable pairing probability conflicts and negations. To bring this one step further and extend the analysis to the 3rd and 4th kcal deltas you start to see the effects become more apparent.

Here is the alternating alternating_AU_hairpin_sequence data for those two kcal deltas.

```
Kcal_Delta=3, EV_old=0.08333333333333333, EV_old_norm=41.666666666666664,  
EV_new=133.33333333333331, structure_count=3  
Ensemble kcal delta secondary structures and energy's  
1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20
```


2: .((((.....))))). FreeEnergy: -2.44: Stack Energy: -1.80
3: ((((((.....)))))) FreeEnergy: -1.70: Stack Energy: -1.70

Kcal_Delta=4, EV_old=0.125, EV_old_norm=62.5, EV_new=200.0, structure_count=4
Ensemble kcal delta secondary structures and energy's

1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20
2: .((((.....))))). FreeEnergy: -2.44: Stack Energy: -1.80
3: ((((((.....)))))) FreeEnergy: -1.70: Stack Energy: -1.70
4: ..((((.....)))).. FreeEnergy: -1.59: Stack Energy: -1.10

From this data you can see that the vast majority of the stack stays paired throughout, with only the end pairs not bonding properly 50% of the time. Compare this to the unidirectional_AU_hairpin_sequence which has the following structures.

Kcal_Delta=3, EV_old=0.08333333333333333, EV_old_norm=41.666666666666664,
EV_new=133.33333333333331, structure_count=3

Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
2: ((((((.....)))))) FreeEnergy: -1.20: Stack Energy: -1.20
3: .((((.....))))). FreeEnergy: -1.44: Stack Energy: -0.80

Kcal_Delta=4, EV_old=0.21428571428571422, EV_old_norm=107.1428571428571, EV_new=342.8571428571428,
structure_count=7

Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
2: ((((((.....)))))) FreeEnergy: -1.20: Stack Energy: -1.20
3: .((((.....))))). FreeEnergy: -1.44: Stack Energy: -0.80
4: FreeEnergy: +0.00: Stack Energy: +0.00
5: ..((((.....)))).. FreeEnergy: -0.47: Stack Energy: +0.00
6: ((.(((.....))).) FreeEnergy: +0.30: Stack Energy: +0.00
7: (((.(((.....))).)) FreeEnergy: +0.30: Stack Energy: +0.00

You can see that there is significantly more variation in the ensemble for the unidirectional AU sequence, with the 4 kcal delta actually showing that some of the preferred base pairs won't be made frequently. The design is also predicted to just not fold on one occasion in this kcal delta, as seen by the dots in structure 4 of Kcal_Delta=4. If you look at the secondary structure of elements 6 and 7 of Kcal_Delta=4 real quick, you can see in these structures that the 3:14 nucleotides are predicted to not pair in separate structures when you get to the 4 kcal group.

Intermediate Application

Now that we have an introductory understanding of how the pairing probability can be read let's open up the filter of the pairing probabilities a bit. This will allow us to look at this design deeper to see what is happening in the background or as people say, let's see how the sausage is made, as we dig deeper into the design ensemble. In fact in case you were not aware, the nupack pairing prob data we are looking at in the next lines is actually what we traditionally have called the noise in the dot plot or the light gray dots you see in them that are distinct from the black solid dots of strong pairs.

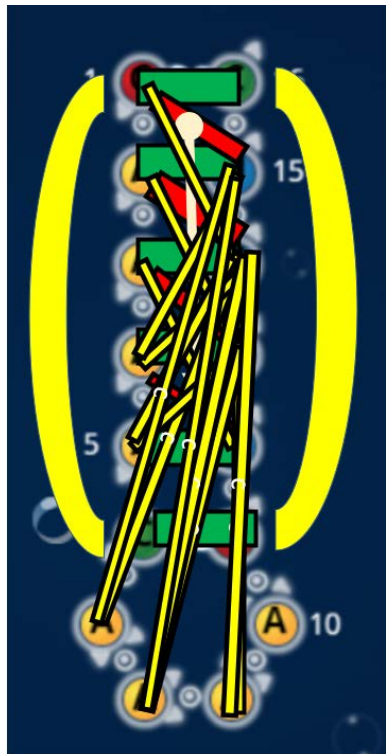
Pair_1: 1,6 0.0000560867
Pair_2: 1,14 0.0000406212

Pair_3: 1,15 0.0003830565
Pair_4: 1,16 0.9268202371

Pair_5: 2,13 0.0000514555
 Pair_6: 2,14 0.0004217807
 Pair_7: 2,15 0.9791552059
 Pair_8: 3,12 0.0000461409
 Pair_9: 3,13 0.0003397172
 Pair_10: 3,14 0.9872231286
 Pair_11: 3,15 0.0000987779
 Pair_12: 4,12 0.0002270456
 Pair_13: 4,13 0.9898084436
 Pair_14: 4,14 0.0000932655
 Pair_15: 4,15 0.0000115877
 Pair_16: 5,12 0.9924448121

Pair_17: 5,13 0.0000637435
 Pair_18: 5,14 0.0000119309
 Pair_19: 6,11 0.9565416761
 Pair_20: 7,14 0.0000134503
 Pair_21: 7,15 0.0000213655
 Pair_22: 8,13 0.0000126749
 Pair_23: 8,14 0.0000245322
 Pair_24: 8,15 0.0000131457
 Pair_25: 9,13 0.0000128704
 Pair_26: 9,14 0.0000122988
 Pair_27: 11,16 0.0000152770

If you don't have experience with this data then the diagram below will probably look like insanity at first, as the new yellow bits are every nucleotide pairing prob for the new probabilities since opening up the probability threshold filter.



unidirectional AU hairpin sequence

Let's look at this real quick and pull apart the many layers of probabilities and see what is happening inside and then we will change the orientation of the 1:16 GC pair so it is now a CG pair. It will be at this point that the connection to the world of mechanics of materials will be made deeper. We will go layer by layer, so think of the different steps in filtering we did on the pairing probabilities as the layers. We first exposed very strong probabilities, followed by slightly weaker prob's which are actually starting to approach the strength of the primary binding sites

but not quite there yet, and then we expose the very weak probabilities that are orders of magnitude smaller than the first 2 layers.

Let's call the very top layer or main pairing bond layer the 3rd layer or layer 3. The 2nd layer is the mild pairing probabilities that form strong-ish prob's with nucleotides but are just slightly weaker than what is required to successfully form a strong base pairing that shifts it into level 3. Then there is level 1 which is the true noise in the probabilities and this is the probabilities of nucleotides that have a super weak attraction to anything that it could pair with and we will later see that often layer 1 is just a factor of nucleotide distance.

Lets assign to level 3 a cutoff value of .7 as this is a value that I recently pulled from Jandersonlee's arcplot documentation. I have always struggled to understand a good range for this as I have always seen it as a range, but nailing down an exact value to use had been rough. Super strong bonds like a GC pair in a stack will have around .9 for the prob but it could be lower. One thing to not forget thow is that the question has always been this at what point does it exit the first layer and his documentation and practical experience show that .7 seem to be a community agreed on number and it makes a lot of sense to me based on what I hav observed over time. Level 1's threshold would be any value that is less than 0.0001, and it is worth noting that this is 4 positions to the right of the decimal which represents a super weak force or attraction. It follows then that anything in the middle between .0001 and .7 can be considered Layer 2. This then also points to the concept that despite a strong bong being .7 or more, you can still have bonds that play a significant role even down to .0001 which is ridiculously weak yet powerful in the background. It should not be forgotten that every RNA nucleotide in a RNA sequence is always attracted to every other nucleotide at all times and its only the forces that win that end up forming the bonds that we observe in the predicted and measured secondary structures. RNA is a massively dynamic fluidic structure is a good way to sum it up.

Visualizing these focus and attractions is a difficult thing and for this book we will use arrows of various size and thickness to show this. Starting with layer 1 we see the diagram of a large pinkish arrow with two pointed ends below. The large arrows are rotated to show the direction of attraction in both directions. This orientation depicts the direction of the nucleotide pairing probabilities of the nuc pair it is above in the diagram.



```

Sequence = GAUAUCAAAAGAUUAC
MFE Secondary Structure = ((((((.....))))))
MFE Free Energy=-4.200000286102295
MFE Stack Energy=-4.200000286102295
Ensemble Defect (ED) = 0.01440257489139193
Kcal_Delta=1, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1
Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

Kcal_Delta=2, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1
Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

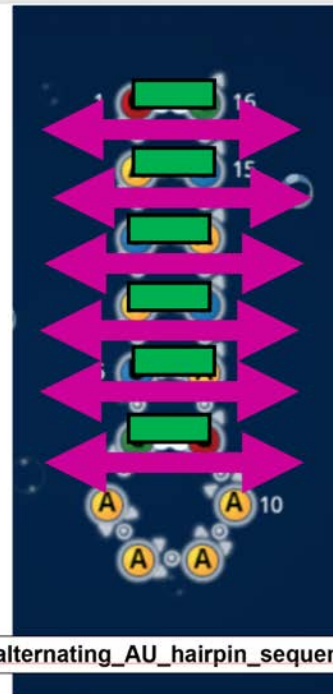
Kcal_Delta=3, EV_old=0.0833333333333333, EV_old_norm=41.666666666666664,
EV_new=133.33333333333331, structure_count=3
Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20
2: .(((.....))) FreeEnergy: -2.44: Stack Energy: -1.80
3: (((.....))) FreeEnergy: -1.70: Stack Energy: -1.70

Kcal_Delta=4, EV_old=0.125, EV_old_norm=62.5, EV_new=200.0, structure_count=4
Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20
2: .(((.....))) FreeEnergy: -2.44: Stack Energy: -1.80
3: (((.....))) FreeEnergy: -1.70: Stack Energy: -1.70
4: -(((.....))) FreeEnergy: -1.59: Stack Energy: -1.10

non-zero pair probs
0.0001

Pair_1: 1,16 0.9300543613
Pair_2: 2,15 0.9828239976
Pair_3: 3,14 0.9955238083
Pair_4: 4,13 0.9971616253
Pair_5: 5,12 0.9978022090
Pair_6: 6,11 0.9814199397

```



alternating_AU_hairpin_sequence

In this picture you can see that each pairing probability is in a single direction of alignment. This direction of alignment in relation to the alignment of the nucleotide base pair is considered an axial alignment in mechanics. We know that the RNA strand is wanting to tear itself apart so this diagram shows that we have designed a RNA molecule whose predicted pairing probabilities show why that RNA strand is predicted to come together so strongly and be able to take hold with the specific nucleotide configuration. Being able to read and interpret what is happening with that information is where mechanics and some of the basic laws of motion come in handy to give us the scientific tools we need to fully understand the event unfolding over time.

It is at this point that I want to start shifting how we talk about pairing probabilities and separate out the components and show them to you, as well as really break down the concepts to the laws of physics and give us the tools I mentioned a moment ago to further the discussion.

The 3 Laws of Motion

1. **1st Law: "An object will remain at rest or in uniform motion in a straight line unless acted upon by an external force."**

a. This is what is known as inertia and this rule is often called the Law of Inertia

2. **2nd Law: Force is equal to the change in momentum over time.**

a. The time rate of change of the momentum of a body is equal in both magnitude and direction to the force imposed on it. Force is a vector and has thus magnitude and direction, mass is considered a scalar property and only increases or decreases the scale of the results of calculations. The equation for this is $F=ma$ and this then shows that force is the representation of a mass being accelerated. The acceleration is a constant and never changes, it is however the mass of each object that changes and thus scales the observed acceleration whose level is represented by the term and concept of forces. Forces are governed by the rules of vectors in turn. This rule also applied to angular momentum that results from moments being acted upon. A moment is the same as a vector but for rotation.

An important note follows that must be understood when discussing acceleration in relation to RNA and this situation. Acceleration is important to understand here as it is vital to understanding how the forces work in relation to RNA. The 2nd law as most know it is a little white lie as they say for what Newton actually said was "force is the time rate of change of momentum," not force equals force times momentum. The difference is that momentum is the product of mass times velocity. Momentum is also a product of mass times velocity and Newton wrote the 2nd law in terms of instantaneous momentum in the form of $F = (dp/dt)$ where \mathbf{p} is the vector for momentum whose formula is $\mathbf{p} = \text{mass} \times \text{velocity}$ and velocity is defined as $v = \text{acceleration} \times \text{time}$. Velocity is how fast an object is and acceleration is the rate of change in location at a single point of time, and only when we give it a time duration do we now know how fast it accelerated through that point, but without time we only know an object moved from point A to point B. Now if we do the math we find the following. $F = (\text{mass} \times \text{velocity} / \text{time}) = [\text{mass} \times (\text{acceleration} \times \text{time})] / \text{time} = (\text{mass} \times \text{acceleration} \times \text{time}) / \text{time}$ which factors out to $F = \text{mass} \times \text{acceleration}$. So now we need to think of RNA and force with this in mind, we are evaluating the forces on the nucleotides that are attracting it to and causing it to move towards other molecules. We want to know the force that moved it from its point at rest to its next point at rest after it moved and that is 2 points in space and thus acceleration. We don't know how quick it happened so we don't know the velocity, and thus the acceleration does not change when looking at a nucleotide pair, but the differences in mass of each nucleotide and that is the change in mass causing varying forces..

3. 3rd Law. For every action there is an equal and opposite reaction.

- a. This explains how forces react when they interact with an object

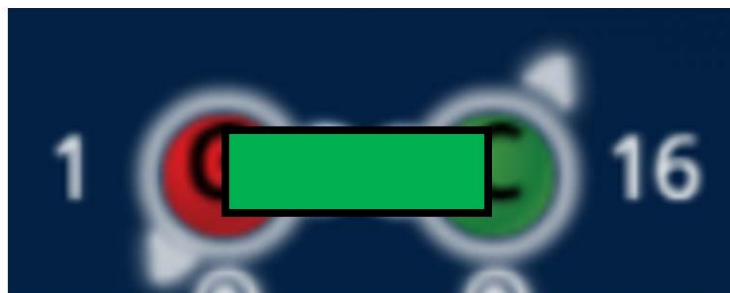
Here is a good book <https://bigthink.com/starts-with-a-bang/most-important-equation-physics/>

Now how does all this apply and why did I go into so much detail for the 2nd Law? Well let's look at the axial alignment of the nucleotide base pair 1,16. You can see that it is an axial orientation for the pairing probability. The hydrogen bonds in the individual nucleotides are generating an attraction defined by organic chemistry and it is found that "a base pair is formed when two bases interact noncovalently, each respectively involving one of these three edges, and form a considerably planar geometry with at least two interbase hydrogen bonds (H-bonds)" ([https://pubs.acs.org/doi/10.1021/acsomega.8b03689#:~:text=Noncoding%20RNA%20molecules%20are%20composed,bonds%20\(H%2Dbonds\).](https://pubs.acs.org/doi/10.1021/acsomega.8b03689#:~:text=Noncoding%20RNA%20molecules%20are%20composed,bonds%20(H%2Dbonds).))

Now the bond of this nucleotide needs to be analyzed further. Starting with just the GC nucleotides only with no other bodies acting upon them, imaging they are isolated by themselves, just the two of them.



Now at this point I want you to think of the pairing probability in relation to the hydrogen bonds. The higher the probability the higher the bond. Lets now at this point start saying bond instead of predicted pairing probability, but still keep that previous phrase in your head as we are just using a synonym in a way. Let's then also now add the green bar to represent the bond of the nucleotides to aid the visual analysis.



Three-Body Thought Experiment

Below is my take on the three body physics thought experiment, but repurposed for RNA. It is a representation of a theoretical single nucleotide base pair and thus a single base pair long stack/stem. The single and dual body thought experiments are crucial in mechanics to build one understanding of the interaction of the forces and help reinforce how the system works. This 3-body experiment can be thought of as dual dual-bodys. One body of each group is out of the range and influence of the other, while a single body shares the domain as both the other nucleotides that are outside of each other's domain. This is achieved by anchoring a nucleotide from each group with an immovable mass-less anchor that prevents it from moving and then moving the single free nucleotide between them. This significantly reduces any questions about interaction of other bodies. Let's start then by first understanding them the molecular weights of the nucleotides and thus the forces of attractions we would expect to see from nucleotides with each other in relation to other nucleotides.

<https://www.thermofisher.com/us/en/home/references/ambion-tech-support/rna-tools-and-calculators/dna-and-rna-molecular-weights-and-conversions.html>

Nucleotide	MW (g/mol)
Ribonucleotide Triphosphates (Avg. MW = 499.5)	
ATP	507.2
CTP	483.2
GTP	523.2
UTP	484.2
Ribonucleotide Monophosphates (Avg. MW = 339.5)	
AMP	347.2
CMP	323.2
GMP	363.2
UMP	324.2

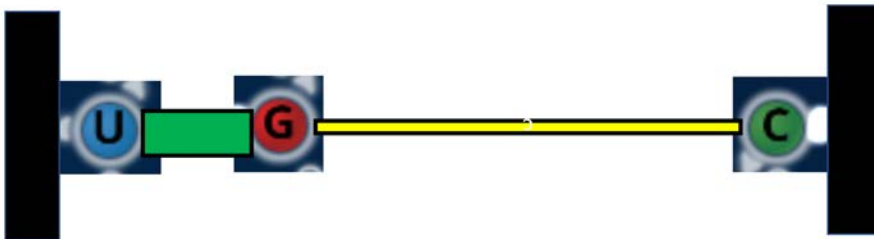
We see that the G is the heaviest so it will have a greater mass and greater resistance to change in position. The C is lighter and will have less resistance to motion. Now imagine a theoretical new G added somewhere off to the right side of the C. What we know through

numerous studies is that the new G will generate a bonding probability. As will be demonstrated when two complementary base pairs get closer and closer (say from a hairpin with the stack getting shorter and shorter) the pairing probabilities will increase and when the distance get farther apart the probabilities will decrease.

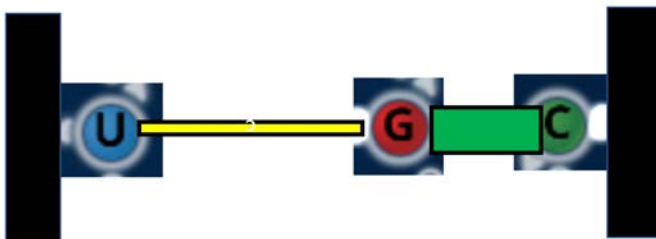
So thus as we move the new G closer and closer being sure to maintain the same plane and axial orientation the C will start to have a higher and higher predicted bond and at some point it will have a competing bond with the U and try to pair with the new one if the conditions are right. To better explain this lets use a different nucleotide configuration that is more logical such as the UG pair with a C nucleotide some distance off, and the U nucleotide is anchored to an immovable object with 0 mass that can affect anything at the 5' location as it would normally be in a RNA strand to other nucleotides in the strand.



The UG bond can be seen and the C is far off and has only the slightest predicted bonding probability. Now as the C nucleotide approaches the G the predicted bonding starts to increase.



And then there is a point where G will leave the U and bond with the C as the bond is much stronger



The movement of the G nucleotide in the example is a result of the bonding effects that cause the G nucleotide to be attracted to and thus move to the C nucleotide, for example. An important thing to keep in mind in this thought experiment is that the binding site on each nucleotide is only on one side and the other side contains a backbone. Thus when the theoretical G's attraction is too far away from the U's attraction, the pull from the C is now stronger and this results in a tug back. and moved toward the opposite nucleotide it actually had to force that previous bond to sever and then pull on the hydrogen bonding sites to enact the change in direction of travel. (may need to put this at the end of this section)

A change in position over time is referred to as velocity and the G had for a moment a measurable velocity. In physics the rate of change in velocity is referred to as acceleration and the G went from a positional reference of 0 to some other number in a period of time and went from a speed of 0 to some other speed and this is measured by acceleration. The G nucleotide has mass and thus the forces represented by the nucleotides travel is represented by the 2nd law of physics or $F = ma$. The G nucleotide was in a state of rest until acted on by the C nucleotide which caused it to gain inertia and pull away from its stationary state. Now the force of the pull of the G nuc on the C nuc that is anchored will cause the G to pull away from the U.

If you were to remove the anchor for the C nuc then it would be free to move on its own, and if you moved the C nuc closer it would eventually start to pull itself into the G and the hydrogen bonds with the A would be replaced with the C for the G nuc. These two different examples used the same setup to show the force pulling in both ways. This is the application of the 3rd law of physics.

This will now allow us an efficient way to discuss the behavior of RNA as it folds into a single state structure, and then later it will help us understand how a RNA riboswitch's switching action is governed by the laws of motion and explained through mechanics of material. We will see in this and the following chapter the application and demonstration of this before moving on to more complex bodies.

Also, now that we have connected the behaviors of RNA to the first 3 laws of physics we can now use the correct language when referring to the behavior and forces acting on the RNA nucleotides bodies as they change distance from each other. Lets now think of predicted pairing probabilities as numbers that represent forces that push and pull on other nucleotides. If I speak of the forces in relation to nucleotides I am speaking of the forces represented by the pairing probabilities.

Now with this all in mind let's jump back to the image from page 13 ?????, but before we go there I want to assure you that I am not just randomly jumping around with concepts. I needed to ensure that you understood what has become scattered common knowledge on the high level, and then present the mystery of pairing probabilities with a lot of noise. I did this so that as I walked you through the mental exercise of connecting the nucleotide pairing energies and attraction to the laws of motion, you would have that in your mind and then be able to transition more smoothly into the next part of the conversation. This next segment of our conversation is

one in which we peel back further the layers of the forces that the nucleotides exert on each other. Lets then begin what I would like to call the “Transformation of a Hairpin.

I have found that often when I use the term transformation when discussing a shape or structure people do not understand what I mean. What I mean is this, a transformation is an idea, and what if... Imagine that you have a triangle with angle A = 45, angle B = 45 and angle C = 45. We know this as a right triangle. Now what if that triangle changed shape and the angles changed, but they changed over time to have angle a = 30, angle B = 30, and angle C = 60. This was not an instantaneous change but a slow change over time as each point of the triangle moved along the plan from its origin to its destination. With that in mind I invite you to join an investigation of the transformation of a hairpin.

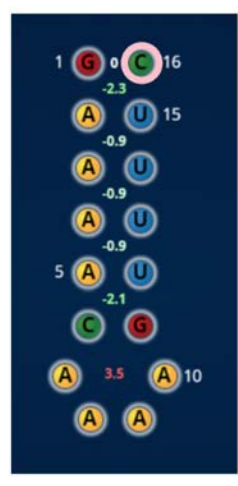
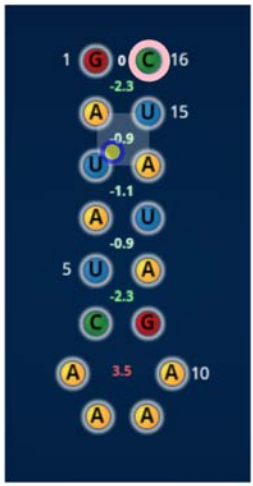
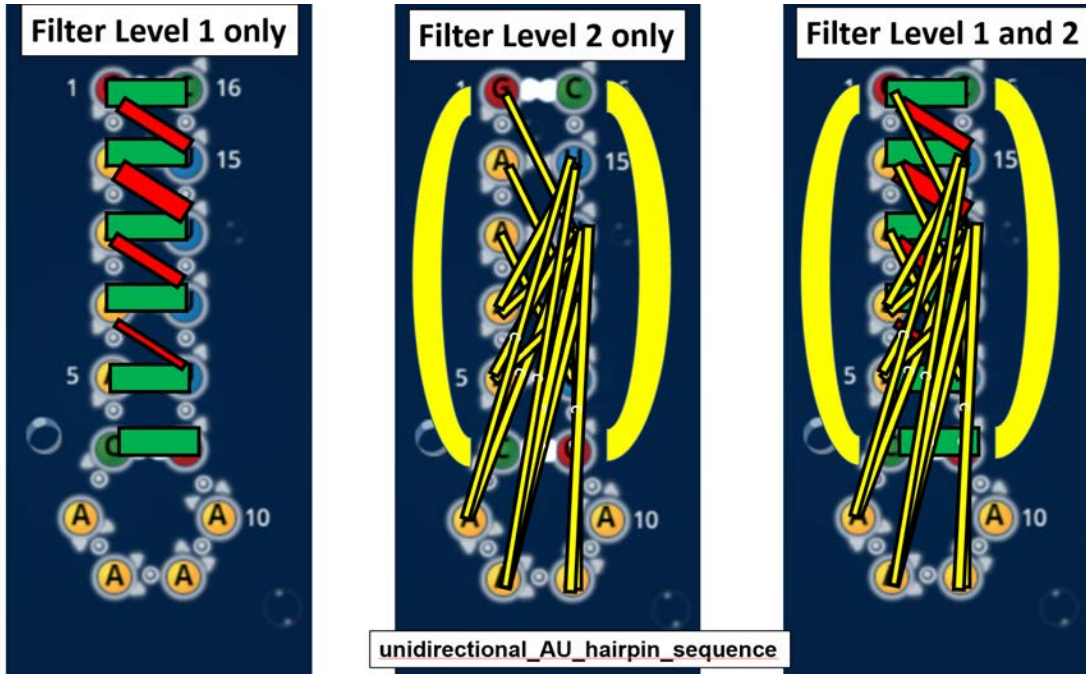
Transformation of a Hairpin

Let's start with the investigation with the “unidirectional_AU_hairpin sequence” and all the probabilities we found and discussed earlier. Below you will find a diagram of the pairing probability influences we first introduced you too in the earlier chapters, but this is taken a step further. I'll explain what is going on but as a delivery mechanism I want to walk you through what you are seeing so that you can understand it on a more fundamental level than if i just told you.



On initial assessment all that can be seen are a mess of predicted pairing probabilities and thus as we have been learning it can be derived that there are a mess of forces both large and small. I want to show you that it is this mess of forces that created the contents of the ensemble that we discussed before. Further this ensemble is defined by the pairing forces and the ensemble is a great analogy or vehicle for the explanation of the pairing probabilities.

Lets separate them all by their filtering layers first and then look at each filter group separately and as a whole. The diagram above is all of the filtering levels added on top of each other and thus is full of noise. Next will be the individual levels andAs I have shown in diagrams in the previous chapters, the colored lines and boxes represent the predicted pairing probabilities from NUPACK (that I visualized as best I could.



MFE Secondary Structure = ((((((.....))))))
 MFE Free Energy=-3.200000286102295
 MFE Stack Energy=-3.200000286102295

Ensemble Defect (ED) = 0.021009203060565033

- | | |
|--|---|
| 1. 1: ((((((.....)))))) FreeEnergy: -3.20 Stack Energy: -3.20 | 3: .(((.....))). FreeEnergy: -1.44: Stack Energy: -0.80 |
| 2. 1: ((((((.....)))))) FreeEnergy: -3.20 Stack Energy: -3.20 | 4: FreeEnergy: +0.00: Stack Energy: +0.00 |
| 2: (((.....))) FreeEnergy: -1.20 Stack Energy: -1.20 | 5: ..(((.....))).. FreeEnergy: -0.47: Stack Energy: +0.00 |
| 3. 1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20 | 6: ((.....)).. FreeEnergy: +0.30: Stack Energy: +0.00 |
| 2: (((.....))) FreeEnergy: -1.20: Stack Energy: -1.20 | 7: (((.....))) FreeEnergy: +0.30: Stack Energy: +0.00 |

First level

Pair_1: 1,15 0.0003830565
Pair_2: 1,16 0.9268202371
Pair_3: 2,14 0.0004217807
Pair_4: 2,15 0.9791552059
Pair_5: 3,13 0.0003397172
Pair_6: 3,14 0.9872231286
Pair_7: 4,12 0.0002270456
Pair_8: 4,13 0.9898084436
Pair_9: 5,12 0.9924448121
Pair_10: 6,11 0.9565416761

2nd level

Pair_1: 1,6 0.0000560867	Pair_15: 4,15 0.0000115877
Pair_2: 1,14 0.0000406212	Pair_16: 5,12 0.9924448121
Pair_3: 1,15 0.0003830565	Pair_17: 5,13 0.0000637435
Pair_4: 1,16 0.9268202371	Pair_18: 5,14 0.0000119309
Pair_5: 2,13 0.0000514555	Pair_19: 6,11 0.9565416761
Pair_6: 2,14 0.0004217807	Pair_20: 7,14 0.0000134503
Pair_7: 2,15 0.9791552059	Pair_21: 7,15 0.0000213655
Pair_8: 3,12 0.0000461409	Pair_22: 8,13 0.0000126749
Pair_9: 3,13 0.0003397172	Pair_23: 8,14 0.0000245322
Pair_10: 3,14 0.9872231286	Pair_24: 8,15 0.0000131457
Pair_11: 3,15 0.0000987779	Pair_25: 9,13 0.0000128704
Pair_12: 4,12 0.0002270456	Pair_26: 9,14 0.0000122988
Pair_13: 4,13 0.9898084436	Pair_27: 11,16 0.0000152770
Pair_14: 4,14 0.0000932655	

alternating_AU_hairpin_sequence

MFE Secondary Structure = ((((((.....))))))

MFE Free Energy=-4.200000286102295

MFE Stack Energy=-4.200000286102295

Ensemble Defect (ED) = 0.01440257489139193

Kcal_Delta=1, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1

Ensemble kcal delta secondary structures and energy's

1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

Kcal_Delta=2, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1

Ensemble kcal delta secondary structures and energy's

1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

Kcal_Delta=3, EV_old=0.08333333333333333, EV_old_norm=41.666666666666664,

EV_new=133.33333333333331, structure_count=3

Ensemble kcal delta secondary structures and energy's

1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

2: .(((.....))) FreeEnergy: -2.44: Stack Energy: -1.80

3: (((.....))) FreeEnergy: -1.70: Stack Energy: -1.70

Kcal_Delta=4, EV_old=0.125, EV_old_norm=62.5, EV_new=200.0, structure_count=4

Ensemble kcal delta secondary structures and energy's

- 1: ((((((.....)))))) FreeEnergy: -4.20: Stack Energy: -4.20
- 2: .(((.....))). FreeEnergy: -2.44: Stack Energy: -1.80
- 3: ((((((.....)))))) FreeEnergy: -1.70: Stack Energy: -1.70
- 4: ..(((.....))). FreeEnergy: -1.59: Stack Energy: -1.10

non-zero pair probs

0.0001

Pair_1: 1,16 0.9300543613

Pair_2: 2,15 0.9828239976

Pair_3: 3,14 0.9955238083

Pair_4: 4,13 0.9971616253

Pair_5: 5,12 0.9978022090

Pair_6: 6,11 0.9814199397

unidirectional_AU_hairpin_sequence

MFE Secondary Structure = ((((((.....))))))

MFE Free Energy=-3.200000286102295

MFE Stack Energy=-3.200000286102295

Ensemble Defect (ED) = 0.021009203060565033

Kcal_Delta=1, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1

Ensemble kcal delta secondary structures and energy's

- 1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20

Kcal_Delta=2, EV_old=0.0625, EV_old_norm=31.25, EV_new=100.0, structure_count=2

Ensemble kcal delta secondary structures and energy's

- 1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
- 2: ((((((.....)))))) FreeEnergy: -1.20: Stack Energy: -1.20

Kcal_Delta=3, EV_old=0.08333333333333333, EV_old_norm=41.666666666666664,

EV_new=133.33333333333331, structure_count=3

Ensemble kcal delta secondary structures and energy's

- 1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
- 2: ((((((.....)))))) FreeEnergy: -1.20: Stack Energy: -1.20
- 3: .(((.....))). FreeEnergy: -1.44: Stack Energy: -0.80

Kcal_Delta=4, EV_old=0.21428571428571422, EV_old_norm=107.1428571428571,

EV_new=342.8571428571428, structure_count=7

Ensemble kcal delta secondary structures and energy's

- 1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
- 2: ((((((.....)))))) FreeEnergy: -1.20: Stack Energy: -1.20
- 3: .(((.....))). FreeEnergy: -1.44: Stack Energy: -0.80
- 4: FreeEnergy: +0.00: Stack Energy: +0.00
- 5: ..(((.....))). FreeEnergy: -0.47: Stack Energy: +0.00
- 6: ((.....)). FreeEnergy: +0.30: Stack Energy: +0.00
- 7: (((.....)).) FreeEnergy: +0.30: Stack Energy: +0.00

non-zero pair probs

Pair_1: 1,15 0.0003830565
Pair_2: 1,16 0.9268202371
Pair_3: 2,14 0.0004217807
Pair_4: 2,15 0.9791552059
Pair_5: 3,13 0.0003397172
Pair_6: 3,14 0.9872231286
Pair_7: 4,12 0.0002270456
Pair_8: 4,13 0.9898084436
Pair_9: 5,12 0.9924448121
Pair_10: 6,11 0.9565416761

Diving right in, let's start with the unidirectional AU sequence, we see that the 1kcal delta ensemble group only consists of one secondary structure and it is the same as the MFE. For reference this is the MFE secondary structure and the energies for the structure from NUPACK as a refresher from earlier in the chapter.

unidirectional_AU_hairpin_sequence

MFE Secondary Structure = ((((((.....))))))
MFE Free Energy=-3.200000286102295
MFE Stack Energy=-3.200000286102295

Upon closer inspection we also see that the free energy and the stack energy are both the same as those found for the MFE secondary structure as seen below.

Kcal_Delta=1, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1
Ensemble kcal delta secondary structures and energy's
1: ((((((.....)))))) FreeEnergy: -3.20: Stack Energy: -3.20

When you shift then to the 2kcal delta group we find that it has more than the 1 kcal group in that it has 2 elements now. The first element has the same structure and energies as the MFE, same as the 1 kcal delta. This is because the 1kcal group is contained within the 2kcal group, and another reason is that the first secondary structure of every alternate structure list that NUPACK generates is the MFE secondary structure. The 2nd element though has a slightly different secondary structure than the MFE, and here is where we will start the discussion around the behavior of the structure in the voice and through the lens of physics and mechanics.

Element 2 if of the 2Kcal group from the unidirectional AU sequence has a secondary structure with a dot bracket notation of ((((((.....)))))) and this translates to the following image based on the nucleotides that are being used in each position.



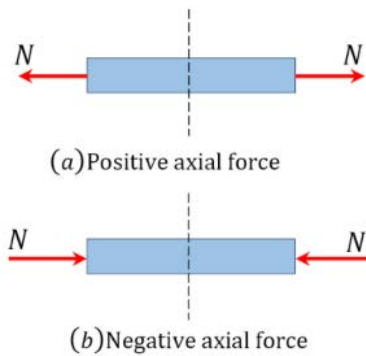
This diagram shows that the loop of the hairpins has opened up compared to the MFE structure. The change is small, but you can see that 1 single nucleotide pair has stopped bonding and is no longer paired to each other or any other nucleotides now. This nucleotide pair is pair 6,11 which is a CG nucleotide pair. This pair is usually very strong, and it is one of the so-called Watson-Crick's complementary base pairs and is indeed the strongest. In order for that bond to be predicted to fail then something must be predicted to be pulling against them with a force that would exceed the CG's relative forces. The hairpin is not completely broken at this point as all the pairs at 5, 12 and on are still predicted to bind.



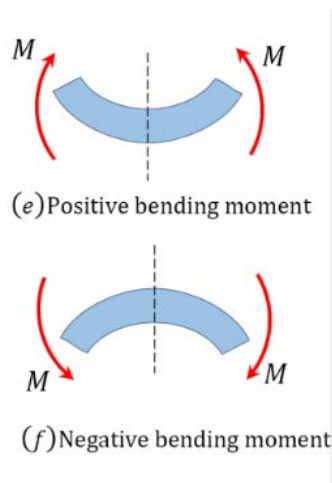
unidirectional_AU_hairpin_sequence

To talk a little bit more about the 6, 11 CG pair, that pair as found in the MFE structure when they are bound to each other are demonstrating the predicted forces at that nuc pair which are strong and create a bond that goes from from side to side in the figure and does not go vertical. The technical term for the force that a nuc pair is exhibiting is referred to as a "normal force". If you reference (<https://pressbooks.library.upei.ca/statics/chapter/3-types-of-internal-forces/>) "The normal force at any section of a structure is defined as the algebraic sum of the axial forces

acting on either side of the section.”. The diagram below shows examples of the two types of axial forces.



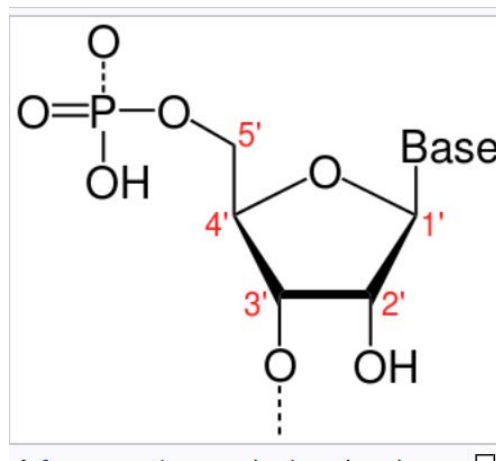
The nucleotides are bound through axial forces acting along the midline of the stack down the middle. The external hairpin loop is also representing a bending moment which is defined as “The bending moment (BM) is defined as the algebraic sum of all the forces’ moments acting on either side of the section of a beam or a frame.” Below is what a bending moment looks like and in our situation we have a positive bending moment as the loop of the hairpin bends upwards in the figure, as well as the 5’ and 3’ end closing at the top of the image vs the bottom of the image.



Now remember what we said earlier about the laws of mechanics that for every force there is an equal and opposite reaction. With this approach I postulate that the RNA strand wants to pull itself straight and the forces of the bonds in the stack thus have a strong positive axial force pulling it apart. The hydrogen bonds then are providing a negative axial force, and at the same time there is a bending moment that is applying forces to pull the rna strand up. The forces pulling the strand upwards which is now forming a curve, are also now trying to straighten itself out. This is because forces that are straightening itself out are the forces found inside the loop pushing out which are in part generated or associated by the positive axial forces.

RNA Nucleotide Components

RNA that we design in the labs starts as a string of A's in the Eterna interface and we can then decide what nucleotide to put in what spot we choose. The reason we have the ability to put any nucleotide in any position, without respect to how well it will bond to another nucleotide, is in the way the nucleotides are strung together in sequential order. The way that RNA connects to each other in a physical chain regardless of the type is by binding the 3' binding site of the starting RNA to the 5' location of the next one. In wikipedia you can see an example to follow as I explain the ' or prime positions. We will use the picture of parts of RNA molecules below to demonstrate. The picture below is only just a part of the RNA molecule as a RNA molecule consists of the backbone and the nucleotide part. What you see below is the backbone, and this consists of a sugar molecule that is attached to a phosphate group. There is also a nitrogen-containing base location visible and this is where the nucleotide base is located. This picture is thus of a complete generic RNA molecule with no specific nucleotide base assigned.

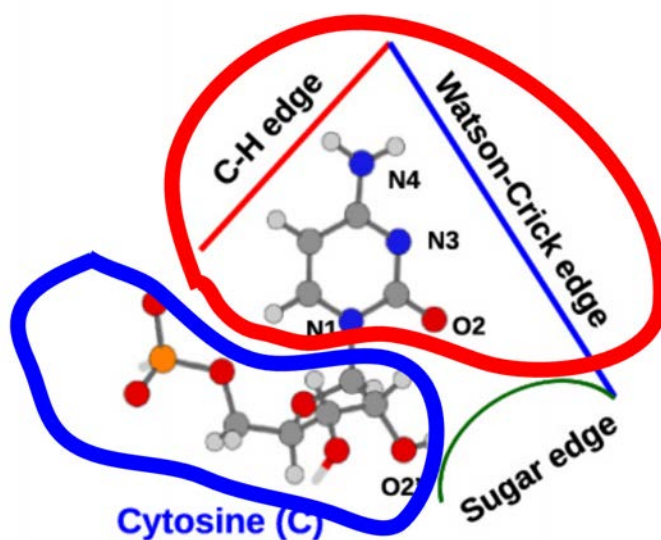


Note how the sugar molecule has edges and points as part of its hexagonal shape. Each point is labeled with a prime number or ' and this simply is a way to quickly reference which point we are talking about as each one has a different makeup and this is what is being referred to when someone says 5' (5 prime) or 3' (3 prime). From the diagram we can see that the phosphate group is connected directly to the 5' point. This phosphate group has a single oxygen that actually is what connects this molecule to the next molecule's empty 3' point. The contents of the molecule discussed so far in this section is what makes up what we call the "backbone". This backbone then is what the nitrogen-containing nucleotide base is attached to, and this point of attachment is the 1' location on the backbone. The thing that makes an A, U, C, or G unique is this base that mounts to the backbone as every molecule has the same backbone structure.

Now going back to the hydrogen bonds each base has different atoms at each hydrogen binding site so thus not all bases can pair with each other which is why there is a fight for selection of strongest forces when the sequence starts to fold. The question then now become how do these hydrogen bonds work and that I think is an important topic to discuss as it will inform thoughts on how the RNA is able to twist and turn, as well as its limitation in those areas. This part of the

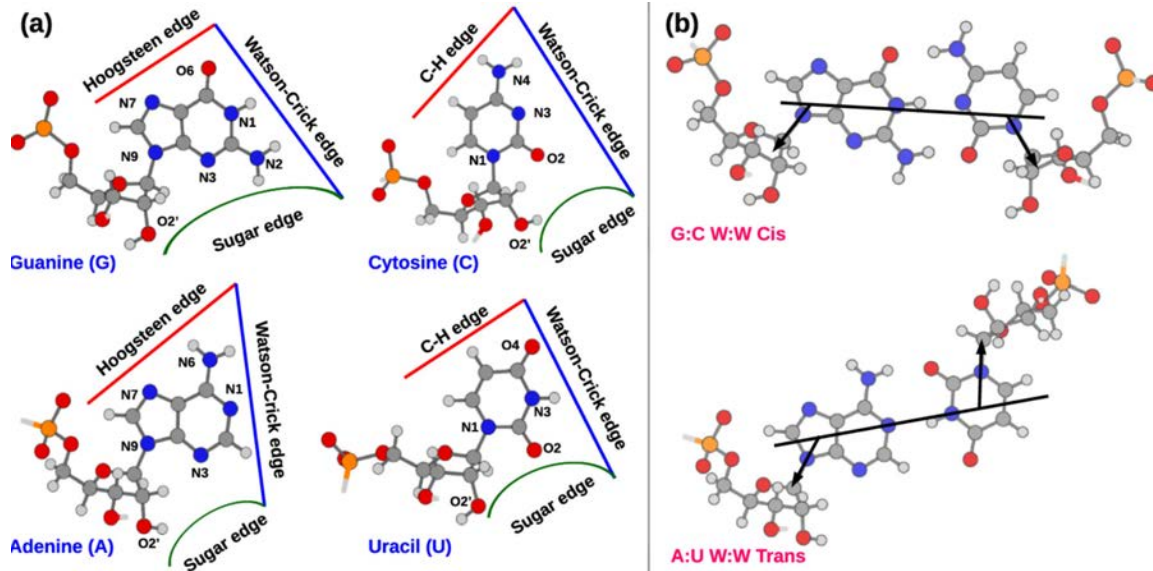
discussion will rely heavily on snippets from research papers to help provide supporting examples.

We will start this discussion with the Cytosine molecule as pictured below and then show a CG bond with its very strong bonding characteristics. We will then use this to jump into how the other nucleotides are arranged, as well as how many nucleotide pairs actually have multiple binding configurations that lend themselves to various mechanical configurations. RNA actually has 3 sides of the nucleotide base that can bond with other nucleotides, those are the C-H or Hoogsteen edge (Hoogsteen is for purines and C-H is for pyrimidines), the Watson-Crick edge, and the sugar edge ¹. The image below is from the paper in the footnote and you can see the backbone circled in blue and the nucleotide base circled in red.



In theory you can have a successful bond between any two nucleotide bases along any of the 3 binding sides as long as the acceptor and donor binding sites line up and that is part of the difficulty in RNA design I think as you need to get this RNA to line up just right, but some base's have multiple options for the a single nucleotide base pairing and these I think fight each other a bit. If you look at a diagram from the paper on hydrogen pair strengths you will start to see what I mean by this.

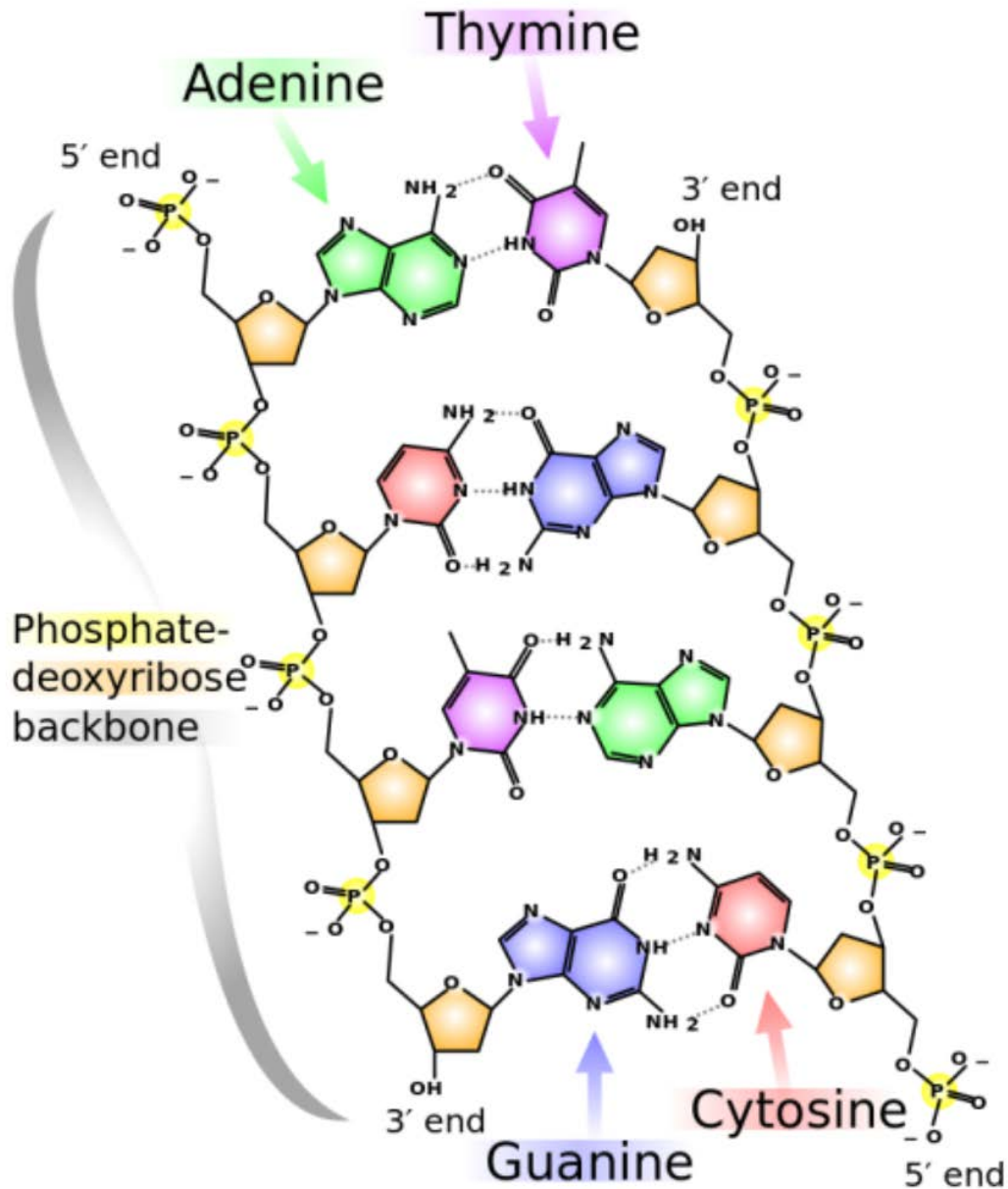
¹ Halder et al, "Estimating Strengths of Individual Hydrogen Bonds in RNA Base Pairs: Toward a Consensus between Different Computational Approaches", ACS Omega 2019 4 (4), 7354-7368



In the diagram above you can see section (a) has all 4 types of RNA we see. The red dots are oxygen atoms, the blue dots are nitrogen, and the whitish-gray dots are hydrogen. Now the oxygen and nitrogen are considered acceptors and the hydrogen is considered a donor. A nucleotide base bond between two bases occurs when a donor from one base and an acceptor from another base line up and are attracted to each other sufficiently. The more sets that line up and have an attraction the stronger the bond for each pair of bases, and this correlates to the strengths we see in how bonds form as CG has the most bonds and the GU and AU have less and they are weaker pairing from experience.

The CG pair needs 3 bonds to form and is the strongest of the pairs when formed as the Guanine has 2 hydrogen sites and 1 oxygen site along the Watson-Crick edge, while the Cytosine has 1 hydrogen, 1 oxygen, and 1 nitrogen across the Watson-Crick edge. Thus the G has 2 donors and 1 receiver, while the C has 2 receivers and 1 donor and this match up in the 1-to-1 match of donor and receptor, in the correct order, makes a good bond. Now, if we look at the rest of the edges of the Guanine and Cytosine we see something interesting that helps us understand how the GC is so strong. The Guanine's Hoogsteen edge has 2 receivers and 1 donor, and the sugar edge has 1 donor and 1 receiver (i'm not counting the free hydrogen on the oxygen for the backbone as that is only in play at the end of a RNA sequence), while the cytosine has 3 receivers in the C-H edge and just the single receiver on the sugar edge. There are no other ways for a CG pair to form generally other than the Watson-crick plane. What is interesting is that Guanine's Hoogsteen edge has a similar orientation as the Cytosine's Watson-Crick edge, and this means that the GG pair is a potentiality but it would have a weird geometry sticking out weirdly. I think that this type of behavior is taken advantage of by pseudoknots and may be what allows them to twist and turn like they do. Hadler et al says "On the basis of the identities of the interacting edges, Leontis and Westhof have organized all these base pairs into 12 geometric classes, viz., WW, WH, HH, HS, WS, and SS (cis and trans). (8) It is theoretically possible to have a total of 144 different types of base pairing geometries. If you look at the Adenine and Uracil with this in mind you can imagine the number of possibilities for

orientations. Putting this all together you can see how the nucleotides in the example below are arranged and how not all the donor-acceptor sites match up. For reference Thymine and Uracil have the same orientation of nucleotide sites and backbone. Thymine is found in DNA and Uracil is found in RNA.



To swing back to the discussion on mechanics of keeping how the hydrogen sites create the bonding forces, let's consider how a RNA strand is joined together as a big long string. This long tangential explanation about the hydrogen bonds and bending moments was so we could better follow and understand what is going on in the RNA before we take this next giant step. This big

long string of molecules contains forces that are influencing it to not fold into a some shape that is not necessarily a single structure as there are multiple options to choose from. As a RNA hairpin loop forms there are forces that struggle to straighten out the strand, but it is the stacks and the stack energies that are keeping it together when it does fold. The stack energies exist because of the hydrogen donor-acceptor sites for each nucleotide base. It is these energies that are acting like glue or bolts to hold the hairpin loop and stack.

In this hairpin loop there thus exists a negative bending moment at all times when nucleotide arraignment is not optimal to form a stable structure. This is because the bonds that hold the hairpin are acting like a lock of sorts. I believe that as soon as the bonding is sufficient enough the design no longer will exert a force that will tear it apart in essence. I believe that this is where RNA gets tricky, think of it as a complex system that when not in motion and and all bonds are perfect it sits in a state of perfect equilibrium. In this state it can be thought of as not having any forces exerted or acted upon, so it stays stable. Once there is a perturbation, such as something that might change an energy delta somewhere, a new mass for instance or maybe a random mRNA oligo with its own energies and thus mass, will force it to change shape as the forces are now no longer in equilibrium.

RNA seems to want to obtain and stay at a state of equilibrium, and by equilibrium I mean that there are no forces against or acting on it. Thus if a RNA sequence was just A's it would be a belt swinging around into whatever tertiary structure it wanted, but it would not form a base pair and thus no energy pulls. It thus is in a state of equilibrium in energy and thus forces I think, as equilibrium is the point where an object stays at rest with no outside action and that is analogous to the single-body example I think. So since no hydrogen bonds are actually made, the hydrogen site of one base may attract another same base for a little distance but there is nothing that will form a good bond and thus it can't bind well and it stops. This is still not generating a bonding force and thus maintains a relative state of equilibrium in my opinion. If anything it would in theory just wiggle around. Once the nucleotides are mutated then the equilibrium state shifts as there is overall energy shift that is the result of a cascade of changes. This results in movement and a pull of the sequence in on itself thus lending to the most simple object in RNA being the hairpin.

The hairpin is the most simple and consistent structure, as there is no other way for a RNA strand that is short to exist with the forces present without forming a hairpin. This also informs us about why RNA tends to have so many loops, as the RNA needs to wind its way around and a sharp bend is not energetically beneficial, unlike an arc as part of a bending moment. The sequence has to loop its way around to form the bonds it is attracted that will allow it to come to a point of rest via equilibrium. Thus a negative bending moment is observed at the hairpin loop due to normal positive axial forces and this is the RNA strand trying to push itself straight, and a positive bending moment is observed due to the negative axial forces from the hydrogen bonds. The signs of the forces then cancel each other out at the point that equilibrium is reached.

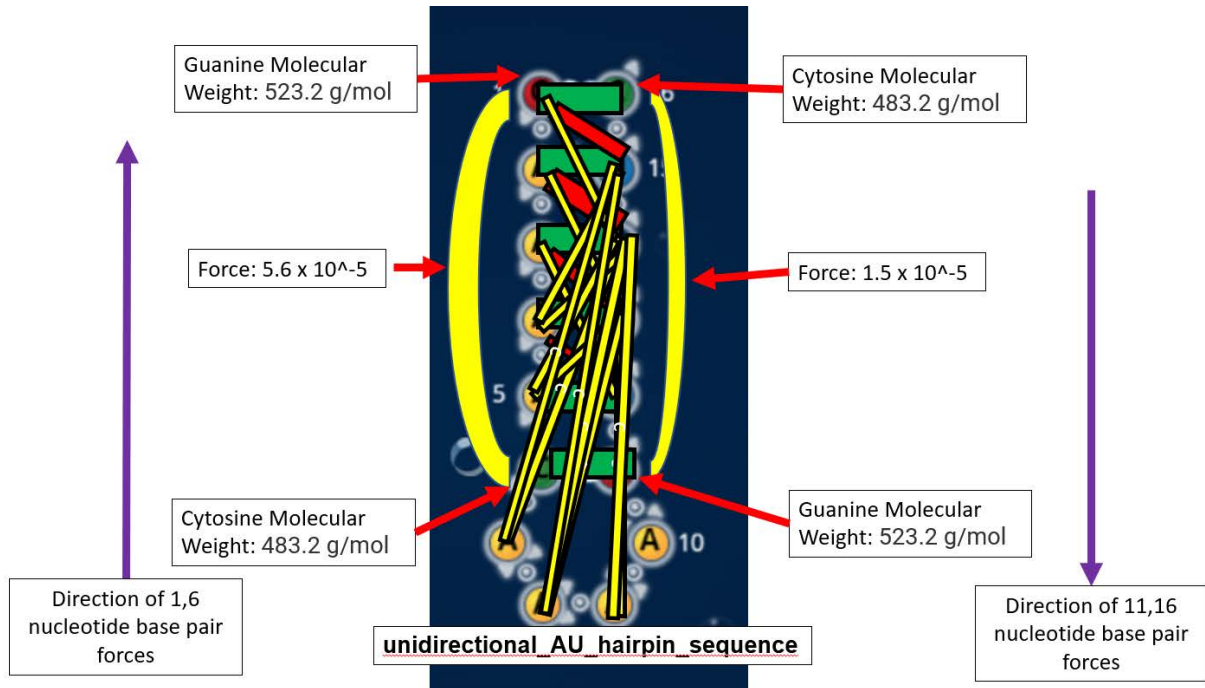
Now the two nucleotides are pulling the strand in and toward a middle center line and this is where the forces are focused and they pull the base of each nuc towards the other. As this force

is moving from a direction that goes in to out of the solid object it is considered a positive axial force. This positive axial force and the negative bending moment explanation do not violate the 1st, 2nd, and 3rd laws of physics, but are indeed supported by them. This concept I believe can thus be expanded further to explain all of RNA's behavior including switches. The one force we have not discussed yet is shear forces and their discussion will further support this assertion.

Shear forces are defined as the algebraic sum of all the transverse forces acting on either side of the section of a beam or a frame. The phrase "on either side" is important, as it implies that at any particular instance the shearing force can be obtained by summing up the transverse forces on the left side of the section or on the right side of the section. What this means is it's a force where you have two opposing axial forces pushing on opposite sides of a structure in a way that tries to split it in half. We will encounter this force in a little bit as we go through the transformation of a hairpin through various nucleotide configurations.

Now keeping in mind that the axial forces at each nucleotide base keeps the base bonded and resist the forces of the opposing bases that are pulling the sugar molecule that connects the string that connects to the other bases, why is that nucleotide pair 6,11 in the stack expected to fail from our earlier discussion in the chapter?

If you look at the level 2 filtering images and data you can see that 1,6 and 11,16 are predicted to have forces between them. Specifically 1,16 is predicted to have a pairing probability of 0.000056 and 11,16 is predicted to have a pairing probability of 0.000015. Now they are low but it is notable that 1,6 is the 2nd highest probability in filter group 2. Now if we also look at filter group 1 data, we see that the 6,11 nucleotide pair has the second weakest predicted probability in the filter group, as the 1,16 pair having the weakest. The 6,11 pair has a .95 prob and the inner stacks have .99 and such, the 1,16 pair has a .93 probability. That is a lot of numbers and it may seem like they don't mean anything or are a jumble, but they are important and paint a bigger picture for us. Lets look closely at this and very observantly scrutinize every force and let's look at a diagram to help.



Looking at this diagram of the known variables we see that the left side or the 5' side of the RNA sequence has a force that is being exerted between base 1 and base 6 in the direction of the 3' end, so from the hairpin loop to the end of the stack. The other side of the RNA has a direction of forces that is opposite. I make this assumption due to the masses of the nucleotides and the fact that a greater mass has greater resistance to change and thus it is sound to suppose that the nucleotide with less mass would move first or fastest if not both and thus the greater amount of forces would be in that direction of travel as force is mass times acceleration.

The same applies then to the right side for why the forces are moving in that way. Now we will be looking at some very small differences in pairing probabilities but the important thing for now is to think of things more in the sense of a vector in that there is direction and magnitude of speed. It is helpful to think of these probabilities in such a way and it is convenient and relevant that forces are considered vectors in physics when performing the calculations for forces.

So we see that the 1,16 has a pair prob of .93 and the 6,11 is .95 and for reference the inner stack have probabilities around .99 which is basically superglue levels in RNA terms. Now you would assume that since 1,16 has a lower pairing prob it would thus fail first, this is a good thought and would be correct, I think, if the structure of the design was different. However, due to the directions of forces and such the 6,11 fails first according to the order of the subopt calculations secondary structures. Let's see how this subopt calculates that the first inner nucleotide pair of the loop will unbind first at the 6,11 site. You can see that there are first filter level probabilities for the 1:15, 2:14, 3:13 nucleotide pairs and so on that are showing there are forces there pulling the 3' nucleotide segment toward the 5' segment but also down at the same time due to the forces from the nucleotides that are forming attractions at what appears to be 45 degree angles that are shown as red bars.

There should ideally be an equal force in both directions for a stack that creates a normal bond with axial forces, but the direction of the forces shown as red bars from the end to the hairpin loop is creating shearing forces or at least shearing force potentials. You can see how the predicted probs for these angled attractions decrease in magnitude the closer you get to the hairpin due to what I think is an equilibrium of forces.

So these shearing forces are applying a small bonus axial force which is additive but has in turn decreased the overall structural strength potential Think local vs global even at this small size. I think this axial force provides a bonus of sorts at this time to the 1:16 pair compared to the 6:11 pair which has just the axial force and is very close to the negative moment of the loop. So with the 6 nuc being more willing to move due to its mass and no 1st or 2nd level forces acting on it that would create a reinforcing or interfering force that would act like a temporary reinforcement it should not require much to lose that 6:11 bonding. The result of all of this is that the 16 and 11 nucs both have strong pairing configuration and forces, however since the 1:16 has a bonus and the 6:11 does not have one the 6:11 pairs gives up its bonds first with the pull from the 1 nuc on the 6 nuc we observed. This now is exactly what was predicted to occur in the first stage of the transition of the hairpin we observed as we looked at the next secondary structure in the subopt alternate structures. Now we need to remember that the only way we reach these structures in the ensemble is by watching as the RNA travels across different energy levels. This means that the energies and bonds present will be different and we know that the closer we move in a positive free energy direction the weaker the RNA bonds and at 0 free energy we have no binding until we start going positive according to the energy models.

So we have the assumption that RNA is a string that is trying to pull itself out straight and this would only happen as the energy levels went in a positive direction of change as the RNA bonds would need to fail. So if the bonds are starting to fail then it makes sense that the next step would have the bond at 1,16 start failing more in the progression to the unzipping of the hairpin which you can see that the forces already seem to desire to do that. So imagine if you will that the following is what is happening.

The 1:6 bond fails and the shearing forces start to act which will push the 3' down toward the hairpin and whatever forces are there. The 1:16 bond fails and the downward forces then cause the 6:11 pair to pop back together. This is due to the already present equilibrium of forces so the distance would be equalized as well. If those end nucs pop out then that would cause equalizing forces that would push it back in place until it stayed. Now this is only a small step until the next energy level shift and the design will now be at a 0 free energy and folds open. This is because at this time there is no potential for any deltas of energy and the entire structure is at an equilibrium of 0 energy forces... and thus a straight line. Then to take it further, as it goes into positive energy it starts to fold in on itself again and is now zipping itself together in the direction from hairpin loop to end of the stack and thus the stacks that are predicted to come close to the loop, along with a bunch of the end nucs being unpaired as a result. It's important to remember that these structure are just snapshots during its transition and we need to understand why it's moving in this way to understand RNA fully and mechanics gives us the tools to interpret and

Speak about what is happening behind the curtains so to speak, These early concepts are essential in understanding the next steps we go through while we observe how the structure changes shape as it transformed through the many shapes of the ensemble.

Chapter 5: Observation of the transformation of a hairpin towards mechanical equilibrium while adding single nucleotide *in silico*

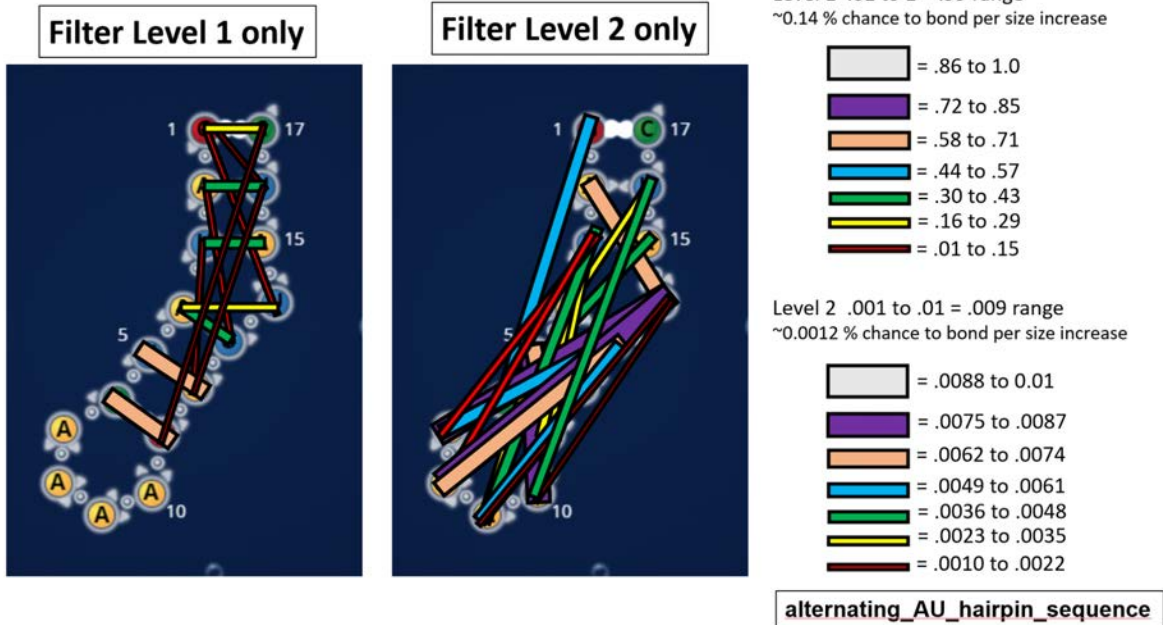
In the last chapter we gained an initial understanding of how you can use the subopt structures found in the ensemble to gain a feeling for how stable a design is and a teasing of what behavior the RNA strand will take on. For instance, in the example we closed with, you we saw that there seems to be a predisposition to having a shearing force along the side of the stack that has the 1 - 5 nucleotides. This shearing force's presence seems to be backed up by the explanations and with physics as to how the secondary structures in the ensemble look the way they do. I would now like to help you take the next step and see how manipulating the nucleotides reveals that our assumptions were correct and informs us of the behavior of RNA mechanics. This will hopefully flip a switch that helps you see RNA in a whole new way with ease and incorporate the theory of mechanics into your interactions with RNA molecules. I will start this exploration with data from in silico calculations. The next book will be a follow up to this book, containing real world examples and experimental data with shape data as well.

Lets jump straight into this by asking "What will happen if we add a single A nucleotide in between nucleotide 13 and nucleotide 14. Initially this results in what was nucleotide 14 advancing one number to now become nucleotide 15. This occurs up the line to the last nucleotide in the sequence with each one after this increasing the base number by one. Now let's interpret what is happening based on pairing probability data, and while we do it let's separate our answers accordingly for alternating and unidirectional sequences. We will base our assumptions on what we learned from the previous chapter with the alternating AU sequence being our launchpoint for the discussion.

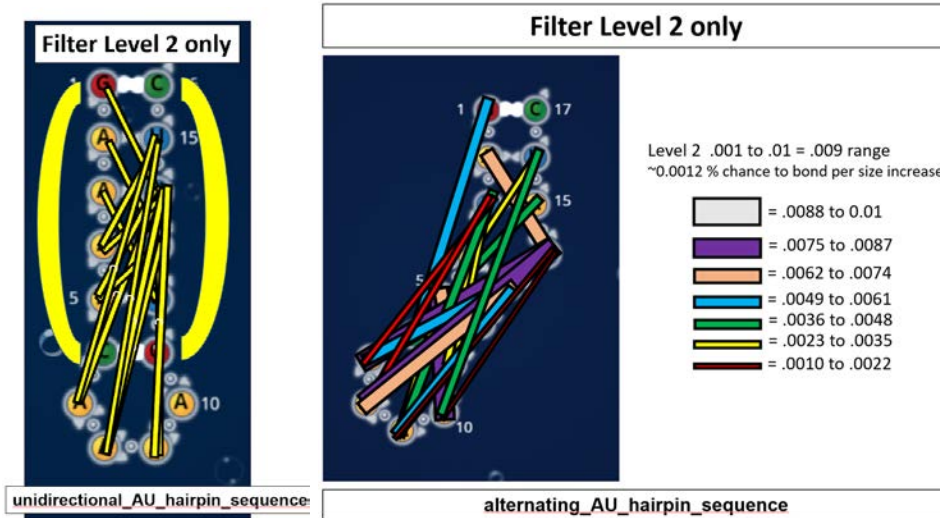
For the alternating AU sequence we would think that the RNA would want to stay in a solid bound hairpin as the predicted forces do not indicate any imbalance and quite strong lateral forces are predicted between the bound nucleotides predominantly. There are also relatively few Level 2 errant forces that would act on the RNA to influence alternate secondary structures in the structure. Now if we add the nucleotide the next image is what this new sequence MFE secondary structure looks like.



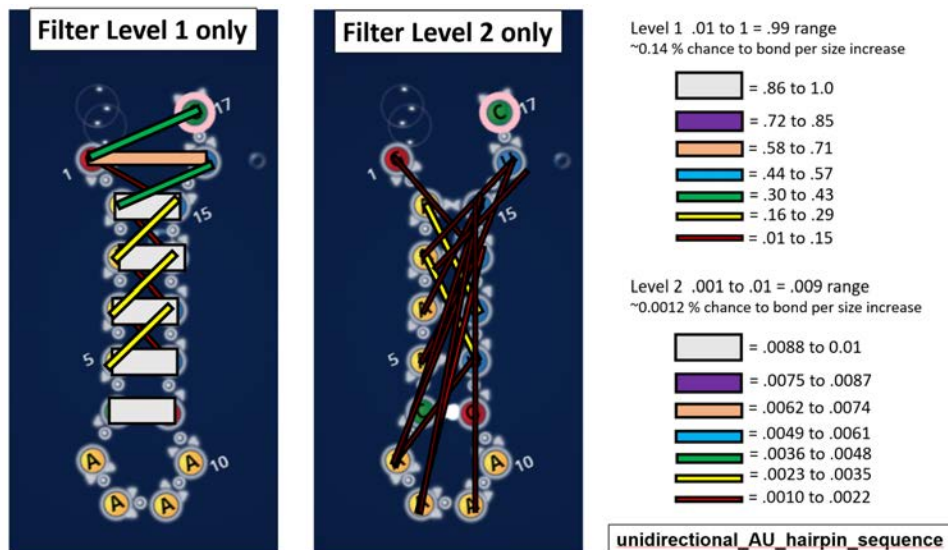
You can see that it now appears to be a stack with a single bulge at nucleotide 14. This is the same nucleotide we just added and the design forced that nucleotide out of the stack and there are no level 1 bonds that are lost as 3:15 is still valid as it was once 3:14 and 4:13 is still valid. What ended up happening is that now we essentially have two stacks on a hinge with a single nucleotide that is unable to fully pair with another base pair. It has essentially been rejected by the hairpin, but it is held in by the forces from the potential 4:14 and 1:14 bonds. These two forces are opposite each other at an angle and they seem ideal to cause the design to hold on to that lone A nucleotide. Here is the force data from nupack that I have interpreted for you with colored bars of specific sizes to represent the predicted forces present.



What I think is interesting about this design (and this is getting into a bit more of my theories) is that while the hinge did in fact create two different stacks, it still seems to act as one single long stack depending on how the forces are interacting. In fact if you compare it to the unidirectional AU sequence's Level 2 you can see that both of Level 2's look similar in that the nucleotide number 14 has multiple attraction forces with nucleotides in the loop. Nuc 1 is even attracted to nuc 4 still.



What if we were to do the same thing to the unidirectional AU design? We already know it is unstable and have an idea of what it will do if it loses any more stability, so let's see if our theories pan out.



You can see that what happened here is that the hairpin did not create a bulge like the alternating direction design. Instead it caused the bonds at the 3' and 5' ends to break and nuc 17 has shifted up and nuc 1 has moved away. The frame of reference for nuc 1 is that nuc 17

moved away from it, but for nuc 17's frame I think that nuc 1 is observed as moving down. This shape is exactly what we had predicted. Nuc 1 was pulled away from its previous bond with nuc 16 which is not nuc 17. Also once that bond was broken the two sides of the hairpin stacks shifted in the same directions as a shearing force. We had predicted shearing forces and the pull of nuc 1 to nuc 4, as well as nuc 16 (now nuc 17) to nuc 11. The forces from 1:6 and 16:11 in the original unidirectional design were pulling the 1:16 nuc pair apart. Once nuc 16 was forced to shift up one, the bond could not be made strong enough and thus they broke. We saw that the 1:16 (now 1:17) pair in the alternating AU design after the nuc addition was able to maintain the 1:17 bond after the addition. I hope that this helped you understand how to interpret and read the pairing probability for RNA and thus the mechanical forces present in the structure that affects the proper folding of the design. My one desire for the outcome of this document is that when you design or analyze RNA, you remember to keep in mind the physical forces present, and that you allow these thoughts on the Mechanics of RNA to help inform your decisions in a way that helps you design a molecule that behaves in the exact way you intended. This will help allow us to generate new medicines and medical tests that may help save many lives.