A mechanical overview of RNA in an energy focused world - Draft rev1
Author: Jennifer Pearl
Date:5/20/2022

Chapter 1: What is the ensemble?

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 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 silica* and *in vivo* testing. To start with, I truly believe that RNA is a mechanical system that lives in an energy world, and thus us bound by the rules of mechanics of materials more than the effects of the energies in the loops. I believe that this is why riboswitches are so difficult. When a switch is made there are shifts in energies and the hydrogen atoms that bind the nucleotides together are pulled apart and pushed together somewhere elses they have a strong attraction. That shape they like to take 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”. This 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. As you move further from the MFE or lowest theoretical energy state you can get, the number of possible secondary structures increases. This is because the overall energy state of the design has changed so teh 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 were too weak are just now strong enough to pair up, and thus predicted secondary structure can be different by orders of magnitude or just slightly. Now with this in mind, the overall energy state of the RNA is calculated with 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 my response of “usually kinda but not really when you think of it” makes sense. When you think of this if you focus all your energies on the MFE or just energies you may miss something. It is more of a kaleidoscope of what could happen, than a dial in type thing. You have to think of it in terms of at any one time all the RNA nucleotides are tugging and pulling on every other molecule. It does not matter how far away they are as there are always tiny forces. These tiny forces are represented in teh 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-crimes canonical base par 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 probability's. I'm not really sure at what point energys 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. It is with certain geometry's 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 this increasing it's 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.

The application of analysis of pairing probabilities 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 pairing probabilities calculated by NUPACK and Vienna 2 that resulted in all the winning designs submitted by Sara during the DOE. There are a few ways to look at the data.

Chapter 2: Analysis of Partition Function: Paring probabilities and the ensemble

My 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. The most important aspect I think is the ensemble, as the only standard metric that is currently part of Vienna2 and NUPACK and this directly citing a paper that I have ever seen track with a design score is ensemble diversity or ensemble defect. In fact when I first noticed this back in 2016, I did a search into the old Etrna 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 the ensemble.

Now why start a chapter on pairing probabilities with ensemble defect rant? The answer is that pairing probabilities essentially determine the ED and the ED really points to a RNA's ability to fold right and/or just be stable. 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.

What are these pairing probabilities that are so important and how do they work? Now remember that RNA is bound by nucleotide pairs whose bonds are generated by hydrogen atoms and the number of atoms is the strength of the bond. The number of hydrogen atoms is determined by the nucleotide type and each type has a specific number with each Watson-Cricks base pair having stronger or weaker bonds corresponding with the bonding we see in Eterna.

When i think of pairing probabilities i think of the dot plots you can look at in Eterna as those are a graphical representation for the pairing probabilities data. A dark spot is a strong predicted bond and a light spot is a weak pairing and a white spot is no pairing. The data outputted from NUPACK and VIENA2 is formatted as numerical values representing the statistical calculation of chance to bond. It is normalized to the value of 1 with 1 being a very strong bond abd the smaller the number gets the weaker the bond abd likely to pair. 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, not large mind you, but enough to nudge the design. When we get

the MFE secondary structure we are then looking at the most probable shape it will take at the lowest and thus strongest energy levels that can hold the RNA strong. 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 kcal delta as well as find out how many possible secondary structures there are. Remember I said a few times that pairing probabilities define the secondary structure, well you can see that in action now and understand what is happening.

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 the secondary structure vary depending on the delta from MFE, and the design will fluctuate between all the possibilities secondary structures in that kcal delta group of the ensemble, and if it goes back to the MFE it will be back at the MFE secondary structure. Now some designs lend to very few different alternate secondary structures and some lend to a lot 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 by mechanics of materials for dynamic loads.

Chapter 3: Why is mechanics of material so important if it's not till chapter 3? Also, what is mechanics of materials and how does it explain RNA behavior?

Mechanics of materials is the study of the motion and movement of objects unders dynamic load. What this means is it explains how the spring action in a spring is happening from the trigonometric and mathematical perspective. When a 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 mechanics of materials. 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.

So the reason I say all this is that, if you have ever seen my work, I strongly argue that RNA should be thought of more as a mechanical object than an energy object, and we should consider how mechanical systems work when evaluating a design. A large part of this chapter will actually be 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 this would be a good example to use as a case study for my earlier thoughts and even dive into some more thoughts and theories I have.

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. 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.

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 model that change. 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. The critical mass in this instance is when the design has a just low enough negative free energy to want to pull together and not form a straight line if we were to start with all A's in the nuc string for example. 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 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 teh 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 energys. 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 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 quit 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 beind 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, its a bit of a thought experiment if you will.

Chapter 4: 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 to explain stack energies as to do so without the loop bit is very similar to teaching people everything about the classification of animals, but always telling people to ignore the swingy bit at the end as that serves no useful purpose. I hope to demonstrate why much of our practical knowledge works for in game energies and thus allow us to progress to much more complex systems. Mechanics of materials 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.

In order to fully understand what is happening

Imagine that you have a length hairpin with alternating GC closing pairs, alternating orientations of AU in the rest of the stack. This design would fold well in Eterna. Let’s do a comparison of15nt long hairpin with the primary structure of “GAUAUCAAAAGAUAUC” that we will call teh alternating_AU_hairpin_sequence and ”GAAAACAAAAGUUUUC” 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.o using rna06 you get the nucleotide pairs pairing probabilities that are greater than non-zero as well as 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

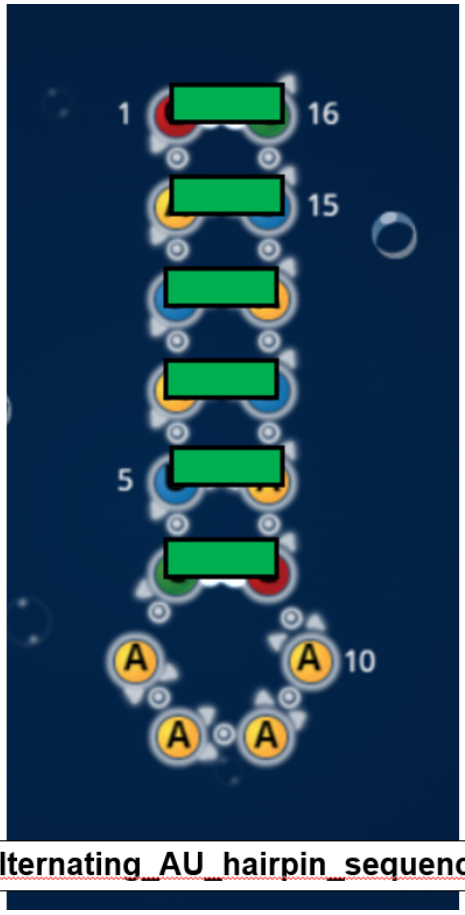
Kcal_Delta=2, EV_old=0.0625, EV_old_norm=31.25, EV_new=100.0, structure_count=2
Ensemble kcal delta secondary structures and energy's
1: ((((((....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
2: (((((.....))))) FreeEnergy: -1.20: Stack Energy: -1.20

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

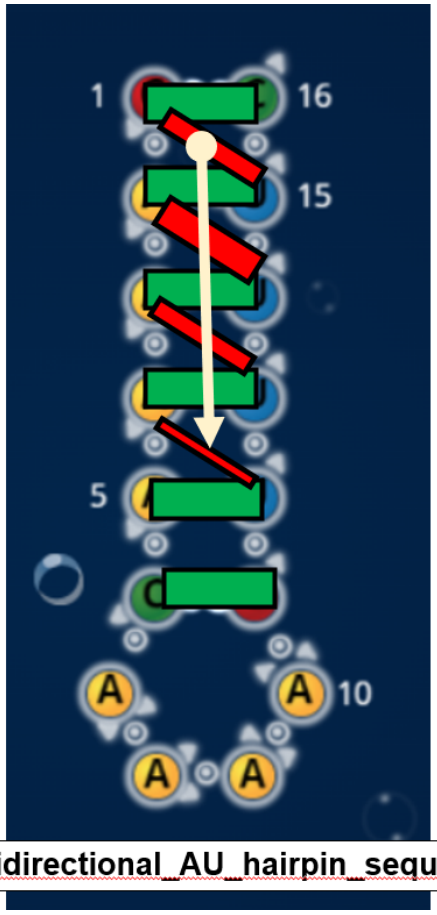
Kcal_Delta=4, EV_old=0.21428571428571422, EV_old_norm=107.1428571428571, EV_new=342.8571428571428, structure_count=7
Ensemble kcal delta secondary structures and energy's
1: ((((((....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
2: (((((.....))))) FreeEnergy: -1.20: Stack Energy: -1.20
3: .(((((....))))). 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 paring 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 wanted complementary base pairs that necessary to properly form the hairpin represented as green squares and the intensity of the predicted pairing probability and thus bonding strength is represented by the thickness of the bar. The undesired forces are represented as red bars with the same rules.

In the alternating_AU_hairpin_sequence you can see that the pairing probs at the filtering level we have chosen are only for the nucleotides in the only stack for the target structure. This is the definition right now of a preferred pairing. When you look at the unidirectional_AU_hairpin_sequence you can see that there are teh same preferred pairs, but there are also predicted pairs of nucleotides shifted one down for nucleotides 1 through 4. These are causing internal forces and this is the point where mechanics of materials shifts into gear to explain the behavior of RNA after the correct energy levels have been reached to establish the preferred pairing.

Think of each bar as a bolt that is used to fasten a piece of metal to another piece of metal. But in this instance the middle is the weak point as that is where the hydrogen bonds are formed. The hydrogen bonds forming a connection or a pair introduces a structural rigidity to the RNA. 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 the hydrogen molecules. Considering that all the nucleotides are bolted down so to say, the entire stack is now a single physical load with a mass from the nucleotides and the rigidity of the bindings. We can ignore the loop for now.

Think then of the undesired forces or undesired pairing probabilities in this design as internal forces. You can see how they 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 secondary structure for each group. That means that there is only the MFE for both 1 and 2 kcal deltas, which those ranges I found are a real good quick guide for stability. 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 is starting to increase now in the nupack data when there are conflicting pairing probabilities.

If you 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 teh 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

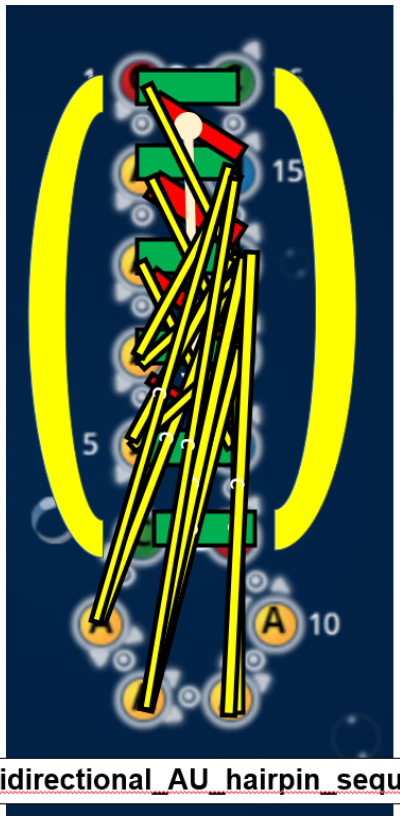
Kcal_Delta=4, EV_old=0.21428571428571422, EV_old_norm=107.1428571428571, EV_new=342.8571428571428, structure_count=7
Ensemble kcal delta secondary structures and energy's
1: ((((((....)))))) FreeEnergy: -3.20: Stack Energy: -3.20
2: (((((.....))))) FreeEnergy: -1.20: Stack Energy: -1.20
3: .(((((....))))). 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 this sequence, with the 4 kcal delta actually showing that some of the preferred base pairs won't be made frequently and the design is predicted to just not fold on 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 and 4,13 nucleotides are predicted to not pair in separate structures when you get to the 4 kcal group.

Now lets open up the filter of the pairing probabilities a bit for this design to see what is now happening in the background in what we traditionally call the noise in the dot plot or the light gray dots you see in them that are almost too hard to see.

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

So this will 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

Lets 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. We will go layer by layer, so think of the different steps in filtering we did on teh pairing probabilities as the layers. We first exposed very strong probabilities, followed by slightly weaker prob's by ones that started to approach the strength of the primary binding sites, then we exposed 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 prob's with nucleotides but are just not strong enough to actually form a base pairing that shifts it into level 3. Then there is level 1 which is the noise in the probabilities and this is the probes that of nucleotides that have a super weak attraction to anything that it could pair with, we will later see that often layer 1 is just a factor of nucleotide distance.

Lets assign 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 range. Super strong bonds like a GC pair in a stack will have around .9 for the prob but it could be lower, but the question has always been at what point does it exit the first layer and his documentation and practical experience show that .7 seam to be a community agreed on number and I kinda like it so its head cannon for me now. Level 1 then

would be any design that is less than 0.0001. This is 4 positions to the right of the decimal and this is super weak force. Anything in the middle of .0001 and .7 can be considered layer 2.

Starting with layer 1 we see this diagram below with pink large arrows that are horizontal that look like this below indicates the orientation of the nucleotide paring probability of the nuc pair it is under.

Sequence = GAUAUCAAAAGAUAUC
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

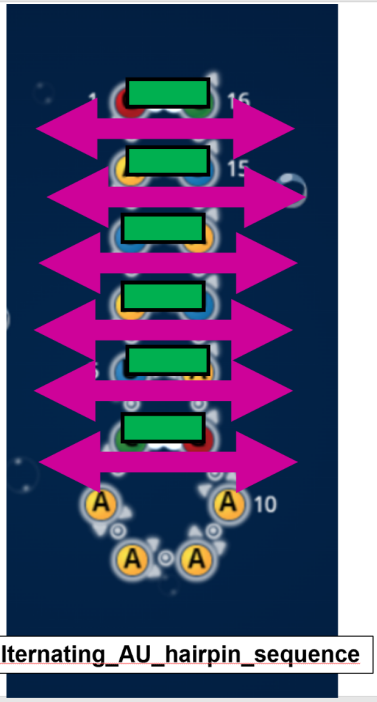
Kcal_Delta=2, EV_old=0.0, EV_old_norm=0.0, EV_new=0.0, structure_count=1
Ensemble kcal delta secondary structures and energy's
1: ((((((....)))))) FreeEnergy: -4.20: Stack Energy: -4.20

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



In this picture you can see that the 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. We know that the RNA strand is wanting to tear itself apart so we have predicted energy pairing probabilities that show why the RNA strand comes together and is able to take hold with specific nucleotide configurations. It is at this point that I want to start shifting how we talk about pairing probabilities and separate the components and really break it down to the laws of physics for a moment and give us the tools to further discuss this.

1. 1st Law
    a. “An object will remain at rest or in uniform motion in a straight line unless acted upon by an external force.” This is what is known as inertia and this rule is often called the Law of Intertia

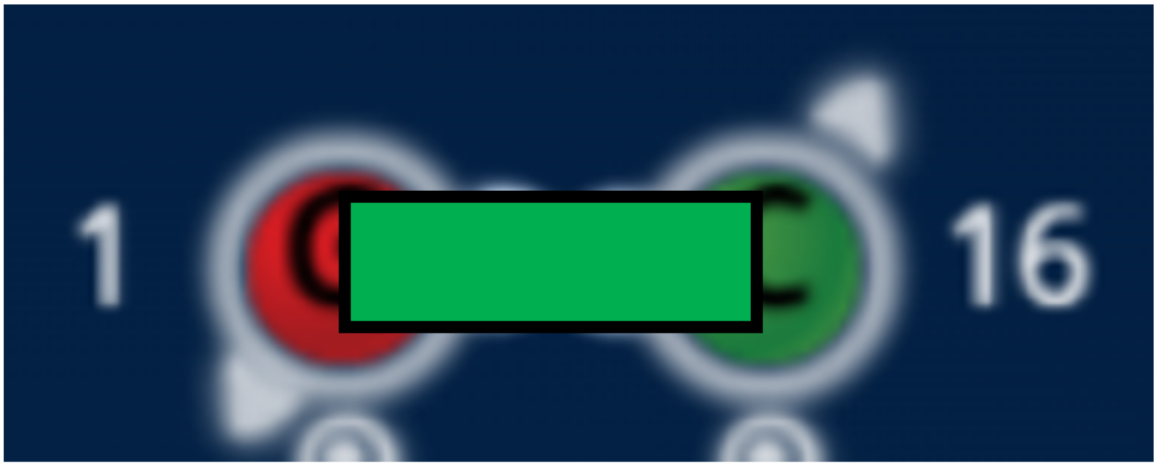
2. 2nd Law of how for
    a. Force is equal to the change in momentum over time. The time rate of change of the momentum of a body is equal in both magnitude and direction to the force imposed on it. Force is a vector and has thus magnitude and direction, mass is considered a scalar property and only increases or decreases the scale of the results of calculations. The equation for this if F=ma and this then shows that force is the representation of a mass being accelerated. The acceleration rate is a constant and never changes, it is however the mass that changes and thus scales the observed acceleration who’s 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.

3. 3rd Law. For every action there is an equal and opposite reaction. This explains how forces react when they interact with an object.

Now how does all this apply and why did I go into so much detain for the 2nd Law? Well lets 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).

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 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 start saying bond instead of predicted pairing probability, but still keep that previous phrase in your head. Let then now add the green bar to represent the bond of the nucleotides.



This is a now a representation of a theoretical single nucleotide base pair and thus a single base pair long stack/stem. There are no other bonds to worry about and the RNA will just stay there locked in based on energy levels. Let's now look at the molecular weights of the nucleotides

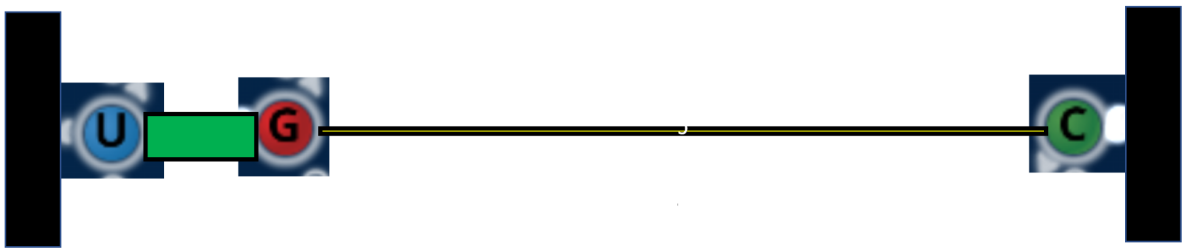
https://www.thermofisher.com/us/en/home/references/ambion-tech-support/rna-tools-and-calculators/dna-and-rna-molecular-weights-and-conversions.html

ucleotide	MW (g/mol)
Ribonucleotide Triphosphates (Avg. MW = 499.5)	
ATP	507.2
CTP	483.2
GTP	523.2
	484.2
Ribonucleotide Monophosphates (Avg. MW = 339.5)	
AMP	347.2
CMP	323.2

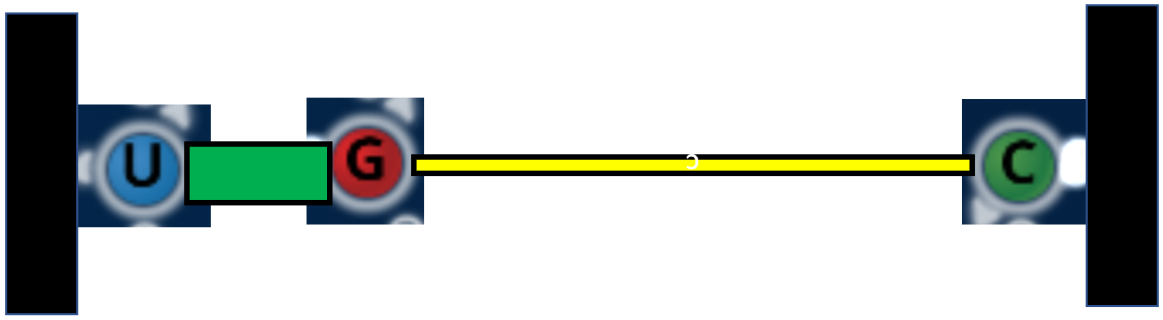
GMP	363.2
UMP	324.2

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 complentary 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 probabilites will decrease.

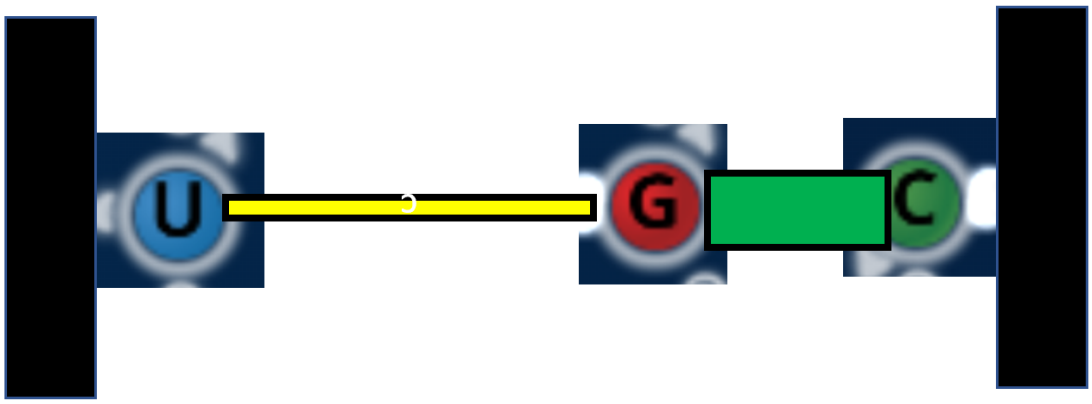
So thus as we move the new G closer and closer being sure to maintain the same plane and axial orientation the C will start to have a higher and higher predicted bond and at some point it will have a competing bond with the previous G and try to pair with the new one if the conditions are right. To better explain this lets use a different nucleotide configuration that is more logical suich 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.



The UG bond can be see 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 started 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



Now the bonding effects cause the G nucleotide to change position and move to the C nucleotide. A change in position over time is referred to a 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.

Now we have connected teh energy beavoirs of RNA to the first 3 laws of physics we can now use that to allow us to connect it more fully to statics, then, dynamics, and then mechanics of materials.