Natural RNA - What does it looks like?

I wanted to know what natural occurring switches looked like. Rhiju introduced me to the Rfam database and helped me getting started. Here is what I learned.

So far we eterna players have only seen simplified RNA sequences, not all the more weird base pairing. At bottom of this intro I have caught a few of these oddball cases from my eterna gaming viewpoint. I'm sure we will find more. Anyone who has additional examples, please add to this post or share why these basepairs happen.

I think this database is good start and shortcut to getting our hands on a little of the switch data we have been dreaming about.

Now go play... :)

Written November 2012

Eli Fisker

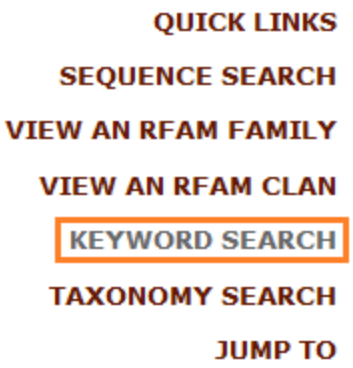
Document shortcuts

Rfam intro - Collection of natural RNA

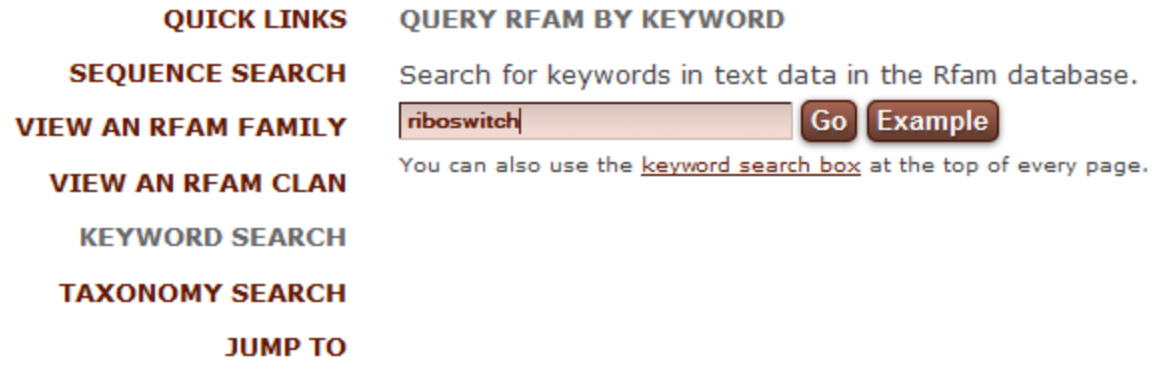
This is an an introduction on how to get a look at pictures of natural occurring RNA switches.

1) Open the database Rfam

2) Choose keyword search in the Menu.



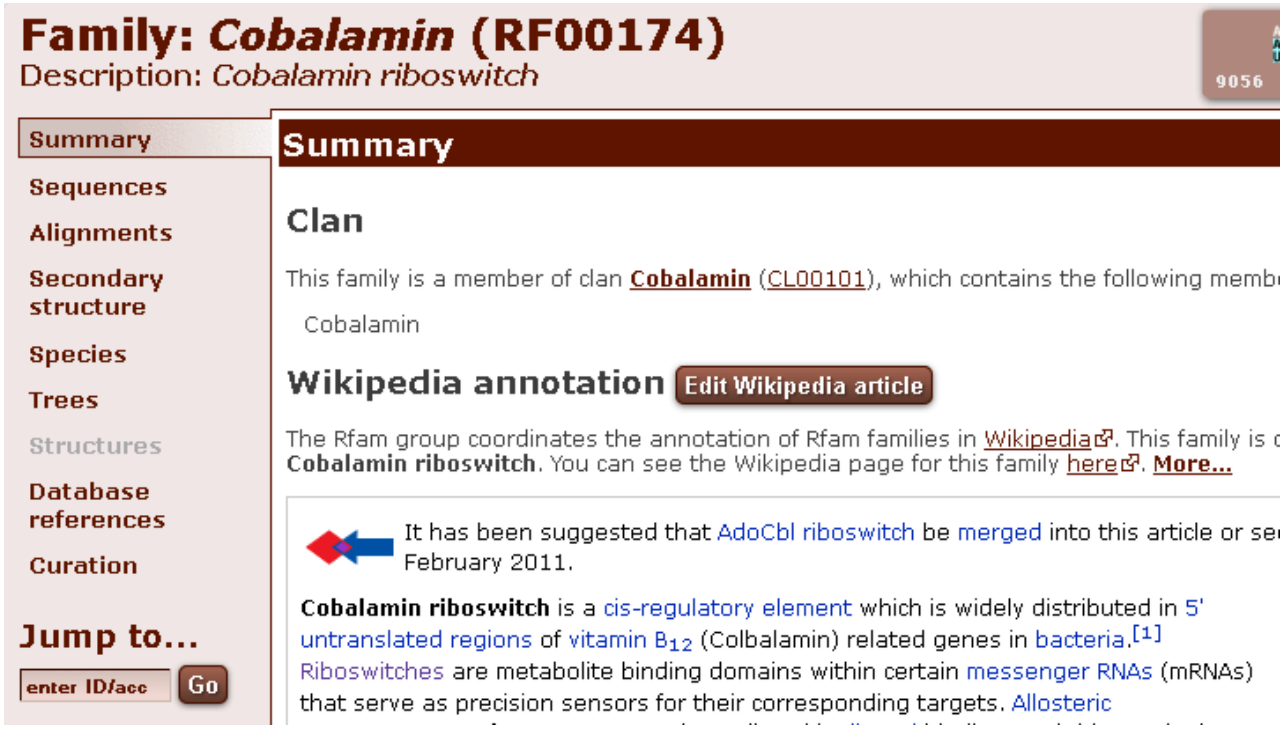
3) Find switches by searching with keyword "switch" or "riboswitch". Riboswitch will give you the biggest number of hits.



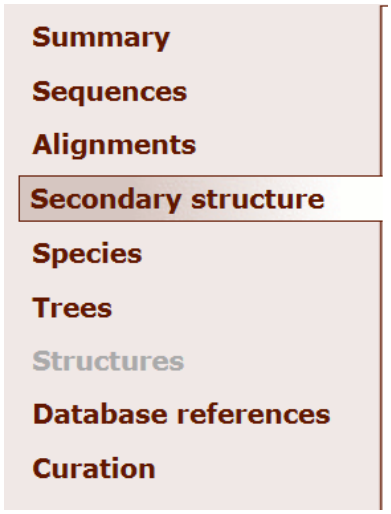
4) Click

5) Choose either an accession number or use an ID name.

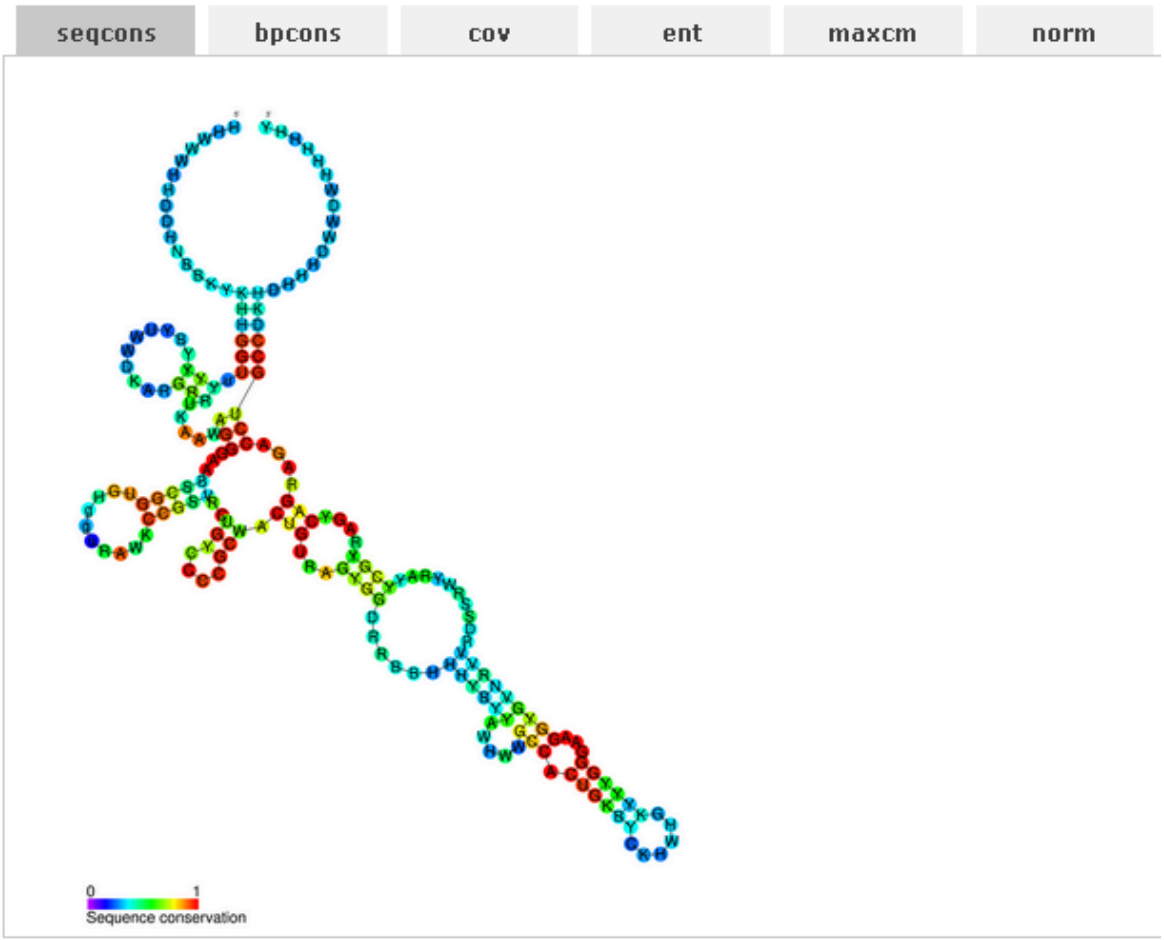
You end up in a picture like this:



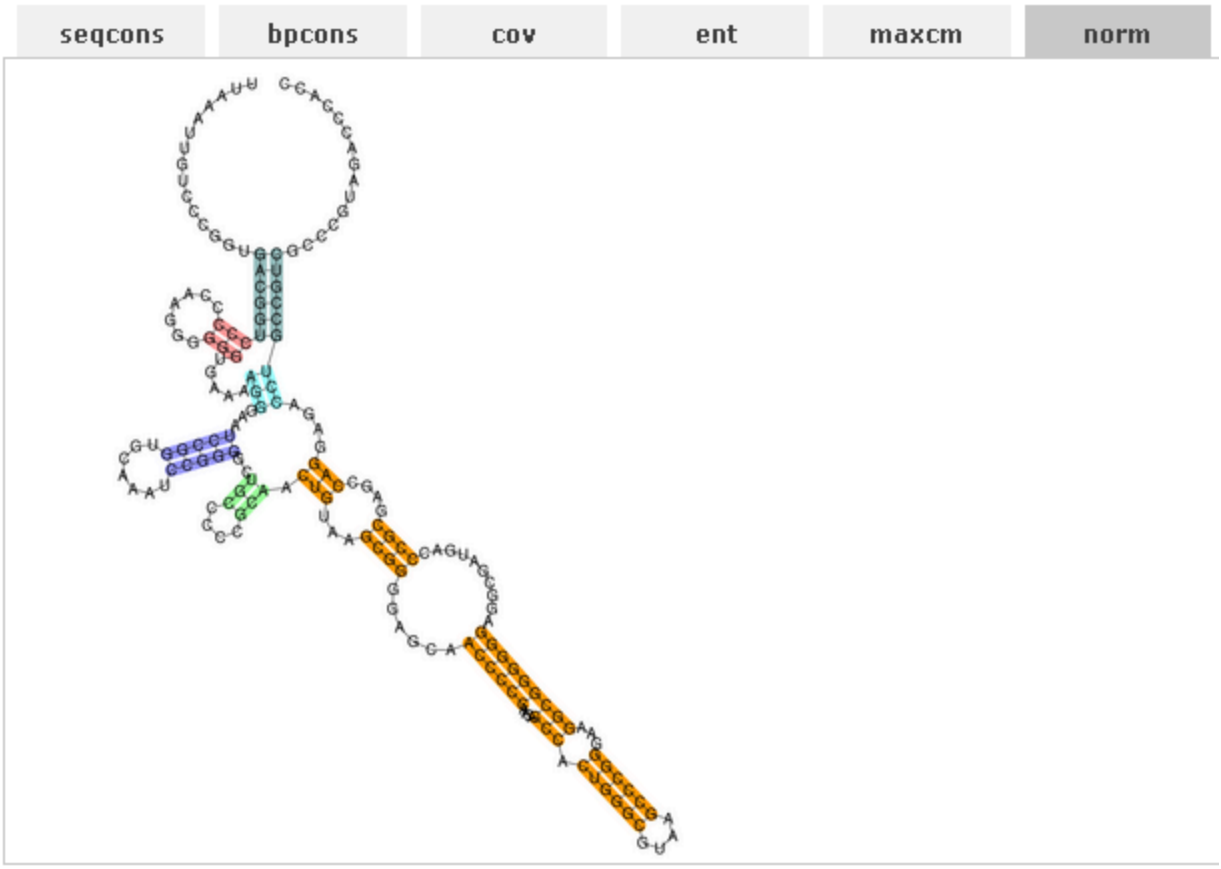
7) Choose Secondary structure.



And you will get the structure.



Choose norm to get the RNA sequence letters as we know them. And now you have a picture of a RNA switch.



Note on what you see. Rhiju said: “The structures in this database are for the RNAs in the 'ON' state, and typically just for the ligand-binding part of the switch -- its poorly understood what the structure of the OFF state is, in most cases.”

So basically this is only one of the two switch states we see - the shape that binds up to the molecule.

Rhiju also said in the scientist chat 14 november 2012: “and we're finding that the 'OFF' states of natural RNAs often form multiple secondary structures, each with significant populations.”

8) If you find some other useful functions in this database or get other ideas on how to watch Natural occurring RNA switches, please share. Post a picture or more to demonstrate how to do it and tell what you use it for.

Weird RNA

Some of the RNA in Rfam have a very different to usual Eterna base pairing. I found a small description on why. It will explain a few of the oddball cases I stumbled upon while I was having a look at the database.

GC and AU basepairs are known as canonical or Watson-Crick base pairing. Those are the strong bindings. One of the non-canonical basepairs we already know is GU. Here is an intro to the others.

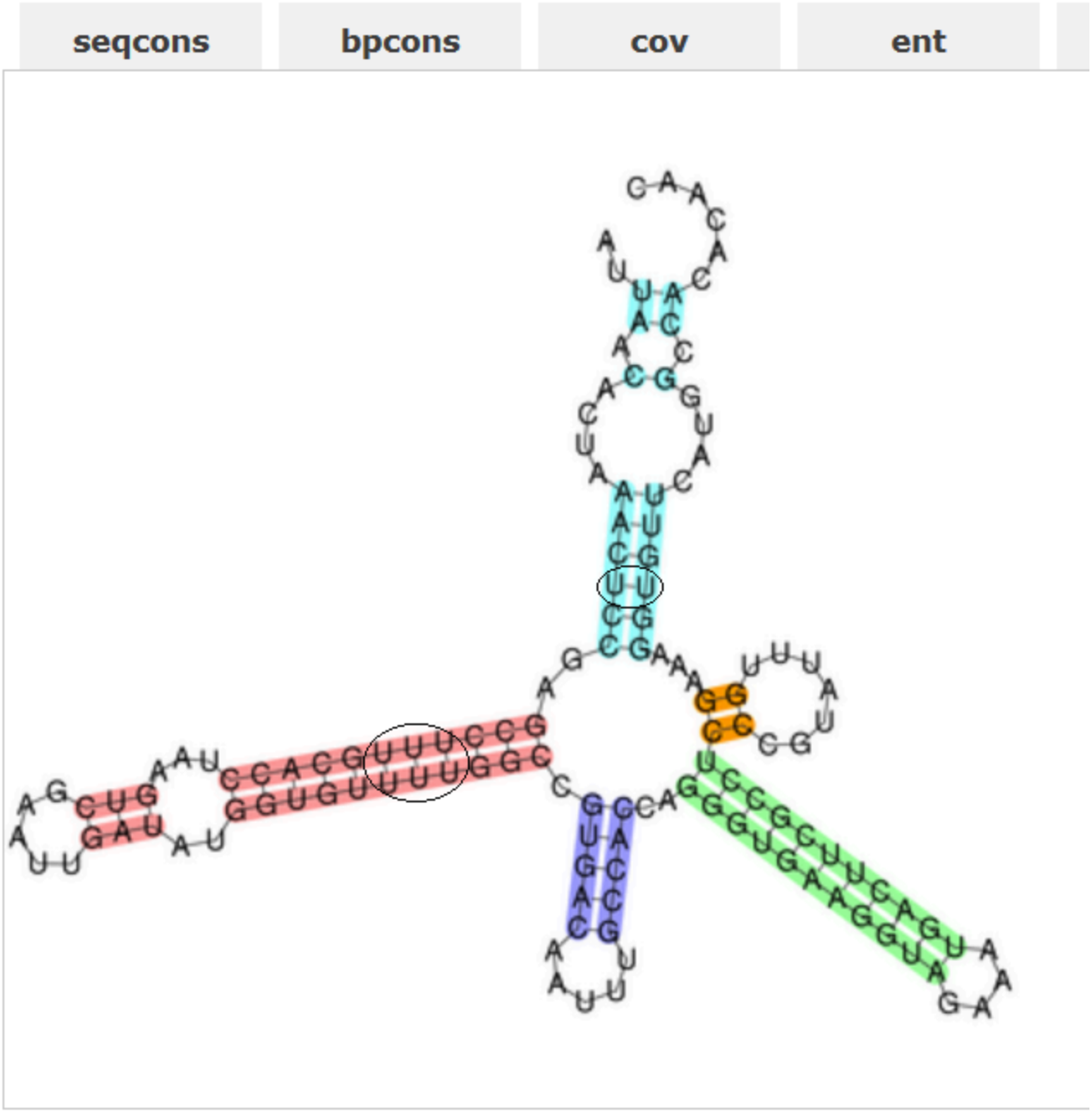
"By contrast, the other eight possible pairings (AA, AC, AG, CC, CU, GG, and UU) are much weaker - energetically close to no pairing at all. In the case of the strong pairings, the two surfaces fit together, so to speak, whereas in the case of the weak pairings, there is a mismatch between the respective molecular shapes."

George Church, Regenesis, page 30

Now you should be a bit better prepared for what you are going to see.

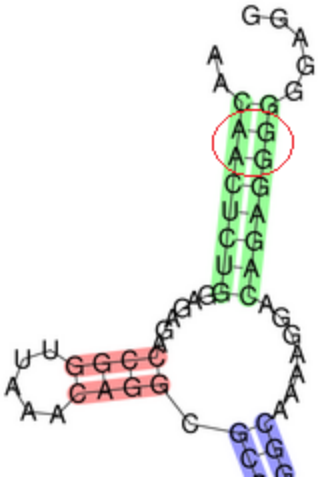
For a short intro to different types of RNA, see under the section on What RNA's do.

Notice the U’s opposite U’s. (marked with black circle)



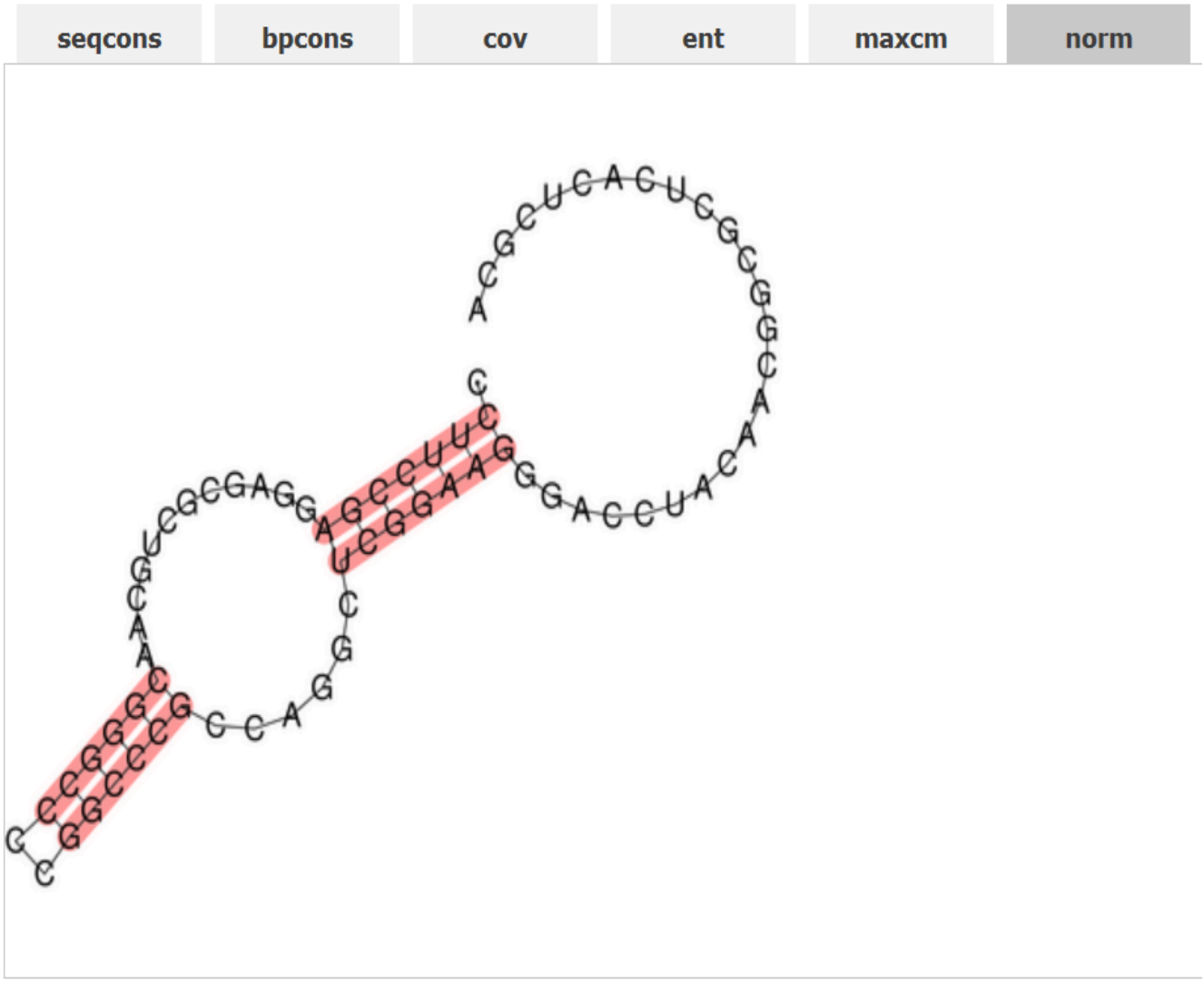
[Moco RNA motif]

This one have AG-pairs.



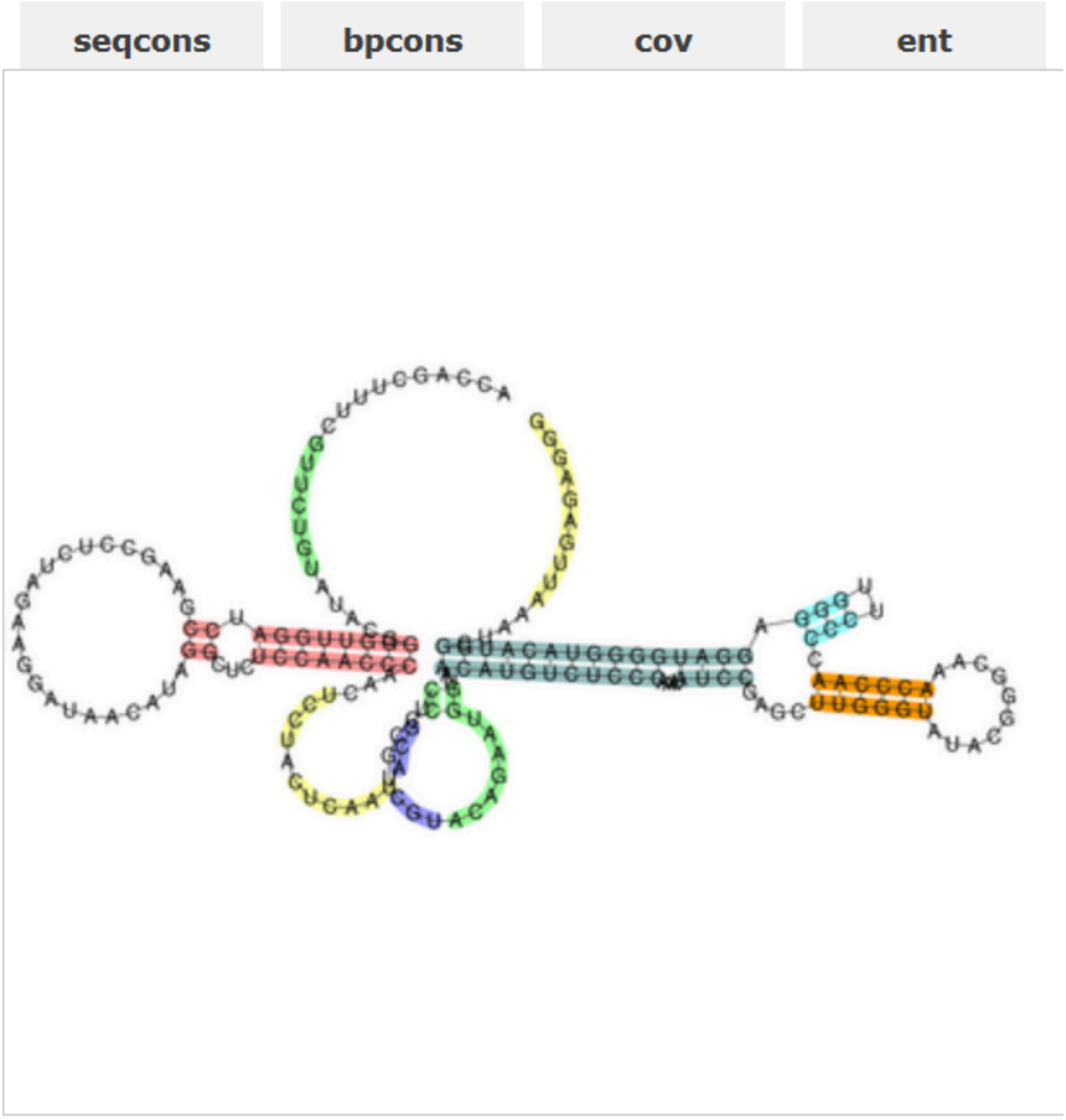
[Family: Glycine (RF00504)]

This one have the weirdest “endloop” I have yet seen. Is that physical possible? No wonder why this whole stem is GC’ed out. And it have lots of double pairs.



Family: SAH_riboswitch (RF01057)

For the design below, I wonder: Do the two yellow strings pair up in the other shape? Same question for the green ones. I can’t get the sequences to match totally from the base pairing I know, but they looks like close matches that way. If this is the case, we get a shadow picture of the unbound shape.



[Family: drz-agam-2-2 (RF01788)]

Anyone who can answer on those two last examples, please do.

Observations from Rfam

I have been playing a little with RFAM as I was interested in seeing how natural switches looked. So far I have mostly checked riboswitches, but looked a few MicroRNA's and Listeria sRNA too.

By the way, I mentions riboswitches a lot. Go read a cool article about them here. First time that riboswitches have been seen for real.

Length of stem

I was wondering what was the typical stem length for natural switch RNA. For now what I have witnessed of stem length is from 1 base pair to a 12 base pair long stem. (There might be longer stems I have missed). The more typical and frequent occurring stem length is from about 3 to 7.

Actually this riboswitch RNA seem in general to be a mixture of different stem lengths. However there are a few with almost similar stem length.

The short stemmed ones like 3-4 basepairs long typically have GC-pairs at the closing bases.

Loop content

What seems to be in highest abundance is A, U, G.

Actually I was surprised about how many reds there was in internal loops. I suddenly thought I understood. Some of the stems are longer in the database than the more common 4 basepairs stem from our data. So for stem to loop regions this will happen:

A GC-pair from stem will get split when the rna switches and one of the nucleotides will end up in loop region and the other not. So either the G or the C will have to be deposited in the loop. The huge presence of G's in internal loops is the logic consequence of C's being mostly unwelcome there. As C are generally more interacting, and stem causing.

Closing basepairs of stems

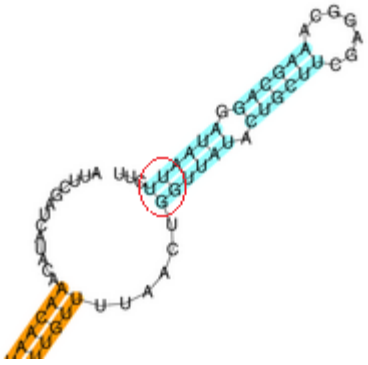
Often a AU in one end and a GC-pair in the other. Not a 100 % rule, but an often happening thing. GU can substitute for AU.

What I see in our own RNA switches is that half of the closing basepairs in an internal loop and at least one closing base pair in a multiloop is AU. I think the AU-s there are helping the switch not to attach itself too much, as a multiloop with all GC's is a strong thing. Same with a stem closed at both ends with GC-pairs. Also the last thing would mean more C's and G's in loop region of the other switch shape.

It is easier breaking a stem open, if not both ends are equally well sealed off. And AU pairs at inner side of multiloops I guess are aiding the switching.

GU content

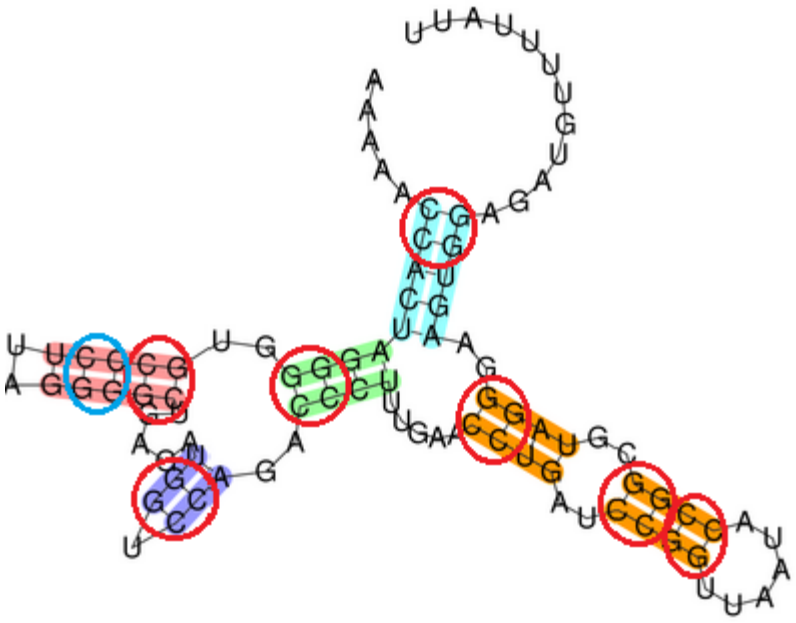
Gu content is relatively low. GU mainly occurs at closing base pair and in short strings too. So that switch strategy I have posted in chat might be premature then. I based it on what data I had, which was too little. The GU-pairs in stem are spread out. Not beside each other. And mainly in longer strings. I did find one double same turning GU at a string end, and a triple one inside a string. Ok. I also found one with twisted GU'pairs. So no clear pattern here.



http://rfam.janelia.org/family/RF01482#tabview=tab3

Repetition

And in general there are have lots of double same turning basepairs. Or 3 of the same nucleotides in line. Those are what stands out to me. I'm aware that and all sorts of other combo's can turn up as periodic repeats, like this thought example GAC or CUA if repeated. Not all Riboswitches are equally



repetitious.

http://rfam.janelia.org/family/RF00059#tabview=tab3

I'm seeing incredible many double GC-blocks, particularly at ends of stems. Red circles

MicroRNA mir

The few I checked have a very line like structure. Long line.

Listeria sRNA

Also has a distinctive shape to them.

Are the different RNA's sorted based on shape or phylogenetic relationship (sequence likeness)? Or both?

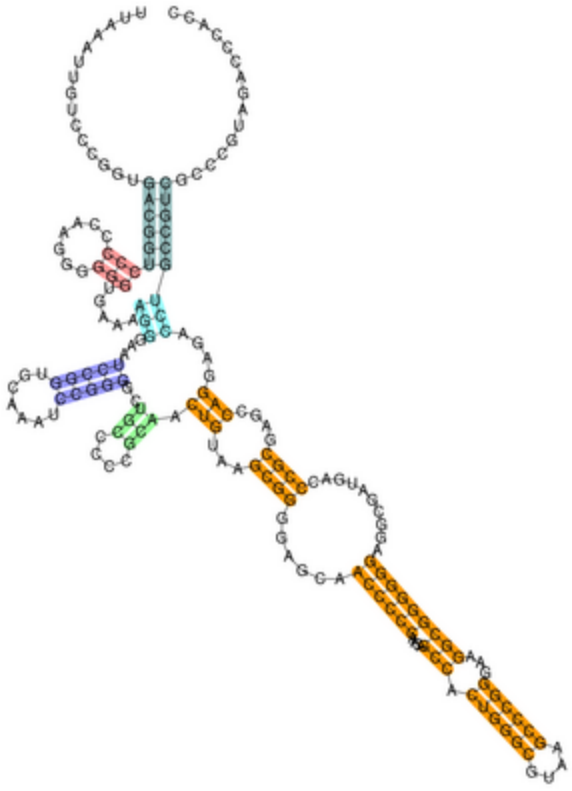
Thoughts on riboswitches in general

Riboswitches are quite varied in shape and surprisingly different in GC and AU content. Some switches are unusually high in GC-pair content. Others in AU content.

Unusually high GC-content in stem region. Compared to our switch lab and single shape lab

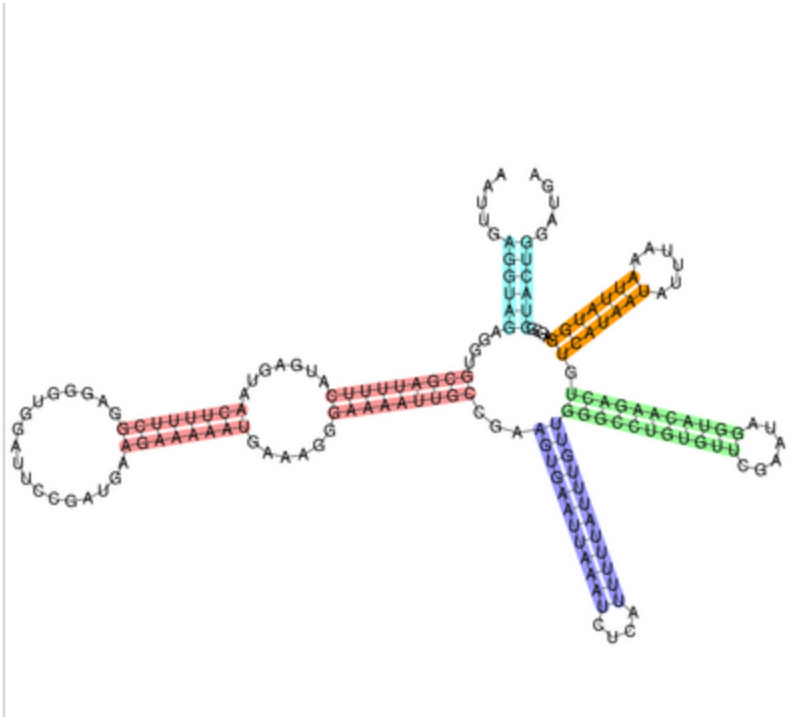
Family: *Cobalamin* (RF00174)
Description: *Cobalamin riboswitch*

.



http://rfam.janelia.org/family/RF00174#tabview=tab3

This one is very high in AU content. But have long strings to explain it. It also have relatively big loops. I wonder if that is typical of high AU content switches?

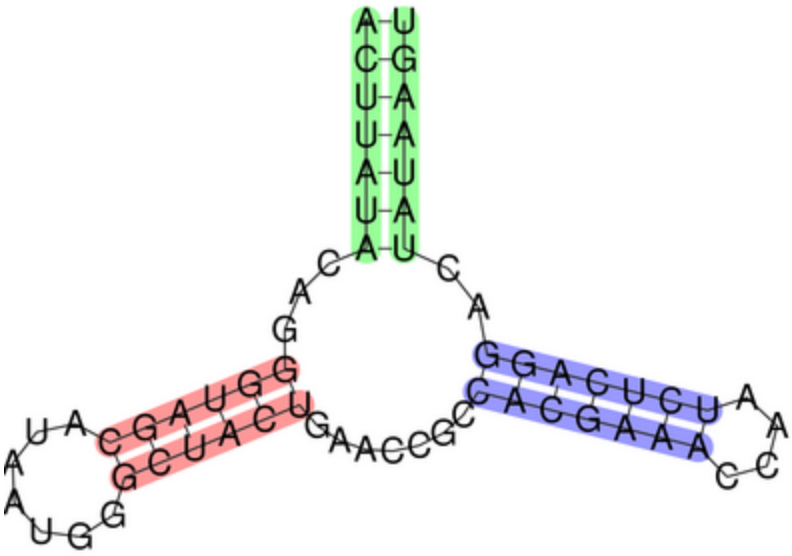


http://rfam.janelia.org/family/RF00168#tabview=tab3

Other overall tendencies. I think the tendency is that the ones with lots of GC-content in the stems have more short stems. Though there are long stems in them too. They just have a higher abundance of short stems.

Small switches

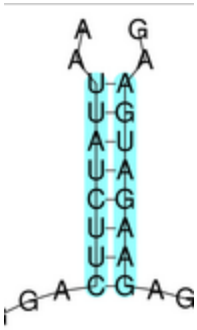
Do smaller switches have less repetitive pattern?



http://rfam.janelia.org/family/RF01510#tabview=tab3

Thoughts in the weird department

When I saw the riboswitch necks I thought I was seeing ghosts. Generally there are lots of AU-basepairs in the start/finishing section (neck) when compared to other strings. That is if they have a clear defined neck. Is switch "neck" different from switch stem?



http://rfam.janelia.org/family/RF00050#tabview=tab3