How to read the SHAPE data - extended

by Eli Fisker

Back to How to read SHAPE data - basic

I have later made a simpler introduction on how to read SHAPE data.

Document shortcuts

Here comes a shape data update as reading shape data for switch lab is a bit different to reading shape data for single shape lab.

When in the lab result page, click on a design and choose this symbol, to see the shape data.

What is shape data?

SHAPE[1] data is the chemical mapping of how the RNA bases are paired up after the RNA is synthesised. Very simply put, the RNA shape gets tested with different chemicals to test which bases bind up or not. For a more thorough explanation check this forum post.

Shape data is a help when reading the results we get back from lab. It makes you able to see which part of the design that folded up as it should and what parts that you can improve.

How to interpret the shape data?

So to sum up how to read SHAPE data: Basically you want loops and single strands to be yellow and stings and basepairs to be blue.

Blue = binding
Yellow = not binding

[1] SHAPE stands for selective hydroxyl acetylation by primer extension.

What are the advantages of watching SHAPE data in the continuous mode of colors?

The continuous mode gives you much more details about how well the bind is. This is very useful when you want to do a mod of a lab design or are watching the lab results. Basically you should go for changing the bases that are yellow in stacks where they are supposed to be blue.

Jieux: the basic one goes into "light blue" bases in loops not being counted.
Jieux: but what about light blue bass in stacks?
Jieux: seems non-continuous mode is a better way to view the data.
Jieux: cause all those shades make no sense to me.
Eli Fisker: The light blue give you more tendency about how well the bind is
Eli Fisker: and that is very useful when you want to do a mod
Jieux: darker blue=better bond yes?
Eli Fisker: yep
Jieux: ok
Eli Fisker: And weaker blue = weaker bond

Also see the answer Rhiju gave on Jieux's question. He gives details on the actual scoring for the SHAPE data.

How can opposite sides of a pair have different colors of blues?

Jieux asked about this. Sometimes a basepair can even have opposite colors of SHAPE data. One blue and one yellow. Sometimes it is because the one nucleotide pairs up somewhere else and the structure misfolds, but that only explains a minor number of the cases.

Rhiju gives an explanation here.

Why does a base pair in SHAPE data have both a blue and a yellow base, when blue means paired base and yellow means unpaired?

Omei had a fine explanation of why a base can appear to unpaired while it's partner base appears to be paired.

Also check the continued forum discussion

Switch lab

In single shape lab the shape data shows nucleotides as binding and not binding. SHAPE data for switch lab also shows if the nucleotides change from the one shape and into the other. Or as Rhiju says it:

The switch scoring system is based on the changes in the SHAPE data in the -FMN vs. +FMN condition, so it is different from before.

-FMN - meaning the unbound shape without the FMN molecule present
+FMN - meaning the molecule-bound shape with the FMN molecule present.

Though the SHAPE data has different origin between the two types of lab, this stays the same: Strings have to be overall blue and and single strands of nucleotides should be yellow.

Shape reactivity

As Rhiju says: *High SHAPE reactivity correlates with unpaired nucleotides, and low SHAPE reactivity with paired nucleotides. We are trying to design switch puzzles where several nucleotides change their status between paired or unpaired -- or vice versa. Then we check if the SHAPE data goes in the right direction*.

Here is what it means:

SHAPE reactivity = How well protected the nucleotides are tested the RNA molecule with chemicals. And Rhiju adds: We are also now using the chemical DMS (dimethyl sulfate) which modifies unpaired A or C nucleotides, as an additional test of the switched structure.

SHAPE in strings

Blue nucleotides means low SHAPE reactivity, which is good as it means the nucleotide have bound up and stays protected, when the whole RNA molecule gets tested with chemicals.

Yellow nucleotides means high SHAPE reactivity, which is bad. As it means the nucleotide is not protected and gets eaten way from the string by chemicals.

SHAPE in single strands and loop ring area

Single strands should be yellow as yellow means the nucleotide goes unpaired and stays single and does not end up in a string. Blue in single strands, means the nucleotides paired up somewhere they are not supposed to be.

As Jennifer Pearl summed up: loops are light, stacks are dark.

Below I have added pictures. Thanks to Mat for the suggestion of examples.

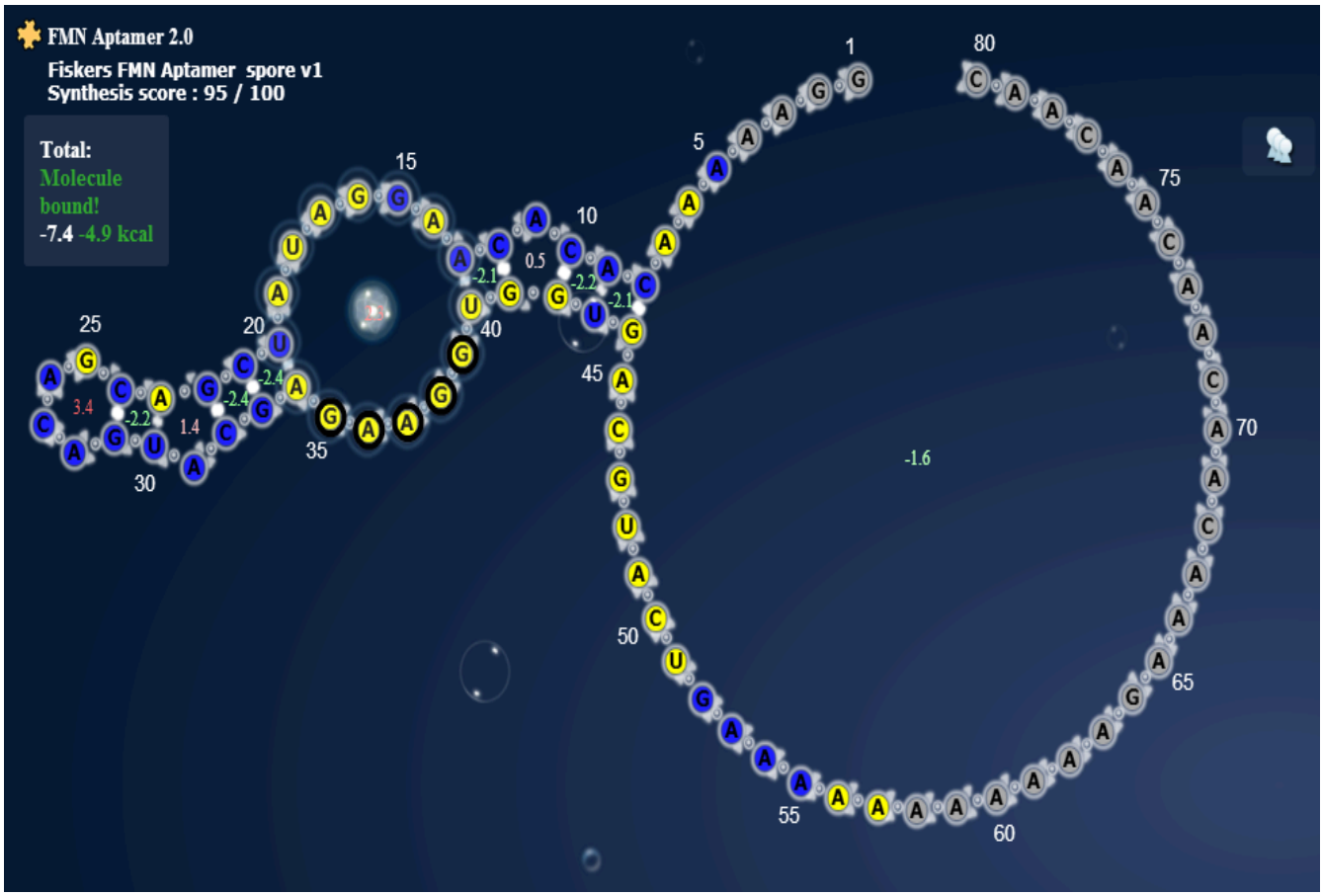
However ring bases can't be as harshly judged as them being not yellow and being bluish, ain't always bad, as it means they might be binding with a stem base and thus adding stability
to the overall secondary structure or even the tertiary structure

SHAPE data

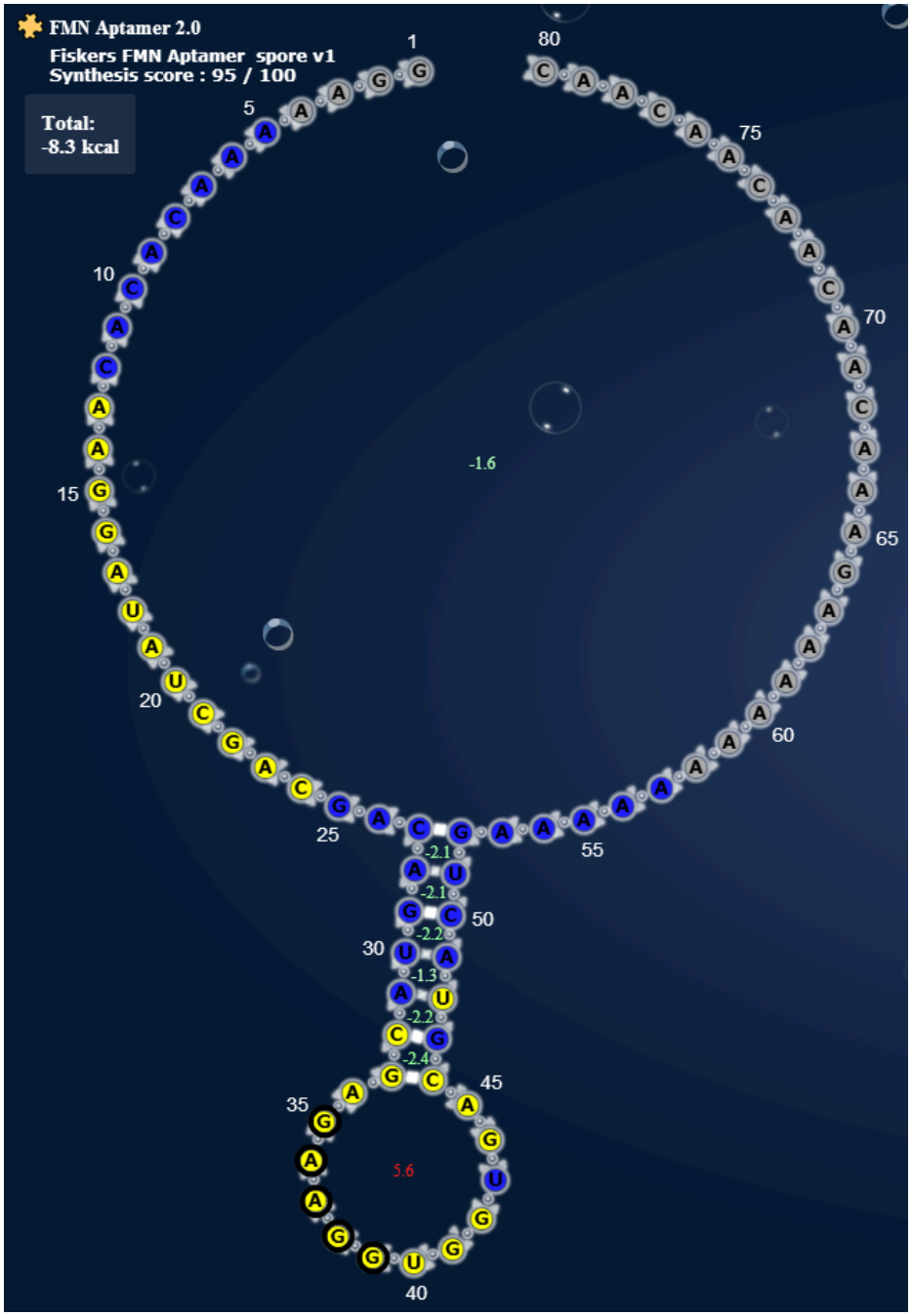
Loop to loop has to be yellow to yellow

Follow the marked nucleotides at 35-39 as they jump from one loop region to another in the next shape.

Molecule-bound shape



Unbound shape

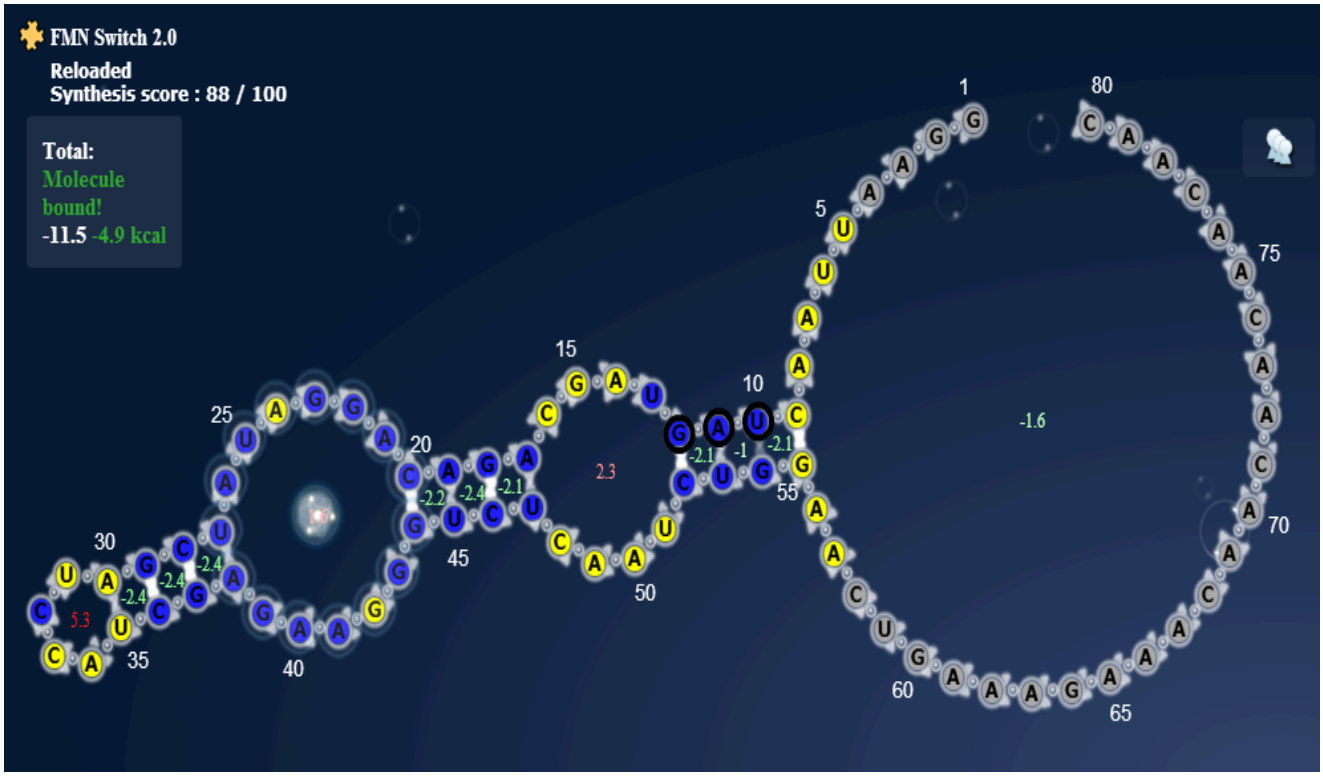


.

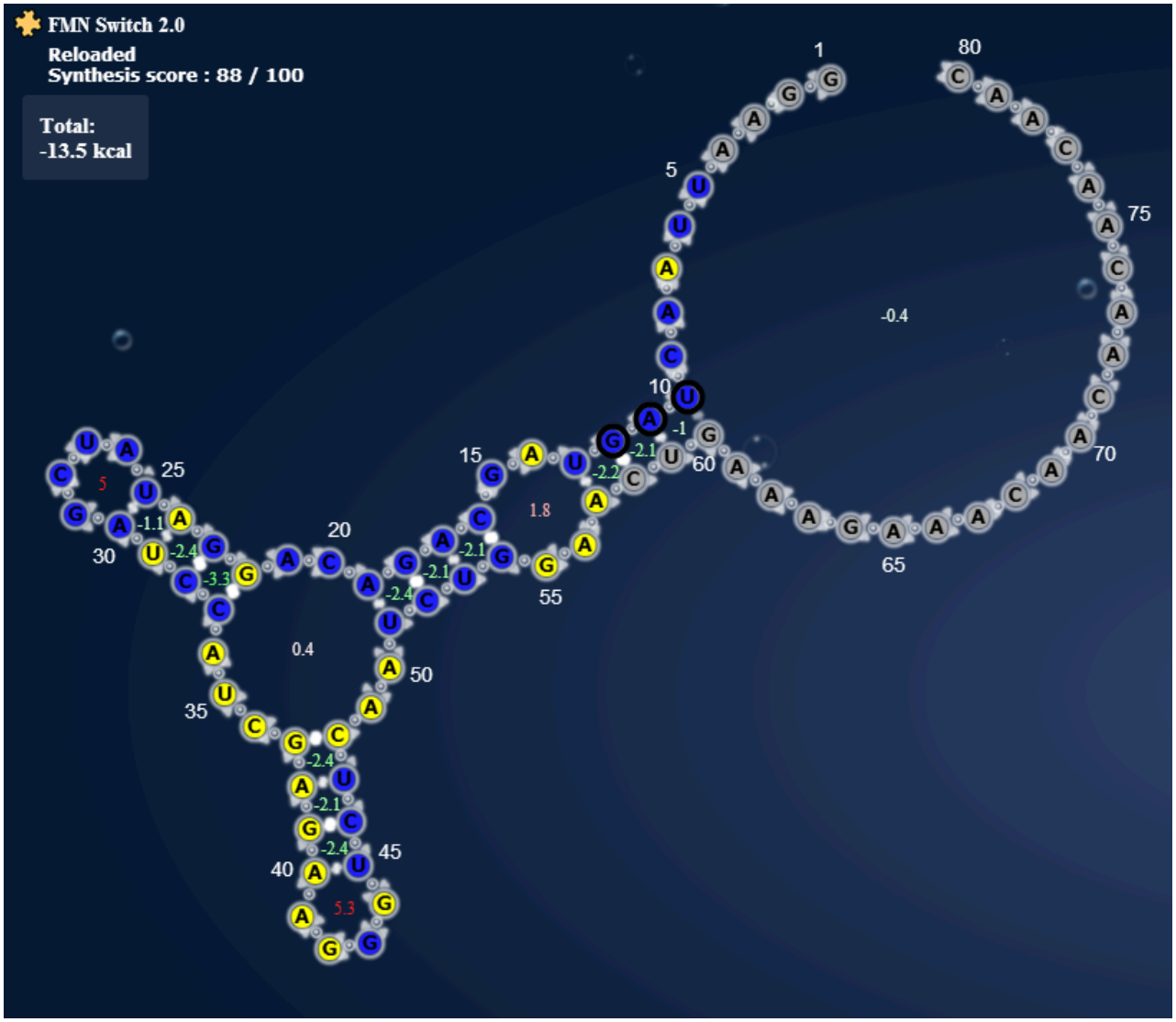
Stack to stack has to be blue to blue.

Follow the marked nucleotides blue nucleotides, as they jump from stack to stack. You can use the numbers too to see where the nucleotides were before. (Nucleotide 10-12)

Molecule-bound shape



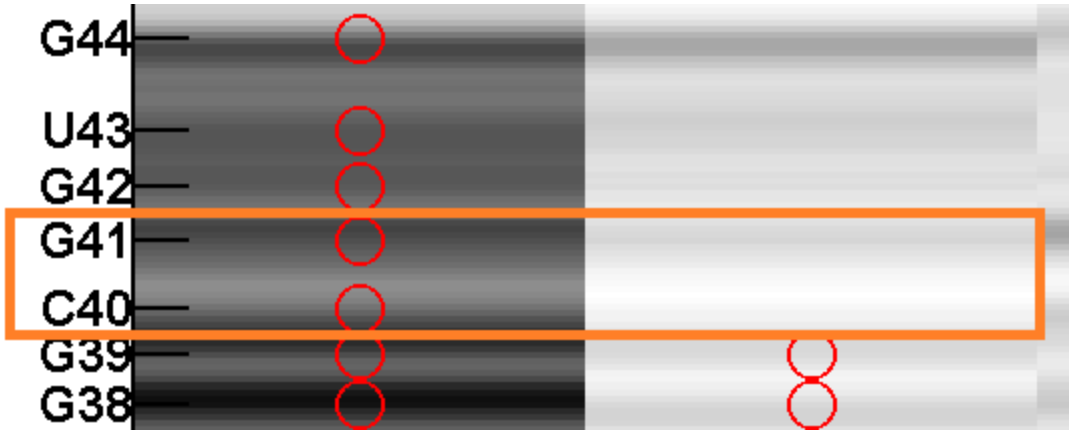
Unbound shape:



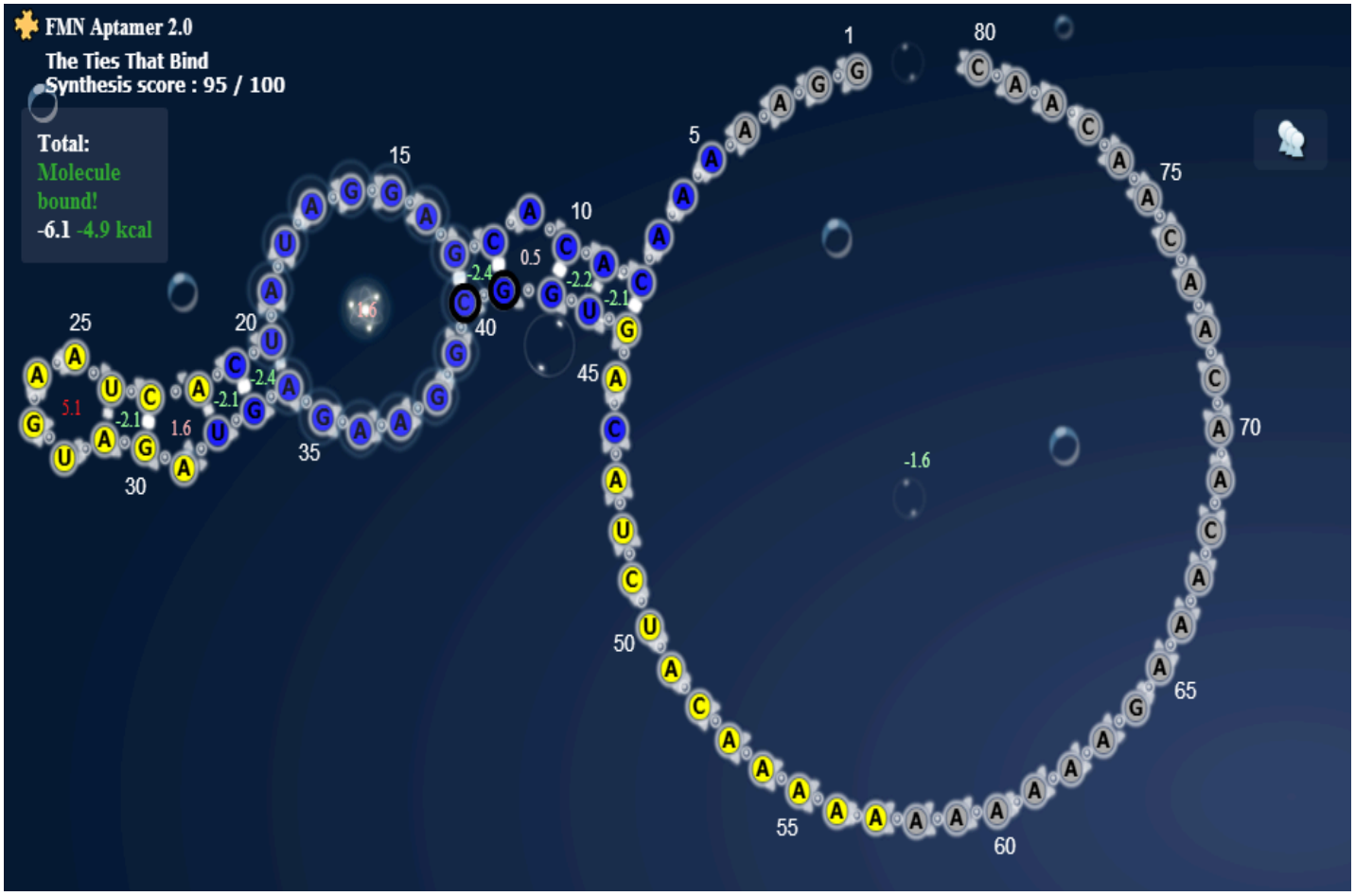
Raw data

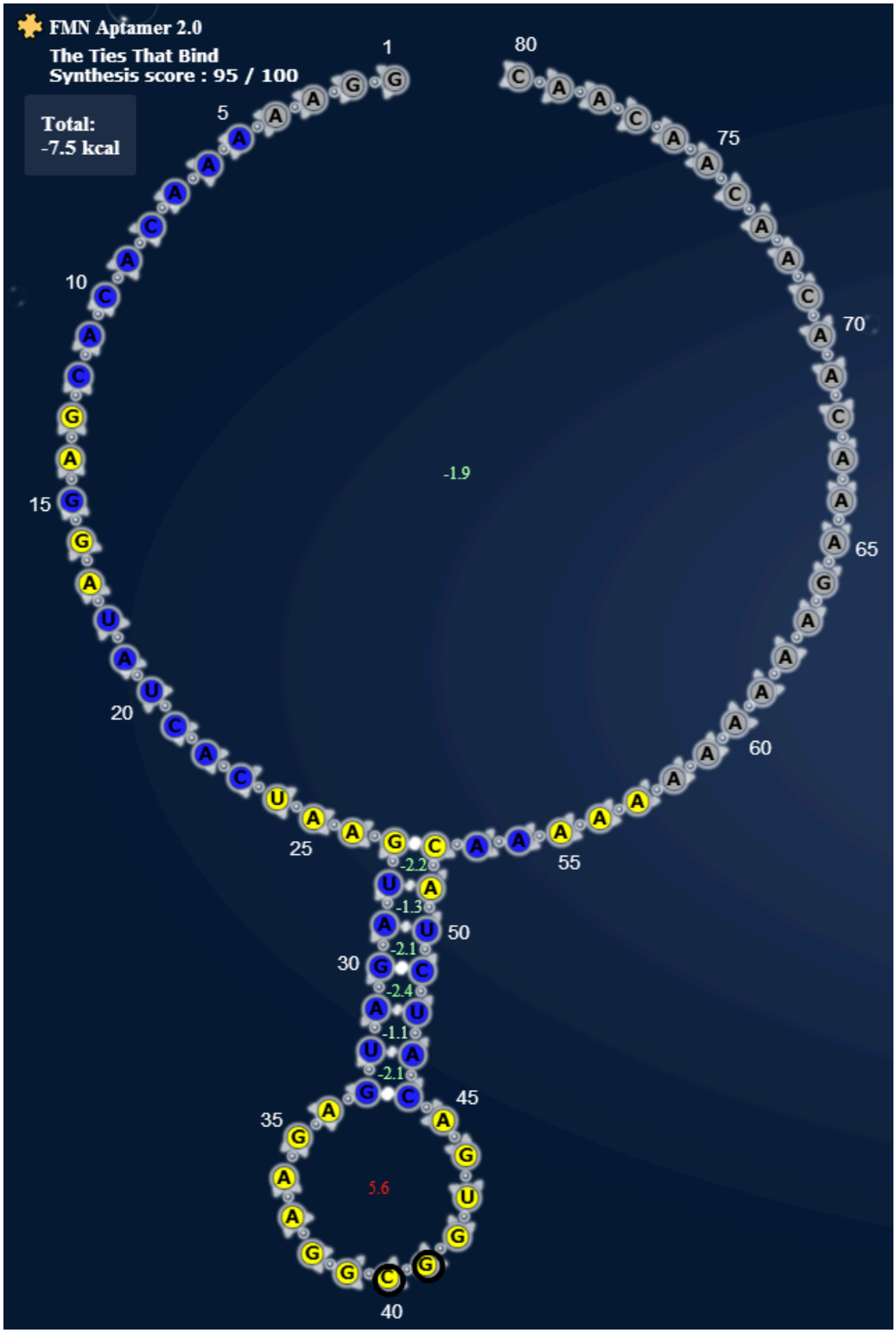
stack to loop has to go from dark to light

Column to the left shows the bound shape, the second column the shape with the molecule.



Molecule-bound shape





Example from Drax FMN aptamer 2.0 winner, [The ties that binds](). [Raw data for design]().
loop to stack, light to dark

Problems with the data

For now even the slightest switch in the raw data counts as a protected and blue nucleotide in the shape data. That should get fixed with the new interface in the coming EteRNA 2.0 update. But for now we get the data as blue and yellow, as if they were black and white and missing the finer grey tones. This problem even persist in the advanced color mode of shape data.

Here is an example from Rhiju's explanation of the Raw data. He said:

Then, nt 24-26. Hmm, doesn't look as good, as 24 and 25 stay exposed. Dang, get one point (for nt 26), out of 3.

But 25 isn't exposed according to the shapedata. It is blue and not yellow.

Same with this:
Fourth, nts 34-36, we get 2 points. Why didn't C36 get protected?

But 36 is blue and protected in the shapedata.

As Rhiju answered to this: If part of the design is supposed to become protected and it goes from very yellow to less yellow, that counts as OK in the 'switch score'.

So even the slightest hint of a shift for a nucleotide, between the two shapes, will be shown as a full blown shift.

Rhiju: I think the confusion is because the yellow/blue coloring is too 'coarse' to reflect the changes -- there's a way to color the structure that goes from yellow to white to blue in a more gradual fashion.

If part of the design is supposed to become protected and it goes from very yellow to less yellow, that counts as OK in the 'switch score'.

There's going to be a new switch interface coming out soon -- I'll relay to Jee that we need some way to clearly visualize the differences in reactivities upon binding... if you or other players think of something, please also post or send Jee an e-mail.

SHAPE Lessons

SHAPE data - version advanced

Nando has written a very fine explanation - with pictures - about why shape data show odd results. Get to understand what is going on, beyond paired and unpaired.

Written Sep 2012, last updated Jan 2014