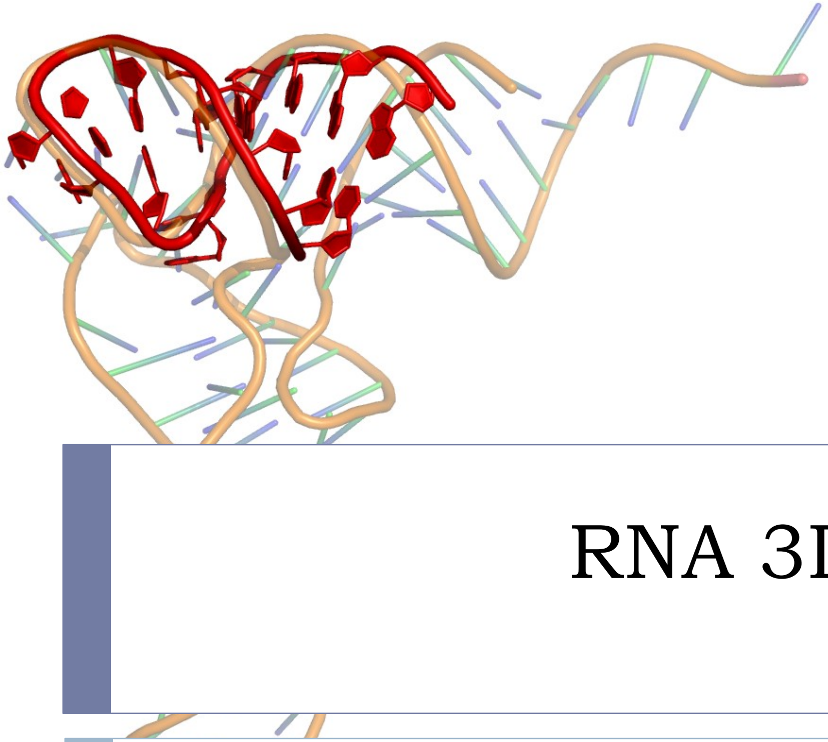
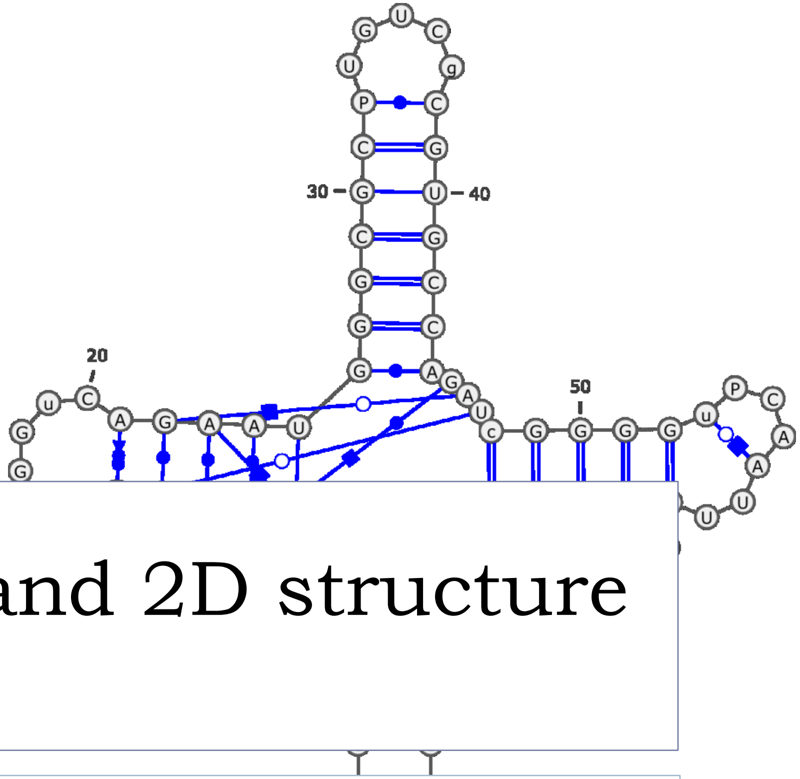




RNA 3D and 2D structure

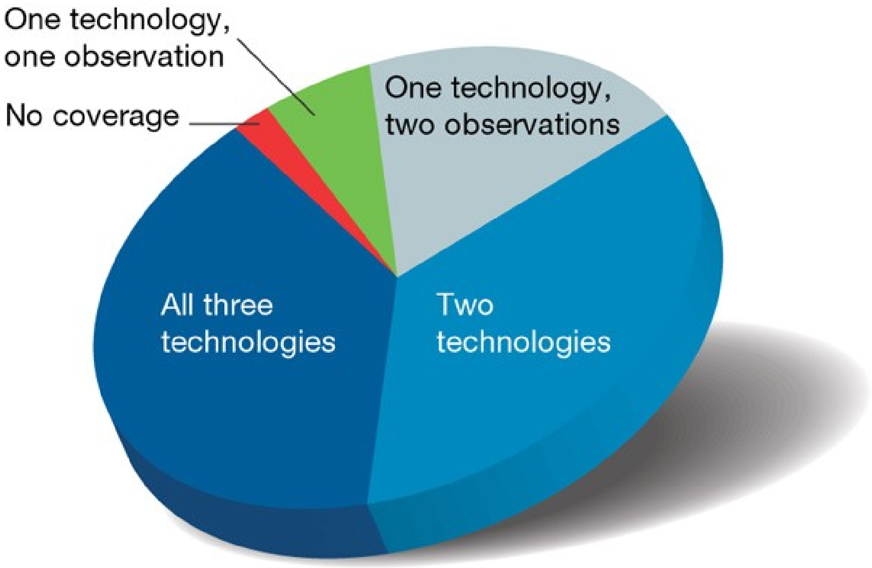
Yann PONTY
CNRS/Ecole Polytechnique

A redundant talk… sorry!

- Gap between analysis tools and viz. tools (M. Brudno)
- Challenge of scale (C. Nielsen)

Why RNA is **so COOL!**

- Ubiquitous
- Pervasively expressed



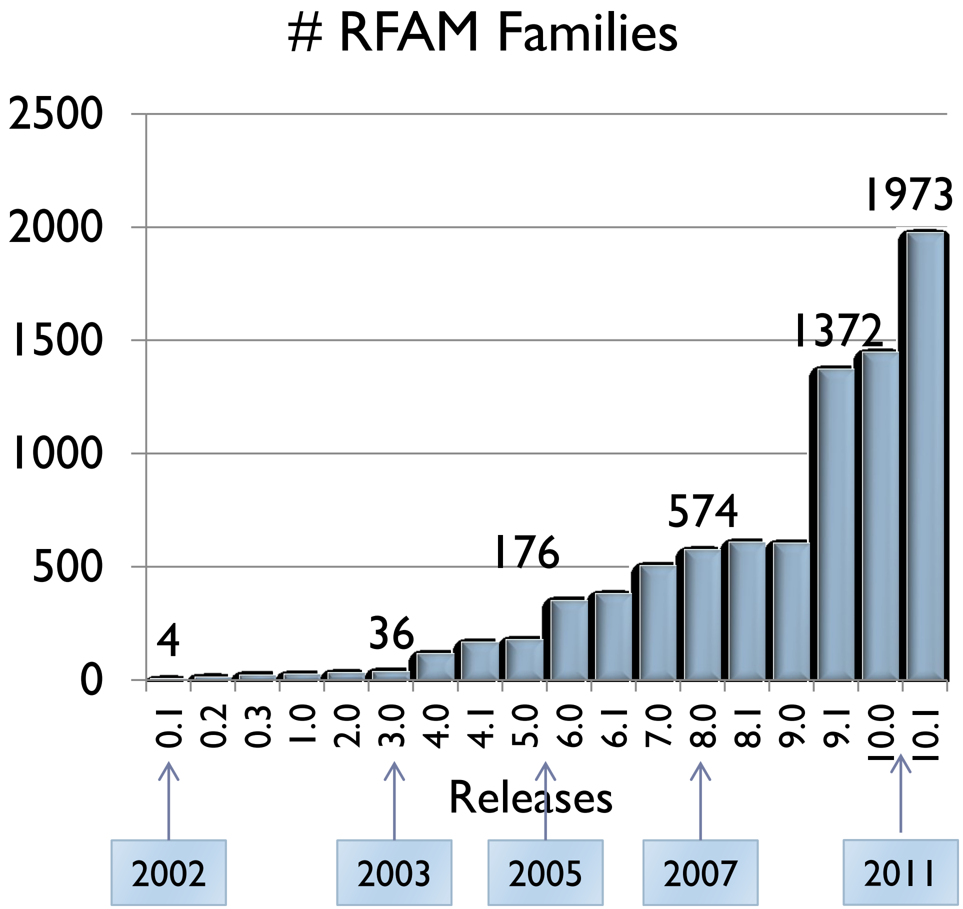
The human genome is pervasively transcribed, such that the majority of its bases are associated with at least one primary transcript and many transcripts link distal regions to established protein-coding loci.

ENCODE Analysis of 1% of the human genome Nature 2007

Why RNA is **so COOL!**

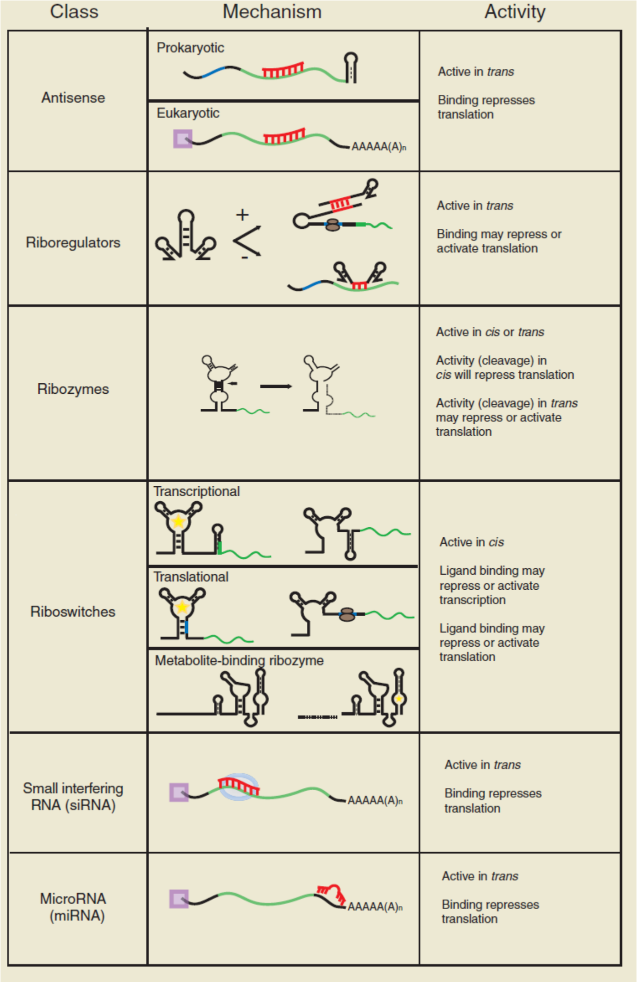
- Ubiquitous
- Pervasively expressed
- Versatile

- Carriers
- Transporter
- Enzymatic
- Processing
- Regulatory
- ssRNA genomes (HIV)
- Immune system?? (CRISPR)
- More soon… (lincRNAs)



Why RNA is **so COOL!**

- Ubiquitous
- Pervasively expressed
- Versatile
- Easy to handle
  - Synthetic biology

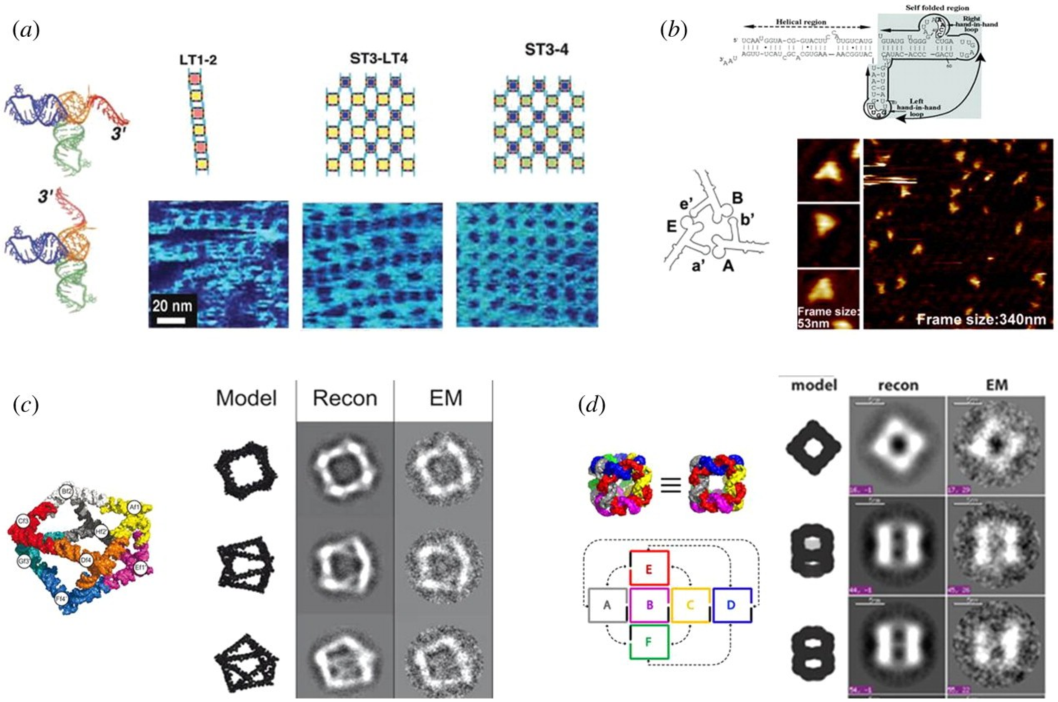


Class	Mechanism	Activity
Antisense	Prokaryotic Eukaryotic	Active in *trans* Binding represses translation
Riboregulators		Active in *trans* Binding may repress or activate translation
Ribozymes		Active in *cis* or *trans* Activity (cleavage) in *cis* will repress translation Activity (cleavage) in *trans* may repress or activate translation
Riboswitches	Transcriptional Translational Metabolite-binding ribozyme	Active in *cis* Ligand binding may repress or activate transcription Ligand binding may repress or activate translation
Small interfering RNA (siRNA)		Active in *trans* Binding represses translation
MicroRNA (miRNA)		Active in *trans* Binding represses translation

[Isaacs, F J *et al.* Nature Biotech. 2006]

Why RNA is **so COOL!**

- Ubiquitous
- Pervasively expressed
- Versatile
- Easy to handle
  - Synthetic biology
  - Nanotechs



[Li H *et al*, Interface Focus 2011]

Why RNA is **so COOL!**

- Ubiquitous
- Pervasively expressed
- Versatile
- Easy to handle
  - Synthetic biology
  - Nanotechs
  - Therapeutics (RNAi)

RNAi : Proof of concept

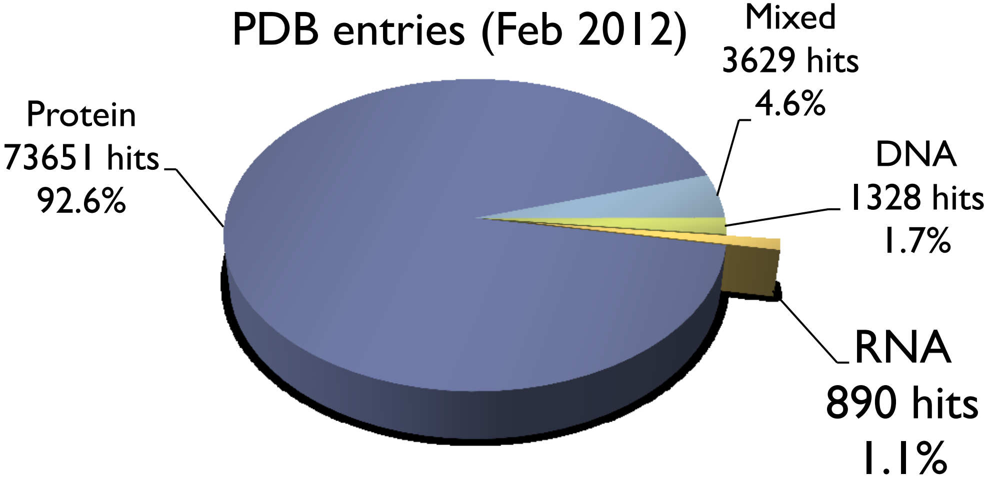
Injecting nanoparticle-vehicled siRNAs in solid-cancer patients:

- siRNA enters tumorous cells
- siRNA interacts with targeted mRNA
- siRNA regulates protein expression

[Davis M I *et al*, Nature 2010]

Why RNA is **so COOL!**

- Ubiquitous
- Pervasively expressed
- Versatile
- Easy to handle
  - Synthetic biology
  - Nanotechs
  - Therapeutics (RNAi)
  - Computationally fun (but still challenging)



(Initial) lack of structural data

Experiment-based energy models
+ Secondary structure
+ Efficient combinatorial algorithms

⇒ Mature *in silico* prediction tools (Mfold, RNAfold…)

Why structure matters

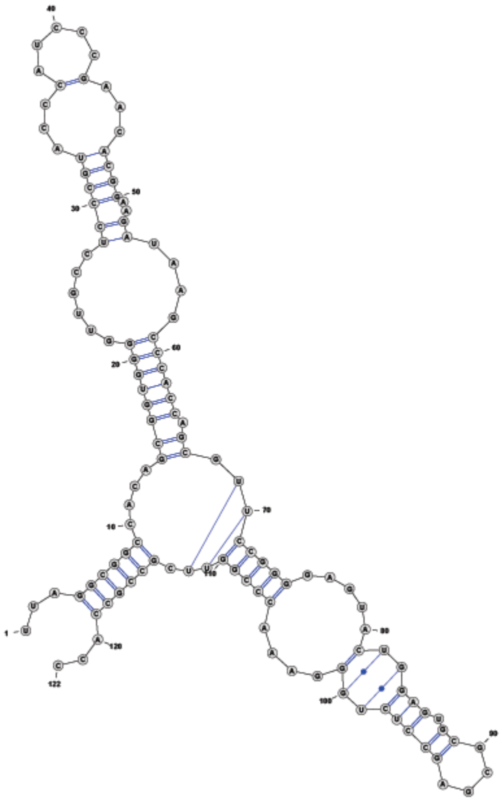
- RNA is single stranded
- Structurally diverse
- Structure more conserved than sequence
- Functionally versatile

Use structure as a proxy for function, favor mechanistic explanations.

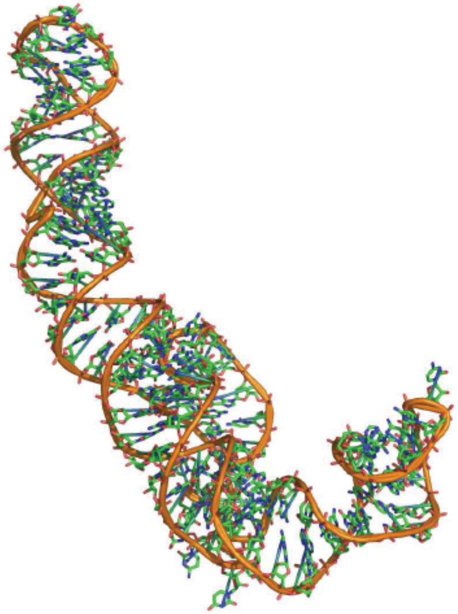
Three levels of RNA structure

UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAAGAUAAGCC
CACCAGCGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC

Primary structure



Secondary structure



Tertiary structure

Source: 5s rRNA (PDBID: 1K73:B)

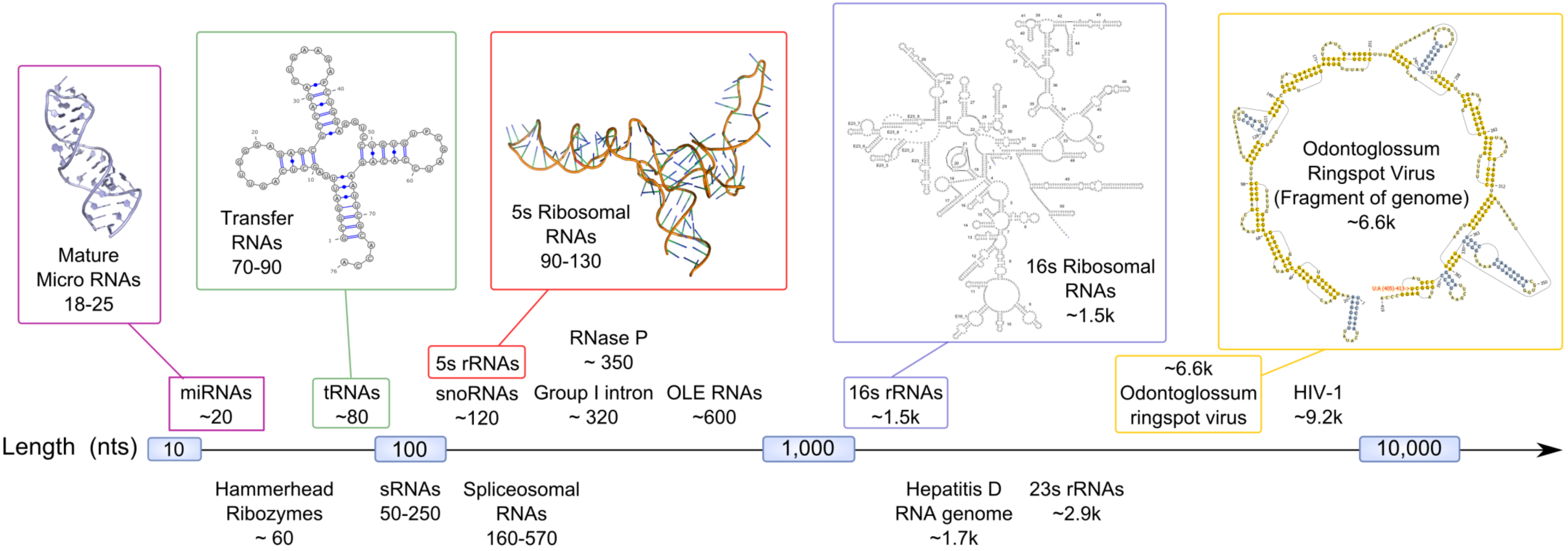
Current visualization of RNA

Exemplary use cases

Visualization helps ncRNA scientists

- Refine structural model based on experimental data
- Assert reliability of predicted structures
- Detect structural homology
- Curate structure-informed alignments
- Communicate functional hypotheses
- …

A challenging diversity of scale



Mature Micro RNAs 18-25

Transfer RNAs 70-90

5s Ribosomal RNAs 90-130

16s Ribosomal RNAs ~1.5k

Odontoglossum Ringspot Virus (Fragment of genome) ~6.6k

miRNAs ~20

tRNAs ~80

5s rRNAs

snoRNAs ~120

RNase P ~ 350

Group I intron ~ 320

OLE RNAs ~600

16s rRNAs ~1.5k

~6.6k Odontoglossum ringspot virus

HIV-1 ~9.2k

Hammerhead Ribozymes ~ 60

sRNAs 50-250

Spliceosomal RNAs 160-570

Hepatitis D RNA genome ~1.7k

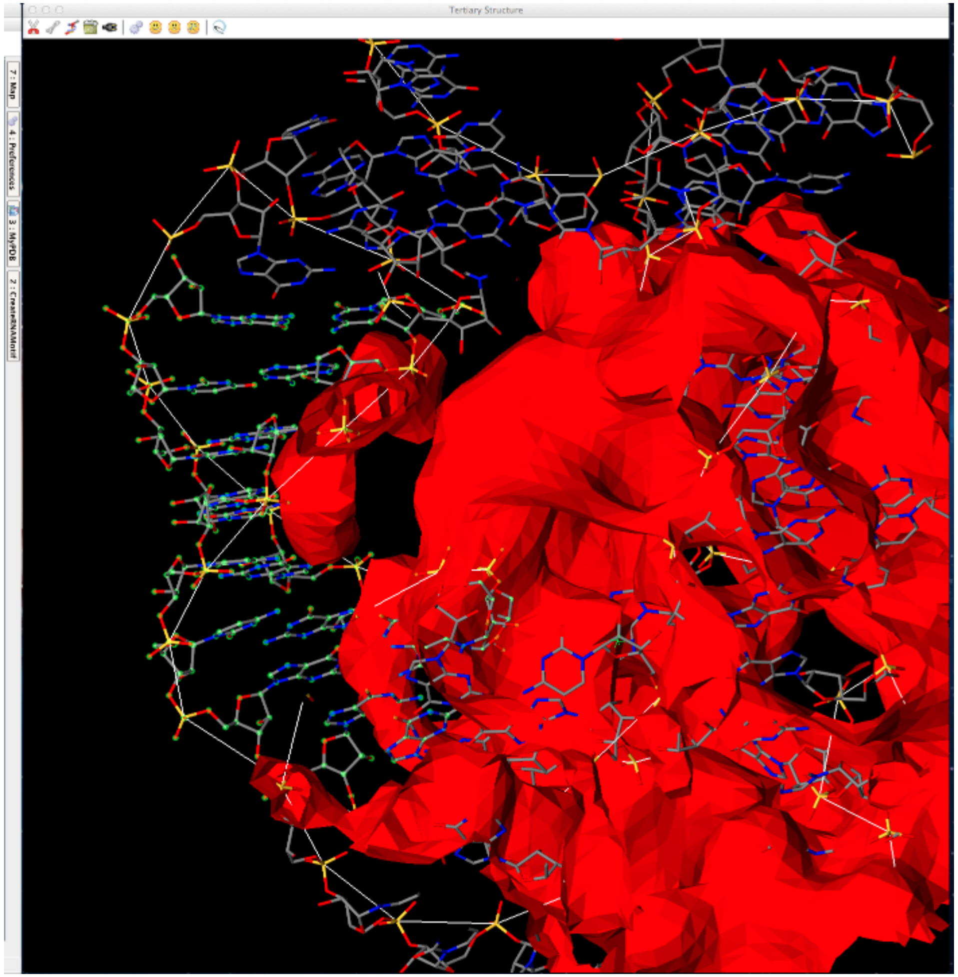
23s rRNAs ~2.9k

- Length of structured RNAs from 18 to over 9k nts.
- 2D schematics vs 3D objects (Top-down vs Bottom-up)
- Local vs Global

Fitting 3D model to density maps

- Cryo-EM maps
  - UCSF Chimera
    [Goddard et al, J Struct Biol 2006]
  - Coot
    [Emsley P *et al*, Act Crys D 2010]
  - Assemble
    [Jossinet *et al*, Bioinf. 2010]
- Semi-automated

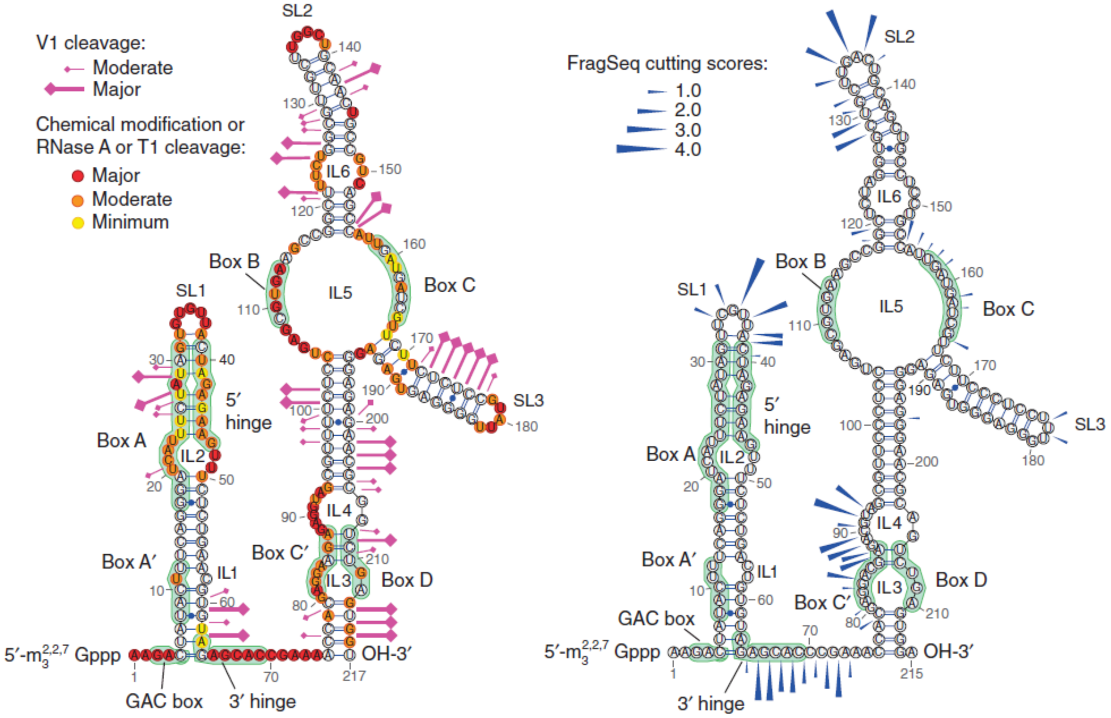
  rCrane [Keating et al, PNAS 2010]



[Assemble, Jossinet *et al* Bioinf. 2010]

Fitting chemical probing data to 2D model

- High-throughput secondary structure determination
- Interactively visualize reactivity data within structural context

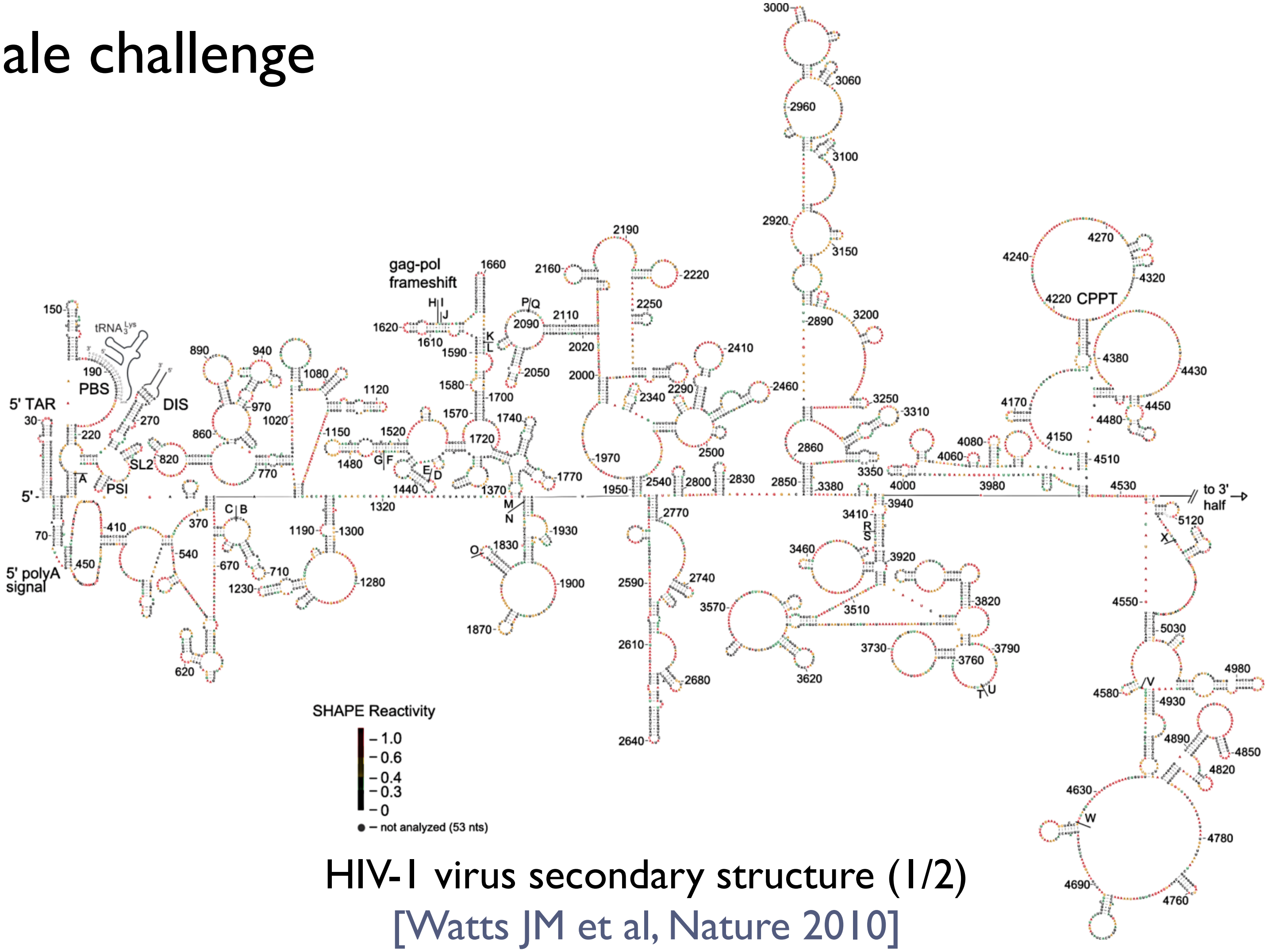


FragSeq method [Underwood *et al*, Nature Methods 2010]

(Images: VARNA)

Fitting chemical probing data to 2D model

- Scale challenge

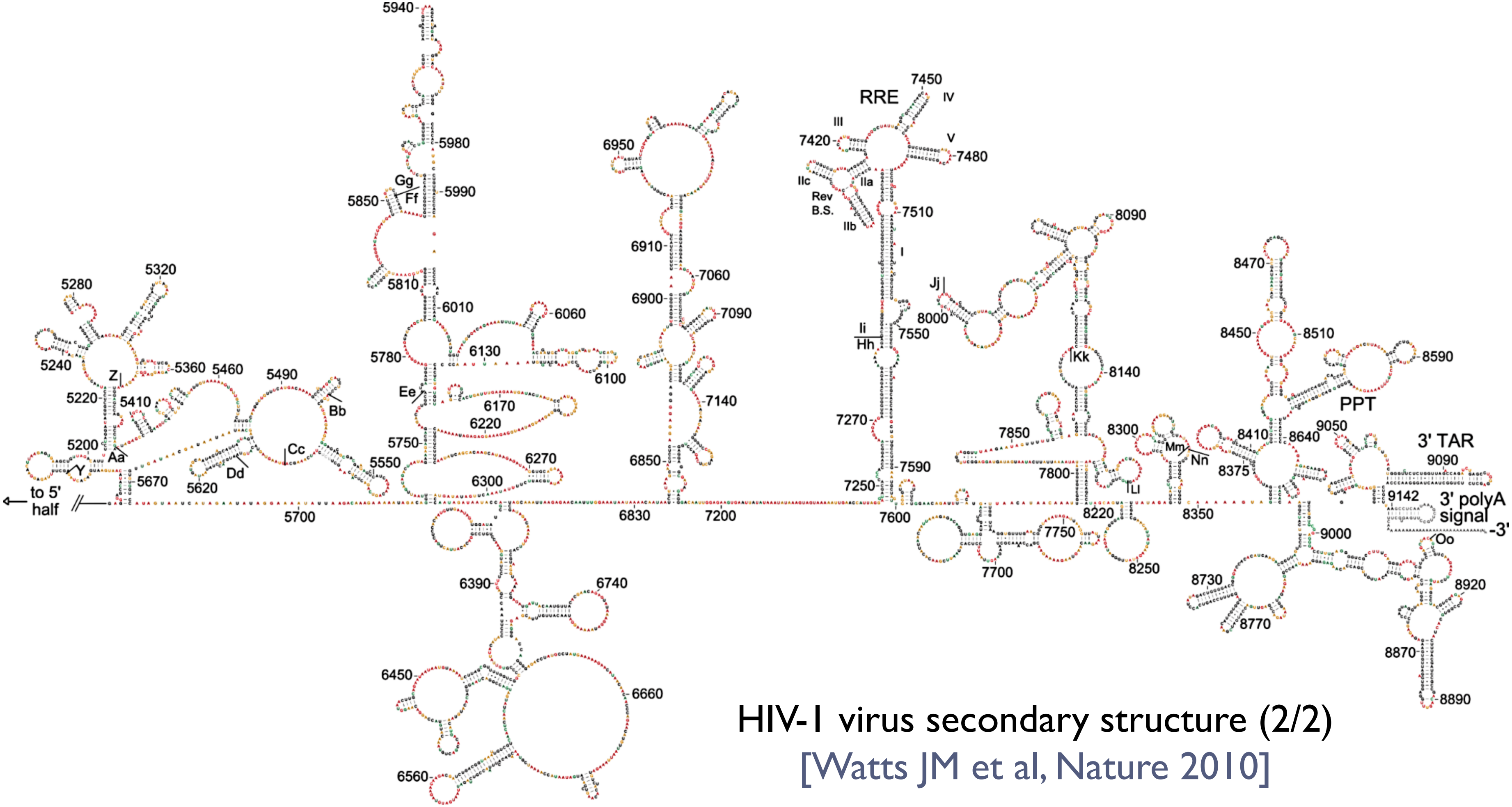


HIV-1 virus secondary structure (1/2)

[Watts JM et al, Nature 2010]

Fitting chemical probing data to 2D model

- Scale challenge

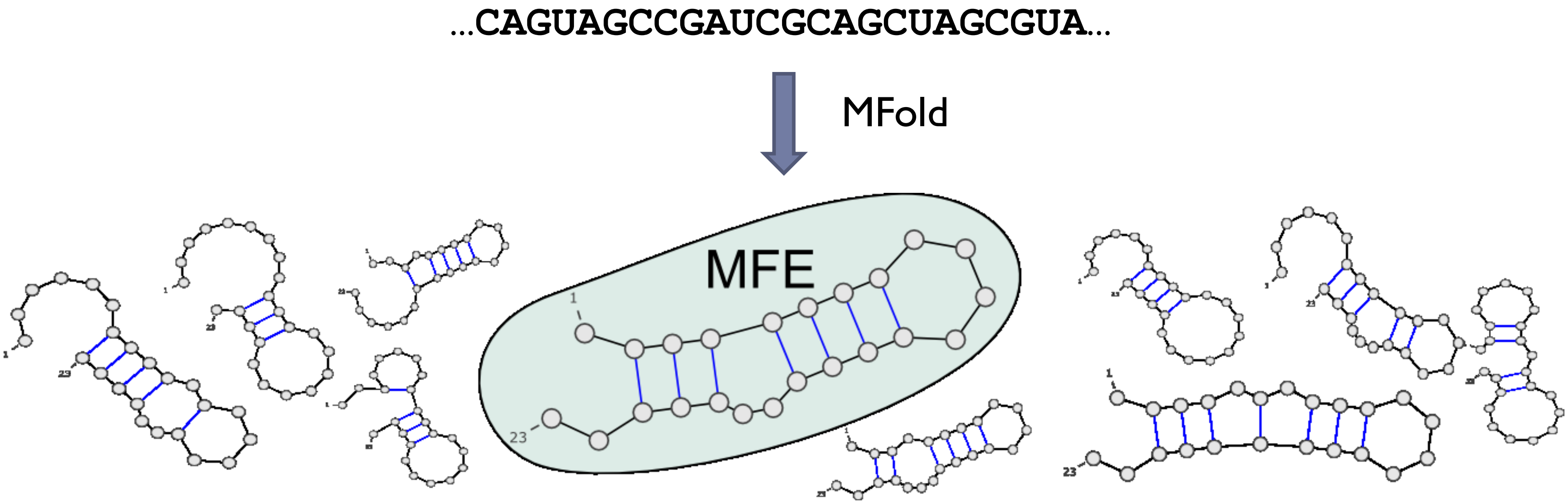


HIV-1 virus secondary structure (2/2)

[Watts JM et al, Nature 2010]

Ensemble approaches in RNA folding

- RNA *in silico* paradigm shift:
  - From single structure, minimal free-energy folding…



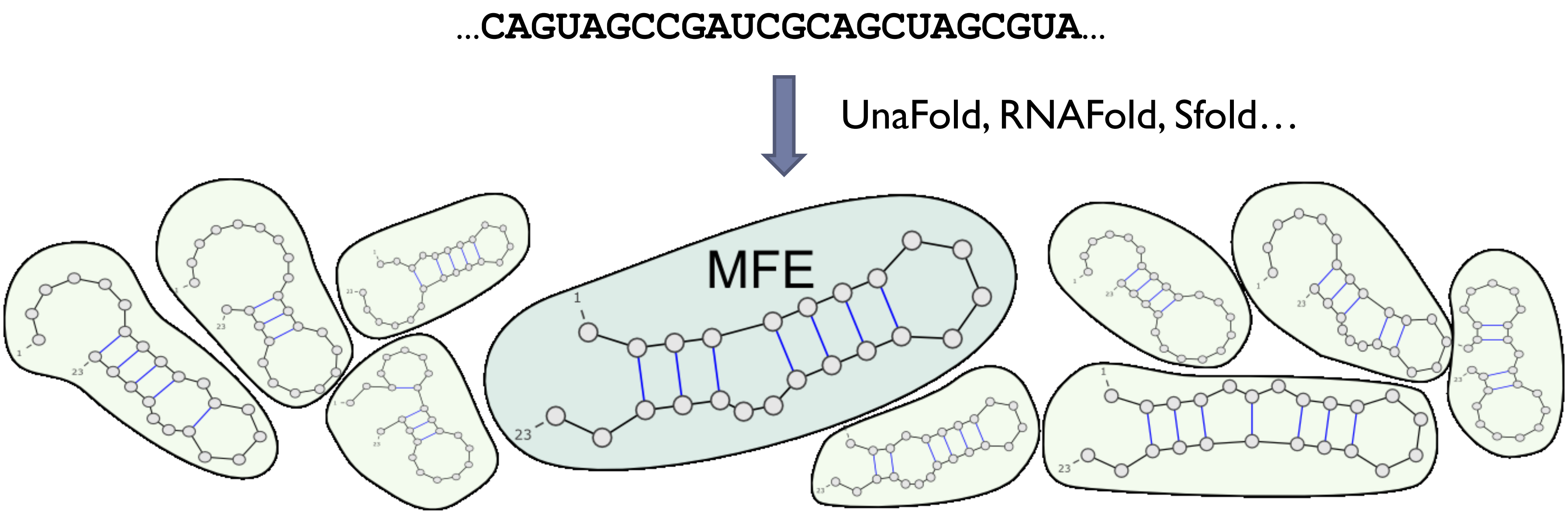
…**CAGUAGCCGAUCGCAGCUAGCGUA**…

MFold

MFE

Ensemble approaches in RNA folding

- RNA *in silico* paradigm shift:
  - From single structure, minimal free-energy folding…
  - … to ensemble approaches.



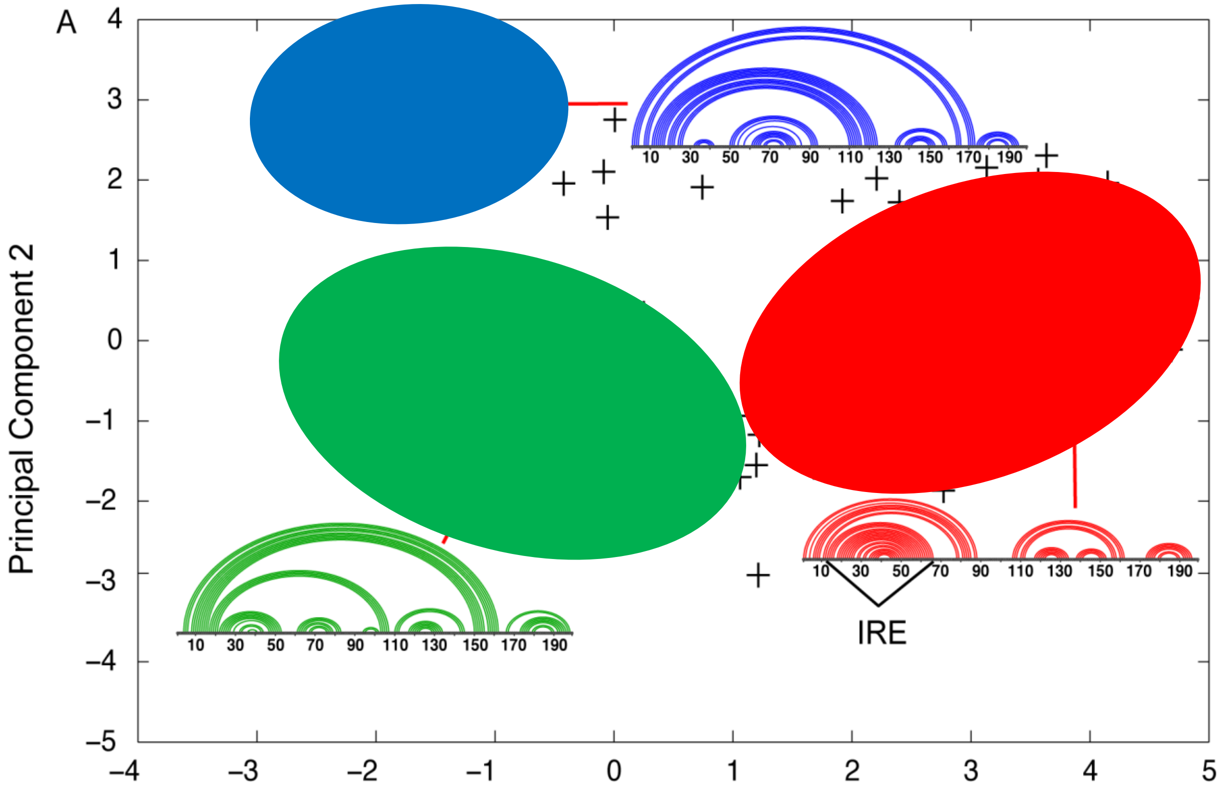
…**CAGUAGCCGAUCGCAGCUAGCGUA**…

UnaFold, RNAFold, Sfold…

Ensemble diversity? Structure likelihood? Evolutionary robustness?

Sensitivity to mutations

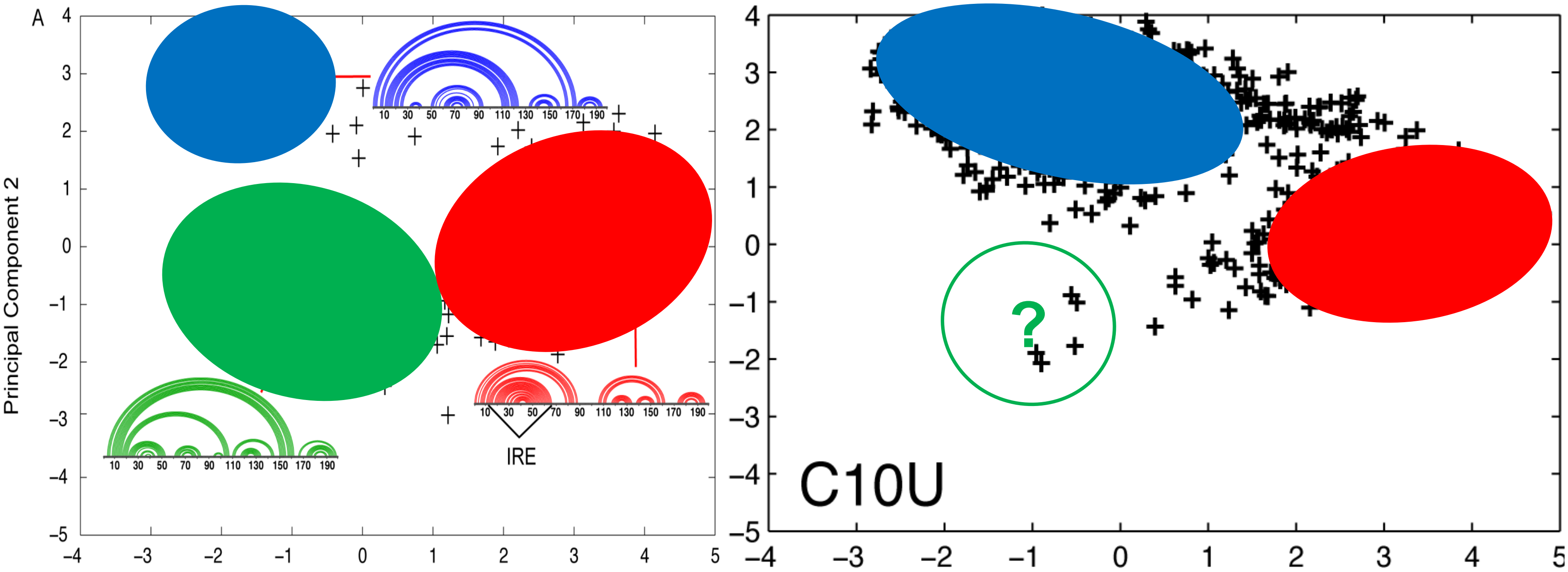
- Boltzmann Sampling → PCA → Clustering



[Halvorsen M *et al*, PLOS Gen 2010]

Sensitivity to mutations

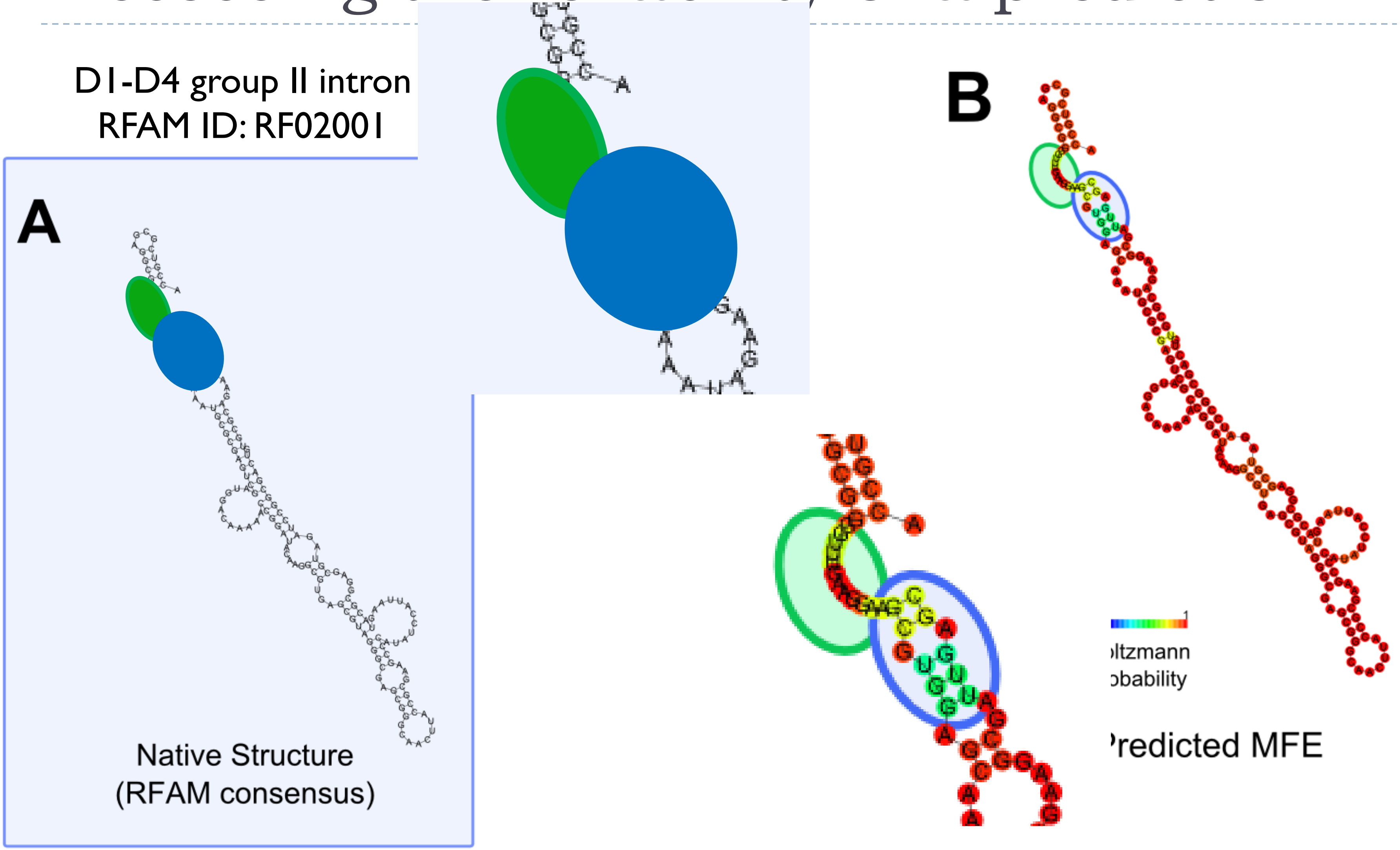
- Boltzmann Sampling → PCA → Clustering



[Halvorsen M *et al*, PLOS Gen 2010]

Assessing the reliability of a prediction

D1-D4 group II intron
RFAM ID: RF02001



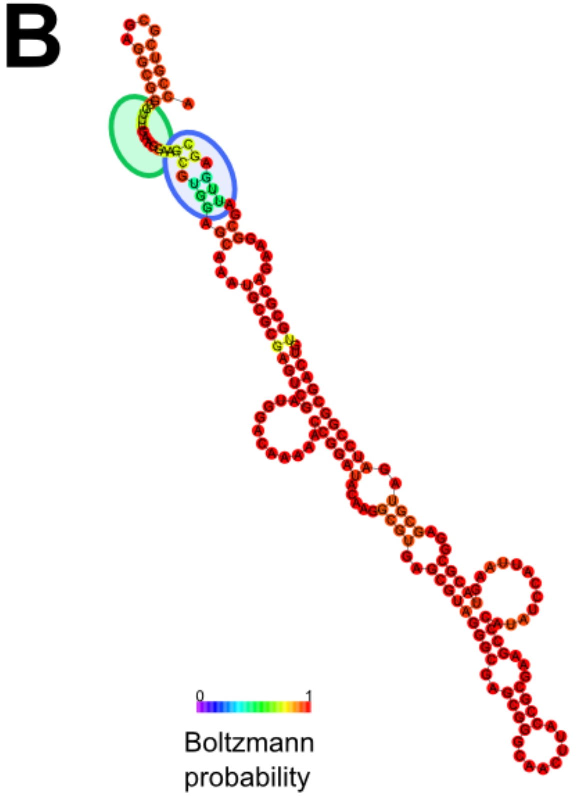
Native Structure
(RFAM consensus)

'redicted MFE

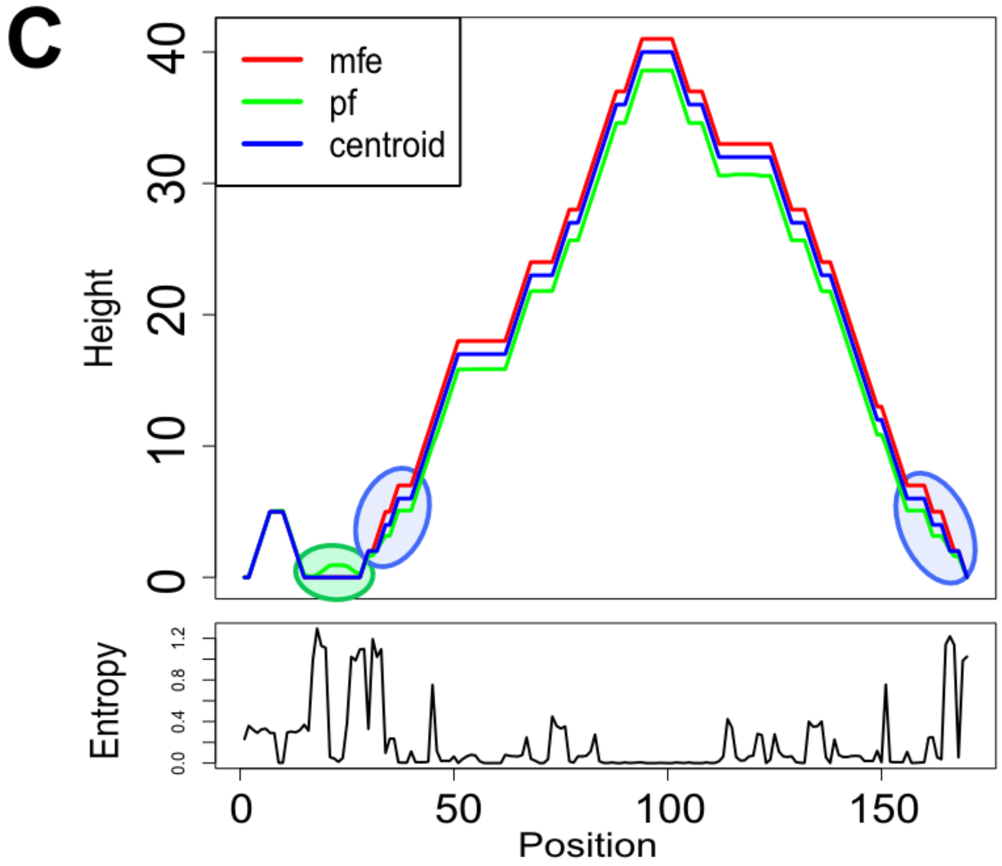
RNAFold [Gruber AR et al. NAR 2008]

Assessing the reliability of a prediction

D1-D4 group II intron
A. Capsulatum sequence



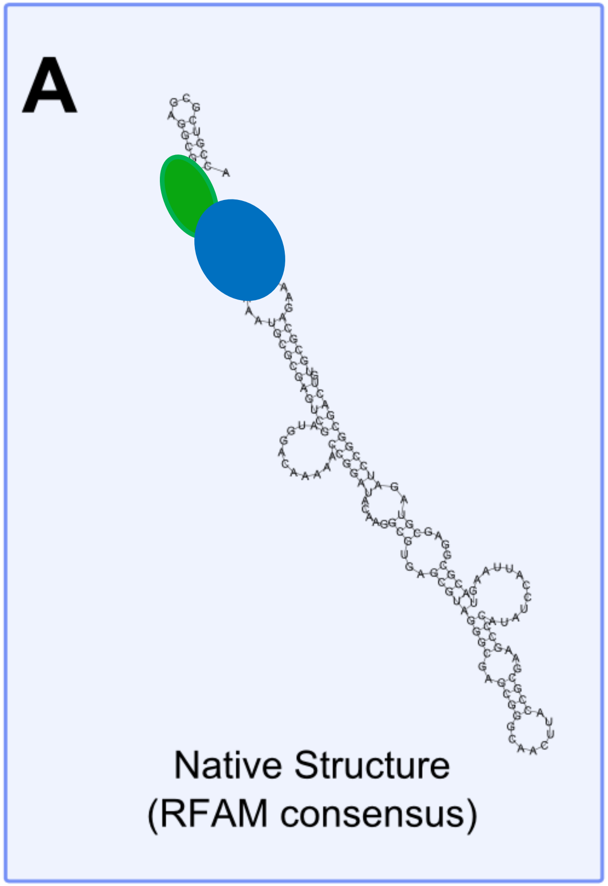
Predicted MFE



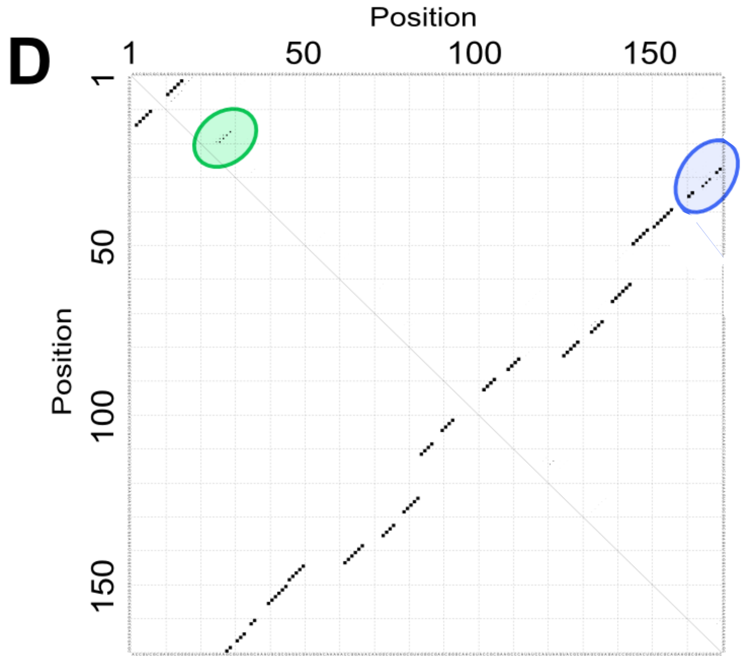
RNAFold [Gruber AR et al. NAR 2008]

Assessing the reliability of a prediction

D1-D4 group II intron
A. Capsulatum sequence



Native Structure
(RFAM consensus)

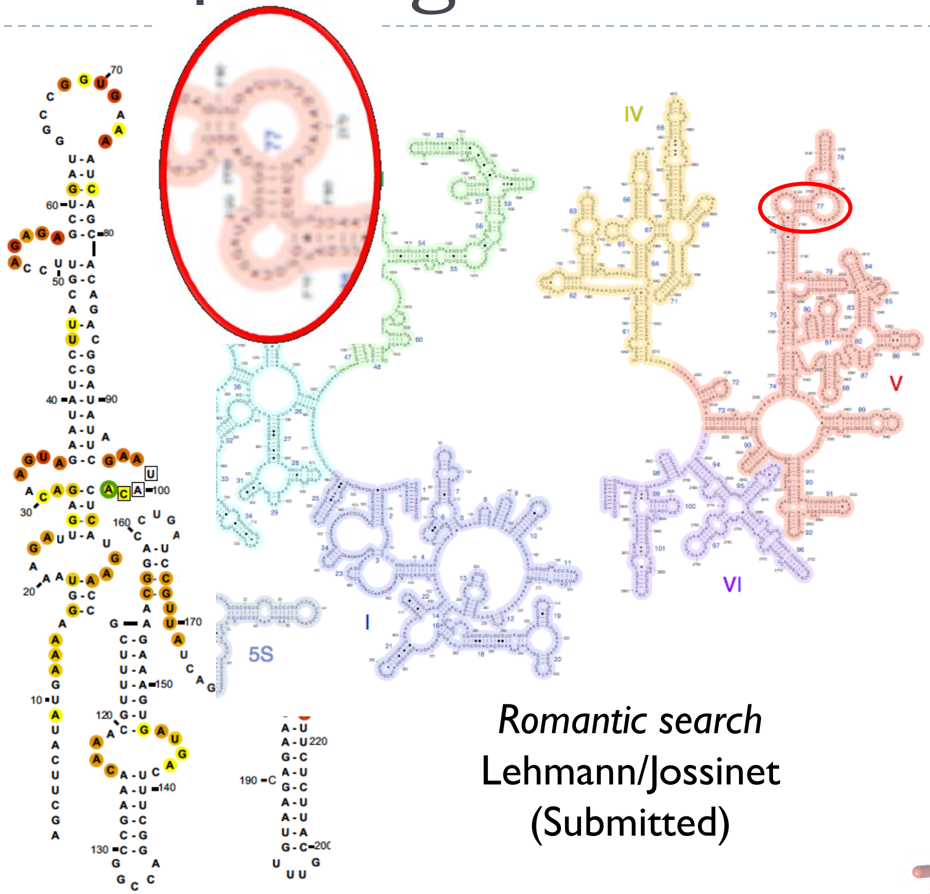


Boltzmann probability - All pairs

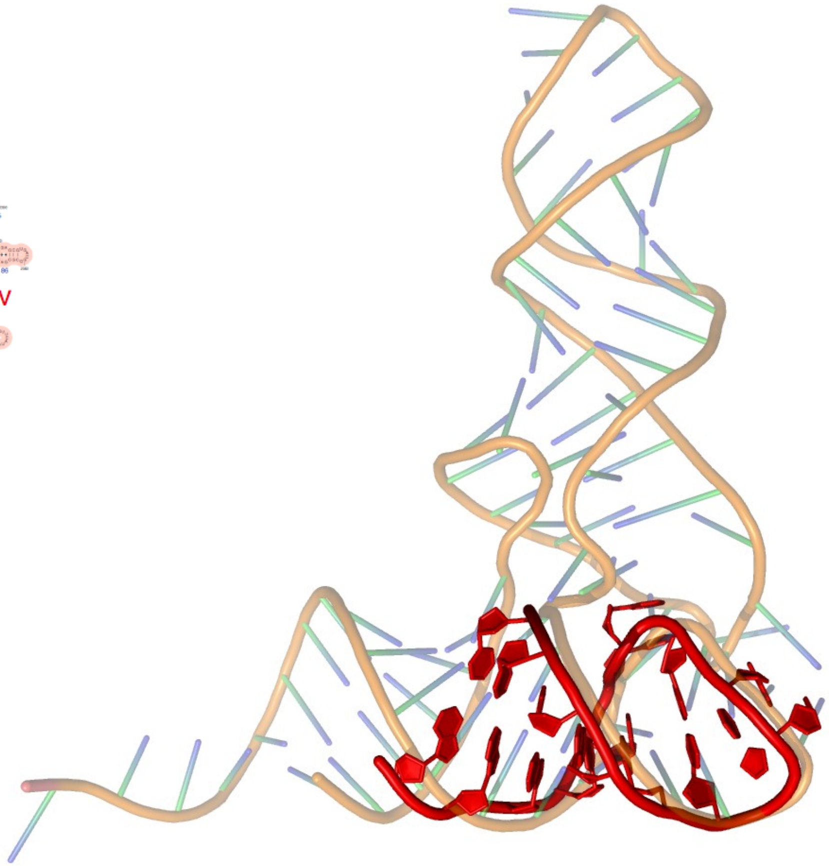
- Low BP probabilities indicate uncertain regions
- BP>99% → Avg. PPV>90% (BP>90% → PPV>83%)
- Visualizing probs in the context of structure helps refining predicted structures.

RNAFold [Gruber AR et al. NAR 2008]

Comparing structures visually



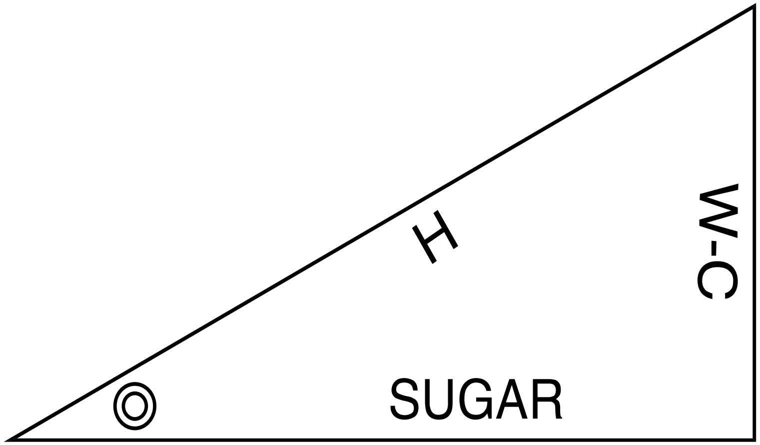
Romantic search
Lehmann/Jossinet
(Submitted)



Fragment of *T thermophylus* tRNA-Phe vs yeast's (PDB: 4TNA & 3BBV)
DARTS [Dror O *et al*, NAR 06] + Pymol

Towards novel representations

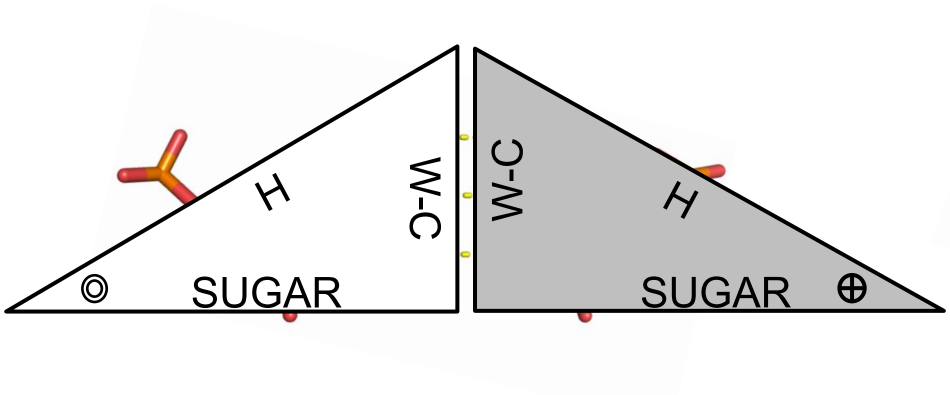
Non canonical/tertiary interactions



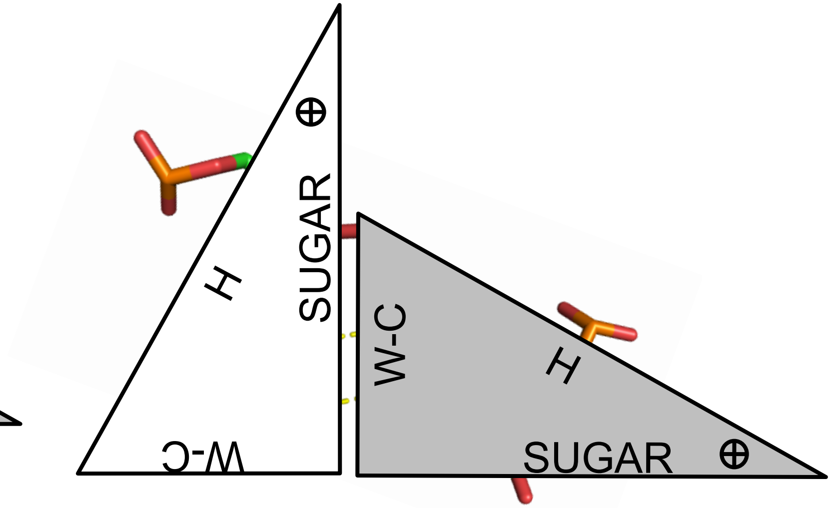
RNA nucleotides bind through edge/edge interactions.

Non canonical are weaker, but cluster into modules that are structurally constrained, evolutionarily conserved, and functionally essential.

Non canonical/tertiary interactions



Canonical G/C pair
(WC/WC cis)

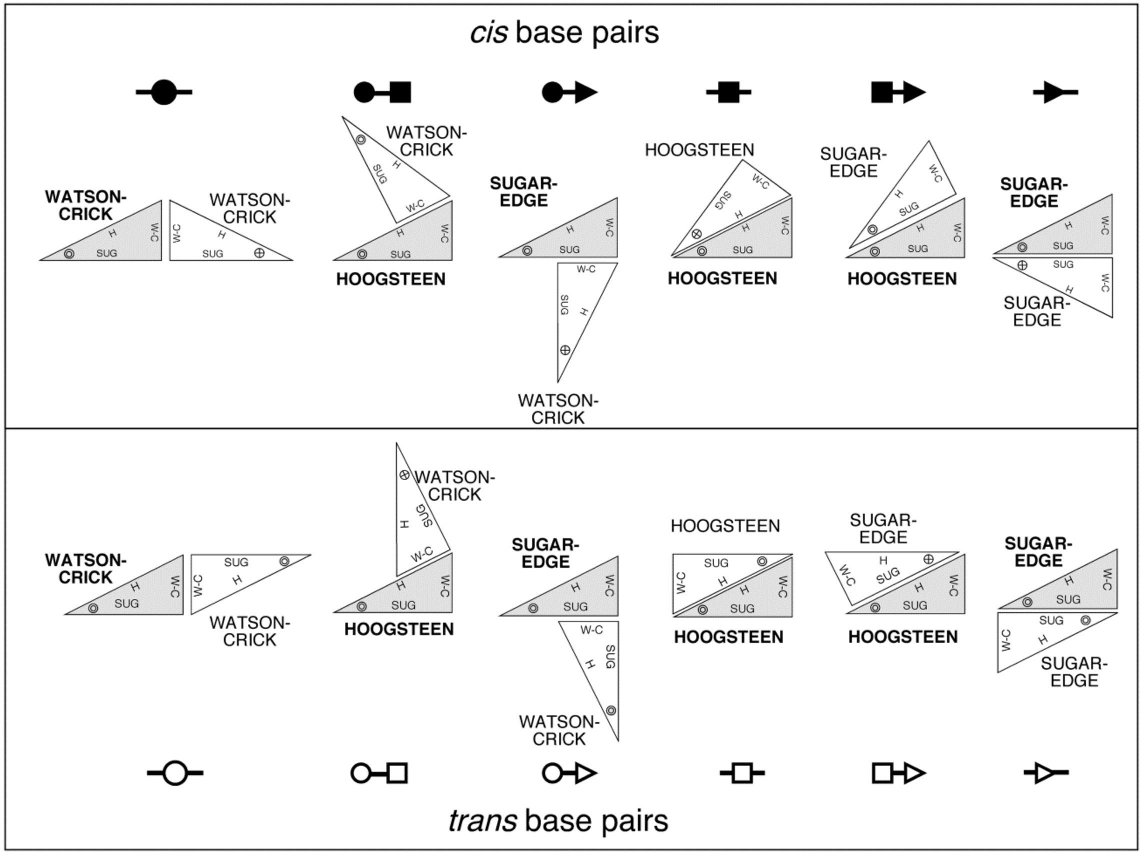


Non Canonical G/C pair
(Sugar/WC trans)

RNA nucleotides bind through edge/edge interactions.

Non canonical are weaker, but cluster into modules that are structurally constrained, evolutionarily conserved, and functionally essential.

Leontis/Westhof nomenclature:
A *visual grammar* for tertiary motifs

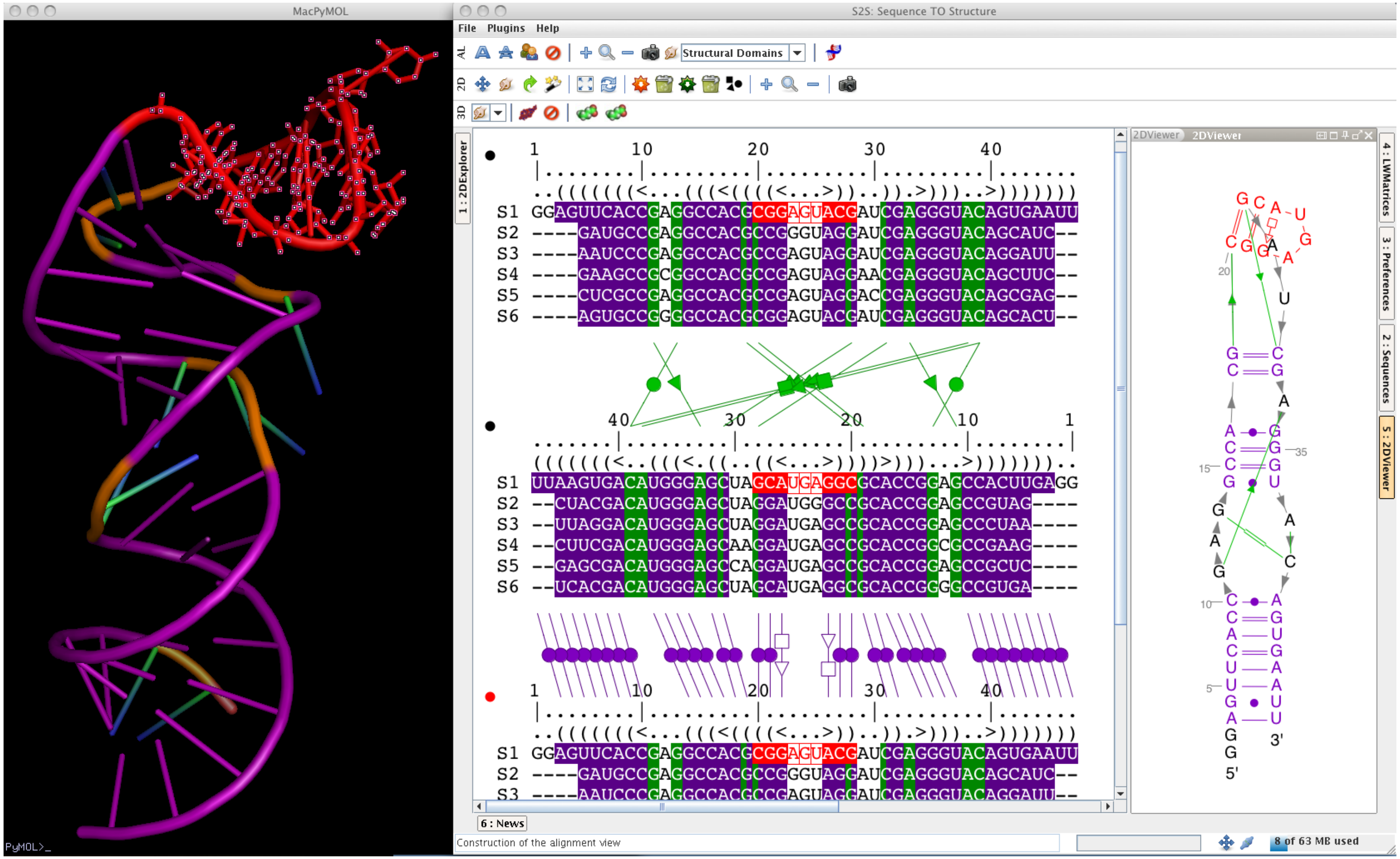


cis	*trans*	
●	○	Watson-Crick edge
■	□	Hoogsteen edge
◀	◁	Sugar edge

—	AU *cis* Watson-Crick base pair
=	GC *cis* Watson-Crick base pair
-○-	GU *cis* Watson-Crick base pair
G or Ⓖ	*syn* oriented
Ⓑ	bifurcated base pair
Ⓦ	water-mediated base pair
----	single H-bond interaction
├─┤	indicates base stacking
→	indicates change in strand orientation

[Leontis/Westhof, NAR 2002]

Leontis/Westhof nomenclature:
A *visual grammar* for tertiary motifs



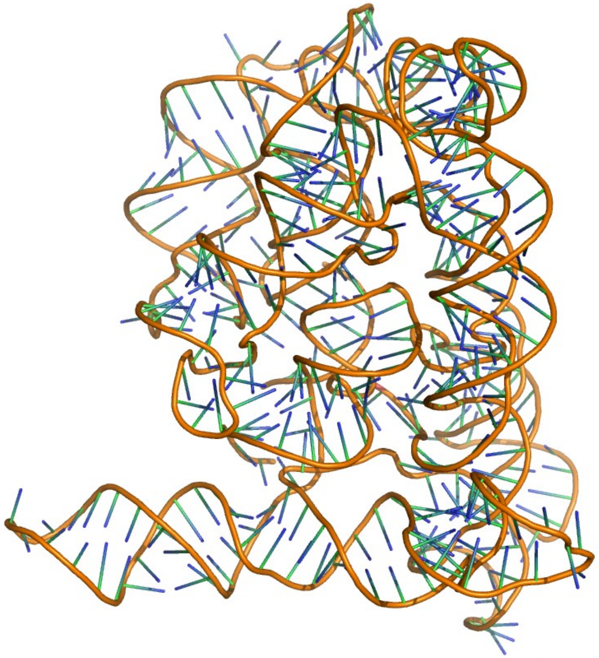
```
 1        10        20        30        40
 |.........|.........|.........|.........|........
 ..((((((((<...(((<(((((<...>))..)).>)))..>))))))))
S1 GGAGUUCACCGAGGCCACGCGGAGUACGAUCGAGGGUACAGUGAAUU
S2 ----GAUGCCGAGGCCACGCCGGGUAGGAUCGAGGGUACAGCAUC--
S3 ----AAUCCCGAGGCCACGCCGAGUAGGAUCGAGGGUACAGGAUU--
S4 ----GAAGCCGCGGCCACGCCGAGUAGGAACGAGGGUACAGCUUC--
S5 ----CUCGCCGAGGCCACGCCGAGUAGGACCGAGGGUACAGCGAG--
S6 ----AGUGCCGGGGCCACGCGGAGUACGAUCGAGGGUACAGCACU--

        40        30        20        10         1
 .......|.........|.........|.........|.........|
 ((((((((<..(((<.((..((<...>)))))>)))...>))))))))..
S1 UUAAGUGACAUGGGAGCUAGCAUGAGGCGCACCGGAGCCACUUGAGG
S2 --CUACGACAUGGGAGCUAGGAUGGGCCGCACCGGAGCCGUAG----
S3 --UUAGGACAUGGGAGCUAGGAUGAGCCGCACCGGAGCCCUAA----
S4 --CUUCGACAUGGGAGCAAGGAUGAGCCGCACCGGCGCCGAAG----
S5 --GAGCGACAUGGGAGCCAGGAUGAGCCGCACCGGAGCCGCUC----
S6 --UCACGACAUGGGAGCUAGCAUGAGGCGCACCGGGGCCGUGA----

 1        10        20        30        40
 |.........|.........|.........|.........|........
 ..((((((((<...(((<(((((<...>))..)).>)))..>))))))))
S1 GGAGUUCACCGAGGCCACGCGGAGUACGAUCGAGGGUACAGUGAAUU
S2 ----GAUGCCGAGGCCACGCCGGGUAGGAUCGAGGGUACAGCAUC--
S3 ----AAUCCCGAGGCCACGCCGAGUAGGAUCGAGGGUACAGGAUU--
```

Construction of the alignment view

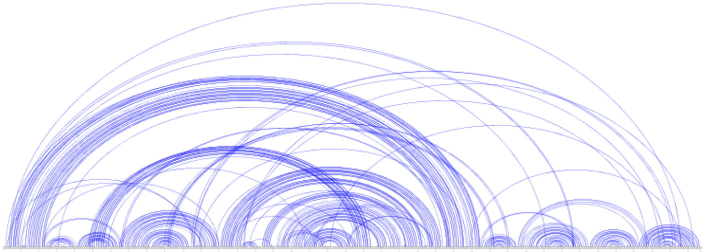
8 of 63 MB used

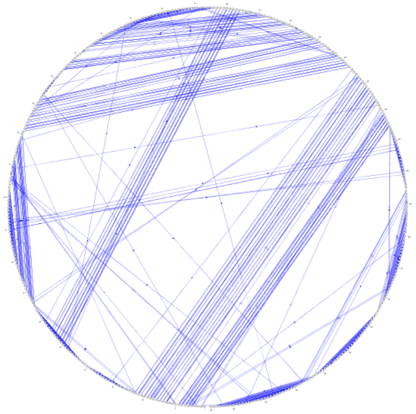
S2S software [Jossinet/Westhof, RNA 2005]

Layout algorithms are challenged by tertiary interactions



Group II Intron (PDB ID: 3GIS)
[Toor N *et al*, RNA 2010]

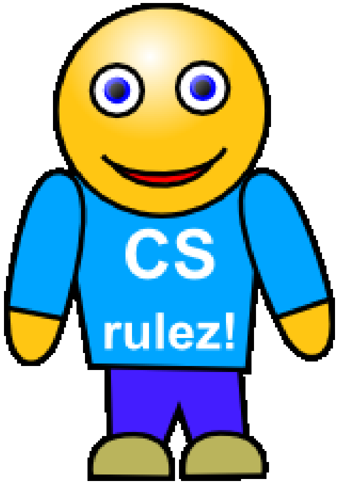




New layout algorithms are needed!
(Multiple views?)

Once upon a time…

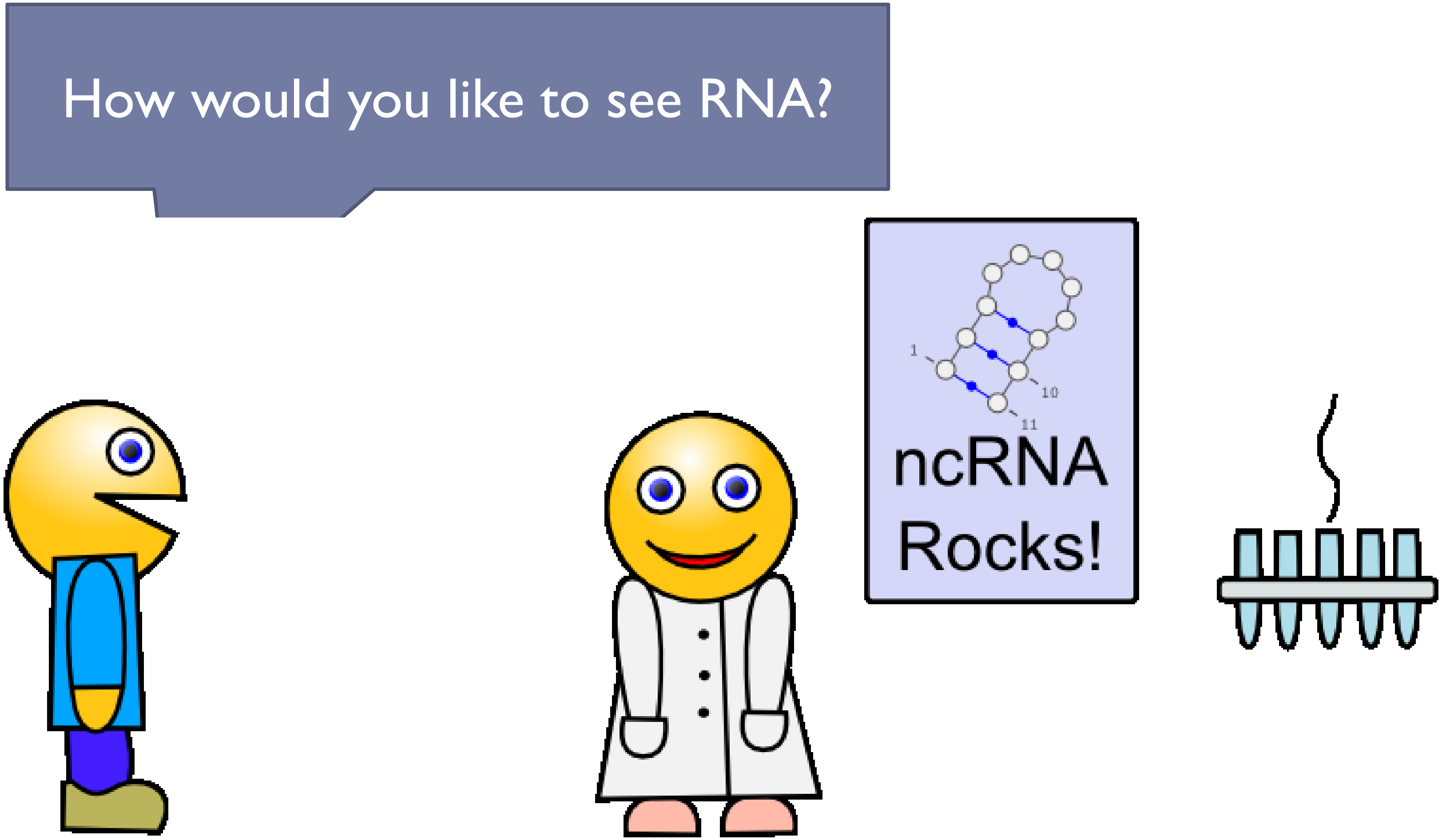
I can draw graphs,
why not draw RNA 2^{ary} structures?



Once upon a time…

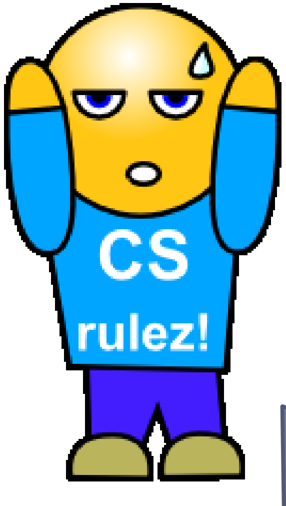
…

Once upon a time…



Once upon a time…

Once upon a time…



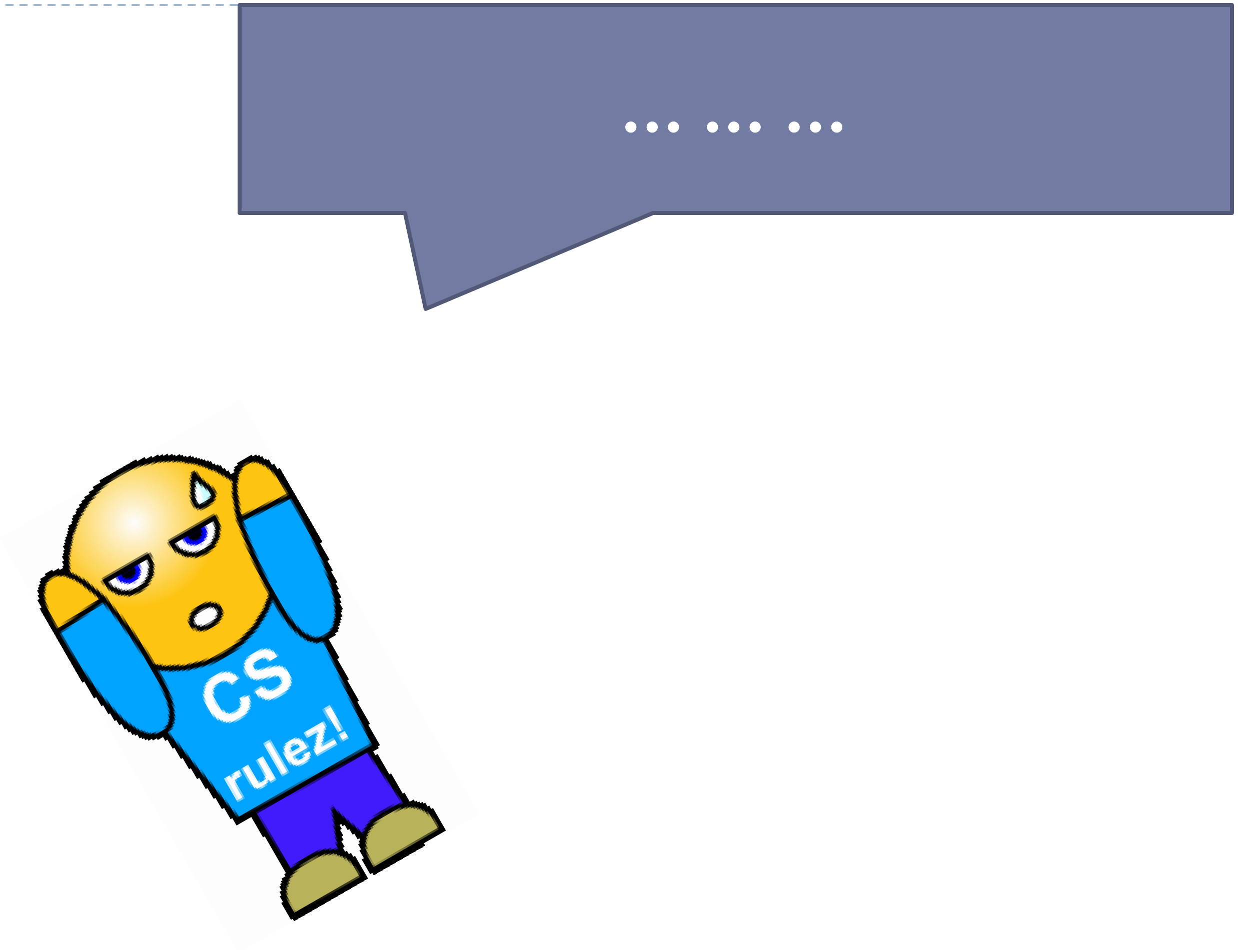
Common sense rules:

- Layout should be non overlapping
- Inner loops = Circular support
- Helices = Straight lines
- Consecutive bases = Equally distant

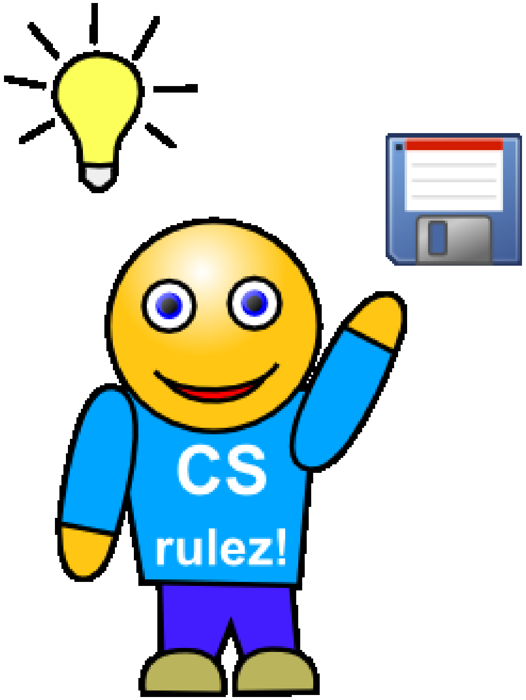
Satisfying these rules makes the problem NP-Hard, but we can still decently approximate it, assuming that …

… APX … greedy … dynamic programming … P=NP(?)…

Once upon a time…



Once upon a time…

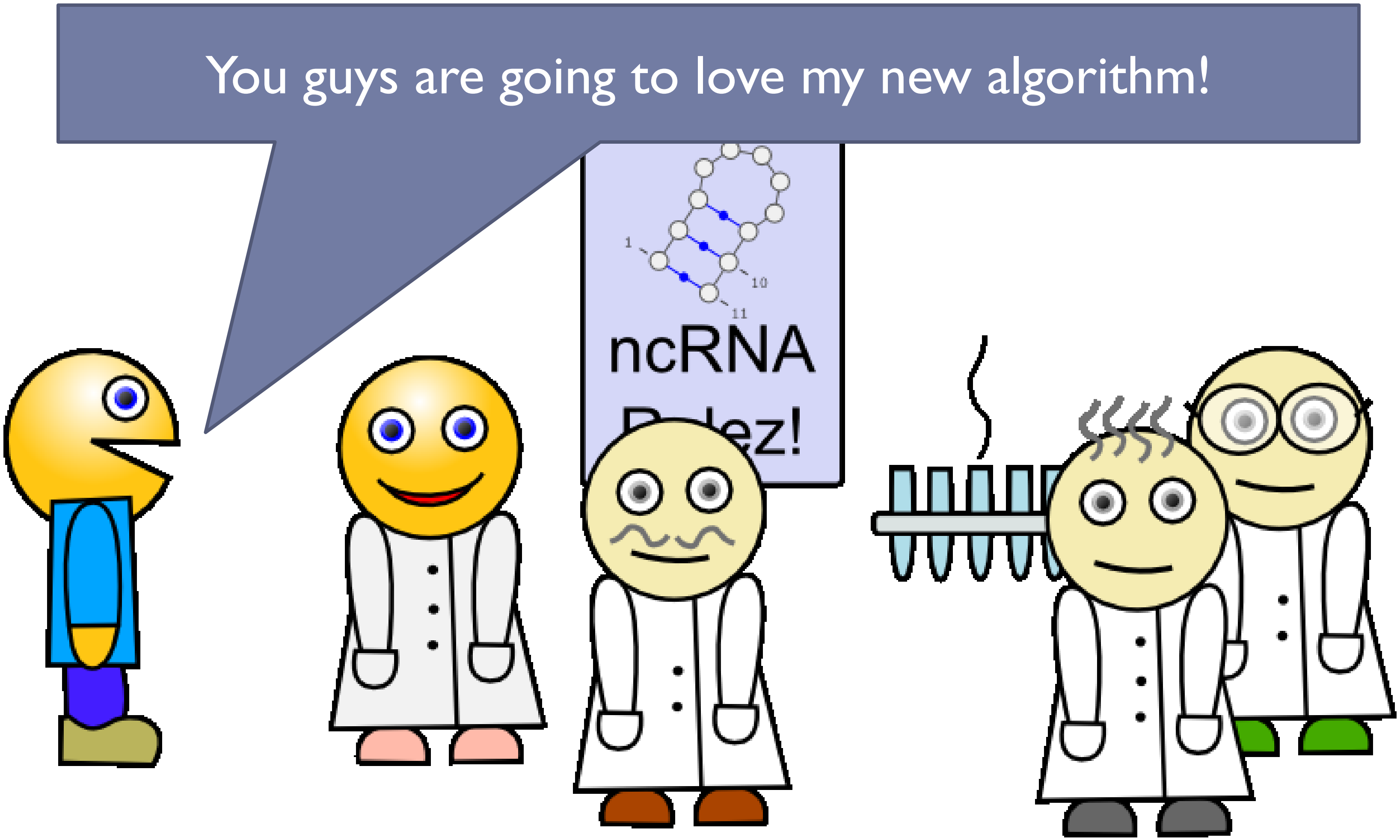


Common sense rules:

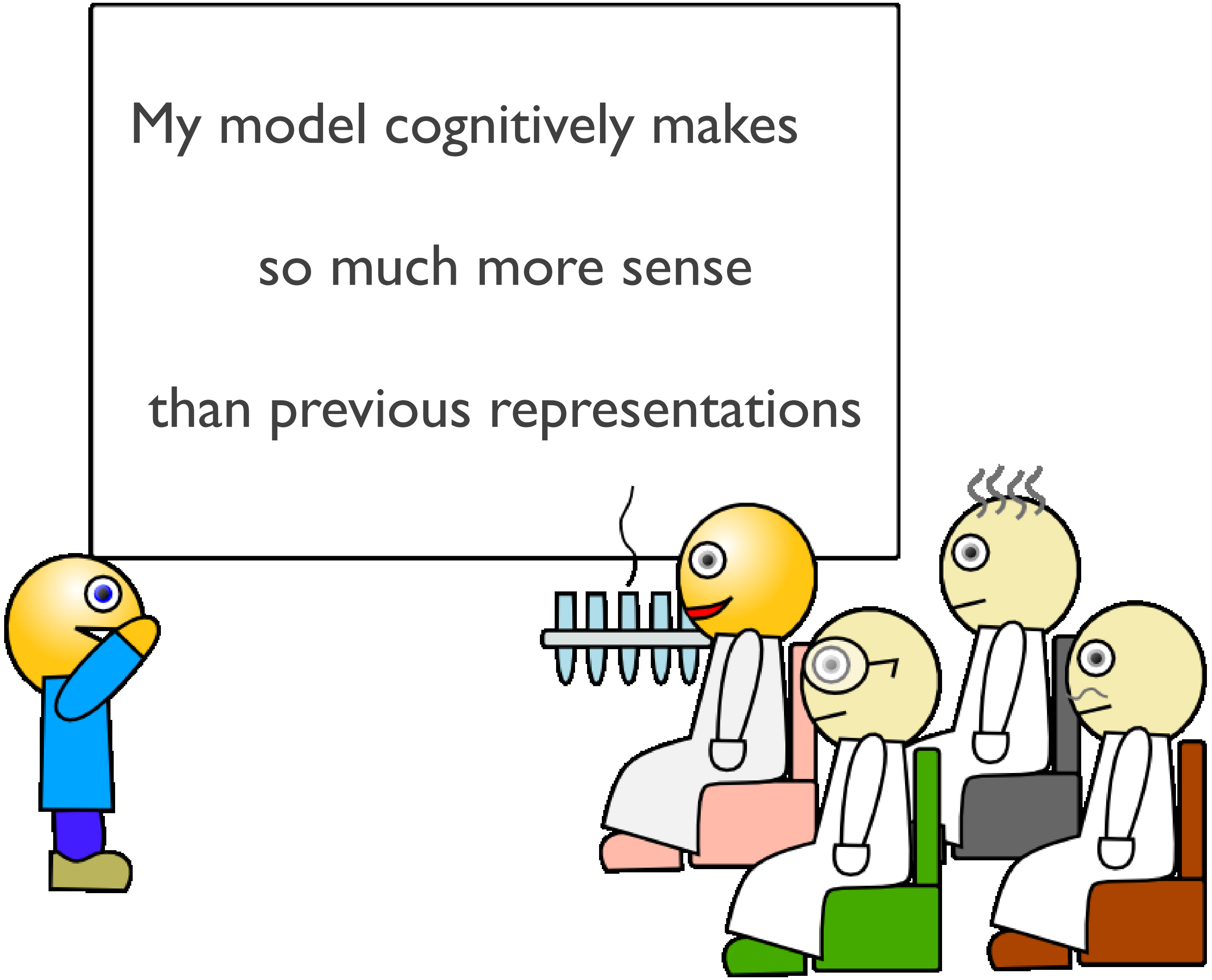
- Layout should be non overlapping
- Inner loops = Circular support
- Helices = Straight lines
- Consecutive bases = Equally distant

\+ Ninja algorithmic skills
\+ Hard work
= Pretty decent algorithm

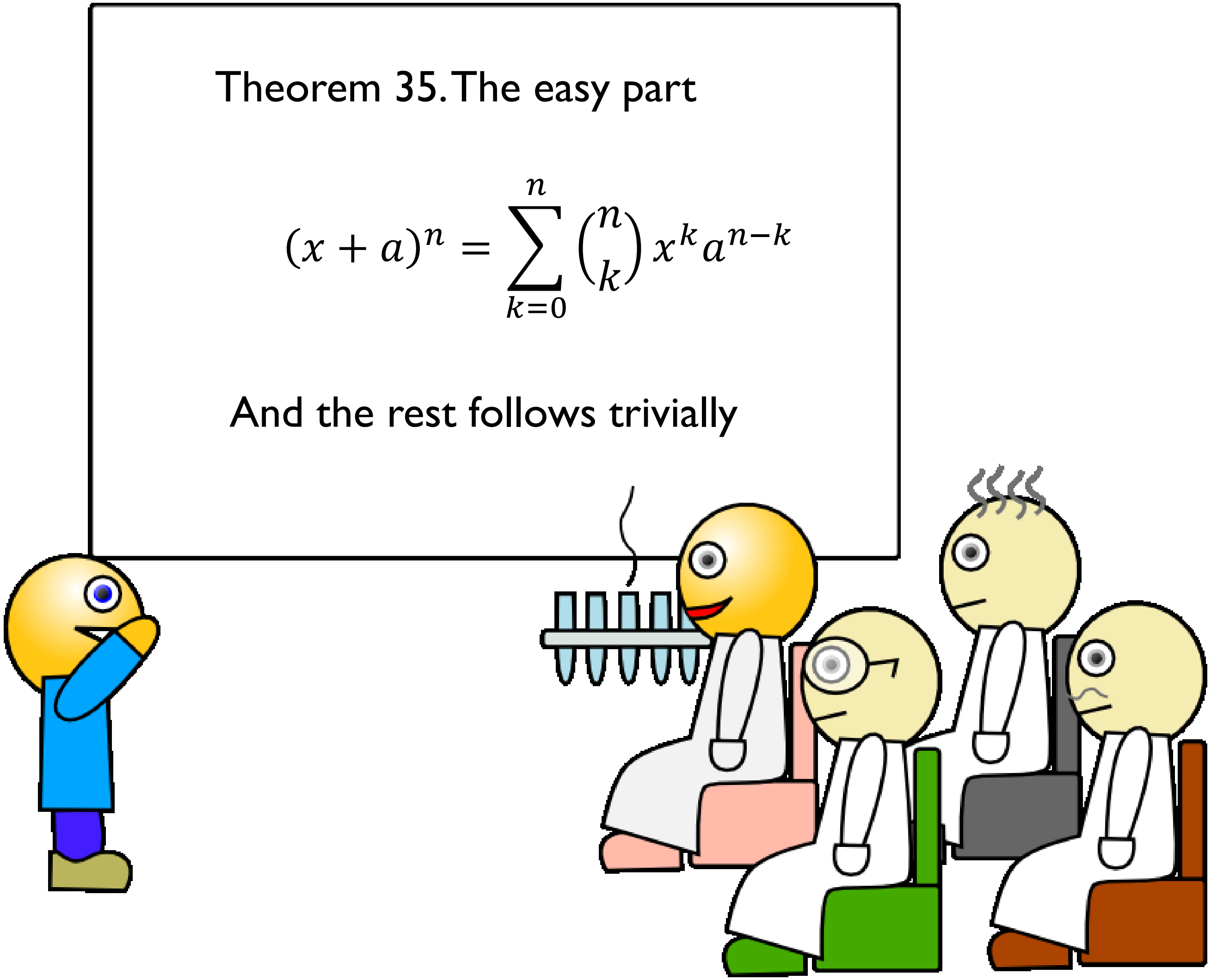
Once upon a time…



Once upon a time…

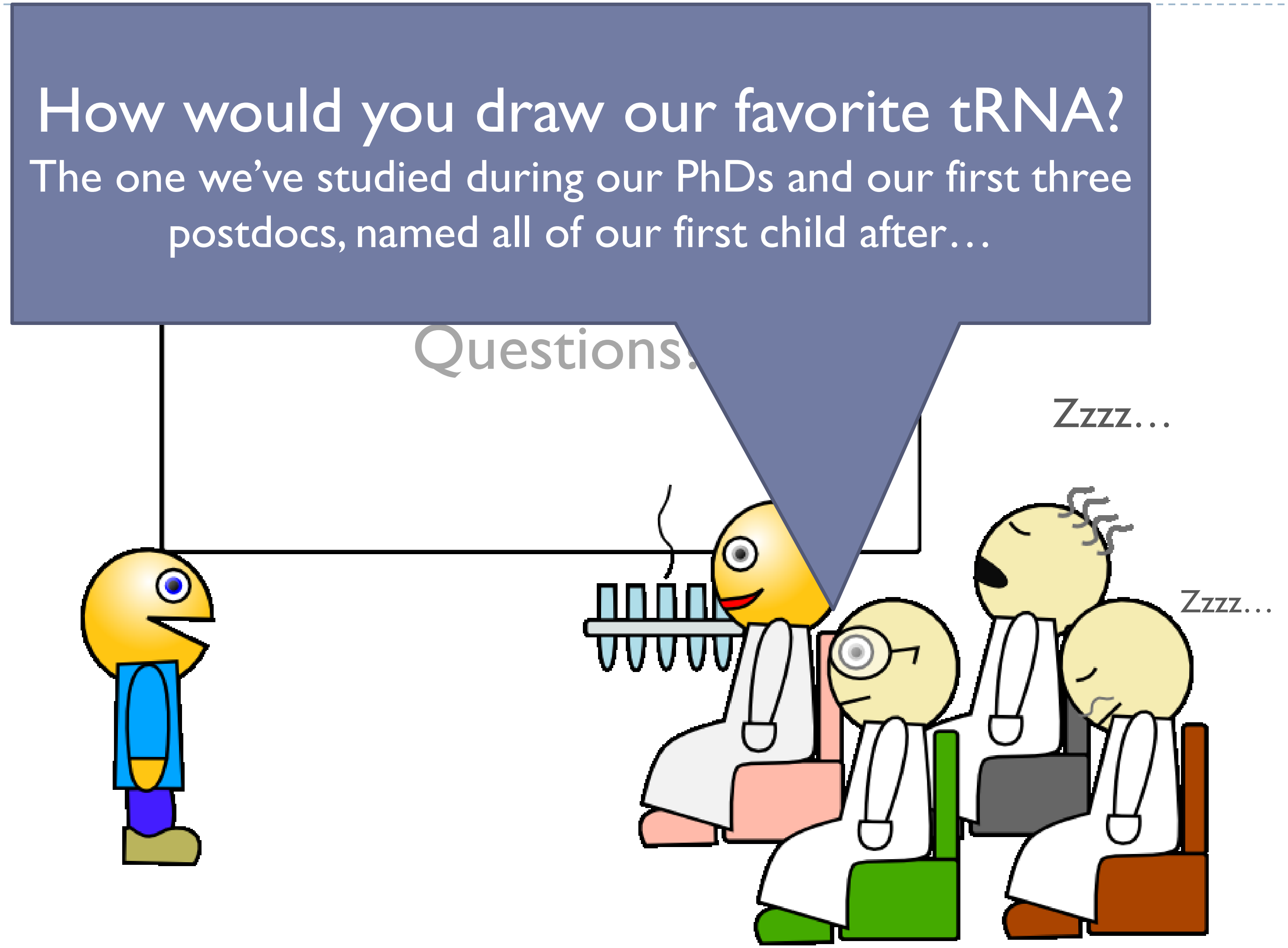


Once upon a time…

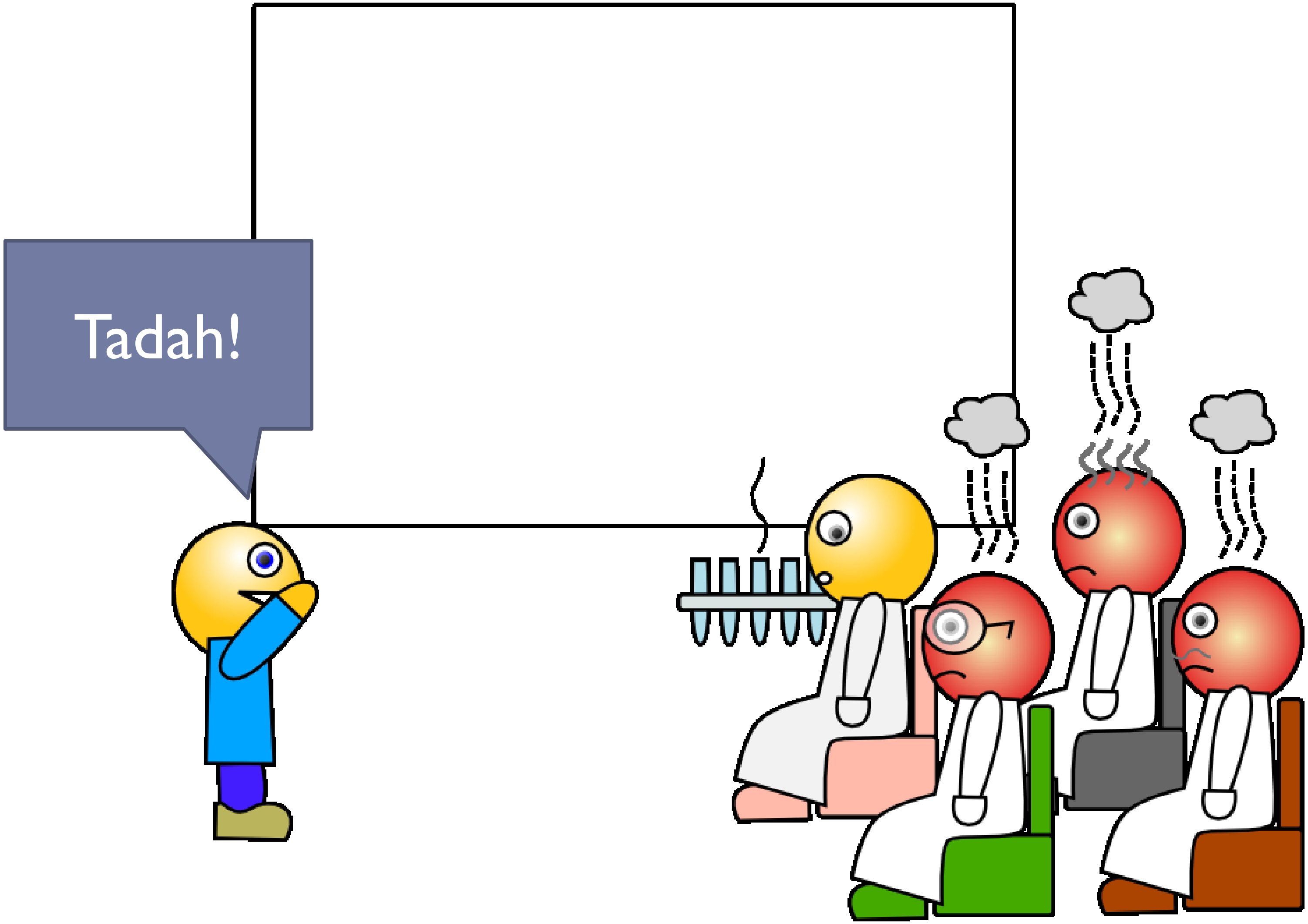


Once upon a time…

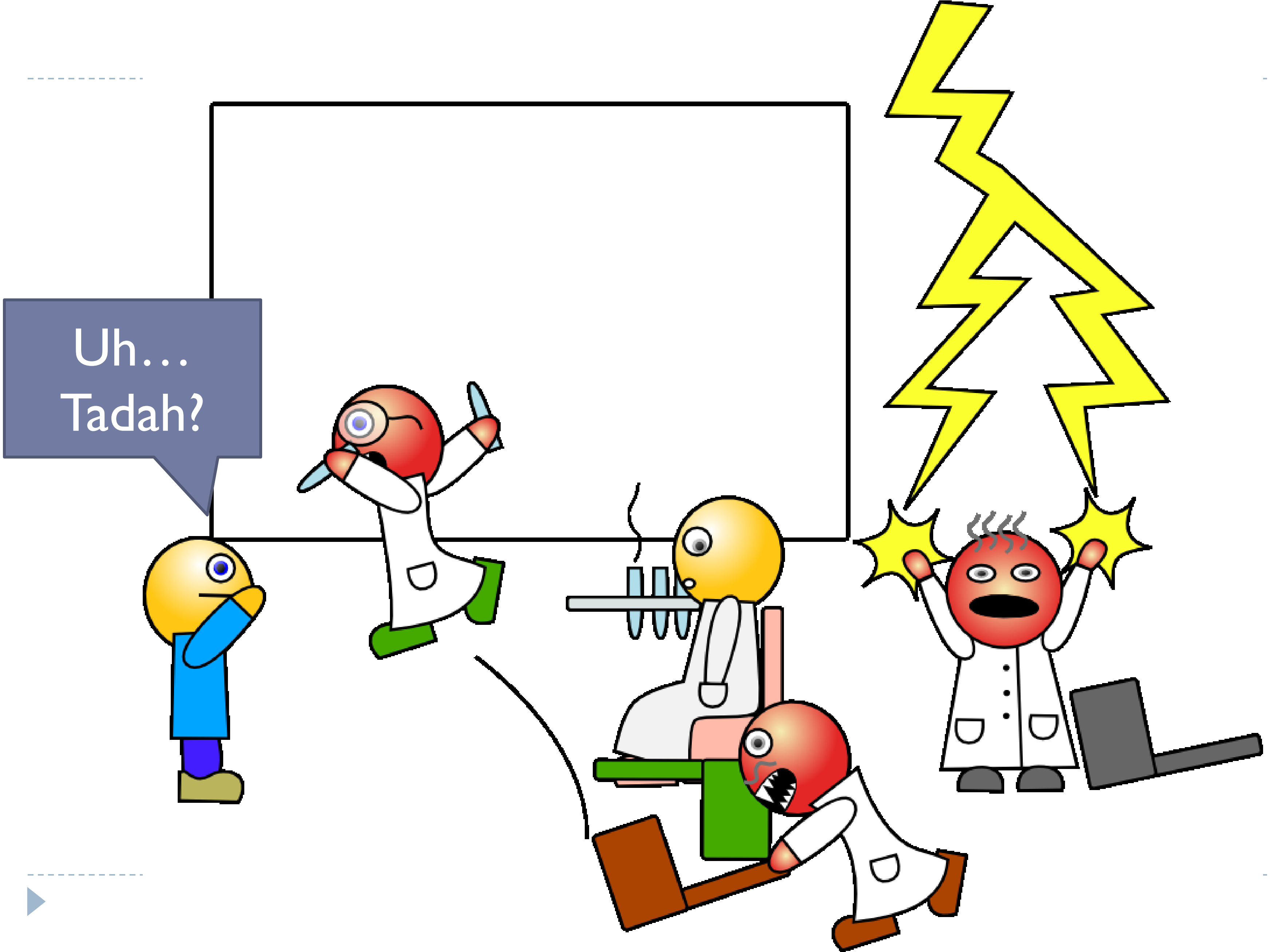
Once upon a time…



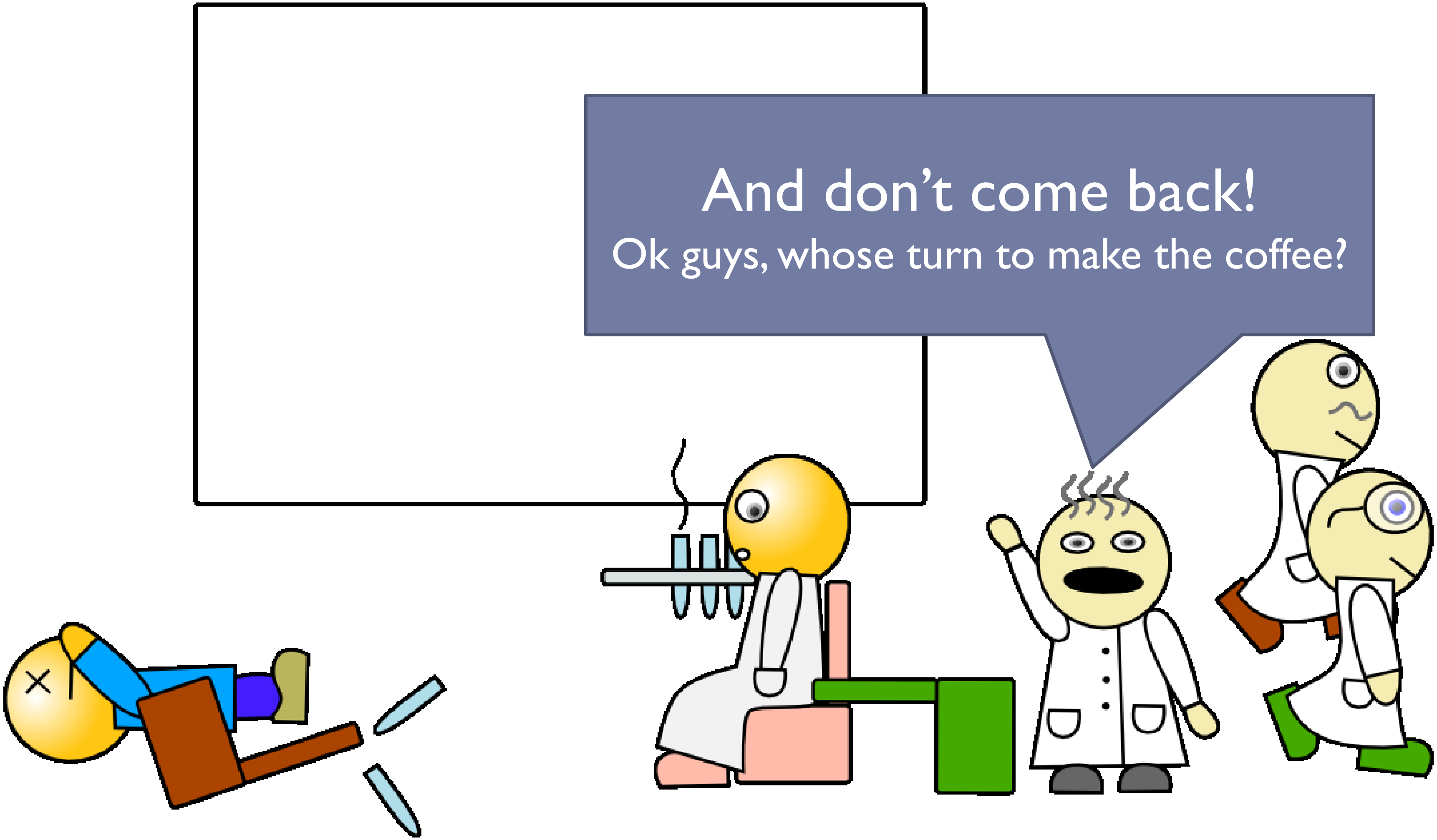
Once upon a time…



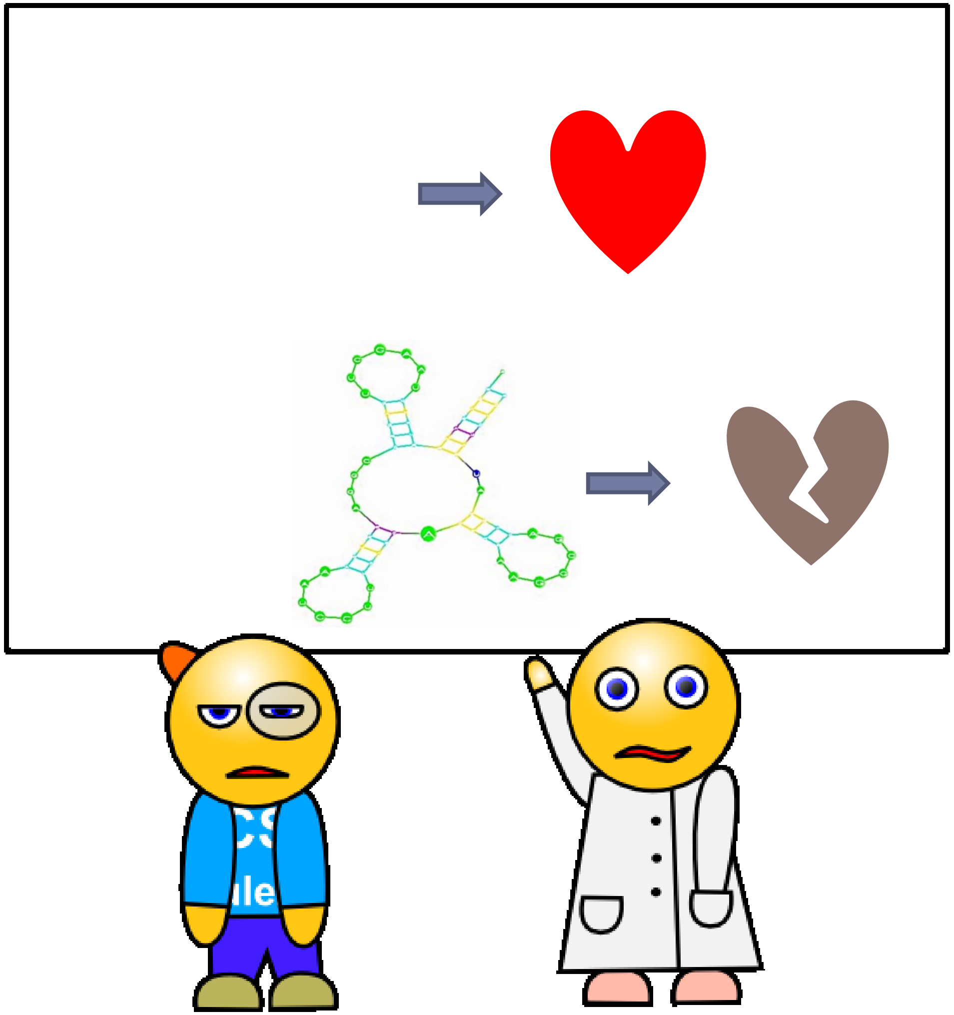
Once upon a time…



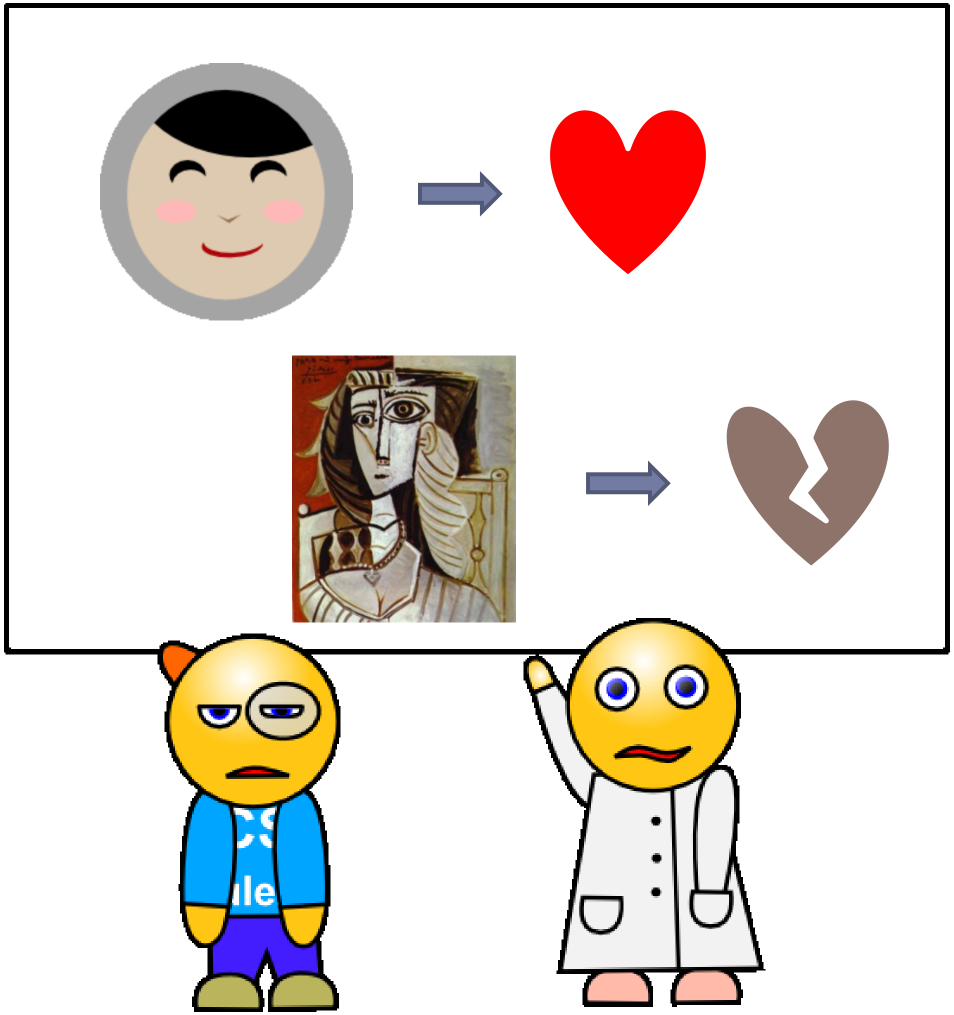
Once upon a time…



Once upon a time…



Once upon a time…



What I learned

Don’t mess with the RNA biologists:

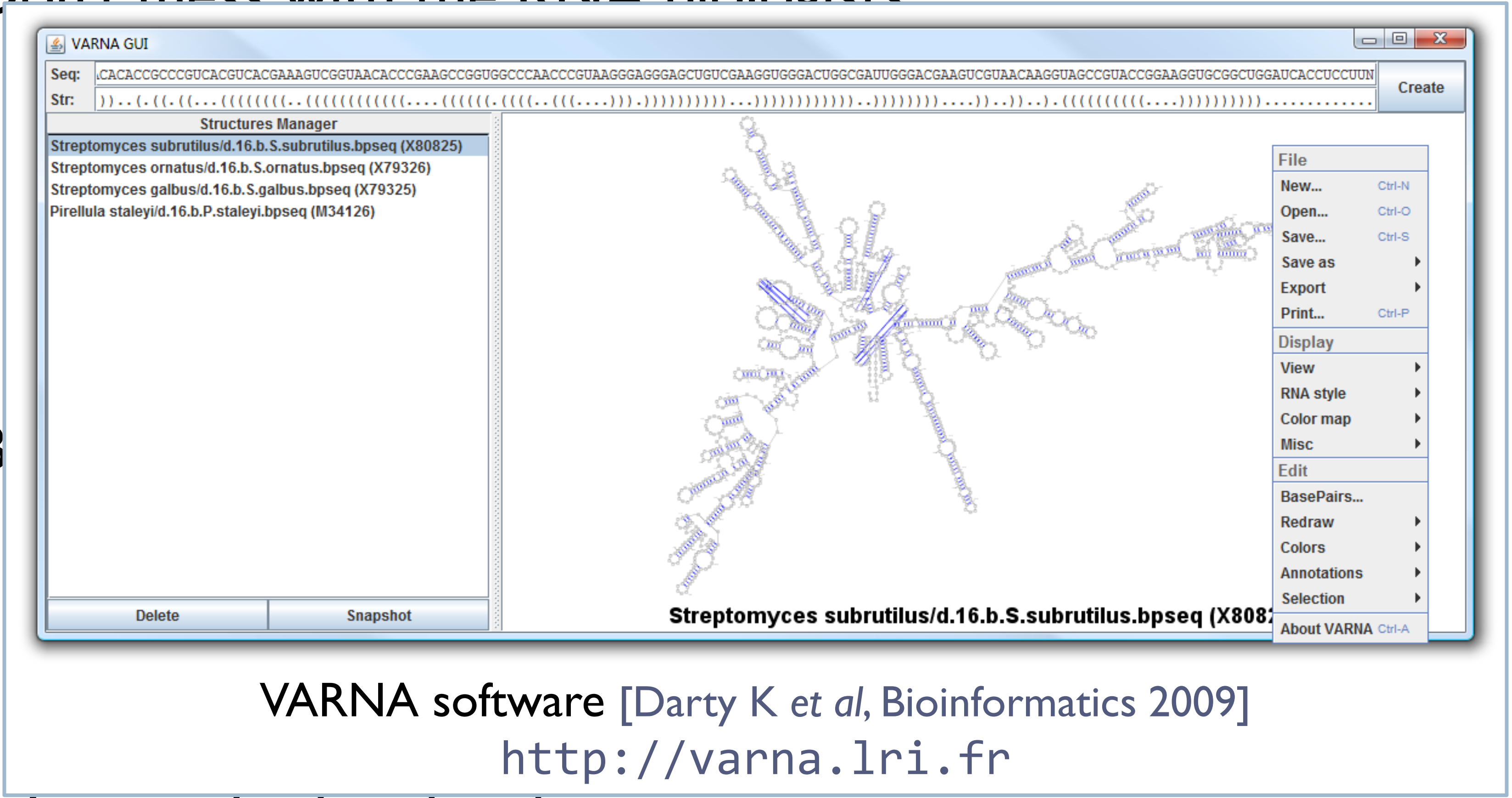
- Offer as many algorithms as humanly possible
- Interactive editing gestures for “historical” layouts
- Templating mechanisms

But indulge your inner geek:

- Cross-platform
- Open source
- Generic component within third-party tool
- Java applet for data bases…

What I learned

Don't mess with the RNA biologists:



VARNA GUI

Seq: CACACCGCCCGUCACGUCACGAAAGUCGGUAACACCCGAAGCCGGUGGCCCAACCCGUAAGGGAGGGAGCUGUCGAAGGUGGGACUGGCGAUUGGGACGAAGUCGUAACAAGGUAGCCGUACCGGAAGGUGCGGCUGGAUCACCUCCUUN

Str:))..(.((.((...((((((((..((((((((((((....((((((.((((..(((....))).))))))))))...))))))))))))..))))))))....))..))..).((((((((((....))))))))))..............

Structures Manager

- Streptomyces subrutilus/d.16.b.S.subrutilus.bpseq (X80825)
- Streptomyces ornatus/d.16.b.S.ornatus.bpseq (X79326)
- Streptomyces galbus/d.16.b.S.galbus.bpseq (X79325)
- Pirellula staleyi/d.16.b.P.staleyi.bpseq (M34126)

Streptomyces subrutilus/d.16.b.S.subrutilus.bpseq (X808

VARNA software [Darty K *et al*, Bioinformatics 2009]

`http://varna.lri.fr`

B

- Java applet for data bases…

Conclusion

Conclusion

Increasing need for visualization:

- More and bigger structural models
- Emerging need for interactive methods:
  - Identification of functional modules
  - Model fitting to probing data

Integrated RNA-specific visualization methods/tools needed for:

- RNA/RNA Interactions
- Automated layout of tertiary motifs (modules)
- Visualization of structure ensembles (Qualitative vs Quantitative)
- Kinetics, folding pathways
- Structure/sequence evolution

Acknowledgements

VARNA crew

- Raphael Champeimont (U Paris 6)
- Kevin Darty (U Paris Sud)
- Alain Denise (U Paris Sud)

VIZBI conference

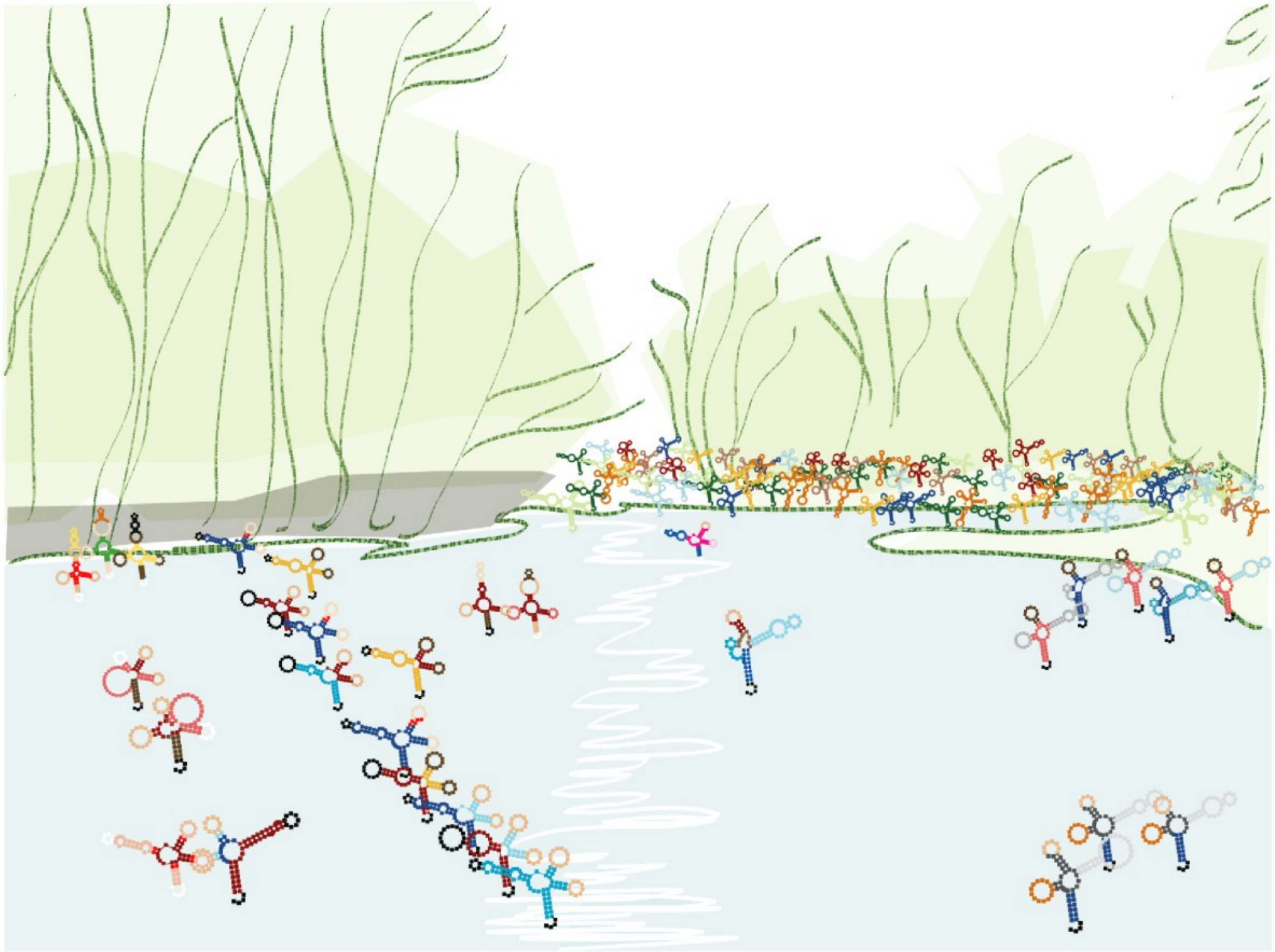
- Jim Procter (+JalView)
- Sean O'Donoghue

VIZBI RNA chapter crew

- Kornelia Aigner (Uni Düsseldorf)
- Fabian Dressen (Uni Düsseldorf)
- Valérie Fritsch (Uni Strasbourg)
- Tanja Gesell (Uni Vienna)
- Fabrice Jossinet (Uni Strasbourg)
- Gerhard Steger (Uni Düsseldorf)
- Eric Westhof (Uni Strasbourg)

Every VARNA user out there…

Questions?



tRNA cloverleaf shape members (skating on a winter pond)
RNArt by S. Konermann *in* Voss B *et al*, BCM Biology 2006