RNA 3D and 2D structure

Yann PONTY
CNRS/Ecole Polytechnique
A redundant talk... sorry!

•

•

•

•

•

•
Why RNA is so COOL!

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
Why RNA is so COOL!

- Ubiquitous
- Pervasively expressed
- Versatile
- Easy to handle

Why RNA is so **COOL!**

- Ubiquitous
- Pervasively expressed
- Versatile
- Easy to handle

---

**Li 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

- 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

⇒ !in silico!

Protein: 73651 hits, 92.6%
Mixed: 3629 hits, 4.6%
DNA: 1328 hits, 1.7%
RNA: 890 hits, 1.1%

PDB entries (Feb 2012)
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
Current visualization of RNA
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

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

Diagram showing:
- miRNAs (~20)
- tRNAs (~80)
- 5s rRNAs (~350)
- snoRNAs (~130)
- Group I intron (~320)
- OLE RNAs (~600)
- 16s rRNAs (~1.5k)
- 16s Ribosomal RNAs (~1.5k)
- 16s Odontoglossum ringspore virus (~6.6k)
- HIV-1 (~9.2k)
- Mature Micro RNAs (~10-25)

Legend:
- Hammerhead Ribozymes
- sRNAs (~50-250)
- Spliceosomal RNAs (~160-570)
- Hepatitis D RNA genome (~1.7k)
- 23s rRNAs (~2.9k)
Fitting 3D model to density maps

- Cryo-EM maps
- Assemble [Jossinet et al, Bioinf. 2010]
- Semi-automated rCrane [Keating et al, PNAS 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]
Fitting chemical probing data to 2D model
Fitting chemical probing data to 2D model

[Refer to the image for the detailed structure and data presented in the chemical probing data and the 2D model.]
Ensemble approaches in RNA folding

- *in silico*

![RNA structure diagram]

...CAGUAGCCGAUCGCAGCUAGCGUA...
Ensemble approaches in RNA folding

- *in silico*
- From single structure, minimal free-energy folding...
- ...to ensemble approaches.

...CAGUAGCCGAUUCGCAGCUAGCGUA...

Ensemble diversity? Structure likelihood? Evolutionary robustness?

UnaFold, RNAFold, Sfold...
Sensitivity to mutations

Halvorsen et al., PLOS Gen 2010

Boltzmann Sampling $\rightarrow$ PCA $\rightarrow$ Clustering
Sensitivity to mutations

Halvorsen et al., PLOS Gen 2010

Boltzmann Sampling → PCA → Clustering
Assessing the reliability of a prediction

Native Structure (RFAM consensus)

Predicted MFE

![Diagram of RNA structures](image)
Assessing the reliability of a prediction

A. *Capsulatum*

**B**

Predicted MFE

**C**

- mfe
- pf
- centroid

Boltzmann probability

Position

Height

Entropy
Assessing the reliability of a prediction

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.

D1 - D4 group II intron A. Capsulatum sequence

RNAFold [Gruber AR et al. NAR 2008]
Comparing structures visually

Fragment of Thermophilus tRNA-Phe vs yeast's (PDB: 4TNA & 3BBV)

DARTS [Dror O et al., NAR 06] + Pymol

Romantic search

! T thermophilus

! et al!
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

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 base pairs

trans base pairs
Leontis/Westhof nomenclature: A visual grammar for tertiary motifs

S2S software [Jossinet/Westhof, RNA 2005]
Layout algorithms are challenged by tertiary interactions

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

New layout algorithms are needed!
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…

………..

CS rulez!
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…

You guys are going to love my new algorithm!
Once upon a time…

My model cognitively makes so much more sense than previous representations.
Once upon a time…

\[ x + a^n = n^k \]

\[ x^n k^a - k^n = 0 \]

Theorem 35. The easy part

And the rest follows trivially
Once upon a time…

Questions?

Thanks for listening.
Once upon a time…

Thanks for listening.

Questions?

How would you draw our favorite tRNA? The one we've studied during our PhDs and our first three postdocs, named all of our first child after…

Zzzz…

Zzzz…
Once upon a time…
Once upon a time...
Once upon a time…

And don't come back!

Ok guys, whose turn to make the coffee?
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
- Templat 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:
- 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 databases...

VARNA software
[Darty et al, Bioinformatics 2009]
http://varna.lri.fr
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)
- 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…