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)

![Graph showing the increase in RFAM Families from 2002 to 2011](image-url)
Why RNA is so COOL!

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

Why RNA is so COOL!

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

[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…)

PDB entries (Feb 2012)

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

UUAGGCCGCCACACGC
GGUGGGGUUGCCUCC
CGUACCCAUUCCGAA
CACCGAAGAUAGCC
CACAGCGUUCGGG
GAGUACUGGAGUGCG
CGAGCUCUGGAAA
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

- 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
  - Coot
    - [Emsley P et al, Act Crys D 2010]
  - 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

(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...
Ensemble approaches in RNA folding

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

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)

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

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.

A. Capsulatum sequence

D1-D4 group II intron

Native Structure
(RFAM consensus)

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

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

*Romantic search*
Lehmann/Jossinet
(Submitted)
Towards novel representations
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

[Leontis/Westhof, NAR 2002]
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 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\text{ary} structures?
Once upon a time…
Once upon a time…

How would you like to see RNA?
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…

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…

Theorem 35. The easy part

\[(x + a)^n = \sum_{k=0}^{n} \binom{n}{k} x^k a^{n-k}\]

And the rest follows trivially
Once upon a time…

Thanks for listening.

Questions?

Zzzz…

Zzzz…
Once upon a time...

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…

Questions:

Zzzz…

Zzzz…
Once upon a time…

Tadah!
Once upon a time…

Uh… Tadah?
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
- 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:

- 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...

VARNA software [Darty K 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)

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)