RNA World Hypothesis and RNA folding

By Lixin Dai
Outline:

- RNA World Hypothesis
- RNA structure
  - primary
  - secondary
  - tertiary
- Folding of RNA structure
- Problems with the folding
- Solutions

October 16, 2002
RNA World Hypothesis:

Central dogma:

- Protein or RNA? – The chicken & egg paradox
- How did life begin?
- What was the most primitive self-replicating system?

Protein needs RNA as a template; RNA needs protein to be synthesized

Diagram:

DNA → transcription → RNA → translation → Protein

 reversetranscription
A possible solution: catalytic RNA

If RNA can be catalytic, then perhaps the simplest life form was an RNA molecule that could self-replicate.

According to the RNA World Hypothesis:
- RNA should be capable of many reactions
- Ribozymes are molecular fossils of the RNA world
- RNA components of cellular enzymes which are not catalytic are possibly descendants of the RNA world and still have remnants of catalytic activity. e.g. the catalytic core of large subunit of ribosome RNA is protein free!
RNA structure:

The presence of uracil in place of thymine, and the 2'-OH in the ribose constitutes the two chemical differences between RNA and DNA.

Base pairing in
DNA: A-T; G-C
RNA: A-U; G-C

Unusual pairing in RNA:
G-U wobble pair
A-G or A-C can also pair at special condition

Most DNA double helices are Type B and RNA helices are Type A.
RNA structure:

- **Primary structure** (e.g. tRNA\textsubscript{phe}):
  
  \[
  \begin{array}{cccccc}
  10 & 20 & 30 & 40 & 50 \\
  GCGAUUUAG & CUCAGUUUGG & AGAGCGCCAG & ACUGAAUAUC & UGGAGGUCCU \\
  
  60 & 70 & 80 \\
  GUGUUGCAGUCC & CACAGAAUUC & GCACC \\
  \end{array}
  \]

- **Secondary structure**:
  - a.a. stem
  - D arm
  - T arm
  - a.c. arm

- **Tertiary structure**
RNA structure:

- **Primary structure (e.g. tRNA\textsuperscript{phe}):**

  10  20  30  40  50  60  70  80
  GCGAUUUAG  CUCAGUUGGG  AGAGGCGCCAG  ACUGAAUAUC  UGGAGGUCCU
  GUGUUCGAUC  CACAGAAUUC  GCACC

- **Secondary structure:**
  - a.a. stem
  - D arm
  - T arm
  - a.c. arm

- **Tertiary structure**
RNA structure:

- **Primary structure** (e.g. tRNA\textsuperscript{phe}):
  
<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCGGAUUUAG</td>
<td>CUCAGUUGGG</td>
<td>AGAGCGCCAG</td>
<td>ACUGAAUAUC</td>
<td>UGGAGGUCCU</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>60</th>
<th>70</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUGUUUCGAUC</td>
<td>CACAGAAUUC</td>
<td>GCACC</td>
</tr>
</tbody>
</table>

- **Secondary structure:**
  - a.a. stem
  - D arm
  - T arm
  - a.c. arm

- **Tertiary structure**
RNA structure:

- **Primary structure** (e.g. tRNA\textsubscript{phe}):

  
<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCGGAUUUAG</td>
<td>CUCAGUUGGG</td>
<td>AGAGCGCCAG</td>
<td>ACUGAAUAUC</td>
<td>UGGAGGUCCU</td>
</tr>
</tbody>
</table>
  
  | 60 | 70 | 80 |
  | GUGUUCGAUC | CACAGAAUUC | GCACC |

- **Secondary structure**:
  - a.a. stem
  - D arm
  - T arm
  - a.c. arm

- **Tertiary structure**
Folding of RNA secondary structure:

- Free energy minimization:
  - Favorable contributions: hydrogen bonds of base pairs, favorable “stacking” interaction of bases, some base-pairs created in irregular structures
  - Unfavorable contributions: symmetric bulges in helices, asymmetric bulges in helices, increasing loop size at the end of helix, multiple branches from single loop

- Simple energies: \( \Delta G \)
  - G-C = -3
  - A-U = -2
  - G-U = -1

- MFOLD program from Dr. Michael Zuker
Folding of RNA tertiary structure:

- 2-D structure can fold to tight 3-D structure by tertiary interactions:
  - WC base-pairs (include G-U pair)
  - base stacking
  - Loop-Loop interactions
  - Loop & receptors interactions
  - Metal ion binding
- X-ray crystallography & NMR
- Computer-based Prediction by energy minimization
- MC-SYM program from Dr. François Major
Problems in folding RNA

- Secondary structure:
  - Multiple solutions for the same primary structure.
  - The structure with lowest free energy is not always the correct structure. Hard to predict \textit{in vivo} condition.

\[ \Delta G = -22.9 \]  
\[ \Delta G = -22.9 \]  
\[ \Delta G = -21.9 \]
Problems in folding RNA

- Tertiary structure:
  - Many RNA molecules are very difficult to crystallize.
  - Too many possibilities for a single secondary structure.
  - Unpredictable tertiary interactions.
Solutions:

- Bioinformatic analysis
  - Multi-sequences alignment
    - Seq1: AGGCUGAAAGGCAG
    - Seq2: CUACUGAAAGGCCC
  - Phylogenetic comparison (conserved motifs)
  - Co-variation analysis
    - Seq1: AGGCUGAAAGGCAG
    - Seq2: CUACUGAAAGGUGC

- Experimental data

- Computer programming
  - Interactive Computer Graphics (ICG)
  - Distance Geometry