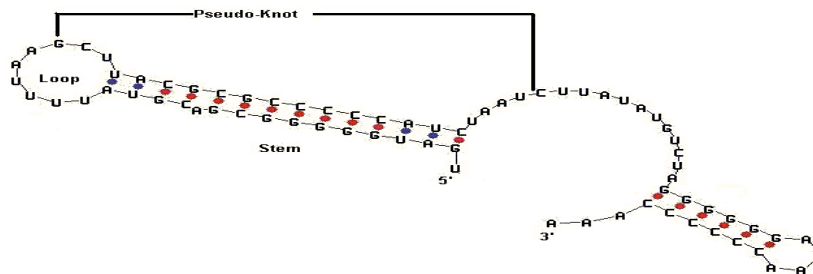


RNA secondary structures

- ✓ RNA is made of 4 nucleotides
 - A, G, C, U
- ✓ RNA molecules are single-stranded but they can interact with themselves to form secondary structures
- ✓ The most typical RNA secondary structure is the hairpin
- ✓ Hairpins are usually made of a stem and a loop as shown here



Predicting RNA Secondary Structures

- ✓ It is possible to predict the secondary structure of an RNA sequence
- ✓ The accuracy is ~70%
- ✓ Predictions are easier and more accurate on small molecules (fewer than 500 nucleotides)
- ✓ More accurate predictions can be made using multiple-sequence alignments

Single-Sequence Structure Prediction for RNA

- ✓ Make single-sequence predictions using the Mfold server
- ✓ The server predicts the most stable secondary structure
 - The most stable structure is the one with the lowest potential energy
- ✓ Mfold can also return suboptimal structures
 - Similar optimal and suboptimal structure indicate good accuracy

Forcing Mfold

- ✓ If you have a known structure, you can force it into Mfold as a constraint
 - Secondary structures can be found experimentally
- ✓ Mfold cannot accept pseudo-knots as constraints

Enter **constraint information** in the box at the right. (optional) You may:

1. force bases $i, i+1, \dots, i+k-1$ to be double stranded by entering:
 $F \ i \ 0 \ k$ on 1 line in the constraint box.
2. force consecutive base pairs $i, j, i+1, j-1, \dots, i+k-1, j-k+1$ by entering:
 $F \ i \ j \ k$ on 1 line in the constraint box.
3. force bases $i, i+1, \dots, i+k-1$ to be single stranded by entering:
 $P \ i \ 0 \ k$ on 1 line in the constraint box.
4. prohibit the consecutive base pairs $i, j, i+1, j-1, \dots, i+k-1, j-k+1$ by entering:
 $P \ i \ j \ k$ on 1 line in the constraint box.
5. prohibit bases i to j from pairing with bases k to l by entering:
 $P \ i-j \ k-l$ on 1 line in the constraint box.

The RNA sequence is .
 Folding temperature is fixed at 37°.
 Ionic conditions: 1M NaCl, no divalent ions.
 Enter the **percent suboptimality** number.

F 10 22 3

The Best RNA Secondary Structure Predictions Come from MSAs

- ✓ The best predictions are based on multiple-sequence alignments
- ✓ They take advantage of covariation
- ✓ No online resource is yet available but you can use
 - selab.janelia.org/software.html

```

AGGACACAUAAGAGAUAUAGACACCACAGAUCCACCACACA
AGGACACAUTAAGAUAUAGACACCACAGAGUACCACACAC
ACACAUAAGA GAUUAUAGACACCACAGAGCAUCACACAGGA
ACACAUGUTUAUUUCAUAGACACCACAGAGUACCACACAC
    
```

Searching Databases for tRNAs

- ✓ Transfer RNAs are small RNA molecules used to build proteins
 - Each protein has 64 tRNAs
 - One for each triplet of the genetic code
 - All tRNAs have the same structure
- ✓ They are easy to find in a genome
 - tRNAscan from Sean Eddy (shown)
 - Visit selab.janelia.org for examples



Finding RNA Genes

- ✓ Many genes do not code for proteins
- ✓ They simply code for RNA
 - tRNA, ribosomal RNA, and miRNA (micro-RNAs)
- ✓ tRNA and ribosomal RNA genes are easy to find
- ✓ The other non-coding genes (such as miRNAs) are very difficult to find
- ✓ We do not know how many RNA genes are in the human genome
 - 1500-1800 predicted/verified in the ENCODE pilot project regions of the human genome (~ 1%).
[Washietl et al. Genome Res. 2007. 17: 852-864]

Micro-RNAs (miRNAs)

- ✓ Micro-RNAs are small, non-coding RNAs involved at almost every stage of cell regulation
- ✓ Micro-RNAs can silence genes through RNA interference
- ✓ Nobel Prize 2006 : Fire and Melo for discovering **RNA interference**
- ✓ miRNAs are the future of medicine; find them at these sites:
 - sirna.cgb.ki.se
 - micorna.sanger.ac.uk

1500-1800 predicted/verified in the ENCODE pilot project regions of the human genome (~ 1%).
[Washietl et al. Genome Res. 2007. 17: 852-864]

Going Farther

- ✓ RNA is the “poor cousin” of bioinformatics
- ✓ The NCBI has no RNA section
- ✓ The RNA world is receiving ever-increasing attention
- ✓ Many excellent public resources are available for RNA

- Find a list of links at the "RNA world" website:
<http://www.fli-leibniz.de/RNA.html>

For example: RNA structures (atomic coordinates) at NDB=Nucleic Acid Database: <http://ndbserver.rutgers.edu>