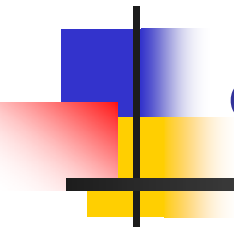


Stacking Energies and RNA Structure Prediction

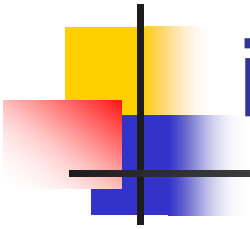
Bioinformatics Senior Project

Adrian Lawsin

December 2008

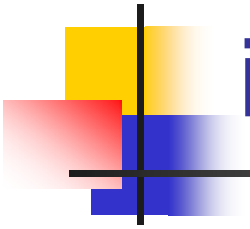
Table of Contents

Importance of Stacking Energies in RNA Structure Prediction



- In nature, compounds try to achieve maximum stability.
- Stability is achieved by minimizing the molecule's free energy.
- Molecules convert (store) free energy when it creates bonds.

Importance of Stacking Energies in RNA Structure Prediction

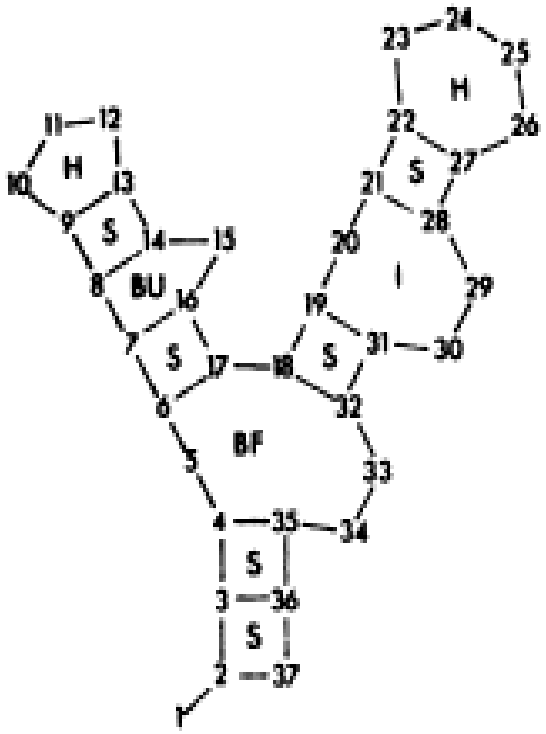


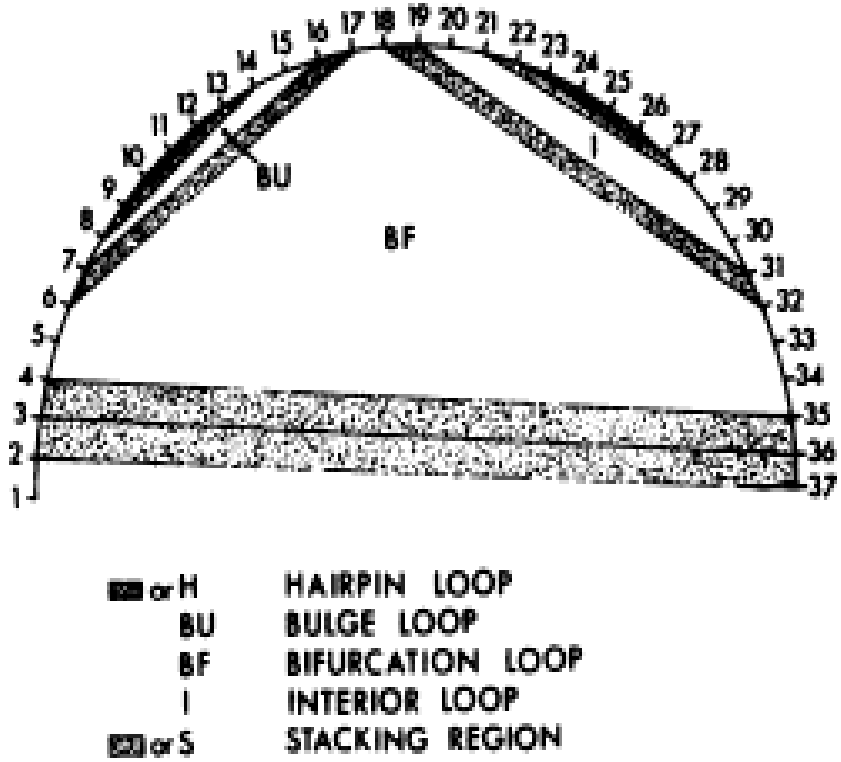
- Current algorithms in RNA structure (bond) prediction are based on free energy minimization.
- It is assumed that stacking base pairs and loop entropies contribute additively to the free energy of an RNA sequence's secondary structure.

Major Types of Stacking Energies – RNA Structures

- RNA secondary structure can be viewed as a conglomeration of several smaller structures.
- These are:
  - Stacking (Base Pairs) Regions
  - Hairpin Loops
  - Interior Loops
  - Bulge Loops
  - Bifurcation (Multi-Stem) Loops
  - Single (Free) Bases

Major Types of Stacking Energies – RNA Structures





Major Types of Stacking Energies – RNA Structures

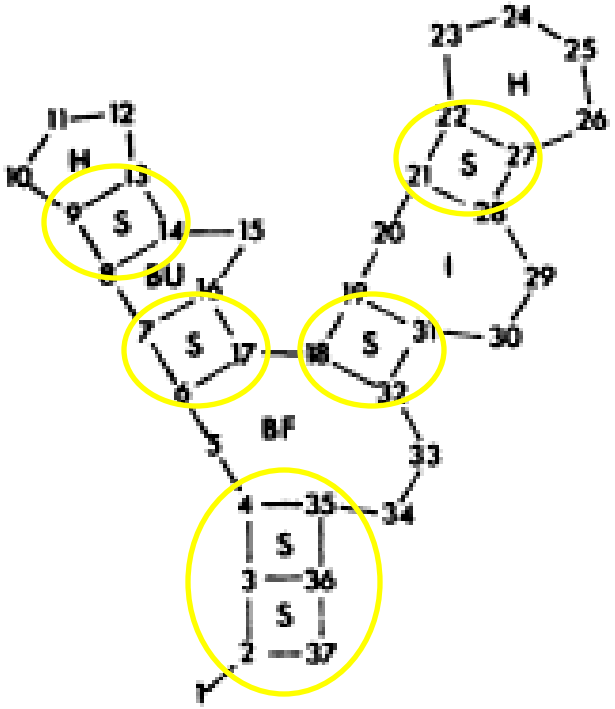
- Each of these structures has a corresponding energy that contributes to the overall stability of the molecule.
- Due to space constraints, we will only offer parts of most lists. The complete list of all the energies is available at:

  http://www.bioinfo.rpi.edu/zukerm/rna/energy/

Major Types of Stacking Energies – RNA Structures

- Most of the research estimating the energies has been done by Prof. D.H. Turner at the University of Rochester.
- He based the energy values through melting studies of synthetically constructed oligoribonucleotides.
- The listed values are at 37° - the human body's internal core temperature.

Stacking (Base Pairs) Regions



Stacking (Base Pairs) Regions

Stacking (Base Pairs) Regions

- Total free energy of the entire Stacking Region is given by the addition of each pair of adjacent base pairs.
- This includes energy contributions for both base pair stacking and hydrogen bonding.
- This breaks down for 2 or more consecutive G-U pairs and pairs that are not Watson-Crick (WC) base pairs.

Stacking (Base Pairs) Regions

- The *Stacking Energies* table uses the arrangement for a stack:

  5′ – WX – 3′

  3′ – ZY – 5′

  The corresponding energy would appear in the W^{th} row and the Z^{th} column of 4 by 4 tables, and in the X^{th} row and the Y^{th} column of that table.

Stacking (Base Pairs) Regions

Stacking energies at 37°:

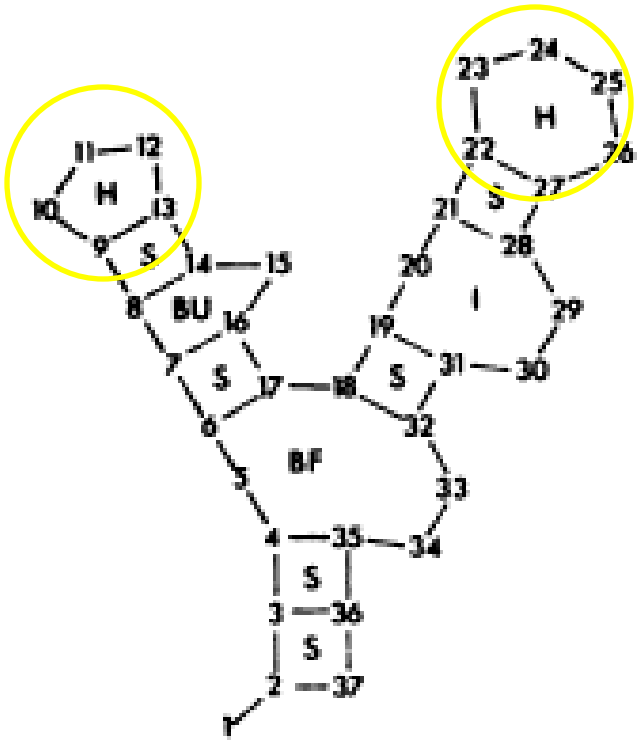
Excerpt from Table of Stacking Energies

Data Arangement:

```
                Y
          ------------------
     (X)  A    C    G    U
          ------------------
              5' ==> 3'
                 AX
                 AY
              3' <== 5'
     (A)    .    .    .    .
     (C)    .    .    .    .
     (G)    .    .    .    .
     (U) -0.70 -0.10 -0.70 -0.10
```

```
          Y                   Y                   Y                   Y
  ------------------  ------------------  ------------------  ------------------
   A    C    G    U    A    C    G    U    A    C    G    U    A    C    G    U
  ------------------  ------------------  ------------------  ------------------
      5' --> 3'           5' --> 3'           5' --> 3'           5' --> 3'
         AX                  AX                  AX                  AX
         AY                  CY                  GY                  UY
      3' <-- 5'           3' <-- 5'           3' <-- 5'           3' <-- 5'
   .    .    .    .    .    .    .    .    .    .    .    .    .    .    .  -0.90
   .    .    .    .    .    .    .    .    .    .    .    .    .    .  -2.10   .
   .    .    .    .    .    .    .    .    .    .    .    .    .  -1.70   .  -0.50
   .    .    .    .    .    .    .    .    .    .    .    .  -0.90   .  -1.00   .
```

Hairpin Loops



Hairpin Loops

Hairpin Loops

- A Hairpin Loop is a structure that looks like a hairpin; after a Stack there is an opening at the end. The hairpin loop starts at the end of the stacking region where the base pairing stops.

Hairpin Loops

- Hairpin Loop Energies are the sum of up to 3 terms:
  - 1. Loop size (number of single stranded bases) – given in the *hairpin* column of the *LOOP Destabilizing Energy* table. For loops larger than 30, 1.75RTln(size/30) is added (R= universal gas constant, T = absolute temperature).

Hairpin Loops

Loop destabilizing energies at 37°:

DESTABILIZING ENERGIES BY SIZE OF LOOP (INTERPOLATE WHERE NEEDED)
hp3 ave calc no tmm;hp4 ave calc with tmm; ave all bulges

SIZE	INTERNAL	BULGE	HAIRPIN
1	.	3.90	.
2	4.10	3.10	.
3	5.10	3.50	4.10
4	4.90	4.20	4.90
5	5.30	4.80	4.40
6	5.70	5.00	4.70
7	5.90	5.20	5.00
8	6.00	5.30	5.10
9	6.10	5.40	5.20
10	6.30	5.50	5.30
11	6.40	5.70	5.40
12	6.40	5.70	5.50
13	6.50	5.80	5.60
14	6.60	5.90	5.70
15	6.70	6.00	5.80
16	6.80	6.10	5.80
17	6.80	6.10	5.90
18	6.90	6.20	5.90
19	6.90	6.20	6.00
20	7.00	6.30	6.10
21	7.10	6.30	6.10
22	7.10	6.40	6.20
23	7.10	6.40	6.20
24	7.20	6.50	6.30
25	7.20	6.50	6.30
26	7.30	6.50	6.30
27	7.30	6.60	6.40
28	7.40	6.70	6.40
29	7.40	6.70	6.50
30	7.40	6.70	6.50

Hairpin Loops

- 2. A favorable (negative) stacking interaction occurs between the closing base pair of the hairpin loop and the adjacent mismatched pair, given in the *Hairpin Loop Terminal Stacking Energy* table.
  - This energy is not added in triloops (loops of size 3).

Hairpin Loops

Terminal mismatch stacking energies for hairpin loops at 37°:

Data Arangement:

```
                Y
        ------------------
    (X)  A    C    G    U
        ------------------
             5' ==> 3'
                AX
                AY
             3' <== 5'
    (A)    .    .    .    .
    (C)    .    .    .    .
    (G)    .    .    .    .
    (U) -0.70 -0.10 -0.70 -0.10
```

```
          Y                   Y                   Y                   Y
  ------------------  ------------------  ------------------  ------------------
   A    C    G    U    A    C    G    U    A    C    G    U    A    C    G    U
  ------------------  ------------------  ------------------  ------------------
      5' --> 3'           5' --> 3'           5' --> 3'           5' --> 3'
         AX                  AX                  AX                  AX
         AY                  CY                  GY                  UY
      3' <-- 5'           3' <-- 5'           3' <-- 5'           3' <-- 5'
    .    .    .  -1.00  .    .    .  -0.70  .    .    .  -1.80 -0.80 -1.00 -1.70 -1.00
    .    .  -1.10  .    .    .  -1.10  .    .    .  -2.30  .   -0.70 -0.70 -0.70 -0.70
    .  -1.40  .  -1.20  .  -1.00  .  -0.90  .  -2.10  .  -2.00 -1.50 -1.00 -1.00 -1.00
 -0.80  .  -0.80  .  -0.70  .  -0.70  .  -1.50  .  -1.50  .   -0.80 -0.80 -0.80 -0.80
```

Excerpt from Table of

Terminal Mismatch Stacking Energies

For Hairpin Loops

Hairpin Loops

- 3. Certain tetraloops have special bonus energies, as given in the *Tetra-loop Bonus Energies* table.

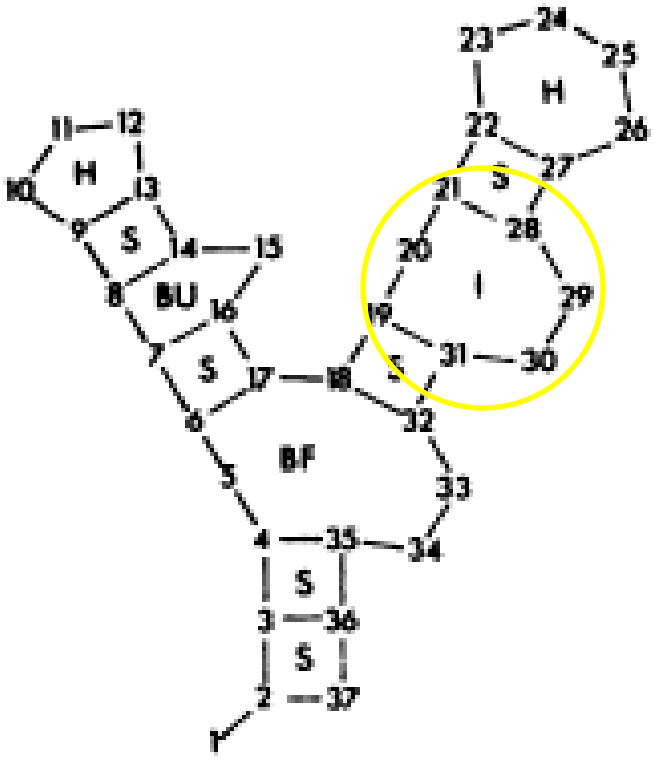
Hairpin Loops

Tetra-loop bonus energies at 37°:

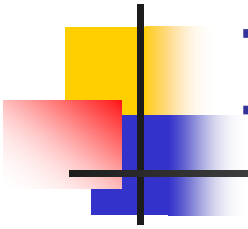
Seq	Energy
AGAAAU	-2.00
AGCAAU	-2.00
AGAGAU	-2.00
AGUGAU	-2.00
AGGAAU	-2.00
AUUCGU	-2.00
AUACGU	-2.00
AGCGAU	-2.00
AUCCGU	-2.00
AGUAAU	-2.00
ACUUGU	-2.00
AAUUUU	-2.00
AUUUAU	-2.00
CGAAAG	-2.00
CGCAAG	-2.00
CGAGAG	-2.00
CGUGAG	-2.00
CGGAAG	-2.00
CUUCGG	-2.00
CUACGG	-2.00
CGCGAG	-2.00
CUCCGG	-2.00
CGUAAG	-2.00
CCUUGG	-2.00
CAUUUG	-2.00
CUUUAG	-2.00
GGAAAC	-2.00
GGCAAC	-2.00
GGAGAC	-2.00
GGUGAC	-2.00
GGGAAC	-2.00
GUUCGC	-2.00

Excerpt from Table of Tetra-Loop Bonus Energies

Interior Loops



Interior Loops



Interior Loops

- Interior Loops occur in the middle of Stacking Regions, breaking it up.
- They are closed by 2 base pairs.
- Similar to Hairpin Loops, Interior Loops Energies are composed of the sum of up to 3 terms.

Interior Loops

- 1. Loop size – given in the *interior* column of the *LOOP Destabilizing Energy* table. For loops larger than 30, 1.75RTln(size/30) is added (R= universal gas constant, T = absolute temperature).

Interior Loops

Loop destabilizing energies at 37°:

DESTABILIZING ENERGIES BY SIZE OF LOOP (INTERPOLATE WHERE NEEDED)
hp3 ave calc no tmm;hp4 ave calc with tmm; ave all bulges

SIZE	INTERNAL	BULGE	HAIRPIN
1	.	3.90	.
2	4.10	3.10	.
3	5.10	3.50	4.10
4	4.90	4.20	4.90
5	5.30	4.80	4.40
6	5.70	5.00	4.70
7	5.90	5.20	5.00
8	6.00	5.30	5.10
9	6.10	5.40	5.20
10	6.30	5.50	5.30
11	6.40	5.70	5.40
12	6.40	5.70	5.50
13	6.50	5.80	5.60
14	6.60	5.90	5.70
15	6.70	6.00	5.80
16	6.80	6.10	5.80
17	6.80	6.10	5.90
18	6.90	6.20	5.90
19	6.90	6.20	6.00
20	7.00	6.30	6.10
21	7.10	6.30	6.10
22	7.10	6.40	6.20
23	7.10	6.40	6.20
24	7.20	6.50	6.30
25	7.20	6.50	6.30
26	7.30	6.50	6.30
27	7.30	6.60	6.40
28	7.40	6.70	6.40
29	7.40	6.70	6.50
30	7.40	6.70	6.50

Interior Loops

- 2. Special terminal stacking energies for the mismatched base pairs adjacent to **both** closing base pairs. Each of these energies is taken from the *Interior Loop Terminal Stacking Energy* table.

Interior Loops

Terminal mismatch stacking energies for interior loops at 37°:

Data Arangement:

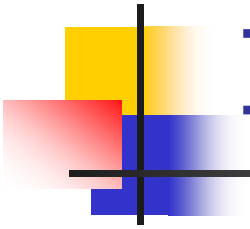
```
                Y
          ------------------
     (X)  A    C    G    U
          ------------------
              5' ==> 3'
                 AX
                 AY
              3' <== 5'
     (A)    .    .    .    .
     (C)    .    .    .    .
     (G)    .    .    .    .
     (U) -0.70 -0.10 -0.70 -0.10
```

```
          Y                   Y                   Y                   Y
  ------------------  ------------------  ------------------  ------------------
   A    C    G    U    A    C    G    U    A    C    G    U    A    C    G    U
  ------------------  ------------------  ------------------  ------------------
      5' --> 3'           5' --> 3'           5' --> 3'           5' --> 3'
         AX                  AX                  AX                  AX
         AY                  CY                  GY                  UY
      3' <-- 5'           3' <-- 5'           3' <-- 5'           3' <-- 5'
   .    .    .   -1.00   .    .    .   -1.00   .    .    .   -2.20 -1.00 -1.00 -2.20 -0.50
   .    .   -1.50   .    .    .   -1.50   .    .    .   -2.70   .   -1.00 -1.00 -0.20 -1.00
   .   -1.50   .   -1.00   .   -1.50   .   -1.00   .   -2.70   .   -2.20 -2.20 -0.50 -1.00 -0.50
  -1.00   .   -1.50   .   -1.00   .   -1.50   .   -2.20   .   -2.70   .   -0.30 -1.00 -0.30 -2.00
```

Excerpt from Table of

Terminal Mismatch Stacking Energies

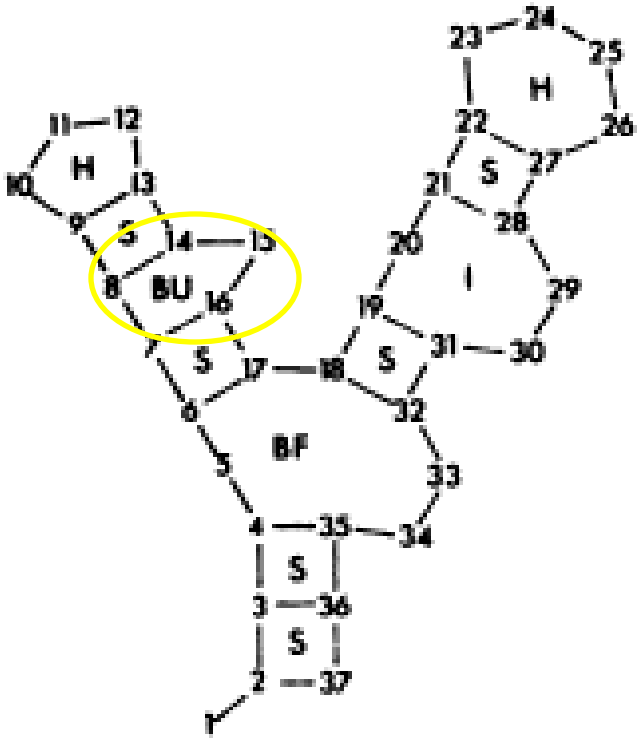
For Interior Loops



Interior Loops

- 3. For non-symmetric interior loops, there is a penalty (positive term). Although the data is incomplete, the maximum penalty is +3.00.

Bulge Loops



Bulge Loops

Bulge Loops

- A Bulge Loop is a special case of an internal loop that has only one of the sides unpaired.
- Bulge Loop's destabilizing energies are given in the *bulge* column of the *LOOP Destabilizing Energy* table. Again, for loops larger than 30, 1.75RTln(size/30) is added (R= universal gas constant, T = absolute temperature).

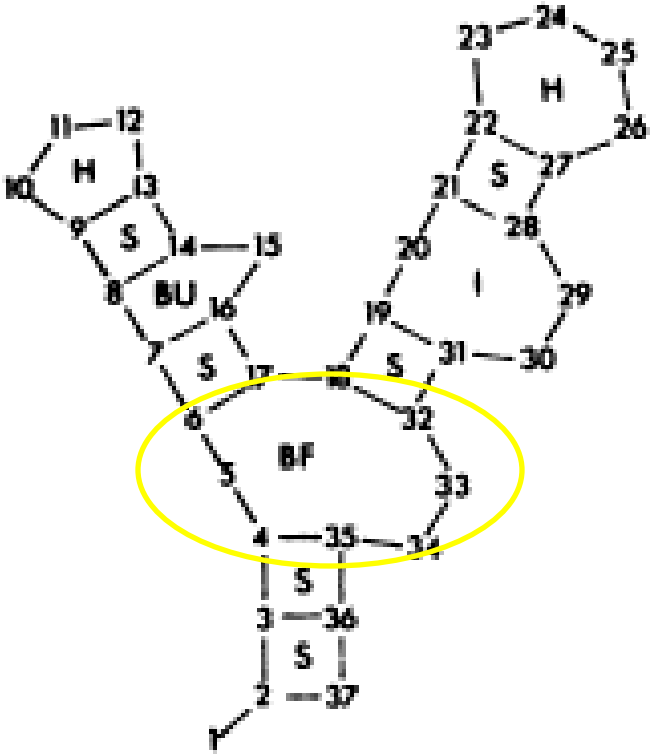
Bulge Loops

Loop destabilizing energies at 37°:

DESTABILIZING ENERGIES BY SIZE OF LOOP (INTERPOLATE WHERE NEEDED)
hp3 ave calc no tmm;hp4 ave calc with tmm; ave all bulges

SIZE	INTERNAL	BULGE	HAIRPIN
1	.	3.90	.
2	4.10	3.10	.
3	5.10	3.50	4.10
4	4.90	4.20	4.90
5	5.30	4.80	4.40
6	5.70	5.00	4.70
7	5.90	5.20	5.00
8	6.00	5.30	5.10
9	6.10	5.40	5.20
10	6.30	5.50	5.30
11	6.40	5.70	5.40
12	6.40	5.70	5.50
13	6.50	5.80	5.60
14	6.60	5.90	5.70
15	6.70	6.00	5.80
16	6.80	6.10	5.80
17	6.80	6.10	5.90
18	6.90	6.20	5.90
19	6.90	6.20	6.00
20	7.00	6.30	6.10
21	7.10	6.30	6.10
22	7.10	6.40	6.20
23	7.10	6.40	6.20
24	7.20	6.50	6.30
25	7.20	6.50	6.30
26	7.30	6.50	6.30
27	7.30	6.60	6.40
28	7.40	6.70	6.40
29	7.40	6.70	6.50
30	7.40	6.70	6.50

Bifurcation (Multi-Stem) Loops



Bifurcation (Multi-Stem) Loops

Bifurcation (Multi-Stem) Loops

- Bifurcation, or Multi-Stem, Loops are loops that form at least two separate branches.
- There is not a lot of experimental information available, but for now:
  - The free energy function is:

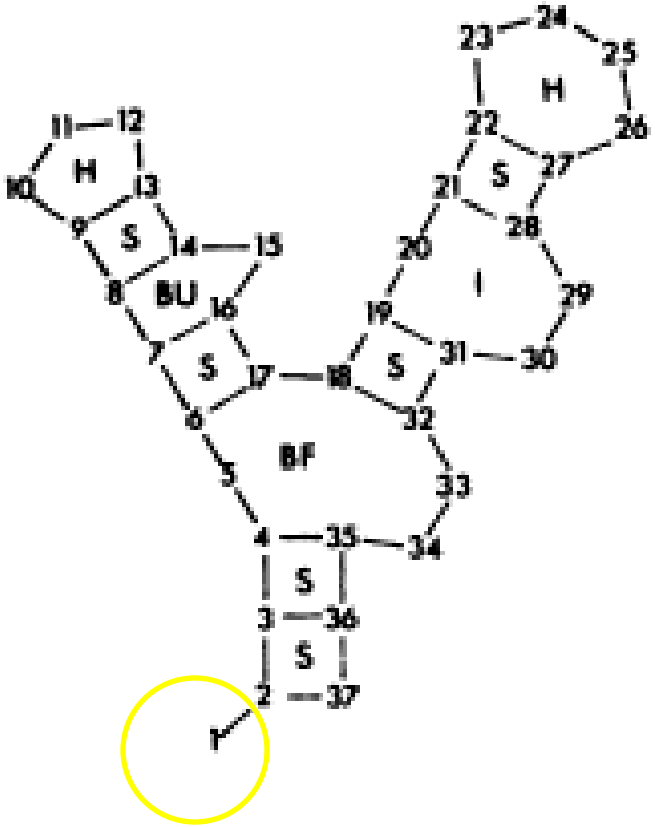
    $E = a + n_1 \times b + n_2 \times c$

    Where a, b, and c are constants, n_1 is the # of single stranded bases in the loop and n_2 is the # of stacks that form the loop.

Bifurcation (Multi-Stem) Loops

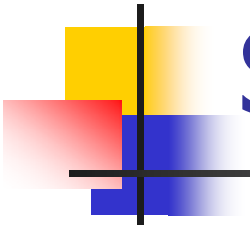
- a, b, and c are called the offset (value of 4.60), free base penalty (value of .40) and helix penalty (value of .10), respectively.

Single (Free) Bases



Single (Free) Bases

Single (Free) Bases



- Single, or Free, bases are single stranded nucleotides that are not in any loop.
- Again, like Bifurcation Loops, not much experimental information is available.
- When a single stranded base is adjacent to the closing base pair of a stack, a *Single Base Stacking Energy* is added.

Single (Free) Bases

- When a single-stranded base is adjacent to 2 stacks, only the most favorable single-base stacking term is added.

Single (Free) Bases

Excerpt from Table of

Single Base Stacking Energies

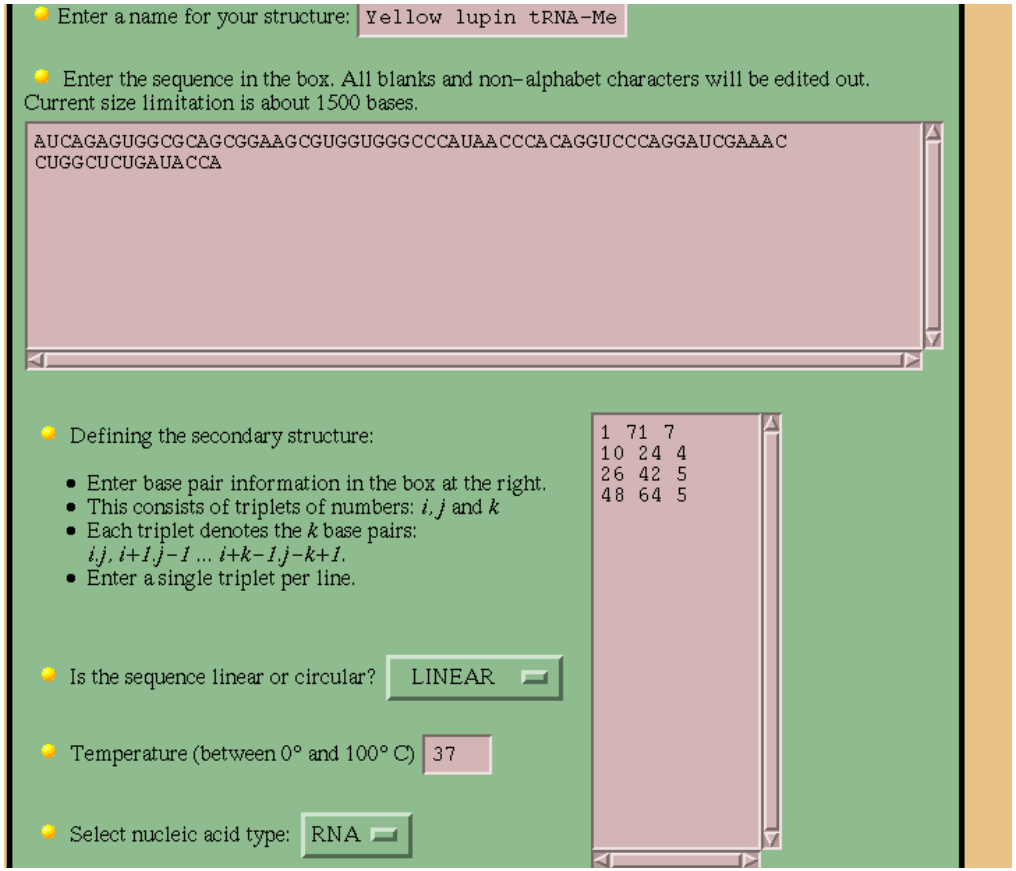
Single base stacking energies at 37°:

X	X	X	X
A C G U	A C G U	A C G U	A C G U
5' --> 3'	5' --> 3'	5' --> 3'	5' --> 3'
AX	AX	AX	AX
A	C	G	U
3' <-- 5'	3' <-- 5'	3' <-- 5'	3' <-- 5'
. . . .	. . . .	. . . .	-0.80 -0.50 -0.80 -0.60
X	X	X	X
A C G U	A C G U	A C G U	A C G U
5' --> 3'	5' --> 3'	5' --> 3'	5' --> 3'
CX	CX	CX	CX
A	C	G	U
3' <-- 5'	3' <-- 5'	3' <-- 5'	3' <-- 5'
. . . .	. . . .	-1.70 -0.80 -1.70 -1.20	. . . .
X	X	X	X
A C G U	A C G U	A C G U	A C G U
5' --> 3'	5' --> 3'	5' --> 3'	5' --> 3'
GX	GX	GX	GX
A	C	G	U
3' <-- 5'	3' <-- 5'	3' <-- 5'	3' <-- 5'
. . . .	-1.10 -0.40 -1.30 -0.60	. . . .	-0.80 -0.50 -0.80 -0.60

Efn Server

- http://mfold.bioinfo.rpi.edu/cgi-bin/efn-form1.cgi
- By entering an RNA sequence and its secondary structure, the free energy of the molecule is calculated.

Efn Server



Sample:

Enter Data

Efn Server

Folding energy of Yellow lupin tRNA-Met-i at 37° C.

- Computed for snark -

Free energy = -22.3 kcal/mole.(*Thermodynamic details*)

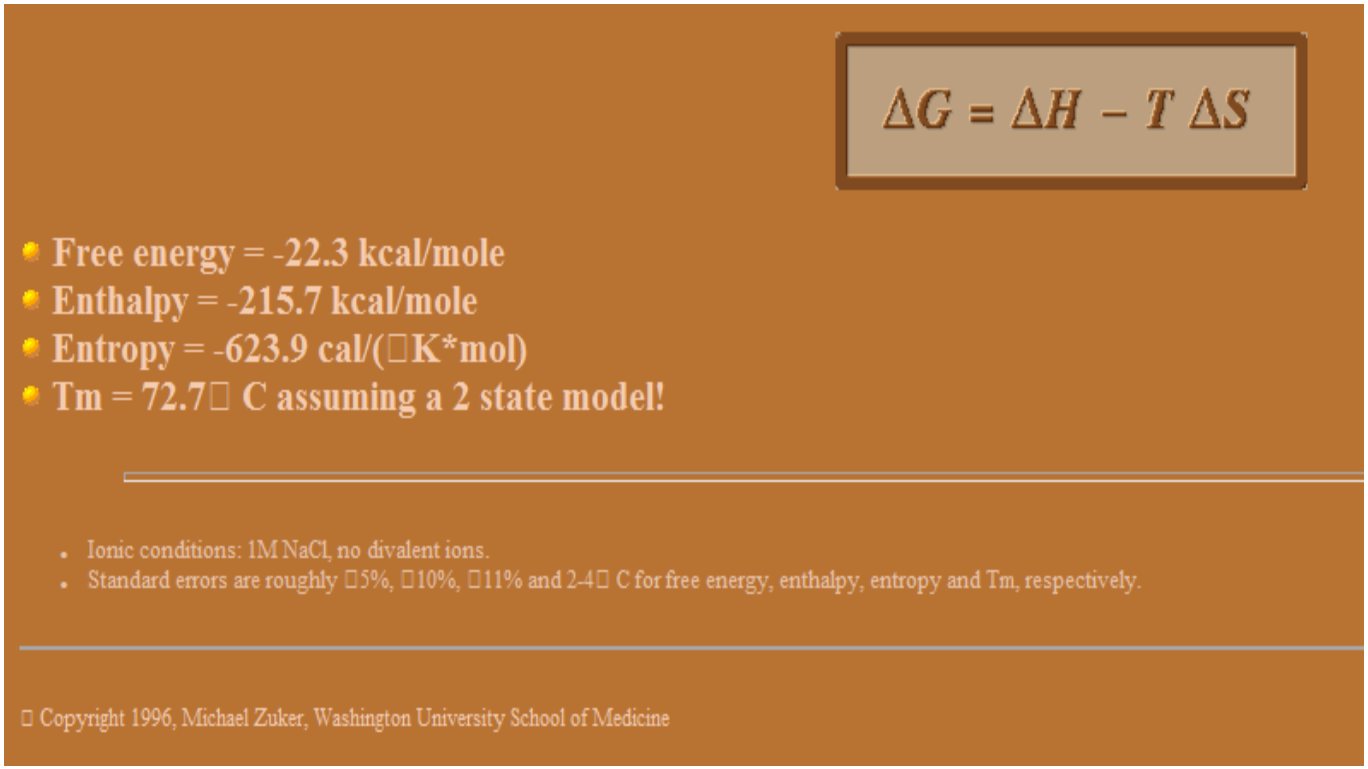
PostScript figure for this structure.

```
FOLDING BASES    1 TO   75 OF  Yellow lupin tRNA-Met-i
 ENERGY  =     -22.3

             10
----       UG      AGC
    AUCAGAG    GCGC
    UAGUCUC    UGCG   G
ACCA       --      AAG
    70           20

                 30
              G      CCC
                 GUGGG
                 CACCC   A
              -       AAU
                  40

                       50
                 AGGUC     AUC
                      CCAGG
                      GGUCC   G
                 -----     AAA
                         60
```

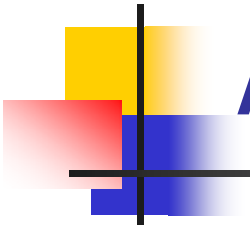
Home page zukerm@rpi.edu

Results

Efn Server



Energy Details

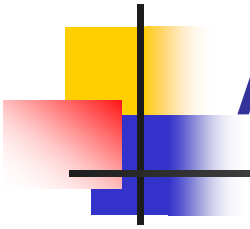


Application

- As previously stated, free energy minimization is at present the most accurate and most generally applicable approach of RNA structure prediction.
- However, current algorithms cannot predict Pseudoknots (overlapping stacking regions).

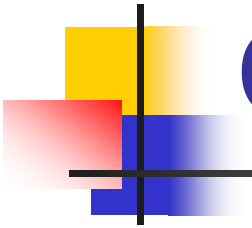
Application

- However, current algorithms that predict the structure of a single RNA molecule (like mfold and the Vienna RNA Package) can predict the structure of an RNA-RNA interaction with a little modification (RNAhybrid and RNAduplex).



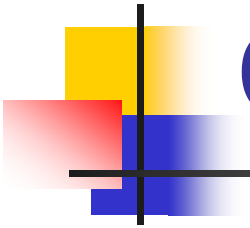
Application

- In the simplest approaches, the RNA molecules are concatenated and treated as one molecule.
- The “new” molecule is then folded normally.



Conclusion

- Since these RNA-RNA algorithms are based on the single RNA algorithm, it has the same weaknesses, mainly the lack of predicting pseudoknots.
- On top of this, there is a conditional probability that the RNA molecules will interact at all.



Conclusion

- Also, there is lack of knowledge concerning the energetics of RNA-RNA interactions within loops.
- Similarly, kissing-interactions (between loops) need to be measured more thoroughly to improve energy parameters.
- Likewise, how protein factors affect RNA-RNA binding energies need to be investigated.

Resources

- Alkan, Can, Emre Karakoc, Joseph H. Nadeau, S. Cenk Sahinalp, and Kaizhong Zhang. "RNA-RNA Interaction Prediction and Antisense RNA Target Search." Journal of Computational Biology 13 (2006): 267-82.
- Delisi, Charles, and Donald M. Crothers. "Prediction of RNA Secondary Structure." Proceedings of the National Academy of Sciences of the United States of America 68 (1971): 2682-685.
- Dima, Ruxandra I., Changbong Hyeon, and D. Thirumalai. "Extracting Stacking Interaction Parameters for RNA from the Data Set of Native Structures." Journal of Molecular Biology 347 (2005): 53-69.
- Matthews, David H., Jeffrey Sabina, Michael Zuker, and Douglass H. Turner. "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure." Journal of Molecular Biology 288 (1999): 911-40.

Resources

- Muckstein, Ulrike, Hakim Tafer, Jorg Hackermuller, Stephan H. Bernhart, Peter F. Stadler, and Ivo L. Hofacker. "Thermodynamics of RNA-RNA binding." Bioinformatics 22 (2006): 1177-182.
- Zuker, Michael, and Patrick Stiegler. "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information." Nucleic Acids Research 9 (1981): 133-48.
- Zuker, Michael. "Efn server: Compute the free energy of an RNA/DNA structure." Rensselaer Polytechnic Institute. 28 Nov. 2008 <http://www.bioinfo.rpi.edu/applications/mfold/cgi-bin/efn-form1.cgi>.
- Zuker, Michael. "Turner Lab: Free energy and Enthalpy Tables for RNA Folding." 3 Nov. 2000. Rensselaer Polytechnic Institute. 28 Nov. 2008 <http://www.bioinfo.rpi.edu/zukerm/rna/energy/>.

Contact Information

- Adrian Lawsin

  Bioinformatics Major, Senior Year

  New Jersey Institute of Technology

  email: al8@njit.edu