Project Report on
RNA Folding with Energy Minimization

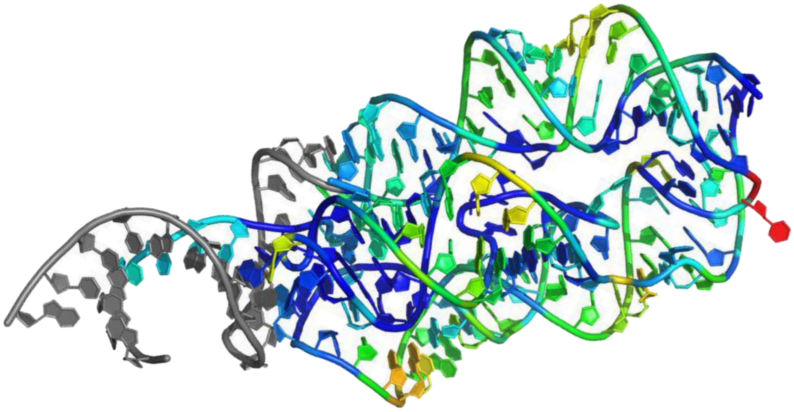


Fig. RNA folding Visualization

Submitted to

Dr. Kaizhong Zhang
Professor, Western University

Submitted by

Harsh Dhillon
hdhill48@uwo.ca

Pankaj Kumar
pkumar@uwo.ca

COMPUTER SCIENCE DEPARTMENT
WESTERN UNIVERSITY

Abstract

RNA molecule which was once consider as the intermediate step between DNA and protein has become one of the important subject of research today. And RNA folding is the most fundamental process which underlying RNA function. RNA folding is simpler. Only four nucleotides, each made of a base, a ribose, and a phosphate, are used as building blocks of the structure but determining the RNA secondary structure from sequence data by computational predictions is a longstanding problem. In our report we are discussing about different algorithm which can be used to determine the RNA secondary structure.

Introduction

RNA, abbreviated for Ribonucleic Acid which was once considered as an intermediate step between DNA and protein, is a complex polymeric molecule essential in various biological roles in coding, decoding, regulation and expression of genes. It has become one of the central subject of research in recent year. It is found to be important in many biological processes. Ribonucleic Acid one of the three major biological macromolecules that are essential for all known forms of life, along with DNA and Proteins. Central dogma of molecular biology states that the flow of genetic information in a cell is from DNA though RNA to proteins. DNA makes RNA with the help of transcription and RNA makes protein with the help of translation, which plays major role in assisting various functions in a cell.

Ribonucleic Acid is typically single-standard biopolymer and is made of ribonucleotides that are linked by phosphodiester bonds. RNA consists of a chain of Ribose Nucleotides which are nitrogenous bases appended to a ribose sugar. The nitrogenous bases in RNA are Adenine (A), Guanine(G), Cytosine(C) and Uracil(U), which replaces Thymine(T) in DNA.

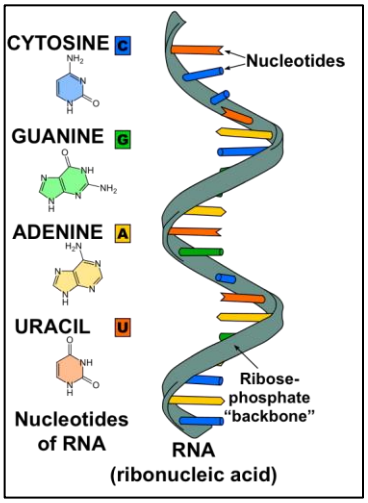


Figure 1.0 – Structure of RNA

RNA are versatile cellular machines that can do what both DNA and protein do and more. They can switch gene on and off. RNA's shape determines its function in a cell. If we could predict exactly what shape any RNA sequence will fold into, we could harness RNA’s potential. For instance, we might be able to design an RNA that turns off cancer causing genes.

RNA folding is a process by which a linear ribonucleic acid molecule acquires a secondary structure through intra-molecular interactions. RNA is single stranded while DNA exists as fully base paired double helix. RNA forms complex and intricate base-pairing interactions due to its increased ability to form hydrogen bonds stemming from the presence of an extra hydroxyl group in the ribose sugar. The folded domains of RNA molecules are often the sites of specific interactions with proteins in forming RNA-Protein complexes such as ribonucleoproteins.

RNA folding is hierarchical[1] and its folding is sequential in the secondary structure. Secondary structure of RNA can be predicted successfully from experimental thermodynamic data on secondary structure. There are only four basic secondary structure elements in RNA (helices, loops, bulges, and junctions). Secondary structure of RNA is much more stable than the tertiary structure. Because the energies involved in the formation of secondary structure are larger than those involved in tertiary interactions. So, secondary structural elements can exist and be stable by themselves. Thus, the energetic contributions of the secondary and tertiary structural elements are separable, and it is possible to treat the energy of tertiary interactions as a perturbation on the energy stabilizing the secondary structure.

RNA secondary structure of Bacillus Subtilis:

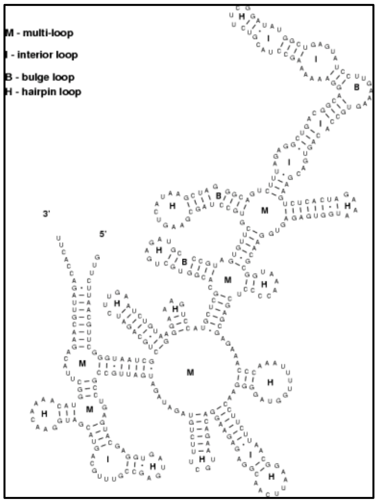


Figure 2.0 – Subtilis secondary structure can be represented in circular manner which is shown in figure 3.0. Lines represent various base pairs present in respective folding. There are no pseudo knots in this representation which is must condition to obtain RNA folding through energy minimization. Any number on boundary of circle represent location of a base in respective sequence

Circular representation of the secondary structure for Figure 2.0

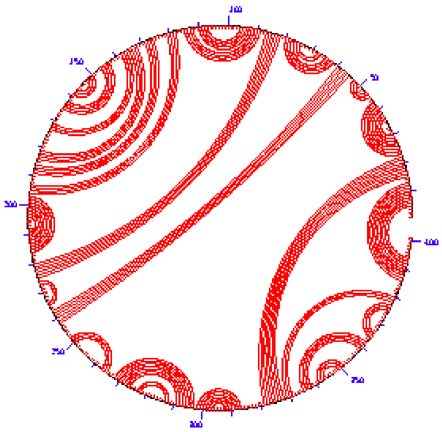


Figure 3.0 – Source: http://www.bioinfo.rpi.edu/%7Ezukerm/Bio-5495/RNAfold-html/node2.html

There are two main methods to predict RNA secondary structure. First method is Comparative sequence analysis[2]. This method infers base-pairs by determining canonical pairs that are common among multiple homologous sequences. To predict the secondary structure of a single sequence, the most popular methods use Free Energy Minimization with computer algorithms based on dynamic programming.

Comparative sequence analysis

Comparative sequence analysis is a powerful method for aiding human gene identification, inferring function of a gene's product, and identifying novel functional elements such as those involved in transcriptional regulation. The information that can be inferred when comparing sequences is dependent on the evolutionary distance between the two organisms. Organisms that are closely related are more likely to share a higher degree of sequence similarity. Distantly related organisms such as yeast and worm share less sequence similarity and are likely to show sequence conservation in coding regions alone. During evolution, secondary structure of functional RNA conserved better than primary. Invariance in certain sections identifies them as being important to structure and function.

Specific pairs are proven by the existence of compensating base-pair changes, where, for example, a GC pair in one sequence is replaced by an AU pair in another sequence. Comparative analysis is quite robust when a number of homologous sequences are available. Over 97% of base-pairs predicted for ribosomal RNA were demonstrated in subsequent crystal structures[3]. Comparative analysis has also been used to infer tertiary structure contacts[4]. Comparative analysis, however, requires multiple sequences, can be time consuming, and requires significant insight.

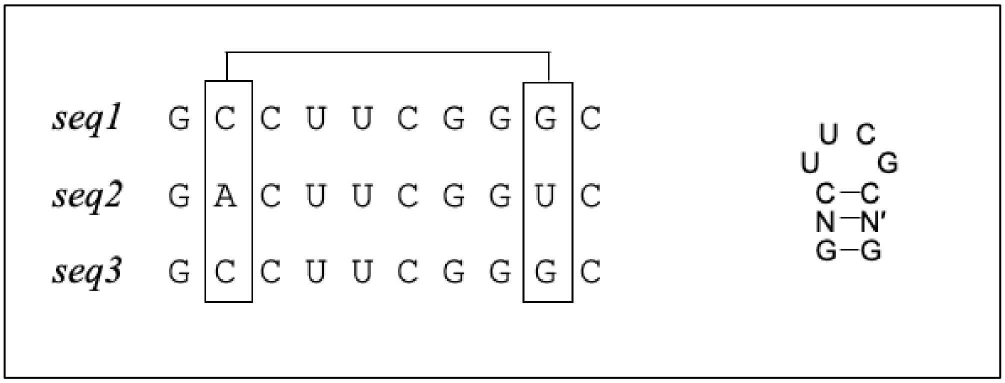


Figure 4.0 – Watson-Crick complementarity

Free Energy Minimization - Dynamic Programming

RNA folding prediction by energy minimization is widely used to predict RNA secondary structure from sequence. For a significant number of RNA molecules, the secondary structure of the RNA molecule is indicative of its function and its computational prediction by minimizing its free energy is important for its functional analysis. A general method for free energy minimization to predict RNA secondary structures is dynamic programming. The secondary structure of an RNA molecule is a representation of the pattern, given an initial RNA sequence, of complementary base-pairings that are formed between the nucleic acids. The sequence, represented as a string of four letters, is a single strand consisting of the nucleotides A, C, G, and U, which are generally assumed to pair to form a secondary structure with minimum free energy. Therefore, aside from the comparative method in case a multiple sequence alignment of homologous sequences is available, if a practitioner would like to predict secondary structure from an RNA sequence that has no homologs or that belongs to a poorly annotated family in a comprehensive RNA database such as Rfam [5], then free energy minimization should be the method of choice.

Since 1970s, folding prediction problem of RNA sequence has been an active area of research. Dynamic programming methods for this problem were initially developed in Nussinov et. Al [6] followed by Waterman Smith et al. [7] and Nussinov et al. [8] later by Zuker et. Al [9].

Energy during RNA Folding

RNA comprises of 4 nucleotide bases (A,U, G & C) as discussed earlier. These bases are attached to ribose sugars in the RNA backbone. Hydrogen bonds can form between complementary nucleotide bases. Complimentary nucleotides which are bond together is G&C and A&U hydrogen bonds. As a result, energy is released upon formation of the bond. Therefore, RNA molecules become more stable.

Calculating the energies in folding. Suppose we have a Nucleotide pairing energy output table:

	Y:	A	C	C	U
X:	A				-2.1
	C			-3.3	
	G		-2.4		-1.4
	U	-2.1		-2.1	

These are the energy values when a hydrogen bond is formed.

A – U = -2.1   U – A = -2.1   G – C = -2.4

U – A = -2.1   C – G = -3.3

So, if we calculate the total energy released as a result of the formation of 5 H-bonds will be -2.1-2.1-2.4-2.1-3.3 = -12.0 kcal/mol.

The Nussinov Folding Algorithm

The simplest approach to predicting the secondary structure of RNA molecules is to find the configuration with the greatest number of bases paired. The number of possible configurations to be inspected grows exponentially with the length of the sequence. We can employ dynamic programming algorithm to find an efficient solution using Nussinov Folding Algorithm. This algorithm was proposed in 1978.

Nussinov is a recursive algorithm. It calculates the best structure for small subsequences and works its way outward to larger and larger subsequences. The key idea of the recursive calculation is that there are only four possible ways of the getting the best structure for i, j from the best structures of the smaller subsequences. It computes the highest number of nucleotide coupling with 2 structure.

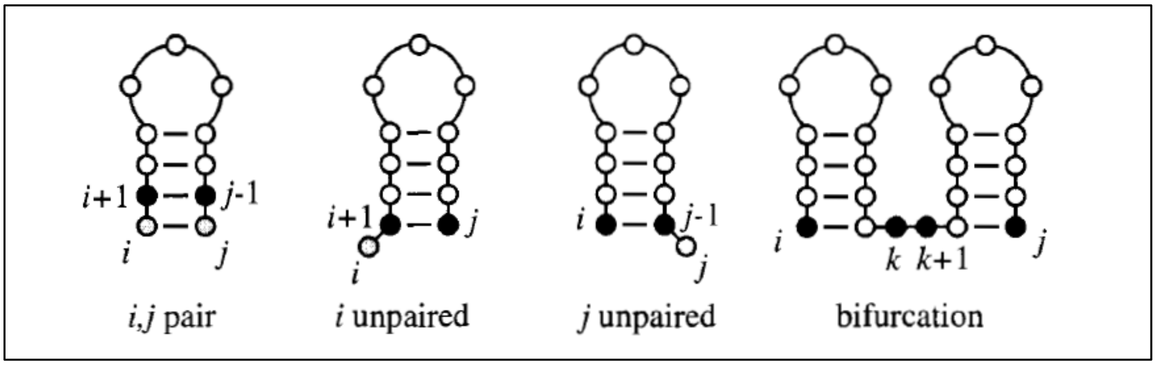


Figure 5.0

Idea: The score $S(i, j)$ is the maximum of the following 4 possibilities.

1. $\boldsymbol{S}(\boldsymbol{i}+\boldsymbol{1},\ \boldsymbol{j}) - i$ unpaired– Add unpaired position i onto best structure for subsequence $i+1,\ j$; Take lower element.
2. $\boldsymbol{S}(\boldsymbol{i},\ \boldsymbol{j}-\boldsymbol{1}\ ) - j$ unpaired– Add unpaired position i onto best structure for subsequence $\boldsymbol{i},\ \boldsymbol{j}-\boldsymbol{1}$; Take left element.
3. $\boldsymbol{S}(\boldsymbol{i}+\boldsymbol{1},\ \boldsymbol{j}-\boldsymbol{1}) +\ \boldsymbol{e}(\boldsymbol{i},\boldsymbol{j})$– i, j paired– Add unpaired position i, j onto best structure for subsequence $\boldsymbol{i}+1,\ \ \boldsymbol{j}-\boldsymbol{1}$; Take the diagonally left Lower Down
4. Bifurcation, or combining two optimal substructures ranging from i <k <j. $\boldsymbol{Max}\ \boldsymbol{i}<\boldsymbol{k}<\boldsymbol{j}\,\{\ \boldsymbol{S}(\boldsymbol{i},\boldsymbol{k})\ +\ \boldsymbol{S}(\boldsymbol{k}+\boldsymbol{1}),\ \boldsymbol{j}\}$– bifurcation – combine two optimal substructure i, k and $k+1,\ j$. Take left row, bottom column.

Given a sequence $x\ =\ (x1, \ldots, xL)$ of length L. We set S($\boldsymbol{i},\boldsymbol{j}$) = 1, if $xi\ -\ xj$ is a canonical base pair and 0, else.

The dynamic programing algorithm has two stages:

In the Matrix Fill stage, we will recursively calculate scores γ(i, j) which are the maximal number of base pairs that can be formed for subsequences $(xi, \ldots, xj)$.

In the Matrix Traceback stage, we traceback through the calculated matrix to obtain one of the maximally base paired structures.

Flow chart for Nussinov algorithm

The flow chart for Nussinov algorithm is shown in fig 6.0 and described below:

1. Obtain RNA sequence for which we have to predict RNA secondary structure.
2. Create a scoring matrix and place sequence – The matrix will be empty initially and then, place the RNA sequence on top as well as left side of the matrix.
3. Set the diagonals of the matrix to zero as well as the lower diagonal.
4. Devise Scoring Scheme – The scoring scheme of Nussinov Algorithm can be done by looking at the bottom, at the left, at the diagonal as well as the two rows beyond the left and bottom.
5. Set each matrix position by calculating 4 conditions.

6. Set the maximum value from the 4 conditions into matrix position.

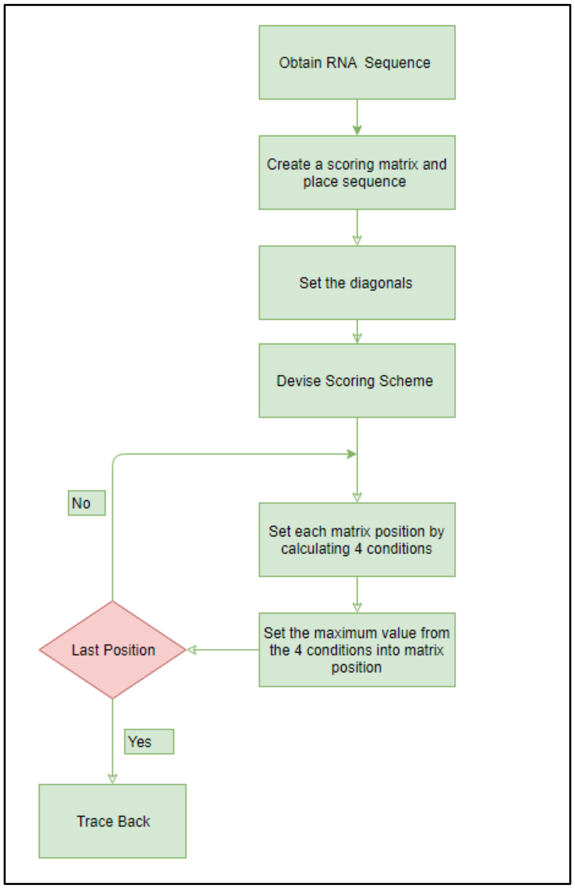

Obtain RNA Sequence
Create a scoring matrix and place sequence
Set the diagonals
Devise Scoring Scheme
No
Set each matrix position by calculating 4 conditions
Last Position
Set the maximum value from the 4 conditions into matrix position
Yes
Trace Back

Figure 6.0

7. Fill all the matrix position that are above the diagonal in the matrix and check if it's last position, if yes go to 8th step else go to 5th step.
8. Trace back – Once the matrix is filled, a new step has to be done i.e. called the trace back which will help to extract meaning from the score which we have put into the matrix.
9. Trace back strategy is used to recover the optimal structure.

The Matrix Fill Stage:

Algorithm (Nussinov RNA folding, fill stage):

- Input: Sequence $x = (x1, x2, . . . , xL)$
- Output: Maximal number $S(i, j)$ of base pairs for $(xi , . . . , xj)$.
- Initialization:

$S(i, i) = 0$ *for I = 2 to L.*

$S(i, i\text{-}1) = 0$ *for I = 2 to L;*

```
for $n = 2\ to\ L\ do$
    for $j = n\ to\ L\ do$
      Set $i = j - n + 1$
      $S(i,j) = \max \begin{cases} S(i+1,j) \\ S(i,j-1) \\ S(i+1,j-1) + e(i,j) \\ \boldsymbol{Maxi < k < j}\{S(i,k) + s(k+1),\ j\} \end{cases}$
Return $S(1,\ L)$
```

Consider the sequence x = GGGAAAUCC. Here is the matrix after initialization will be:

S(i, i) = 0 *for I = 2 to L.*

S(i, i-1) = 0 *for I = 2 to L;*

After using recursively, the four possibilities from:

$$S(i,j) = \max \begin{cases} S(i+1,j) \\ S(i,j-1) \\ S(i+1,j-1) + e(i,j) \\ \boldsymbol{Maxi < k < j}\{S(i,k) + s(k+1),\ j\} \end{cases}$$

We get

		1	2	3	4	5	6	7	8	9
		G	**G**	**G**	**A**	**A**	**A**	**U**	**C**	**C**
1	**G**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**2**	**3**
2	**G**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**2**	**3**
3	**G**		**0**	**0**	**0**	**0**	**0**	**1**	**2**	**2**
4	**A**			**0**	**0**	**0**	**0**	**1**	**1**	**1**
5	**A**				**0**	**0**	**0**	**1**	**1**	**1**
6	**A**					**0**	**0**	**1**	**1**	**1**
7	**U**						**0**	**0**	**0**	**0**
8	**C**							**0**	**0**	**0**
9	**C**								**0**	**0**

The Matrix Traceback Stage

Algorithm Traceback (i, j)

Input: Matrix S and positions i, j.

Output: Secondary structure maximizing the number of base pairs.

Initial call: traceback $(i = 1,\ j = L)$.

```
if $i < j$ then
   if $S(i, j) = S(i + 1, j)$ then                          // case (1)
         traceback $(i + 1, j)$
   else if $S(i, j) = S(i, j - 1)$ then                     // case (2)
         traceback $(i, j - 1)$
   else if $S(i, j) = S(i + 1, j - 1) + w(i, j)$ then       // case (3)
         print base pair $(i, j)$
         traceback $(i + 1, j - 1)$
   else for $k = i + 1$ to $j - 1$ do                        // case (4)
      if $S(i, j) = S(i, k) + S(k + 1, j)$ then
            traceback $(i, k)$
            traceback $(k + 1, j)$
            break
end
```

Here is the traceback through $S(i :\downarrow, j :\rightarrow)$:

		1	2	3	4	5	6	7	8	9
		G	**G**	**G**	**A**	**A**	**A**	**U**	**C**	**C**
1	**G**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**2**	**3**
2	**G**	**0**	**0**	**0**	**0**	**0**	**0**	**1**	**2**	**3**
3	**G**		**0**	**0**	**0**	**0**	**0**	**1**	**2**	**2**
4	**A**			**0**	**0**	**0**	**0**	**1**	**1**	**1**
5	**A**				**0**	**0**	**0**	**1**	**1**	**1**
6	**A**					**0**	**0**	**1**	**1**	**1**
7	**U**						**0**	**0**	**0**	**0**
8	**C**							**0**	**0**	**0**
9	**C**								**0**	**0**

And the resulting secondary structure is shown in figure 7.0:

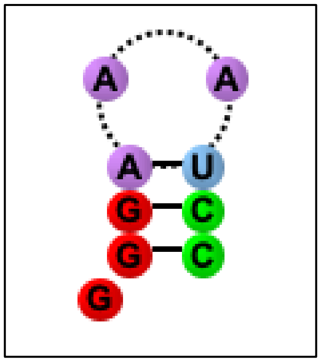


Figure 7.0

There are few limitation of The Nussinov Algorithm:

- Only one structure predicted as base pair maximization cannot differentiate structures sufficiently well. It doesn't create viable secondary structure.
- It does not provide sub-optimal solutions.
- Crossing structures cannot be predicted using Nussinov algorithm.
- Base pair maximization does not yield biologically relevant structures.

Zuker Algorithm

RNA folding algorithm is dictated by biophysics rather than by counting and maximizing the number of base pairs. The most sophisticated secondary structure prediction method for single RNAs is the ZUKER algorithm, an energy minimization algorithm which assumes that the correct structure is the one with the lowest equilibrium free energy (ΔG).

In Zuker's algorithm we need to evaluate the free or Gibbs energy that is available in RNA secondary structure. In Thermodynamics, the Gibbs free energy G describes the energetics of molecules in aqueous solution. The change ∆G of the free energy in a chemical process, such as nucleic acid folding, determines the direction of the process:

$$G = H - TS$$

where H is the enthalpy (potential to perform work), T the absolute temperature and S the entropy (measure of disorder).

- $\Delta G = 0$ indicates equilibrium
- $\Delta G > 0$ indicates an unfavorable process and
- $\Delta G < 0$ indicates a favorable process.

Hence, biomolecules in solution arrange themselves so as to minimize the free energy of the entire system (biomolecules + solvent)

The ΔG of an RNA secondary structure is approximated as the sum of individual contributions from:

- Loops,
- Stacked base pairs,
- Other secondary structure elements

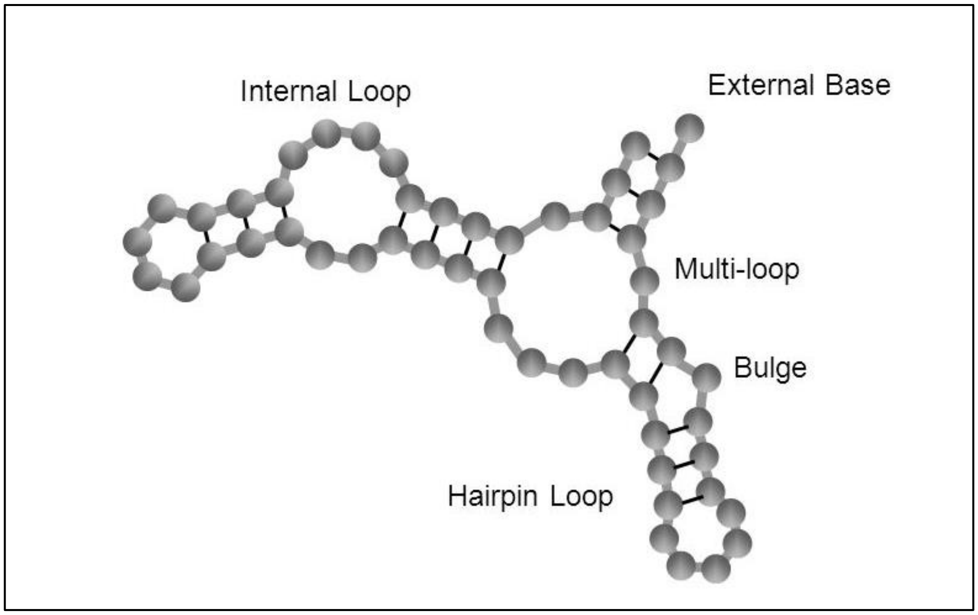


Figure 8.0

We can see that the important difference between Nussinov algorithm and Zuker Algorithm is that Zuker's algorithm focuses on loop rather than base pairs. Also, we have two different kinds of energy in the RNA secondary structure which are:

1. **Stabilized Energy** - The energy released as a result of coupling complementary nucleotide. It is denoted with negative energy.

2. **Destabilized Energy** – Nucleotide of bulge, hairpin, Internal loops etc. are not coupled. These uncoupled nucleotides tend to make RNA structure unstable. It is denoted with positive energy.

And we need to find Sum of Energy which is the sum of Stabilized Energy and Destabilized energy and these sum of energies help us in determining the quality of secondary RNA structure. In Zuker's algorithm, the energy which is given out after two complementary nucleotides form a hydrogen bond. i.e. negative value. And the more negative value is, the more stable the structure will be of secondary RNA structure. There will be some nucleotide that are not coupled. They contribute instability and they are positive values.

Zuker's Algorithm is developed a more involved dynamic program that uses loop-dependent rules. We must use two matrices W and V.

For $i < j$, let $W(i, j)$ denote the minimum folding energy of all non-empty folding of the subsequence $x_i, \dots, x_j$.

Additionally, let $V(i, j)$ denote the minimum folding energy of all non-empty folding of the subsequence $x_i, \dots, x_j$, containing the base pair (i, j). The following obvious fact is crucial:

$$W(i, j) \leq V(i, j) \text{ for all } i, j.$$

These matrices are initialized as follows:

$$W(i, j) = V(i, j) = \infty \; for\ all\ i, j\ with\ j - 4 < i < j$$

Loop Dependent Energies

Energy contributions of the various structure elements are:

- $\mathrm{eh}(\mathrm{i}, \mathrm{j})$ be the energy of the hairpin loop closed by the base pair (i, j).
- $\mathrm{es}(\mathrm{i}, \mathrm{j})$ be the energy of the stacked pair (i, j) and $(\mathrm{i} + 1, \mathrm{j} - 1)$,
- $\mathrm{ebi}(\mathrm{i}, \mathrm{j}, \mathrm{i}', \mathrm{j}')$ be the energy of the bulge or interior loop that is closed by (i, j), with $(\mathrm{i}', \mathrm{j}')$ accessible from (i, j), and

- Generally, multi loop contribution will be too expensive in prediction. We need to use a constant energy function for multi-loops. We need to use a simplified contribution scheme.

$$\mathrm{multiloop}\ \mathrm{eM}(\mathrm{i},\mathrm{j},\mathrm{k},\mathrm{k'}) = \mathrm{a} + \mathrm{bk} + \mathrm{ck'}$$

$\mathrm{a}, \mathrm{b}, \mathrm{c}$ = weights/Constants
a = energy contribution for closing of loop
k = number of inner base pairs
$\mathrm{k'}$= number of unpaired bases within loop

In general, multi-loop energy depends on everything like inner base pairs, closing base pair, and sequence. We can simplify multi-loop as dependency is only on most inner base pair k and number of unpaired bases $\mathrm{k'}$. For example, if we calculate the general and simplified multi-loop energy it will be:

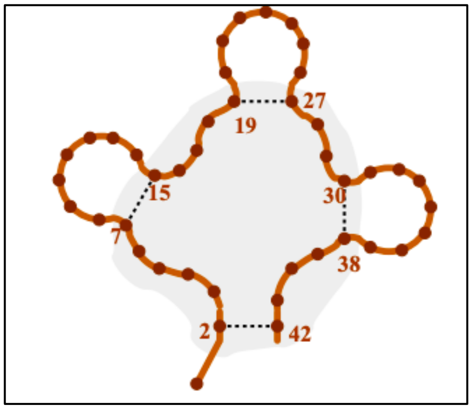
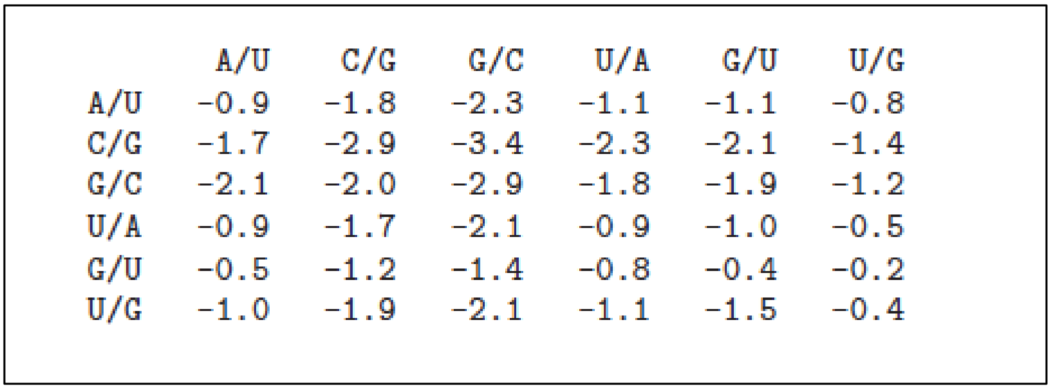


Figure 9.0

In general: eM (2, 42, 7, 15, 19, 27, 30, 38)
simplified: eM (2, 42, k, k') = $a + bk + ck'$,
where

k = 3: inner base pairs within loop
k'= 12: unpaired bases within multi-loop

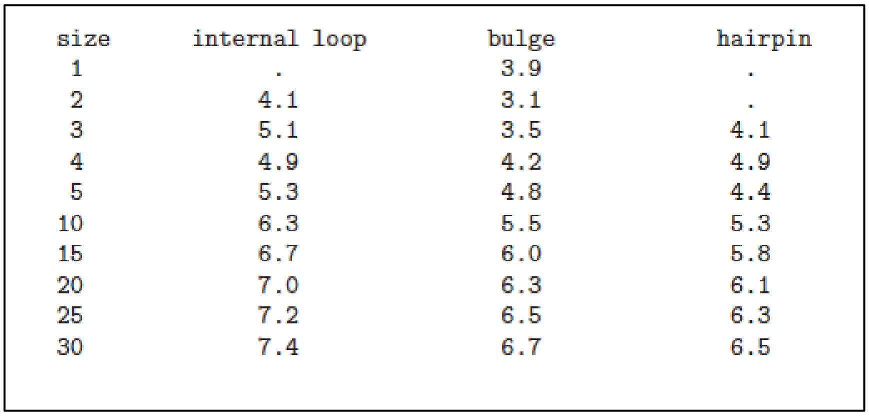
Predicted free-energy values (kcal/mole at 37oC) for base pair stacking:



	A/U	C/G	G/C	U/A	G/U	U/G
A/U	-0.9	-1.8	-2.3	-1.1	-1.1	-0.8
C/G	-1.7	-2.9	-3.4	-2.3	-2.1	-1.4
G/C	-2.1	-2.0	-2.9	-1.8	-1.9	-1.2
U/A	-0.9	-1.7	-2.1	-0.9	-1.0	-0.5
G/U	-0.5	-1.2	-1.4	-0.8	-0.4	-0.2
U/G	-1.0	-1.9	-2.1	-1.1	-1.5	-0.4

Figure 10.0

Predicted free-energy values (kcal/mole at 37oC) for features of predicted RNA secondary structures, by size of loop:



size	internal loop	bulge	hairpin
1	.	3.9	.
2	4.1	3.1	.
3	5.1	3.5	4.1
4	4.9	4.2	4.9
5	5.3	4.8	4.4
10	6.3	5.5	5.3
15	6.7	6.0	5.8
20	7.0	6.3	6.1
25	7.2	6.5	6.3
30	7.4	6.7	6.5

Figure 11.0

The Main Recursion

For all i, j with $1 \leq i < j \leq L$:

$$W(i,j) = \min \begin{cases} W(i+1,j) \\ W(i,j-1) \\ V(i,j) \\ \min_{i<k<j}\{W(i,k) + W(k+1,j)\}, \end{cases}$$

and

$$V(i,j) = \min \begin{cases} eh(i,j) \\ es(i,j) + V(i+1, j-1) \\ VBI(i,j), \\ VM(i,j), \end{cases}$$

Where,

$$VBI(i,j) = \min_{\substack{i < i' < j' < j \\ i' - i + j - j' > 2}} \{ebi(i,j,i',j') + V(i',j')\},$$

and

$$VM(i,j) = \min_{i<k<j-1} \{W(i+1,k) + W(k+1,j-1)\} + a.$$

Zuker’s Algorithm Flow Chart

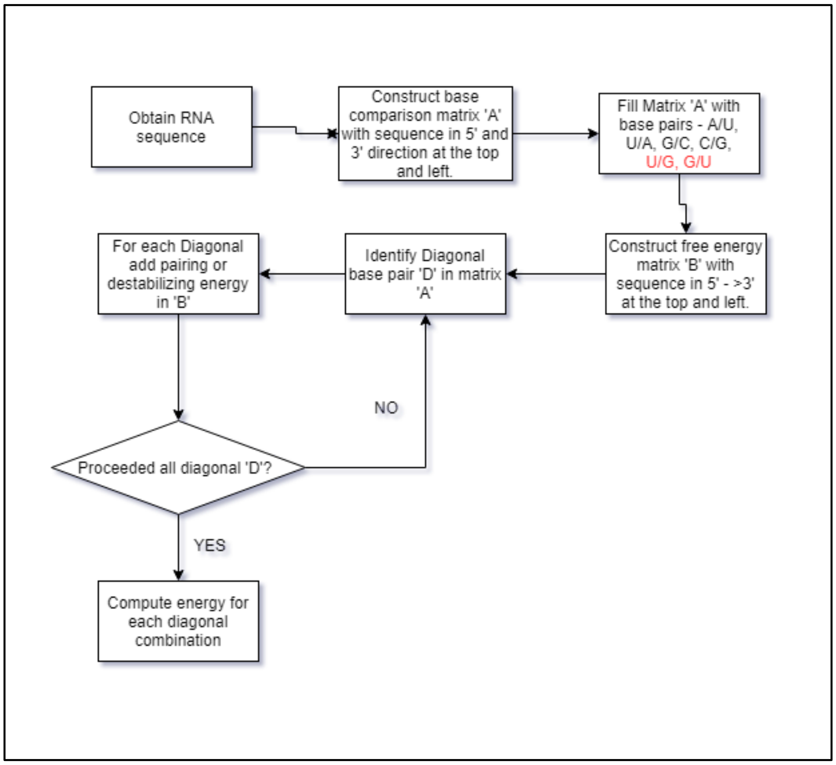

Obtain RNA sequence
Construct base comparison matrix 'A' with sequence in 5' and 3' direction at the top and left.
Fill Matrix 'A' with base pairs - A/U, U/A, G/C, C/G, U/G, G/U
Construct free energy matrix 'B' with sequence in 5' - >3' at the top and left.
Identify Diagonal base pair 'D' in matrix 'A'
For each Diagonal add pairing or destabilizing energy in 'B'
NO
Proceeded all diagonal 'D'?
YES
Compute energy for each diagonal combination

Figure 12.0

Example of Energy Calculation:

Here is any example of the full energy calculation for an RNA stem loop:

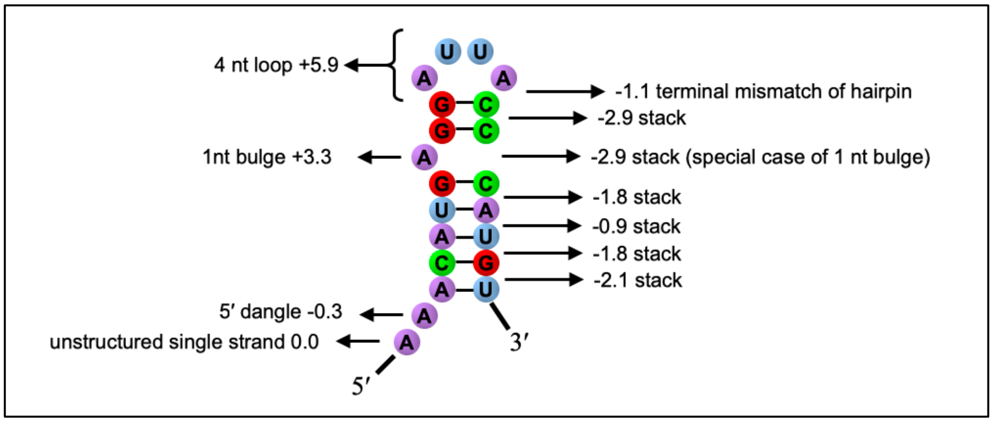


Figure 13.0

Total ΔG = -4.6kcal/mol.

Time Analysis of Zuker's Algorithm:

The minimum folding energy $Emin$ is given by $W(1, L)$.

There are $O(L2)$ pairs (i, j) satisfying $1 \le i < j \le L$.

The computation of:

1. W takes $O(L3)$ steps,
2. V takes $O(L2)$ steps,
3. VBI takes $O(L4)$ steps, and
4. VM takes $O(L3)$ steps,

and so, the total run time is $O(L4)$.

The most practical way to reduce the run time to $O(L3)$ is to limit the size of a bulge or interior loop to some fixed number d, usually about 30. This is achieved by limiting the search by considering all possible ways to define a bulge or interior loop that involves a base pair (i', j') and is closed by (i, j). In each situation, we have a contribution from the bulge or interior loop and a contribution from the structure that is on the opposite side of (i', j') to $2 < i' - i + j - j' - 2 \le d$.

References -

1. Tinoco, I., Jr & Bustamante, C. (1999). How RNA folds. J. Mol. Biol. 293, 271–281.
2. Pace, N. R., Thomas, B. C. & Woese, C. R. (1999). Probing RNA structure, function, and history by comparative analysis. In The RNA World (Gesteland, R. F., Cech, T. R. & Atkins, J. F., eds), pp. 113–141, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3. Gutell, R. R., Lee, J. C. & Cannone, J. J. (2002). The accuracy of ribosomal RNA comparative structure models. Curr. Opin. Struct. Biol. 12, 301–310.
4. Michel, F., Costa, M., Massire, C. &Westhof, E. (2000). Modeling RNA tertiary structure from patterns of sequence variation. Methods Enzymol. 317, 491–510.
5. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR (2003) Rfam: an RNA family database. Nucleic Acids Res 31: 439–441
6. Nussinov R, Pieczenik G, Grigg JR, Kleitman DJ (1978) Algorithms for loop matchings. SIAM J Appl Math 35:68–82
7. Waterman MS, Smith TF (1978) RNA secondary structure: a complete mathematical analysis. Math Biosci 42:257–266
8. Nussinov R, Jacobson AB (1980) Fast algorithm for predicting the secondary structure of single stranded RNA. Proc Natl Acad Sci U S A 77(11):6309–6313
9. Zuker M, Sankoff D (1984) RNA secondary structures and their prediction. Bull Math Biol 46:591–621