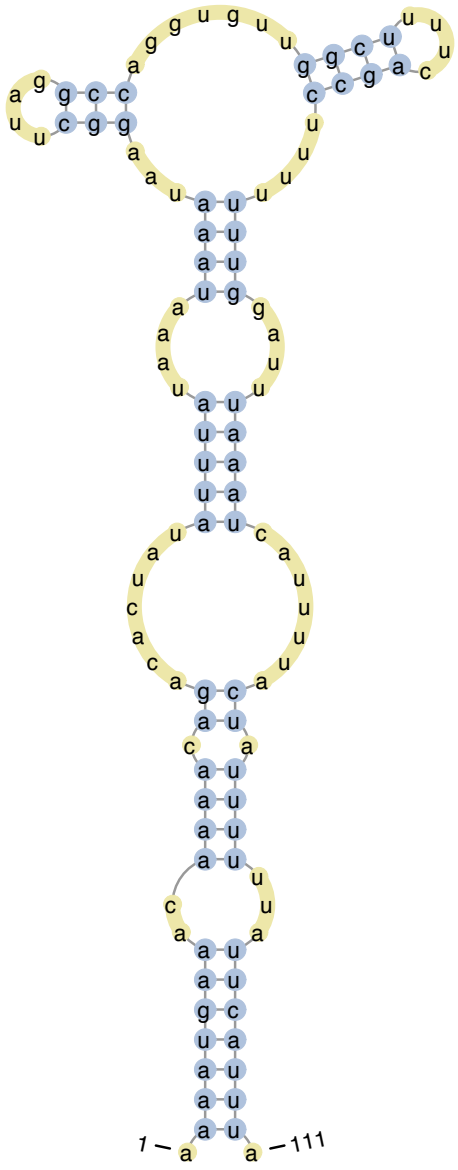
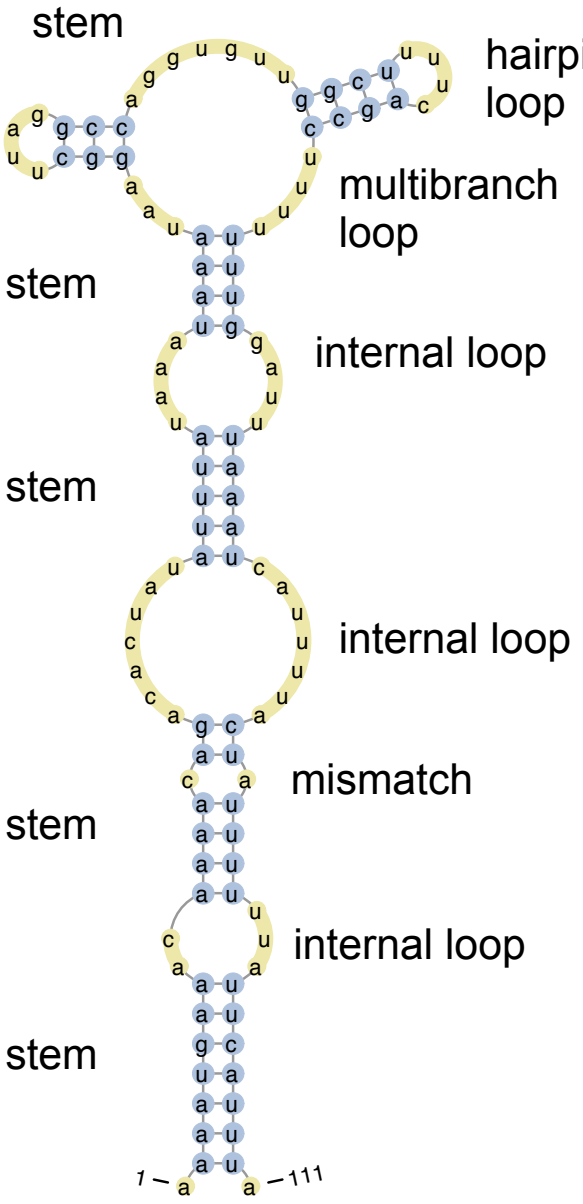
RNA secondary structure: thermodynamics and structure predictions

RNA secondary structure



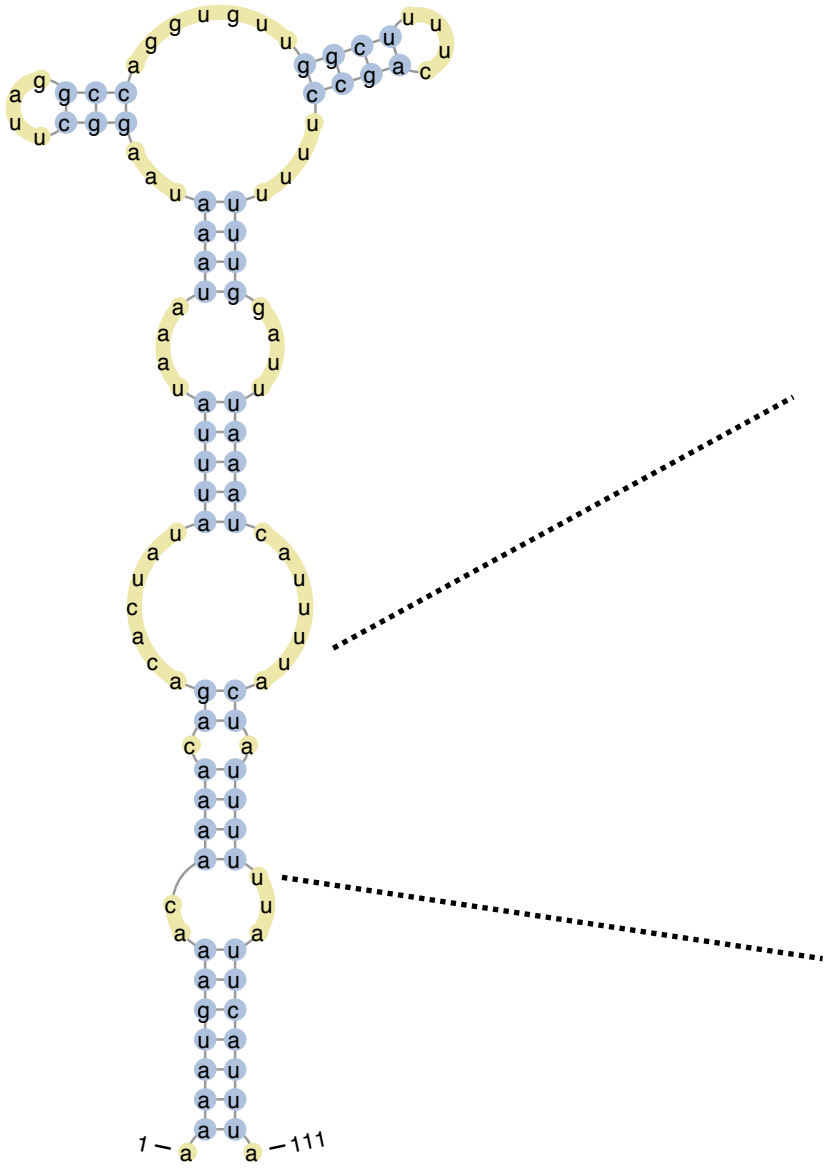
RNA secondary structure



At physiological conditions, all RNAs fold into secondary structures consisting of double-helical stems and single-stranded loops.

Stems stabilize the folding (Gibbs free energy $\Delta G < 0$), loops destabilize it ($\Delta G > 0$)

RNA secondary structure



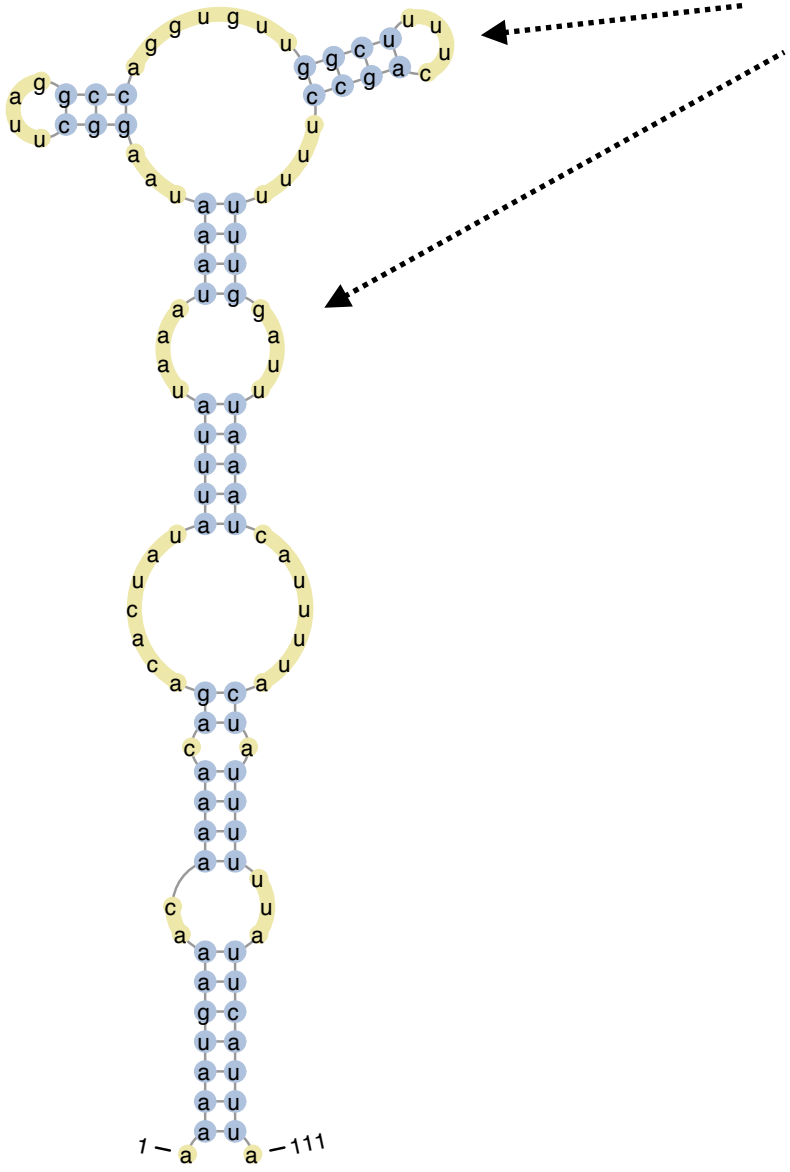
ΔG(stem): stacking of base pairs.

Usually described in terms of the nearest-neighbor model:

```
          ΔG(stem) =
A   A     ΔG(GC/AA) +
 G-C      ΔG(AU/GC) +
 A-U      ΔG(UA/AC) +
C   A     ΔG(AU/CA) +
 A-U      ΔG(AU/AU) +
 A-U      ΔG(AU/AU) +
 A-U      ΔG(AU/AU) +
 A-U      ΔG(UA/UC)
C   U
```

Adjacent mismatches are also stacked on base pairs.

RNA secondary structure



ΔG(loop): the first approximation is zero enthalpic contribution:

$\Delta G = \Delta H - T\Delta S = - T\Delta S \quad (\Delta H = 0)$

ΔS - conformational entropy, can be calculated as

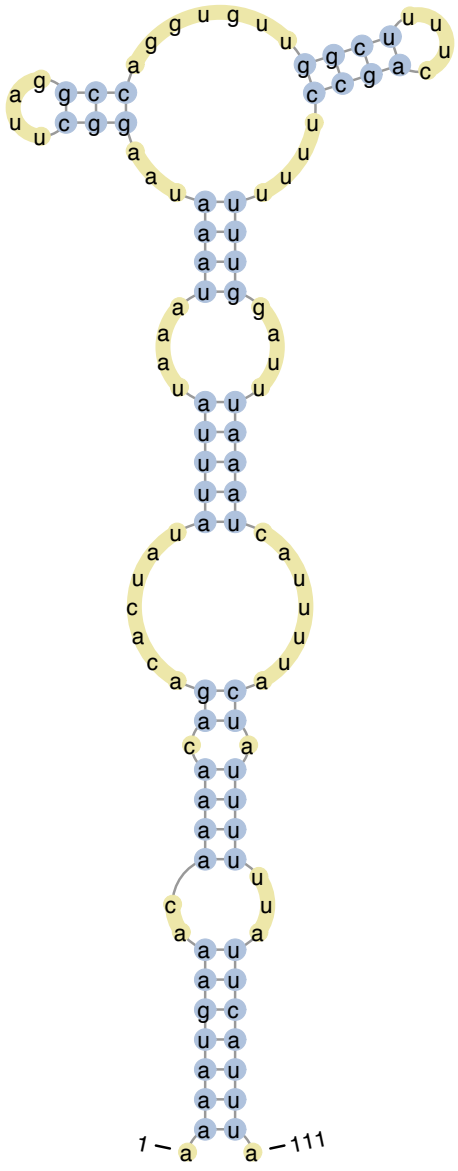
$\Delta S = - R [A + 1.75 \ln N]$, where

R - universal gas constant,
A depends on loop configuration,
N - number of monomers in the loop.

In practice, ΔG of loops can be approximated as

$\Delta G = a + b \times \ln (N)$, where a and b are fitted for various loop topologies.

RNA secondary structure



Apart from the nearest-neighbor stacking free energies in the stems and conformational free energies of loops, there are other contributions taken into account by various RNA secondary structure models, such as:

- sequence-dependent loop contributions (extra stable hairpin tetraloops, triloops, small internal loops etc.);

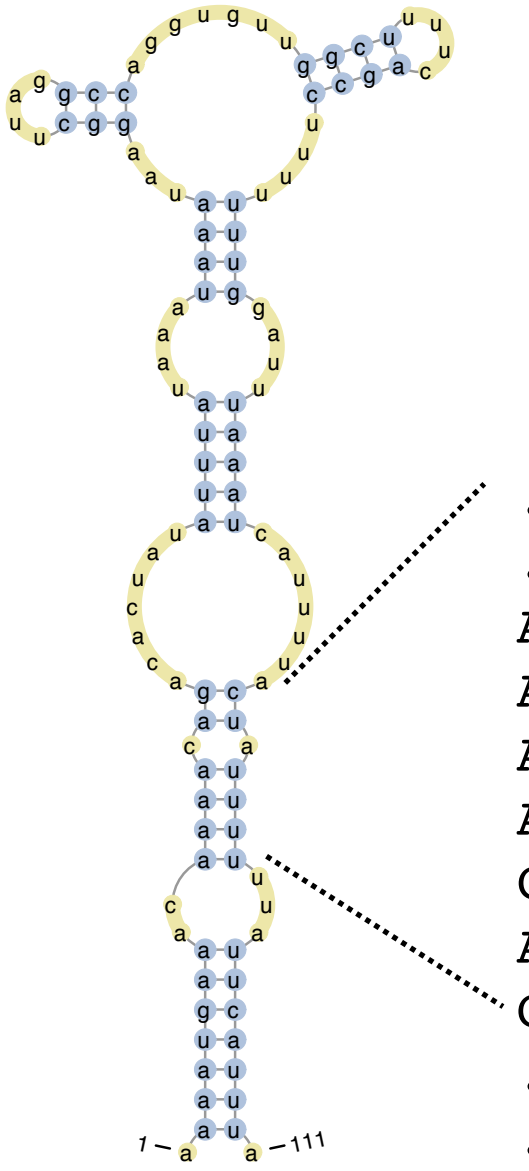
- asymmetry terms in internal loops;

- coaxial stacking of stems;

- closing mismatches in hairpins and internal loops;

- dangling ends.

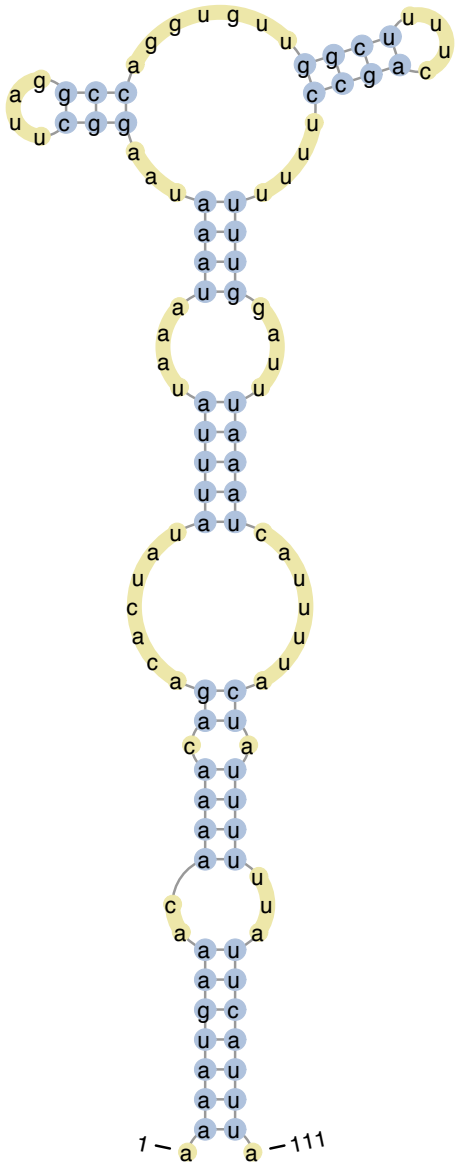
RNA secondary structure prediction



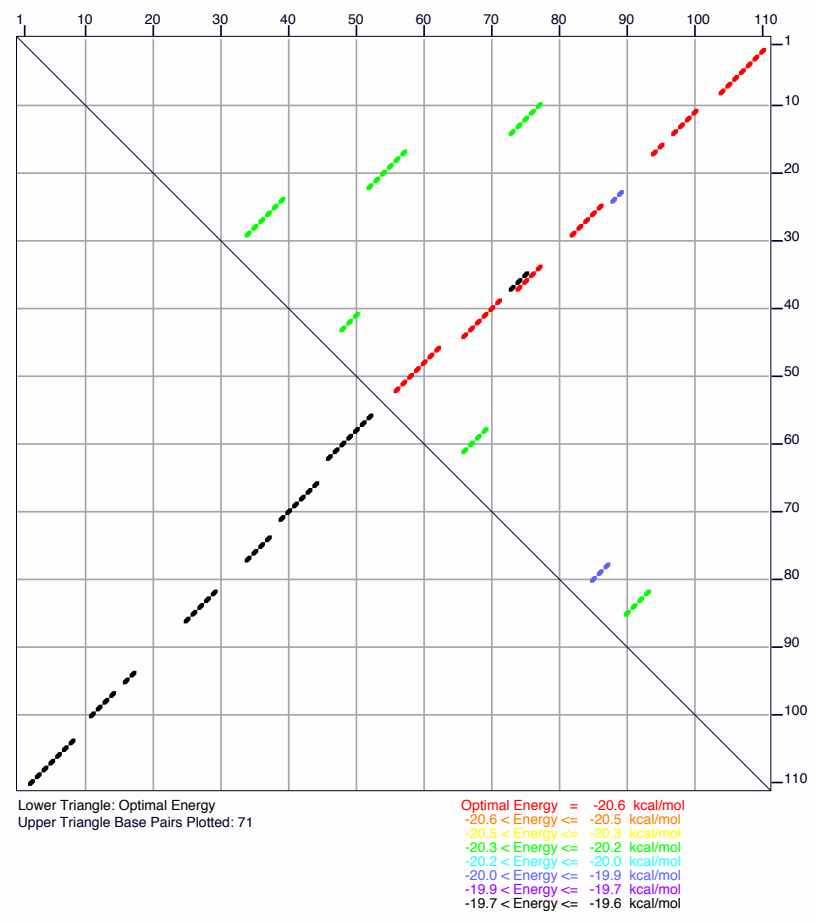
Given the set of free energy parameters, the lowest free energy state can be computed using a dynamic programming algorithm. The approach resembles the alignment algorithm.

```
..A A A A C A G. . .C U A U U U U .
.
.
A
A
A
A
C
A
G
.
.
```

RNA secondary structure prediction



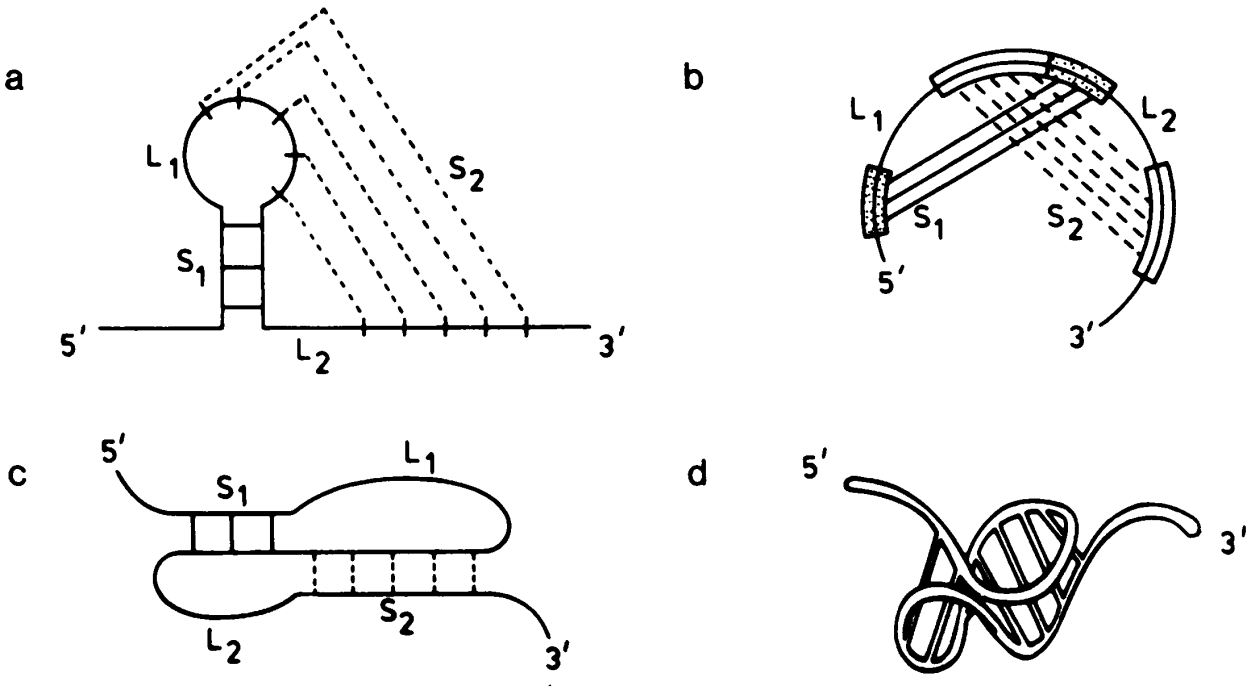
Dynamic programming algorithm can compute both optimal and suboptimal structures that can be shown in various ways in e.g. “dot-plots”:



Prediction of RNA pseudoknots

Pseudoknot formation:

[“H(airpin)-pseudoknot"]

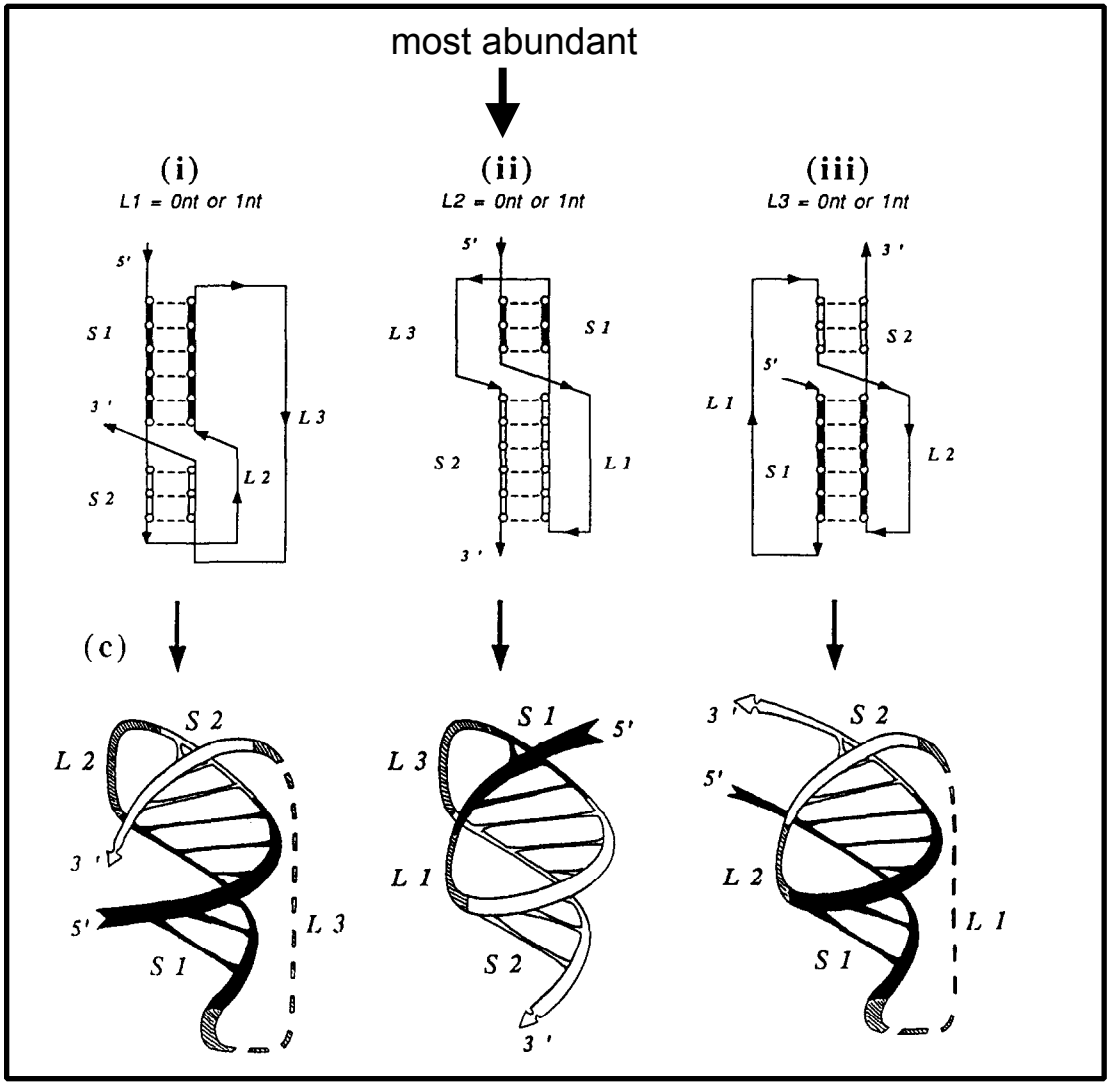


(Pleij et al., 1985)

Prediction of pseudoknot structures is more difficult as compared to the “orthodox” RNA secondary structure. Many programs ignore pseudoknot formation.

Prediction of RNA pseudoknots

H-pseudoknots are usually stabilized by coaxial stacking of two stems. Among three possible stacking topologies, one (below in the centre) is the most abundant “classic” pseudoknot. The pseudoknot loops are topologically different: in classic pseudoknot the loop L1 crosses helical deep groove whereas L2 crosses shallow groove.



(Jaeger et al., 1991)

Prediction of RNA pseudoknots

Conformational free energies of pseudoknot loops depend on both their sizes and lengths of the crossed stems. The approximated values and conformational restrictions are different for loops crossing deep and shallow grooves of RNA helices. Various approximations were suggested for the use in computer programs, one of the first ones is shown below:

ΔG°_{37} :

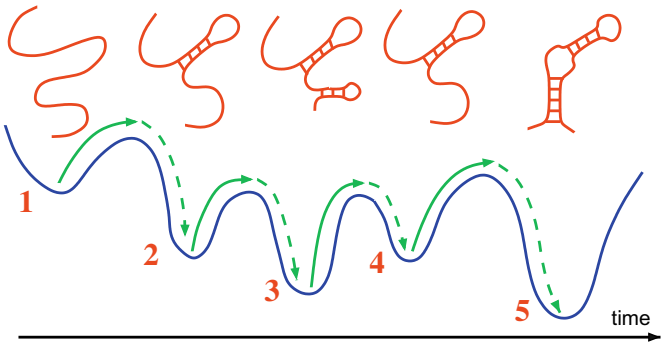
stem size (bp)	loop size (nt) 1	2	3	4	5	6	8	10	15	20	30
(S2)	deep groove (L1)										
2	—	—	—	—	—	—	—	—	—	—	—
3	—	5.0	5.7	6.2	6.5	6.7	7.1	7.4	7.8	8.2	8.6
4	4.7	5.4	5.9	6.2	6.4	6.6	6.9	7.2	7.6	7.9	8.4
5	4.2	4.9	5.4	5.7	5.9	6.1	6.4	6.7	7.1	7.4	7.9
6	3.5	4.2	4.7	5.0	5.2	5.4	5.7	6.0	6.4	6.7	7.2
7	3.5	4.2	4.7	5.0	5.2	5.4	5.7	6.0	6.4	6.7	7.2
8	—	4.2	4.9	5.4	5.7	5.9	6.3	6.6	7.0	7.4	7.8
9	—	4.7	5.4	5.9	6.2	6.4	6.8	7.1	7.5	7.9	8.3
10	—	5.0	5.7	6.2	6.5	6.7	7.1	7.4	7.8	8.2	8.6
(S1)	shallow groove (L2)										
2	—	—	—	—	—	—	—	—	—	—	—
3	—	3.5	4.2	4.7	5.0	5.2	5.6	5.9	6.3	6.7	7.1
4	—	—	4.2	4.9	5.4	5.7	6.1	6.4	7.0	7.3	7.8
5	—	—	—	4.7	5.4	5.9	6.4	6.8	7.4	7.8	8.3
6	—	—	—	5.0	5.7	6.2	6.7	7.1	7.7	8.1	8.6
7	—	—	—	—	5.2	5.9	6.7	7.1	7.8	8.2	8.7
8	—	—	—	—	5.4	6.1	6.9	7.3	8.0	8.4	8.9
9	—	—	—	—	—	5.6	6.8	7.3	8.1	8.4	9.0
10	—	—	—	—	—	5.7	6.9	7.4	8.2	8.6	9.2

(Gultyaev et al., 1999)

RNA folding simulations

It is possible to simulate kinetic RNA folding as a stepwise process.

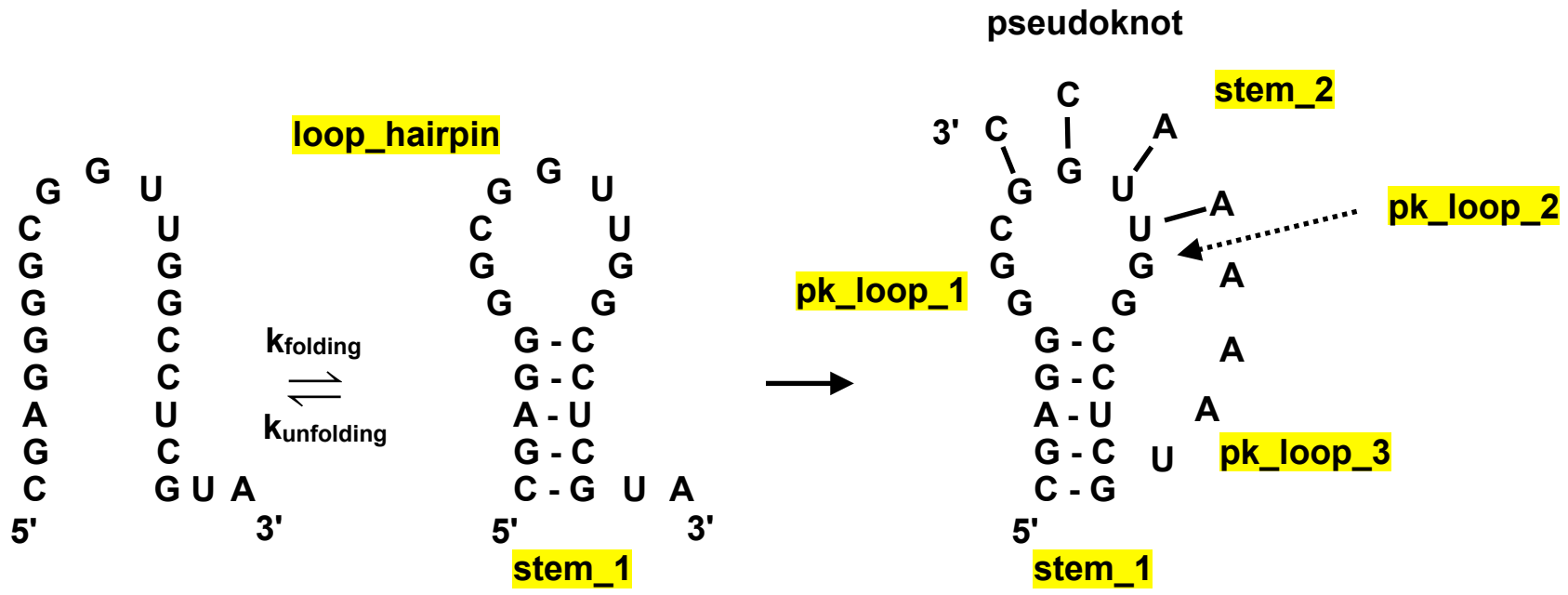
from Isambert (2009):



Energy barriers are mostly determined by positive free energies of loops and disruptions of stems with negative free energies.

RNA folding simulations

It is possible to simulate kinetic RNA folding as a stepwise process, for instance:



$\Delta G = \Delta G_{stem\_1} + \Delta G_{loop\_hairpin}$

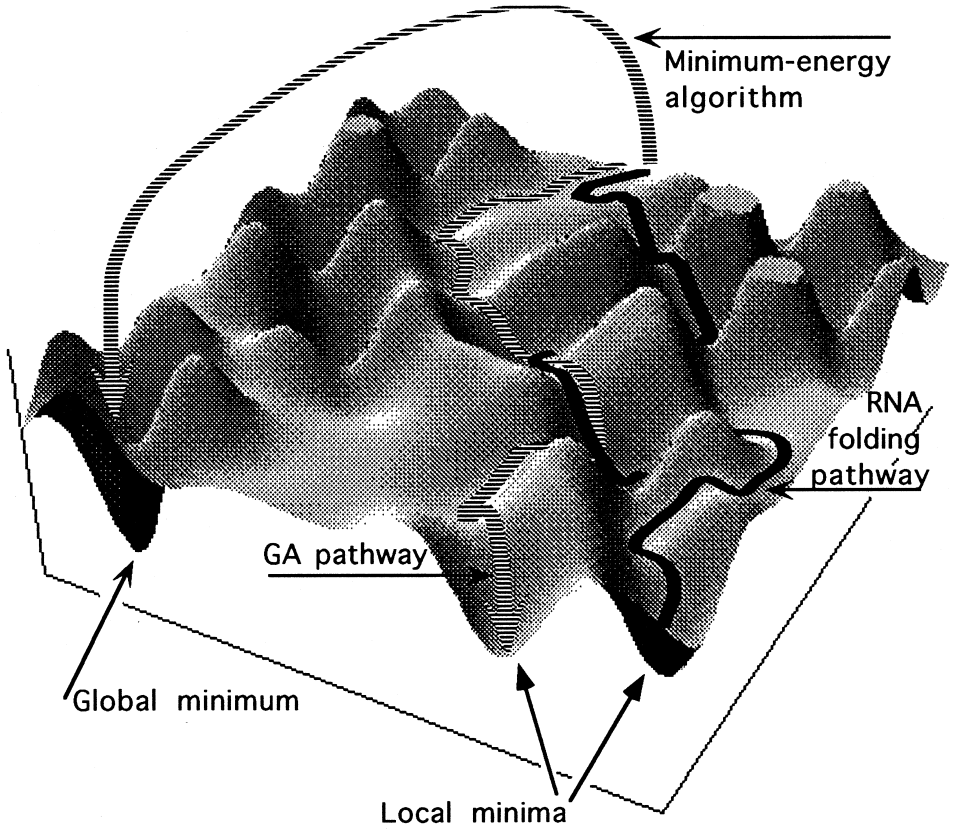
$k_{folding} \sim \exp(-\Delta G_{loop\_hairpin} / RT)$

$k_{unfolding} \sim \exp(\Delta G_{stem\_1} / RT)$

$\Delta G = \Delta G_{stem\_2} + \Delta G_{pk\_loop\_1} + \Delta G_{pk\_loop\_2} + \Delta G_{pk\_loop\_3} - \Delta G_{loop\_hairpin}$

- Such a simulation can use different approximations of elementary folding units, e.g. single base pairs or stems.
- The folding can be simulated by a stochastic algorithm like Monte Carlo (MC) or Genetic Algorithm (GA).

RNA folding simulations



from van Batenburg et al. (1995)

- In a stepwise RNA folding simulation by a stochastic algorithm, folding/unfolding transition steps are chosen with probabilities that depend on transition free energies.

- Advantages: predictions of metastable states in local free energy minima, pseudoknot structures, cotranscriptional RNA folding.
- Disadvantages: lack of knowledge of RNA kinetics and pseudoknot thermodynamics, computational time complexity.

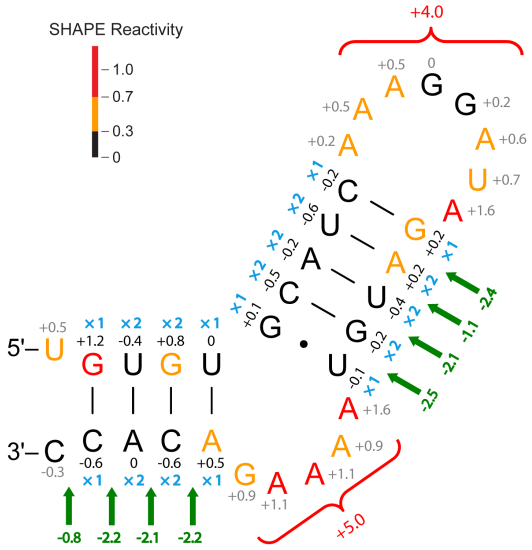
- Such a simulation can use different approximations of elementary folding units, e.g. single base pairs or stems.
- The folding can be simulated by a stochastic algorithm like Monte Carlo (MC) or Genetic Algorithm (GA).

RNA structure prediction using experimental data

The algorithms for RNA structure prediction can be greatly improved by an implementation of constraints that take into account experimental data (e.g. probing).

The most simple way is to force some nucleotides to be in a single-stranded conformation (prohibit their pairing).

Upon development of quantitative measurements of the nucleotide conformational states (SHAPE), more accurate incorporation of probing results has become possible. For instance, using pseudo-energy values. In SHAPE probing, nucleotide reactivities are the measure of “single-strandedness” (high for loops, low for helices).



$$\Delta G_{\mathrm{NN}} = \sum \Delta G_{\mathrm{stacks}} + \sum \Delta G_{\mathrm{loops}}$$

$$\Delta G_{\mathrm{SHAPE}} = \mathbf{1} \times \sum \Delta G_{\mathrm{ends}} + \mathbf{2} \times \sum \Delta G_{\mathrm{interior}}$$

$$\Delta G_{\mathrm{total}} = \Delta G_{\mathrm{NN}} + \Delta G_{\mathrm{SHAPE}}$$

A pseudo-energy can be introduced, for instance, as a term for every nucleotide in the probed sequence:

$\Delta G_{\mathrm{SHAPE}}(i) = m \times \ln[\mathrm{reactivity}(i) + 1] + b,$

where m and b are empirical constants.
E.g. m = 2.6 kcal/mol and b = −0.8 kcal/mol.

A pseudo-energy is added to the free energy and the total potential can be used in an algorithm that searches for free energy minima.
In such an implementation, only the pseudo-energy terms for base-paired nucleotides are added: once at the ends of helices and twice for interior nucleotides in helices, mimicking stacking contributions.

(Low & Weeks, 2010)