Partition function for circular RNA and Application

Ronny Lorenz
ronny@bioinf.uni-leipzig.de

Department of Bioinformatics
University of Leipzig

October 30, 2006

1. Partition function for linear RNAs
   - Structure decomposition
2. Extension for circular RNAs
   - Structure decomposition
3. Application
   - Occurence of circular RNAs
   - Metastable Structures

- let S be the set of all possible secondary structures for a given RNA sequence
- each $s \in S$ is associated with a free energy G_s

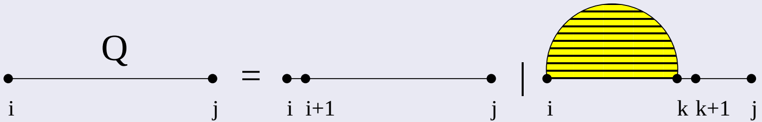
The partition function Q for a given sequence

$$Q = \sum_{s \in S} e^{\frac{-G_S}{RT}} \tag{1}$$

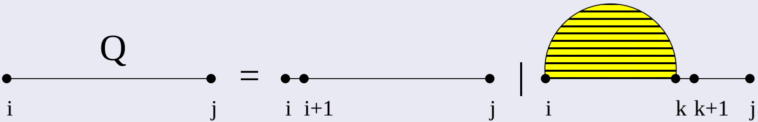
- i and j do not pair
- i and j form a hairpin loop, with closing pair [i,j],
- there exist two positions k,l with $i < k < l < j$ so that the subsequence i,j forms an interior loop with closing pairs i,j and k,l
- i and j are the closing pair of a multiloop

dynamic programming approach, that calculates the partition function for all subsequences $x[i,j]$ and ends with the partition function for the whole sequence...

Decomposition of paired/unpaired i,j

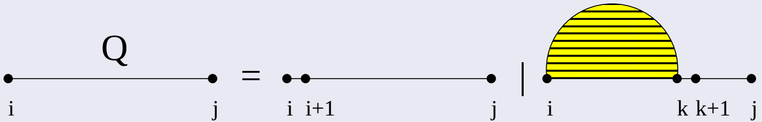


Decomposition of paired/unpaired i,j



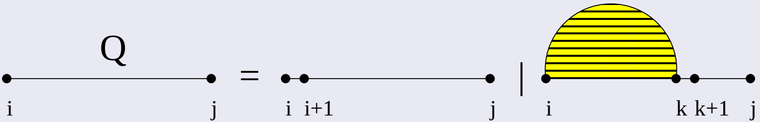
$$Q_{i,j} = 1 + \sum_{1<k<j} Q^b_{i,k} \cdot Q_{k+1,j} \tag{2}$$

Decomposition of paired/unpaired i,j



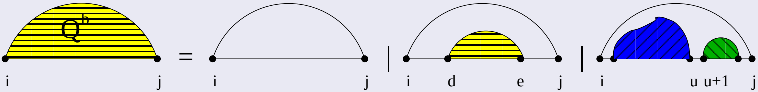
$$Q_{i,j} = 1 + \sum_{1<k<j} Q^{b}_{i,k} \cdot Q_{k+1,j} \quad (2)$$

Decomposition of paired/unpaired i,j

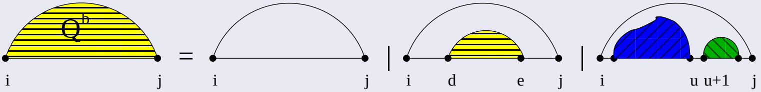


$$Q_{i,j} = 1 + \sum_{1<k<j} Q^{b}_{i,k} \cdot Q_{k+1,j} \tag{2}$$

Decomposition with constraint that i pair with j

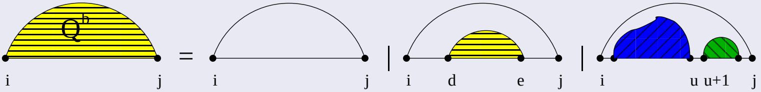


Decomposition with constraint that i pair with j



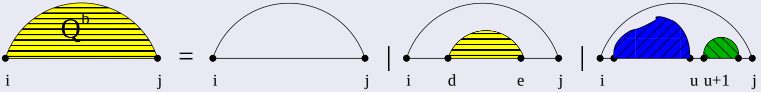
$$Q^b_{i,j} = e^{\frac{-G^{H_{i,j}}}{RT}} + \sum_{i<d<e<j} e^{\frac{-G^{I_{i,d,e,j}}}{RT}} \cdot Q^b_{d,e} + \sum_{i<u<j} Q^m_{i+1,u} \cdot Q^{m1}_{u+1,j-1} \cdot e^{\frac{-a}{RT}} \quad (3)$$

Decomposition with constraint that i pair with j



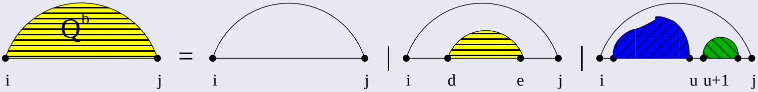
$$Q^b_{i,j} = e^{\frac{-G^{H_{i,j}}}{RT}} + \sum_{i<d<e<j} e^{\frac{-G^{I_{i,d,e,j}}}{RT}} \cdot Q^b_{d,e} + \sum_{i<u<j} Q^m_{i+1,u} \cdot Q^{m1}_{u+1,j-1} \cdot e^{\frac{-a}{RT}} \quad (3)$$

Decomposition with constraint that i pair with j



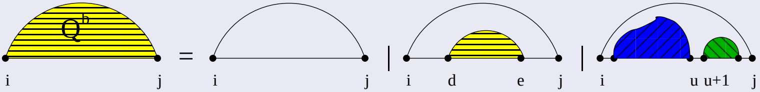
$$Q^b_{i,j} = e^{\frac{-G^{H_{i,j}}}{RT}} + \sum_{i<d<e<j} e^{\frac{-G^{I_{i,d,e,j}}}{RT}} \cdot Q^b_{d,e} + \sum_{i<u<j} Q^m_{i+1,u} \cdot Q^{m1}_{u+1,j-1} \cdot e^{\frac{-a}{RT}} \quad (3)$$

Decomposition with constraint that i pair with j



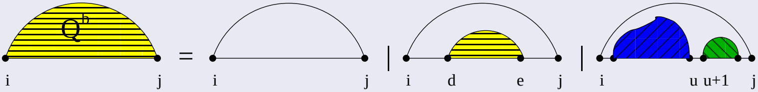
$$Q^b_{i,j} = e^{\frac{-G^{H_{i,j}}}{RT}} + \sum_{i<d<e<j} e^{\frac{-G^{I_{i,d,e,j}}}{RT}} \cdot Q^b_{d,e} + \sum_{i<u<j} Q^m_{i+1,u} \cdot Q^{m1}_{u+1,j-1} \cdot e^{\frac{-a}{RT}} \tag{3}$$

Decomposition with constraint that i pair with j



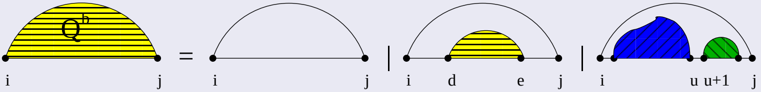
$$Q^b_{i,j} = e^{\frac{-G^{H_{i,j}}}{RT}} + \sum_{i<d<e<j} e^{\frac{-G^{I_{i,d,e,j}}}{RT}} \cdot Q^b_{d,e} + \sum_{i<u<j} Q^m_{i+1,u} \cdot Q^{m1}_{u+1,j-1} \cdot e^{\frac{-a}{RT}} \tag{3}$$

Decomposition with constraint that i pair with j



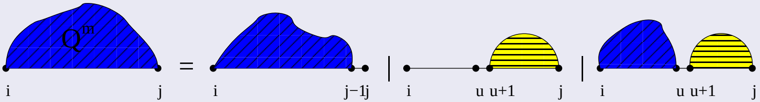
$$Q^b_{i,j} = e^{\frac{-G^{H_{i,j}}}{RT}} + \sum_{i<d<e<j} e^{\frac{-G^{I_{i,d,e,j}}}{RT}} \cdot Q^b_{d,e} + \sum_{i<u<j} Q^m_{i+1,u} \cdot Q^{m1}_{u+1,j-1} \cdot e^{\frac{-a}{RT}} \quad (3)$$

Decomposition with constraint that i pair with j

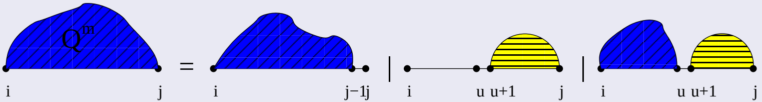


$$Q^b_{i,j} = e^{\frac{-G^{H_{i,j}}}{RT}} + \sum_{i<d<e<j} e^{\frac{-G^{I_{i,d,e,j}}}{RT}} \cdot Q^b_{d,e} + \sum_{i<u<j} Q^m_{i+1,u} \cdot Q^{m1}_{u+1,j-1} \cdot e^{\frac{-a}{RT}} \quad (3)$$

Decomposition of multiloop

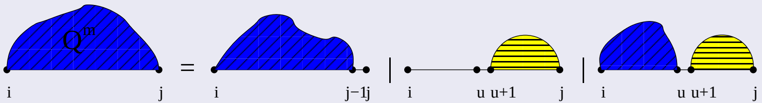


Decomposition of multiloop



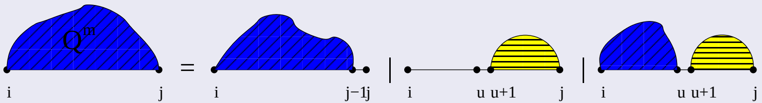
$$Q^m_{i,j} = Q^m_{i,j-1} \cdot e^{\frac{-c}{RT}} + \sum_{i<u<j} e^{\frac{-(u-i-1)\cdot c - b}{RT}} \cdot Q^b_{u+1,j} + \sum_{i<u<j} Q^m_{i,u} \cdot Q^b_{u+1,j} \cdot e^{\frac{-b}{RT}} \quad (4)$$

Decomposition of multiloop



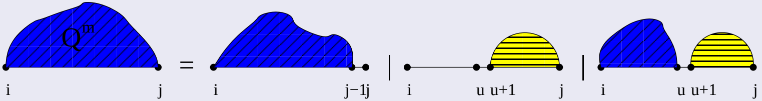
$$Q^m_{i,j} = Q^m_{i,j-1} \cdot e^{\frac{-c}{RT}} + \sum_{i<u<j} e^{\frac{-(u-i-1)\cdot c - b}{RT}} \cdot Q^b_{u+1,j} + \sum_{i<u<j} Q^m_{i,u} \cdot Q^b_{u+1,j} \cdot e^{\frac{-b}{RT}} \qquad (4)$$

Decomposition of multiloop



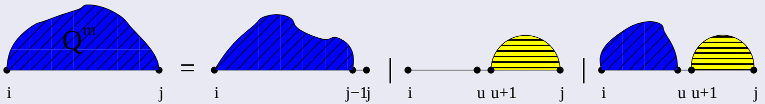
$$
\begin{aligned}
Q^{m}_{i,j} \quad = \quad & Q^{m}_{i,j-1} \cdot e^{\frac{-c}{RT}} + \\
& \sum_{i<u<j} e^{\frac{-(u-i-1)\cdot c - b}{RT}} \cdot Q^{b}_{u+1,j} + \\
& \sum_{i<u<j} Q^{m}_{i,u} \cdot Q^{b}_{u+1,j} \cdot e^{\frac{-b}{RT}}
\end{aligned}
\tag{4}
$$

Decomposition of multiloop



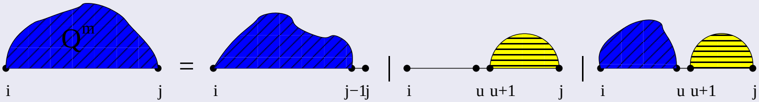
$$Q^m_{i,j} = Q^m_{i,j-1} \cdot e^{\frac{-c}{RT}} + \sum_{i<u<j} e^{\frac{-(u-i-1)\cdot c-b}{RT}} \cdot Q^b_{u+1,j} + \sum_{i<u<j} Q^m_{i,u} \cdot Q^b_{u+1,j} \cdot e^{\frac{-b}{RT}} \quad (4)$$

Decomposition of multiloop



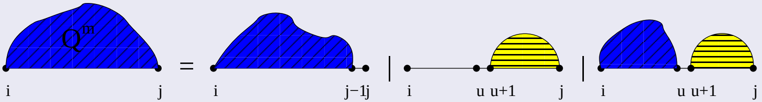
$$Q^m_{i,j} = {\color{blue}Q^m_{i,j-1}} \cdot e^{\frac{-c}{RT}} + \sum_{i<u<j} e^{\frac{-(u-i-1)\cdot c-b}{RT}} \cdot {\color{red}Q^b_{u+1,j}} + \sum_{i<u<j} Q^m_{i,u} \cdot Q^b_{u+1,j} \cdot e^{\frac{-b}{RT}} \quad (4)$$

Decomposition of multiloop



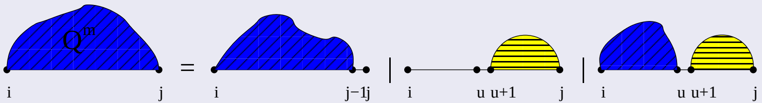
$$
\begin{aligned}
Q^m_{i,j} \quad &= \quad Q^m_{i,j-1} \cdot e^{\frac{-c}{RT}} + \\
&\sum_{i<u<j} e^{\frac{-(u-i-1)\cdot c - b}{RT}} \cdot Q^b_{u+1,j} + \\
&\sum_{i<u<j} Q^m_{i,u} \cdot Q^b_{u+1,j} \cdot e^{\frac{-b}{RT}}
\end{aligned} \tag{4}
$$

Decomposition of multiloop



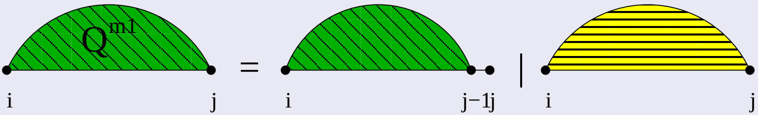
$$Q^m_{i,j} = Q^m_{i,j-1} \cdot e^{\frac{-c}{RT}} + \sum_{i<u<j} e^{\frac{-(u-i-1)\cdot c - b}{RT}} \cdot Q^b_{u+1,j} + \sum_{i<u<j} Q^m_{i,u} \cdot Q^b_{u+1,j} \cdot e^{\frac{-b}{RT}} \quad (4)$$

Decomposition of multiloop

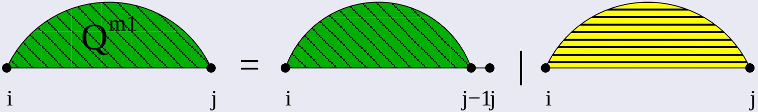


$$
\begin{aligned}
Q^m_{i,j} \quad = \quad & Q^m_{i,j-1} \cdot e^{\frac{-c}{RT}} + \\
& \sum_{i<u<j} e^{\frac{-(u-i-1)\cdot c-b}{RT}} \cdot Q^b_{u+1,j} + \\
& \sum_{i<u<j} Q^m_{i,u} \cdot Q^b_{u+1,j} \cdot e^{\frac{-b}{RT}}
\end{aligned} \tag{4}
$$

Decomposition of multiloop part with exactly one component

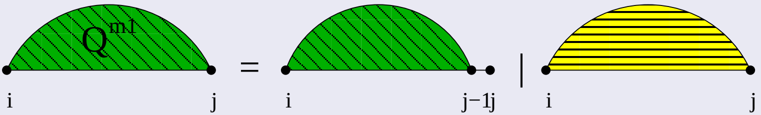


Decomposition of multiloop part with exactly one component



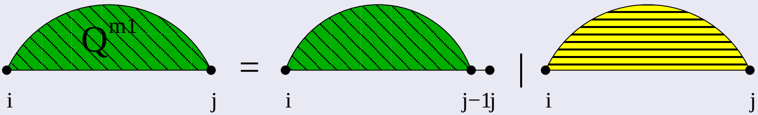
$$Q^{m1}_{i,j} = Q^{m1}_{i,j-1} \cdot e^{\frac{-c}{RT}} + Q^{b}_{i,j} \cdot e^{\frac{-b}{RT}} \tag{5}$$

Decomposition of multiloop part with exactly one component



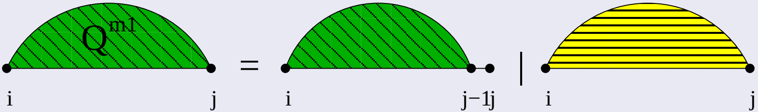
$$Q^{m1}_{i,j} = Q^{m1}_{i,j-1} \cdot e^{\frac{-c}{RT}} + Q^{b}_{i,j} \cdot e^{\frac{-b}{RT}} \quad (5)$$

Decomposition of multiloop part with exactly one component



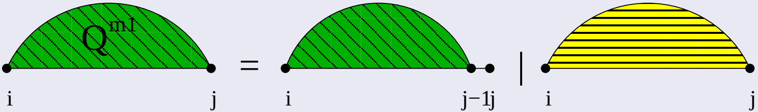
$$Q^{m1}_{i,j} = Q^{m1}_{i,j-1} \cdot e^{\frac{-c}{RT}} + Q^{b}_{i,j} \cdot e^{\frac{-b}{RT}} \quad (5)$$

Decomposition of multiloop part with exactly one component



$$Q^{m1}_{i,j} = Q^{m1}_{i,j-1} \cdot e^{\frac{-c}{RT}} + Q^{b}_{i,j} \cdot e^{\frac{-b}{RT}} \quad (5)$$

Decomposition of multiloop part with exactly one component



$$Q^{m1}_{i,j} = Q^{m1}_{i,j-1} \cdot e^{\frac{-c}{RT}} + Q^{b}_{i,j} \cdot e^{\frac{-b}{RT}} \quad (5)$$

Similarities to linear case:

Similarities to linear case:

- Q is also computed by previously computing $Q_{i,j}$ for all subsequences

Similarities to linear case:

- Q is also computed by previously computing $Q_{i,j}$ for all subsequences
- structure decomposition for $Q_{i,j} \forall 1 \leq i < j \leq n$

Similarities to linear case:

- Q is also computed by previously computing $Q_{i,j}$ for all subsequences
- structure decomposition for $Q_{i,j} \forall 1 \leq i < j \leq n$

Difference:

Similarities to linear case:

- Q is also computed by previously computing $Q_{i,j}$ for all subsequences
- structure decomposition for $Q_{i,j} \forall 1 \leq i < j \leq n$

Difference:

- computing sum of weighted energies for all loops that contain bases x_n and x_1

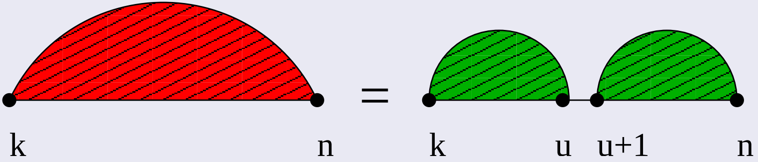
Similarities to linear case:

- Q is also computed by previously computing $Q_{i,j}$ for all subsequences
- structure decomposition for $Q_{i,j} \forall 1 \leq i < j \leq n$

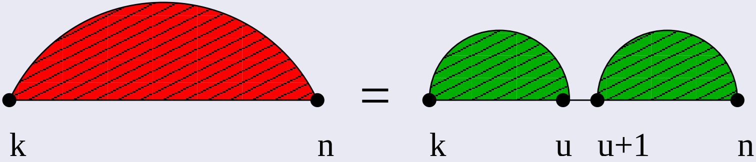
Difference:

- computing sum of weighted energies for all loops that contain bases x_n and x_1
- leads to the computation of so called ”exterior loops”

Computation of M^2 Array

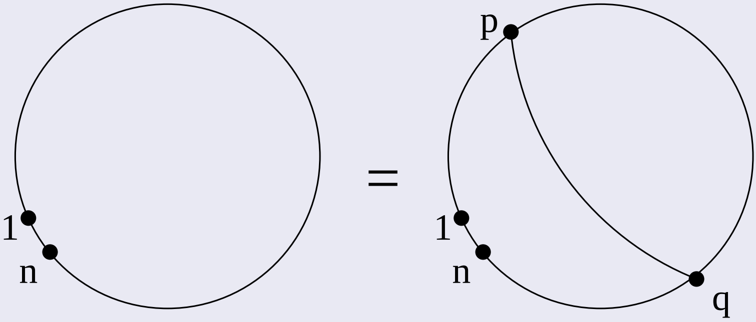


Computation of M^2 Array

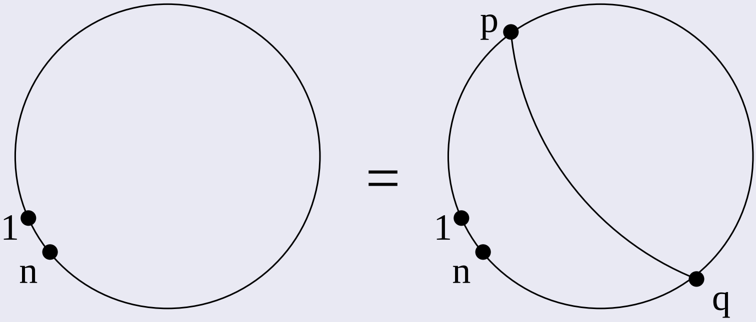


$$Q^{M2}_{k,n} = \sum_{k<u<n} Q^{m1}_{k,u} \cdot Q^{m1}_{u+1,n} \tag{6}$$

Computation of exterior hairpin loop

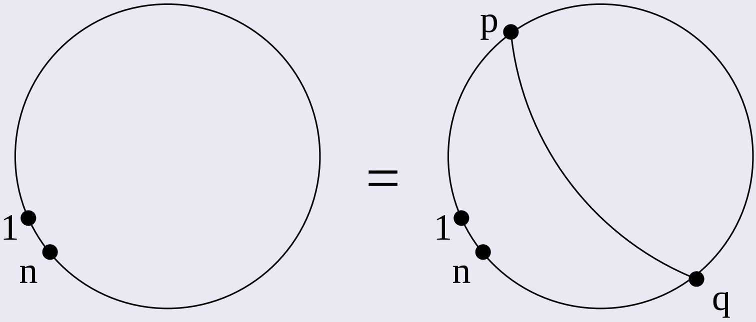


Computation of exterior hairpin loop



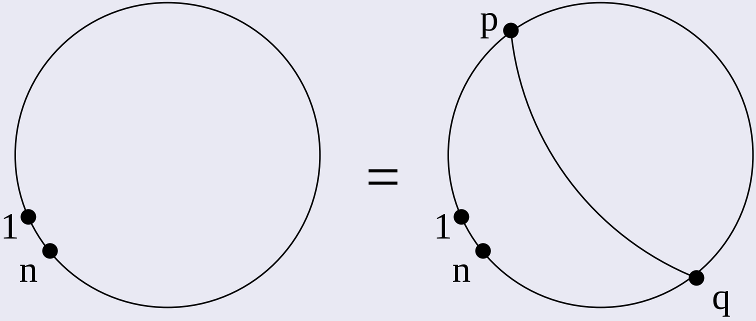
$$Q^o_H = \sum_{p<q} Q^b_{p,q} \cdot e^{\frac{-G^{H_{q,p}}}{RT}} \qquad (7)$$

Computation of exterior hairpin loop



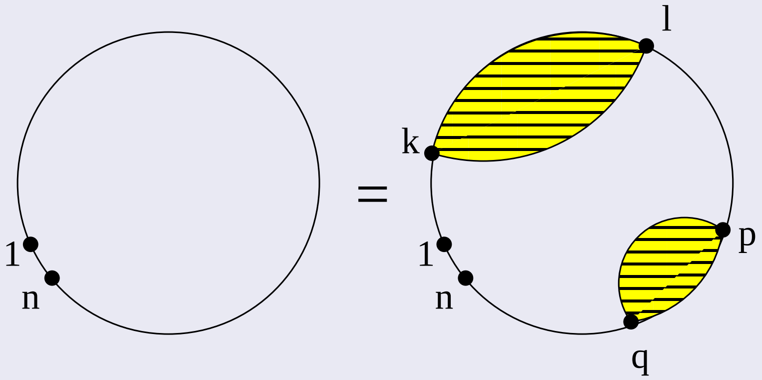
$$Q^o_H = \sum_{p<q} Q^b_{p,q} \cdot e^{\frac{-G^{H_{q,p}}}{RT}} \qquad (7)$$

Computation of exterior hairpin loop

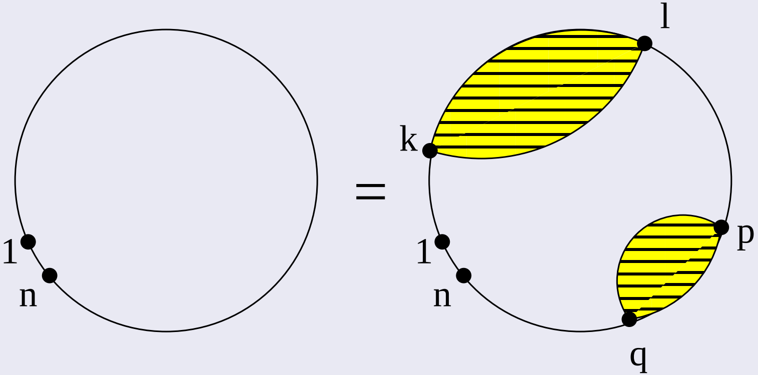


$$Q^o_H = \sum_{p<q} Q^b_{p,q} \cdot e^{\frac{-G^{H_{q,p}}}{RT}} \qquad (7)$$

Computation of exterior interior loop

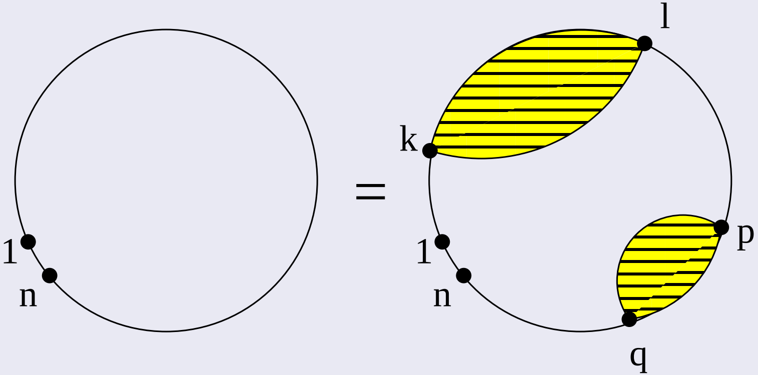


Computation of exterior interior loop



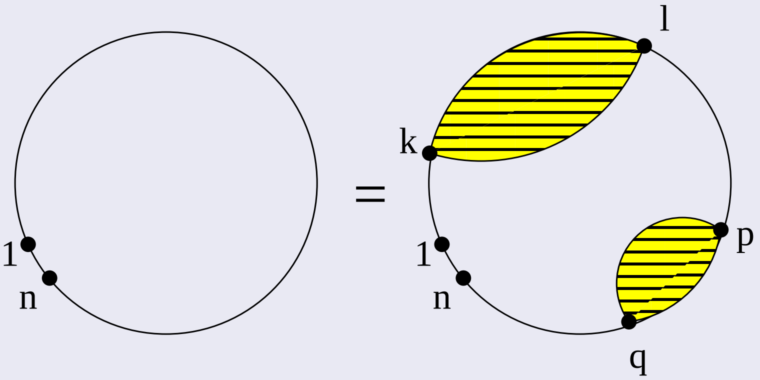
$$Q_I^o = \sum_{k<l<p<q} Q_{k,l}^b \cdot Q_{p,q}^b \cdot e^{\frac{-G^I k,l,p,q}{RT}} \quad (8)$$

Computation of exterior interior loop



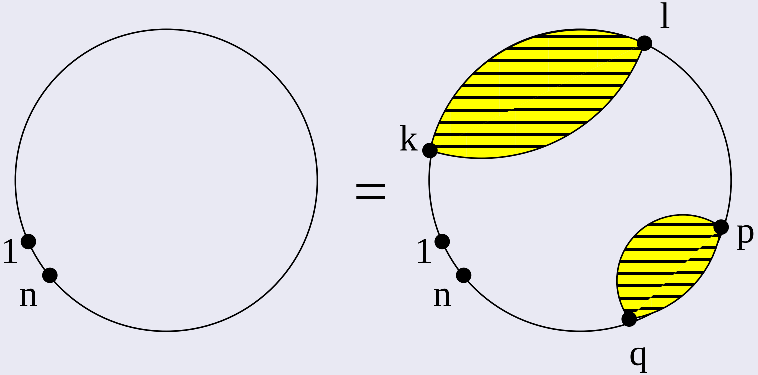
$$Q_I^o = \sum_{k<l<p<q} Q_{k,l}^b \cdot Q_{p,q}^b \cdot e^{\frac{-G^I{}_{k,l,p,q}}{RT}} \quad (8)$$

Computation of exterior interior loop



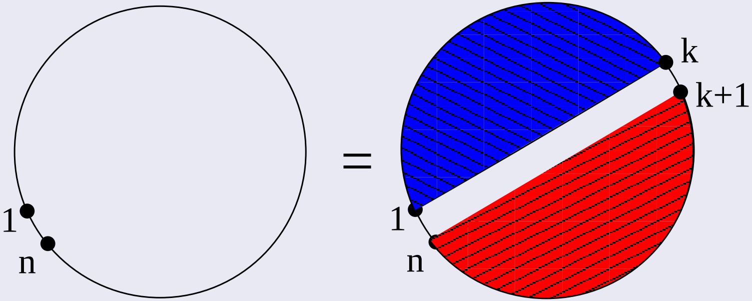
$$Q^o_I = \sum_{k<l<p<q} Q^b_{k,l} \cdot Q^b_{p,q} \cdot e^{\frac{-G^{I_{k,l,p,q}}}{RT}} \quad (8)$$

Computation of exterior interior loop

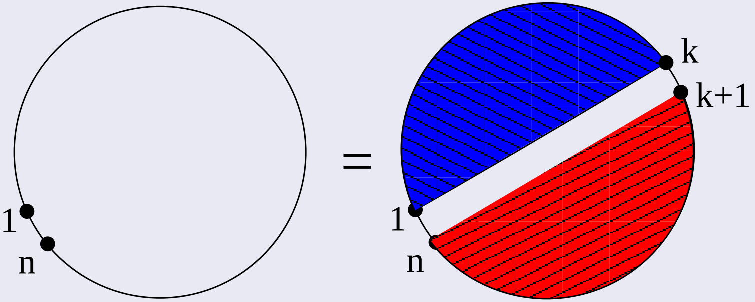


$$Q^o_I = \sum_{k<l<p<q} Q^b_{k,l} \cdot Q^b_{p,q} \cdot e^{\frac{-G^{I}{k,l,p,q}}{RT}} \quad (8)$$

Computation of exterior multiloop

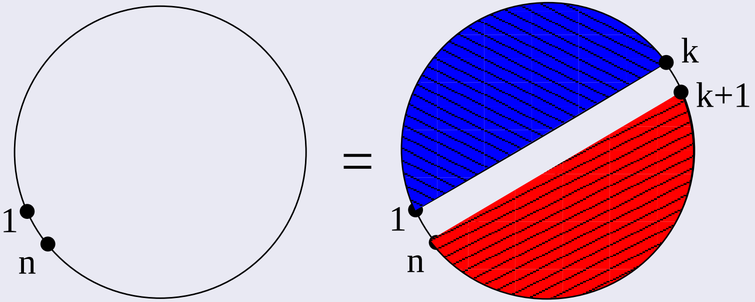


Computation of exterior multiloop



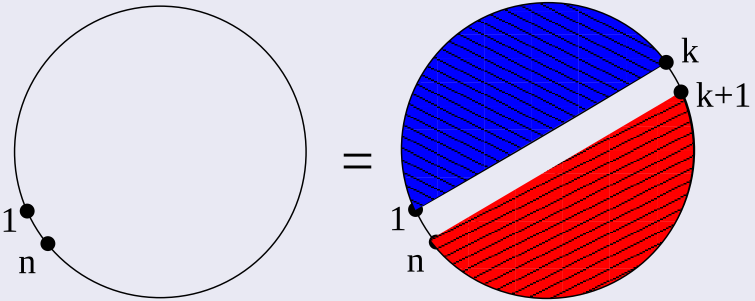
$$Q^o_M = \sum_{1<k<n} Q^m_{1,k} \cdot Q^{M2}_{k+1,n} \cdot e^{\frac{-a}{RT}} \qquad (9)$$

Computation of exterior multiloop



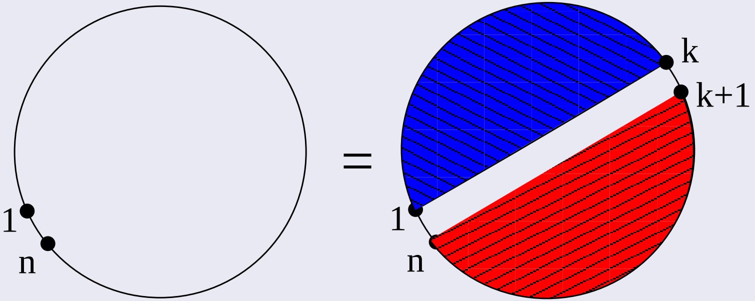
$$Q^o_M = \sum_{1<k<n} Q^m_{1,k} \cdot Q^{M2}_{k+1,n} \cdot e^{\frac{-a}{RT}} \quad (9)$$

Computation of exterior multiloop



$$Q^{o}_{M} = \sum_{1<k<n} Q^{m}_{1,k} \cdot Q^{M2}_{k+1,n} \cdot e^{\frac{-a}{RT}} \qquad (9)$$

Computation of exterior multiloop



$$Q^o_M = \sum_{1<k<n} Q^m_{1,k} \cdot Q^{M2}_{k+1,n} \cdot e^{\frac{-a}{RT}} \quad (9)$$

Partition function for circular case:

$$Q^o = Q^o_H + Q^o_I + Q^o_M \tag{10}$$

This is the step where i am actually in my diploma thesis

This is the step where i am actually in my diploma thesis
I will skip

This is the step where i am actually in my diploma thesis
I will skip

- equations for base pairing probability matrix

This is the step where i am actually in my diploma thesis
I will skip

- equations for base pairing probability matrix
- way of backtracking to obtain secondary structures

Most known RNAs are linear and circular RNAs are not found very often

But they are found in:

Most known RNAs are linear and circular RNAs are not found very often

But they are found in:

- Viroids and satellite RNAs (sizes arround 300-400 nt)

Most known RNAs are linear and circular RNAs are not found very often

But they are found in:

- Viroids and satellite RNAs (sizes arround 300-400 nt)
- Genome of Hepatitis δ virus

Most known RNAs are linear and circular RNAs are not found very often

But they are found in:

- Viroids and satellite RNAs (sizes arround 300-400 nt)
- Genome of Hepatitis δ virus
- spliced Group I Introns

Most known RNAs are linear and circular RNAs are not found very often
But they are found in:

- Viroids and satellite RNAs (sizes arround 300-400 nt)
- Genome of Hepatitis δ virus
- spliced Group I Introns
- Box C/D RNAs in Pyrococcus furiosus

Most known RNAs are linear and circular RNAs are not found very often

But they are found in:

- Viroids and satellite RNAs (sizes arround 300-400 nt)
- Genome of Hepatitis δ virus
- spliced Group I Introns
- Box C/D RNAs in Pyrococcus furiosus
- other

Most known RNAs are linear and circular RNAs are not found very often

But they are found in:

- Viroids and satellite RNAs (sizes arround 300-400 nt)
- Genome of Hepatitis δ virus
- spliced Group I Introns
- Box C/D RNAs in Pyrococcus furiosus
- other

In almost all cases, circular RNAs code not for any mRNA but for their own structure

What can be the aim for application of CPF?

What can be the aim for application of CPF?

- finding metastable structure states

What can be the aim for application of CPF?

- finding metastable structure states
- better prediction of circular RNA structures

What can be the aim for application of CPF?

- finding metastable structure states
- better prediction of circular RNA structures
- other things I did not have thought about yet

Thank you for your attention