Computational Biology

Ribonucleic Acid (RNA)

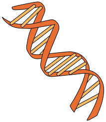
Marc Hellmuth

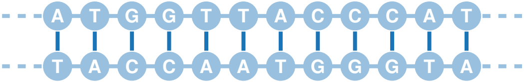
Department of Mathematics
Stockholm University

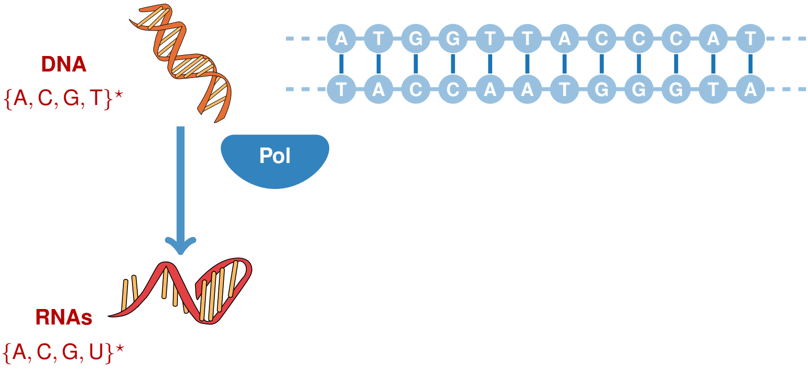
With kind permission of Yann Ponty and Sebastian Will (l'Ecole Polytechnique LIX), I reuse for the RNA-part many of their slides of the AMI2B course.

DNA

$\{A, C, G, T\}^\star$

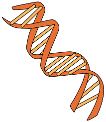


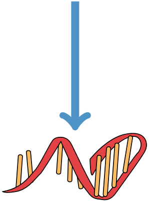



DNA
{A, C, G, T}*
Pol
RNAs
{A, C, G, U}*
A T G G T T A C C C A T
T A C C A A T G G G T A

DNA

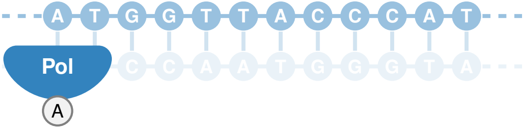
{A, C, G, T}*





RNAs

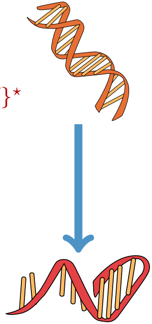
{A, C, G, U}*



Fundamental *dogma* of molecular biology

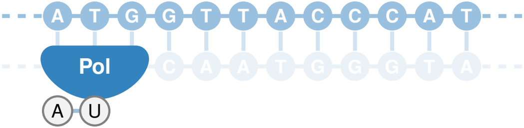
DNA

$\{A, C, G, T\}^\star$



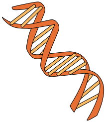
RNAs

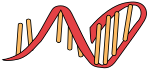
$\{A, C, G, U\}^\star$



DNA

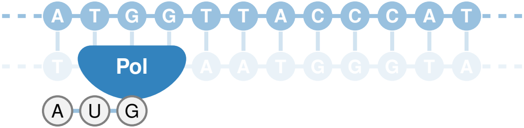
$\{A, C, G, T\}^\star$





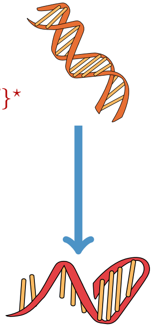
RNAs

$\{A, C, G, U\}^\star$



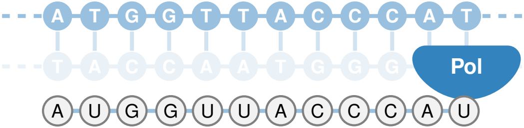
DNA

$\{A, C, G, T\}^\star$



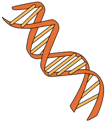
RNAs

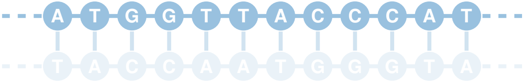
$\{A, C, G, U\}^\star$



DNA

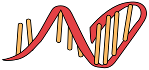
$\{A, C, G, T\}^\star$

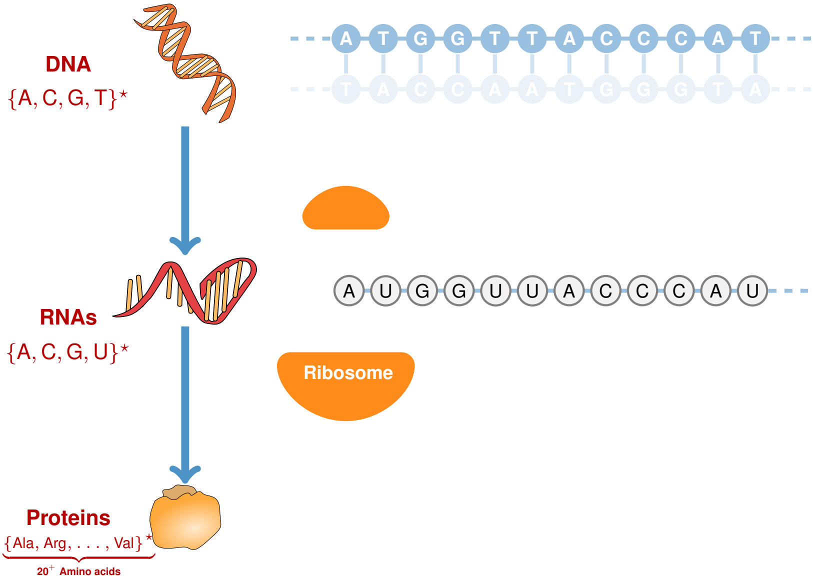


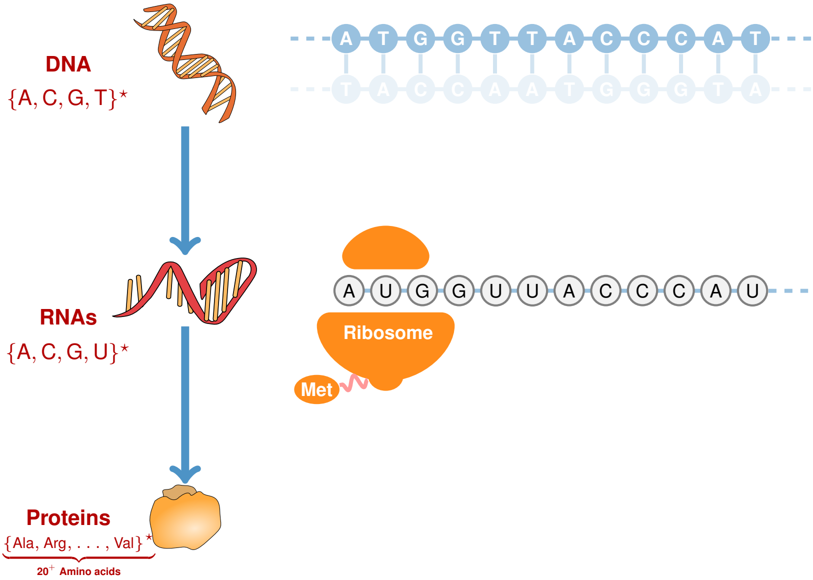


RNAs

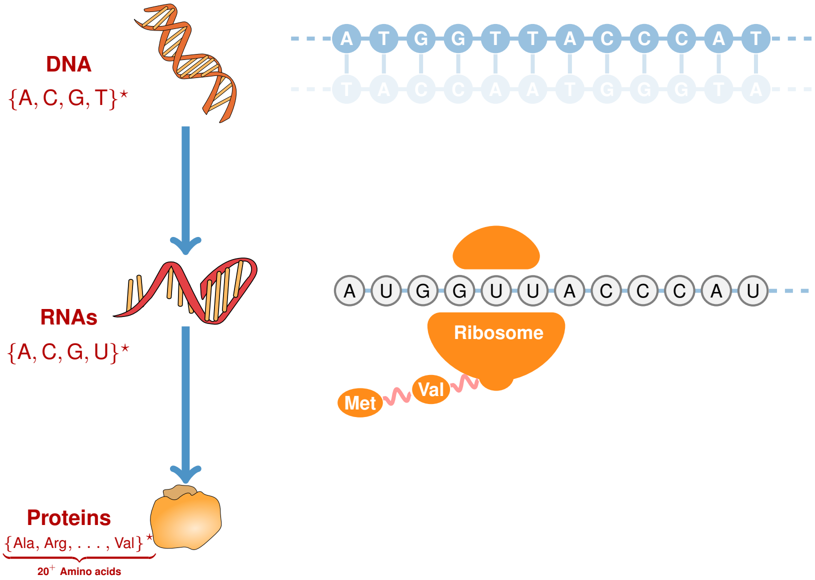
$\{A, C, G, U\}^\star$



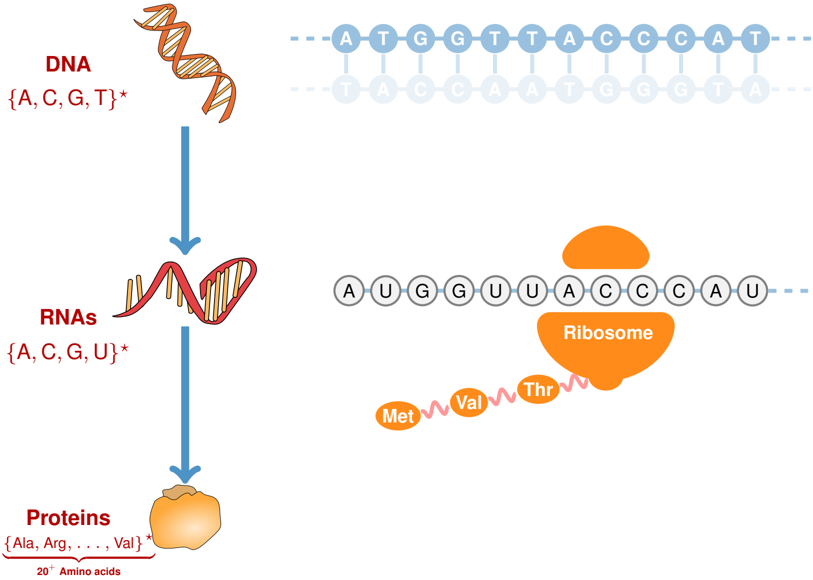

DNA
{A, C, G, T}*
A T G G T T A C C C A T
T A C C A A T G G G T A
RNAs
{A, C, G, U}*
A U G G U U A C C C A U
Ribosome
Proteins
{Ala, Arg, . . . , Val}*
20+ Amino acids


DNA
{A, C, G, T}*
A T G G T T A C C C A T
T A C C A A T G G G T A
RNAs
{A, C, G, U}*
A U G G U U A C C C A U
Ribosome
Met
Proteins
{Ala, Arg, . . . , Val}*
20+ Amino acids

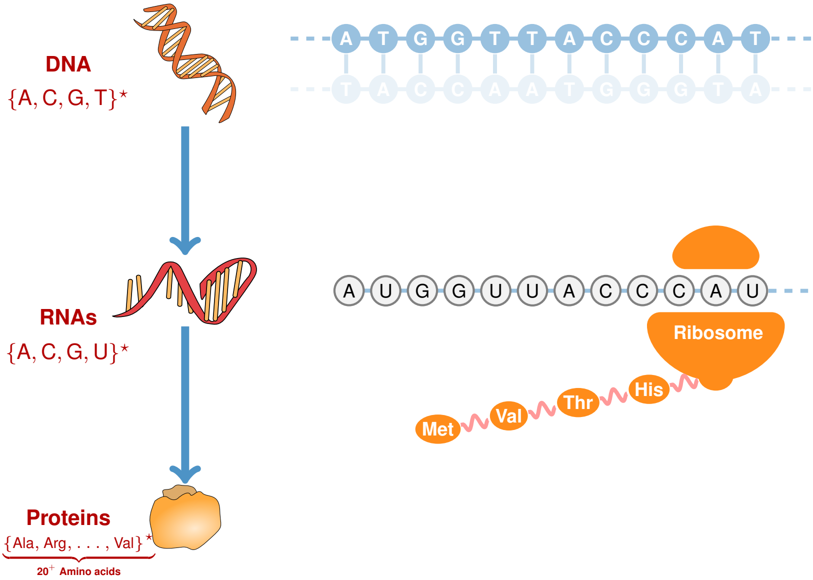
Fundamental *dogma* of molecular biology

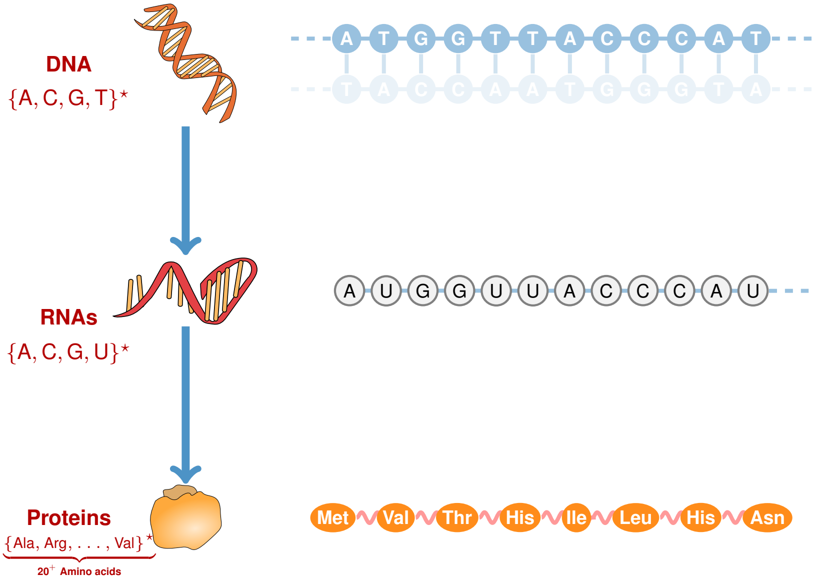


Fundamental *dogma* of molecular biology

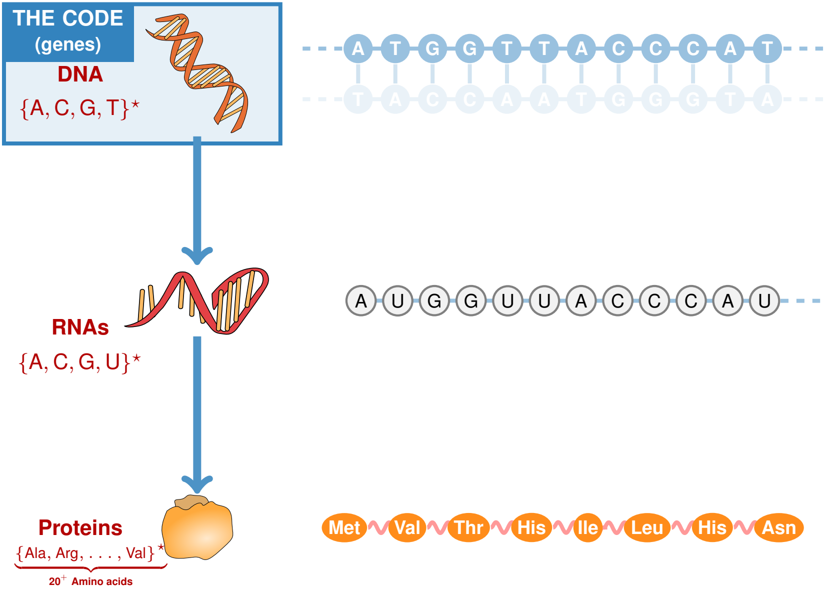


Fundamental *dogma* of molecular biology

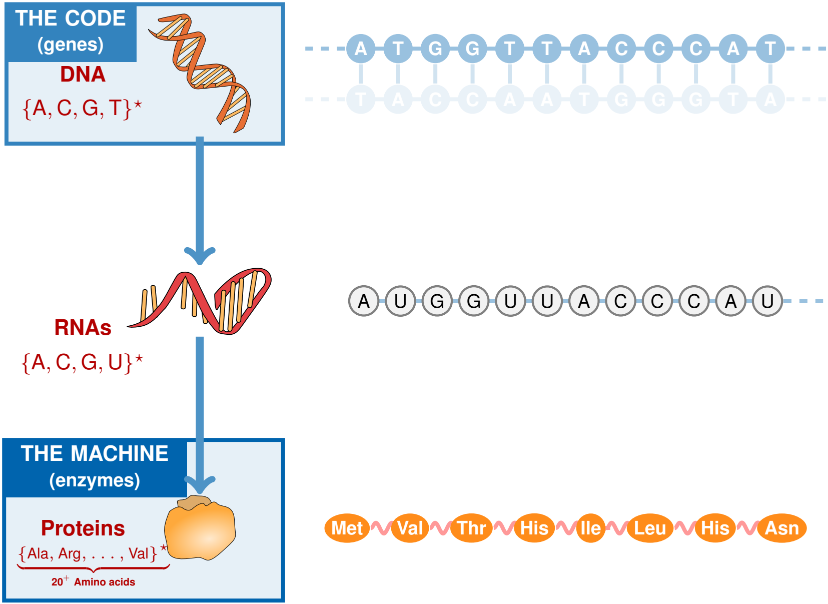



DNA
{A, C, G, T}*
A T G G T T A C C C A T
T A C C A A T G G G T A
RNAs
{A, C, G, U}*
A U G G U U A C C C A U
Proteins
{Ala, Arg, . . . , Val}*
20+ Amino acids
Met Val Thr His Ile Leu His Asn

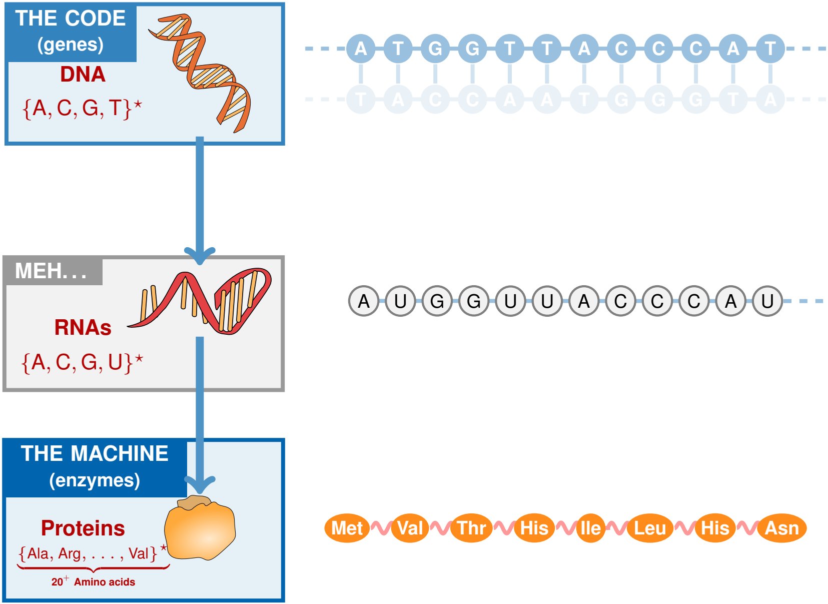
Fundamental *dogma* of molecular biology

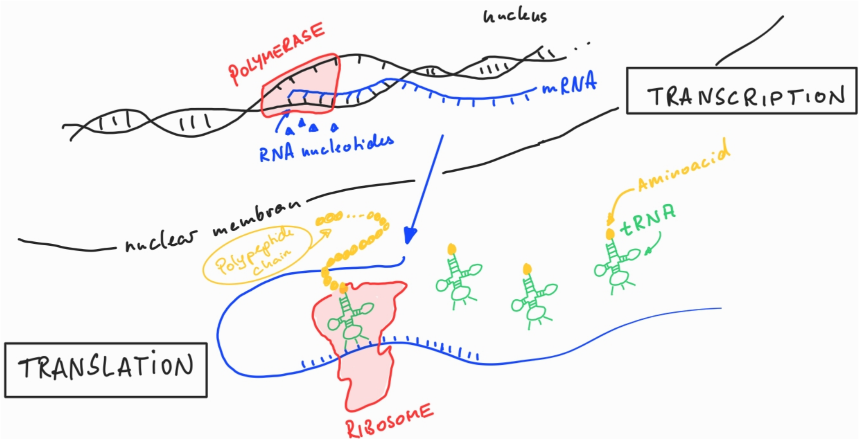


Fundamental *dogma* of molecular biology



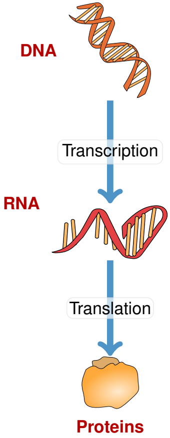
Fundamental *dogma* of molecular biology

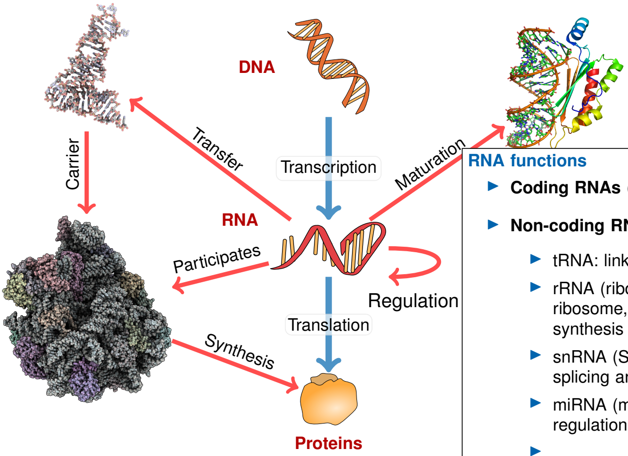




Oh! There are RNAs that are not only messengers . . .

great visualization: `https://www.youtube.com/watch?v=gG7uCskUOrA`


DNA
Transcription
RNA
Translation
Proteins

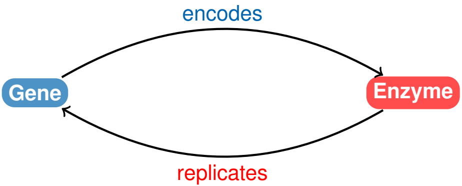


RNA functions

- **Coding RNAs** (Translation - mRNA)
- **Non-coding RNAs**
  - tRNA: linking codons to aminoacids
  - rRNA (ribosomalRNA): part of of the ribosome, essential for protein synthesis
  - snRNA (Small nuclear RNA, $\sim$ 150 nt): splicing and other functions
  - miRNA (microRNA, 21-22 nt): regulation of gene expression
  - ...

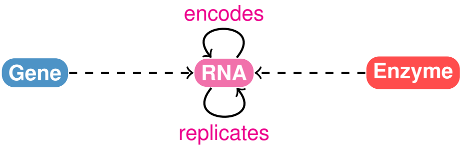
$\sim$ 97% **are non-coding RNAs in eukaryotes**

RNA can even act as genome (virus)



A gene big enough to specify an enzyme would be too big to replicate accurately without the aid of an enzyme of the very kind that it is trying to specify. So the system *apparently cannot get started*.

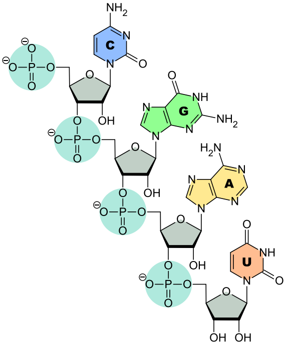
RNA - WORLD - HYPOTHESIS:
self-replicating RNA molecules are precursors to all current life on earth.



A gene big enough to specify an enzyme would be too big to replicate accurately without the aid of an enzyme of the very kind that it is trying to specify. So the system *apparently cannot get started*.

RNA - WORLD - HYPOTHESIS:
self-replicating RNA molecules are precursors to all current life on earth.

- single-stranded polymer
- polymer made of nucleotides+backbone
- nucleotides: guanine (G), adenine (A), uracil (U), cytosine (C)
- backbone: alternating sugar (ribose) and phospat groups (related to phosphoric acid) nucleotides are attached to sugar
- the nucleotides of polymer can bind (A-U, C-G, G-U) via hydrogen bonds, i.e., unlike DNA it is more often found in nature as a single-strand folded unto itself, rather than a paired double-strand.

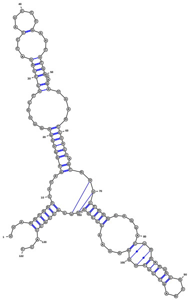


RNA can fold into complex 3D structures that are essential to its function(s).

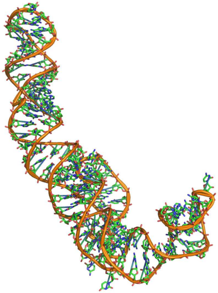
Three* levels of representation:

UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAAGAUAAGCC
CACCAGCGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC

Primary structure



Secondary structure



Tertiary structure

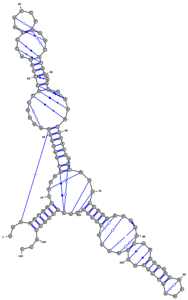
Source: 5s rRNA (PDB 1K73:B)

*Well, mostly...

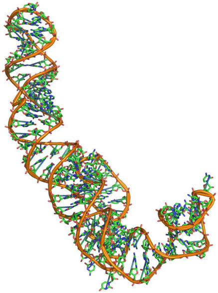
Three* levels of representation:

UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAAGAUAAGCC
CACCAGCGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA
CC

Primary structure



Secondary+ structure



Tertiary structure

Source: 5s rRNA (PDB 1K73:B)

*Well, mostly...

$\mathbb{A} := \{A, C, G, U\}$ and $\mathbb{B} := \{AU, UA, GC, CG, GU, UG\}$

A **primary structure** (of length n) is a sequence $s = s_1 \dots s_n \in \mathbb{A}^n$.

A **secondary structure** $\mathscr{S}$ is a collection of ordered pairs (i, j), where $1 \leq i < j \leq n$, s.t. the following properties hold:

1. If $(i, j), (k, l) \in \mathscr{S}$, then it is not the case that $i < k < j < l$.
2. If $(i, j), (k, l) \in \mathscr{S}$ and $i \in (k, l)$ implies that $i = k$ and $j = l$.
3. If $(i, j) \in \mathscr{S}$, then $j > i + \theta$, where θ is a fixed integer and usually taken to be 3.

A secondary structure $\mathscr{S}$ for a given sequence $s = s_1 \dots s_n \in \mathbb{A}^n$ is a secondary structure fullfilling in addition

4. If $(i, j) \in \mathscr{S}$, then $s_i s_j \in \mathbb{B}$.

If item 4. is fullfilled, then we say that the sequence $s \in \mathbb{A}^n$ realizes $\mathscr{S}$.

A **tertiary structure** is basically the 3D structure, i.e., refers to locations of the atoms in three-dimensional space.

$\mathbb{A} := \{A, C, G, U\}$ and $\mathbb{B} := \{AU, UA, GC, CG, GU, UG\}$

A **primary structure** (of length n) is a sequence $s = s_1 \dots s_n \in \mathbb{A}^n$.

A **secondary structure** $\mathscr{S}$ is a collection of ordered pairs (i, j), where $1 \leq i < j \leq n$, s.t. the following properties hold:

1. If $(i, j), (k, l) \in \mathscr{S}$, then it is not the case that $i < k < j < l$.
2. If $(i, j), (k, l) \in \mathscr{S}$ and $i \in (k, l)$ implies that $i = k$ and $j = l$.
3. If $(i, j) \in \mathscr{S}$, then $j > i + \theta$, where θ is a fixed integer and usually taken to be 3.

A secondary structure $\mathscr{S}$ for a given sequence $s = s_1 \dots s_n \in \mathbb{A}^n$ is a secondary structure fullfilling in addition

4. If $(i, j) \in \mathscr{S}$, then $s_i s_j \in \mathbb{B}$.

If item 4. is fullfilled, then we say that the sequence $s \in \mathbb{A}^n$ **realizes** $\mathscr{S}$.

A **tertiary structure** is basically the 3D structure, i.e., refers to locations of the atoms in three-dimensional space.

$\mathbb{A} := \{A, C, G, U\}$ and $\mathbb{B} := \{AU, UA, GC, CG, GU, UG\}$

A **primary structure** (of length n) is a sequence $s = s_1 \dots s_n \in \mathbb{A}^n$.

A **secondary structure** $\mathscr{S}$ is a collection of ordered pairs (i, j), where $1 \leq i < j \leq n$, s.t. the following properties hold:

1. If $(i, j), (k, l) \in \mathscr{S}$, then it is not the case that $i < k < j < l$.
2. If $(i, j), (k, l) \in \mathscr{S}$ and $i \in (k, l)$ implies that $i = k$ and $j = l$.
3. If $(i, j) \in \mathscr{S}$, then $j > i + \theta$, where θ is a fixed integer and usually taken to be 3.

A secondary structure $\mathscr{S}$ for a given sequence $s = s_1 \dots s_n \in \mathbb{A}^n$ is a secondary structure fullfilling in addition

4. If $(i, j) \in \mathscr{S}$, then $s_i s_j \in \mathbb{B}$.

If item 4. is fullfilled, then we say that the sequence $s \in \mathbb{A}^n$ realizes $\mathscr{S}$.

A **tertiary structure** is basically the 3D structure, i.e., refers to locations of the atoms in three-dimensional space.

$\mathbb{A} := \{A, C, G, U\}$ and $\mathbb{B} := \{AU, UA, GC, CG, GU, UG\}$

A **primary structure** (of length n) is a sequence $s = s_1 \dots s_n \in \mathbb{A}^n$.

A **secondary structure** $\mathscr{S}$ is a collection of ordered pairs (i, j), where $1 \leq i < j \leq n$, s.t. the following properties hold:

1. If $(i, j), (k, l) \in \mathscr{S}$, then it is not the case that $i < k < j < l$.
2. If $(i, j), (k, l) \in \mathscr{S}$ and $i \in (k, l)$ implies that $i = k$ and $j = l$.
3. If $(i, j) \in \mathscr{S}$, then $j > i + \theta$, where θ is a fixed integer and usually taken to be 3.

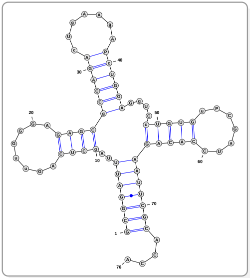
A secondary structure $\mathscr{S}$ for a given sequence $s = s_1 \dots s_n \in \mathbb{A}^n$ is a secondary structure fullfilling in addition

4. If $(i, j) \in \mathscr{S}$, then $s_i s_j \in \mathbb{B}$.

If item 4. is fullfilled, then we say that the sequence $s \in \mathbb{A}^n$ realizes $\mathscr{S}$.

A **tertiary structure** is basically the 3D structure, i.e., refers to locations of the atoms in three-dimensional space.

Various representations for a versatile biomolecule

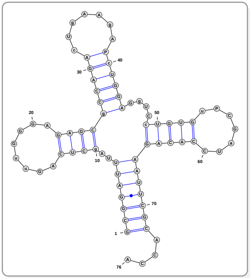


Outer-planar graphs
Hamiltonian-path,
$\Delta(G) \leq 3$, 2-connected*

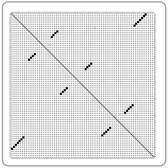
Supporting intuitions

Different representations
Common combinatorial structure

* Additional steric constraints



Outer-planar graphs
Hamiltonian-path,
$\Delta(G) \leq 3$, 2-connected*

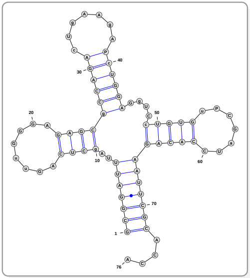


Dot plots
Adjacency matrices*

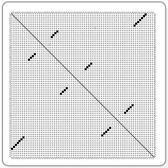
Supporting intuitions

Different representations
Common combinatorial structure

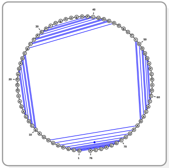
*Additional steric constraints



Outer-planar graphs
Hamiltonian-path,
$\Delta(G) \leq 3$, 2-connected⋆



Dot plots
Adjacency matrices⋆



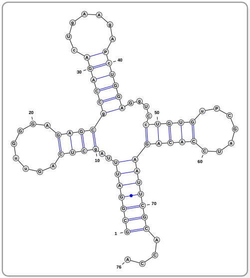
Non-crossing arc diagrams⋆

Supporting intuitions

Different representations

Common combinatorial structure

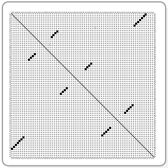
⋆Additional steric constraints



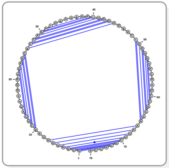
Outer-planar graphs
Hamiltonian-path,
$\Delta(G) \leq 3$, 2-connected*

```
(((((((..((((........))))((((((.......))))))....(((((.......))))))))))))....
```

Motzkin words*



Dot plots
Adjacency matrices*

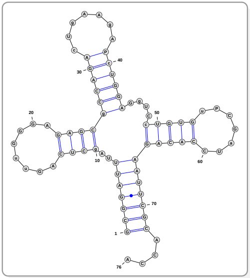


Non-crossing arc diagrams*

Supporting intuitions

Different representations
Common combinatorial structure

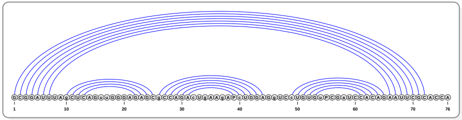
*Additional steric constraints



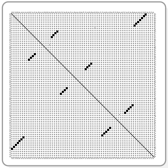
Outer-planar graphs
Hamiltonian-path,
$\Delta(G)\leq 3$, 2-connected*

(((((((..((((........))))((((((.......))))))....(((((.......))))))))))))....

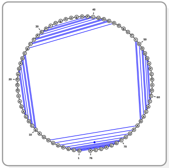
Motzkin words*



Non-crossing arc-annotated sequences*



Dot plots
Adjacency matrices*

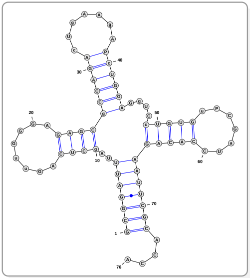


Non-crossing arc diagrams*

Supporting intuitions

Different representations
Common combinatorial structure

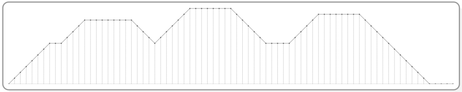
*Additional steric constraints



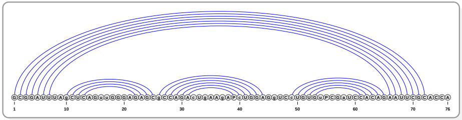
Outer-planar graphs
Hamiltonian-path,
$\Delta(G)\leq 3$, 2-connected*

(((((((..((((........))))((((((.......))))))....(((((.......))))))))))))....

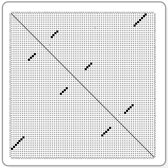
Motzkin words*



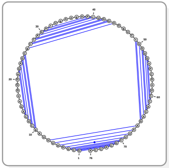
Positive 1D meanders* over $\mathcal{S} = \{+1, -1, 0\}$



Non-crossing arc-annotated sequences*



Dot plots
Adjacency matrices*



Non-crossing arc diagrams*

Supporting intuitions

Different representations
Common combinatorial structure

*Additional steric constraints

Theorem

Let $S(n)$ denote the number of secondary structures of size n and $\theta = 1$. Then $S(0) = 0$, $S(1) = 1$ and for $n \geq 1$,

$$S(n+1) = S(n) + S(n-1) + \sum_{k=2}^{n-1} S(k-1)S(n-k)$$

and

$$S(n) \geq 2^{n-2}.$$

Theorem

Let $S(n,k)$ denote the number of secondary structures of size n that contain exactly k basepairs ($\theta = 1$). Set $S(n,0) = 1$ for all n and $S(n,k) = 0$ for $k \geq \frac{n}{2}$. Then for $n \geq 2$,

$$S(n+1,k+1) = S(n,k+1) + \sum_{j=1}^{n-1} \left[\sum_{i=0}^{k} S(j-1,i)S(n-j,k-i) \right].$$

Corollary

$$S(n) = \sum_{k=0}^{\lfloor n/2 \rfloor} S(n,k).$$

All proofs on whiteboard

A sequence $s \in \mathbb{A}^n$ realizes or is compatible with a secondary structure $\mathscr{S}$ of length n if for any $(i, j) \in \mathscr{S}$ it holds that $s_i, s_j \in \mathbb{B}$.

- *dependency graph or shape graph* $G(\mathscr{S}_1, \ldots, \mathscr{S}_k)$

QUESTION:

On what conditions is it possible to find *one* sequence, which is realizing *all* sec.str. $\mathscr{S}_1, \ldots, \mathscr{S}_k$ of the same size?

equivalent to:

What properties have to be fulfilled in $G(\mathscr{S}_1, \ldots, \mathscr{S}_k)$, s.t. there exists a single sequence, which is realizing all $\mathscr{S}_1, \ldots, \mathscr{S}_k$?
$C(\mathscr{S})$ denotes the set of all sequences that realize Secondary Structure $\mathscr{S}$.

Theorem (Intersection Theorem, Reidys et al. 1995)

For any two secondary structures $\mathscr{S}_1$ and $\mathscr{S}_2$ of same size holds: $C(\mathscr{S}_1) \cap C(\mathscr{S}_2) \neq \emptyset$.

Theorem (Generalized Intersection Theorem, Flamm et al. 2001)

$\bigcap_{i=1}^{k} C(\mathscr{S}_i) \neq \emptyset \iff G(\mathscr{S}_1, \ldots, \mathscr{S}_k)$ *is bipartite.*

All proofs on whiteboard

A sequence $s \in \mathbb{A}^n$ realizes or is compatible with a secondary structure $\mathscr{S}$ of length n if for any $(i, j) \in \mathscr{S}$ it holds that $s_i, s_j \in \mathbb{B}$.

- *dependency graph or shape graph* $G(\mathscr{S}_1, \ldots, \mathscr{S}_k)$

QUESTION:

On what conditions is it possible to find *one* sequence, which is realizing *all* sec.str. $\mathscr{S}_1, \ldots, \mathscr{S}_k$ of the same size?

equivalent to:

What properties have to be fulfilled in $G(\mathscr{S}_1, \ldots, \mathscr{S}_k)$, s.t. there exists a single sequence, which is realizing all $\mathscr{S}_1, \ldots, \mathscr{S}_k$?
$C(\mathscr{S})$ denotes the set of all sequences that realize Secondary Structure $\mathscr{S}$.

Theorem (Intersection Theorem, Reidys et al. 1995)

For any two secondary structures $\mathscr{S}_1$ and $\mathscr{S}_2$ of same size holds: $C(\mathscr{S}_1) \cap C(\mathscr{S}_2) \neq \emptyset$.

Theorem (Generalized Intersection Theorem, Flamm et al. 2001)

$\bigcap_{i=1}^{k} C(\mathscr{S}_i) \neq \emptyset \iff G(\mathscr{S}_1, \ldots, \mathscr{S}_k)$ *is bipartite.*

All proofs on whiteboard

A sequence $s \in \mathbb{A}^n$ realizes or is compatible with a secondary structure $\mathscr{S}$ of length n if for any $(i, j) \in \mathscr{S}$ it holds that $s_i, s_j \in \mathbb{B}$.

- *dependency graph or shape graph* $G(\mathscr{S}_1, \ldots, \mathscr{S}_k)$

QUESTION:

On what conditions is it possible to find *one* sequence, which is realizing *all* sec.str. $\mathscr{S}_1, \ldots, \mathscr{S}_k$ of the same size?

equivalent to:

What properties have to be fulfilled in $G(\mathscr{S}_1, \ldots, \mathscr{S}_k)$, s.t. there exists a single sequence, which is realizing all $\mathscr{S}_1, \ldots, \mathscr{S}_k$?
$C(\mathscr{S})$ denotes the set of all sequences that realize Secondary Structure $\mathscr{S}$.

Theorem (Intersection Theorem, Reidys et al. 1995)

For any two secondary structures $\mathscr{S}_1$ *and* $\mathscr{S}_2$ *of same size holds:* $C(\mathscr{S}_1) \cap C(\mathscr{S}_2) \neq \emptyset$.

Theorem (Generalized Intersection Theorem, Flamm et al. 2001)

$\bigcap_{i=1}^{k} C(\mathscr{S}_i) \neq \emptyset \Leftrightarrow G(\mathscr{S}_1, \ldots, \mathscr{S}_k)$ *is bipartite.*

All proofs on whiteboard

RNA folding: hierarchical process in which secondary structure is broadly considered as sufficient approximation assessing the most relevant characteristics of an RNA molecule

$\mathbb{S}(x)$ = set of all second.structures $\mathscr{S}$ that realize a given RNA sequence x.
$\mathbb{S}(x)$ is called folding space of x.
Structure prediction means to select the "most-likely" structure from elements of $\mathbb{S}(x)$.

"most-likely" = most stable.

Let us start with a simple naive approach, that laid the foundation for many more sophisticated and realistic approaches: Nussinov Algorithm

Nussinov/Jacobson energy model (NJ)

Base-pair maximization (with a twist):

- Additive model on independently contributing base-pairs;
- Canonical base-pairs only: Watson/Crick (A/U,C/G) and Wobble (G/U)

Variant: Weight each pair with #Hydrogen bonds

$w(G\equiv C) = 3 \qquad w(A{=}U) = 2 \qquad w(G{-}U) = 1 \qquad w(\text{other}) = 0$

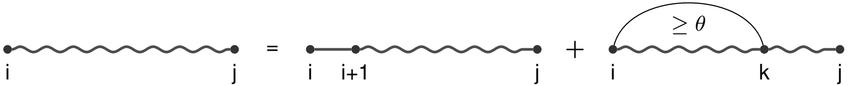
Nussinov/Jacobson energy model (NJ)

Base-pair maximization (with a twist):

- Additive model on independently contributing base-pairs;
- Canonical base-pairs only: Watson/Crick (A/U,C/G) and Wobble (G/U)

Variant: Weight each pair with #Hydrogen bonds

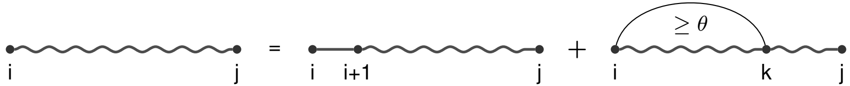
$$w(\text{G}\equiv\text{C}) = 3 \qquad w(\text{A}=\text{U}) = 2 \qquad w(\text{G}-\text{U}) = 1 \qquad w(\text{other}) = 0$$



$$N_{i,t} = 0, \quad \forall t \in [i, i+\theta]$$

$$N_{i,j} = \max \begin{cases} N_{i+1,j} & i \text{ unpaired} \\ \max\limits_{i+\theta<k\leq j} w_{i,k} + N_{i+1,k-1} + N_{k+1,j} & i \text{ paired with } k \end{cases}$$

$w_{i,k} = w(s_i, s_k) \in \{0, 1, 2, 3\}$.



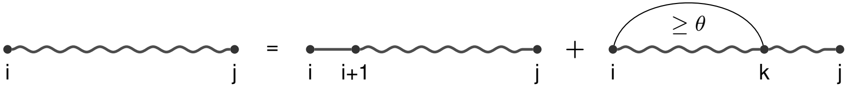
$$N_{i,t} = 0, \quad \forall t \in [i, i+\theta]$$

$$N_{i,j} = \max \begin{cases} N_{i+1,j} & i \text{ unpaired} \\ \max\limits_{i+\theta<k\leq j} w_{i,k} + N_{i+1,k-1} + N_{k+1,j} & i \text{ paired with } k \end{cases}$$

$w_{i,k} = w(s_i, s_k) \in \{0, 1, 2, 3\}$.

Correctness. Goal = Show that MFE over interval $[i, j]$ is indeed found in $N_{i,j}$ after completing the computation. Proceed by induction:

- Assume that property holds for any $[i', j']$ such that $j' - i' < \ell$.
- Consider $[i, j], j - i = \ell$. Let $\text{MFE}_{i,j} :=$ Base-pairs of best struct. on $[i, j]$. Then first position i in $\text{MFE}_{i,j} =$ is either:
  - **Unpaired:** $\text{MFE}_{i,j} = \text{MFE}_{i+1,j}$ $\rightarrow$ free-energy $= N_{i+1,j}$
  - **Paired to k:** $\text{MFE}_{i,j} = \{(i,k)\} \cup \text{MFE}_{i+1,k-1} \cup \text{MFE}_{k+1,j}$.
    (Indeed, any BP between $[i+1, k-1]$ and $[k+1, j]$ would cross (i,k))
    $\rightarrow$ free-energy $= w_{i,k} + N_{i+1,k-1} + N_{k+1,j}$



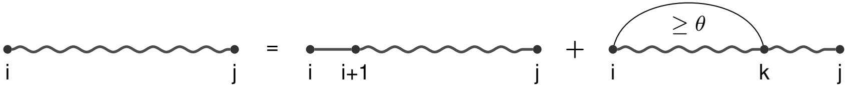
$$N_{i,t} = 0, \quad \forall t \in [i, i+\theta]$$

$$N_{i,j} = \max \begin{cases} N_{i+1,j} & i \text{ unpaired} \\ \max_{i+\theta<k\leq j} w_{i,k} + N_{i+1,k-1} + N_{k+1,j} & i \text{ paired with } k \end{cases}$$

$w_{i,k} = w(s_i, s_k) \in \{0, 1, 2, 3\}$.

Correctness. Goal = Show that MFE over interval $[i, j]$ is indeed found in $N_{i,j}$ after completing the computation. Proceed by induction:

- Assume that property holds for any $[i', j']$ such that $j' - i' < \ell$.
- Consider $[i, j], j - i = \ell$. Let $\text{MFE}_{i,j}$:= Base-pairs of best struct. on $[i, j]$. Then first position i in $\text{MFE}_{i,j}$ = is either:
  - **Unpaired:** $\text{MFE}_{i,j} = \text{MFE}_{i+1,j}$ → free-energy $= N_{i+1,j}$
  - **Paired to k:** $\text{MFE}_{i,j} = \{(i,k)\} \cup \text{MFE}_{i+1,k-1} \cup \text{MFE}_{k+1,j}$.
    (Indeed, any BP between $[i+1, k-1]$ and $[k+1, j]$ would cross (i, k)))
    → free-energy $= w_{i,k} + N_{i+1,k-1} + N_{k+1,j}$



$$N_{i,t} = 0, \quad \forall t \in [i, i+\theta]$$

$$N_{i,j} = \max \begin{cases} N_{i+1,j} & i \text{ unpaired} \\ \max_{i+\theta<k\leq j} w_{i,k} + N_{i+1,k-1} + N_{k+1,j} & i \text{ paired with } k \end{cases}$$

$w_{i,k} = w(s_i, s_k) \in \{0, 1, 2, 3\}$.

Correctness. Goal = Show that MFE over interval $[i, j]$ is indeed found in $N_{i,j}$ after completing the computation. Proceed by induction:

- Assume that property holds for any $[i', j']$ such that $j' - i' < \ell$.
- Consider $[i, j], j - i = \ell$. Let $\text{MFE}_{i,j}$:= Base-pairs of best struct. on $[i, j]$. Then first position i in $\text{MFE}_{i,j}$ = is either:
  - **Unpaired:** $\text{MFE}_{i,j} = \text{MFE}_{i+1,j}$ $\quad\rightarrow$ free-energy $= N_{i+1,j}$
  - **Paired to k:** $\text{MFE}_{i,j} = \{(i,k)\} \cup \text{MFE}_{i+1,k-1} \cup \text{MFE}_{k+1,j}$.
    (Indeed, any BP between $[i+1, k-1]$ and $[k+1, j]$ would cross (i, k))
    $\rightarrow$ free-energy $= w_{i,k} + N_{i+1,k-1} + N_{k+1,j}$

Nussinov/Jacobson DP scheme

$$N_{i,t} = 0, \quad \forall t \in [i, i+\theta]$$

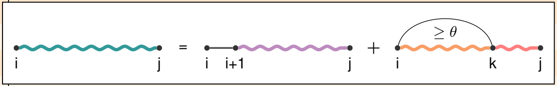
$$N_{i,j} = \max \begin{cases} N_{i+1,j} & i \text{ unpaired} \\ \max\limits_{i+\theta<k\leq j} w_{i,k} + N_{i+1,k-1} + N_{k+1,j} & i \text{ paired with } k \end{cases}$$

$w_{i,k} = w(s_i, s_k) \in \{0, 1, 2, 3\}$.

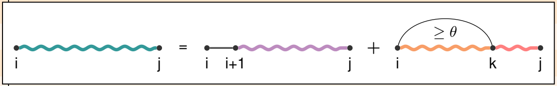
$i = 0, j = \mathrm{SIZE} - 1$;
`Traceback`(matrix N, i, j)

if $i < \mathrm{SIZE} - \theta - 1$ **then**
    if $N_{i,j} = N_{i+1,j}$ **then**   *// i unpaired*
        `Traceback`(N, i+1, j)
    else
        for($k = i + \theta + 1, \ldots, j$)
            if $N[i][j] = w_{i,k} + N[i+1][k-1] + N[k+1][j]$) and neither i, k paired
                print "basepair (i, k)"
                `Traceback`(N, i+1, k-1)
                `Traceback`(N, k+1, j)

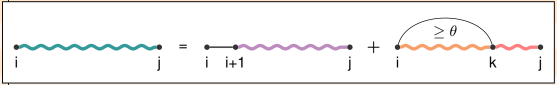
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



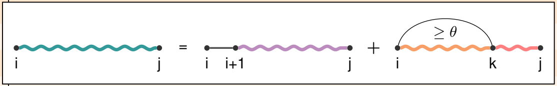
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



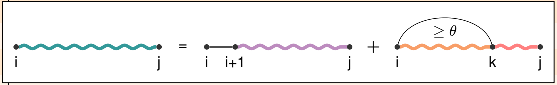
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



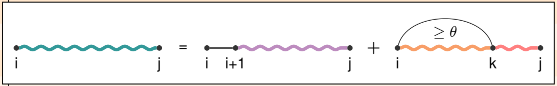
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



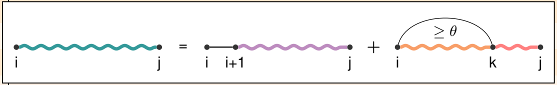
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



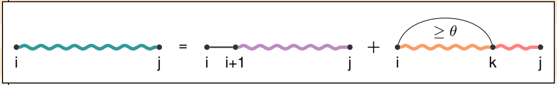
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



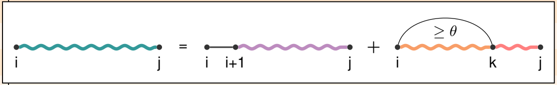
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



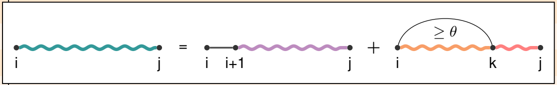
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



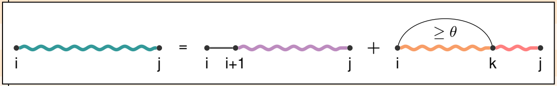
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



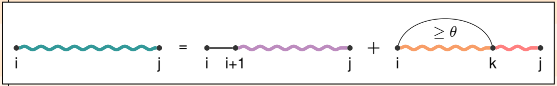
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



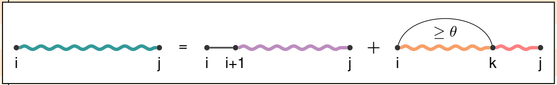
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



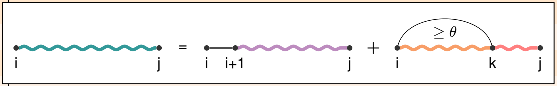
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



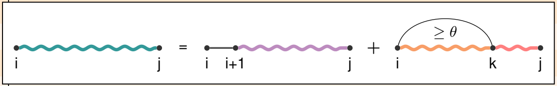
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



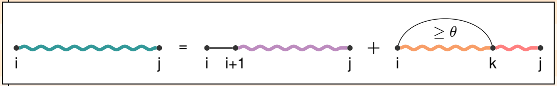
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	.	.	.	.	.	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



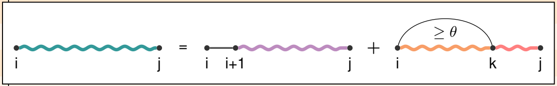
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	.	.	.	.	.	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



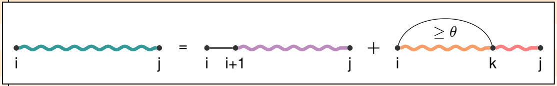
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	.	.	.	.	.	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



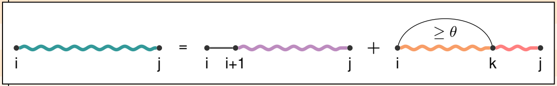
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



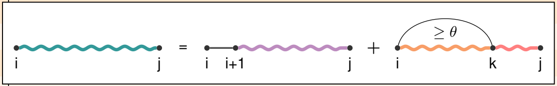
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



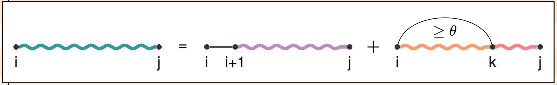
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



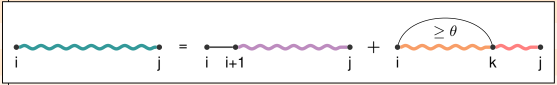
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



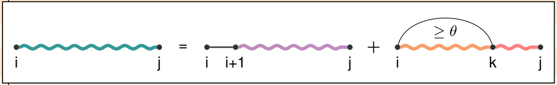
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



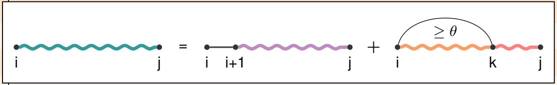
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



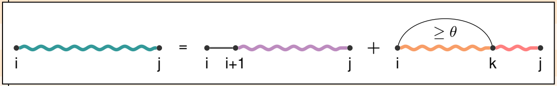
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	.	.	.	.	.	.	.	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



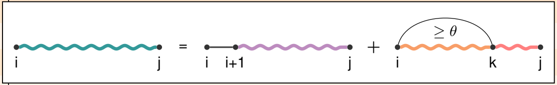
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	(	.	.	.	.	.	)	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



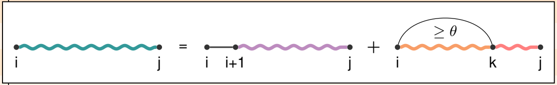
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	(	.	.	.	.	.	)	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



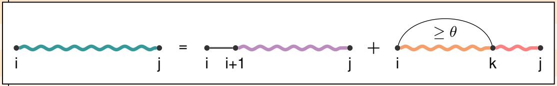
	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	(	.	.	.	.	.	)	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	(	(	.	.	.	)	)	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



	C	G	G	A	U	A	C	U	U	C	U	U	A	G	A	C	G	A
	(	(	(	.	.	.	)	.	(	(	.	.	.	)	)	)	)	.
C	0	0	0	0	0	0	3	4	4	6	6	6	6	9	9	11	14	14
G		0	0	0	0	0	3	4	4	6	6	6	6	7	9	11	11	11
G			0	0	0	0	3	3	3	5	5	5	5	6	8	10	10	10
A				0	0	0	0	2	2	2	2	4	4	5	7	7	8	10
U					0	0	0	0	0	0	2	2	4	5	7	7	8	10
A						0	0	0	0	0	2	2	2	5	5	5	8	8
C							0	0	0	0	0	0	2	5	5	5	8	8
U								0	0	0	0	0	2	3	5	5	6	7
U									0	0	0	0	2	3	5	5	5	7
C										0	0	0	0	3	3	3	5	5
U											0	0	0	0	2	2	2	3
U												0	0	0	0	0	1	2
A													0	0	0	0	0	0
G														0	0	0	0	0
A															0	0	0	0
C																0	0	0
G																	0	0
A																		0



Nussinov drawbacks

Maximizing the nr of bp does not lead to biological meaningful structures:

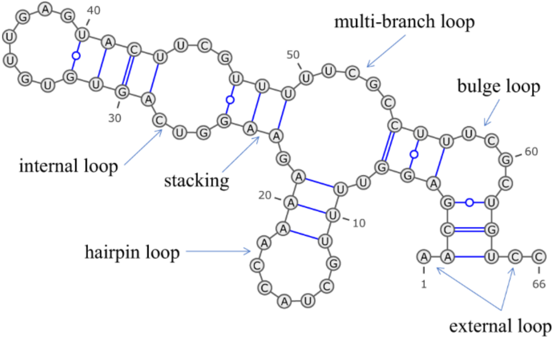
Stacking of bp not considered:

(	(	(	.	)	)	)	stable
(	)	(	)	.	(	)	instable

Size of intern loop not considered:

Nevertheless, although Nussinov-Alg is too simple to be accurate, it is stepping-stone for later algorithms

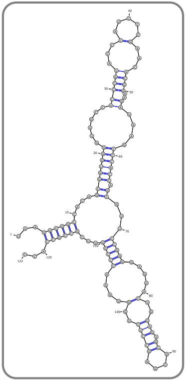
Define energy model for RNA that takes into account local energy contributions from loop and stacking regions.



- More realistic: thermodynamics and statistical mechanics.
- Stability of an RNA sec.str. coincides with thermodynamic stability
- Quantified as the amount of free energy released/used by forming bp.

Based on unambiguous decomposition of 2ᵃʳʸ structure into loops:

- Internal loops
- Bulges
- Terminal (hairpin) loops
- Multi loops
- Stackings



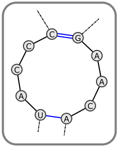
Free-energy Δ G of a loop depend on bases, assymmetry, dangles . . .

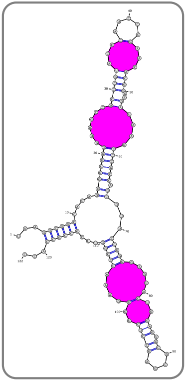
Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.

Based on unambiguous decomposition of 2ary structure into loops:

- Internal loops
- Bulges
- Terminal (hairpin) loops
- Multi loops
- Stackings





Free-energy Δ G of a loop depend on bases, assymmetry, dangles . . .

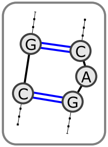
Experimentally determined
+ Interpolated for larger loops.

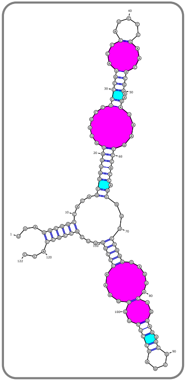
Improved results by taking stacking into account.

Turner energy model

Based on unambiguous decomposition of 2^{ary} structure into loops:

- Internal loops
- Bulges
- Terminal (hairpin) loops
- Multi loops
- Stackings





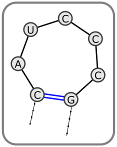
Free-energy Δ G of a loop depend on bases, assymmetry, dangles . . .

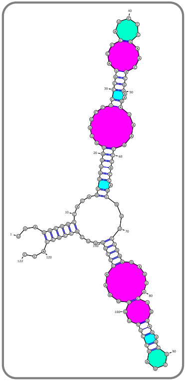
Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.

Based on unambiguous decomposition of 2^{ary} structure into loops:

- Internal loops
- Bulges
- Terminal (hairpin) loops
- Multi loops
- Stackings





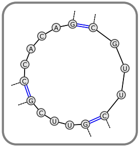
Free-energy Δ G of a loop depend on bases, assymmetry, dangles . . .

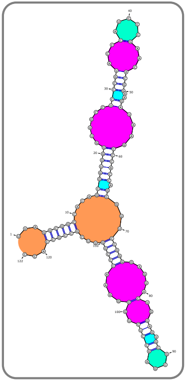
Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.

Based on unambiguous decomposition of 2^{ary} structure into loops:

- Internal loops
- Bulges
- Terminal (hairpin) loops
- Multi loops
- Stackings





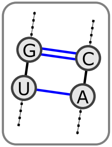
Free-energy Δ G of a loop depend on bases, assymmetry, dangles . . .

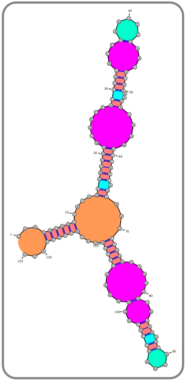
Experimentally determined
+ Interpolated for larger loops.

Improved results by taking stacking into account.

Based on unambiguous decomposition of 2^{ary} structure into loops:

- Internal loops
- Bulges
- Terminal (hairpin) loops
- Multi loops
- Stackings





Free-energy Δ G of a loop depend on bases, assymmetry, dangles . . .

Experimentally determined
+ Interpolated for larger loops.

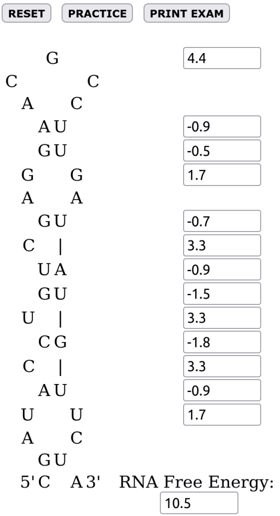
Improved results by taking stacking into account.

The Turner rules are a set of experimentally determined parameters which allow us to predict the stability of RNA secondary structures.

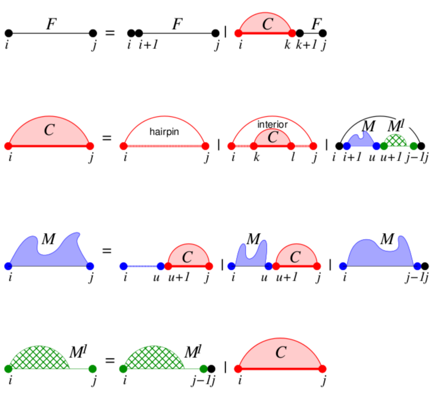
Turner Energy Rules

BOTTOM \ TOP	AU	CG	GC	UA	GU	UG
AU	-0.9	-1.8	-2.3	-1.1	-0.5	-0.7
CG	-2.1	-2.9	-3.4	-2.3	-1.5	-1.5
GC	-1.7	-2	-2.9	-1.8	-1.3	-1.5
UA	-0.9	-1.7	-2.1	-0.9	-0.7	-0.5
GU	-0.9	-1.7	-2.1	-0.9	-0.5	-0.5
UG	-0.9	-1.7	-2.1	-0.9	0.6	-0.5

Bases in Loop	Internal Loop	Bulge Loop	Hairpin Loop
1	0	3.3	0
2	0.8	5.2	0
3	1.3	6	7.4
4	1.7	6.7	5.9
5	2.1	7.4	4.4
6	2.5	8.2	4.3
7	2.6	9.1	4.1
8	2.8	10	4.1



https://www.kelleybioinfo.org/



$$F_{i,j} = \min \begin{Bmatrix} F_{i+1,j}, \\ \min\limits_{i<k\leq j} C_{i,k} + F_{k+1,j} \end{Bmatrix}$$

$$C_{i,j} = \min \begin{Bmatrix} \mathcal{H}(i,j), \\ \min\limits_{i<k<l<j} C_{k,l} + \mathcal{I}(i,j;k,l), \\ \min\limits_{i<u<j} M_{i+1,u} + M^1_{u+1,j-1} + a \end{Bmatrix}$$

$$M_{i,j} = \min \begin{Bmatrix} \min\limits_{i<u<j} (u-i+1)c + C_{u+1,j} + b, \\ \min\limits_{i<u<j} M_{i,u} + C_{u+1,j} + b, \\ M_{i,j-1} + c \end{Bmatrix}$$

$$M^1_{i,j} = \min \begin{Bmatrix} M^1_{i,j-1} + c, \\ C_{i,j} + b \end{Bmatrix}$$

First proposed by Zuker et al.

init: $F_{i,j} = 0$, $C_{i,j} = M_{i,j} = M^1_{i,j} = \infty$.

- $F_{1,n}$ stores the energy value of the thermodynamically most stable structure, its Minimum Free Energy (MFE).
- traceback structure

MFE-folding via loop decomposition: remarks

- $F_{i,j}$: free energy of the opt. sub-struct. on the sub-seq. $s_i \dots s_j$.
- $C_{i,j}$: free energy of the opt. sub-struct. on the sub-seq. $s_i \dots s_j$ given that i and j form a base pair.
- $M_{i,j}$: free energy of the opt. sub-struct. on the sub-seq. $s_i \dots s_j$ given that $s_i \dots s_j$ is part of a multi-loop and has at least one "component".
- $M^1_{i,j}$: free energy of the opt. sub-struct. on the sub-seq. $s_i \dots s_j$ given that $s_i \dots s_j$ is part of a multi-loop and has exactly one component which has the closing pair (i, h) for some h satisfying $i < h \leq j$.

RNA Fold


```
http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi

 RNAfold < trna.fa
>AF041468
GGGGGUAUAGCUCAGUUGGUAGAGCGCUGCCUUUGCACGGCAGAUGUCAGGGGUUCGAGUCCCCUUACCUCCA
(((((((..((((........)))).(((((.......))))).....(((((.......)))))))))))).
-31.10 kcal/mol
```

What is RNA?