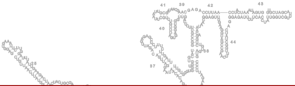
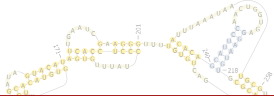




RNA Design

Designability and Structure-Approximating Algorithm in Watson-Crick and Nussinov-Jacobson Energy Models

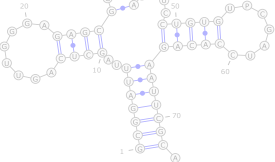
Jozef Haleš⋆, **Alice Héliou**•,†, Ján Maňuch⋆, Yann Ponty⋆,•,†, Ladislav Stacho⋆

⋆ Simon Fraser University/PIMS, Vancouver, Canada

• LIX, CNRS/Ecole Polytechnique

† Amib project-team, Inria Saclay

Slides were made by Yann Ponty



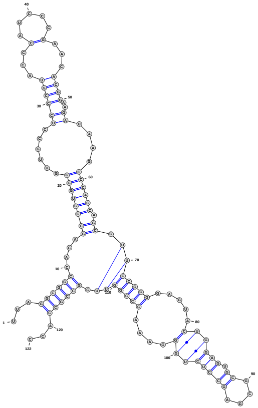
RNA structures

RNA = Linear Polymer = Sequence in $\{A, C, G, U\}^\star$

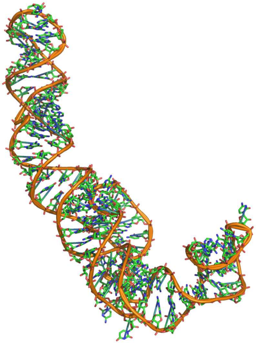
UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAAGAUAAGCC
CACCAGCGUUCCGGG
GAGUACUGGAGUGCG
CGAGCCUCUGGGAAA
CCCGGUUCGCCGCCA

CC

Primary Structure



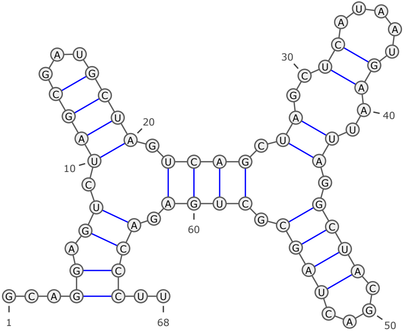
Secondary Structure



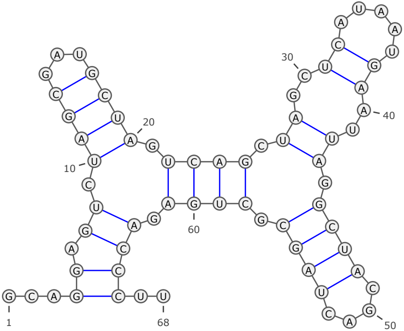
Tertiary Structure

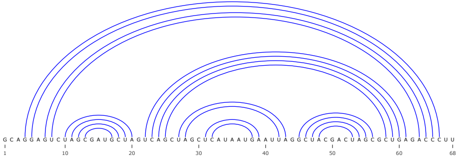
5s rRNA (PDBID: 1K73:B)

Structure = Bunch of **non-crossing** base-pairs.



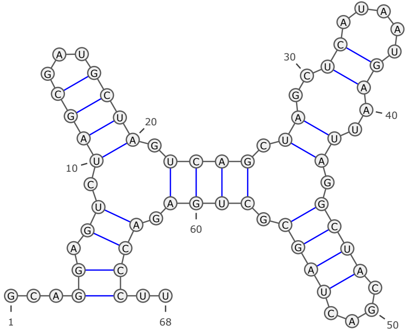
Structure = Bunch of **non-crossing** base-pairs.

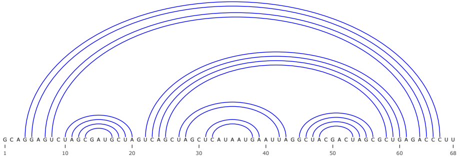




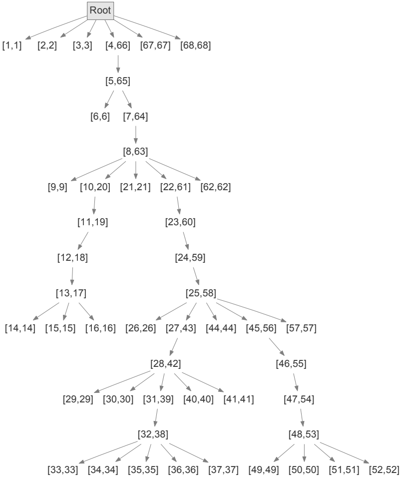
arc diagram

Structure = Bunch of **non-crossing** base-pairs.



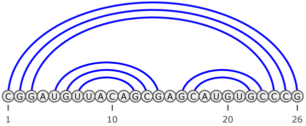


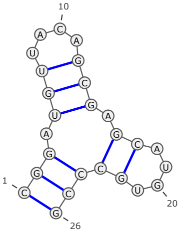
arc diagram



tree representation

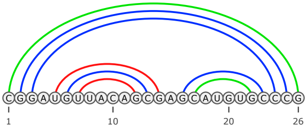
Energy Model

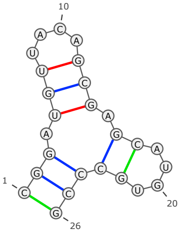




- **RNA structure S:** Non-crossing base-pairs for positions in sequence w
- **Motifs:** Sequence/structure features (e.g. Base-pairs, Stacking pairs, Loops. . .)
- **Energy model:**
  Motif $\rightarrow$ Free-energy contribution $\Delta(\cdot) \in \mathbb{R}^{-} \cup \{+\infty\}$
  Free-Energy $E_w(S)$: Sum over (independently contributing) motifs in S

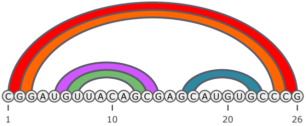
Energy Model

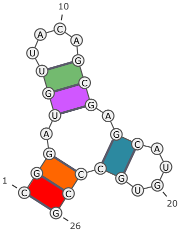




- **RNA structure S:** Non-crossing base-pairs for positions in sequence w
- **Motifs:** Sequence/structure features (e.g. Base-pairs, Stacking pairs, Loops. . .)
- **Energy model:**
  Motif $\rightarrow$ Free-energy contribution $\Delta(\cdot) \in \mathbb{R}^- \cup \{+\infty\}$
  Free-Energy $E_w(S)$: Sum over (independently contributing) motifs in S

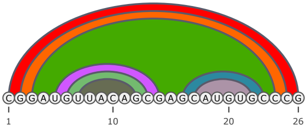
Energy Model

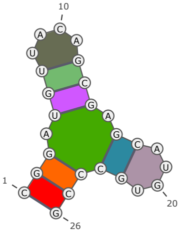




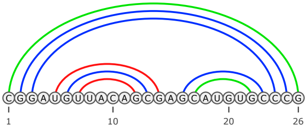
- **RNA structure *S*:** Non-crossing base-pairs for positions in sequence *w*
- **Motifs:** Sequence/structure features (e.g. Base-pairs, Stacking pairs, Loops. . .)
- **Energy model:**
  Motif $\rightarrow$ Free-energy contribution $\Delta(\cdot) \in \mathbb{R}^- \cup \{+\infty\}$
  Free-Energy $E_w(S)$: Sum over (independently contributing) motifs in S

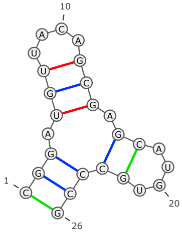
Energy Model





- **RNA structure S:** Non-crossing base-pairs for positions in sequence w
- **Motifs:** Sequence/structure features (e.g. Base-pairs, Stacking pairs, Loops. . .)
- **Energy model:**
  Motif $\rightarrow$ Free-energy contribution $\Delta(\cdot) \in \mathbb{R}^{-} \cup \{+\infty\}$
  Free-Energy $E_w(S)$: Sum over (independently contributing) motifs in S

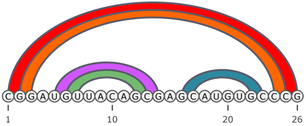


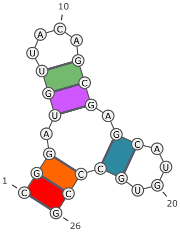


- **RNA structure *S*:** Non-crossing base-pairs for positions in sequence *w*
- **Motifs:** Sequence/structure features (e.g. Base-pairs, Stacking pairs, Loops. . .)
- **Energy model:**
  Motif $\rightarrow$ Free-energy contribution $\Delta(\cdot) \in \mathbb{R}^- \cup \{+\infty\}$
  Free-Energy $E_w(S)$: Sum over (independently contributing) motifs in S

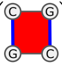
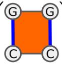
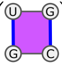
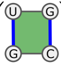
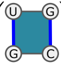
$$E_S = 2\cdot\Delta\begin{pmatrix}\text{U}\\ |\\ \text{G}\end{pmatrix} + 4\cdot\Delta\begin{pmatrix}\text{G}\\ |\\ \text{C}\end{pmatrix} + 2\cdot\Delta\begin{pmatrix}\text{C}\\ |\\ \text{G}\end{pmatrix}$$

Energy Model

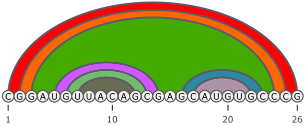


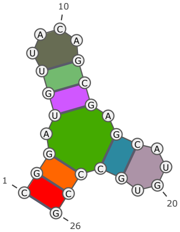


- **RNA structure S:** Non-crossing base-pairs for positions in sequence w
- **Motifs:** Sequence/structure features (e.g. Base-pairs, Stacking pairs, Loops...)
- **Energy model:**
  Motif $\rightarrow$ Free-energy contribution $\Delta(\cdot) \in \mathbb{R}^- \cup \{+\infty\}$
  Free-Energy $E_w(S)$: Sum over (independently contributing) motifs in S

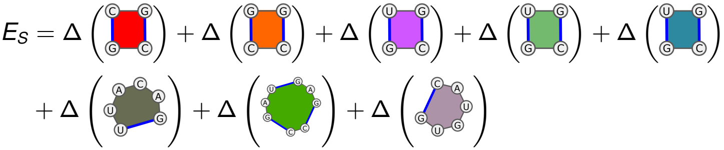
$E_S = \Delta\Big($  $\Big) + \Delta\Big($  $\Big) + \Delta\Big($  $\Big) + \Delta\Big($  $\Big) + \Delta\Big($  $\Big)$

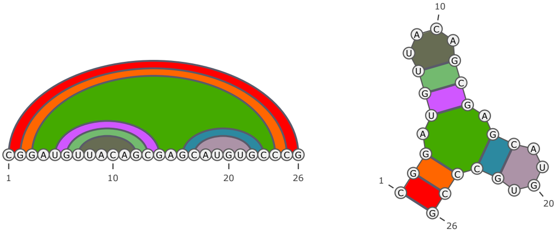
Energy Model





- **RNA structure S:** Non-crossing base-pairs for positions in sequence w
- **Motifs:** Sequence/structure features (e.g. Base-pairs, Stacking pairs, Loops. . .)
- **Energy model:**
  Motif $\rightarrow$ Free-energy contribution $\Delta(\cdot) \in \mathbb{R}^- \cup \{+\infty\}$
  Free-Energy $E_w(S)$: Sum over (independently contributing) motifs in S





- **RNA structure S:** Non-crossing base-pairs for positions in sequence w
- **Motifs:** Sequence/structure features (e.g. Base-pairs, Stacking pairs, Loops. . .)
- **Energy model:**
  Motif $\rightarrow$ Free-energy contribution $\Delta(\cdot) \in \mathbb{R}^- \cup \{+\infty\}$
  Free-Energy $E_w(S)$: Sum over (independently contributing) motifs in S

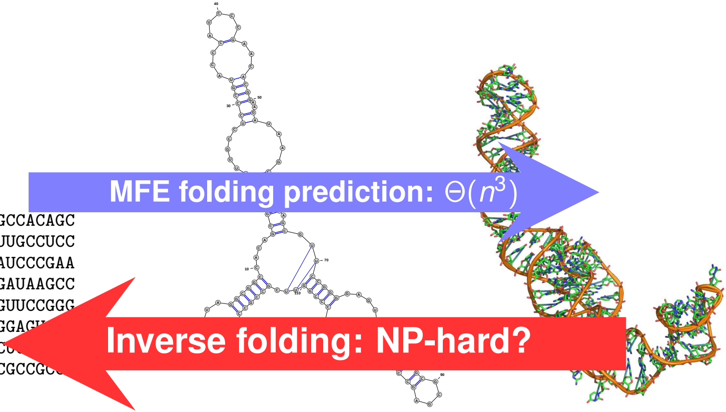
Problem solved **exactly** in $O(n^3)$ time.

[Nussinov Jacobson, PNAS 1980] [Zuker Stiegler, NAR 1981]. . . .

RNA = Linear Polymer = Sequence in $\{A, C, G, U\}^\star$

UUAGGCGGCCACAGC
GGUGGGGUUGCCUCC
CGUACCCAUCCCGAA
CACGGAAGAUAAGCC
CACCAGCGUUCCGGG
GAGUACUGGAG
CGAGCCUCU
CCCGGUUCGCCGC

CC



Primary Structure | Secondary Structure | Structure Tertiaire

5s rRNA (PDBID: 1K73:B)

Design objectives

Positive structural design

Optimize **affinity** of designed sequences towards target structure

Examples: Most stable sequence for given fold...

Negative structural design

Limit affinity of designed sequences towards **alternative structures**

Examples: Lowest free-energy, High Boltzmann probability/Low entropy...

RNA Design Problem

Let $\mathcal{M}$ be an energy model.

Problem (INVERSE-FOLDING($\mathcal{M}, \Sigma, \Delta$) problem)

Input: *Secondary structure S + Energy distance $\Delta > 0$*
Output: *RNA sequence $w \in \Sigma^\star$ — called a design for S — such that:*

$$\forall S' \in \mathcal{S}_{|w|} \setminus \{S\} : \; E_{\mathcal{M}}(w, S') \geq E_{\mathcal{M}}(w, S) + \Delta$$

or $\varnothing$ if no such sequence exists.

Difficult problem: No obvious DP decomposition

- Existing algorithms: Heuristics or Exponential-time
- Complexity of problem unknown (despite [Schnall Levin et al (2008)])
  Reason: Non locality, no theoretical frameworks, too many parameters...

$\Rightarrow$ **Stick to a simplified model!**

RNA Design Problem

Let $\mathcal{M}$ be an energy model.

Problem (INVERSE-FOLDING($\mathcal{M}, \Sigma, \Delta$) problem)

Input: *Secondary structure S + Energy distance* $\Delta > 0$
Output: *RNA sequence* $w \in \Sigma^\star$ *— called a design for S — such that:*

$$\forall S' \in \mathcal{S}_{|w|} \setminus \{S\} : \; E_{\mathcal{M}}(w, S') \geq E_{\mathcal{M}}(w, S) + \Delta$$

or $\varnothing$ *if no such sequence exists.*

Difficult problem: No obvious DP decomposition

- Existing algorithms: Heuristics or Exponential-time
- Complexity of problem unknown (despite [Schnall Levin et al (2008)])
  Reason: Non locality, no theoretical frameworks, too many parameters...

$\Rightarrow$ **Stick to a simplified model!**

RNA Design Problem (simplified)

Simplified formulation for Watson-Crick model $\mathcal{W}$ and $\Delta = 1$:

Problem (INVERSE-FOLDING(Σ) problem)

Input: *Secondary structure S*
Output: *RNA sequence $w \in \Sigma^\star$ — called a design for S — such that:*

$$\text{RNA-FOLD}_{\mathcal{W}}(w) = \{S\}$$

or $\varnothing$ if no such sequence exists.

Designable(Σ)**:** All designable structures

RNA Design Problem (simplified)

Simplified formulation for Watson-Crick model $\mathcal{W}$ and $\Delta = 1$:

Problem (INVERSE-FOLDING(Σ) problem)

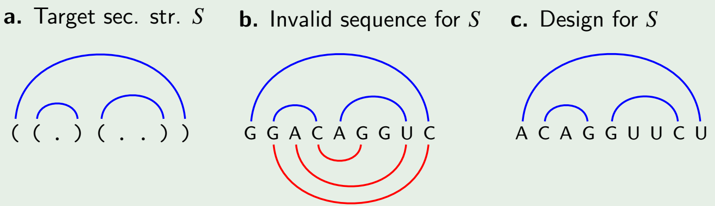
Input: *Secondary structure S*
Output: *RNA sequence $w \in \Sigma^\star$ — called a design for S — such that:*

$$\text{RNA-FOLD}_{\mathcal{W}}(w) = \{S\}$$

or $\varnothing$ if no such sequence exists.

Designable(Σ): All designable structures

Example

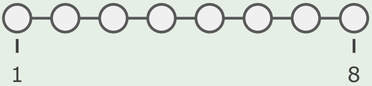


Our Results: Definitions and notations

Given a secondary structure S:

- Unpaired_S = Set of all unpaired positions of S.
- S is **saturated** $\Leftrightarrow \text{Unpaired}_S = \varnothing$.
  Saturated = Set of all saturated structures.
- **Paired degree of base-pair** = #Helices on the loop.
- $D(S)$ = Maximal *paired degree* of nodes in the tree representation of S.

Example



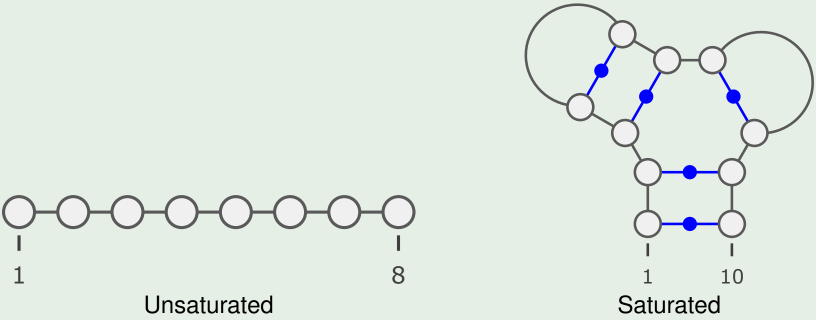
$\text{Unpaired}_S = \{1, 2, 3, 4, 5, 6, 7, 8\}$

Our Results: Definitions and notations

Given a secondary structure S:

- Unpaired_S = Set of all unpaired positions of S.
- S is **saturated** $\Leftrightarrow \text{Unpaired}_S = \varnothing$.
  Saturated = Set of all saturated structures.
- **Paired degree of base-pair** = #Helices on the loop.
- $D(S)$ = Maximal *paired degree* of nodes in the tree representation of S.

Example

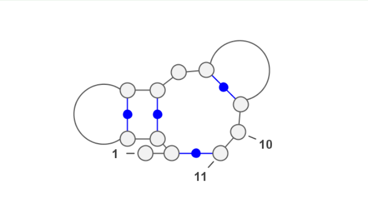


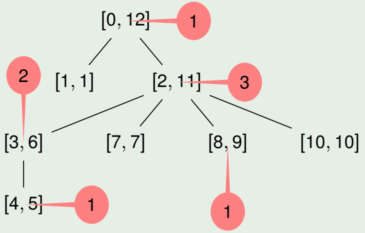
Our Results: Definitions and notations

Given a secondary structure S:

- Unpaired$_S$ = Set of all unpaired positions of S.
- S is **saturated** $\Leftrightarrow$ Unpaired$_S = \varnothing$.
  Saturated = Set of all saturated structures.
- **Paired degree of base-pair** = #Helices on the loop.
- $D(S)$ = Maximal *paired degree* of nodes in the tree representation of S.

Example





$D(S) = 3$

Our Results: Designability over Restricted Alphabets

$\Sigma_{c,u} =$ Alphabet with c pairs of complementary bases and u unpairable bases.

R1 $\Sigma_{0,u} \Rightarrow$ Designable = Empty (single-stranded) structures;

R2 $\Sigma_{1,0} \Rightarrow$ Designable = Saturated with degree≤ 2 + empty structures ;

R3 $\Sigma_{1,1} \Rightarrow$ Designable = Degree≤ 2.

Our Results: Designability over Restricted Alphabets

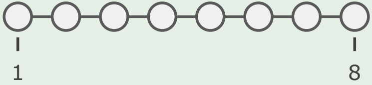
$\Sigma_{c,u}$ = Alphabet with c pairs of complementary bases and u unpairable bases.

R1 $\Sigma_{0,u} \Rightarrow$ Designable = Empty (single-stranded) structures;

R2 $\Sigma_{1,0} \Rightarrow$ Designable = Saturated with degree≤ 2 + empty structures ;

R3 $\Sigma_{1,1} \Rightarrow$ Designable = Degree≤ 2.

Example

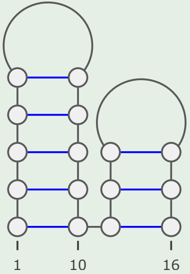


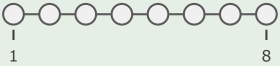
Our Results: Designability over Restricted Alphabets

$\Sigma_{c,u}$ = Alphabet with c pairs of complementary bases and u unpairable bases.

- **R1** $\Sigma_{0,u} \Rightarrow$ Designable = Empty (single-stranded) structures;
- **R2** $\Sigma_{1,0} \Rightarrow$ Designable = Saturated with degree≤ 2 + empty structures ;
- **R3** $\Sigma_{1,1} \Rightarrow$ Designable = Degree≤ 2.

Example





Our Results: Designability over Restricted Alphabets

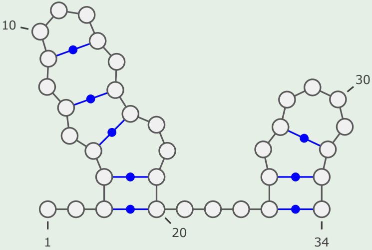
$\Sigma_{c,u} =$ Alphabet with c pairs of complementary bases and u unpairable bases.

R1 $\Sigma_{0,u} \Rightarrow$ Designable $=$ Empty (single-stranded) structures;

R2 $\Sigma_{1,0} \Rightarrow$ Designable $=$ Saturated with degree≤ 2 + empty structures ;

R3 $\Sigma_{1,1} \Rightarrow$ Designable $=$ Degree≤ 2.

Example



+ miRNAs, some lncRNAs. . .

Our Results: Designability over Restricted Alphabets

$\Sigma_{c,u}$ = Alphabet with c pairs of complementary bases and u unpairable bases.

R1 $\Sigma_{0,u} \Rightarrow$ Designable $=$ Empty (single-stranded) structures;

R2 $\Sigma_{1,0} \Rightarrow$ Designable $=$ Saturated with degree≤ 2 + empty structures ;

R3 $\Sigma_{1,1} \Rightarrow$ Designable $=$ Degree≤ 2.

Question: Why not degree 3?

Proof.

□

Our Results: Designability over Restricted Alphabets

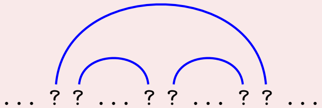
$\Sigma_{c,u}$ = Alphabet with c pairs of complementary bases and u unpairable bases.

- **R1** $\Sigma_{0,u} \Rightarrow$ Designable $=$ Empty (single-stranded) structures;
- **R2** $\Sigma_{1,0} \Rightarrow$ Designable $=$ Saturated with degree≤ 2 + empty structures ;
- **R3** $\Sigma_{1,1} \Rightarrow$ Designable $=$ Degree≤ 2.

Question: Why not degree 3?

Proof.

Within an internal node:



□

Our Results: Designability over Restricted Alphabets

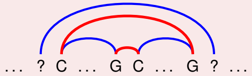
$\Sigma_{c,u}$ = Alphabet with c pairs of complementary bases and u unpairable bases.

- **R1** $\Sigma_{0,u} \Rightarrow$ Designable = Empty (single-stranded) structures;
- **R2** $\Sigma_{1,0} \Rightarrow$ Designable = Saturated with degree≤ 2 + empty structures ;
- **R3** $\Sigma_{1,1} \Rightarrow$ Designable = Degree≤ 2.

Question: Why not degree 3?

Proof.

Within an internal node:



Either we get a repeat...

$\square$

Our Results: Designability over Restricted Alphabets

$\Sigma_{c,u} =$ Alphabet with c pairs of complementary bases and u unpairable bases.

R1 $\Sigma_{0,u} \Rightarrow$ Designable $=$ Empty (single-stranded) structures;

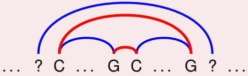
R2 $\Sigma_{1,0} \Rightarrow$ Designable $=$ Saturated with degree≤ 2 + empty structures ;

R3 $\Sigma_{1,1} \Rightarrow$ Designable $=$ Degree≤ 2.

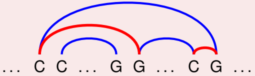
Question: Why not degree 3?

Proof.

Within an internal node:



Either we get a repeat...



...or some parent/child have complementary pairs.

+ Same principle at the root level. □

Our Results: Designability over Restricted Alphabets

$\Sigma_{c,u}$ = Alphabet with c pairs of complementary bases and u unpairable bases.

- **R1** $\Sigma_{0,u} \Rightarrow$ Designable $=$ Empty (single-stranded) structures;
- **R2** $\Sigma_{1,0} \Rightarrow$ Designable $=$ Saturated with degree≤ 2 + empty structures ;
- **R3** $\Sigma_{1,1} \Rightarrow$ Designable $=$ Degree≤ 2.

This can be easily generalized to:

Lemma

For any structure S in Designable$(\Sigma_{c,u})$*, $D(S) \leq 2c$.*

Our Results: Designability over the Complete Alphabet

$\Sigma_{2,0} = \{A, U, C, G\}$ + $\{G - C, A - U\}$ base pairs.

Without unpaired position $\rightarrow$ complete characterization:

R4 $\Sigma_{2,0} \Rightarrow$ Saturated Designable = Degree≤ 4.

With unpaired positions $\rightarrow$ partial characterization:

R5 (Necessary) Designable structure cannot contain "*a multiloop of degree* ≥ 5" (motif m_5) or "*a multiloop with unpaired position of degree* ≥ 3" (motif $m_{3\circ}$).

R6 (Sufficient) Separated = Structure that admit a separated (proper) coloring. Then any Separated **structure is Designable in** $\Sigma_{2,0}$.

R7 If $S \in$ Designable(), then k-stutter $S^{[k]} \in$ Designable($\Sigma_{2,0}$).

Our Results: Designability over the Complete Alphabet

$\Sigma_{2,0} = \{A, U, C, G\}$ + $\{G - C, A - U\}$ base pairs.

Without unpaired position $\rightarrow$ complete characterization:

R4 $\Sigma_{2,0} \Rightarrow$ Saturated Designable = Degree$\leq$ 4.

With unpaired positions $\rightarrow$ partial characterization:

R5 (Necessary) Designable structure cannot contain "*a multiloop of degree* $\geq$ 5" (motif m_5) or "*a multiloop with unpaired position of degree* $\geq$ 3" (motif $m_{3\circ}$).

R6 (Sufficient) Separated = Structure that admit a separated (proper) coloring. Then any Separated **structure is Designable in** $\Sigma_{2,0}$.

R7 If $S \in$ Designable(), then k-stutter $S^{[k]} \in$ Designable($\Sigma_{2,0}$).

Our Results: Designability over the Complete Alphabet

$\Sigma_{2,0} = \{A, U, C, G\}$ + $\{G - C, A - U\}$ base pairs.

Without unpaired position $\rightarrow$ complete characterization:

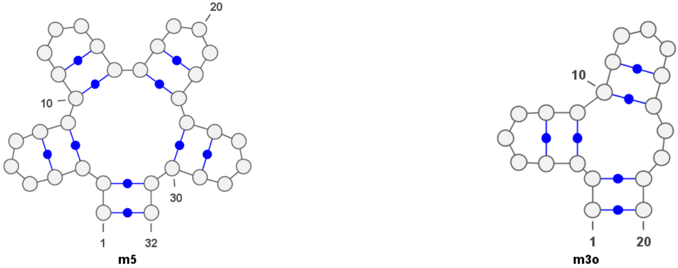
R4 $\Sigma_{2,0} \Rightarrow$ Saturated Designable $=$ Degree≤ 4.

With unpaired positions $\rightarrow$ partial characterization:

R5 (Necessary) Designable structure cannot contain "*a multiloop of degree* ≥ 5" (motif m_5) or "*a multiloop with unpaired position of degree* ≥ 3" (motif $m_{3\circ}$).

R6 (Sufficient) Separated $=$ Structure that admit a separated (proper) coloring. Then any Separated **structure is Designable in** $\Sigma_{2,0}$.

R7 If $S \in$ Designable(), then k-stutter $S^{[k]} \in$ Designable($\Sigma_{2,0}$).



Our Results: Designability over the Complete Alphabet

$\Sigma_{2,0} = \{A, U, C, G\} + \{G - C, A - U\}$ base pairs.

Without unpaired position $\rightarrow$ complete characterization:

R4 $\Sigma_{2,0} \Rightarrow$ Saturated Designable $=$ Degree≤ 4.

With unpaired positions $\rightarrow$ partial characterization:

R5 (Necessary) Designable structure cannot contain "*a multiloop of degree* ≥ 5" (motif m_5) or "*a multiloop with unpaired position of degree* ≥ 3" (motif $m_{3\circ}$).

R6 (Sufficient) Separated $=$ Structure that admit a separated (proper) coloring. Then any Separated **structure is Designable in** $\Sigma_{2,0}$.

R7 If $S \in$ Designable(), then k-stutter $S^{[k]} \in$ Designable($\Sigma_{2,0}$).

Our Results: Separated Coloring

From the tree representation T_S of structure S, color every paired node of T_S:

- black $\to$ G $\cdot$ C;
- white $\to$ C $\cdot$ G;
- grey $\to$ A $\cdot$ U or U $\cdot$ A.

Proper coloring:

1. each internal node has at most one black, one white and two grey children;
2. a grey node has at most one grey child;
3. a black node does not have a white child; and
4. a white node does not have a black child.

Level of a node = #black nodes − #white nodes on the path to the root.

Separated coloring: Levels of grey nodes $\cap$ Levels of unpaired nodes $= \varnothing$

Our Results: Separated Coloring

From the tree representation T_S of structure S, color every paired node of T_S:

- black $\rightarrow$ G $\cdot$ C;
- white $\rightarrow$ C $\cdot$ G;
- grey $\rightarrow$ A $\cdot$ U or U $\cdot$ A.

Proper coloring:

1. each internal node has at most one black, one white and two grey children;
2. a grey node has at most one grey child;
3. a black node does not have a white child; and
4. a white node does not have a black child.

Level of a node = #black nodes $-$ #white nodes on the path to the root.

Separated coloring: Levels of grey nodes $\cap$ Levels of unpaired nodes $= \varnothing$

Our Results: Separated Coloring

From the tree representation T_S of structure S, color every paired node of T_S:

- black $\rightarrow$ G $\cdot$ C;
- white $\rightarrow$ C $\cdot$ G;
- grey $\rightarrow$ A $\cdot$ U or U $\cdot$ A.

Proper coloring:

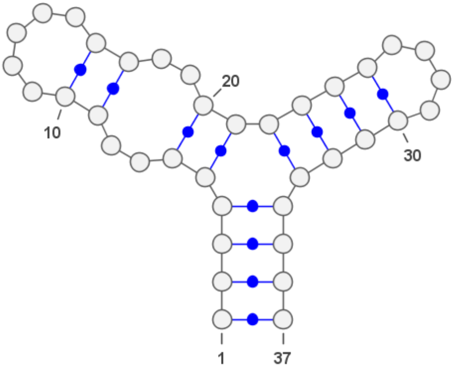
1. each internal node has at most one black, one white and two grey children;
2. a grey node has at most one grey child;
3. a black node does not have a white child; and
4. a white node does not have a black child.

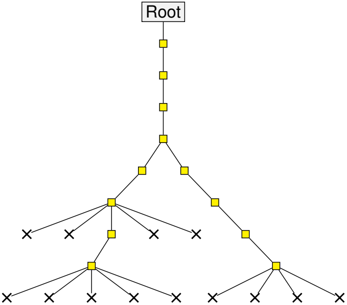
Level of a node = #black nodes $-$ #white nodes on the path to the root.

Separated coloring: Levels of grey nodes $\cap$ Levels of unpaired nodes $= \varnothing$

Our Results: Separated Coloring (example)

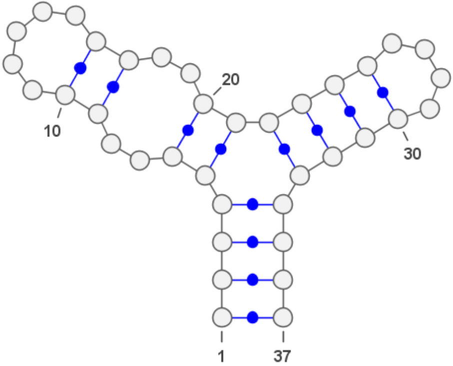
Descendant restrictions: Any node $\rightarrow \leq 1$ black & ≤ 1 White & ≤ 2 Grey;
Grey $\rightarrow 0/1$ Grey; Black $\rightarrow 0$ White; White $\rightarrow 0$ Black.
($\bullet \rightarrow$ GC $\circ \rightarrow$ CG ● $\rightarrow$ AU|UA $\times \rightarrow$ U)

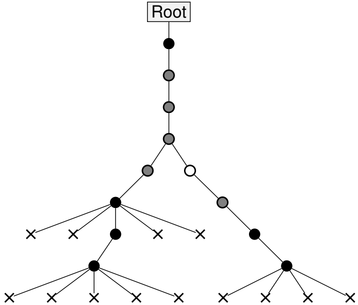




Our Results: Separated Coloring (example)

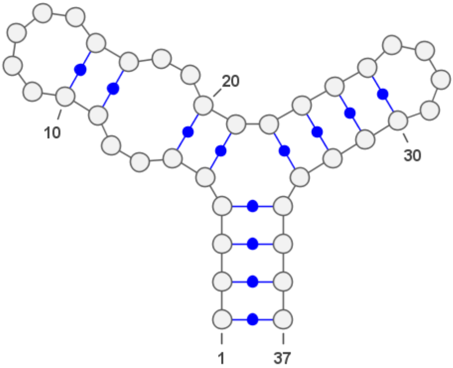
Descendant restrictions: Any node $\rightarrow \leq 1$ black & ≤ 1 White & ≤ 2 Grey;
Grey $\rightarrow 0/1$ Grey; Black $\rightarrow$ 0 White; White $\rightarrow$ 0 Black.
(● $\rightarrow$ GC ○ $\rightarrow$ CG ◉ $\rightarrow$ AU|UA × $\rightarrow$ U)

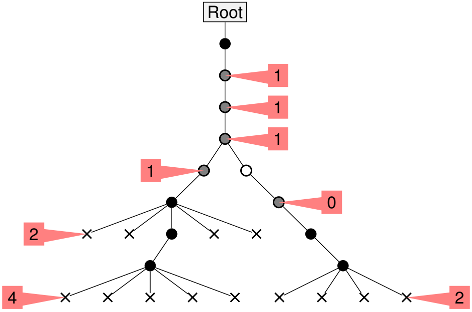




Our Results: Separated Coloring (example)

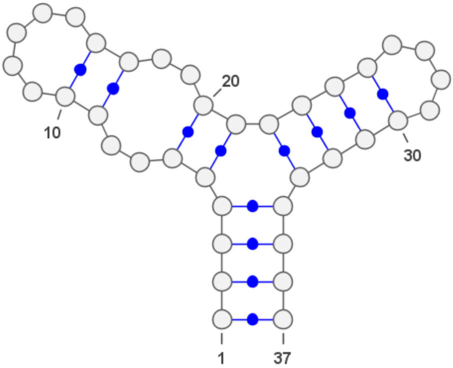
Descendant restrictions: Any node $\rightarrow \leq 1$ black & ≤ 1 White & ≤ 2 Grey;
Grey $\rightarrow 0/1$ Grey; Black $\rightarrow$ 0 White; White $\rightarrow$ 0 Black.
(● $\rightarrow$ GC ○ $\rightarrow$ CG ◉ $\rightarrow$ AU|UA × $\rightarrow$ U)

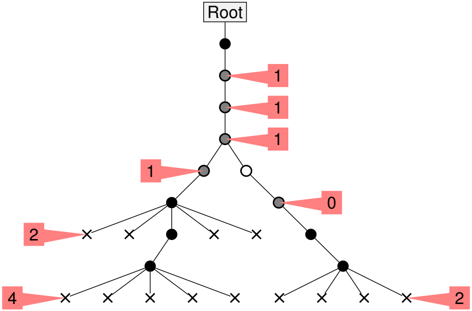




Our Results: Separated Coloring (example)

Descendant restrictions: Any node $\rightarrow \leq 1$ black & ≤ 1 White & ≤ 2 Grey;
Grey $\rightarrow 0/1$ Grey; Black $\rightarrow$ 0 White; White $\rightarrow$ 0 Black.
($\bullet \rightarrow$ GC $\circ \rightarrow$ CG ● $\rightarrow$ AU|UA $\times \rightarrow$ U)

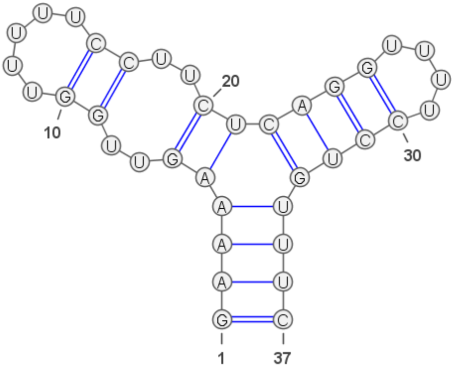


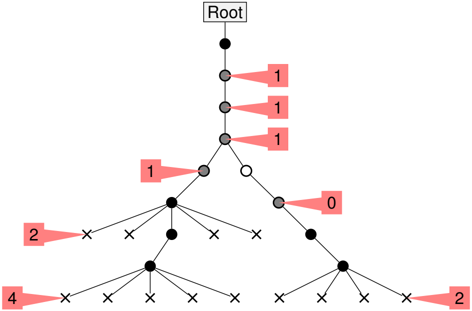


Levels of grey nodes: 0,1
Levels of leaves: 2,4
Separated coloring

Our Results: Separated Coloring (example)

Descendant restrictions: Any node $\rightarrow \leq 1$ black & ≤ 1 White & ≤ 2 Grey;
Grey $\rightarrow 0/1$ Grey; Black $\rightarrow$ 0 White; White $\rightarrow$ 0 Black.
($\bullet \rightarrow$ GC $\circ \rightarrow$ CG ● $\rightarrow$ AU|UA $\times \rightarrow$ U)

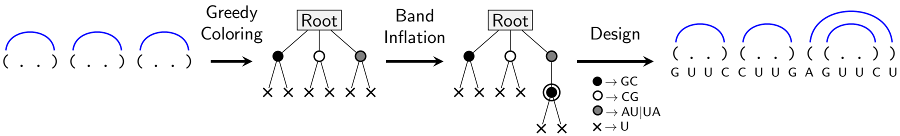




Levels of grey nodes: 0,1
Levels of leaves: 2,4
Separated coloring

$\Rightarrow$ **Design:** GAAAAGUUGGUUUUUCCUUCUCAGGUUUUCCUGUUUC

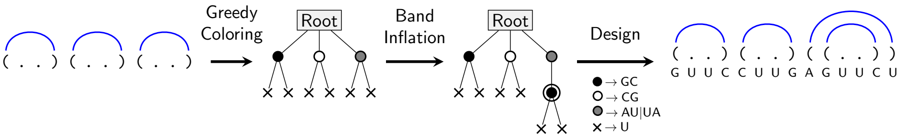
R8 Any structure S without m_5 and $m_{3\circ}$ can be transformed in $\Theta(n)$ time into a designable structure S', by adding at most a single base-pair to its helices.



Main idea: Offset grey vertices and leaves to odd/even levels
→ Coloring is now **separated**

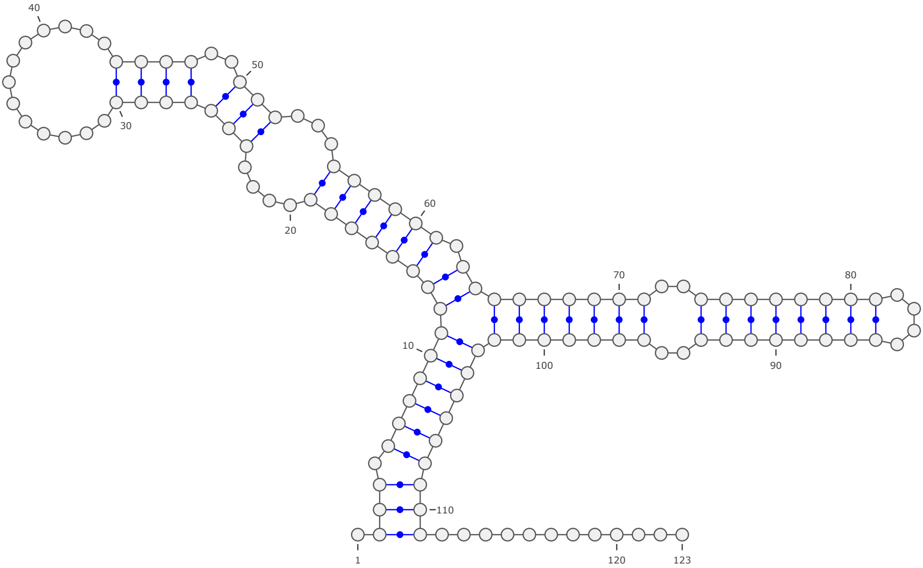
Our Results: Structure-Approximating Algorithm

R8 Any structure S without m_5 and $m_{3\circ}$ can be transformed in $\Theta(n)$ time into a designable structure S', by adding at most a single base-pair to its helices.

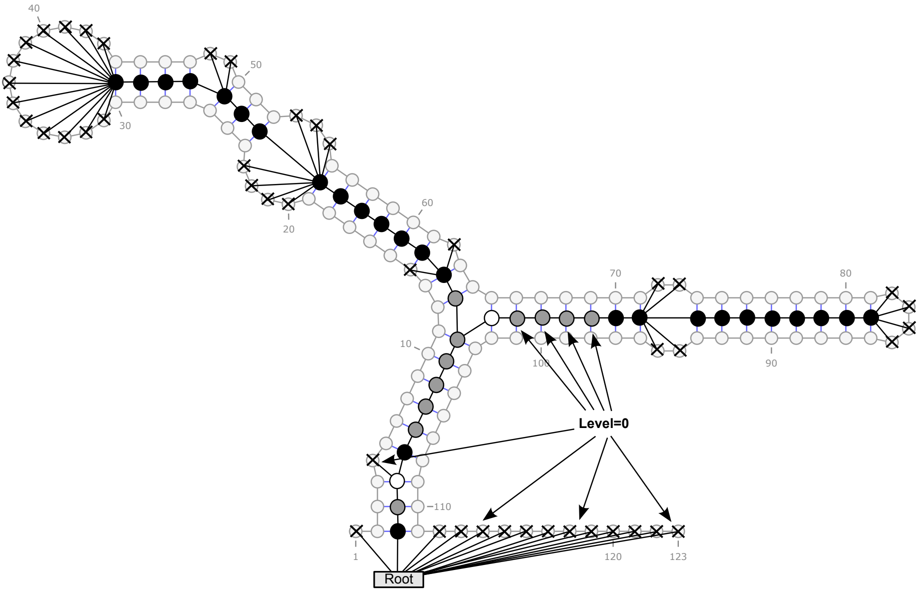


Main idea: Offset grey vertices and leaves to odd/even levels
→ Coloring is now **separated**

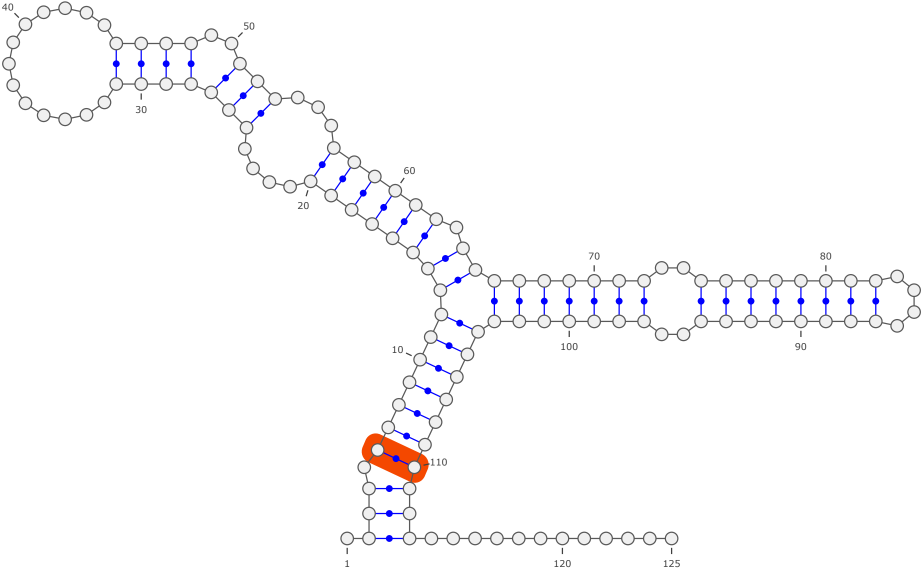
Example



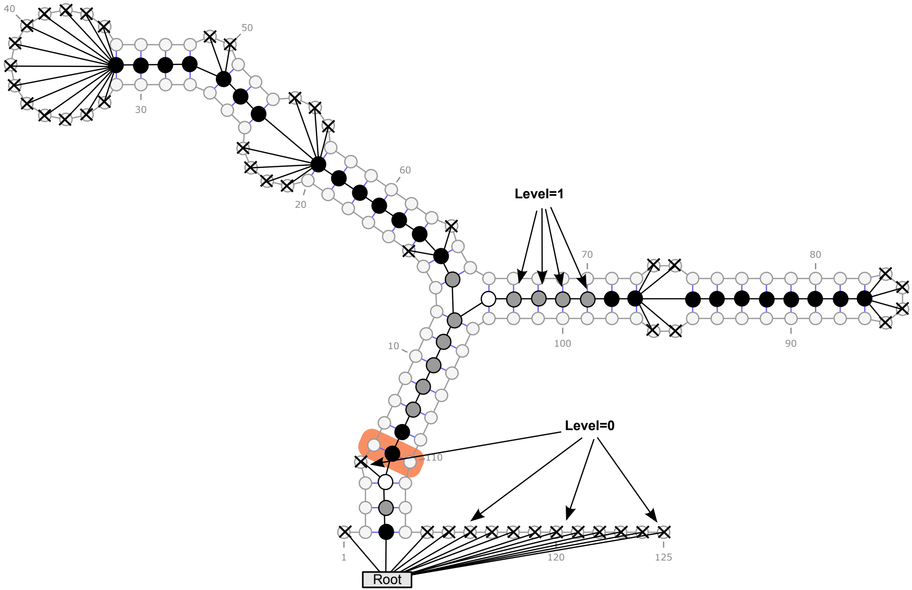
Example



Example



Example



Generalization

Theorem

All the above results hold in any energy models $\mathcal{M}$:

$$E_{\mathcal{M}}(X, Y) = \begin{cases} \alpha & \textit{if } \{X, Y\} = \{\mathrm{G}, \mathrm{C}\} \\ \beta & \textit{if } \{X, Y\} = \{\mathrm{A}, \mathrm{U}\} \\ \gamma & \textit{if } \{X, Y\} = \{\mathrm{G}, \mathrm{U}\} \\ +\infty & \textit{otherwise} \end{cases}$$

such that $\alpha, \beta > \gamma$.

Proof idea: Our results are based on (G, C)-saturated sequences
No $\mathrm{G} - \mathrm{U}$ base pair in optimal fold, since $\alpha > \gamma$.
Numbers of $\mathrm{G} - \mathrm{C}$ and $\mathrm{A} - \mathrm{U}$ base pairs are upper-bounded.
$\Rightarrow$ Any alternative has same number of each base-pair as target structure.

Remarks

- Results also hold in **Nussinov** energy model ($A-U, G-C, G-U$ + weights)
  $\Rightarrow$**Stacking** energy model? **Turner**?

- Characterized classes are mostly **easy**:
  - **Designable** classes $\rightarrow$ Linear time **algorithms**
  - **Non-designable** classes $\rightarrow$ Linear time **membership tests**

- **Forbidden local motifs** (*e.g.* m_5 & $m_{3\circ}$) can be found in any energy model
  $\Rightarrow$ **Designable structures** $\subset$ **Tree-like** objects with **forbidden motifs**
  Proportion of designable structures: $\left(\frac{\beta}{\alpha}\right)^n$, **exponentially decreasing** with n.
  Possible consequences on **RNA neutral network** studies
  + motivation for identifying **new forbidden motifs**

Remarks

- Results also hold in **Nussinov** energy model ($A-U, G-C, G-U$ + weights)
  $\Rightarrow$**Stacking** energy model? **Turner**?

- Characterized classes are mostly **easy**:
  - **Designable** classes $\rightarrow$ Linear time **algorithms**
  - **Non-designable** classes $\rightarrow$ Linear time **membership tests**

- **Forbidden local motifs** (*e.g.* m_5 & $m_{3\,\circ}$) can be found in any energy model
  $\Rightarrow$ **Designable structures** $\subset$ **Tree-like** objects with **forbidden motifs**
  Proportion of designable structures: $\left(\frac{\beta}{\alpha}\right)^n$, **exponentially decreasing** with n.
  Possible consequences on **RNA neutral network** studies
  + motivation for identifying **new forbidden motifs**

Remarks

- Results also hold in **Nussinov** energy model ($A-U, G-C, G-U$ + weights)
  $\Rightarrow$**Stacking** energy model? **Turner**?

- Characterized classes are mostly **easy**:
  - **Designable** classes $\rightarrow$ Linear time **algorithms**
  - **Non-designable** classes $\rightarrow$ Linear time **membership tests**

- **Forbidden local motifs** (*e.g.* m_5 & $m_{3\circ}$) can be found in any energy model
  $\Rightarrow$ **Designable structures** $\subset$ **Tree-like** objects with **forbidden motifs**
  Proportion of designable structures: $\left(\frac{\beta}{\alpha}\right)^n$, **exponentially decreasing** with n.
  Possible consequences on **RNA neutral network** studies
  + motivation for identifying **new forbidden motifs**

Conclusions

- **RNA design** is one of the current challenge of RNA bioinformatics with far-reaching consequences for drug design, synthetic biology. . .

- **RNA inverse folding** is the combinatorial core of design.
  It remains **largely unsolved**, and opens **new lines of research** in Comp. Sci.

Thanks

LIX – Ecole Polytechnique

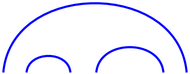
Yann Ponty
Mireille Regnier

Amélie Héliou
Afaf Saaidi
Juraj Michalik

Simon Fraser University

Jozef Haleš
Ján Maňuch
Ladislav Stacho

Our Results: *k*-Stutter (example)



Designable structure: ((.) (. .))

Then 2-stutter is designable as well:

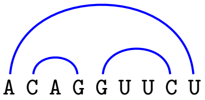
Our Results: *k*-Stutter (example)

Designable structure: A C A G G U U C U

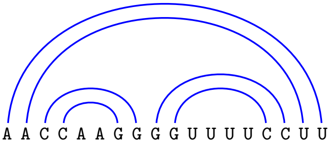
Then 2-stutter is designable as well: ((((. .)) ((. . . .))))

Our Results: *k*-Stutter (example)

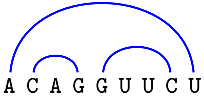
Designable structure:



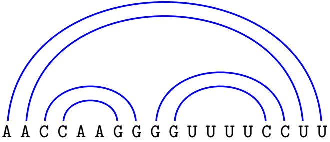
Then 2-stutter is designable as well:



Designable structure:



Then 2-stutter is designable as well:

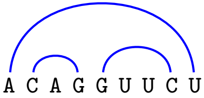


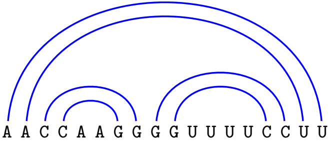
Proof idea: w: Design for S; $S' \neq S^{[k]}$: Alternative folding for k-stutter $w^{[k]}$:

- Compact k consecutive positions $\rightarrow$ Multigraph G such that $\Delta(G) = k$
- Base-pair compatibility graph is bipartite $\rightarrow$ G is also bipartite
- Therefore G is k edge-colorable
- Any restriction of G to a given color c = Valid structure S_c for w
- Either $E_{\mathcal{M}}(S_c) = E_{\mathcal{M}}(S)$ ($\Rightarrow S_c = S$), or $E_{\mathcal{M}}(S_c) > E_{\mathcal{M}}(S)$ (holds for some c)
- Thus $\sum_c E_{\mathcal{M}}(S_c) > k \cdot E(S) = E(S^{[k]})$

$\Rightarrow$ $w^{[k]}$ is design for $S^{[k]}$ (holds for any base-pair additive $\mathcal{M}$)

Our Results: k-Stutter (example)

Designable structure: 

Then 2-stutter is designable as well: 

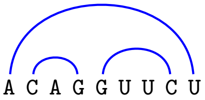
Proof idea: w: Design for S; $S' \neq S^{[k]}$: Alternative folding for k-stutter $w^{[k]}$:

- Compact k consecutive positions $\rightarrow$ Multigraph G such that $\Delta(G) = k$
- Base-pair compatibility graph is bipartite $\rightarrow$ G is also bipartite
- Therefore G is k edge-colorable
- Any restriction of G to a given color c = Valid structure S_c for w
- Either $E_{\mathcal{M}}(S_c) = E_{\mathcal{M}}(S)$ ($\Rightarrow S_c = S$), or $E_{\mathcal{M}}(S_c) > E_{\mathcal{M}}(S)$ (holds for some c)
- Thus $\sum_c E_{\mathcal{M}}(S_c) > k \cdot E(S) = E(S^{[k]})$

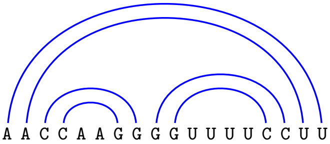
$\Rightarrow$ $w^{[k]}$ is design for $S^{[k]}$ (holds for any base-pair additive $\mathcal{M}$)

Our Results: k-Stutter (example)

Designable structure:



Then 2-stutter is designable as well:



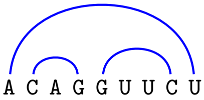
Proof idea: w: Design for S; $S' \neq S^{[k]}$: Alternative folding for k-stutter $w^{[k]}$:

- Compact k consecutive positions $\rightarrow$ Multigraph G such that $\Delta(G) = k$
- Base-pair compatibility graph is bipartite $\rightarrow$ G is also bipartite
- Therefore G is k edge-colorable
- Any restriction of G to a given color c = Valid structure S_c for w
- Either $E_{\mathcal{M}}(S_c) = E_{\mathcal{M}}(S)$ ($\Rightarrow S_c = S$), or $E_{\mathcal{M}}(S_c) > E_{\mathcal{M}}(S)$ (holds for some c)
- Thus $\sum_c E_{\mathcal{M}}(S_c) > k \cdot E(S) = E(S^{[k]})$

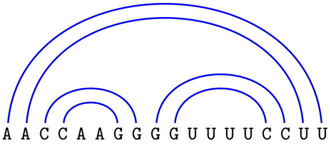
$\Rightarrow$ $w^{[k]}$ is design for $S^{[k]}$ (holds for any base-pair additive $\mathcal{M}$)

Our Results: k-Stutter (example)

Designable structure:



Then 2-stutter is designable as well:

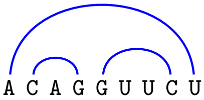


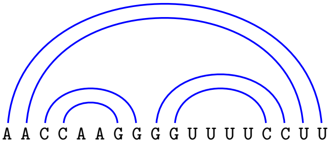
Proof idea: w: Design for S; $S' \neq S^{[k]}$: Alternative folding for k-stutter $w^{[k]}$:

- Compact k consecutive positions $\rightarrow$ Multigraph G such that $\Delta(G) = k$
- Base-pair compatibility graph is bipartite $\rightarrow$ G is also bipartite
- Therefore G is k edge-colorable
- Any restriction of G to a given color c = Valid structure S_c for w
- Either $E_{\mathcal{M}}(S_c) = E_{\mathcal{M}}(S)$ ($\Rightarrow S_c = S$), or $E_{\mathcal{M}}(S_c) > E_{\mathcal{M}}(S)$ (holds for some c)
- Thus $\sum_c E_{\mathcal{M}}(S_c) > k \cdot E(S) = E(S^{[k]})$

$\Rightarrow$ $w^{[k]}$ is design for $S^{[k]}$ (holds for any base-pair additive $\mathcal{M}$)

Our Results: k-Stutter (example)

Designable structure: 

Then 2-stutter is designable as well: 

Proof idea: w: Design for S; $S' \neq S^{[k]}$: Alternative folding for k-stutter $w^{[k]}$:

- Compact k consecutive positions $\rightarrow$ Multigraph G such that $\Delta(G) = k$
- Base-pair compatibility graph is bipartite $\rightarrow$ G is also bipartite
- Therefore G is k edge-colorable
- Any restriction of G to a given color c = Valid structure S_c for w
- Either $E_{\mathcal{M}}(S_c) = E_{\mathcal{M}}(S)$ ($\Rightarrow S_c = S$), or $E_{\mathcal{M}}(S_c) > E_{\mathcal{M}}(S)$ (holds for some c)
- Thus $\sum_c E_{\mathcal{M}}(S_c) > k \cdot E(S) = E(S^{[k]})$

$\Rightarrow$ $w^{[k]}$ is design for $S^{[k]}$ (holds for any base-pair additive $\mathcal{M}$)