

Parameterized algorithms for the RNA energy barrier problem

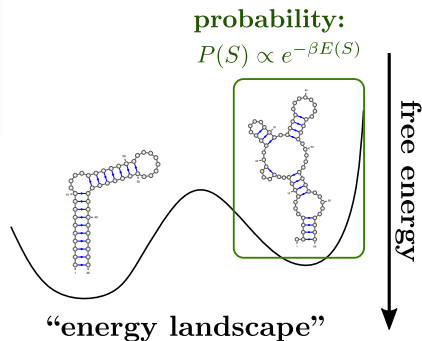
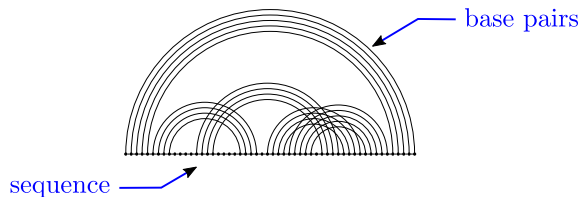
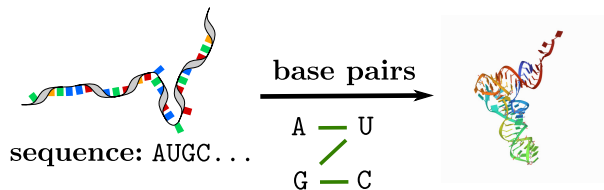
Théo Boury (ENS Lyon), Laurent Bulteau (LIGM), Bertrand Marchand (LIX) and
Yann Ponty (LIX)

SeqBIM 2022

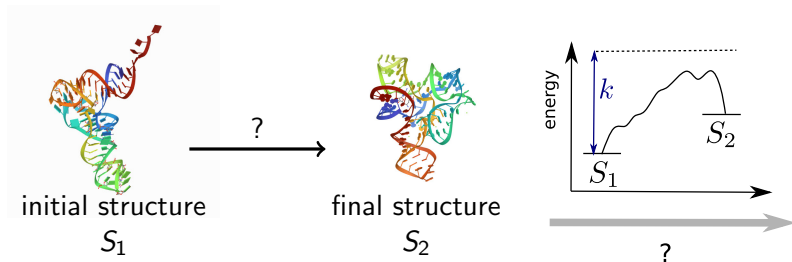
November 18, 2022

Introduction: RNA structure

- ▶ RNA structure consists of base pairs
- ▶ form spontaneously to **lower free energy**



The RNA energy barrier problem

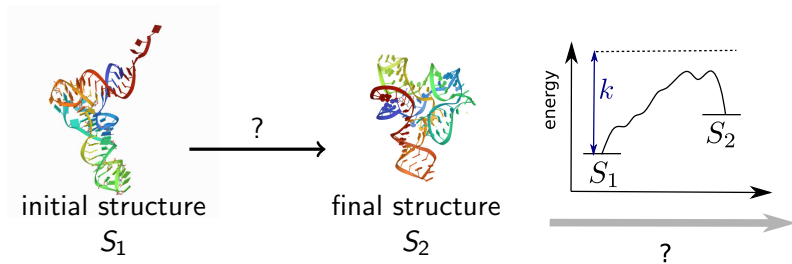


RNA BARRIER

Input: two structures S_1 , S_2 , integer k

Question: Is there a sequence of reconfiguration moves connecting S_1 and S_2 staying below $E(S_1) + k$?

The RNA energy barrier problem

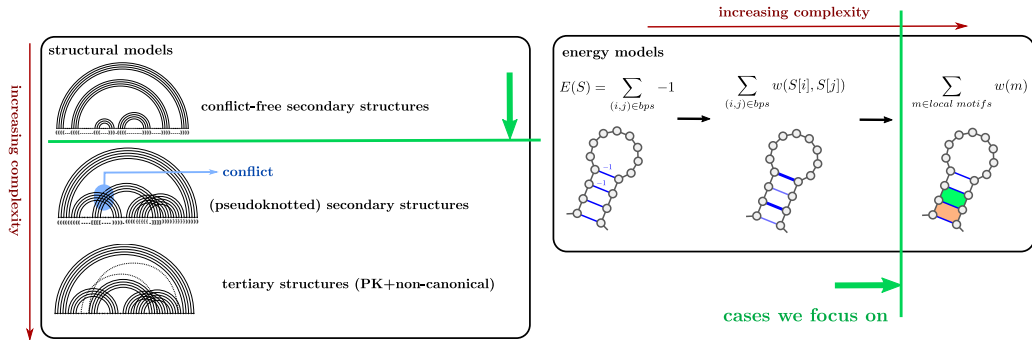


Question: could we use **parameterized** algorithmics ?

→ i.e. looking for **FPT** → $f(p) \cdot \text{poly}(n)$ or **XP** → $O(n^{g(p)})$

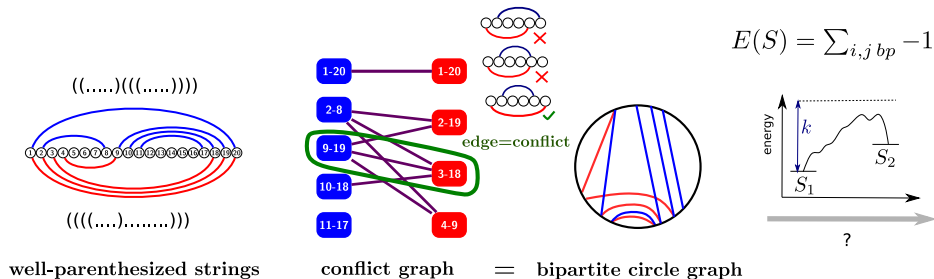
parameter

Structure classes and energy models



- ▶ We restrict the problem to **conflict-free secondary structures** and **base-pair-based energy models**
- ▶ Already **NP-hard** [Mañuch et al., 2011]

Conflict graph

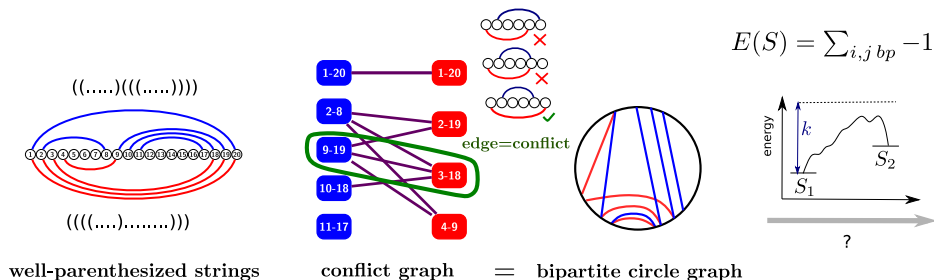


RNA BARRIER

Input: two structures S_1 , S_2 , integer k

Question: Is there a sequence of reconfiguration moves connecting S_1 and S_2 staying below $E(S_1) + k$?

Conflict graph



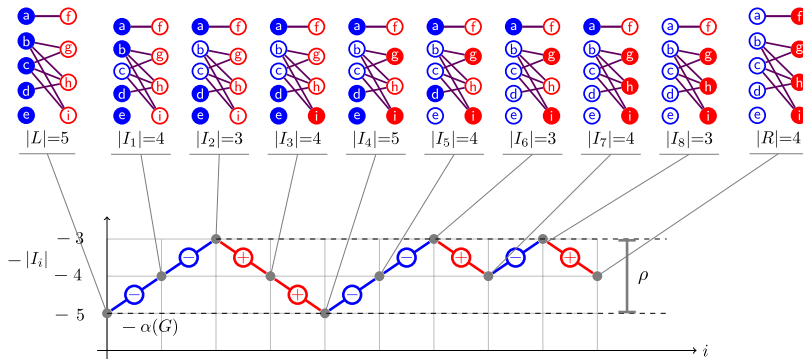
BIPARTITE INDEPENDENT SET RECONFIGURATION

Input: bipartite circle graph G with parts L, R, ρ

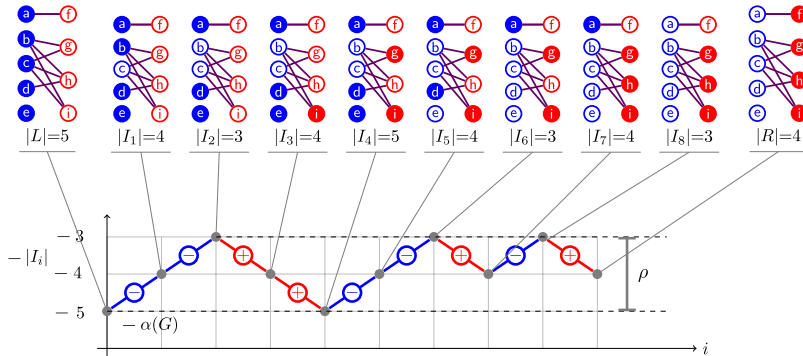
Question: Is there a sequence $I_1 = L, \dots, I_l = R$ of independent sets of G , s.t

- ▶ $I_i \Delta I_{i+1} = 1$
- ▶ $\forall i |I_i| \geq \alpha(G) - \rho$

Example



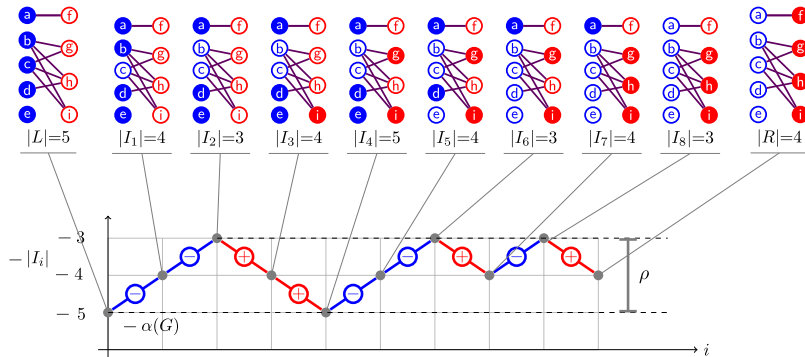
Example



- ▶ Processing each vertex once is enough [Lokshtanov and Mouawad, 2018].
- ▶ **minimum cumulative-cost ordering**

$$\text{barrier}(G) = \min_{\text{valid permutation } \sigma} \max_{X \sqsubseteq \sigma} \left(\sum_{x \in X} c(x) \right)$$

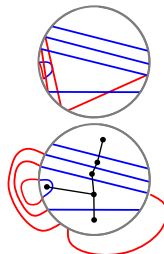
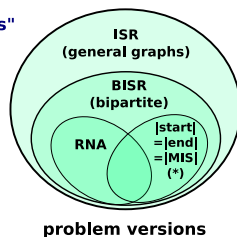
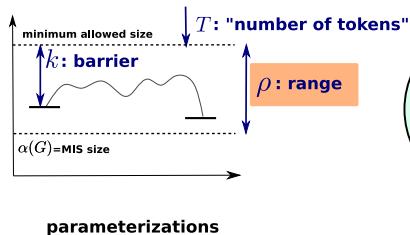
Example



- ▶ Processing each vertex once is enough [Lokshtanov and Mouawad, 2018].
- ▶ **minimum cumulative-cost ordering**

$$\text{barrier}(G) = \min_{\text{valid permutation } \sigma} \max_{X \sqsubseteq \sigma} (|X \cap B| - |X \cap R|)$$

Parameterizations and state of the art



NEW ! ← Φ = number of leaves

- ▶ “reconfiguration” literature → parameter T [Ito et al., 2020]
- ▶ Bioinformatics literature → XP algorithm for parameter k but in $(*)$ case [Thachuk et al., 2010], and **heuristics**.

We studied ρ : the **range** [Bulteau et al., 2021] and the **arboricity** Φ (Théo Boury’s internship)

Our results

problem parameter	$ \text{start} = \text{end} = \text{MIS} $	RNA	BISR
T	not adapted to biology		?
k	XP (n^{2k})	?	?
ρ	?	?	?
$\Phi = \min(\Phi_1, \Phi_2)$?	?	not defined

↓ our work

problem parameter	$ \text{start} = \text{end} = \text{MIS} $	RNA	BISR
T	not adapted to biology		?
k	XP - $O(n^{k+1})$	Para-NP-hard	
ρ	XP - $O(n^{\rho+1})$		
Φ	XP - $(O(f(\Phi)n^{2\Phi+5}))$		not defined
$\rho + \Phi' = \max(\Phi_1, \Phi_2)$	FPT - $O(\Phi'^{3\rho} \cdot n^{O(1)})$		not defined

Our results

problem parameter	$ \text{start} = \text{end} = \text{MIS} $	RNA	BISR
T	not adapted to biology		?
k	XP (n^{2k})	?	?
ρ	?	?	?
$\Phi = \min(\Phi_1, \Phi_2)$?	?	not defined

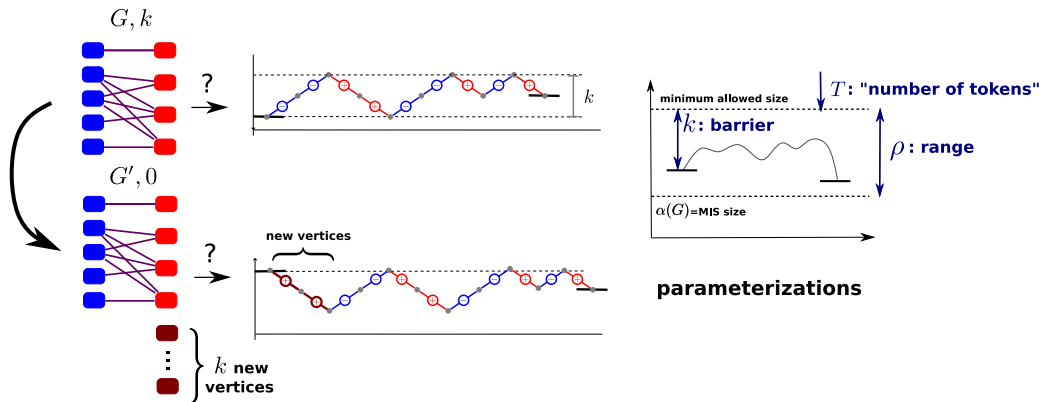
↓ our work

problem parameter	$ \text{start} = \text{end} = \text{MIS} $	RNA	BISR
T	not adapted to biology		?
k	XP - $O(n^{k+1})$	Para-NP-hard	
ρ	XP - $O(n^{\rho+1})$		
Φ	XP - $(O(f(\Phi)n^{2\Phi+5}))$		not defined
$\rho + \Phi' = \max(\Phi_1, \Phi_2)$	FPT - $O(\Phi^{13\rho} \cdot n^{O(1)})$		not defined

► FPT for ρ alone or Φ alone still open

Para-NP-hardness of BISR for barrier parameter

- ▶ Deciding whether a (RNA) bipartite graph has barrier 0 or not is NP-hard → **Para-NP-hardness**
- ▶ **Proof:** reduction from general case.

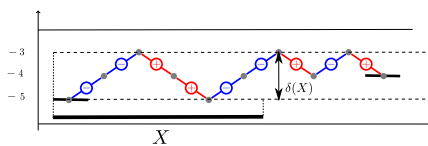


mixed Maximum Independent Set

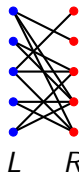
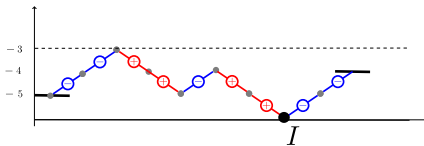
If a **mixed MIS** I exists: can be used as **splitting point**:

$$\text{barrier}(L \rightarrow R) \leq k \Leftrightarrow \text{barrier}(L \rightarrow I) \leq k \ \& \ \text{barrier}(I \rightarrow R) \leq k - \delta(X_I)$$

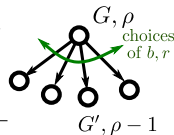
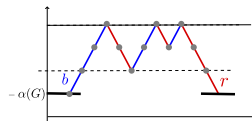
with $X_I = (L \setminus I) \cup (R \cap I)$ and $\delta(X_I) = |B \cap X_I| - |R \cap X_I|$



reshuffle

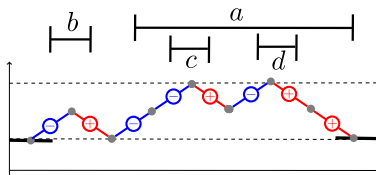
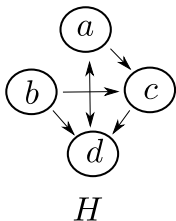
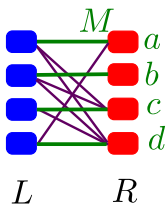


→ allows for $O(n^{2\rho})$ algorithm (**bounded search trees**)



Correspondance with directed pathwidth

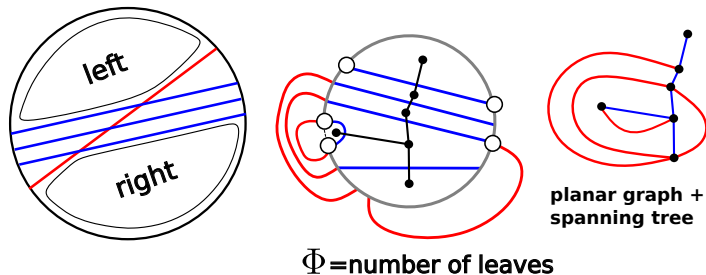
- ▶ Consider G bipartite graph with a **perfect matching**¹ M .
→ **directed graph** H from matching M :



$(l, r) \rightarrow (l', r')$ iff $(l, r') \in E(G)$ and G, ρ BISR yes-instance $\Leftrightarrow dpw(H) \leq \rho$

¹with very little work also works for general case

Arboricity parameterization

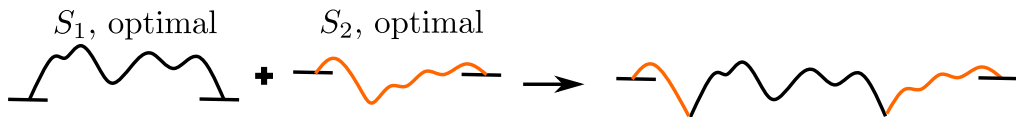


- ▶ \exists poly algorithm when $\Phi = 2$ [Rafiey et al., 2015] \rightarrow M2 internship of **Théo Boury** exploring this parameterization ($\rightarrow O(n^{2\Phi+5})$ generalization)

$$\text{barrier}(G) = \min_{r \in R} [\text{barrier}(N(r)) \cdot r \cdot \text{MERGE}(\text{left}(r), \text{right}(r))]$$

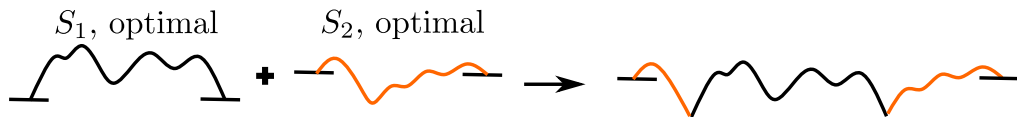
\rightarrow **Dynamic Programming over subtrees**

MERGE procedure



solved in [Abdel-Wahab and Kameda, 1980, Rafiey et al., 2015]

MERGE procedure



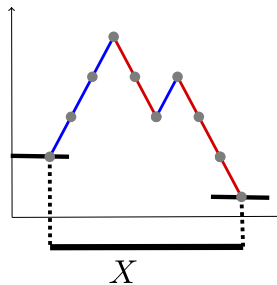
solved in [Abdel-Wahab and Kameda, 1980, Rafiey et al., 2015]

We call **k -safe unitary set** a subset $X \subseteq G$ s.t:

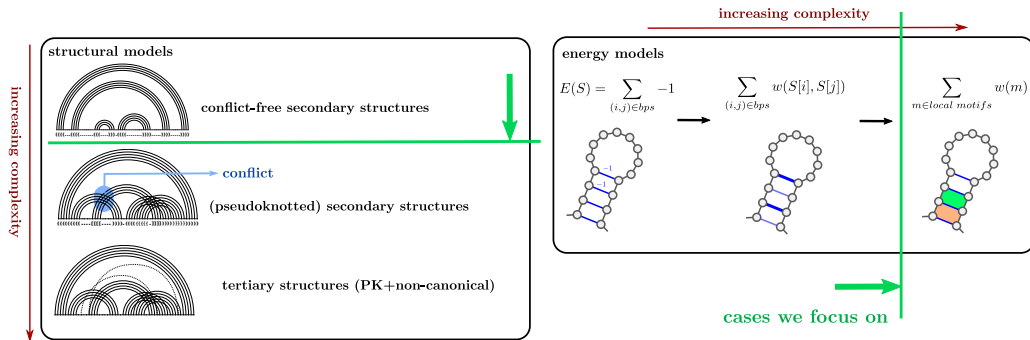
- (1) X is k -feasible
- (2) $\delta(X) \leq 0$
- (3) $\forall X' \subset X, \delta(X') > 0$

Lemma 1: (1)+(2)+(3) \Rightarrow starting with X is safe

Lemma 2: they are **subtrees** $\Rightarrow O(n^\Phi)$ enumeration

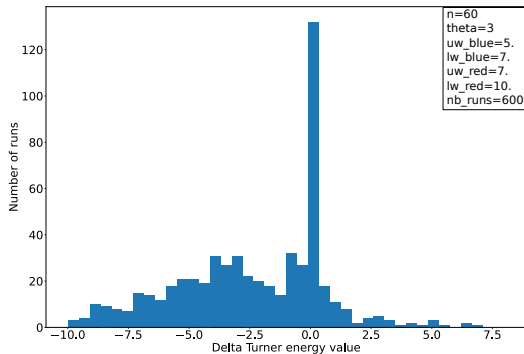


Generalizations



- ▶ Both algorithms are adaptable to **weighted** base pairs
- ▶ Turner model ? → change of paradigm, future work.

Numerical results from implementation



- ▶ Using **RNAEval** (Vienna Package) to score a schedule according to Turner model.

$$\Delta = E_{\text{Turner}}(\text{Weighted model}) - E_{\text{Turner}}(\text{Unitary model})$$

- ▶ Weighted model does provide better schedules






Conclusion

- ▶ Tackling an RNA kinetics problem through parameterized algorithmics
- ▶ connections with reconfiguration, directed pathwidth and scheduling
- ▶ Still a lot of **open problems**
- ▶ For more realistic biological models → new framework needed

Conclusion



- ▶ Tackling an RNA kinetics problem through parameterized algorithmics
- ▶ connections with reconfiguration, directed pathwidth and scheduling
- ▶ Still a lot of **open problems**
- ▶ For more realistic biological models → new framework needed

Thank you

-  Abdel-Wahab, H. M. and Kameda, T. (1980).
On strictly optimal schedules for the cumulative cost-optimal scheduling problem.
[Computing](#), 24(1):61–86.
-  Bulteau, L., Marchand, B., et al. (2021).
A new parametrization for independent set reconfiguration and applications to rna kinetics.
In [IPEC 2021-16th International Symposium on Parameterized and Exact Computation](#).
-  Ito, T., Kamiński, M., Ono, H., Suzuki, A., Uehara, R., and Yamanaka, K. (2020).
Parameterized complexity of independent set reconfiguration problems.
[Discrete Applied Mathematics](#), 283:336–345.
-  Lokshtanov, D. and Mouawad, A. E. (2018).
The complexity of independent set reconfiguration on bipartite graphs.
[ACM Transactions on Algorithms \(TALG\)](#), 15(1):1–19.
-  Maňuch, J., Thachuk, C., Stacho, L., and Condon, A. (2011).

N_p -completeness of the energy barrier problem without pseudoknots and temporary arcs.

[Natural Computing](#), 10(1):391–405.

-  Rafiey, A., Kinne, J., Manuch, J., and Rafiey, A. (2015). Ordering with precedence constraints and budget minimization. [arXiv preprint arXiv:1507.04885](#).
-  Thachuk, C., Maňuch, J., Rafiey, A., Mathieson, L.-A., Stacho, L., and Condon, A. (2010). An algorithm for the energy barrier problem without pseudoknots and temporary arcs. In [Biocomputing 2010](#), pages 108–119. World Scientific.