

siRNA Design and RNA-RNA Interaction Algorithms

Saad Mneimneh

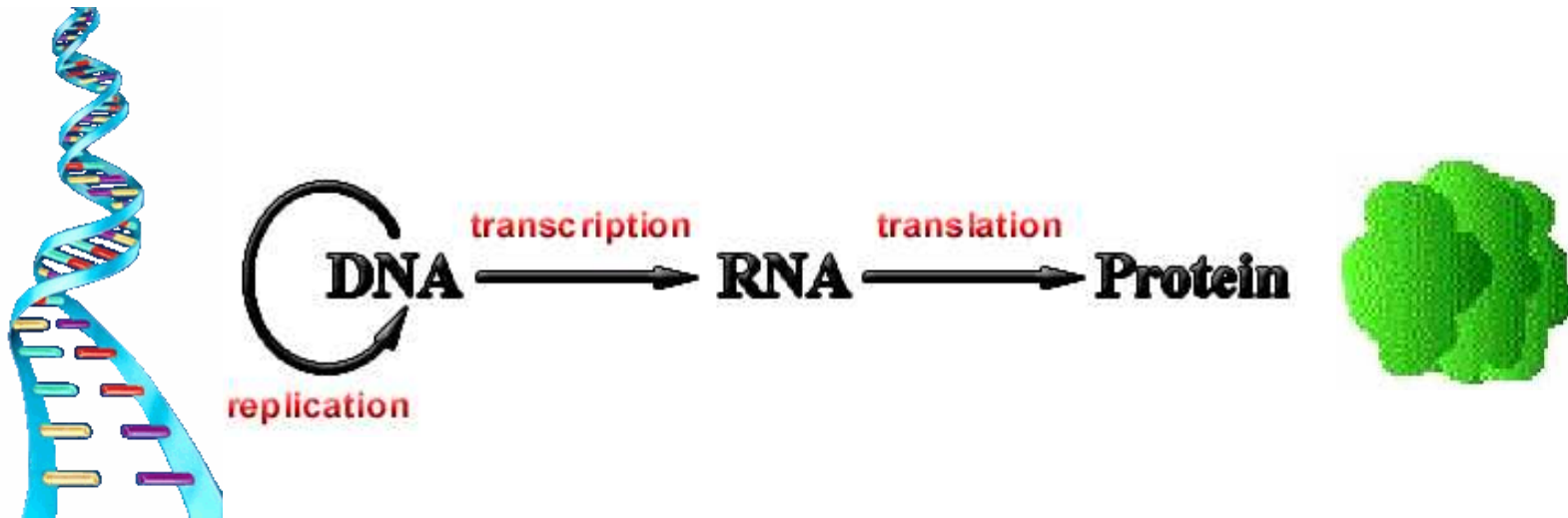
Assistant Professor
Computer Science and Engineering
SMU

Ph.D. MIT 2002

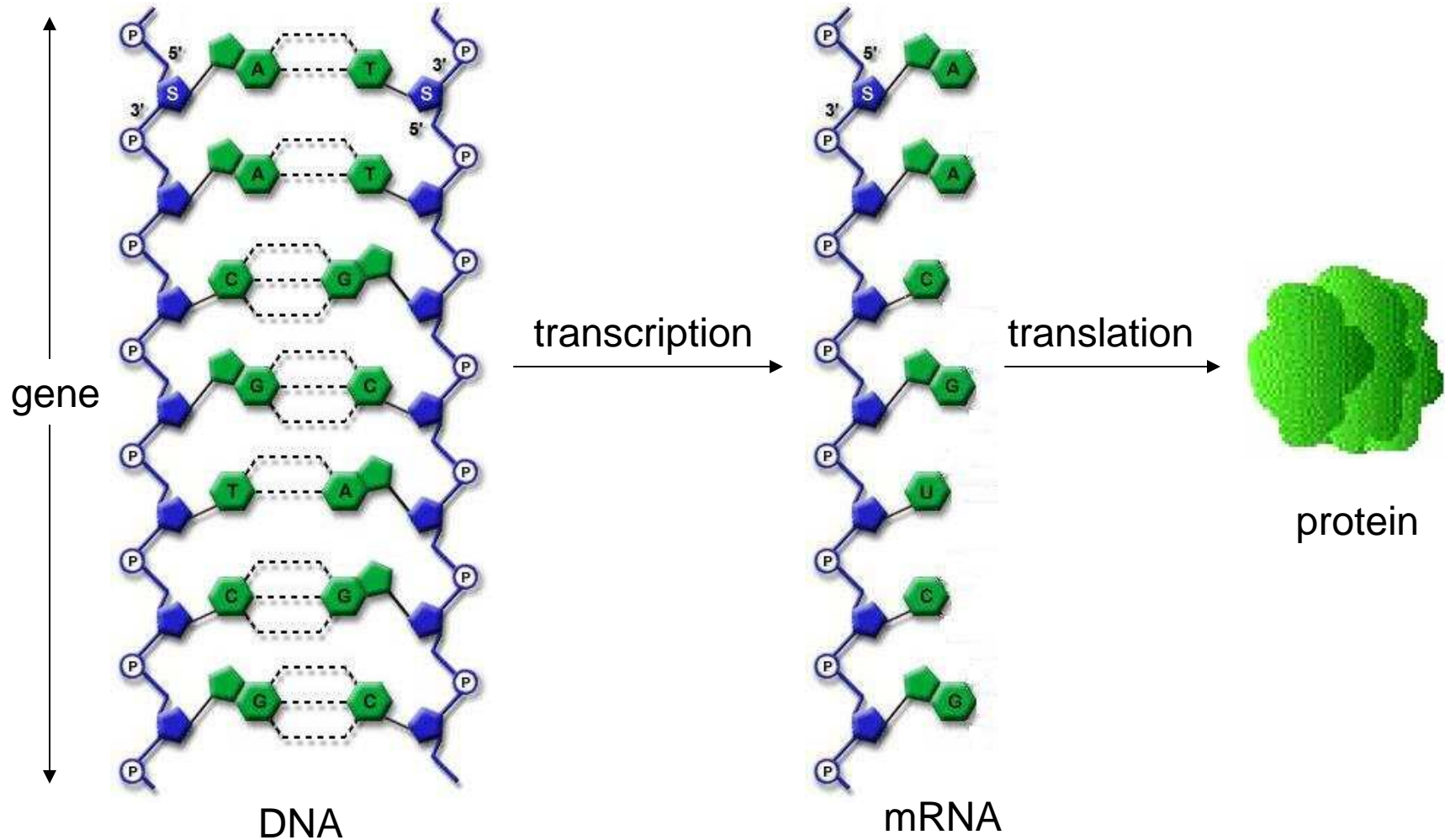
S.M MIT 1997

B.E. AUB 1995

Biology in one slide



... well, in two slides



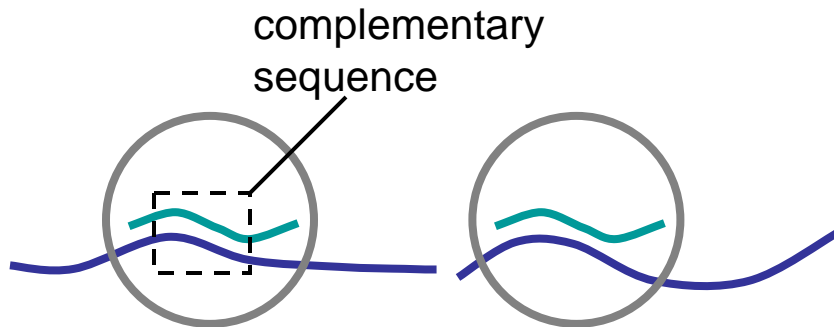
Gene Silencing

(post-transcriptional gene silencing PTGS)



1 small interfering RNA
19-23 nucleotides
introduced

2 siRNA combines with RISC
(RNA Induced Silencing Complex)



3 Binding with target mRNA

4 Target mRNA
cleavage by RISC

An important question

Given a target gene, how to design an siRNA that knocks it down by PTGS?



The simple answer...

- siRNA should be complementary to RNA sequence
- But siRNAs are small (19-23 nucleotides)
- So which part of the RNA sequence should we complement?
- Would any part work?

RNA folding

- Although RNA is a single strand, it usually folds
- The folding makes siRNA design more difficult
- siRNA has to possibly bind to the folded RNA

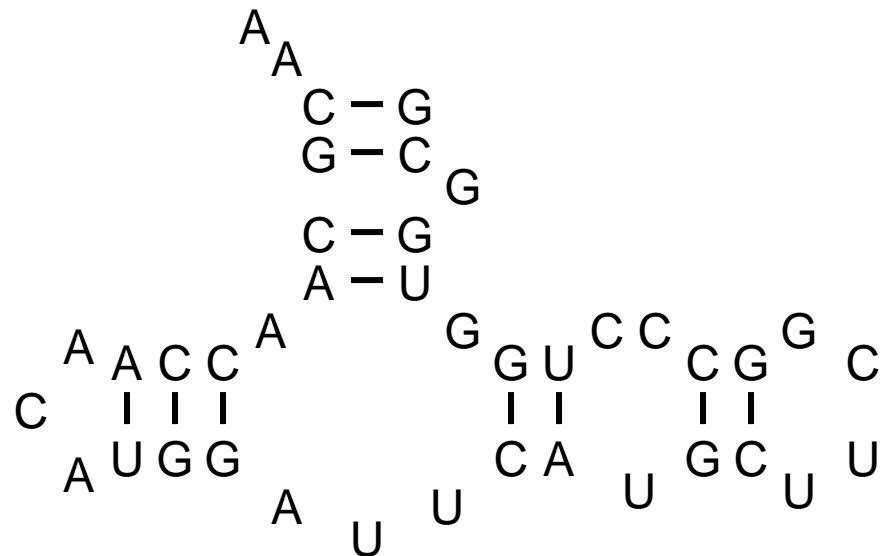
Folded RNA

AACGGAACCAACAUGGAUUCACGCUUCGGCCCUGGUCGCG

secondary structure of RNA

minimum energy structure

maximum number of bonds



Simple algorithms based on dynamic programming can solve the RNA folding problem

Two theories

- **Theory 1:** siRNA effectiveness is sequence specific and does not depend on the secondary structure of RNA
 - siRNA binds to RNA before it folds
- **Theory 2:** siRNA effectiveness depends on the secondary structure of RNA (single-stranded regions)
 - siRNA binds to RNA after it folds
- There is evidence to support both!

State of the art

- Sequence specific design rules first published by Tuschl lab at Rockefeller
- Ambion a company with some more design rules and a software
- Sfold, a software for generating a number of possible RNA secondary structures based on the work of Ding and Lawrence and identifying single stranded regions
- Combination of these and other products also exist

RNA-RNA interaction

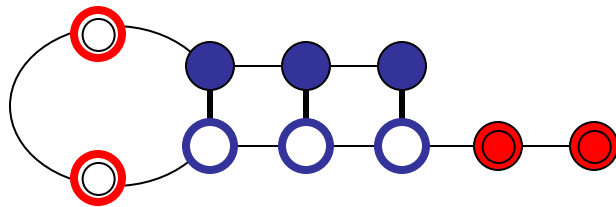
- Assume simultaneous siRNA binding and RNA folding
- Given two RNAs, what is the optimal structure in which they fold and interact simultaneously?

Toy example

RNA₁

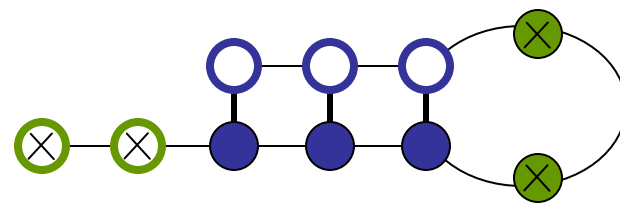


RNA₂



optimal folding of
RNA₁

6



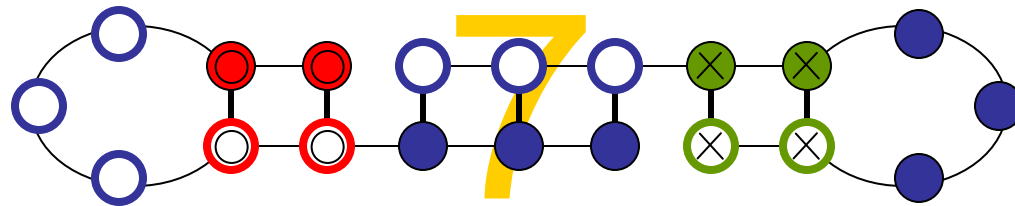
optimal folding of
RNA₂

Toy example (cont.)

RNA₁



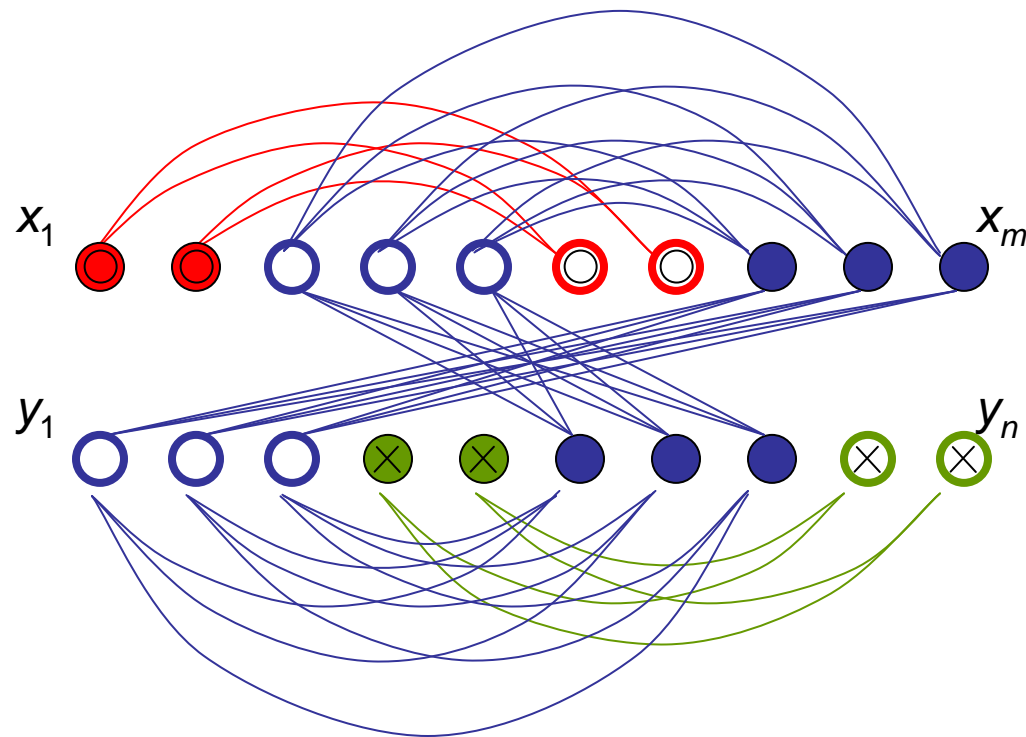
RNA₂



non-optimal folding of
RNA₁

non-optimal folding of
RNA₂

RNA-RNA interaction graph



RNA-RNAi problem

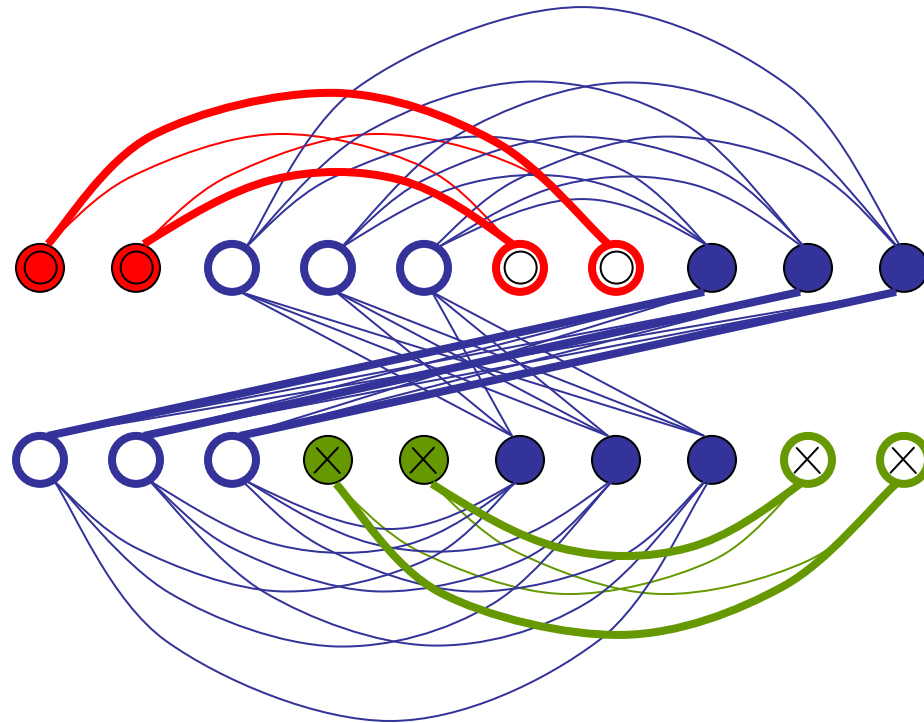
- Given an RNA-RNA interaction graph
- Find a set of edges S that are node disjoint

such that

- if $(x_i, x_j) \in S$ and $(x_k, x_l) \in S$ then NOT $i < k < j < l$
- if $(y_i, y_j) \in S$ and $(y_k, y_l) \in S$ then NOT $i < k < j < l$
- if $(x_i, y_j) \in S$ and $(x_k, y_l) \in S$ then NOT $(i < k \text{ and } j > l)$

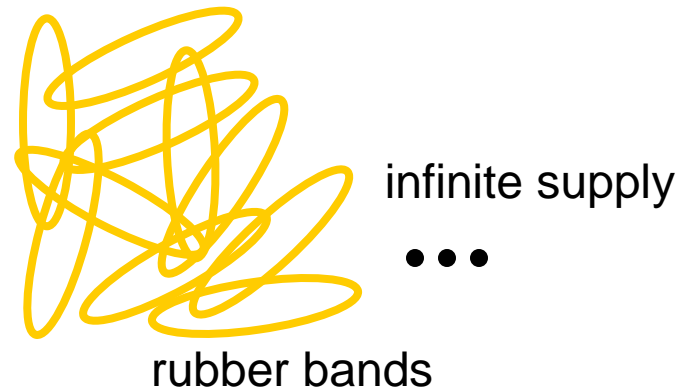
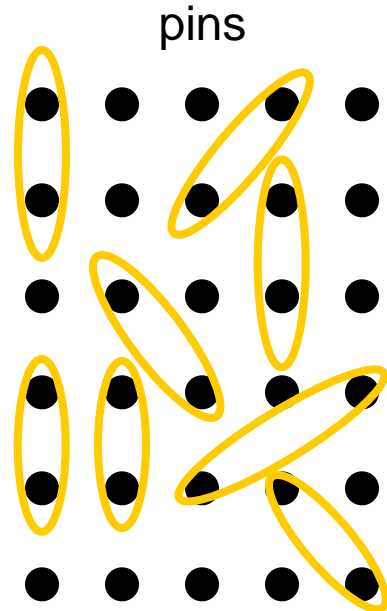
- Maximize $|S|$

Previous example



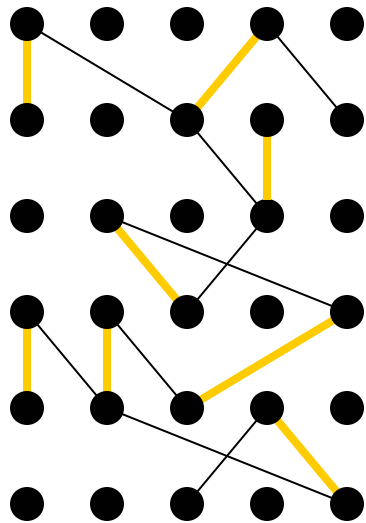
RNA-RNAi is NP-complete

Rubber Bands Game



- Each pair of pins belonging to consecutive levels contribute either 0 or 1 point
- Place rubber bands around pins, rubber band cannot stretch beyond one level
- Rubber bands cannot intersect
- How many points can you make?

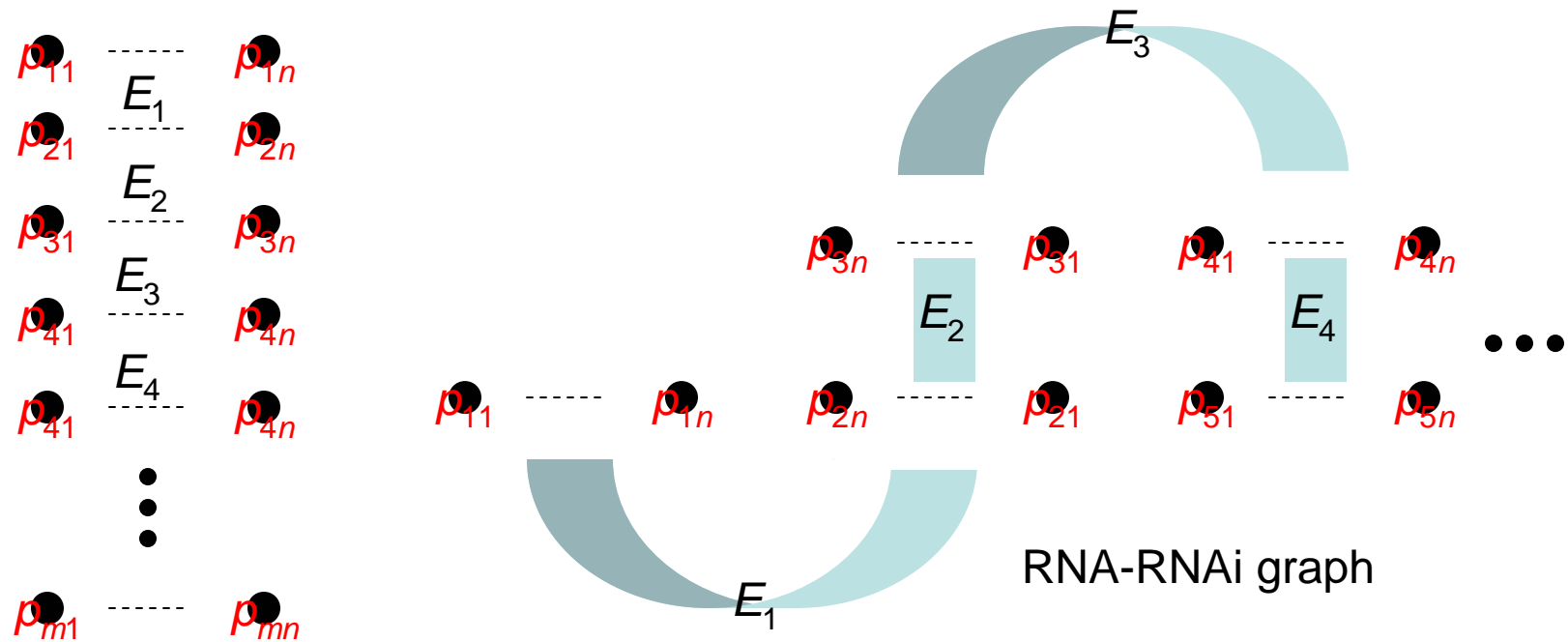
Graph representation



- An edge represents a pair of pins that contribute 1 point
- Need to find a largest set of non-intersecting edges
- This is also NP-complete

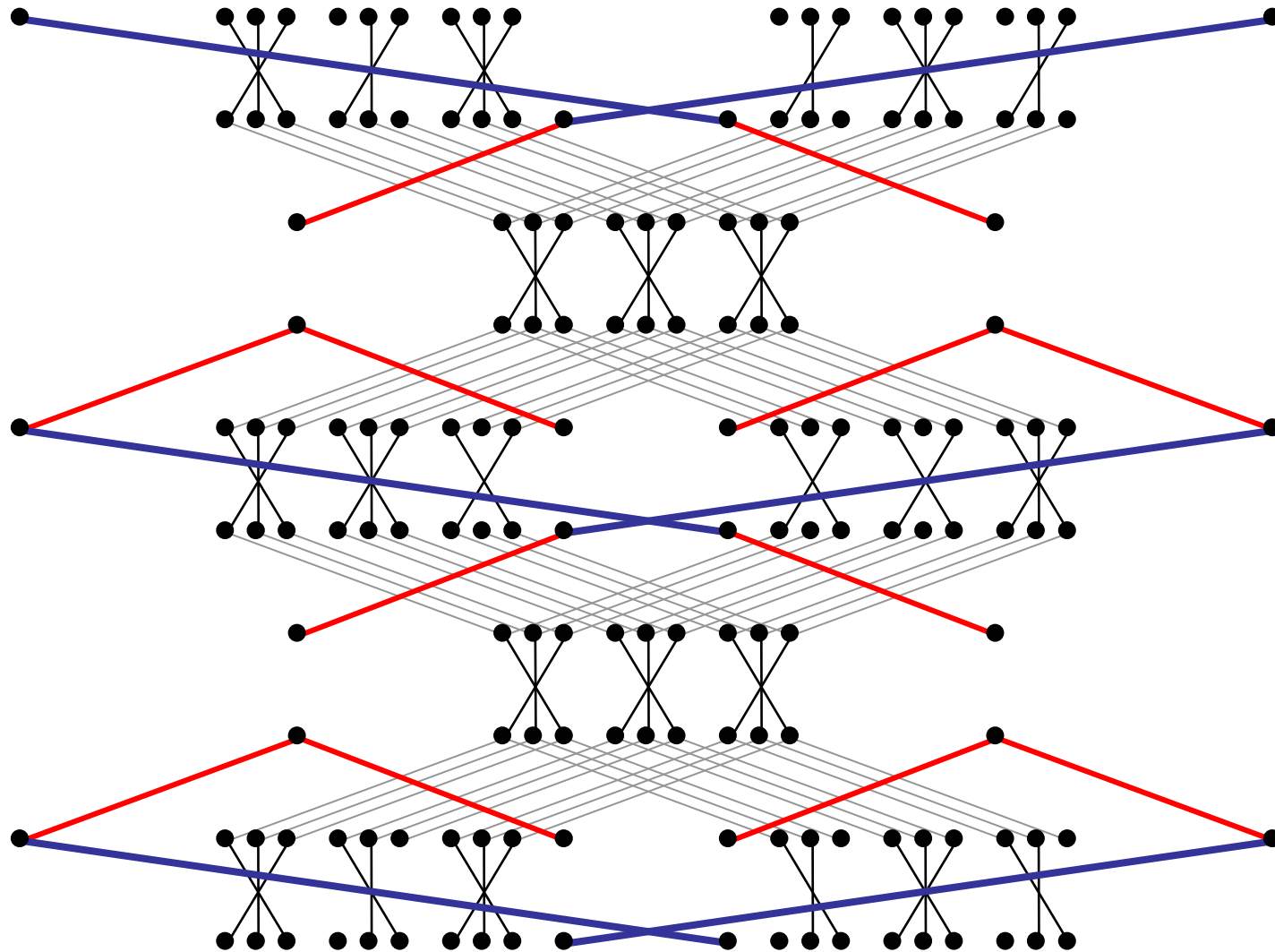


RNA-RNAi solves Rubber Bands



Why is Rubber Bands NP-complete?

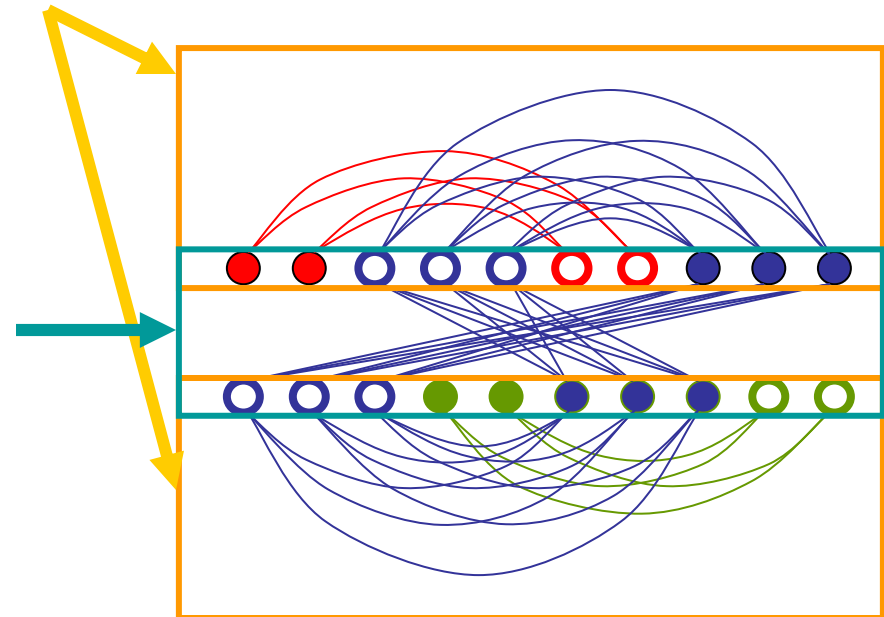
Rubber Bands solves 3SAT



An example reduction from $(X \vee Y \vee Z) (X' \vee Y \vee Z') (X \vee Y' \vee Z)$. This actually presents a weighted version of Rubber Bands game. Thicker edges have higher weights; moreover, the weight of an edge is greater than the sum of weights of all edges with lower weight. The weights can be removed by replacing each edge (and their adjacent nodes) with multiple edges (multiple nodes). SAT formula is satisfiable iff weight of Rubber Bands game is equal to a particular value (obtained from the construction).

1/2-Approximation

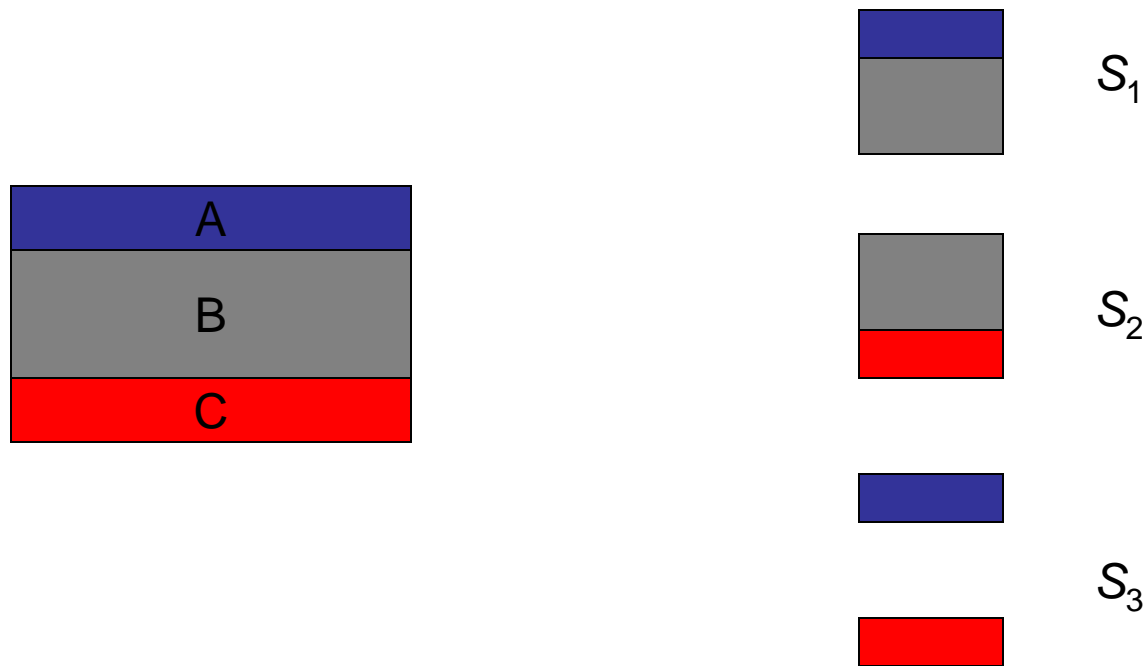
- Optimally fold RNA_1 and RNA_2 independently (i.e. ignore interaction) $O(n^3)$ time
- Optimally align RNA_1 and RNA_2 (i.e. ignore folding) $O(n^2)$ time
- One of the two contains at least 1/2 the optimal number of edges (bonds)



2/3-Approximation

- Optimally fold RNA_1 and RNA_2 ignoring interaction, $O(n^3)$ time
- Optimally fold RNA_1 while interacting with RNA_2 and ignore folding for RNA_2
- Optimally fold RNA_2 while interacting with RNA_1 and ignore folding for RNA_1
- One of these contains at least $2/3$ of the optimal number of edges (bonds)

2/3-Approximation

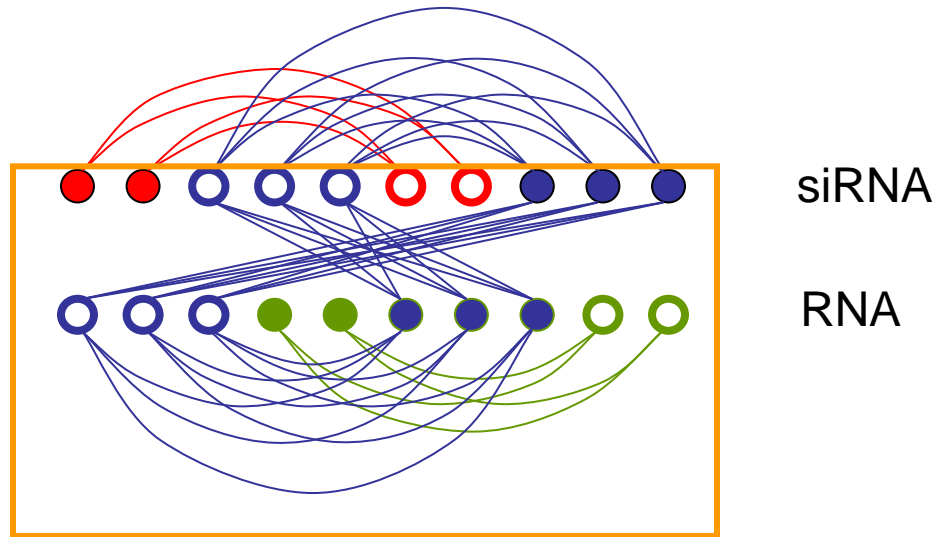


$$2 \cdot OPT = (A+B+C) + (A+B+C) = (A+B) + (B+C) + (A+C) \leq S_1 + S_2 + S_3$$

$$2/3 \cdot OPT \leq (S_1 + S_2 + S_3)/3 \leq \max(S_1, S_2, S_3)$$

siRNA-RNAi

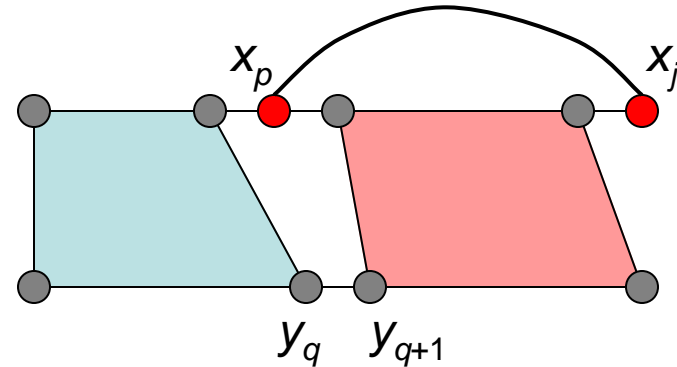
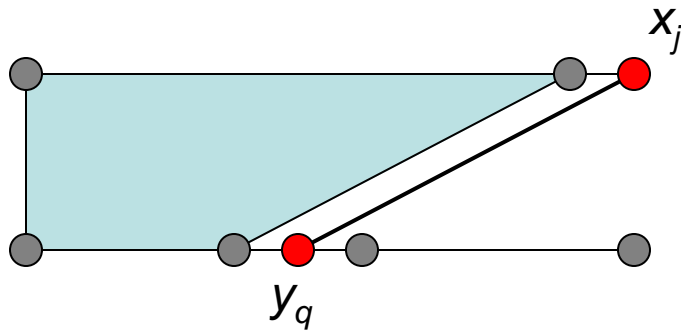
siRNA does not fold



Dynamic Programming

- Sub-problem: Let $V(i,j,k,l)$ = number of bonds that can be achieved in the optimal solution for $x[i..j]$ and $y[k..l]$
- We want to compute $V(1,m,1,n)$
- We can use a dynamic programming approach where $V(1,m,1,n)$ can be computed recursively in terms of smaller sub-problems

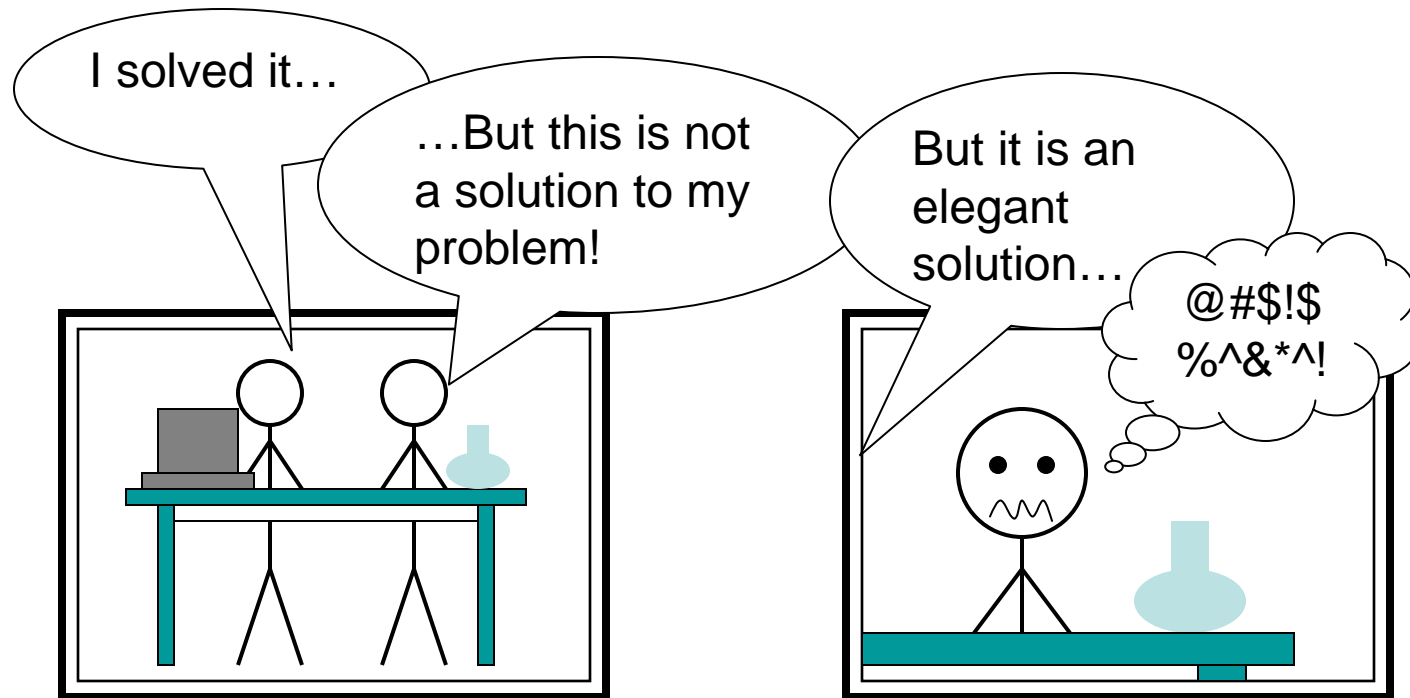
Dynamic programming



$$V(i,j,k,l) = \max \begin{cases} V(i,j-1,k,l) \\ V(i,j-1,k,q-1) + w(x_j, y_q) & k \leq q \leq l \quad O(m) \\ V(i,p-1,k,q) + V(p+1,j-1,q+1,l) + w(x_p, x_j) & i \leq p \leq j-1 \\ & k \leq q \leq l \quad O(mn) \end{cases}$$

- $O(m^2)$ substrings for x
- $O(n^2)$ substrings for y
- $O(m^2n^2) \cdot O(mn) = O(m^3n^3)$ algorithm, polynomial but not that efficient!

Computer scientists & Biologists



- Computer scientists seek mathematical beauty
- Biologists want to solve real problems
- Need to bridge the gap

Conclusion

- RNA-RNAi is NP-complete
- Approximation algorithms exist, but need better ones
- Simpler variations are polynomially solvable
- What computational models are appropriate for RNA-RNA interaction
- Application to siRNA design needs to be investigated

Q&A

Thank you...

