

Designing RNA sequences with targeted secondary structure and nucleotide distribution

Vladimir Reinhartz

Yann Ponty, Jérôme Waldspühl

vladimir.reinhartz@mail.mcgill.ca

July 19, 2013



RNA Design

A

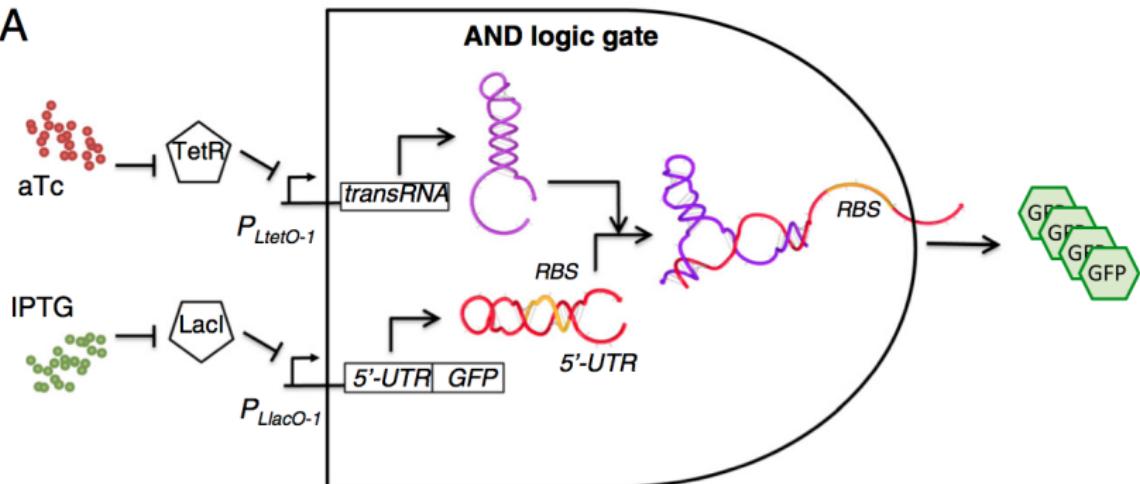
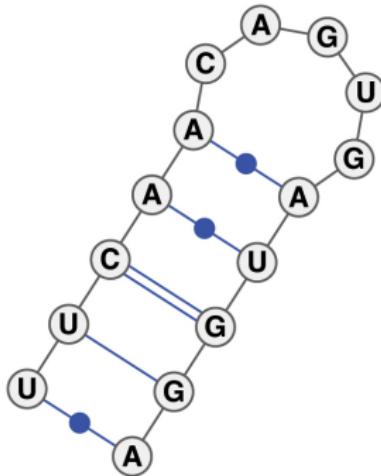
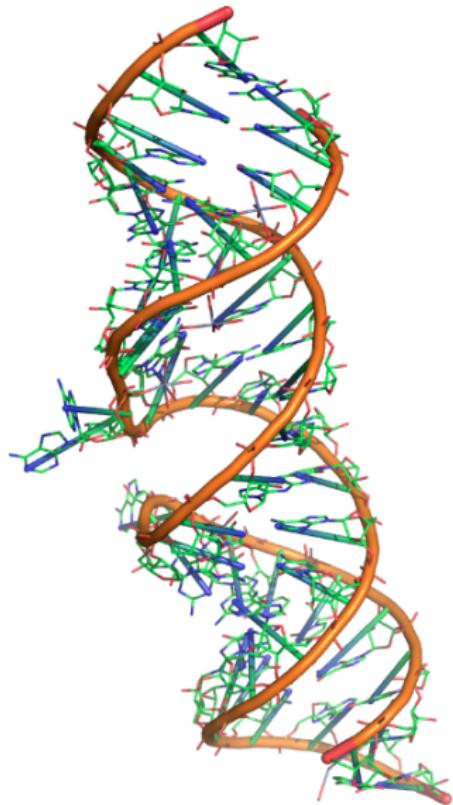


Figure : Rodrigo et al. 2012

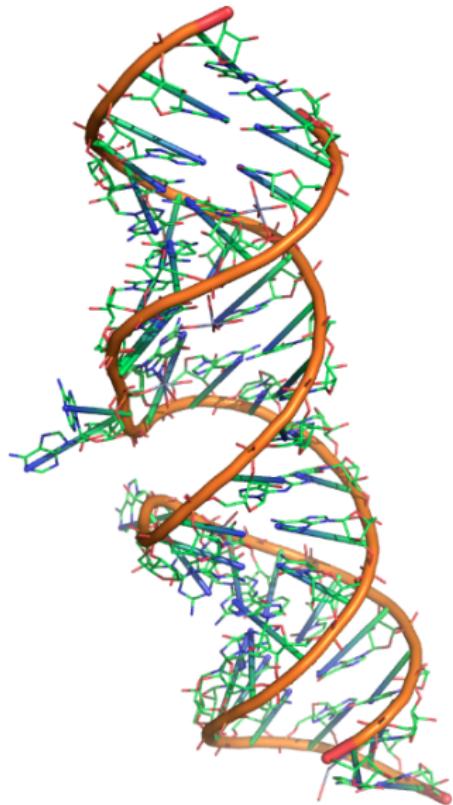
RNA Structure



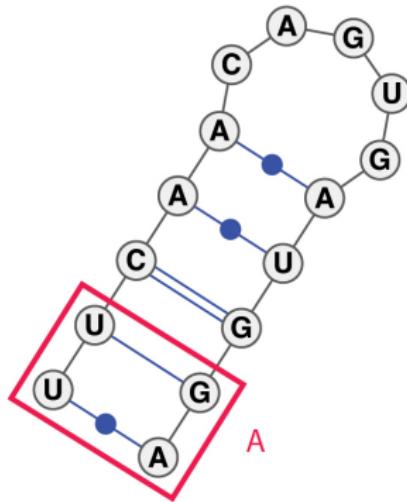
RNA Structure



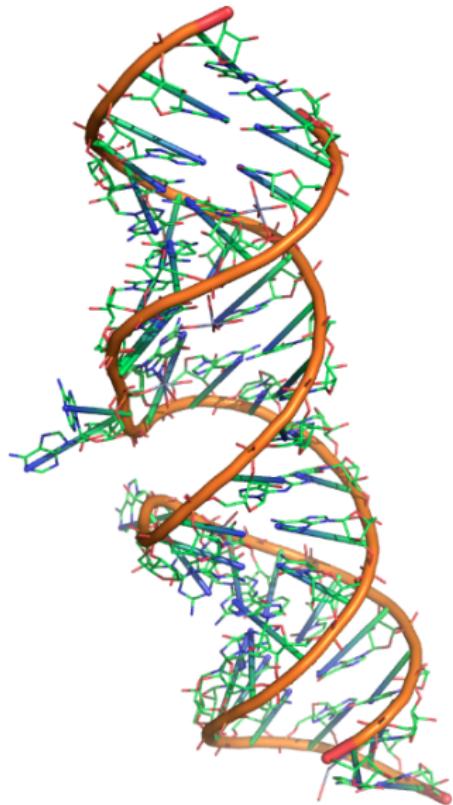
RNA Structure



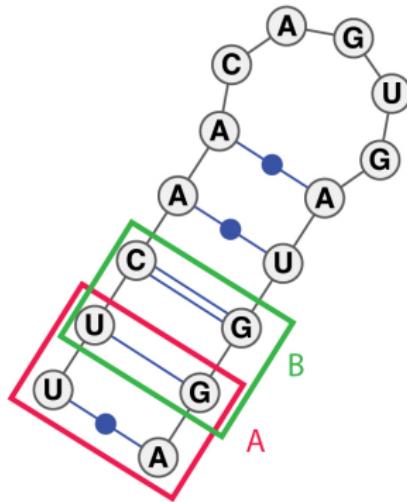
Energy
 $A = -1.3$



RNA Structure



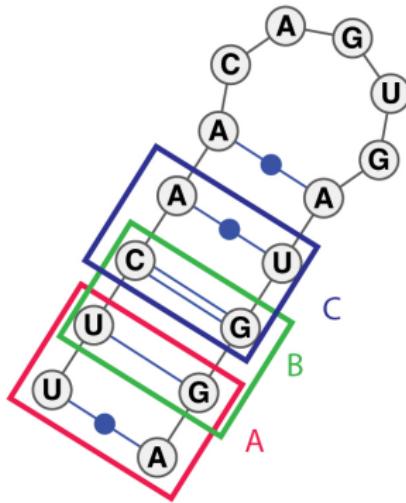
Energy
 $A = -1.3$
 $B = -1.5$



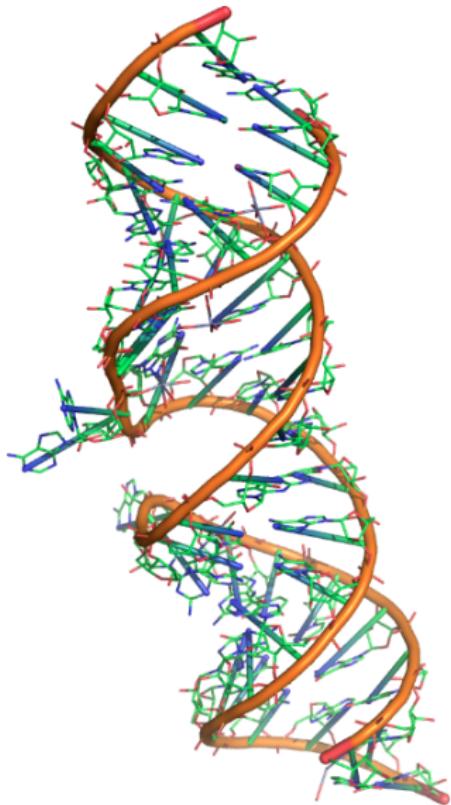
RNA Structure



Energy
A = -1.3
B = -1.5
C = -2.1



RNA Structure



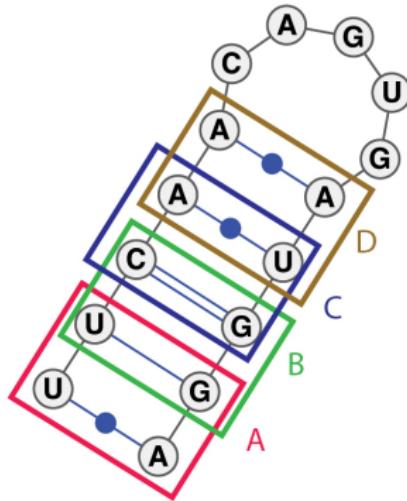
Energy

$$A = -1.3$$

$$B = -1.5$$

$$C = -2.1$$

$$D = \beta$$



Energy Landscape

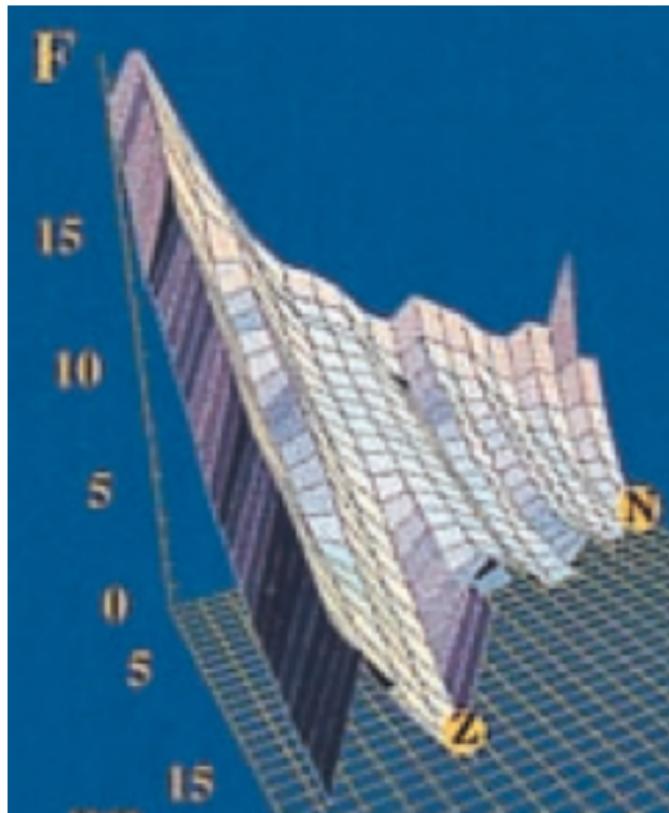
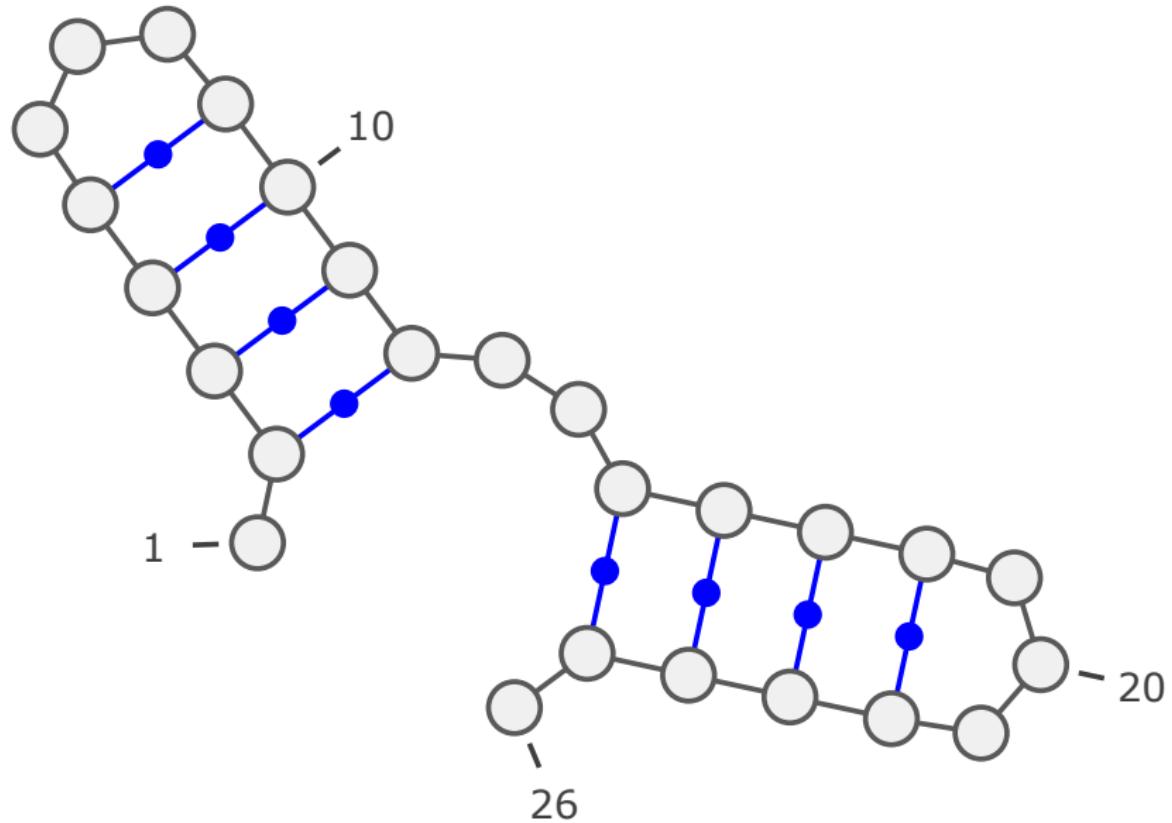
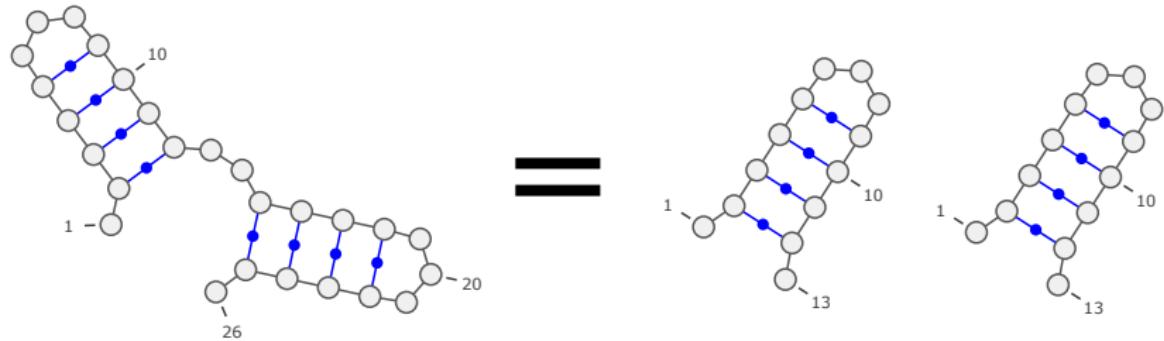


Figure : Chen and Dill, 1999

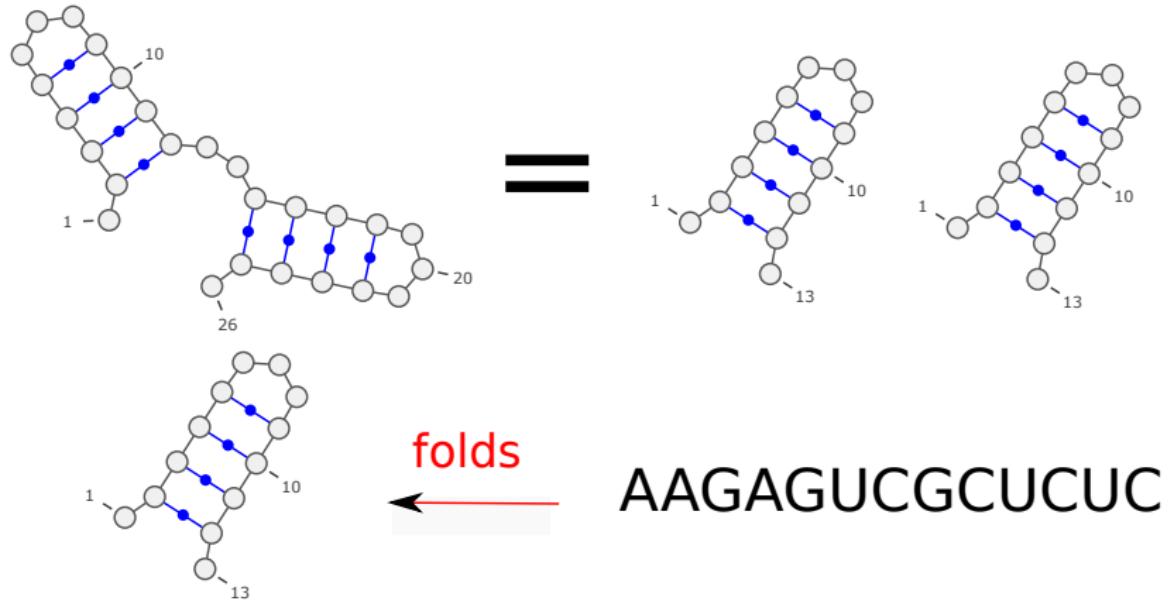
Local Design



Local Design



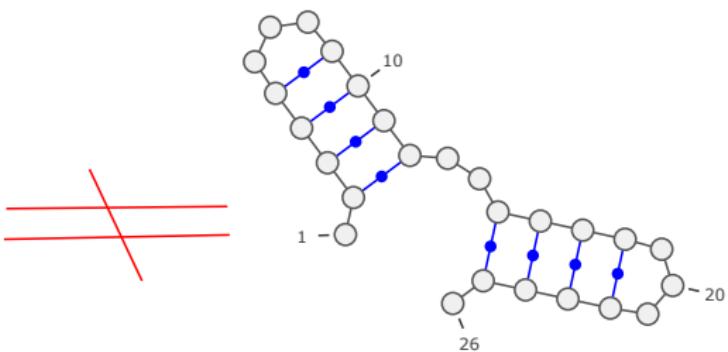
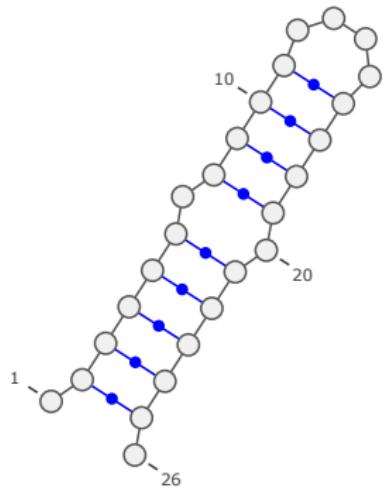
Local Design



Local Design

AAGAGUCGCUCUC **AAGAGUCGCUCUC**

Folds
↓



Probabilistic Model Sequence space

- ▶ We can define a Boltzmann distribution
- ▶ Due to additive properties of the Stacking Energy we can efficiently compute the partition function \mathbb{Z} with a Dynamic Programming scheme

Boltzmann factor:

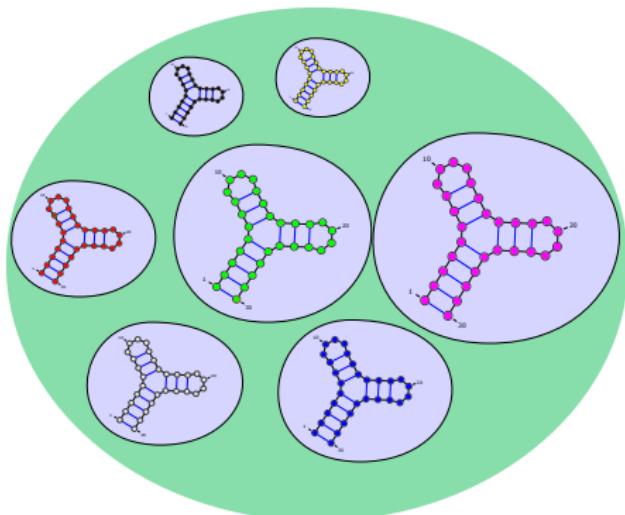
$$\mathcal{B}(s) := e^{\frac{-E(s)}{RT}}$$

Partition function:

$$\mathbb{Z} = \sum_s \mathcal{B}(s)$$

$$E(s) := \underbrace{\text{ES}}_{\text{Stacking Energy}}(s, S)$$

Stacking Energy



IncaRNAtion

RNA-ensign¹

IncaRNAtion

IncaRNAtion

RNA-ensign¹
Seeded

IncaRNAtion
No Seed

IncaRNAtion

RNA-ensign¹

Seeded

Explore mutant space

IncaRNAtion

No Seed

Explore sequence space

IncaRNAtion

RNA-ensign¹

Seeded

Explore mutant space

$\mathcal{O}(n^5)$

Sequence constraints

IncaRNAtion

No Seed

Explore sequence space

$\mathcal{O}(n)$

IncaRNAtion

RNA-ensign¹

Seeded

Explore mutant space

$\mathcal{O}(n^5)$

Sequence constraints

IncaRNAtion

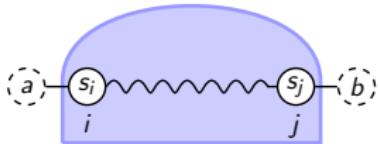
No Seed

Explore sequence space

$\mathcal{O}(n)$

DP Recursion global

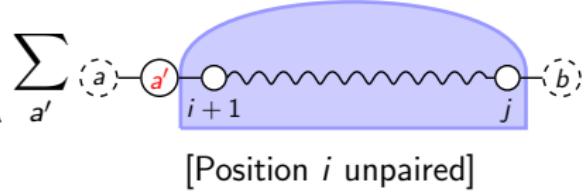
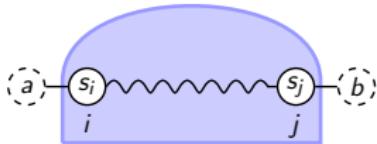
Explore sequence space
Structure fixed



DP Recursion

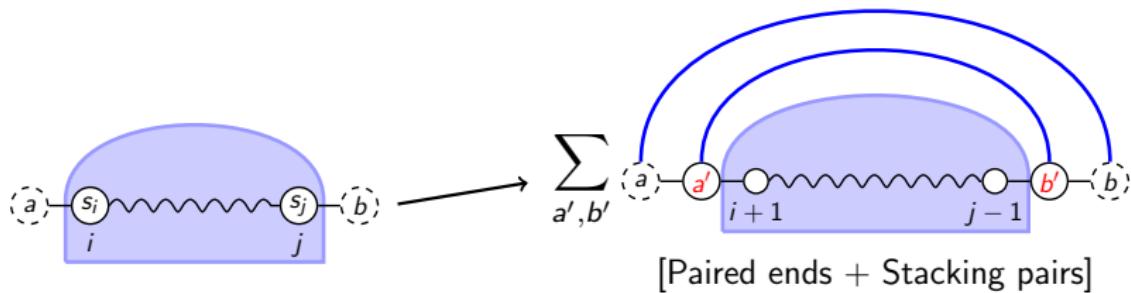
global

Explore sequence space
Structure fixed



DP Recursion global

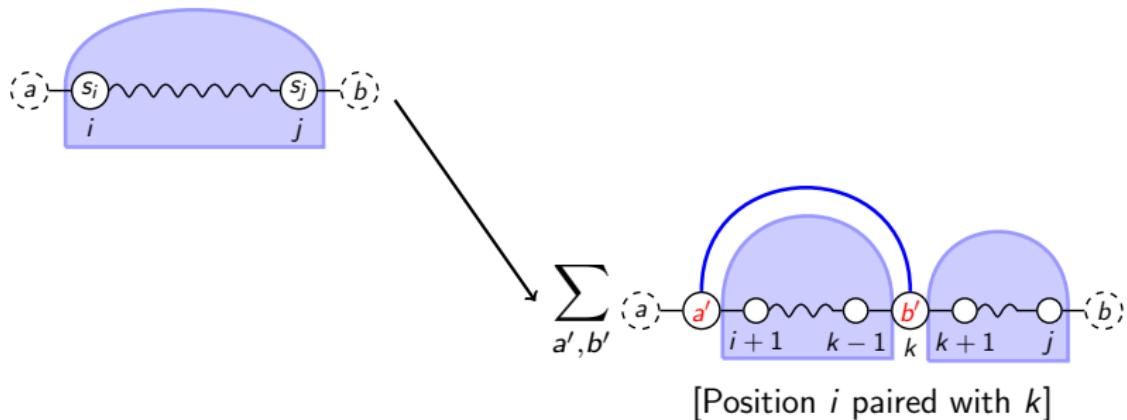
Explore sequence space
Structure fixed



DP Recursion

global

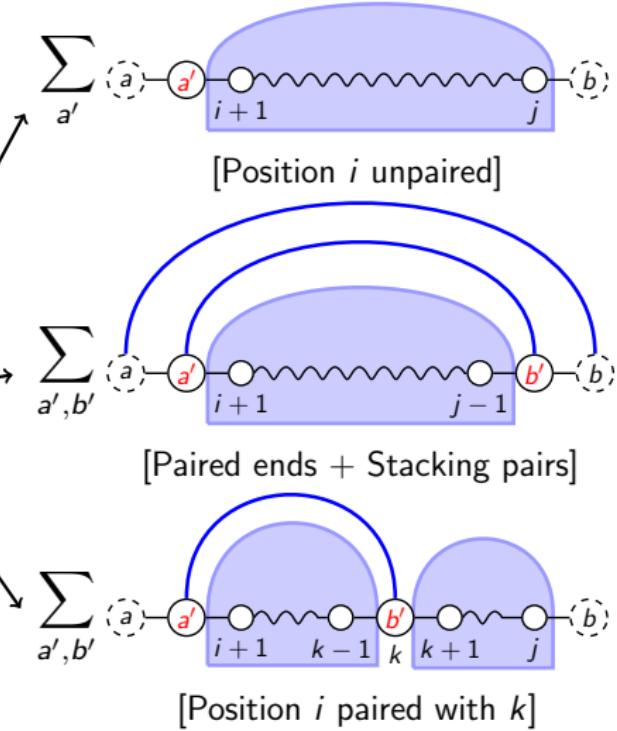
Explore sequence space
Structure fixed



DP Recursion global

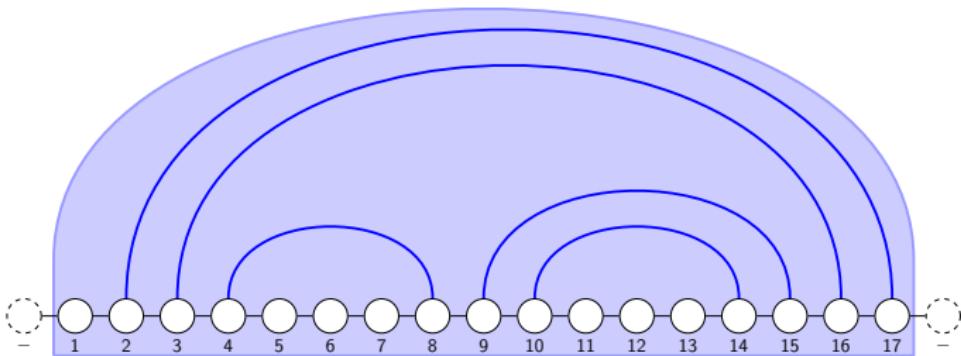
Explore sequence space
Structure fixed

Only 1 case applies



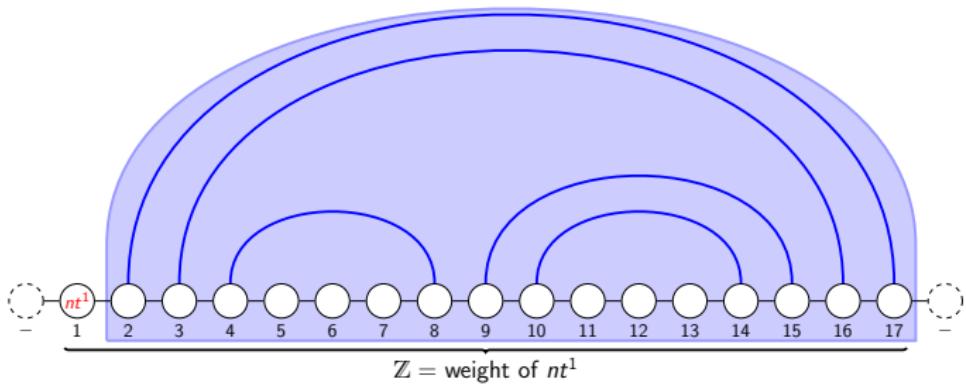
"Backtrack" global

Sequence:



"Backtrack" global

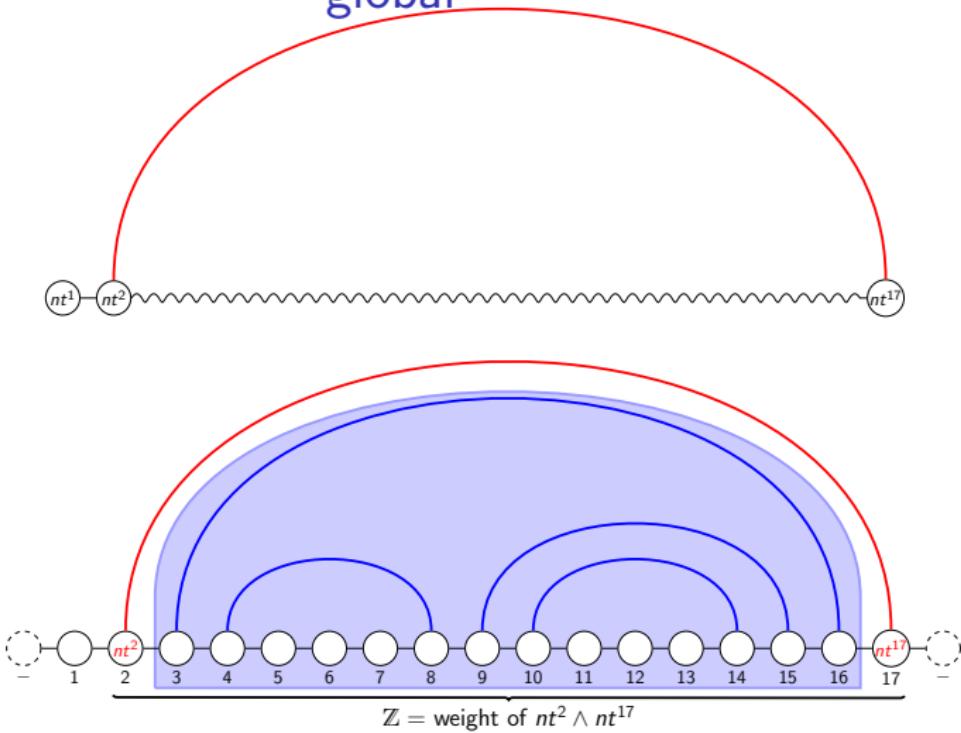
Sequence: nt^1



"Backtrack"

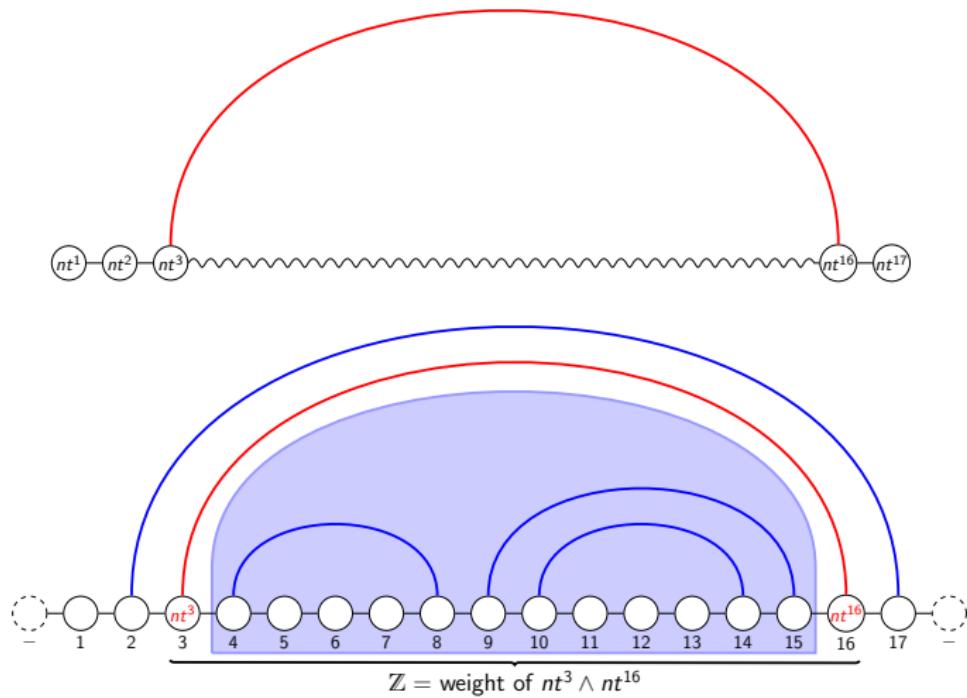
global

Sequence:



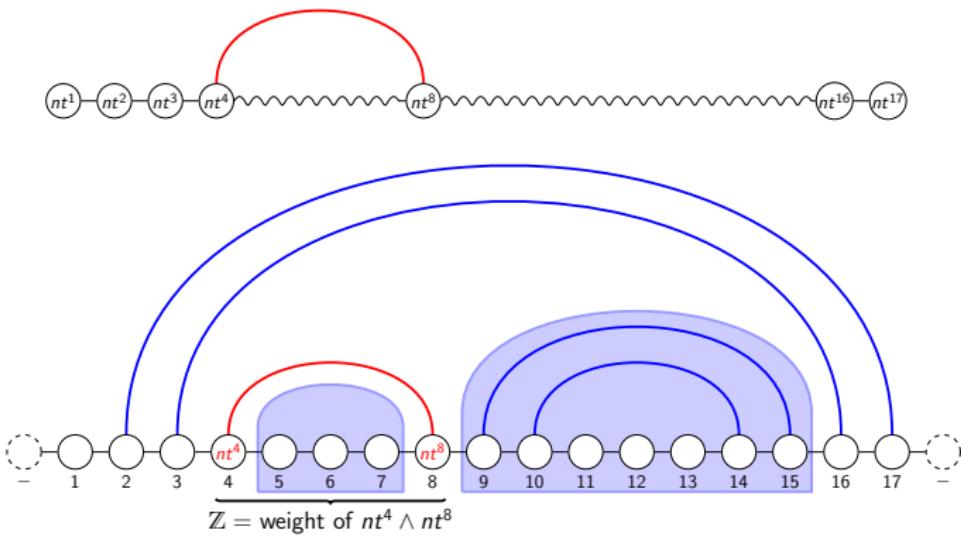
"Backtrack" global

Sequence:



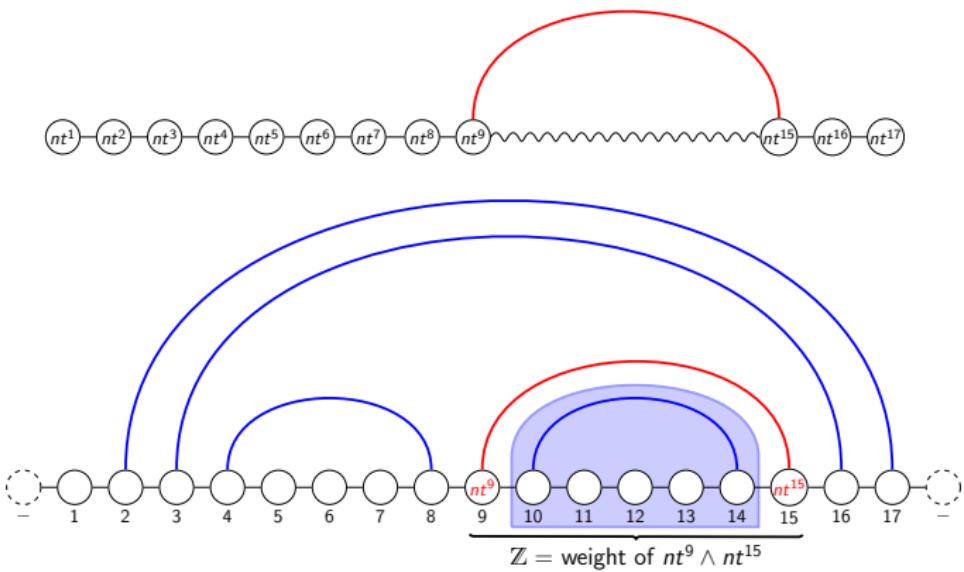
"Backtrack" global

Sequence:



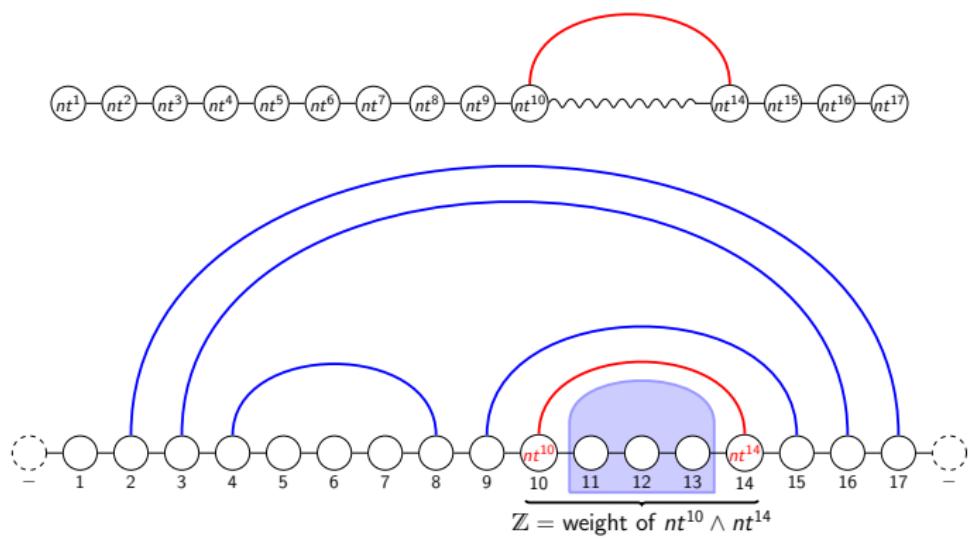
"Backtrack" global

Sequence:



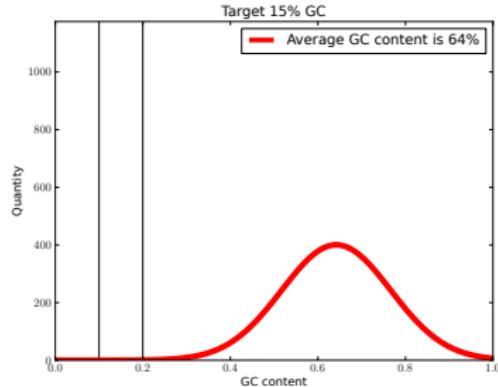
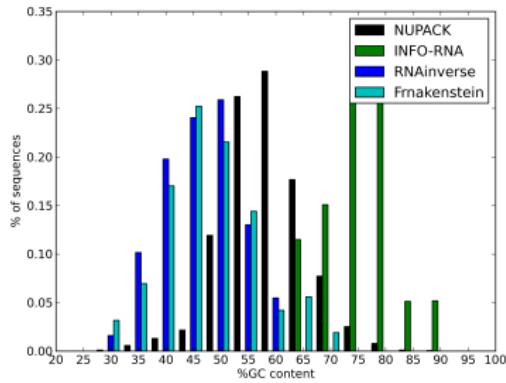
"Backtrack" global

Sequence:



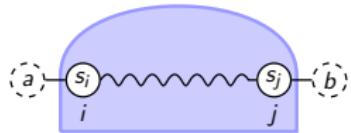
GC Bias

Avg RFAM GC is 48.8%



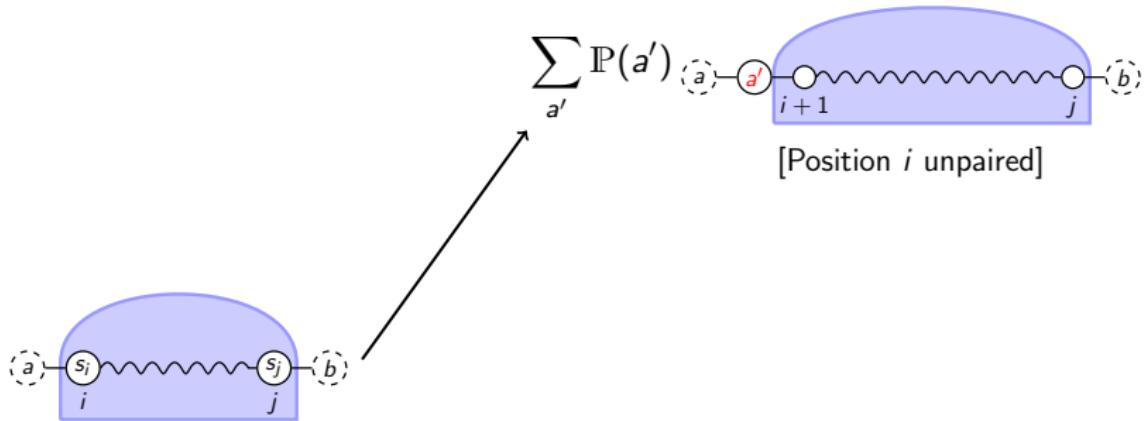
Weighted DP Recursion

global



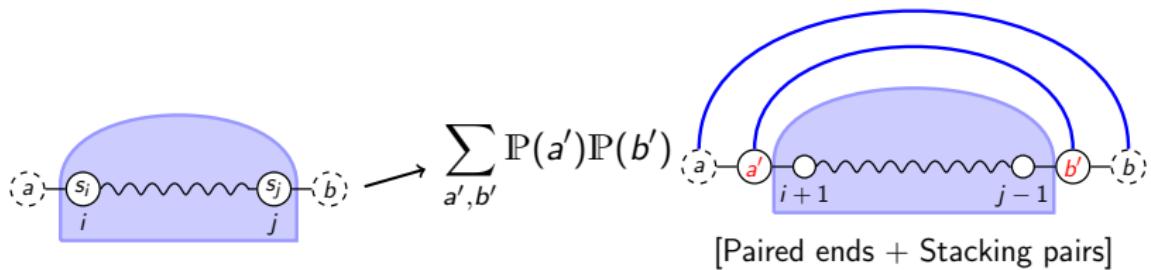
Weighted DP Recursion

global



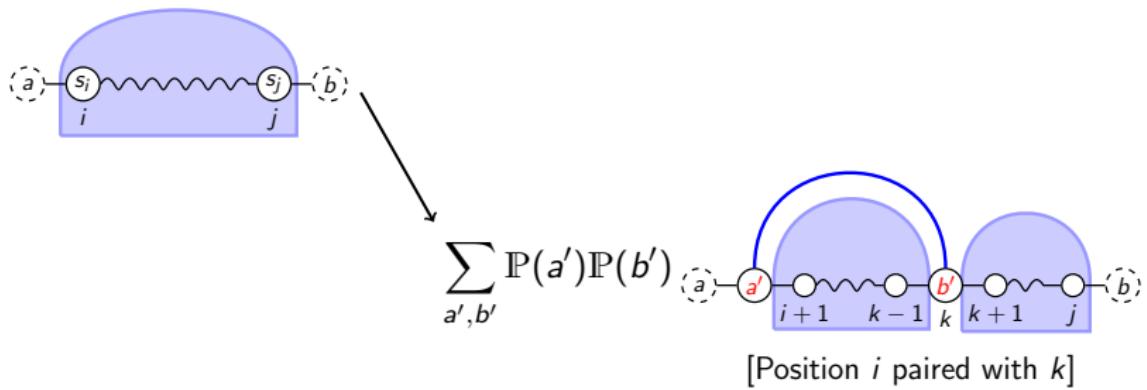
Weighted DP Recursion

global



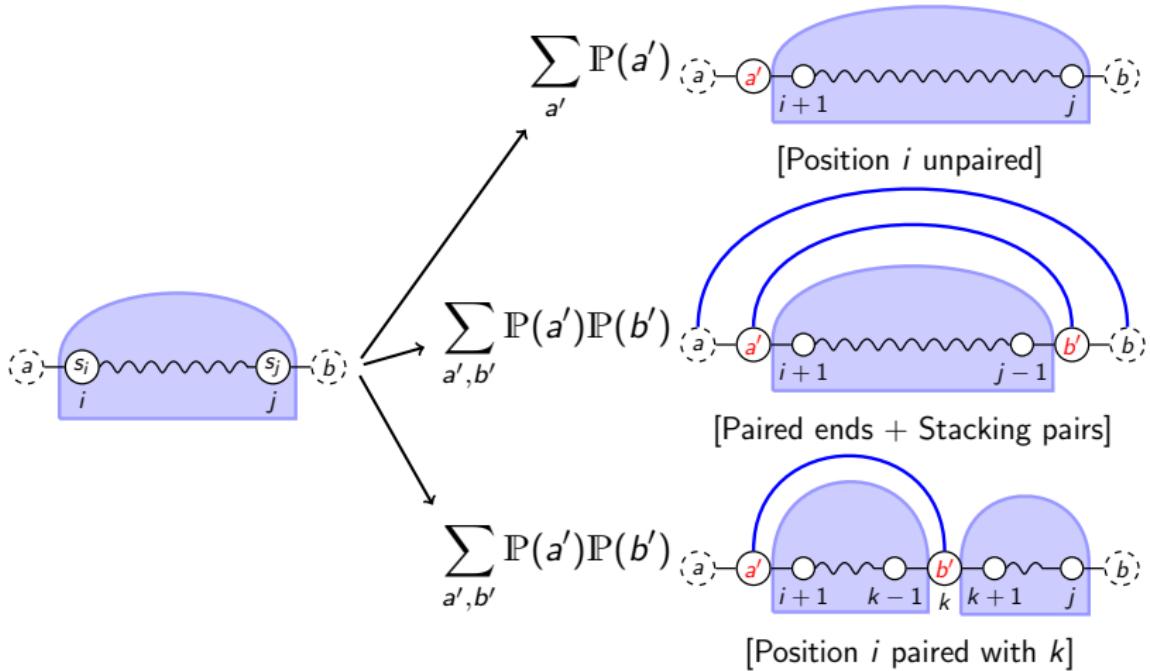
Weighted DP Recursion

global

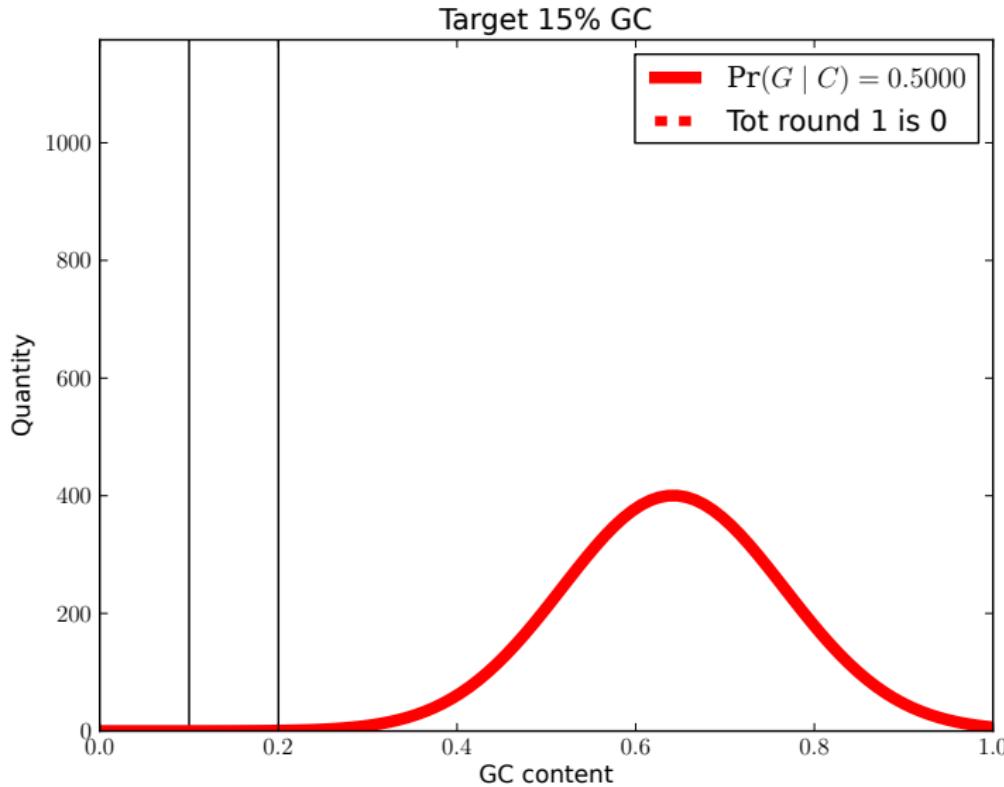


Weighted DP Recursion

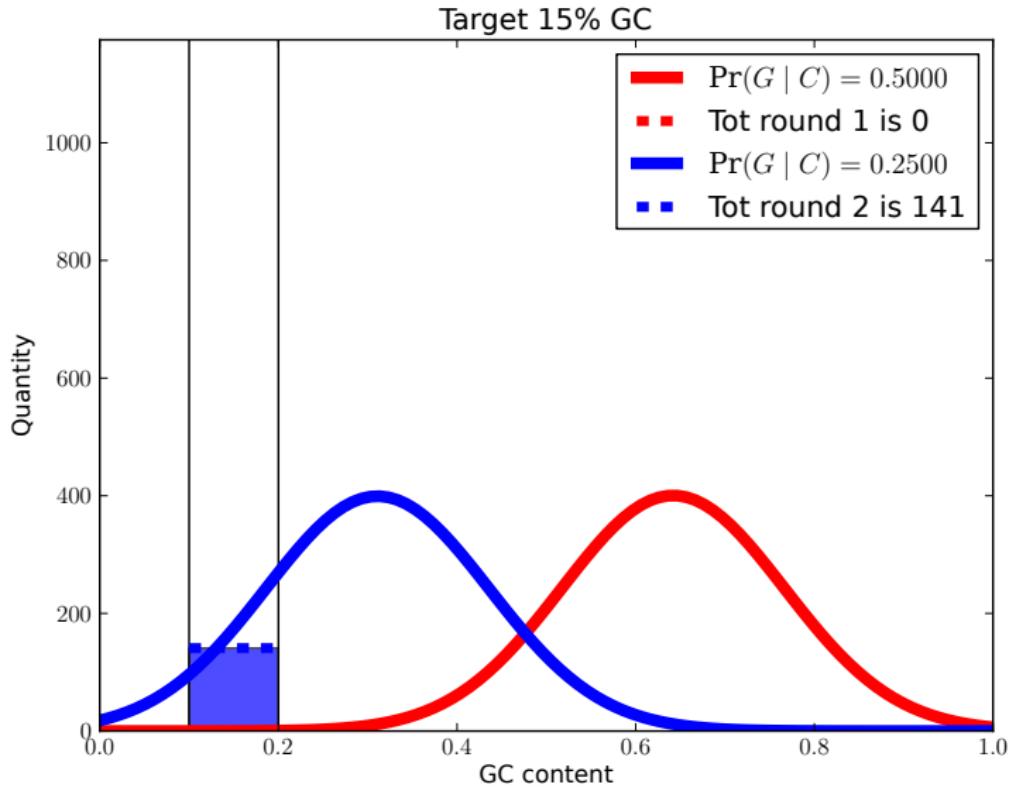
global



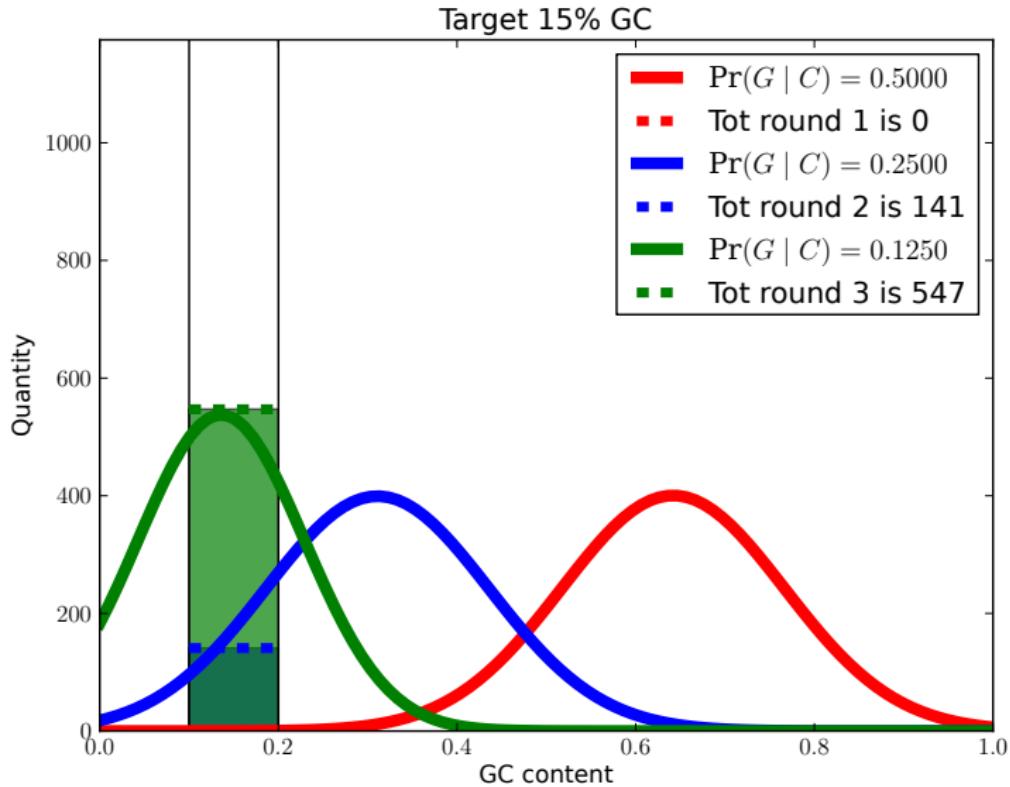
Incarnation distribution Bisection scheme



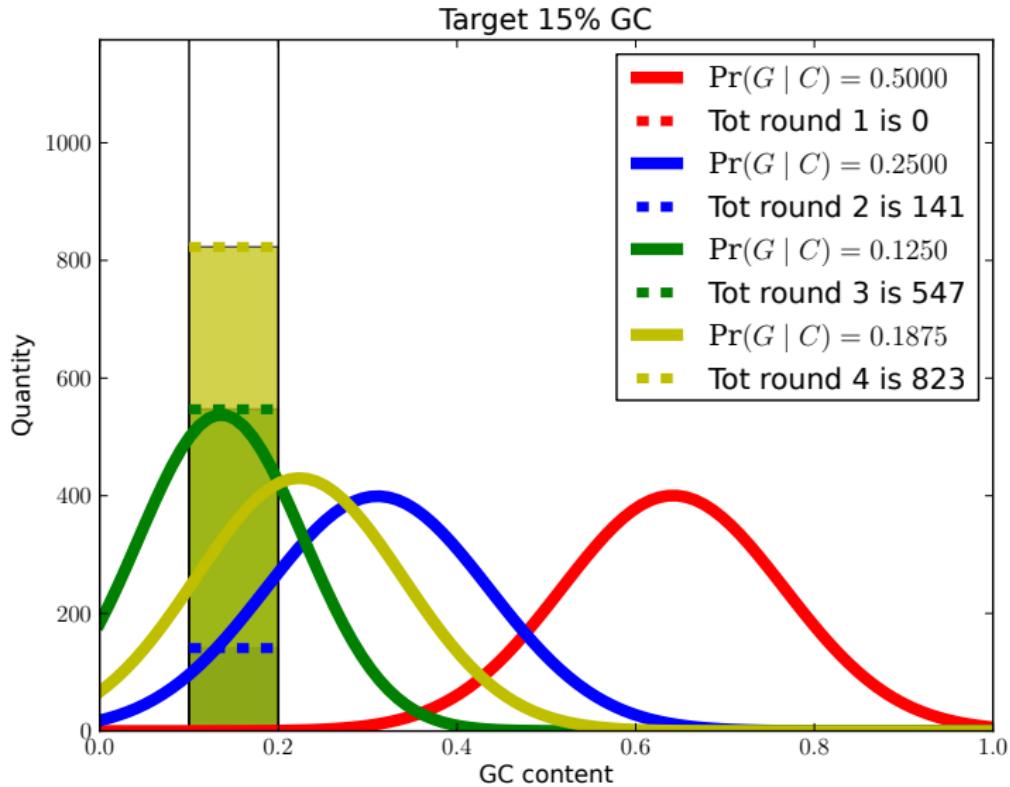
Incarnation distribution Bisection scheme



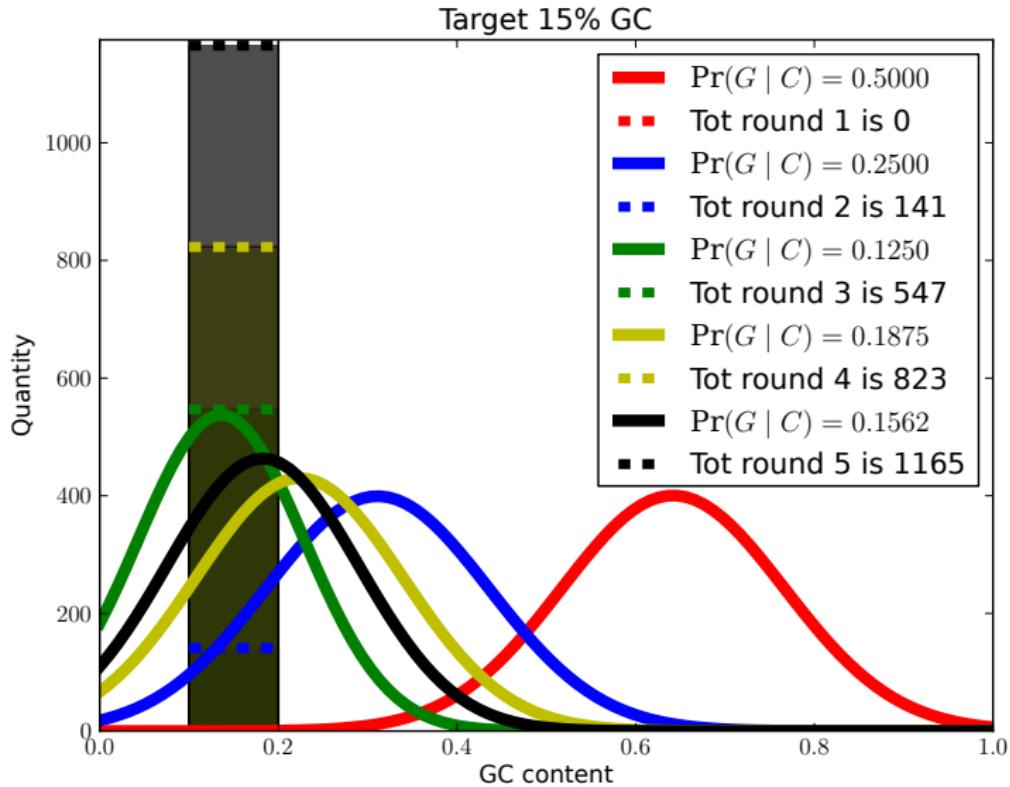
Incarnation nt distribution Bisection scheme



Incarnation nt distribution Bisection scheme



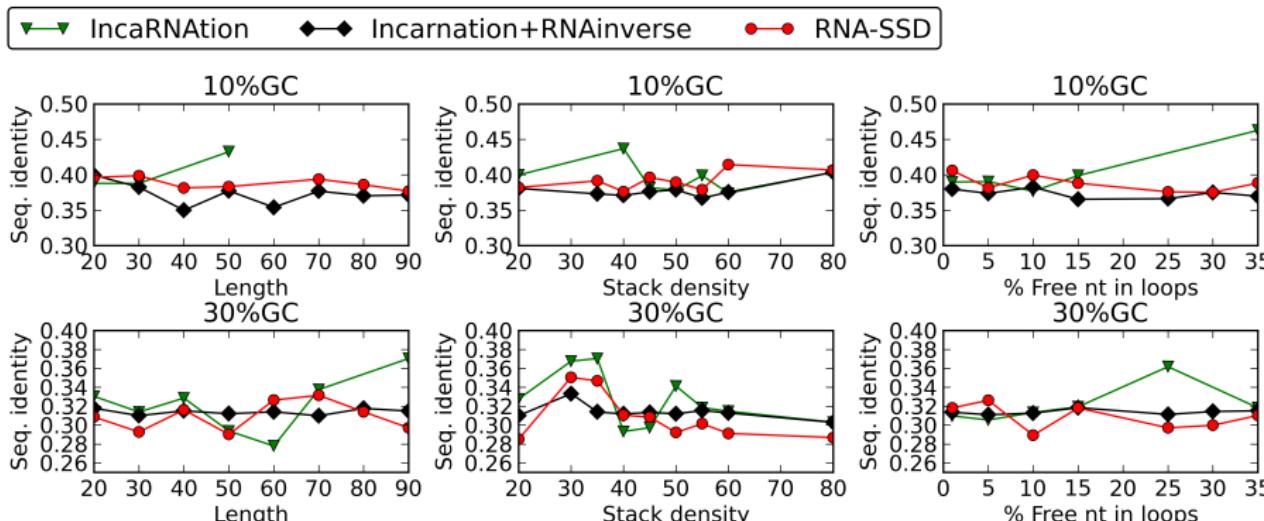
Incarnation nt distribution Bisection scheme



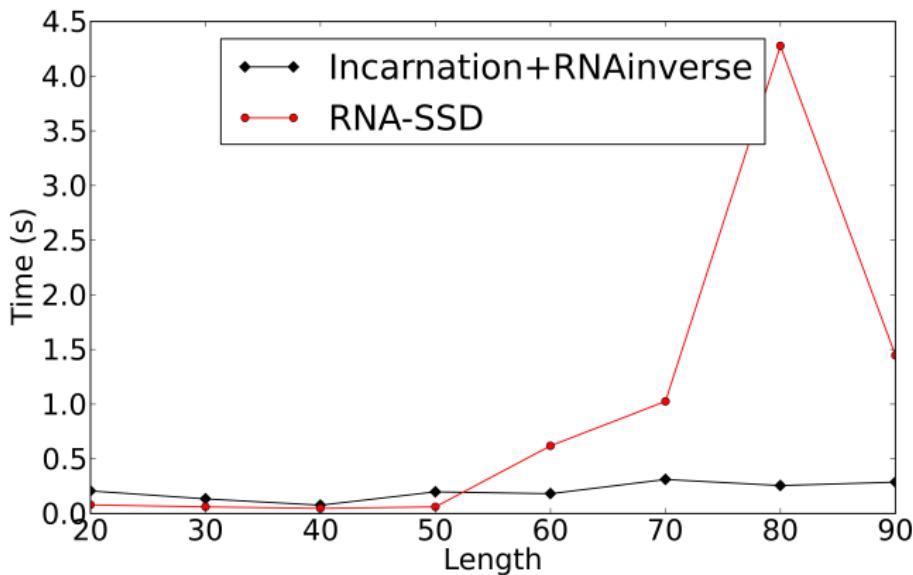
Incarnation + RNAinverse Results

Sequence identity

Only keep length?



Incarnation + RNAinverse time



Future

- ▶ Full energy model
- ▶ Include other seeded methods
- ▶ Convince people to make the molecules in labs

Acknowledgments

- ▶ Coauthors

- ▶ Jérôme Waldspühl
- ▶ Yann Ponty



- ▶ Erasmus TEE fellowship
- ▶ Travel funding to ISMB/ECCB 2013 was generously provided by ISCB.

