

RNA design in theory and practice

Sven Findeiß

sven@bioinf.uni-leipzig.de

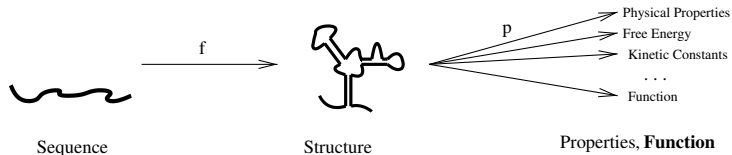
Bioinformatics Group and Interdisciplinary Center for Bioinformatics,
Department of Computer Science,
University Leipzig

AlgoSB 2019

January, 2019

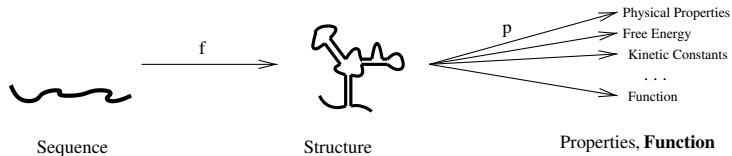
Sequence-Structure-Function Relation

**Basic paradigm of RNA structure biology:
Structure of an RNA sequence typically determines the function.**



Sequence-Structure-Function Relation

**Basic paradigm of RNA structure biology:
Structure of an RNA sequence typically determines the function.**



Sequence \Rightarrow Structure \Rightarrow Function

See a structure...

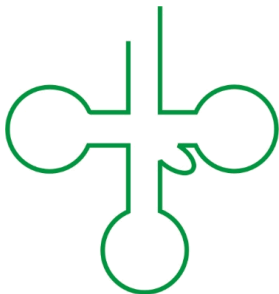
Donath, Alexander, et al., **2010**, Noncoding RNA, 251–293, in Evolutionary Genomics and Systems Biology, edited by Caetano-Anolles, Wiley-Blackwell

See a structure...

... guess the function !!!

Donath, Alexander, et al., **2010**, Noncoding RNA, 251–293, in Evolutionary Genomics and Systems Biology, edited by Caetano-Anolles, Wiley-Blackwell

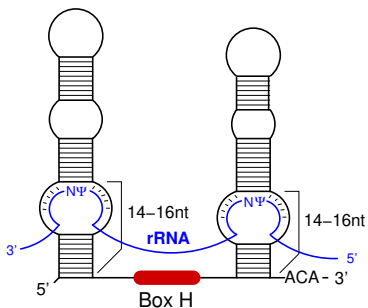
See a structure...



... guess the function !!!

Donath, Alexander, et al., 2010, Noncoding RNA, 251–293, in Evolutionary Genomics and Systems Biology, edited by Caetano-Anolles, Wiley-Blackwell

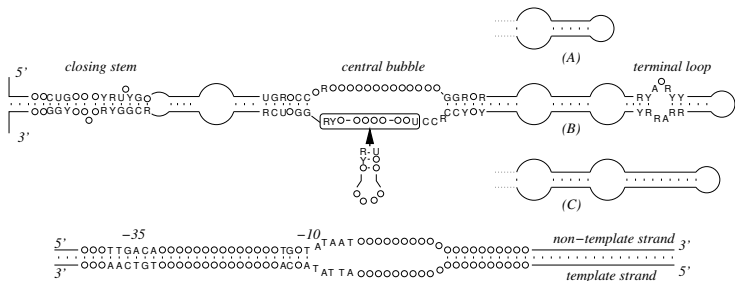
See a structure...



... guess the function !!!

Donath, Alexander, et al., 2010, Noncoding RNA, 251–293, in Evolutionary Genomics and Systems Biology, edited by Caetano-Anolles, Wiley-Blackwell

See a structure...

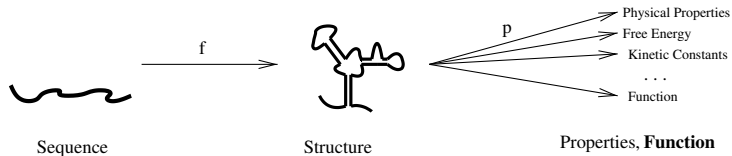


... guess the function !!!

Donath, Alexander, et al., 2010, Noncoding RNA, 251–293, in *Evolutionary Genomics and Systems Biology*, edited by Caetano-Anolles, Wiley-Blackwell

Sequence-Structure-Function Relation

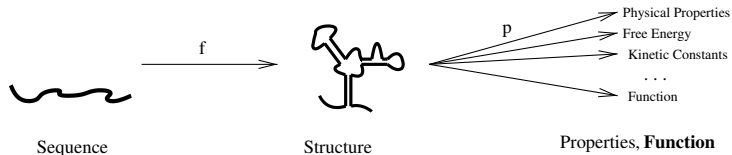
**Basic paradigm of RNA structure biology:
Structure of an RNA sequence typically determines the function.**



Sequence \Rightarrow Structure \Rightarrow Function

Sequence-Structure-Function Relation

**Basic paradigm of RNA structure biology:
Structure of an RNA sequence typically determines the function.**



Sequence \Leftarrow Structure \Leftarrow Function

RNA design aims on the reverse

Sequence \Leftarrow Structure \Leftarrow Function

- 1 Which sequence optimally folds into given target structure(s)?
- 2 How to implement novel functions?

Sequence \Leftarrow Structure \Leftarrow Function

- 1 Which sequence optimally folds into given target structure(s)?
- 2 How to implement novel functions?

Sequence \Leftarrow Structure \Leftarrow Function

- 1 Which sequence optimally folds into given target structure(s)?
- 2 How to implement novel functions?

Complexity of Sequence Design I

- **Nucleic acid sequence sampling:** four possible assignments on each position.

$NNNNNNNNNNNNNN = 4^{12} = 16,777,216$ possible sequences

$N = \{A, C, G, U\}$

Complexity of Sequence Design I

- **Nucleic acid sequence sampling:** four possible assignments on each position.

$$\text{NNNNNNNNNNNN} = 4^{12} = 16,777,216 \text{ possible sequences}$$

$$N = \{A, C, G, U\}$$

- **Sequence constraints:** restriction of individual positions to a subset of possible assignments

$$\text{NGNNNVNNNN} = 1 \times 3 \times 4^{10} = 3,145,728 \text{ possible sequences}$$

$$V = \{A, C, G\}$$

Complexity of Sequence Design I

- **Nucleic acid sequence sampling:** four possible assignments on each position.

$NNNNNNNNNNNN = 4^{12} = 16,777,216$ possible sequences

$N = \{A, C, G, U\}$

- **Sequence constraints:** restriction of individual positions to a subset of possible assignments

$NGNNNVNNNN = 1 \times 3 \times 4^{10} = 3,145,728$ possible sequences

$V = \{A, C, G\}$

- **RNA forms base pairs:** restriction to six possible assignments for paired positions

$(((((...))))). = 4^4 \times 6^4 = 331,776$ possible sequences

$() = \{AU, UA, CG, GC, UG, GU\} \quad . = N = \{A, G, C, U\}$

Complexity of Sequence Design I

- **Nucleic acid sequence sampling:** four possible assignments on each position.

$NNNNNNNNNNNN = 4^{12} = 16,777,216$ possible sequences

$N = \{A, C, G, U\}$

- **Sequence constraints:** restriction of individual positions to a subset of possible assignments

$NGNNNVNNNN = 1 \times 3 \times 4^{10} = 3,145,728$ possible sequences

$V = \{A, C, G\}$

- **RNA forms base pairs:** restriction to six possible assignments for paired positions

$(((((...))))). = 4^4 \times 6^4 = 331,776$ possible sequences

$() = \{AU, UA, CG, GC, UG, GU\} \quad . = N = \{A, G, C, U\}$

- **Combination of sequence and structure constraint:**

$(((((...))))). =$

$NGNNNVNNNN = 3 \times 4^3 \times 2 \times 6^3 = 82,944$ possible sequences

Sequence Selection

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization **J Comput Chem** 32: 439–452

Sequence Selection

(((((...))))).
NGNNNNVNNNNN



- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- EDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization **J Comput Chem** 32: 439–452

Sequence Selection

Calculate the energy of the target structure for each sequence.

((((...)))).
NGNNNNVNNNNN

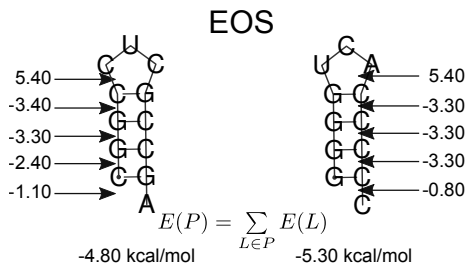
- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization **J Comput Chem** 32: 439–452

Sequence Selection

Calculate the energy of the target structure for each sequence.



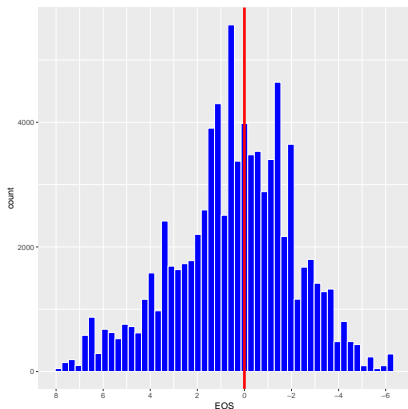
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection



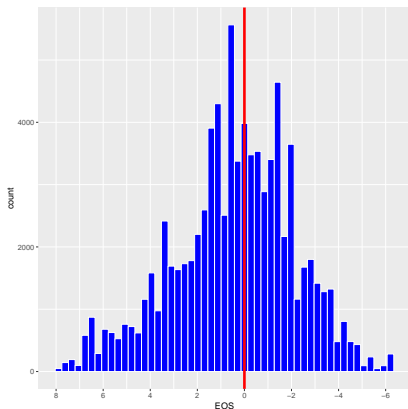
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

$(((((\dots))))).$
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization **J Comput Chem** 32: 439–452

Sequence Selection



EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

((((...)))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization **J Comput Chem** 32: 439–452

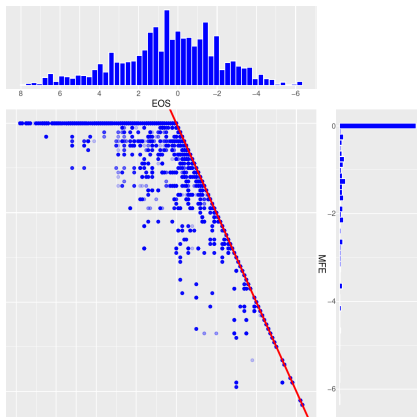
Does the target structure correspond to the actual minimum free energy conformation?

EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Sequence Selection



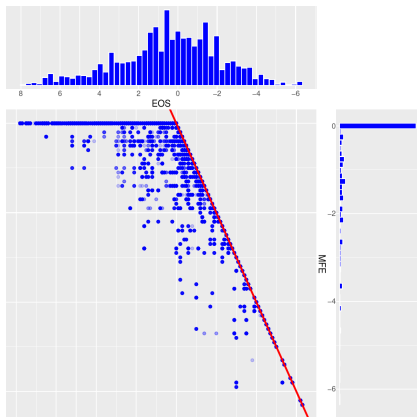
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization **J Comput Chem** 32: 439–452

Sequence Selection



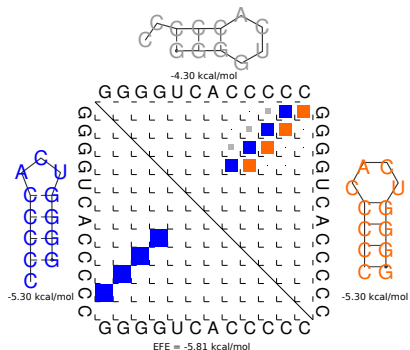
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection



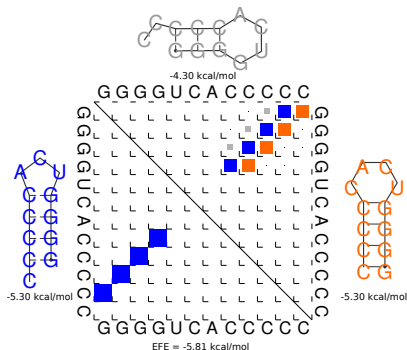
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection



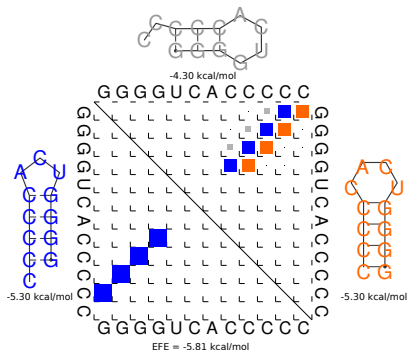
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection



EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection

Does the target conformation dominate the ensemble?

$$EFE = -RT \ln \underbrace{\sum_P \exp\left(-\frac{E(P)}{RT}\right)}_{\text{partition function } Z}$$

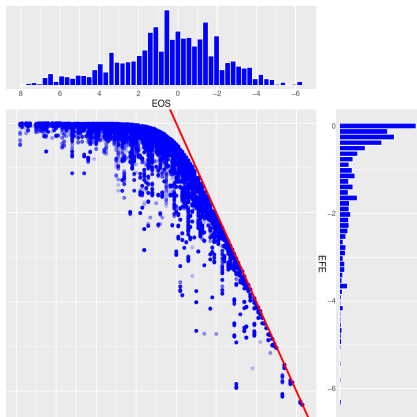
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection



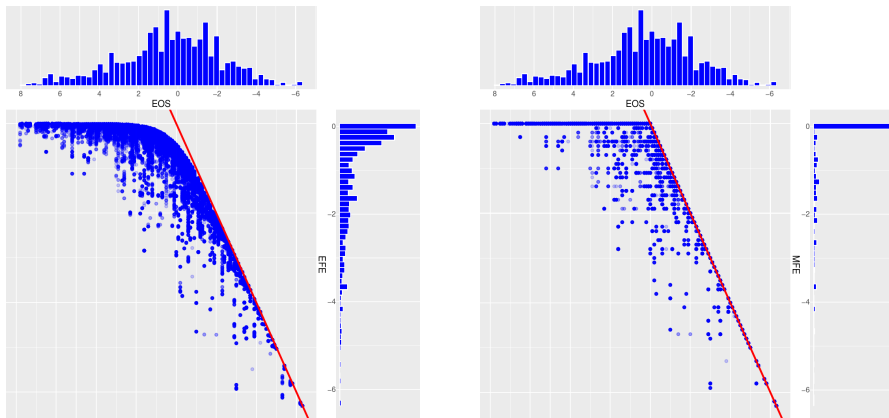
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization **J Comput Chem** 32: 439–452

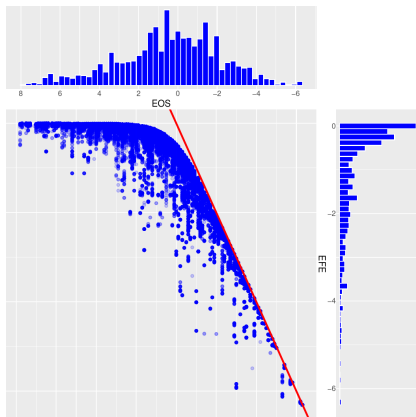
Sequence Selection



EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection



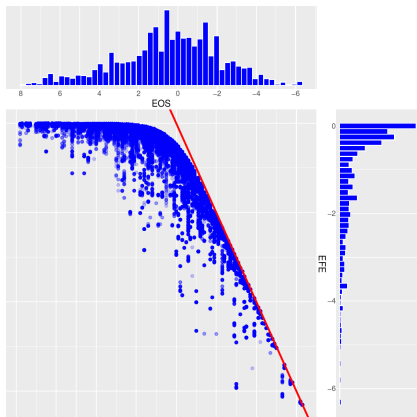
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization **J Comput Chem** 32: 439–452

Sequence Selection



EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection

Combine EFE and EOS to calculate the probability of the target structure in the ensemble.

$$p(P) = \exp\left(-\frac{EOS - EFE}{RT}\right)$$

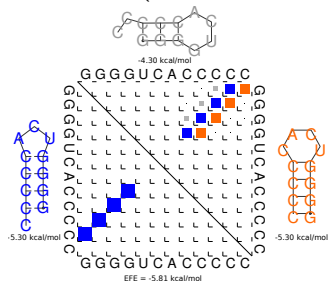
(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Sequence Selection

Combine EFE and EOS to calculate the probability of the target structure in the ensemble.

$$p(P) = \exp\left(-\frac{EOS - EFE}{RT}\right)$$



$$\exp\left(-\frac{-5.30 - (-5.81)}{RT}\right) = 0.44$$

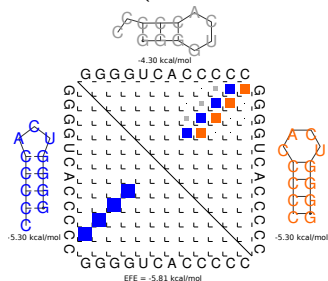
(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Sequence Selection

Combine EFE and EOS to calculate the probability of the target structure in the ensemble.

$$p(P) = \exp\left(-\frac{EOS - EFE}{RT}\right)$$



$$\exp\left(-\frac{-4.30 - (-5.31)}{RT}\right) = 0.09$$

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Sequence Selection

Combine EFE and EOS to calculate the probability of the target structure in the ensemble.

$$p(P) = \exp\left(-\frac{EOS - EFE}{RT}\right)$$

The probability defect estimates the offset

$$PDE = 1 - p(P)$$

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Sequence Selection

Combine EFE and EOS to calculate the probability of the target structure in the ensemble.

$$p(P) = \exp\left(-\frac{EOS - EFE}{RT}\right)$$

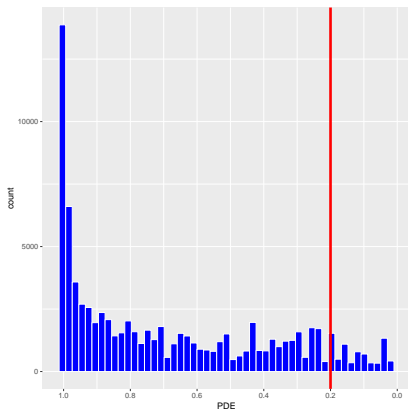
The probability defect estimates the offset

$$PDE = 1 - p(P)$$

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Sequence Selection



EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

((((...))) .
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection

Ensemble defect ensures that structures similar to the target conformation are favored, while distant or contrary structures are prohibited.

EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Sequence Selection

Ensemble defect ensures that structures similar to the target conformation are favored, while distant or contrary structures are prohibited.

$$EDE = \sum_P p(P) \underbrace{d_{BP}(P, T)}_{|BP_P| + |BP_T| - 2|BP_P \cap BP_T|}$$

EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection

Ensemble defect ensures that structures similar to the target conformation are favored, while distant or contrary structures are prohibited.

$$EDE = \sum_P p(P) \underbrace{d_{BP}(P, T)}_{|BP_P| + |BP_T| - 2|BP_P \cap BP_T|}$$

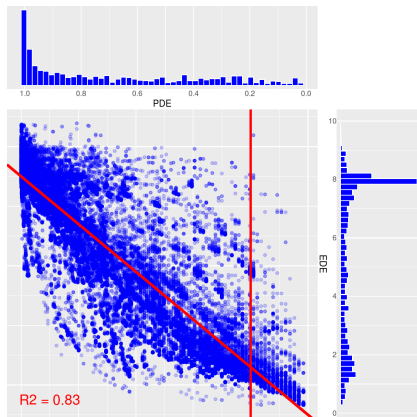
EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection



EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

((((...))) .
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection

Two conditions must be fulfilled

- 1 the target structure must be thermodynamically stable
→ positive design
- 2 contrary structures should be less stable and thus less probable in the ensemble
→ negative design

EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection

Two conditions must be fulfilled

- 1 the target structure must be thermodynamically stable
→ positive design
- 2 contrary structures should be less stable and thus less probable in the ensemble
→ negative design

EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection

Two conditions must be fulfilled

- 1 the target structure must be thermodynamically stable
→ positive design
- 2 contrary structures should be less stable and thus less probable in the ensemble
→ negative design

EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

(((((...))))).
NGNNNNVNNNNN

- 82,944 compatible sequences
- favor low EOS
- favor low EOS = MFE
- MFE state rarely the only stable conformation
- structure ensemble might be diverse
- favor low EOS = EFE
- PDE takes only the exact target state into account
- EDE also counts similar structures

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

Sequence Selection

Two conditions must be fulfilled

- 1 the target structure must be thermodynamically stable
→ positive design
- 2 contrary structures should be less stable and thus less probable in the ensemble
→ negative design

EOS := Energy Of Structure
MFE := Minimum Free Energy
EFE := Ensemble Free Energy
PDE := Probability Defect
EDE := Ensemble Defect

- 1994 RNAinverse
- 2004 RNA-SSD
- 2006 INFO-RNA
- 2011 RNAexinv
- 2011 NUPACK:Design*
- 2012 RNA-ensign
- 2012 DSS-Opt
- 2013 RNAiFold*
- 2013 IncaRNation
- 2014 ERD
- 2014 EteRNA(Bot)*
- 2015 antaRNA
- ...

Zadeh, Joseph N., et al., 2011, Sequence Design via Efficient Ensemble Defect Optimization *J Comput Chem* 32: 439–452

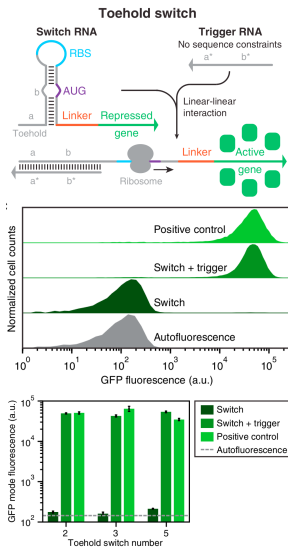
“Nature scores!”

Rhiju Das

Ensemble Defect based Design

sRNA binding regulates **accessibility** of region around the RBS and AUG and therefor **activates expression**

- **NUPACK:Design** for optimization towards **ensemble defect**
- **GFP reporter assays** with flow cytometry
- **Iterative design approach**: Tested various loop sizes, trigger RNA lengths, binding site length and their effect
- **in silico screening** to predict cross-talk, free energy of structural features, duplex formation, stability of toehold switch region



Green, A., et al., 2014, Toehold Switches: De-Novo-Designed Regulators of Gene Expression, *Cell*, 159, 925–939

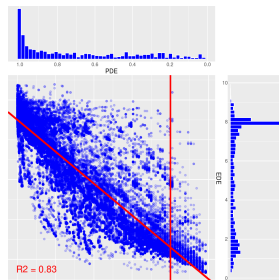
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVN>NNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

- Propagation of individual constraints to multiple positions
GRRRNNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

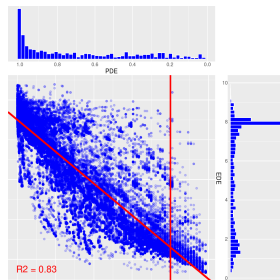
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GNNNNVNNNN

- Propagation of individual constraints to multiple positions
GRRRNVYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

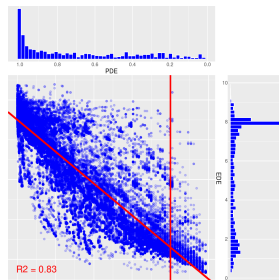
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GNNNNNVNNNNN

- Propagation of individual constraints to multiple positions
GRRRNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

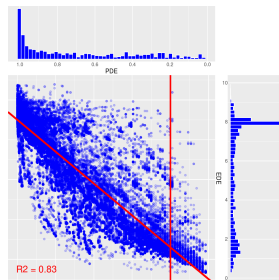
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNNVNNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GNNNNNVNNNYY

- Propagation of individual constraints to multiple positions
GRRRNNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

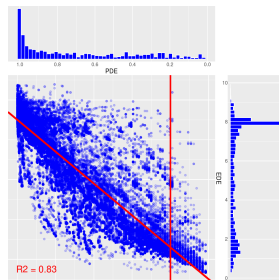
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GNNNNVNNNY

- Propagation of individual constraints to multiple positions
GRRRNVYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

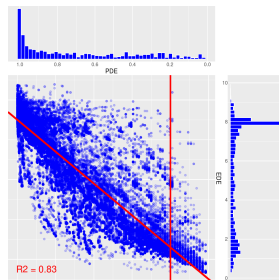
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GRNNNVNNNY

- Propagation of individual constraints to multiple positions
GRRRNVYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

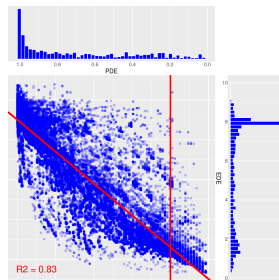
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GRNNNVNNYY

- Propagation of individual constraints to multiple positions
GRRRNVYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

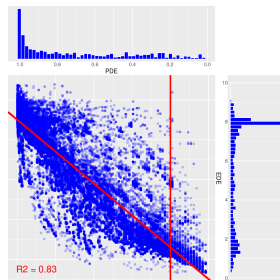
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GRNNNVNNYYY

- Propagation of individual constraints to multiple positions
GRRRNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

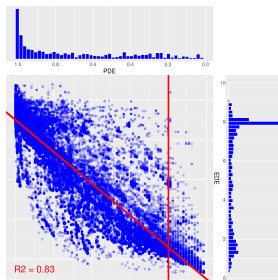
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GRNNNVNNYYY

- Propagation of individual constraints to multiple positions
GRRRNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

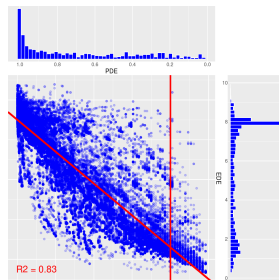
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GRRNNNVNNYYY

- Propagation of individual constraints to multiple positions
GRRRNNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

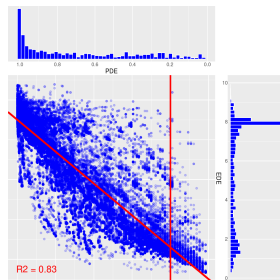
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

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

GRRNNNVNNYYY

- Propagation of individual constraints to multiple positions
GRRRNNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

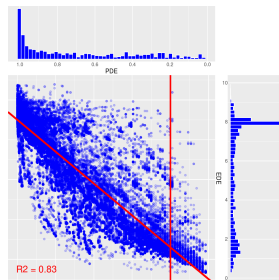
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

(((((...)))

GRRNNNVNYYYY

- Propagation of individual constraints to multiple positions
GRRRNNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

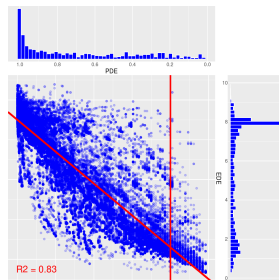
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVN>NNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

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

GRRNNNVNYYYY

- Propagation of individual constraints to multiple positions
GRRRNNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

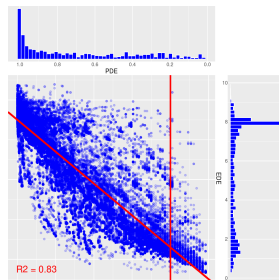
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVN>NNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

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

GRRRNNVN**YYYY**

- Propagation of individual constraints to multiple positions
GRRRN**V**YYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

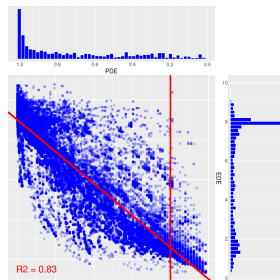
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVN>NNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

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

GRRRNNVNYYYY

- Propagation of individual constraints to multiple positions

GRRRNNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

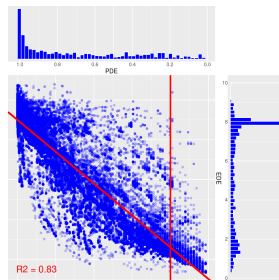
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNNVNNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

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

GRRRNNVYYYYY

- Propagation of individual constraints to multiple positions

GRRRNNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

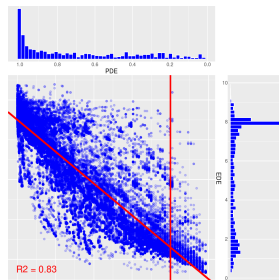
Complexity of Sequence Design II

- Combination of sequence and **one** structure constraint

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

NGNNNVNNNN = $3 \times 4^3 \times 2 \times 6^3 = 82,944$
possible sequences

- Utilize the **ensemble defect as objective** for optimization



- Combination of sequence and **multiple** structural constraints

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

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

GRRRNNVYYYYY

- **Propagation** of individual **constraints** to multiple positions

GRRRNNVYYYYY = $1 \times 2^8 \times 4^2 \times 3 = 12,288$
possible sequences

Assignment of unconstraint paths

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

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

NNNNNNNNNNNNNN

1.....6.....12

- Possible assignments for a base pair:
- Possible assignments for a path of length two:
- Possible assignment for a path of length L :

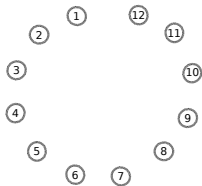
Assignment of unconstraint paths

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

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

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:
- Possible assignments for a path of length two:
- Possible assignment for a path of length L :

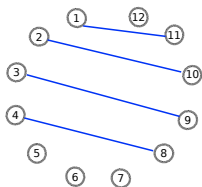
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:
- Possible assignments for a path of length two:
- Possible assignment for a path of length L :

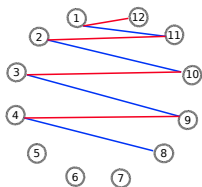
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:
- Possible assignments for a path of length two:
- Possible assignment for a path of length L :

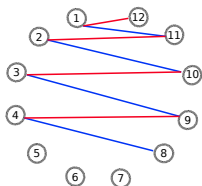
Assignment of unconstraint paths

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

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

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:
- Possible assignments for a path of length two:
- Possible assignment for a path of length L :

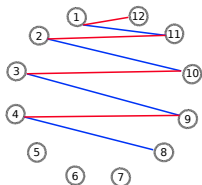
Assignment of unconstraint paths

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

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

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

()
A	U
C	G
G	C
G	U
U	A
U	G

- Possible assignments for a path of length two:

- Possible assignment for a path of length L :

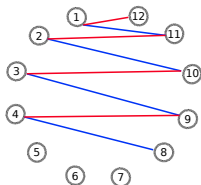
Assignment of unconstraint paths

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

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

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

6

()
A	U
C	G
G	C
G	U
U	A
U	G

- Possible assignments for a path of length two:

- Possible assignment for a path of length L :

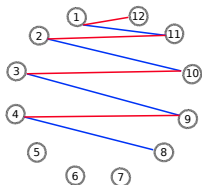
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

6

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

- Possible assignments for a path of length two:

- Possible assignment for a path of length L :

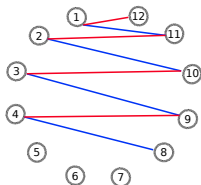
Assignment of unconstraint paths

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

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

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

6

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

- Possible assignments for a path of length two:

- Possible assignment for a path of length L :

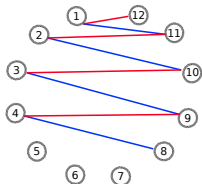
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

- Possible assignments for a path of length two:

- Possible assignment for a path of length L :

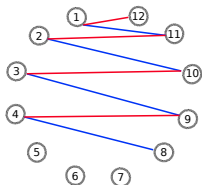
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

- Possible assignments for a path of length two:

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

- Possible assignment for a path of length L :

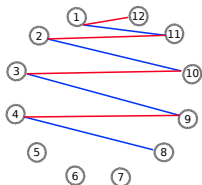
Assignment of unconstraint paths

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

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

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

- Possible assignments for a path of length two:

- Possible assignment for a path of length L :

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

()

A U

C G

G C

G U

U A

U G

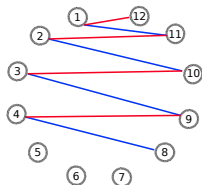
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

- Possible assignments for a path of length two:

- Possible assignment for a path of length L :

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

() .
· ()
A U

C G

G C

G U

U A

U G

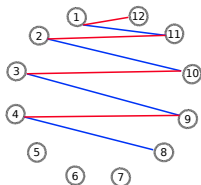
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

- Possible assignments for a path of length two:

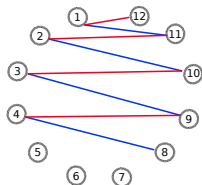
- Possible assignment for a path of length L :

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

() .
 . ()
 A U A
 A U G
 C G C
 C G U
 G C G
 G U A
 G U G
 U A U
 U G U
 U G C

Assignment of unconstraint paths

(((((((. . .))))) .
 (((((((. . . .))))))
 NNNNNNNNNNNNNNN
 1 6 12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

- Possible assignments for a path of length two:

10

- Possible assignment for a path of length L :

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

() .
 . ()
 A U A
 A U G
 C G C
 C G U
 G C G
 G U A
 G U G
 U A U
 U G U
 U G C

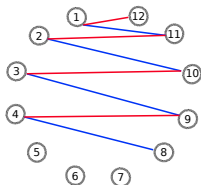
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

- Possible assignments for a path of length two:

10

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

$$BPT^2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

- Possible assignment for a path of length L :

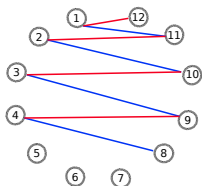
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

- Possible assignments for a path of length two:

$$\sum_{i,j} BPT^2[i,j] = 10$$

- Possible assignment for a path of length L :

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

$$BPT^2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

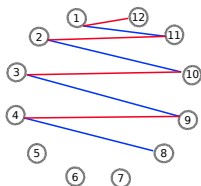
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

- Possible assignments for a path of length two:

$$\sum_{i,j} BPT^2[i,j] = 10$$

- Possible assignment for a path of length L :

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

$$BPT^2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

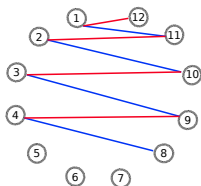
Assignment of unconstraint paths

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

(((((...)))

NNNNNNNNNNNNNN

1.....6.....12



- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

- Possible assignments for a path of length two:

$$\sum_{i,j} BPT^2[i,j] = 10$$

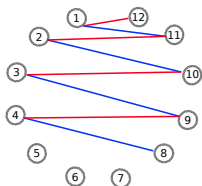
$$BPT^2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

- Possible assignment for a path of length L :

$$\sum_{i,j} BPT^L[i,j]$$

Assignment of unconstraint paths

(((((...))))).
(((...)))
NNNNNNNNNNNNNN
1.....6.....12



$$4^3 \times \sum_{i,j} BPT^8[i,j]$$

- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

- Possible assignments for a path of length two:

$$\sum_{i,j} BPT^2[i,j] = 10$$

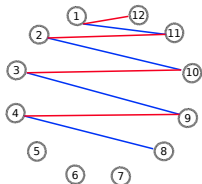
$$BPT^2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

- Possible assignment for a path of length L :

$$\sum_{i,j} BPT^L[i,j]$$

Assignment of unconstraint paths

(((((...))))).
(((...)))
NNNNNNNNNNNNNN
1.....6.....12



$$4^3 \times \sum_{i,j} BPT^8[i,j]$$
$$= 4^3 \times 178$$
$$= 11,392$$

- Possible assignments for a base pair:

$$\sum_{i,j} BPT[i,j] = 6$$

$$BPT = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

- Possible assignments for a path of length two:

$$\sum_{i,j} BPT^2[i,j] = 10$$

$$BPT^2 = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 2 & 0 \\ 0 & 1 & 0 & 2 \end{bmatrix}$$

- Possible assignment for a path of length L :

$$\sum_{i,j} BPT^L[i,j]$$

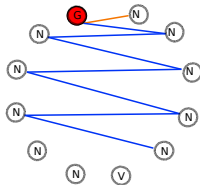
Counting the correct way

(((((. . .))))) .
(((((. . . .)))))
GNNNNNVNNNNN

Flamm, C., et al., 2001, Design of multistable RNA molecules, **RNA**, 7(2), 254-256

Counting the correct way

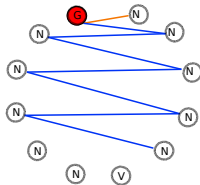
((((...)))).
(((.....)))
GNNNNNVNNNN



Flamm, C., et al., 2001, Design of multistable RNA molecules, *RNA*, 7(2), 254-256

Counting the correct way

((((...))) .
(((.....)))
GNNNNNVNNNN

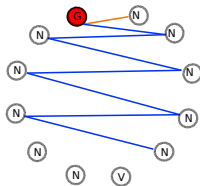


$$BPT^7 = \begin{bmatrix} 0 & 8 & 0 & 13 \\ 8 & 0 & 13 & 0 \\ 0 & 13 & 0 & 21 \\ 13 & 0 & 21 & 0 \end{bmatrix}$$

Flamm, C., et al., 2001, Design of multistable RNA molecules, *RNA*, 7(2), 254-256

Counting the correct way

((((...))) .
(((.....)))
GNNNNNVNNNN

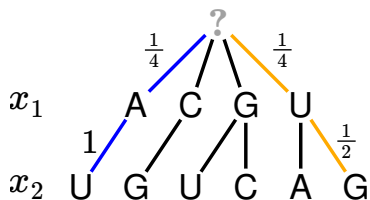


$$BPT^7 = \begin{bmatrix} 0 & 8 & 0 & 13 \\ 8 & 0 & 13 & 0 \\ 0 & 13 & 0 & 21 \\ 13 & 0 & 21 & 0 \end{bmatrix}$$

$$2 \times (13 + 21) \times 4^2 \times 3 = 3,264$$

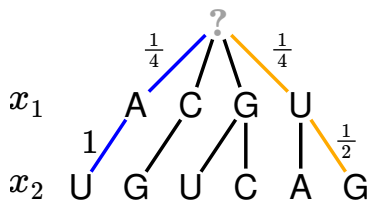
Flamm, C., et al., 2001, Design of multistable RNA molecules, *RNA*, 7(2), 254-256

Uniform Sampling of a Base Pair



Flamm, C., et al., 2001, Design of multistable RNA molecules, *RNA*, 7(2), 254-256

Uniform Sampling of a Base Pair

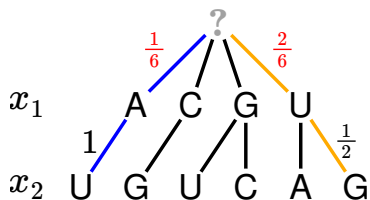


uniform decision \neq uniform sampling

$$P(AU) = \frac{1}{4} \times 1 = \frac{1}{4}$$

$$P(UG) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$$

Uniform Sampling of a Base Pair

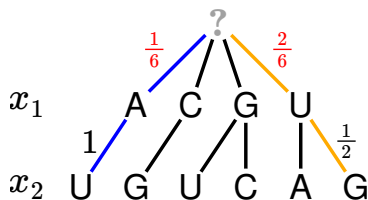


uniform decision \neq uniform sampling

$$P(AU) = \frac{1}{4} \times 1 = \frac{1}{4}$$

$$P(UG) = \frac{1}{4} \times \frac{1}{2} = \frac{1}{8}$$

Uniform Sampling of a Base Pair



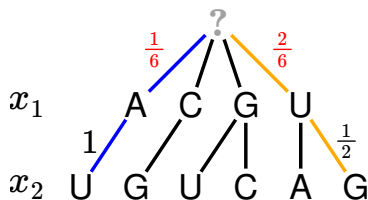
weighted decision = uniform sampling

$$P(x_1 = U) = \frac{N(\{x_1, x_2\} | x_1 = U)}{N(x_1, x_2)} = \frac{2}{6}$$

$$P(AU) = \frac{1}{6} \times 1 = \frac{1}{6}$$

$$P(UG) = \frac{2}{6} \times \frac{1}{2} = \frac{1}{6}$$

Uniform Sampling of a Base Pair



weighted decision = uniform sampling

$$P(x_1 = U) = \frac{N(\{x_1, x_2\} | x_1 = U)}{N(x_1, x_2)} = \frac{2}{6}$$

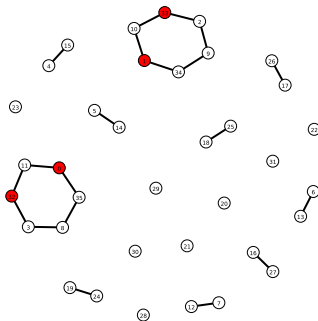
$$P(AU) = \frac{1}{6} \times 1 = \frac{1}{6}$$

$$P(UG) = \frac{2}{6} \times \frac{1}{2} = \frac{1}{6}$$

Example: 35nt long and three structural states

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

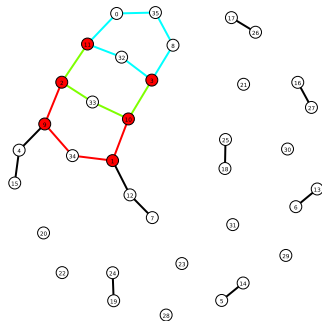
Size of solution space: 1.42658e+14



Example: 35nt long and four structural states

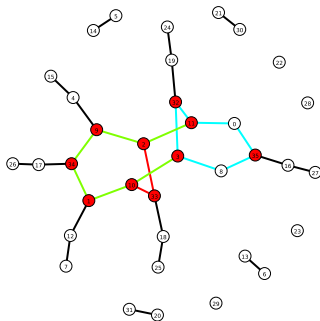
```
(((((.....)))) ..... ((((((.....)))))) .....  
..... ((((((..... ((((((.....)))))) .....))))))  
(((((((((((.....)))))) ((((((.....)))))) .....))))  
..... ((((((.....)))))) ..... ((((((.....)))))) .....  
.....
```

Size of solution space: 1.24018e+13



Example: 35nt long and five structural states

```
(((((.....)))) ..... ((((((.....)))))) .....  
..... ((((((..... ((((((.....)))))) .....))))))  
(((((((((((.....)))))) ((((((.....)))))) .....))))  
..... ((((((.....)))))) ..... ((((((.....)))))) .....  
..... ((((((.....)))))) ..... (((((((.....)))))))))  
Size of solution space: 7.08853e+10
```

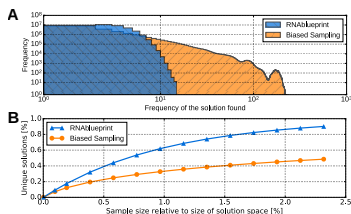




RNAbp blueprint

A tool that implements a graph coloring approach to **sample nucleic acid sequences** . . .

- compatible to **multiple structural and sequence constraints**
- **fairly drawn** from the whole solution space



<https://github.com/ViennaRNA/RNAbp blueprint>

Hammer, S., et al., 2017, RNAbp blueprint: flexible multiple target nucleic acid sequence design. *Bioinformatics* 33(18): 2850-58.

Objective Function

- Not all compatible sequences are equally good when evaluated under certain assumptions.

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

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

GNNNNNVNNNNN

$$f(x) = \underbrace{((EOS_1 + EOS_2) - 2EFE)}_{\text{dominate ensemble of structures}} + \underbrace{\xi |EOS_1 - EOS_2|}_{\text{equally stable}}$$

$$f(\text{GGGAGCCCC}) = 3.82 \quad f(\text{GAAUCAUUU}) = 8.42$$

$$f(\text{GGGUCCCCU}) = 4.71 \quad f(\text{GGGGAUUCU}) = 5.71$$

- Depending on the application the objective function needs to be developed.

Objective Function

- Not all compatible sequences are equally good when evaluated under certain assumptions.

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

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

GNNNNNVNNNN

$$f(x) = \underbrace{((EOS_1 + EOS_2) - 2EFE)}_{\text{dominate ensemble of structures}} + \underbrace{\xi |EOS_1 - EOS_2|}_{\text{equally stable}}$$

$$f(\text{GGGAGCCCC}) = 3.82 \quad f(\text{GAAUCAUUU}) = 8.42$$

$$f(\text{GGGUCCCCU}) = 4.71 \quad f(\text{GGGGAUUCU}) = 5.71$$

- Depending on the application the objective function needs to be developed.

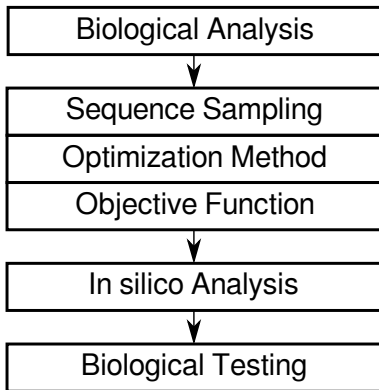
Sequence \Leftarrow Structure \Leftarrow Function

- 1 Which sequence optimally folds into given target structure(s)?
- 2 How to implement novel functions?

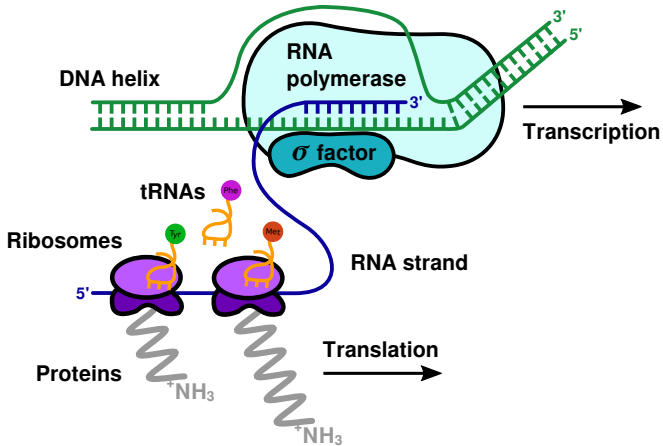
Sequence \Leftarrow Structure \Leftarrow Function

- 1 Which sequence optimally folds into given target structure(s)?
- 2 How to implement novel functions?

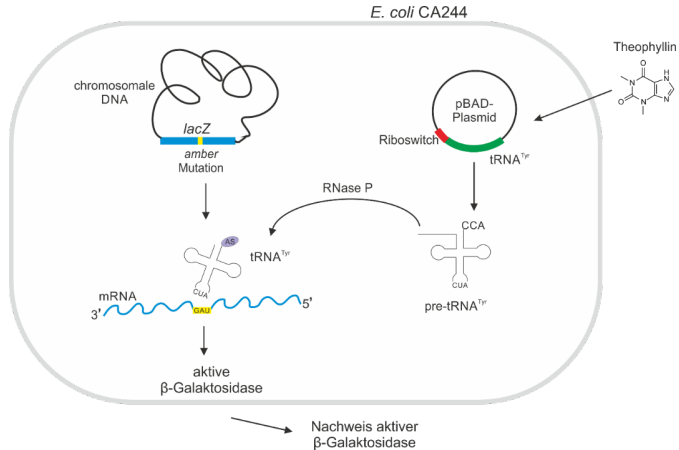
Design Pipeline



Transcription/Translation

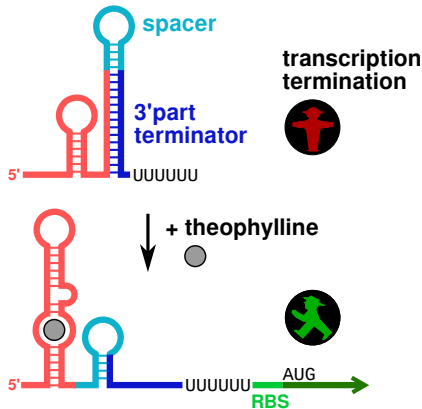
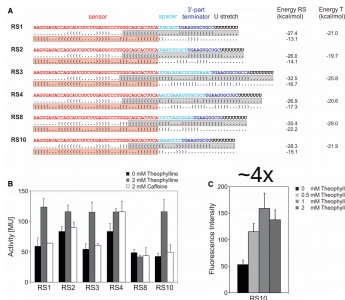


tRNA processing mechanism

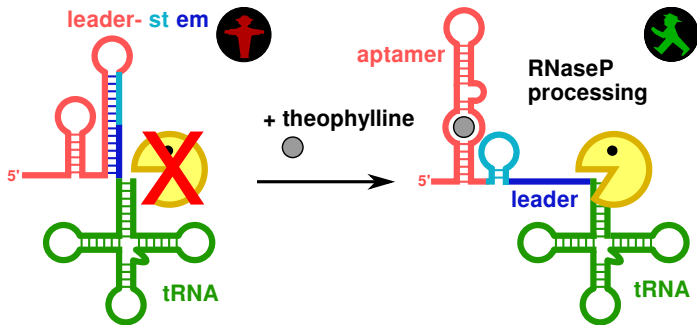


Artificial Riboswitch

- Theophylline triggered
- Transcriptional
- Reporter Gene



RNaseP riboswitch idea



Copy&Paste Design

construct I (RNaseP RS, R273)

> Theophylline RS w/o poly U fused to su3+

AAGUGAUACCAGCAUCGUCUUGAUGCCCUUGGCAGCACUUCA GAAAUCUCUGAAGUGCUGCCAA GGUGGGGUUCC
CGAGCGGCCAAAGGGAGCAGACUCUAAAUCUGCCGUCAUCGACUUCGAAGGUUUCGAAUCCUUC CCCCACCACCA

cleavage



tRNA spacer aptamer

- This construct is without the original 5'-leader sequence of the tRNA

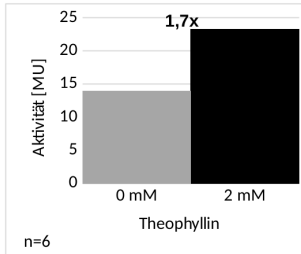
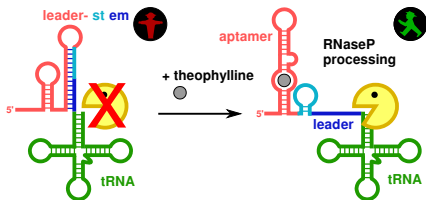
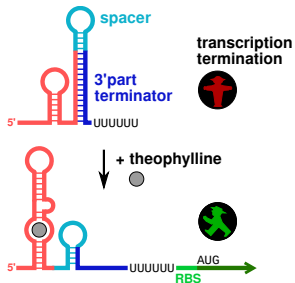
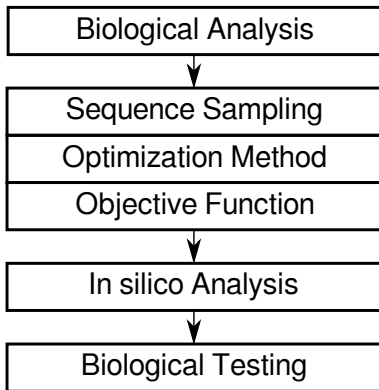


Figure 1: ONPG- assay with construct I (RNase R RS).

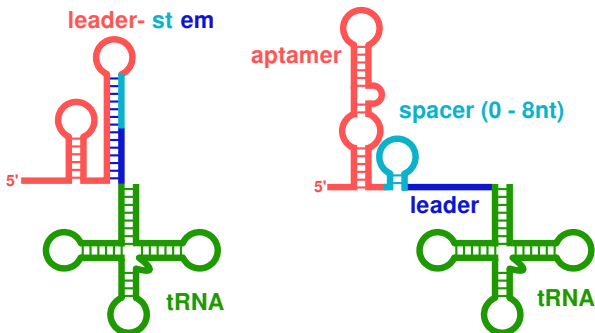
Timing is important!



Design Pipeline



Constrained Sampling



Sequence + Structure constraints →

RNAblueprint

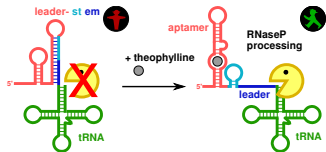


→ Sequences

Objective Functions

- Accessibility score [0,1]:

$$f(x) = P_{x, C_{theo}}(\Theta_{leader}) \cdot (1 - P_x(\Theta_{leader}))$$



- State score [0,1]:

$$f(x) = \underbrace{P_x(\Theta_{leader-stem})}_{\text{leader-stem} \rightarrow 1} \cdot \underbrace{\left(\frac{1 - P_x(\Theta_{aptamer}) + P_x(\Theta_{leader-stem})}{2} \right)^2}_{\text{difference to aptamer} \rightarrow 1}$$

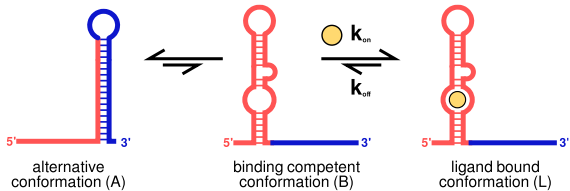
$$\cdot \underbrace{P_{x, C_{theo}}(\Theta_{aptamer})}_{\text{aptamer} \rightarrow 1} \cdot \underbrace{\left(\frac{1 - P_{x, C_{theo}}(\Theta_{leader-stem}) + P_{x, C_{theo}}(\Theta_{aptamer})}{2} \right)^2}_{\text{difference to leader-stem} \rightarrow 1}$$

x ... Sequence

Θ ... Structure

C_{theo} ... Soft-Constraint for Theophylline

LIGAND DEPENDENT RNA SWITCH



Findeiß, S., Hammer, S., Wolfinger, M. T., Kühnl, F., Flamm, C., & Hofacker, I. L. (2018). In silico design of ligand triggered RNA switches. *Methods*.

Thanks to...

lab members:

- Stefan Hammer
- Felix Kühnl
- Peter F. Stadler

- Petra Pregel
- Jens Steuck

collaborators:

- Anna Ender, Leipzig
- Chris Günzel, Leipzig
- Mario Mörl, Leipzig
- Christina Weinberg, Leipzig

- Ilka Axmann, Düsseldorf
- Sebastian Will, Vienna
- Christoph Flamm, Vienna
- Ivo L. Hofacker, Vienna
- Yann Ponty, Palaiseau
- Michael Ryckelynck, Strasbourg

