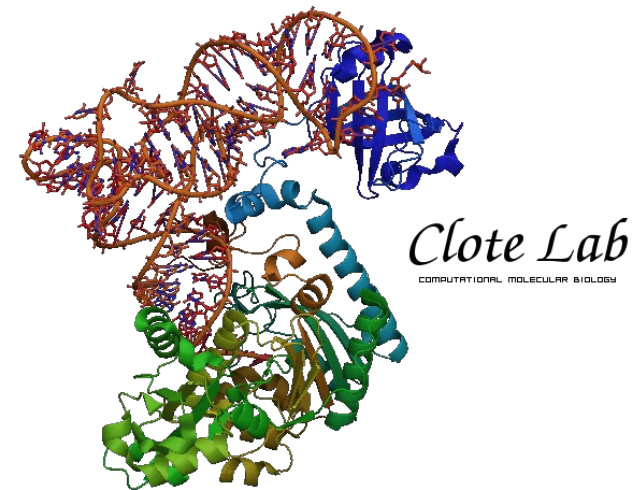


RNAiFold: Complete Inverse Folding for Synthetic Design



Juan Antonio García Martín
23/07/2015

Synthetic RNA design approaches

Computational

Energy model based

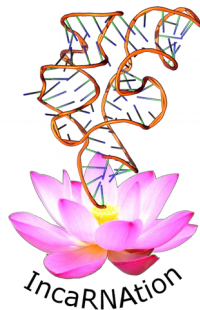
Nearest Neighbor DB



NUPACK nucleic acid package

RNAinverse

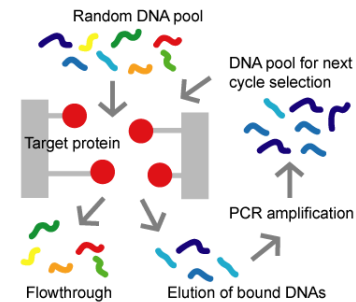
Weighted sampling/Similarity



AD-HOC

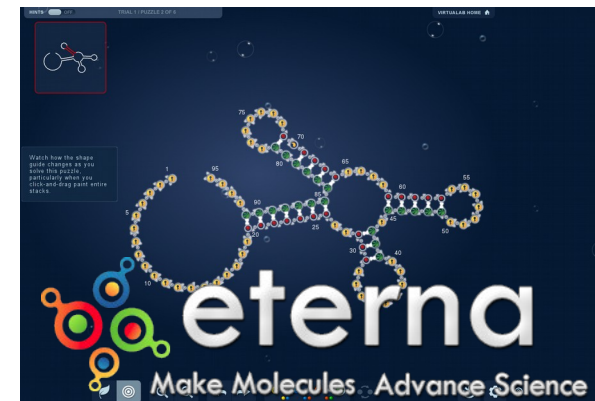
Experimental

SELEX



Collaborative

Crowd science



Objective

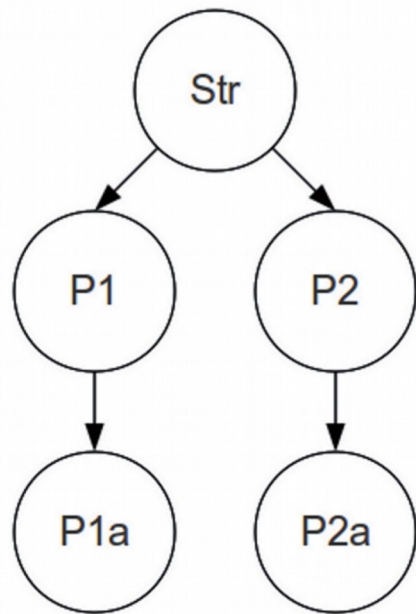
Novel algorithm design and implementation of tools to design synthetic RNA

Approach

Generate a large number of sequences whose minimum free energy structure is identical to the target design structure, and subsequently filter with respect to different criteria in order to select the most promising candidates for biochemical validation

RNAiFold

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

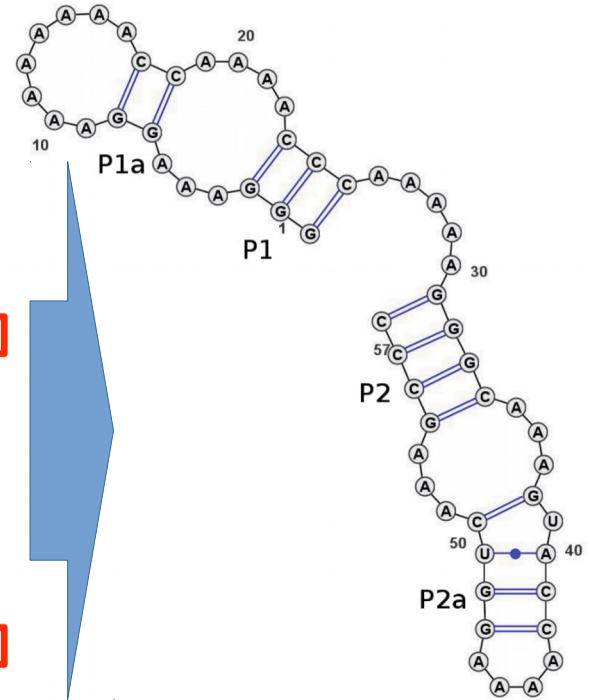


```

..(((.....(((.....))))))
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNGNNGNCNNNNNNNNNN
NNNNNNNNNNCGNNNCGNNNNNNNNN
NNNNNNNNNNGCGNNNCGCNNNNNNNN
NNNNNNNNNNGCGANNCGCNNNNNNNN
NNNNNNNNNNGCGAGNCGCNNNNNNNN
NNNNNNNNNNGCGAGACGCNNNNNNNN
..(((.....(((.....))))))
NNNNNCNNNNGCGAGACGCNNGNNN
NNNNGCNNNNGCGAGACGCNNGCNN
NNNCGCNNNNGCGAGACGCNNGCGN
NNGCGCNNNNGCGAGACGCNNGCGC
NNGCGCANNNGCGAGACGCNNGCGC
NNGCGCAAAGCGAGACGCNNGCGC
NNGCGCAAAGCGAGACGCANGCGC
NNGCGCAAAGCGAGACGCAAGCGC
ANGCGCAAAGCGAGACGCAAGCGC
AAGCGCAAAGCGAGACGCAAGCGC
..(((.....(((.....))))))
  
```

MFE helix check

MFE helix check



Additional features

- Cofolding
- Template sequence
- GC%
- Limit #base pairs per type
- Consecutive bases
- Incompatible base pairs

First complete inverse folding



Performance

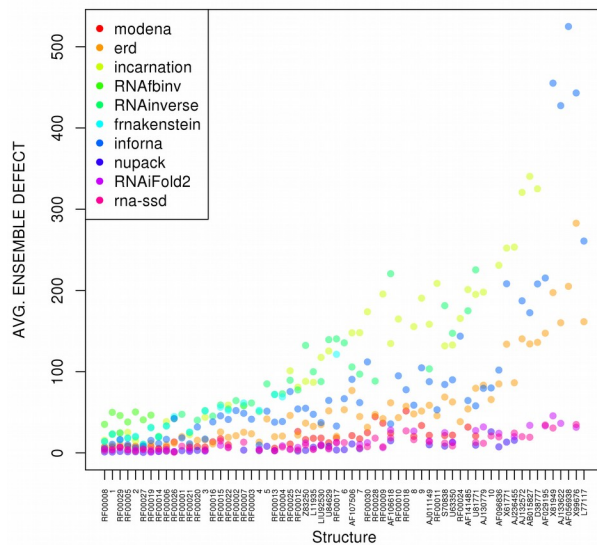
Table 1. Rfam CP Results.

| Parameters | | CP | | INFO-RNA | | MODENA | | RNA-SSD | | RNAinverse | |
|-------------|----------|-------------|--------------|-----------|------------|-----------|-------------|-----------|-------------|------------|------------|
| RF id | <i>n</i> | sol | time | sol | time | sol | time | sol | time | sol | time |
| RF00001.121 | 117 | 38 | 21.5 | 50 | 0.0 | 6 | 36.8 | 22 | 1.0 | 41 | 233.1 |
| RF00002.2 | 151 | 44 | 29.5 | 4 | 62.6 | 20 | 39.4 | 6 | 12.2 | 0 | - |
| RF00003.94 | 161 | 0 | - | 1 | 72.1 | 29 | 70.2 | 0 | - | 0 | - |
| RF00004.126 | 193 | 50 | 1.5 | 50 | 0.1 | 34 | 52.9 | 50 | 2.0 | 50 | 48.3 |
| RF00005.1 | 74 | 50 | 0.2 | 50 | 0.0 | 33 | 12.4 | 50 | 0.1 | 50 | 0.1 |
| RF00006.1 | 89 | 50 | 0.3 | 50 | 0.0 | 37 | 15.1 | 50 | 0.6 | 50 | 4.3 |
| RF00007.20 | 154 | 50 | 5.6 | 50 | 0.0 | 34 | 44.4 | 50 | 1.1 | 50 | 12.4 |
| RF00008.11 | 54 | 50 | 0.1 | 50 | 0.0 | 26 | 8.7 | 50 | 0.0 | 50 | 0.0 |
| RF00009.115 | 348 | 48 | 20.8 | 0 | - | 29 | 214.1 | 26 | 48.2 | 0 | - |
| RF00010.253 | 357 | 0 | - | 0 | - | 0 | - | 0 | - | 0 | - |
| RF00011.18 | 382 | 0 | - | 0 | - | 0 | - | 0 | - | 0 | - |
| RF00012.15 | 215 | 50 | 2.7 | 15 | 25.0 | 27 | 64.5 | 28 | 28.8 | 1 | 139.4 |
| RF00013.139 | 185 | 50 | 1.6 | 50 | 0.8 | 12 | 51.5 | 49 | 2.8 | 50 | 19.8 |
| RF00014.2 | 87 | 50 | 0.3 | 50 | 0.0 | 33 | 17.5 | 49 | 0.1 | 50 | 0.0 |
| RF00015.101 | 140 | 49 | 1.3 | 50 | 0.2 | 38 | 29.1 | 40 | 0.6 | 50 | 52.4 |
| RF00016.15 | 129 | 0 | - | 0 | - | 0 | - | 0 | - | 0 | - |
| RF00017.90 | 301 | 50 | 19.3 | 50 | 0.0 | 28 | 208.1 | 50 | 7.0 | 50 | 10.0 |
| RF00018.2 | 360 | 47 | 12.1 | 1 | 697.0 | 28 | 331.5 | 0 | - | 0 | - |
| RF00019.115 | 83 | 50 | 0.2 | 50 | 0.0 | 32 | 14.9 | 50 | 0.2 | 50 | 0.3 |
| RF00020.107 | 119 | 0 | - | 0 | - | 0 | - | 0 | - | 0 | - |
| RF00021.10 | 118 | 50 | 0.3 | 50 | 0.0 | 37 | 27.8 | 49 | 0.2 | 50 | 0.2 |
| RF00022.1 | 148 | 50 | 0.7 | 50 | 0.0 | 38 | 32.6 | 24 | 0.9 | 35 | 225.5 |
| RF00024.16 | 451 | 0 | - | 0 | - | 0 | - | 0 | - | 0 | - |
| RF00025.12 | 210 | 50 | 1.4 | 9 | 47.9 | 33 | 54.2 | 29 | 2.9 | 0 | - |
| RF00026.1 | 102 | 50 | 0.4 | 33 | 5.5 | 38 | 15.2 | 50 | 1.4 | 44 | 173.2 |
| RF00027.7 | 79 | 50 | 0.1 | 50 | 0.0 | 32 | 17.4 | 50 | 0.1 | 50 | 0.4 |
| RF00028.1 | 344 | 39 | 6.2 | 0 | - | 0 | - | 4 | 71.2 | 0 | - |
| RF00029.107 | 73 | 50 | 0.3 | 50 | 0.0 | 37 | 10.4 | 50 | 0.2 | 50 | 0.3 |
| RF00030.30 | 340 | 46 | 6.8 | 1 | 57.3 | 22 | 186.8 | 34 | 39.3 | 0 | - |
| sum | - | 1111 | 133.2 | 813 | 271.5 | 683 | 1555.5 | 860 | 220.9 | 771 | 919.7 |
| avg | - | 38.3 | 5.7 | 28.0 | 12.9 | 23.6 | 67.6 | 29.7 | 10.0 | 26.6 | 54.1 |

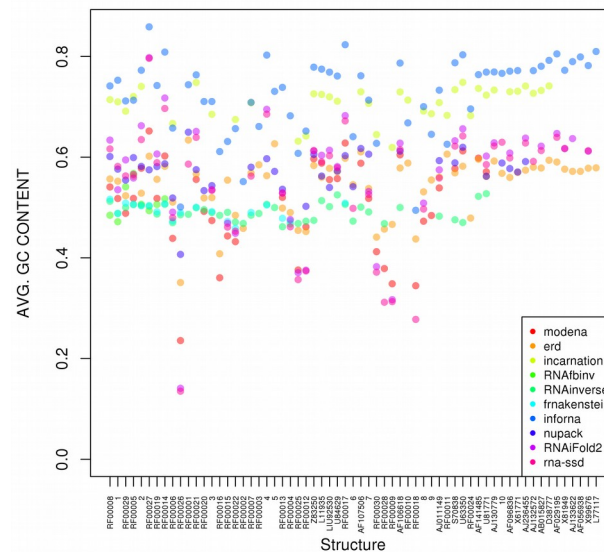
Output sequences

| Method | ERD | FRNA | Incarnation | Info-RNA | MODENA | Nupack | RNA-SSD | RNAfbinv | RNAiFold2 | RNAinverse |
|------------------|-------|-------|-------------|----------|--------|--------|---------|----------|-----------|------------|
| Output (%) | 100% | 30% | 60% | 95% | 60% | 57% | 90% | 13% | 65% | 65% |
| Target (%) | 85% | 38% | 0% | 57% | 45% | 70% | 82% | 0% | 100% | 18% |
| Avg str len | 397 | 122 | 352 | 393 | 234 | 256 | 400 | 74 | 363 | 208 |
| Avg output | 117 | 325 | 41,535 | 195 | 50 | 22 | 1 | 2 | 55,476 | 935 |
| P(S) | 3.32% | 1.70% | 0.06% | 3.17% | 11.30% | 30.01% | 2.24% | 0.36% | 23.21% | 0.78% |
| Native cont. (%) | 85±9 | 61±15 | 63±13 | 76±12 | 89±9 | 98±1 | 85 | 32±6 | 93±2 | 57±12 |
| Avg E | -0.41 | -0.24 | -0.46 | -0.63 | -0.46 | -0.44 | -0.30 | -0.14 | -0.56 | -0.23 |
| Pos entropy | 0.33 | 0.71 | 0.41 | 0.44 | 0.15 | 0.07 | 0.36 | 0.88 | 0.12 | 0.80 |
| MH diversity | 0.16 | 0.35 | 0.21 | 0.22 | 0.07 | 0.03 | 0.18 | 0.45 | 0.06 | 0.38 |
| Vienna diversity | 0.11 | 0.23 | 0.15 | 0.16 | 0.05 | 0.02 | 0.11 | 0.30 | 0.05 | 0.26 |
| Exp bp dist | 0.09 | 0.21 | 0.27 | 0.16 | 0.06 | 0.01 | 0.08 | 0.38 | 0.03 | 0.24 |
| Ens def | 0.14 | 0.32 | 0.39 | 0.22 | 0.08 | 0.02 | 0.14 | 0.56 | 0.04 | 0.37 |
| Exp num bp | 0.28 | 0.29 | 0.34 | 0.30 | 0.26 | 0.29 | 0.28 | 0.28 | 0.27 | 0.28 |
| GC-content (%) | 55% | 49% | 71% | 72% | 50% | 57% | 36% | 51% | 57% | 49% |

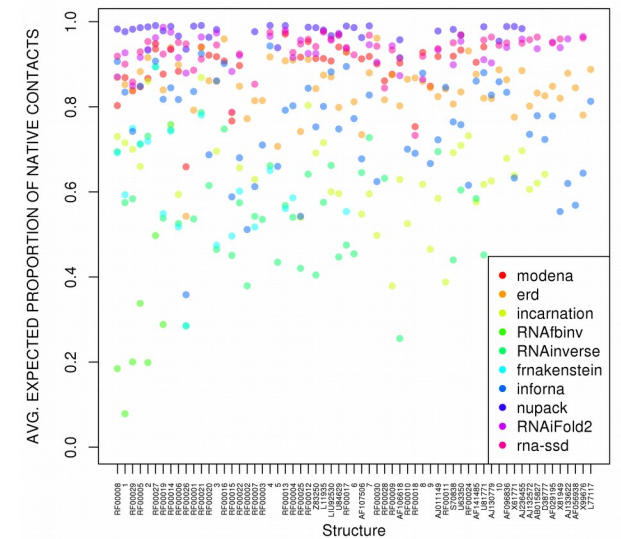
AVG. ENSEMBLE DEFECT



AVG. GC CONTENT



AVG. EXPECTED PROPORTION OF NATIVE CONTACTS



RNAiFold 1.0: Limitations

- Memory usage
 - 32 bit software limits memory to 4GB
- Sequence length
 - Complete search is unpractical for very long structures
- COMET availability
 - Available only via webserver



RNAiFold 2.0

- Implemented in C++ using OR-Tools (open source).
- Modular and scalable

| Software | ↓ | WS | PK | H | MT | PT | T | EM | D | SeqC | StrC | AaC | O | Num |
|--------------|---|----|----|----|----|----|---|---------|---------|---------------|------|-----|-------------|------|
| RNAiFold 2.0 | ✓ | ✓ | — | ✓ | — | ✓ | ✓ | '99,'04 | 0,1,2,3 | ✓ | ✓ | ✓ | mfe | MAX |
| RNAinverse | ✓ | ✓ | — | — | — | — | ✓ | '99,'04 | 0,1,2,3 | IUPAC★ | — | — | mfe, prob | 100 |
| RNA-SSD | — | ✓ | — | — | — | — | ✓ | '99 | 1 | IUPAC★ | — | — | mfe | 10 |
| Info-RNA | ✓ | ✓ | — | — | — | — | — | '04 | 1 | IUPAC | — | — | mfe, prob | 50 |
| NUPACK | ✓ | ✓ | — | ✓★ | — | — | ✓ | '99,'04 | 0,1,2 | ✓ | — | — | ens def | 10 |
| MODENA | ✓ | — | ✓ | — | — | — | — | I | def | — | — | — | mfe, prob | ? |
| Frnakenstein | ✓ | — | — | — | ✓ | — | ✓ | I | def | — | — | — | various | ? |
| IncaRNation | ✓ | — | — | — | — | — | ✓ | '04★ | — | IUPAC | — | — | pf sampling | — |
| ERD | ✓ | ✓ | — | — | — | — | ✓ | I | def | IUPAC★ | — | — | mfe | MAX★ |
| RNAdesign | ✓ | — | — | — | ✓ | — | ✓ | '04 | def | — | — | — | various | — |
| RNAfbinv | ✓ | — | — | — | — | ✓ | — | '99, I | def | local A,C,G,U | — | — | mfe | — |

- Partial target structure(**)
- Sequence constraints
 - GC-content (*)
 - Number of base pairs(***)
 - Maximum, minimum and consecutive nucleotides by region(***)
 - Structural compatibility and incompatibility constraints (***)
- Amino acid coding constraints (***)
 - Partial and overlapping regions
 - Restrict search to amino acids similar in BLOSUM62 similarity score
 - Maximization of BLOSUM62 similarity score

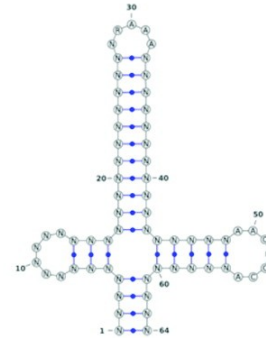
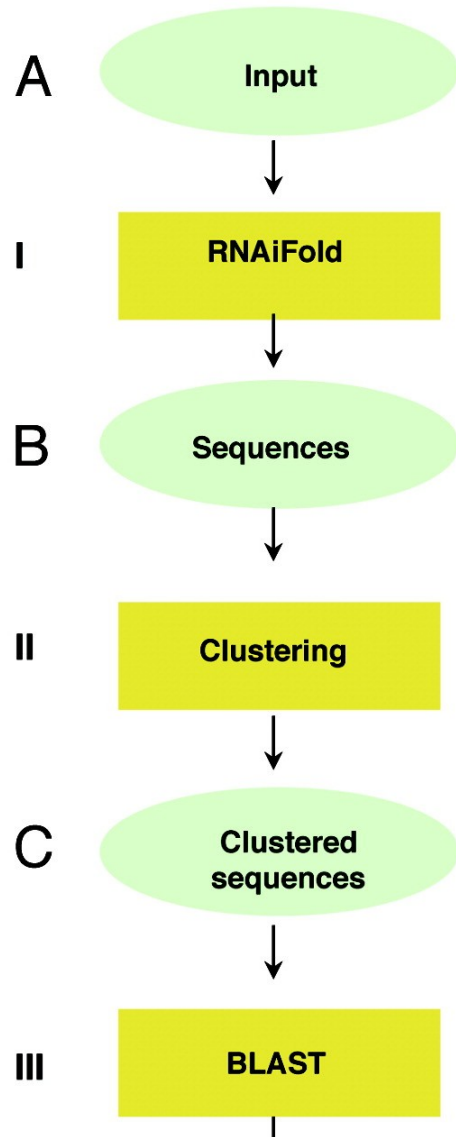
Supported by:

(*) Some other methods

(**) Only one other software

(***) RNAiFold only

Applications (I): IRES discovery

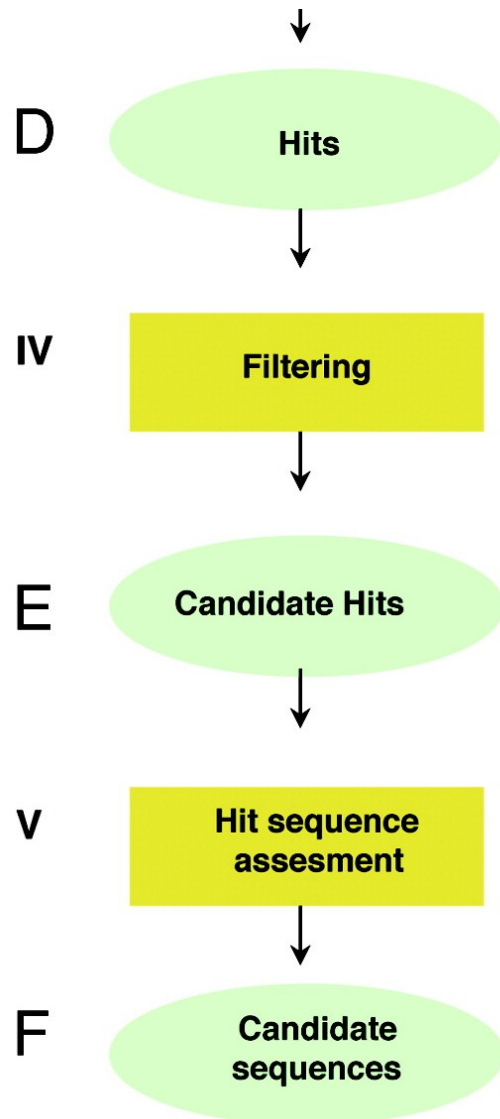


```

    CCGGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGG
    GCGGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGCC
    GACGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGUC
    CACGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGUG
    GCGGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGGG
    CCCGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGGG
    ACCGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGGU
    UCCGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGGA
    GUCGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGAC
    CUCGGCCAAAAAGGCGUGGCGGGGAAAAACCCCGCGCGCCGCCAACCCAGGCGGCGAG
  
```



Applications(I): IRES discovery



---Sequence:
CGCGGCCAAAAAGGCGCGUGGCGGGGUAACCCCGCCGCGCCUGUAACCCACGCAGCGCG

gb|CP001840.1| Bifidobacterium bifidum PRL2010, complete genome 42 0.045
 gb|CP002220.1| Bifidobacterium bifidum S17, complete genome 42 0.045
 gb|CP001001.1| Methylobacterium radiotolerans JCM 2831, complete... 39 0.35
 gb|CP002279.1| Mesorhizobium opportunistum WSM2075, complete genome 38 0.97
 gb|HQ404344.1| Uncultured archaeon clone LJ108_1 16S ribosomal R... 36 2.7
 ref|NM_177289.1| Mus musculus core-binding factor, runt domain, ... 36 2.7 ref|
 NM_001109873.1| Mus musculus core-binding factor, runt domai... 36 2.7

1) Keywords:

- factor, receptor, mRNA, transcription, homeo box, leucine zipper, translation, kinase, phosphatidyl inositol, helicase, control, ribosome, protein

2) Location:

- 5' UTR
 - Beginning of the coding region

---Sequence:
AGAGGCCAAAAAGGCGCGUGGCGGGGUGAAACCCCGCCGCGCCGAAAACCCCAUUCGGCUCU
 ref|NM_079437.3| Drosophila melanogaster TBP-associated factor 6... 36 2.7

---Sequence:
CGGUGGCCAAAAAGGCGCGUGGCGGGGAAAAACCCCGCCGCGCCCGCAACCCAGCGGGACCG
 ref|NM_214436.1| Sus scrofa beta-galactosamide alpha-2,6-sialylt... 43 0.016

---Sequence:
GGUGGCCAAAAAGGCGCGUGGCGGGGUAACCCCGCCGCGCCCGCAACCCAGCGGGCGCC
 ref|NM_001205770.1| Bos taurus tumor necrosis factor (ligand) su... 38 0.97

1. Codon usage bias

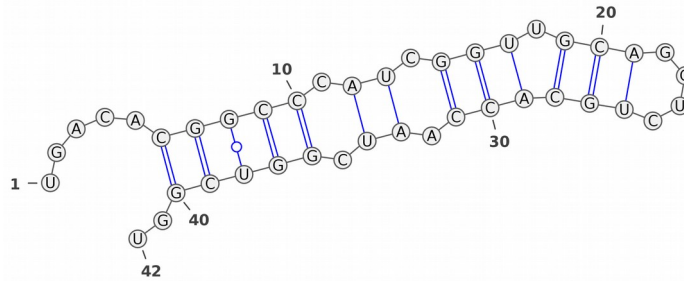
2. Total Variation:

$$\frac{1}{\sum_{i=1}^8 N(i)} \sum_{i=1}^8 N(i) \cdot \left(\frac{1}{2} \sum_{j \in C(i)} |P_i(j) - Q_i(j)| \right)$$

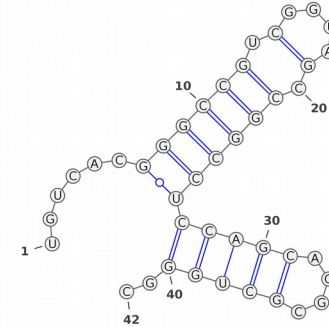
aaggaaatcgacctaggagaaatcacatccaccaactctgtaaaaattcc
 cctggatctcaccctgcgctcccattggttgtgtggaggga**gtgcaac**
 ccactgtgcccgaaaacccccctccgctctcgaaggattcccagttactg

Applications(II): SECIS insertion

Model: formate dehydrogenase F (fdhF)

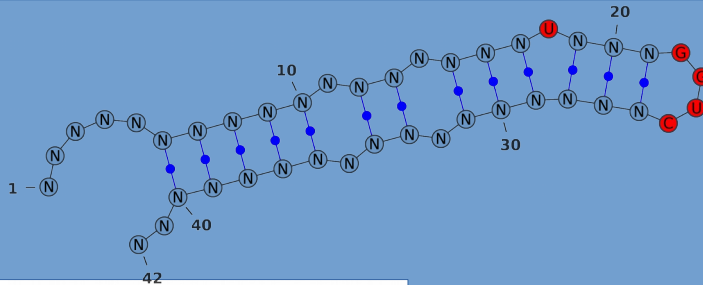


S. enterica (CDS70432.2)
UHGPSVAGLHQSVG



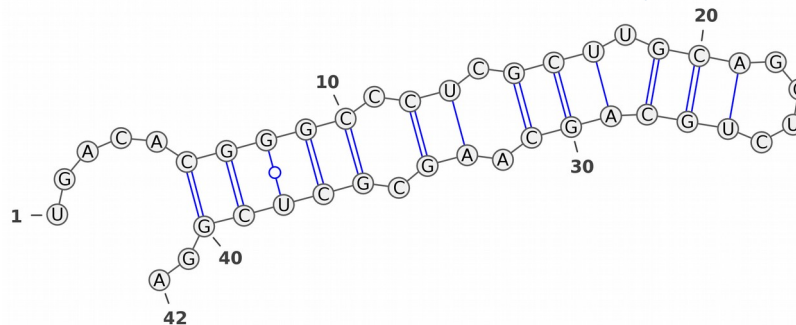
R. ornithinolytica (AJF73661.1)
CHGPSVAGLQQALG

RnaiFold 2.0
Constraints



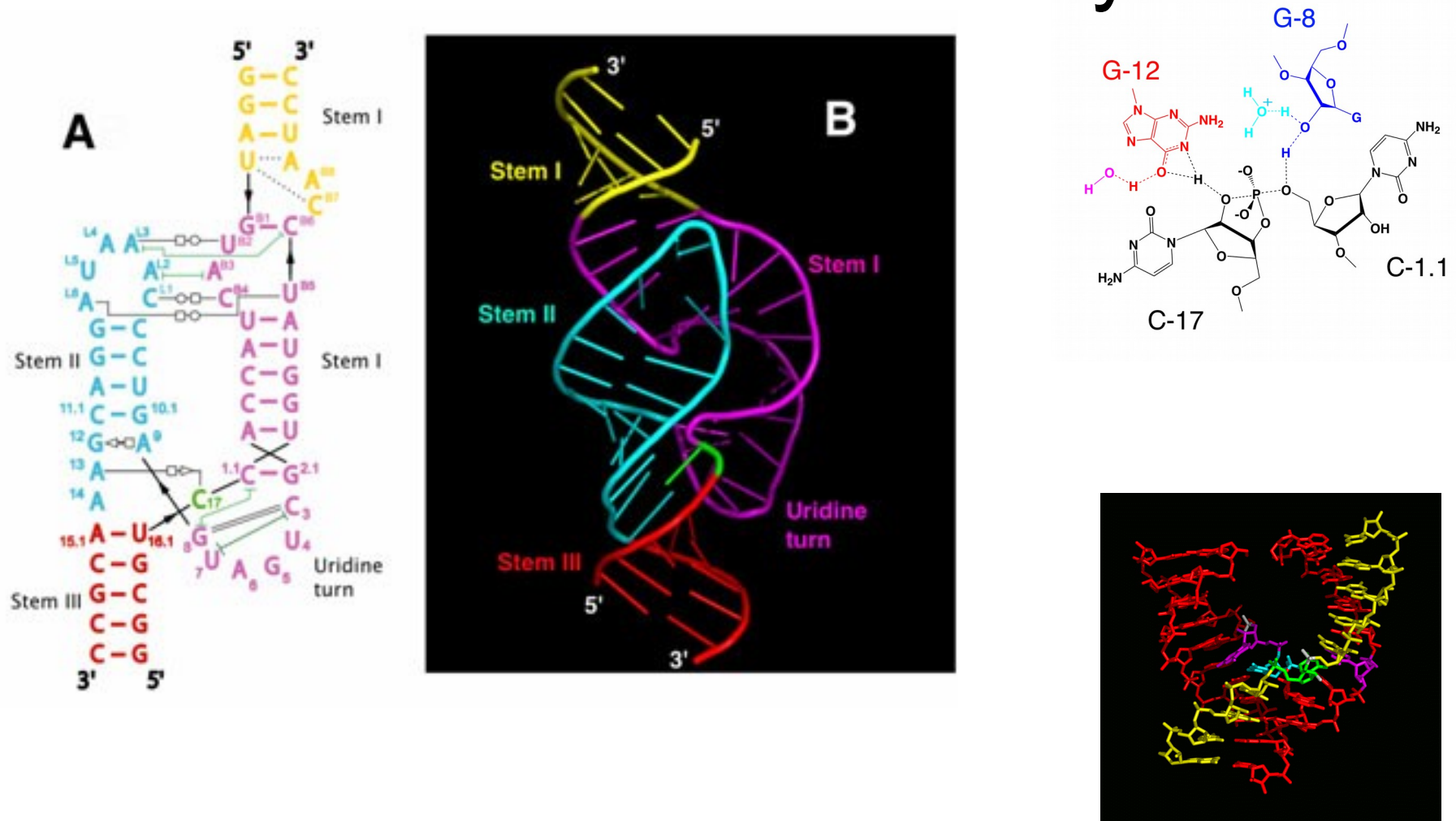
Amino acid target:
UHGPSVAGLQQALG
BLOSUM62 similarity \rightarrow -1
Maximize BLOSUM62 score

Solution



UHGPSVAGLQQALG
Score: 68 out of 71

Applications: Computational design of Hammerhead Ribozymes

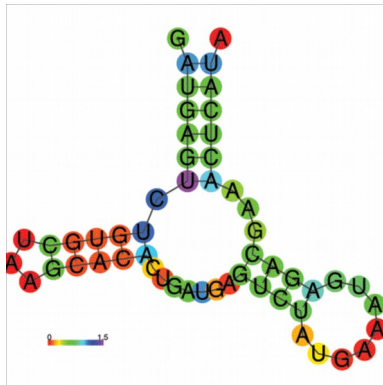


Figures: Monika Martick et al (2008),

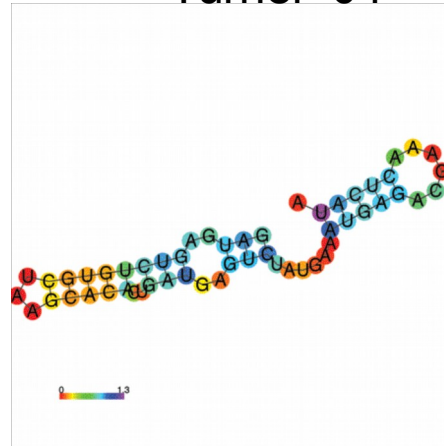
1. Sequence conservation

Peach Latent Mosaic Viroid
(PLMVd) AJ005312.1/282-335

Turner '99

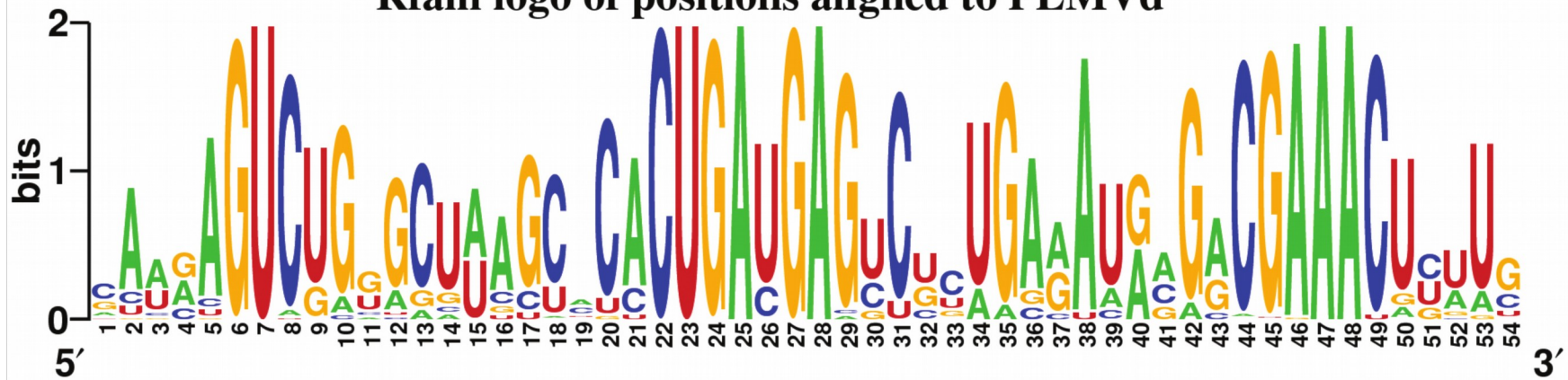


Turner '04



| Rank | Position (1-54) | Nucleotide | Frequency | Percentage |
|------|-----------------|------------|-----------|------------|
| 1 | 7 | U | 1 | 100% |
| 2 | 23 | U | 1 | 100% |
| 3 | 27 | G | 1 | 100% |
| 4 | 22 | C | 1 | 100% |
| 5 | 48 | A | 1 | 100% |
| 6 | 47 | A | 1 | 100% |
| 7 | 28 | A | 1 | 100% |
| 8 | 25 | A | 1 | 100% |
| 9 | 24 | G | 0.988095 | 99% |
| 10 | 46 | A | 0.988095 | 99% |
| 11 | 6 | G | 0.987342 | 99% |
| 12 | 45 | G | 0.97619 | 98% |
| 13 | 49 | C | 0.97561 | 98% |
| 14 | 29 | G | 0.964286 | 96% |
| 15 | 44 | C | 0.964286 | 96% |
| 16 | 38 | A | 0.957746 | 96% |
| 17 | 8 | C | 0.949367 | 95% |
| 18 | 35 | G | 0.942857 | 94% |
| 19 | 42 | G | 0.928571 | 93% |
| 20 | 31 | C | 0.892857 | 89% |

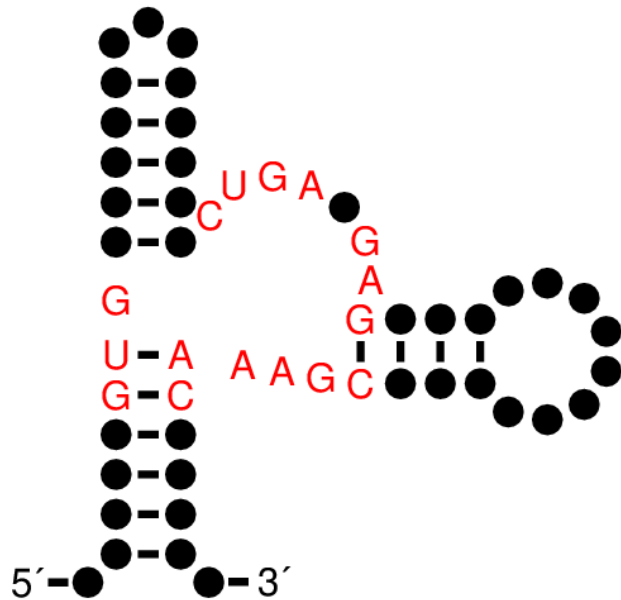
Rfam logo of positions aligned to PLMVd



Dotu I, Garcia-Martin JA, Slinger BL, Mechery V, Meyer MM, Clote P.

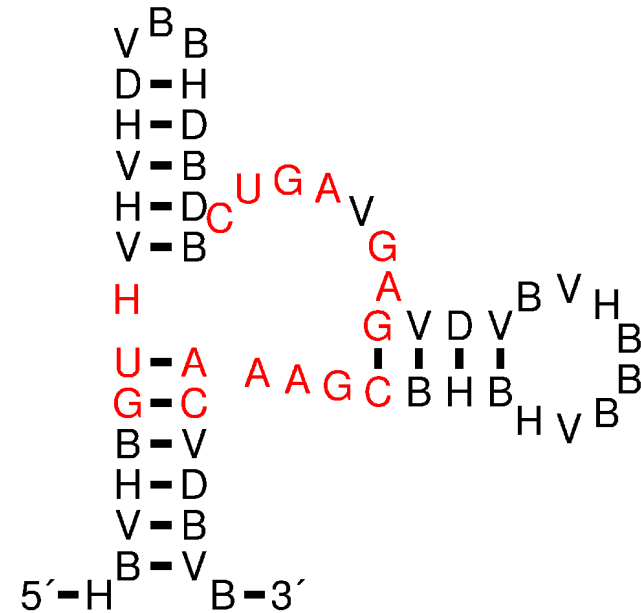
Complete RNA inverse folding: computational design of functional hammerhead ribozymes. *Nucleic Acids Res.* 2015;42(18):11752-62.

2. Generating sequences



RNAiFold

No possible solution



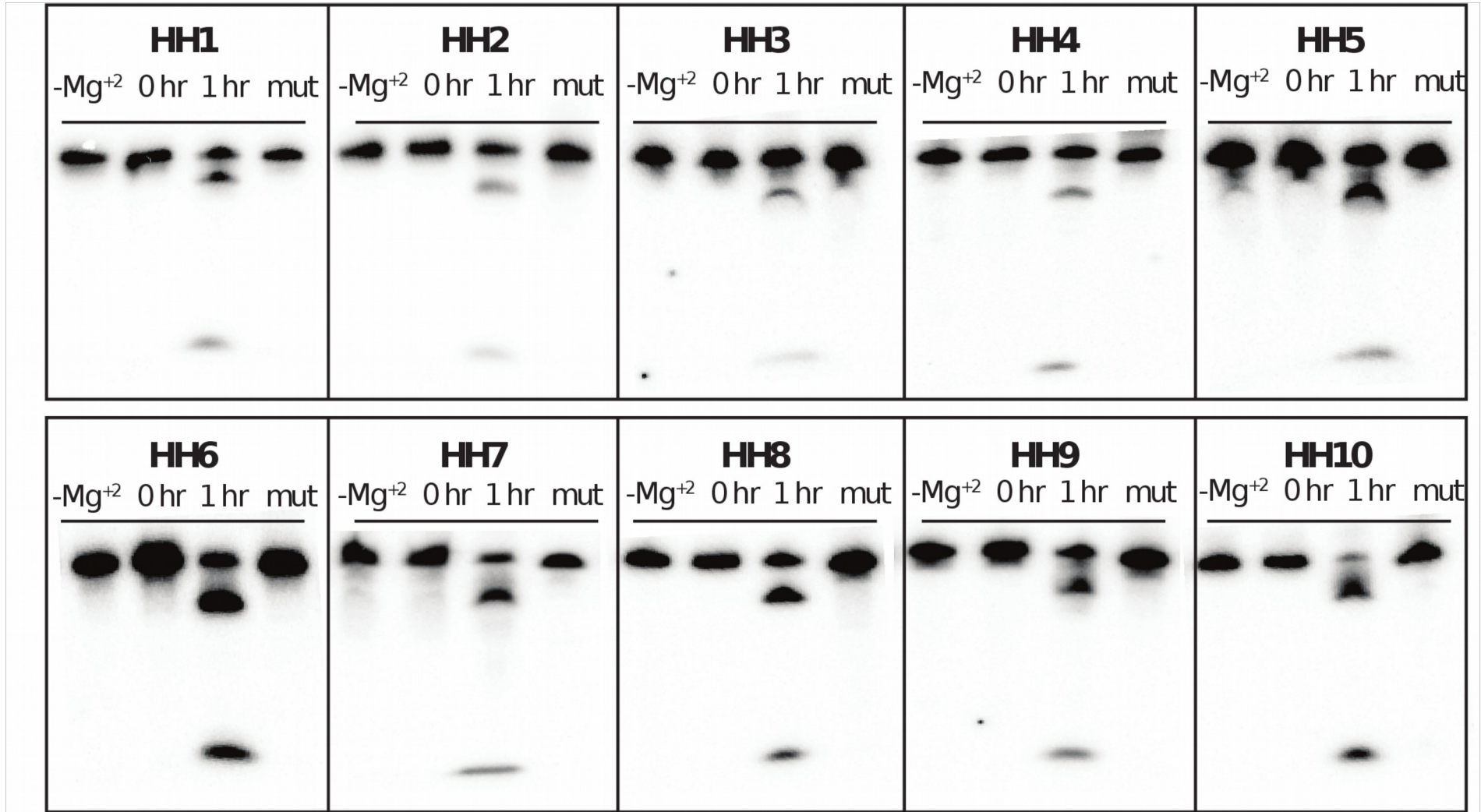
RNAiFold

~1.4 million sequences

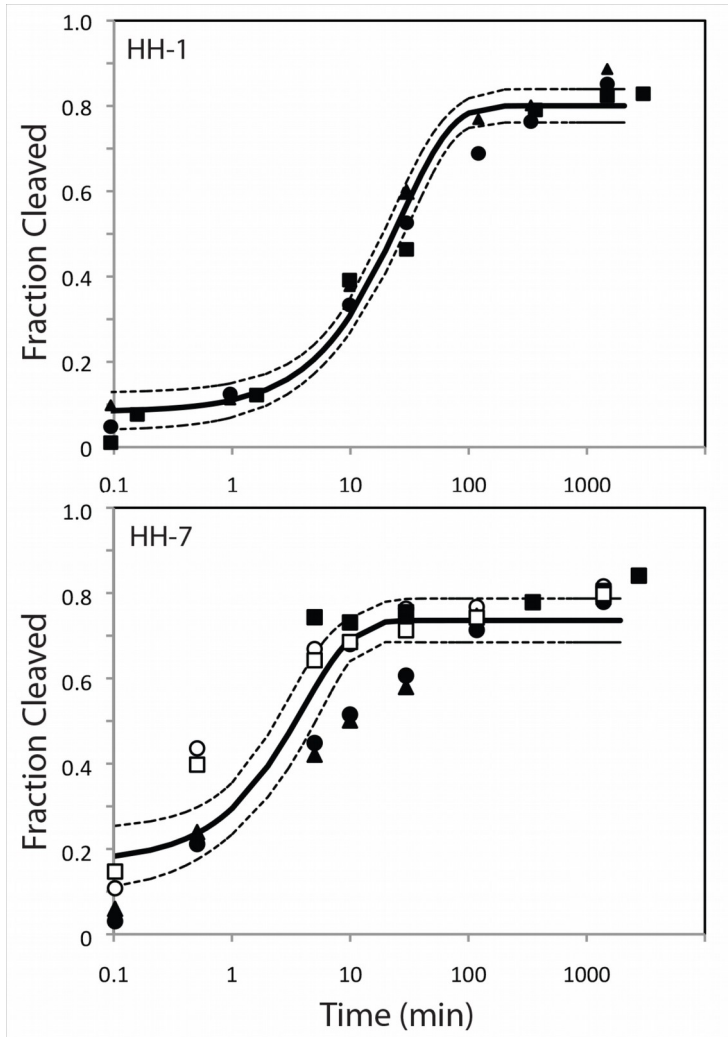
3. Filtering and selection

| ID | Selection criteria |
|------|--|
| HH1 | GC-content 30 – 39%, $P(S_0, s) \geq 40\%$, smallest (binary) entropy distance for active site |
| HH2 | GC-content 40 – 49%, $P(S_0, s) \geq 40\%$, smallest (binary) entropy distance for active site |
| HH3 | GC-content 40 – 49%, $P(S_0, s) \geq 40\%$, second smallest (binary) entropy distance for active site |
| HH4 | GC-content 50 – 59%, $P(S_0, s) \geq 40\%$, smallest (binary) entropy distance for active site |
| HH5 | GC-content 60 – 69%, $P(S_0, s) \geq 40\%$, smallest (binary) entropy distance for active site |
| HH6 | GC-content 30 – 39%, $P(S_0, s) \geq 40\%$, largest (binary) entropy distance for active site |
| HH7 | GUC in positions 6-8, smallest ensemble defect |
| HH8 | smallest ensemble defect, entropy and expected bp distance and highest Boltzmann probability |
| HH9 | $P(S_0, s) \leq 20\%$, smallest (binary) entropy distance for active site |
| HH10 | smallest (binary) entropy distance for active site |

4. Experimental validation



Kinetics



| ID | K_{obs} | F_{max} | MSE | Pos Ent | Ens Def | EBPD Dis Act |
|------|-----------|-----------|--------|---------|---------|--------------|
| HH1 | 0.037 | 0.79 | 0.0029 | 0.27 | 4.17 | 0.0501207 |
| HH2 | 0.0057 | 0.74 | 0.003 | 0.29 | 4.55 | 0.0386253 |
| HH3 | 0.0027 | 0.65 | 0.0039 | 0.26 | 4.12 | 0.0410984 |
| HH4 | 0.0127 | 0.55 | 0.0048 | 0.40 | 6.76 | 0.0354213 |
| HH5 | 0.0085 | 0.52 | 0.0066 | 0.38 | 6.24 | 0.033132 |
| HH6 | 0.102 | 0.73 | 0.0047 | 0.41 | 8.14 | 0.059864 |
| HH7 | 0.25 | 0.74 | 0.0107 | 0.12 | 2.38 | 0.0406728 |
| HH8 | 0.02 | 0.68 | 0.0124 | 0.08 | 1.45 | 0.0662421 |
| HH9 | 0.025 | 0.76 | 0.0015 | 0.25 | 4.53 | 0.0328018 |
| HH10 | 0.14 | 0.77 | 0.01 | 0.29 | 4.98 | 0.0269354 |

Correlations: Pos. entropy (-0.46) Ens. defect (-0.37)

RNAiFold web server

RNA Inverse Folding

Constraint Programming (CP) and Large Neighborhood Search (LNS) algorithms for inverse folding

Welcome to the Clote Lab Inverse Folding website !

[Servers](#) [Inv. fold](#) [Inv. cofold](#) [Structures](#) [Downloads](#) [References](#)

The *RNA inverse folding problem* is the problem, given a *target* secondary structure in dot bracket notation, of determining one or more RNA sequences, whose minimum free energy (MFE) structure is the target structure. Here, the MFE structure is computed using `RNAfold` from the [Vienna RNA Package](#). In addition, the user may provide sequence *constraints*, stipulating that certain positions be occupied by specific nucleotides, or that (for instance) the solution sequence has a GC-content within a certain user-specified range. This website provides access to two algorithms for the inverse folding problem:

RNA-CPdesign

- **RNA-CPdesign.** Given a target structure and optional sequence constraints, CPdesign uses *Constraint Programming (CP)* to determine one or more RNA sequences that fold into the given target structure. CP performs a complete exploration of the search space, and, thus can also prove that no sequence folds into the target structure exists. Since computation time may be exhorbitant, the latter is only feasible for sufficiently small structures.

RNA-LNSdesign

- **RNA-LNSdesign.** Given a target structure and optional sequence constraints, LNSdesign uses *Large Neighborhood Search (LNS)* to determine one or more RNA sequences that fold into the target structure. LNS is a heuristic, that calls CP as a subroutine, which is suitable for larger structures. Since LNSsearch is a heuristic algorithm, it cannot prove the nonexistence of a solution to an inverse folding problem.

Supported by NSF DMS-1016618. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

RNA CP design

RNA inverse folding using Constraint Programming(CP)

[Home](#)

[RNA CP design Server](#)

[Results](#)

[Sequences](#)

[Manual](#)

[Contacts](#)

[RNAiFold](#)

Server

Email address: (optional):

Either upload a file containing a single RNA structure, paste in an RNA structure, or fill in fields in the verbose form.

File upload Paste input Verbose input

Fill the following fields: [Use sample](#)

Fasta comment

Target structure (required)

Sequence

Structure of compatible base pairs

Structure of incompatible base pairs

Temperature (°C): Maximum solutions:

Dangling end treatment:

Limit GC pairs: Min: Max:

Limit AU pairs: Min: Max:

Limit GU pairs: Min: Max:

Maximum consecutive nucleotides:

A: C: G: U:

GC content: Min: Max:

Supported by NSF DMS-1016618. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

RNAiFold 2.0 web server

Email address: (optional):

Step1 Step2 Step3

Find sequences in the seed alignment whose predicted minimum free energy structure corresponds to the alignment consensus structure

Select a Rfam family : [Use sample](#)

Energy model

Dangling end treatment

[Continue to Step 2](#)

Step1 Step2 Step3

Select one sequence as model to determine conserved positions

Choose a target structure from the list:

Structure .(((((((.....))))).(((.....))))).
 Sequence: GAUGAGUCUGUGCUAAGCACACUGAUGAGUCUAUGAAAUGAGACGAAACUCAUA

Force inverse of non conserved positions

Specify a minimum conservation threshold [Get conserved positions](#)

HBVHGBGUDVHVHDVBBHDBBCUGAVGAGVDVVBVGBBAVHGBGCGAAACVDBVB
 .(((((((.....))))).(((.....))))).

[Continue to Step 3](#)

| Position | A | C | G | U |
|----------|----------|----------|----------|----------|
| 1 | 0.112676 | 0.478873 | 0.295775 | 0.112676 |
| 2 | 0.792208 | 0.103896 | 0.025974 | 0.077922 |
| 3 | 0.506494 | 0.077922 | 0.064935 | 0.350649 |
| 4 | 0.428571 | 0.194805 | 0.376623 | 0.000000 |
| 5 | 0.857143 | 0.077922 | 0.012987 | 0.051948 |
| 6 | 0.012987 | 0.000000 | 0.987013 | 0.000000 |
| 7 | 0.000000 | 0.000000 | 0.000000 | 1.000000 |
| 8 | 0.038961 | 0.948052 | 0.000000 | 0.012987 |
| 9 | 0.014925 | 0.000000 | 0.208955 | 0.776119 |
| 10 | 0.085366 | 0.036585 | 0.878049 | 0.000000 |
| 11 | 0.060976 | 0.097561 | 0.487805 | 0.353659 |
| 12 | 0.170732 | 0.000000 | 0.792683 | 0.036585 |
| 13 | 0.048780 | 0.817073 | 0.121951 | 0.012195 |
| 14 | 0.060976 | 0.073171 | 0.097561 | 0.768293 |
| 15 | 0.512195 | 0.012195 | 0.000000 | 0.475610 |
| 16 | 0.768293 | 0.109756 | 0.073171 | 0.048780 |

Step1 Step2 Step3

Find compatible sequences using inverse folding

Either upload a file containing a single RNA structure, paste in an RNA structure, or fill in fields in the verbose form.

File upload Paste input Verbose input

Fill the following fields:

Fasta comment

Target structure (required)

Sequences

5 solutions found.
 Download [RNAiFold 2.0 results](#) or [RNAiFold 2.0 input](#) files.

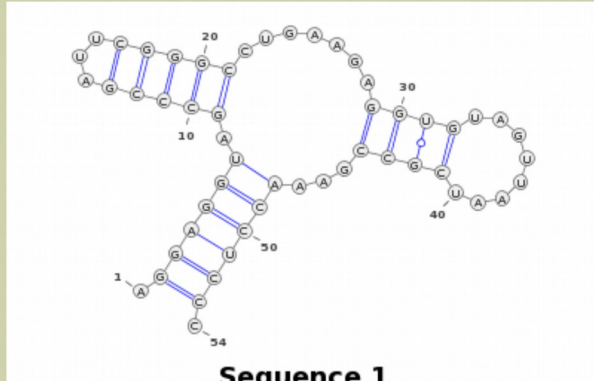
Target structure:
 > AJ005312.1_282-335
 .(((((((.....))))).(((.....))))).

Solution 1 :

AGGAGGUAGCCCGAUUCGGGCCUGAAGAGGUGUAGUUAUCGCCGAAACUCC

GC content: 0.57 - AUs: 2 - GCs: 12 - GUs: 1
 Probability of MFE structure:0.346724
 Expected pointwise entropy:0.0797598
 Morgan-Higgs structural Diversity:2.21595
 Vienna structural Diversity:1.17149
 Expected base pair distance:0.751023
 Ensemble defect:1.43592

[BLAST this sequence](#)



Sequence 1

RNAiFold 2.0

Some ongoing work

- Optimize inverse folding
 - New search heuristics and structure decomposition strategies
 - Constraint evaluation order
- Add new functionalities
 - Design constraints
 - Incorporate target functions and energy models
 - Evaluate new measures of structural diversity
- Expand objective functions (e.g. ensemble defect)

ONGOING WORKS



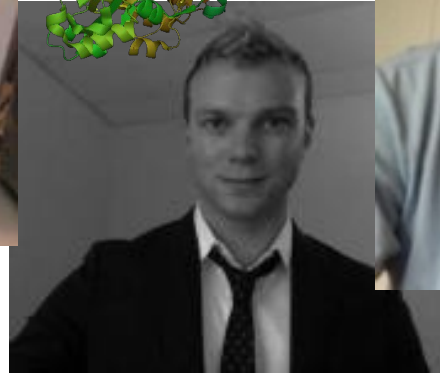
Conclusions

- Advantages of using a CP approach
 - Scalability
 - Flexibility for constrained design
 - Completeness (can also determine that NO solution exists)
- Applications of RNAiFold
 - Properties of the sequence ensemble
 - Discovery of functional RNAs
 - Synthetic design

Peter Clote

Thanks!!

Ivan Dotú



Amir Bayegan

Evan Senter



NSF DBI-1262439

The Meyer Lab



Betty Slinger Michelle Meyer