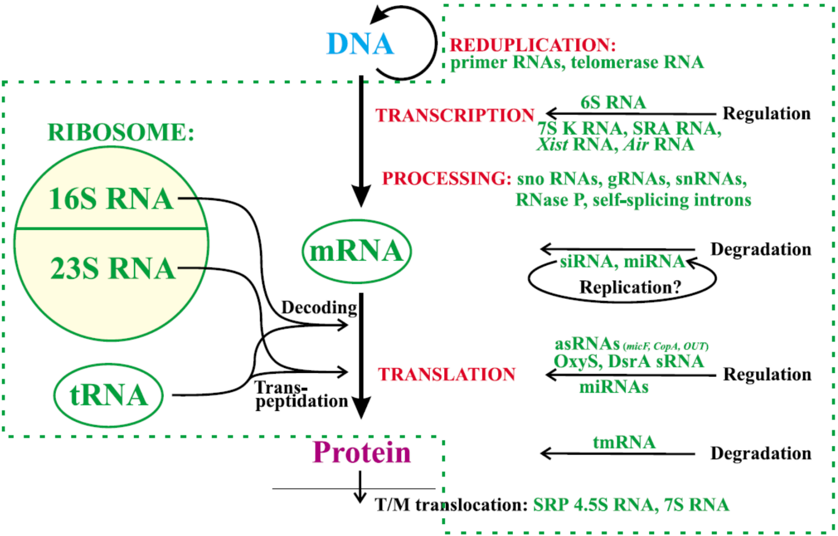
A Brief Overview of RNA Bioinformatics

Sebastian Will
University of Vienna

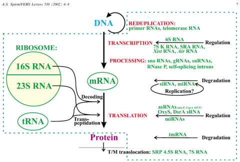
Freiburg sRNA Meeting 2019

Same Goal: Functions and Role of RNAs

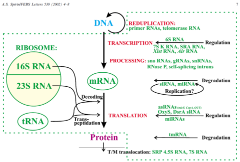
A.S. Spirin/FEBS Letters 530 (2002) 4–8 7



Different Methods



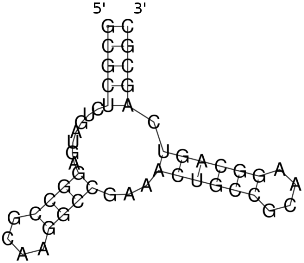
Different Methods



The Central Dogma (of RNA Bioinformatics)

Sequence $\Longrightarrow$ Structure $\Longrightarrow$ Function

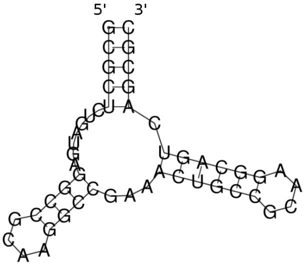
5' - GCGCUCUGAUGAGGCCGCAAGGCCGAAACUGCCGCAAGGCAGUCAGCGC - 3'



The Central Dogma (of RNA Bioinformatics)

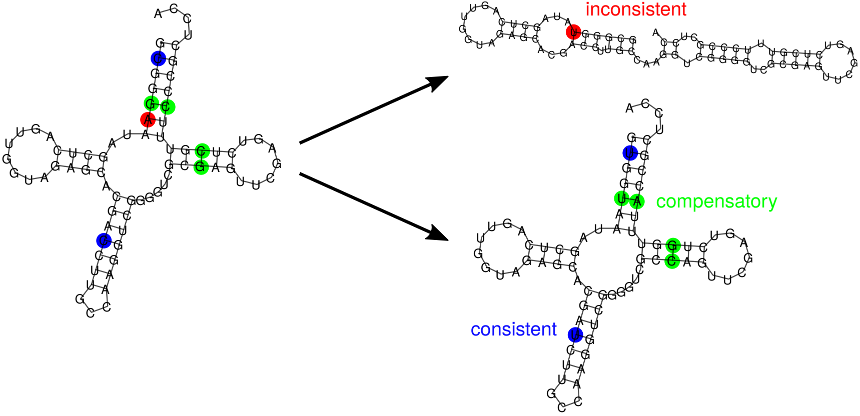
Sequence $\Longrightarrow$ Structure $\Longrightarrow$ Function

5' - GCGCUCUGAUGAGGCCGCAAGGCCGAAACUGCCGCAAGGCAGUCAGCGC - 3'



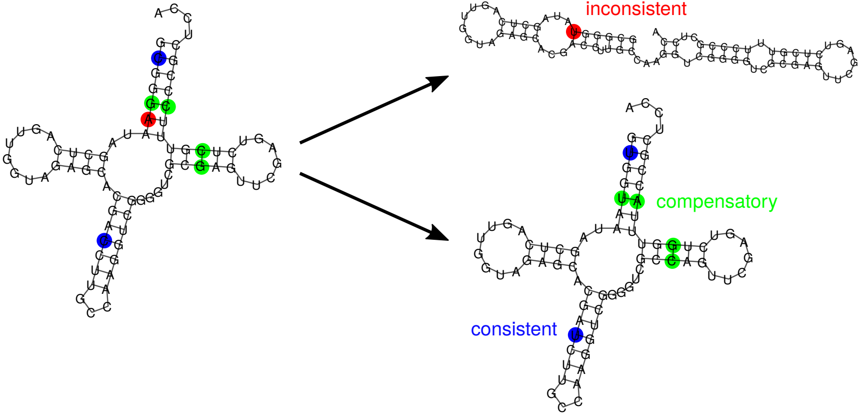
Structure as *Proxy* of Function

RNA Structure is Emergent



Almost identical sequences — very different structures
Very different sequences — same structure

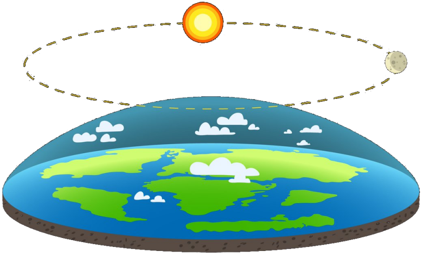
RNA Structure is Emergent



No Shortcut “Sequence $\Longrightarrow$ Function”

More Dogmas

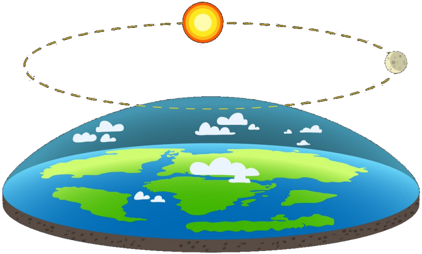
The world is simple!*



More Dogmas

The world is simple!*

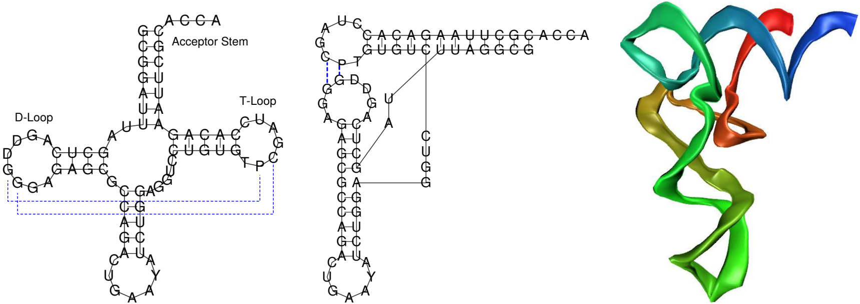
**in first approximation*



More Dogmas

The world is simple!*

**in first approximation*



Viable Shortcut

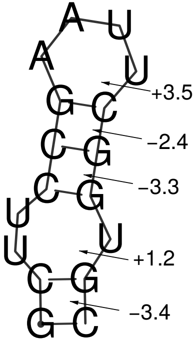
Sequence $\Longrightarrow$ 2D Structure $\Longrightarrow$ Function

Energies of RNA structures can be calculated

Nearest Neighbor Model (NNM)

- Free energies = sum of loop energies
- Loop energies measured experimentally (based on UV melting curves)
- Loop energies depend on
  - loop type
  - size
  - base composition

  ⇒ large energy parameter tables



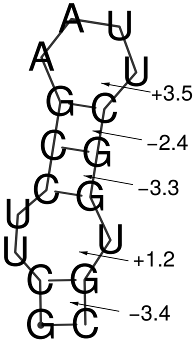
Total energy
−4.4 kcal/mol

Energies of RNA structures can be calculated

Nearest Neighbor Model (NNM)

- Free energies = sum of loop energies
- Loop energies measured experimentally (based on UV melting curves)
- Loop energies depend on
  - loop type
  - size
  - base composition

  ⇒ large energy parameter tables



Total energy
−4.4 kcal/mol

+ distinguish RNA structures by free energy
+ define *minimum free energy (MFE)*
+ basis for entire tool set for RNA structure

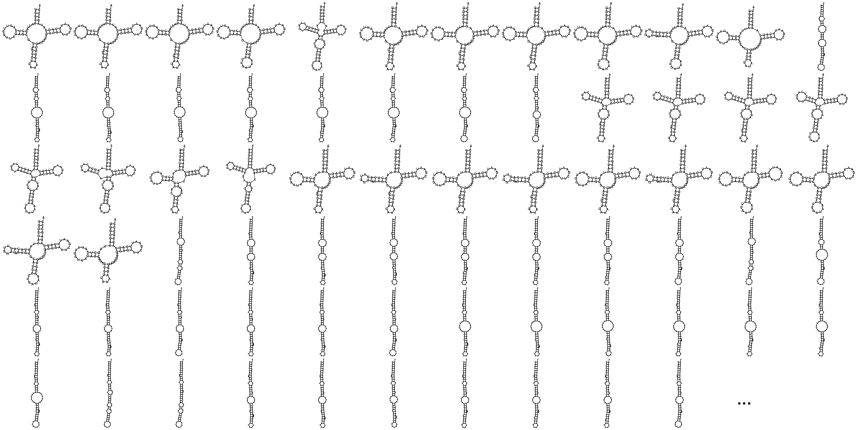
- limitations (stay tuned)

From NNM to Structure Prediction

GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCCUGAUAAGGGUGAGGUCGCUGAUUCGAAUUCAGCAUAGCCCA

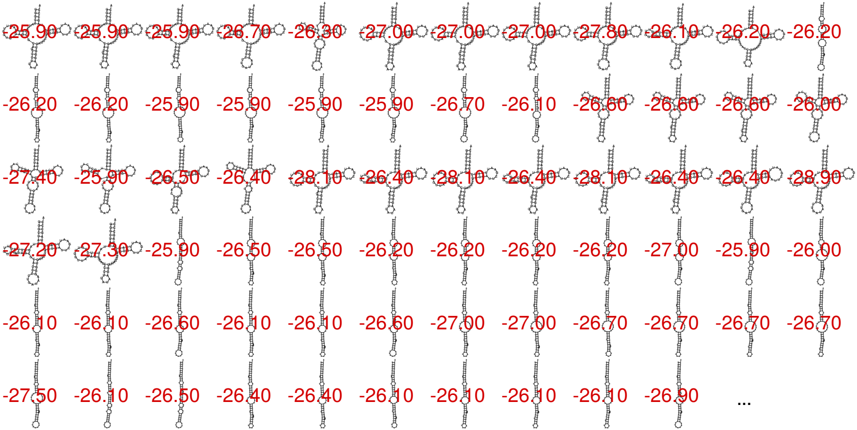
From NNM to Structure Prediction

GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCCUGAUAAGGGUGAGGUCGCUGAUUCGAAUUCAGCAUAGCCCA



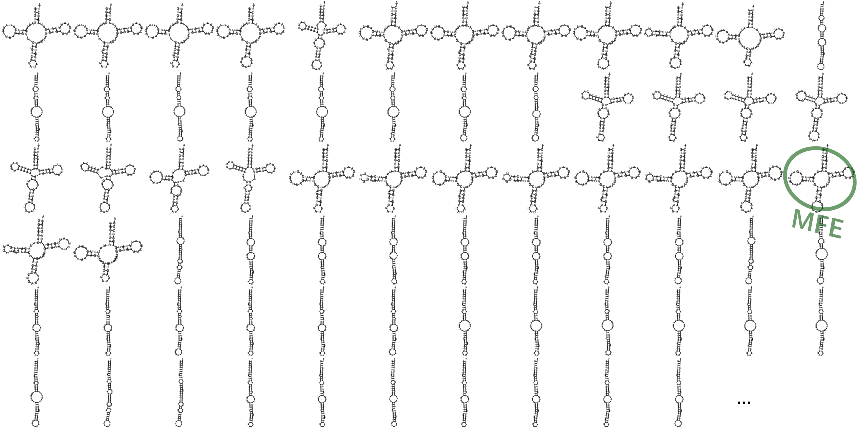
From NNM to Structure Prediction

GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCCUGAUAAGGGUGAGGUCGCUGAUUCGAAUUCAGCAUAGCCCA



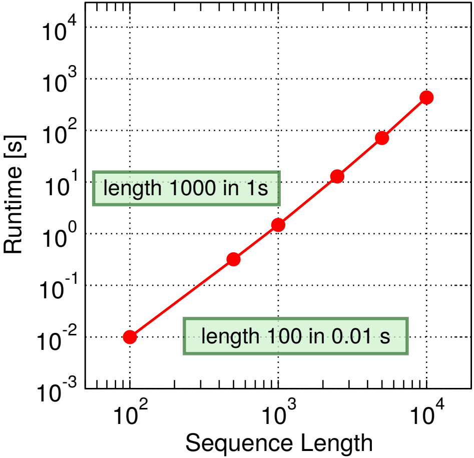
From NNM to Structure Prediction

GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCCUGAUAAGGGUGAGGUCGCUGAUUCGAAUUCAGCAUAGCCCA



Structure Prediction: Fast and Accurate

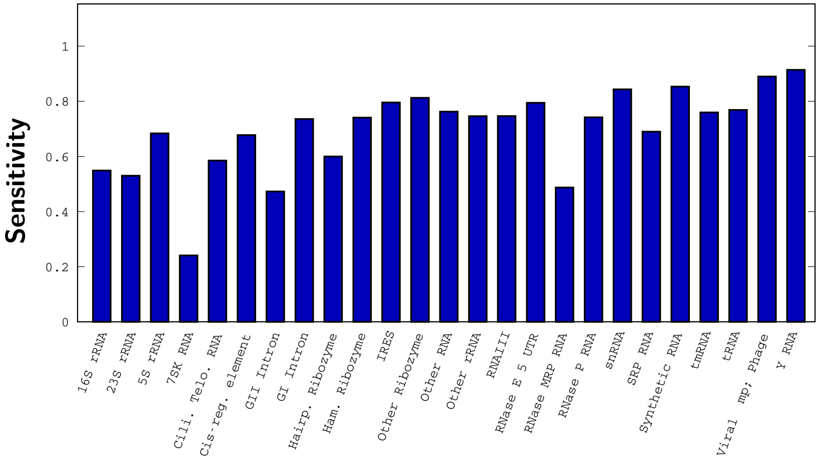
Performance of RNAFold (Vienna RNA Package 2.0)



[adapted from Vienna RNA Package 2.0, ALMOB 2011]

- **Very fast** folding algorithms: **0.01 seconds at length 100**

Structure Prediction: Fast and Accurate

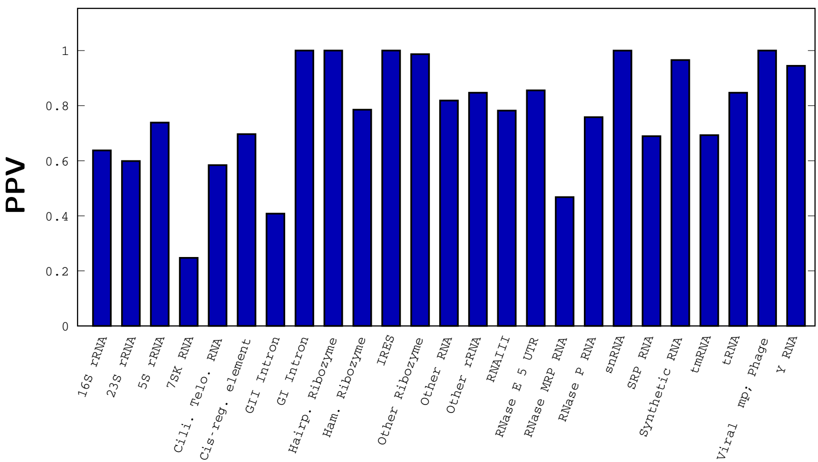


Performance of RNAFold (Vienna RNA Package 2.0)

[adapted from Vienna RNA Package 2.0, ALMOB 2011]

- **Very fast** folding algorithms: **0.01 seconds at length 100**
- **Very useful** accuracy: $\sim$ **70% predicted base pairs correct**

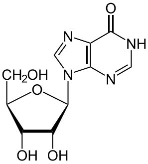
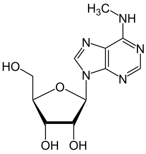
Structure Prediction: Fast and Accurate


Performance of RNAFold (Vienna RNA Package 2.0)

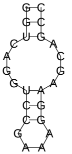
[adapted from Vienna RNA Package 2.0, ALMOB 2011]

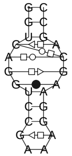
- **Very fast** folding algorithms: **0.01 seconds at length 100**
- **Very useful** accuracy: **∼ 70% predicted base pairs correct**

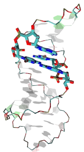
Limitations

- Modified bases  ,  , . . .

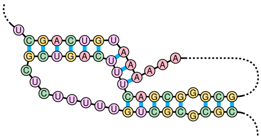
- Non-canonical base pairs

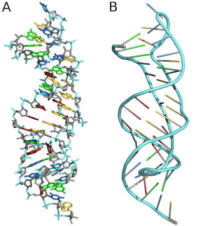




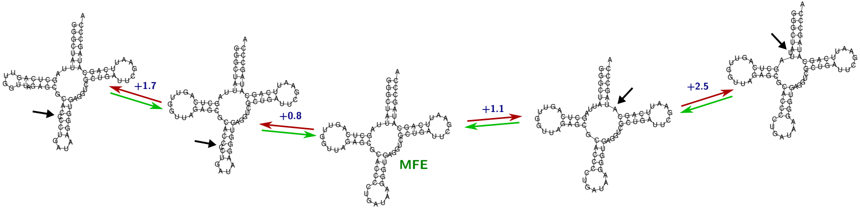


Pseudoknots





RNAs Refold at 'Room Temperature'



The MFE misleads! Look at

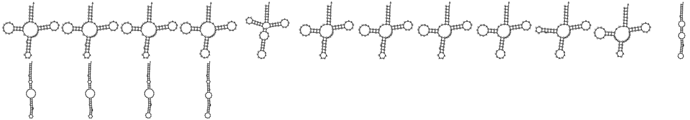
- suboptimal structures
- structure ensembles
- kinetics (co-transcriptional!)

Suboptimals and Probabilities

GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCCUGAUAAGGGUGAGGUCGCUGAUUCGAAUUCAGCAUAGCCCA

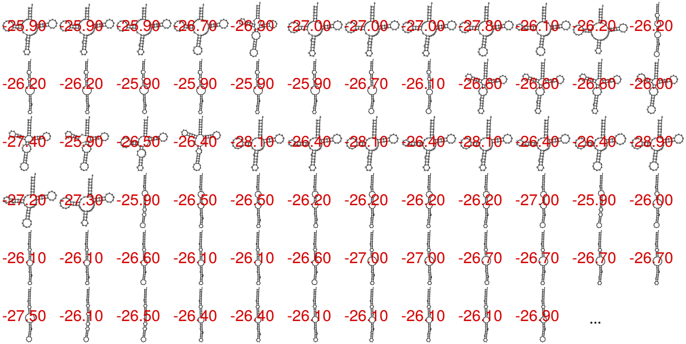
Suboptimals and Probabilities

GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCCUGAUAAGGGUGAGGUCGCUGAUUCGAAUUCAGCAUAGCCCA



Suboptimals and Probabilities

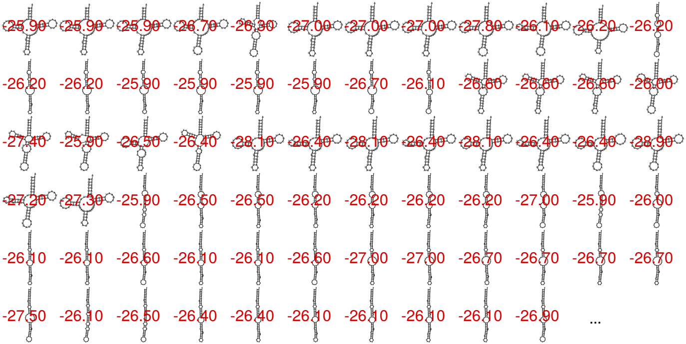
GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCCUGAUAAGGGUGAGGUCGCUGAUUCGAAUUCAGCAUAGCCCA



Energies → Structure Probabilities

Suboptimals and Probabilities

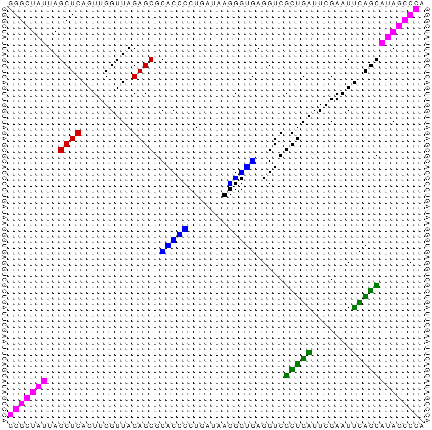
GGGCUAUUAGCUCAGUUGGUUAGAGCGCACCCCUGAUAAGGGUGAGGUCGCUGAUUCGAAUUCAGCAUAGCCCA

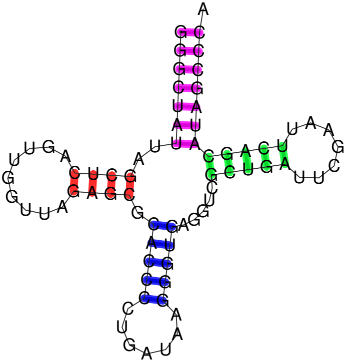


Energies → Structure Probabilities

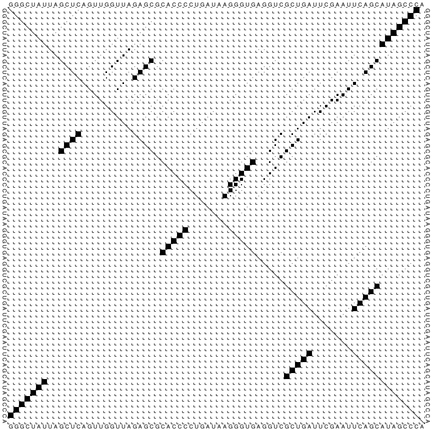
Structure Probabilities → Base Pair Probabilities

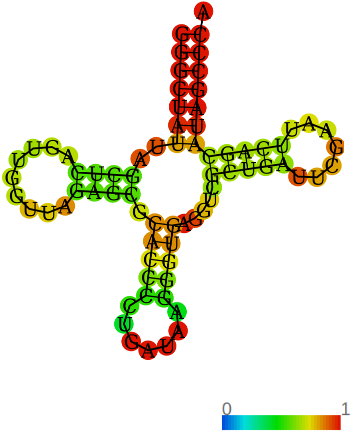
Dotplots and Reliabilities





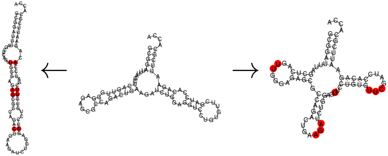
Dotplots and Reliabilities





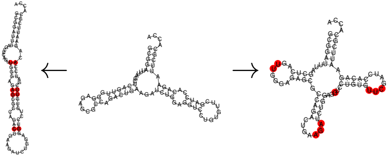
Integrating Prior Knowledge

- Knowledge on base pairing:

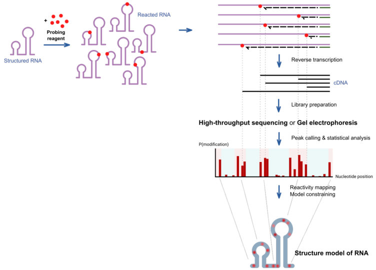


Integrating Prior Knowledge

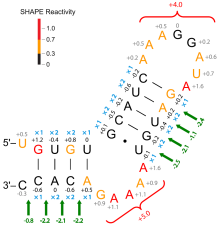
- Knowledge on base pairing:



- Structure probing experiments (e.g. SHAPE)



Adapted from Ptrw08, A schematic figure explaining the steps in a typical chemical probing experiment to assay the structure of RNA molecules, CC BY-SA 4.0



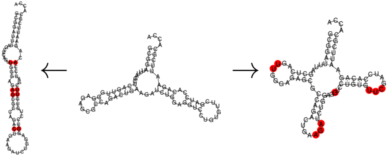
$$\Delta G_{\text{NN}} = \sum \Delta G_{\text{stacks}} + \sum \Delta G_{\text{loops}}$$

$$\Delta G_{\text{SHAPE}} = \mathbf{1\times} \sum \Delta G_{\text{ends}} + \mathbf{2\times} \sum \Delta G_{\text{interior}}$$

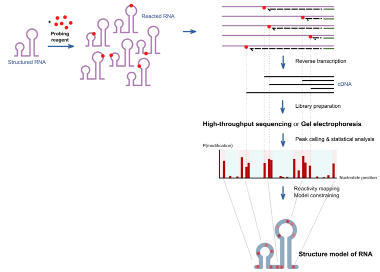
$$\Delta G_{\text{total}} = \Delta G_{\text{NN}} + \Delta G_{\text{SHAPE}}$$

Integrating Prior Knowledge

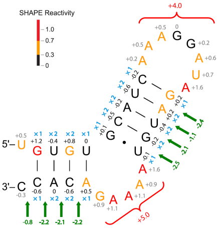
- Knowledge on base pairing:



- Structure probing experiments (e.g. SHAPE)



Adapted from *Ptrw08, A schematic figure explaining the steps in a typical chemical probing experiment to assay the structure of RNA molecules, CC BY-SA 4.0*



$$\Delta G_{\text{NN}} = \sum \Delta G_{\text{stacks}} + \sum \Delta G_{\text{loops}}$$

$$\Delta G_{\text{SHAPE}} = 1 \times \sum \Delta G_{\text{ends}} + 2 \times \sum \Delta G_{\text{interior}}$$

$$\Delta G_{\text{total}} = \Delta G_{\text{NN}} + \Delta G_{\text{SHAPE}}$$

- Homology

```
Examp1    ----------CCGG-AAA-CCGAACGCAGCACCGCGG------AU-CUGGAACGC--
Examp2    ----------CGCU-AG--AACAAC-------UAUCU------GU-AGCGCGAAAAC
Examp3    ---------AUUGUGUA--GCAUU------AGUUUGC-------GUGCAAAGAACGC
Examp4    -------UGCCAUCGCAUUAGCACC---U-AGCCGCAUUUUCUGGCGAUGAUG----
Examp5    AGCACCGAACCGCAU----GCGAACUGAG-AA--CGCAACC----AUGCGCGCACC-
```

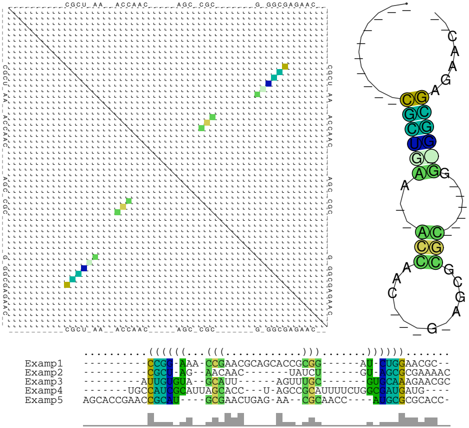
Comparative Analysis with Alifold

```
Examp1   ----------CCGG-AAA-CCGAACGCAGCACCGCGG------AU-CUGGAACGC--
Examp2   ----------CGCU-AG--AACAAC-------UAUCU------GU-AGCGCGAAAAC
Examp3   ---------AUUGUGUA--GCAUU------AGUUUGC-------GUGCAAAGAACGC
Examp4   -------UGCCAUCGCAUUAGCACC---U-AGCCGCAUUUUCUGGCGAUGAUG----
Examp5   AGCACCGAACCGCAU----GCGAACUGAG-AA--CGCAACC----AUGCGCGCACC-
```

Comparative Analysis with Alifold

```
Examp1    ----------CCGG-AAA-CCGAACGCAGCACCGCGG------AU-CUGGAACGC--
Examp2    ----------CGCU-AG--AACAAC-------UAUCU------GU-AGCGCGAAAAC
Examp3    ---------AUUGUGUA--GCAUU------AGUUUGC-------GUGCAAAGAACGC
Examp4    -------UGCCAUCGCAUUAGCACC---U-AGCCGCAUUUUCUGGCGAUGAUG----
Examp5   AGCACCGAACCGCAU----GCGAACUGAG-AA--CGCAACC----AUGCGCGCACC-
```

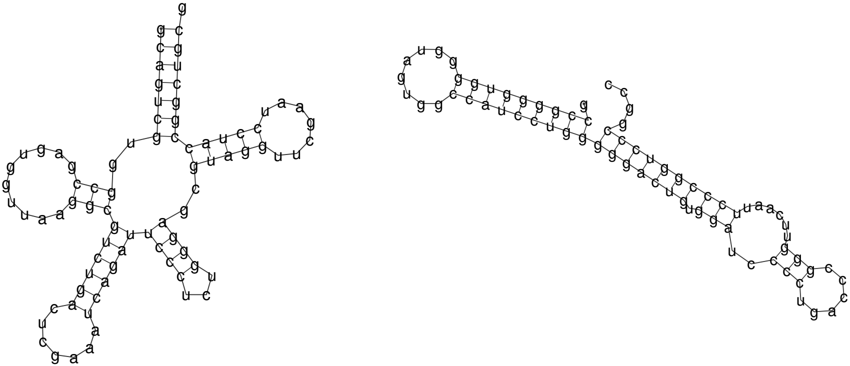
⇓ Alifold



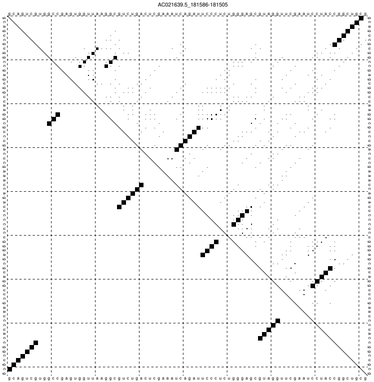
```
          ..........((((((...(((..............)))........))))))........
Examp1    ----------CCGG-AAA-CCGAACGCAGCACCGCGG------AU-CUGGAACGC--
Examp2    ----------CGCU-AG--AACAAC-------UAUCU------GU-AGCGCGAAAAC
Examp3    ---------AUUGUGUA--GCAUU------AGUUUGC-------GUGCAAAGAACGC
Examp4    -------UGCCAUCGCAUUAGCACC---U-AGCCGCAUUUUCUGGCGAUGAUG----
Examp5    AGCACCGAACCGCAU----GCGAACUGAG-AA--CGCAACC----AUGCGCGCACC-
```

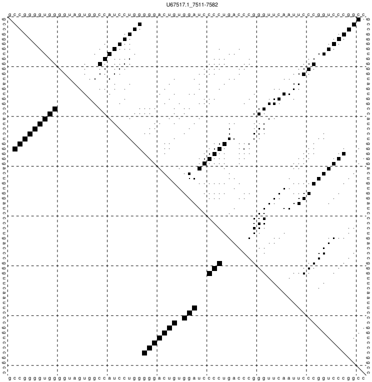
[Fig. adapted from Vienna RNA Package 2.0, ALMOB 2011, alifold exanple]

Simultaneous Alignment and Folding (with LocARNA)

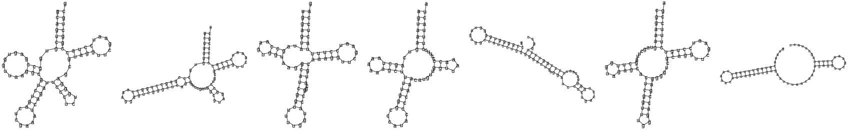


Simultaneous Alignment and Folding (with LocARNA)

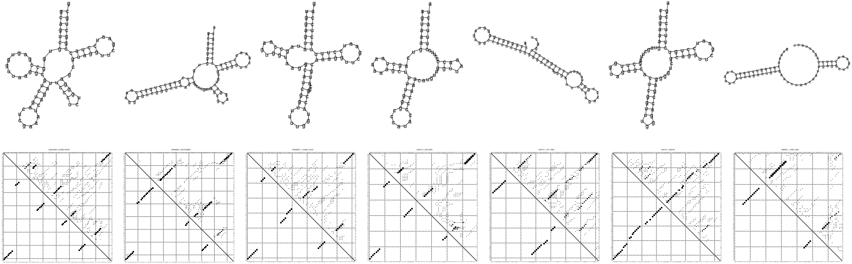




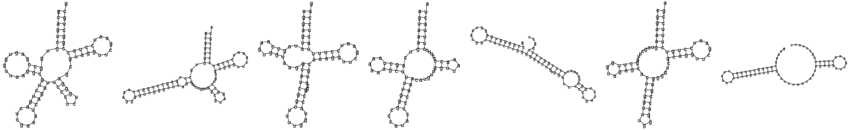
Simultaneous Alignment and Folding (with LocARNA)

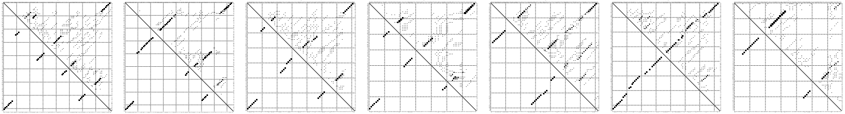


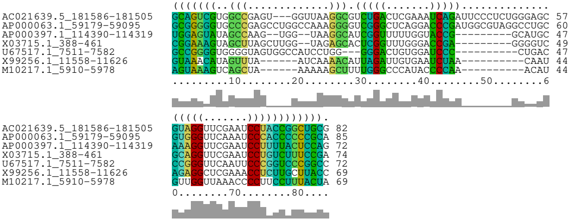
Simultaneous Alignment and Folding (with LocARNA)



Simultaneous Alignment and Folding (with LocARNA)

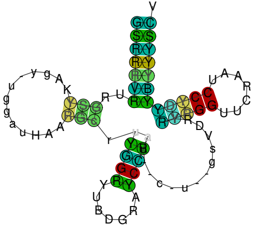




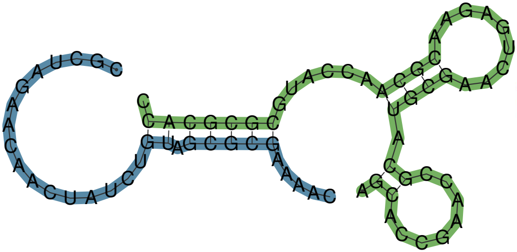


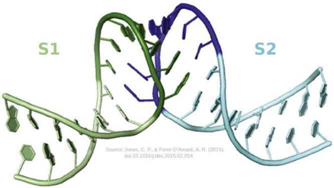
```
                              (((((((..(((.............))).(((((.......))))).............. 
AC021639.5_181586-181505      GCAGUCGUGGCCGAGU---GGUUAAGGCGUCUGACUCGAAAUCAGAUUCCCUCUGGGAGC 57
AP000063.1_59179-59095        GCGGGGGUGCCCGAGCCUGGCCAAAGGGGUCGGGCUCAGGACCCCGAUGGCGUAGGCCUGC 60
AP000397.1_114390-114319      UGGAGUAUAGCCAAG--UGG--UAAGGCAUCGGUUUUUGGUACCG----------GCAUGC 47
X03715.1_388-461              CGGAAAGUAGCUUAGCUUGG--UAGAGCACUCGGUUUGGGACCGA----------GGGGUC 49
U67517.1_7511-7582            GCCGGGGUGGGGUAGUGGCCAUCCUGG---GGGACUGUGGAUCCC----------CUGAC 47
X99256.1_11558-11626          GUAAACAUAGUUUA------AUCAAAACAUUAGAUUGUGAAUCUAA----------CAAU 44
M10217.1_5910-5978            AGUAAAGUCAGCUA------AAAAAGCUUUUGGGCCCAUACCCCAA----------ACAU 44
                              .........10........20........30........40........50........6

                              (((((.......))))))))))))).
AC021639.5_181586-181505      GUAGGUUCGAAUCCUACCGGCUGCG 82
AP000063.1_59179-59095        GUGGGUUCAAAUCCCACCCCCCGCA 85
AP000397.1_114390-114319      AAAGGUUCGAAUCCUUUUACUCCAG 72
X03715.1_388-461              GCAGGUUCGAAUCCUGUCUUUCCGA 74
U67517.1_7511-7582            CCGGGUUCAAUUCCCGGUCCCGGCC 72
X99256.1_11558-11626          AGAGGCUCGAAACCUCUUGCUUACC 69
M10217.1_5910-5978            GUUGGUUAAACCCCUUCCUUUACUA 69
                              0........70........80....
```

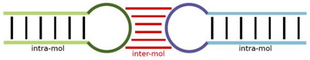


Interaction Prediction

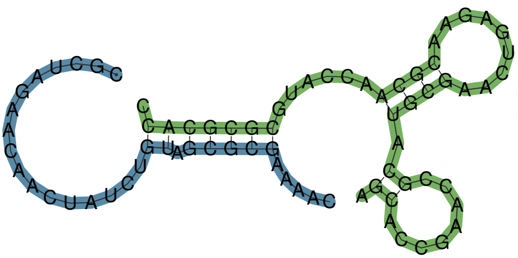


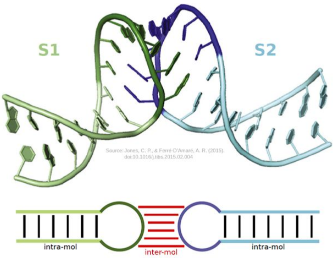


Source: Jones, C. P., & Ferré-D'Amaré, A. R. (2015). doi:10.1016/j.tibs.2015.02.004



Interaction Prediction



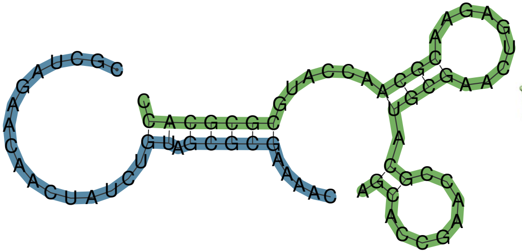


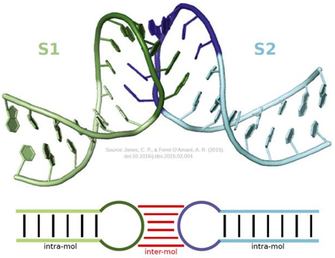
Source: Jones, C. P., & Ferré-D'Amaré, A. R. (2015). doi:10.1016/j.tibs.2015.02.004

- Similar to structure prediction: use NNM!

[Cofold example figure adapted from Vienna RNA Package 2.0, ALMOB 2011]

Interaction Prediction

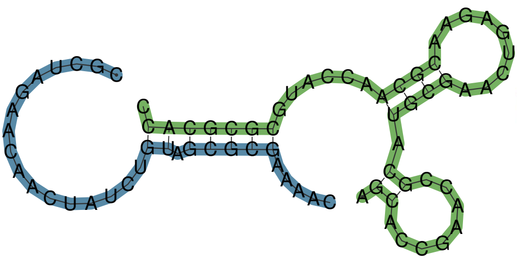


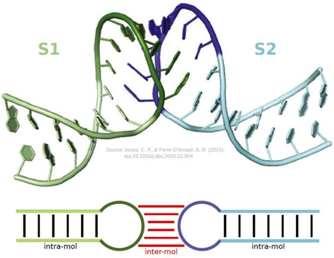


- Similar to structure prediction: use NNM!
- Predict intra- and inter-molecular structure
  - strong restrictions (cofold), **no** KHP → **fast**
  - more freedom (Alkan et al.), KHP → **slow**

[Cofold example figure adapted from Vienna RNA Package 2.0, ALMOB 2011]

Interaction Prediction





Source: Jones, C. P., & Ferré-D'Amaré, A. R. (2015). doi:10.1016/j.tibs.2015.02.004

- Similar to structure prediction: use NNM!
- Predict intra- and inter-molecular structure
  - strong restrictions (cofold), **no** KHP → **fast**
  - more freedom (Alkan et al.), KHP → **slow**
- IntaRNA: reasonable abstraction → **fast**
  - Use unpairing probabilities
  - E.g. genome-wide prediction of sRNA targets

[Cofold example figure adapted from Vienna RNA Package 2.0, ALMOB 2011]

RNA Bioinformatics—Take home

- **Secondary structure** as proxy for **RNA function**
- **Nearest neighbor model (NNM)** enables prediction of MFE structures and **probabilities**
- Solid fundament to construct methods for
  - Integrating prior knowledge
  - Simultaneous alignment and folding
  - Prediction of RNA interactions
  - ...pseudoknots, modifications, non-canonical base pairs, 3D structure, kinetics, design
- Building blocks of **pipelines to learn about RNA function**

  e.g. sRNA target prediction