Theoretical and practical metagenomic approaches to viral discovery

Practical Session: ViennaRNA on single molecules

Kevin Lamkiewicz, Manja Marz

24.10.2019

European Virus Bioinformatics Center

Introducing ViennaRNA

Nothing in common!

Tools and Scripts

Tools and Scripts

- ViennaRNA
  - RNAfold
  - RNAsubopt
  - RNAcofold
  - RNAduplex
  - RNAalifold
  - RNALfold
  - ...

Tools and Scripts

- ViennaRNA
  - RNAfold
  - RNAsubopt
  - RNAcofold
  - RNAduplex
  - RNAalifold
  - RNALfold
  - ...
- LocaRNA

Tools and Scripts

- ViennaRNA
  - RNAfold
  - RNAsubopt
  - RNAcofold
  - RNAduplex
  - RNAalifold
  - RNALfold
  - ...
- LocaRNA
- MAFFT, VARNA, ...

Hands on!

RNAFOLD

```
1 # Use RNAfold on sequence.fasta in order
2 # to fold all sequences in the file
3 $> RNAfold  < sequence.fasta
```

RNAFOLD

```
# Use RNAfold on sequence.fasta in order
# to fold all sequences in the file
$> RNAfold  < sequence.fasta
```

```
# Redirect the output of RNAfold into a file
$> RNAfold  < sequence.fasta > sequence.fold
```

RNAfold mit Partition Function

```
# Use the -p parameter to calculate the partition function
# on top of the minimum free energy secondary structures
$> RNAfold -p < sequence.fasta > sequence.fold

```

RNAfold mit Partition Function

```
# Use the -p parameter to calculate the partition function
# on top of the minimum free energy secondary structures
$> RNAfold -p < sequence.fasta > sequence.fold

# This will produce a centroid structure;
# the structure with minimal average distance to all
# sampled structures.
# . unpaired
# , weakly paired
# | strongly paired w/o preference
# { } weakly paired
# ( ) strongly paired
```

MEA STRUCTURE

Maximum expected accuracy Structure

Reminder: With the *partition function* we are able to calculate basepair probabilities. The structure with the heighest sum of all probabilities, is called the *MEA structure*.

MEA STRUCTURE

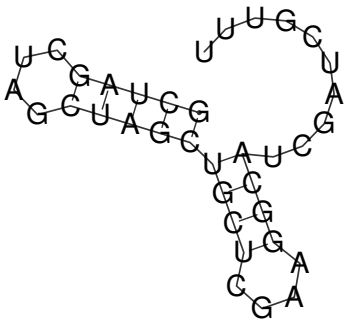
Maximum expected accuracy Structure

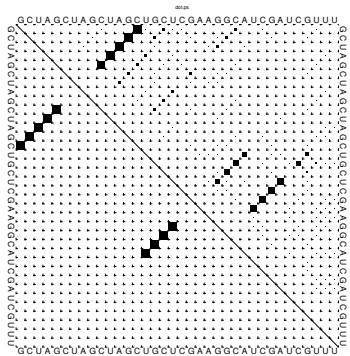
Reminder: With the *partition function* we are able to calculate basepair probabilities. The structure with the heighest sum of all probabilities, is called the *MEA structure*.

```
# Use -p and --MEA to calculate the MFE,
# the MEA and the centroid structure.
# Note: The MFE and the MEA structure do not have to be the same!
# Note2: --MEA usually implies -p
$> RNAfold -p --MEA < sequence.fasta > sequence.fold
```

Structure and Dotplots

```
1 # the same command again; you will get an rna.ps and a dot.ps
2 $> RNAfold -p --MEA < sequence.fasta > sequence.fold
```





RNAFOLD+

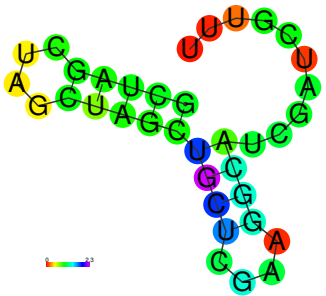
Let's use some colors!

The `ViennaRNA Package` offers some `Perl`-scripts, which can enrich the PostScript files of our structures.

RNAFOLD+

Let's use some colors!

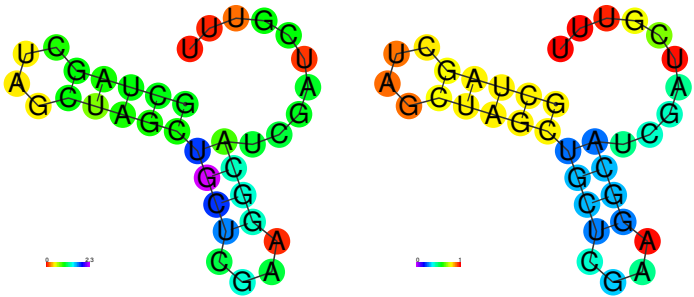
The `ViennaRNA Package` offers some `Perl`-scripts, which can enrich the PostScript files of our structures.



RNAFOLD+

Let's use some colors!

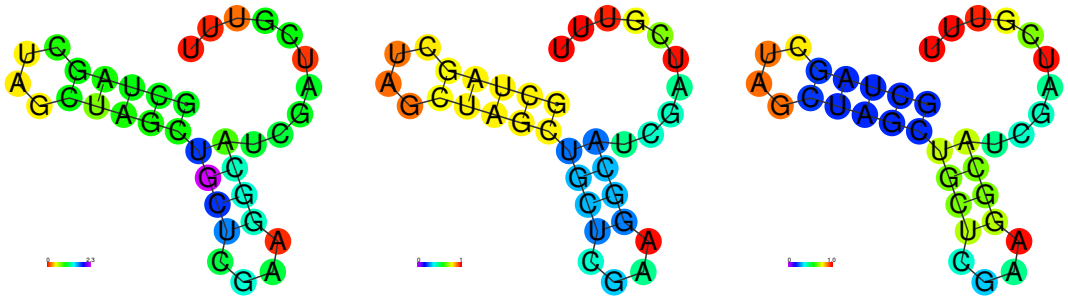
The `ViennaRNA Package` offers some `Perl`-scripts, which can enrich the PostScript files of our structures.



RNAFOLD+

Let's use some colors!

The `ViennaRNA Package` offers some `Perl`-scripts, which can enrich the PostScript files of our structures.



RNAfold Perl Scripts

```
# Low entropy regions have little structural flexibility,
# which means the reliability of the predicted structure is high.
# High entropy indicate many structual alternatives
# which might be functional important but make the prediction
# more difficult - and thus less reliable.
$> ./relplot.pl rna.ps dot.ps > entropy.ps
```

RNAfold Perl Scripts

```
# Low entropy regions have little structural flexibility,
# which means the reliability of the predicted structure is high.
# High entropy indicate many structual alternatives
# which might be functional important but make the prediction
# more difficult - and thus less reliable.
$> ./relplot.pl rna.ps dot.ps > entropy.ps

# -p colors the nucleotides based on their base-pairing
# probability
$> ./relplot.pl -p rna.ps dot.ps > probability.ps

```

RNAfold Perl Scripts

```
# Low entropy regions have little structural flexibility,
# which means the reliability of the predicted structure is high.
# High entropy indicate many structual alternatives
# which might be functional important but make the prediction
# more difficult - and thus less reliable.
$> ./relplot.pl rna.ps dot.ps > entropy.ps

# -p colors the nucleotides based on their base-pairing
# probability
$> ./relplot.pl -p rna.ps dot.ps > probability.ps

# -a colors the nucleotides based on their accessbility
# (e.g. the probability of being unpaired)
$> ./relplot.pl -a rna.ps dot.ps > access.ps
```

Suboptimal Structures and Constraints

RNASUBOPT

Sometimes we're interested in suboptimal structures.

```
# In general RNAsubopt is used exactly like RNAfold.
# With -p one calculates the partition function
$> RNAsubopt [OPTIONS] < sequence.fasta > sequence.subopt

```

RNASUBOPT

Sometimes we're interested in suboptimal structures.

```
# In general RNAsubopt is used exactly like RNAfold.
# With -p one calculates the partition function
$> RNAsubopt [OPTIONS] < sequence.fasta > sequence.subopt

# With the -e parameter, one can define a certain
# energy range. Using this, RNAsubopt returns
# all structures that are in range of this parameter.
$> RNAsubopt -e 2 < sequence.fasta > sequence_e2.subopt
```

RNAfold with constraints

- Structure of RNA is partly known (e.g. via SHAPE experiments)
- `RNAfold` is able to consider this knowledge

```
# Enable -C to include constraints. --noPS prevents the generation
# of the rna.ps and dot.ps files.
$> RNAfold --noPS -C < constrained.fasta > constrained.fold

```

RNAfold with constraints

- Structure of RNA is partly known (e.g. via SHAPE experiments)
- `RNAfold` is able to consider this knowledge

```
# Enable -C to include constraints. --noPS prevents the generation
# of the rna.ps and dot.ps files.
$> RNAfold --noPS -C < constrained.fasta > constrained.fold

# . (no constraint for this base)
# | (corresponding base has to be paired)
# x (base is unpaired)
# < (base i is paired with base j>i)
# > (base i is paired with base j<i)
# and matching brackets ()  (base i pairs with base j)
```

Calculating structures alignment-based

RNAALIFOLD

`RNAalifold` calculates a *consensus* RNA secondary structure for several aligned RNA sequences.

```
# RNAalifold accepts CLUSTAL, Stockholm, FASTA or MAF
# formats for the input alignment.
# --color will produce a colored version of the structure plot
# --aln produces a colored alignment based on the structure

$> RNAalifold --aln --color < input.aln > consensus.alifold
```

LocARNA vs MAFFT

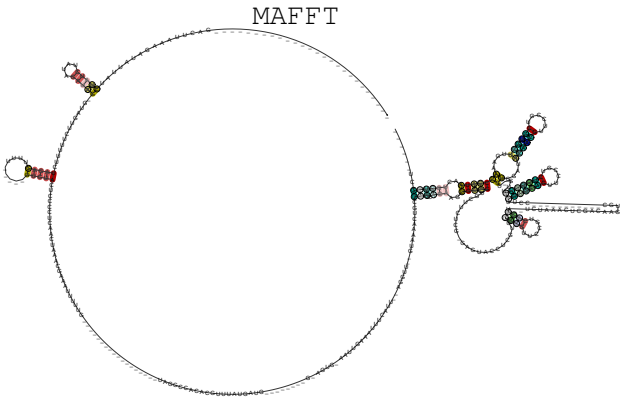
In order to create a multiple sequence alignment, we can use `MAFFT` and/or `LocARNA` (and many more...)

```
# mafft creates a multiple sequence alignment based on
# sequence conservation only
$> mafft --clustalout cov_5utr.fa > cov_5utr_mafft.aln

# locarna folds and aligns the sequences simultanously,
# yielding better results for sequence that share a
# structural conservation
# However, locarna needs quite some time compared to sequence-based
# alignment tools.
$> mlocarna --thread 4 cov_5utr.fasta > cov_5utr.locarna
```

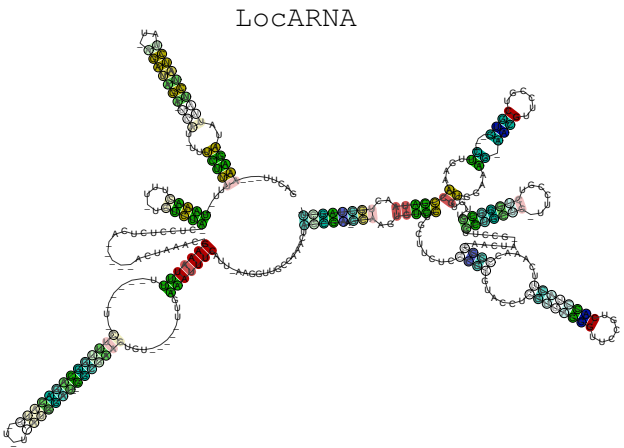
RNAALIFOLD RESULTS

```
# alirna.ps and aln.ps give information of the structural conservation
$> RNAalifold --aln --color < cov_5utr_mafft.aln \
      > cov_5utr_mafft.alifold

# NOTE: both PostScript files will be overwritten!
# locarna saves the alignment in a subdirectory
$> RNAalifold --aln --color < cov_5utr.out/results/result.aln \
      > cov_5utr_locarna.alifold
```

STRUCTURES

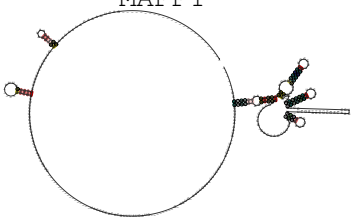


STRUCTURES

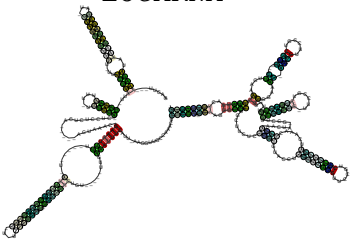


STRUCTURES

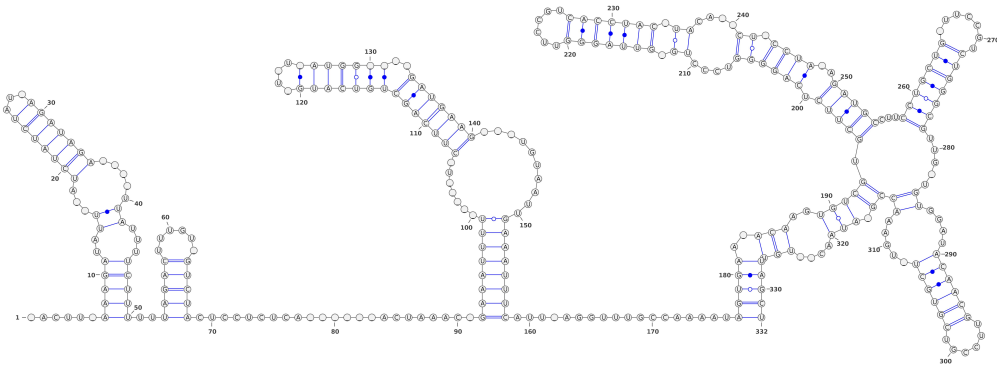
MAFFT



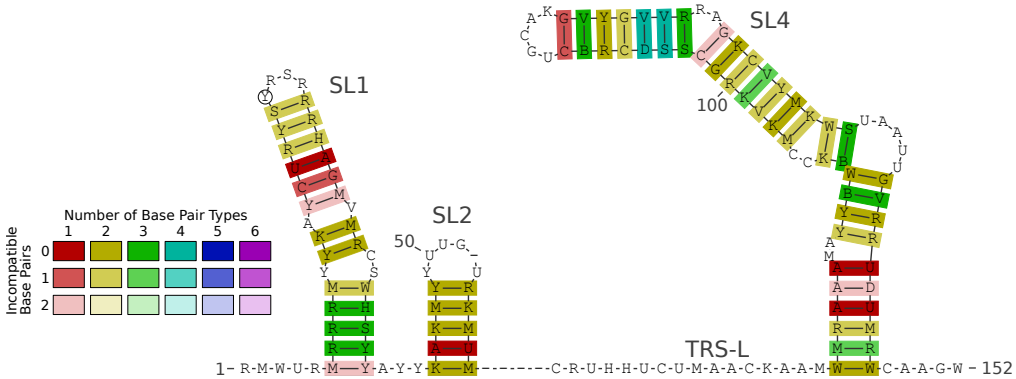
LocARNA



VARNA: FOR NICE FIGURES!



RNAALIFOLD COLOR CODE



COFFEE BREAK