



News from Vienna

Stephan Bernhart

Institute for Theoretical Chemistry
University of Vienna

Bled, 2006



Outline

RNAfold - Unified

Precursors

Shortcomings

Unified Algorithm

Results

Local Folding Programs

Why

RNAplfold

Alifold



RNAcofold

Fundamentals

- Standard loop based energy model
- Algorithm based on RNAfold
- Concatenate sequences, fold them, treat gap between sequences special
- $\mathcal{O}((n_1 + n_2)^3)$ processor, $\mathcal{O}((n_1 + n_2)^2)$ memory



RNAup

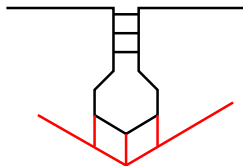
Fundamentals

- Compute full fold of long/both molecule(s)
- Compute probability of being unpaired
- Compute inter-molecular energy contributions
- Combine with unpaired probability
- $\mathcal{O}(w^2(n_1)^3)$ processor, $\mathcal{O}((n_1)^2)$ memory



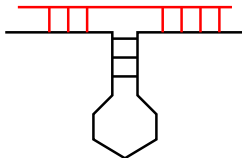
What's missing

- RNAfold can not compute pseudoknots



What's missing

- RNAcofold can not compute pseudoknots
- In RNAup binding site can not be split or structured





What's missing

- RNAcofold can not compute pseudoknots
- In RNAup binding site can not be split or structured
- No pair probability in RNAup (yet)

Combine RNAup and RNAcofold

- Algorithm able to compute pseudo knots and split binding sites
- Include base pair probability
- Time complexity should be $\mathcal{O}((n_1 + n_2)^3)$

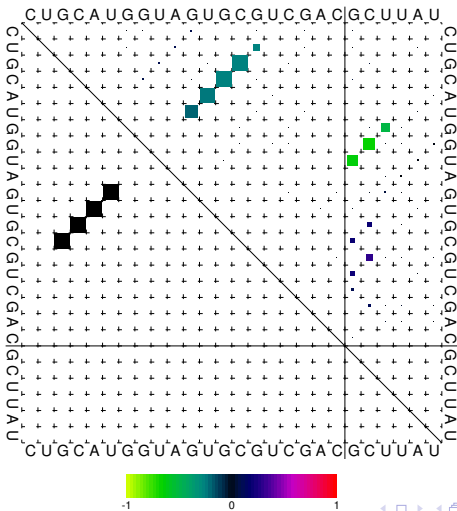


Combine RNAup and RNAcofold

- Compute RNAcofold
- Compute intermolecular energy contributions
- Compute probability of being unpaired, given pseudoknot
- Combine these informations to compute pair probabilities

Preliminary Results

Difference Dot Plot created by MakeDiffplot.pl



Conclusions

- (much) slower than RNAfold or RNAup
- still $\mathcal{O}(n^3)$
- Applications include RNA openers, DNA computing
- In principle possible to let both molecules form intramolecular base pairs (at great performance cost)



Local folding Algorithms

- Huge amount of sequence data
- RNAfold too slow for large sequences, e.g. Genomes
- Transcripts often not too long
- Boundaries of transcripts not always known
- Sliding window approach



Sliding window Approach



AUGCUGCAU GAUGAGUACAGUACGAU

With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap



Sliding window Approach



UJGCUGCAUGCUGAGUACAGUACGAU

With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap



Sliding window Approach



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap



Sliding window Approach

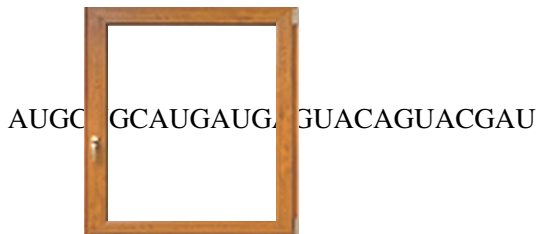


With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach

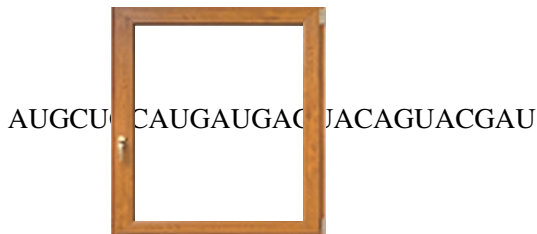


With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach

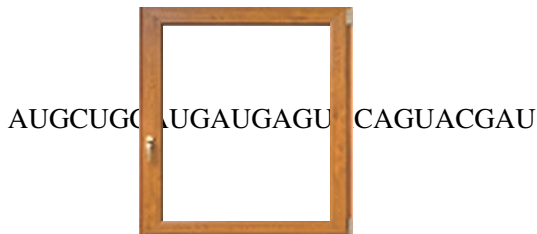


With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach

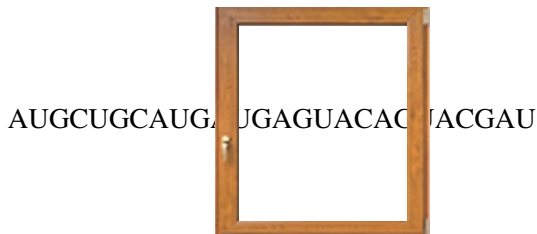


With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach

AUGCUGCAUGAUGAGUACAGUACGAU



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach

AUGCUGCAUGAUC GUACAGUA GAU



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap



Sliding window Approach

AUGCUGCAUGAUGA UACAGUACCAU



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach

AUGCUGCAUGAUGAGUACAGUACGA



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap

Sliding window Approach

AUGCUGCAUGAUGAGUA AGUACGAU



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap



Sliding window Approach

AUGCUGCAUGAUGAGUAC GUACGAU



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap



Sliding window Approach

AUGCUGCAUGAUGAGUACAUUACGAU



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap



Sliding window Approach

AUGCUGCAUGAUGAGUACAGUACGAU



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap



Sliding window Approach

AUGCUGCAUGAUGAGUACAGUACGAU



With window size l , sequence size n and stepsize s we get:

$$\mathcal{O}\left(l^3 \frac{n}{s}\right)$$

where $s = l - o$, o =overlap



RNAplfold

partition function

for $s = 1$, sliding window approach takes $\mathcal{O}(l^3 n)$

- Compute partition function and mean pair probability for $s = 1$
- $Q(i, j + 1) = Q(i, j) + \sum_{i \leq k < j} Q^B(k, j + 1)Q(i, k - 1)$



RNAplfold

partition function

for $s = 1$, sliding window approach takes $\mathcal{O}(l^3 n)$

- Compute partition function and mean pair probability for $s = 1$

- $Q(i + 1, j + 1) = Q(i + 1, j) + \sum_{i+1 \leq k < j} Q^B(k, j + 1)Q(i + 1, k - 1)$



RNAfold

partition function

for $s = 1$, sliding window approach takes $\mathcal{O}(l^3 n)$

- Compute partition function and mean pair probability for $s = 1$
- $Q(i + 1, j + 1) = Q(i + 1, j) + \sum_{i+1 \leq k < j} Q^B(k, j + 1)Q(i + 1, k - 1)$
- RNAfold Algorithm computes $Q(i, j) \quad \forall \quad i < j \in n$
- For every step of size 1, only $\sum_{i+1 \leq k < j} Q^B(k, j + 1)Q(i + 1, k - 1)$ has to be computed
- Thus, partition function for every possible window of size l can be computed in $\mathcal{O}(l^2 n)$



RNAplfold

pair probability

with windows of length l starting at u , mean pair probability π_{ij}^l

$$\pi_{ij}^l = \underbrace{\frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \frac{z_{1,i-1}^{u,l} \zeta_{i,j}^{u,l} z_{j+1,n}^{u,l}}{z_{1,n}^{u,l}}}_{\pi_{ij}^{*l}} + \frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \sum_{k < i} \sum_{m > j} p_{km}^{u,l} \mathbb{I}_{ij,km}$$



RNAfold

pair probability

with windows of length l starting at u , mean pair probability π_{ij}^l

$$\pi_{ij}^l = \underbrace{\frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \frac{z_{1,i-1}^{u,l} \zeta_{i,j}^{u,l} z_{j+1,n}^{u,l}}{z_{1,n}^{u,l}}}_{\pi_{ij}^{*l}} + \frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \sum_{k < i} \sum_{m > j} p_{km}^{u,l} \equiv_{ij,km}$$

Number of windows containing pair i, j



RNAfold

pair probability

with windows of length l starting at u , mean pair probability π_{ij}^l

$$\pi_{ij}^l = \underbrace{\frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \frac{z_{1,i-1}^{u,l} \zeta_{i,j}^{u,l} z_{j+1,n}^{u,l}}{z_{1,n}^{u,l}}}_{\pi_{ij}^{*l}} + \frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \sum_{k < i} \sum_{m > j} p_{km}^{u,l} \equiv_{ij,km}$$

For every window containing pair i, j



RNAplfold

pair probability

with windows of length l starting at u , mean pair probability π_{ij}^l

$$\pi_{ij}^l = \underbrace{\frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \frac{Z_{1,i-1}^{u,l} \zeta_{i,j}^{u,l} Z_{j+1,n}^{u,l}}{Z_{1,n}^{u,l}}}_{\pi_{ij}^{*l}} + \frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \sum_{k < i} \sum_{m > j} p_{km}^{u,l} \equiv_{ij,km}$$

probability for base pair i, j in window u , given no base pair encloses i, j



RNAfold

pair probability

with windows of length l starting at u , mean pair probability π_{ij}^l

$$\pi_{ij}^l = \underbrace{\frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \frac{z_{1,i-1}^{u,l} \zeta_{i,j}^{u,l} z_{j+1,n}^{u,l}}{z_{1,n}^{u,l}}}_{\pi_{ij}^{*l}} + \frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \sum_{k < i} \sum_{m > j} p_{km}^{u,l} \equiv_{ij,km}$$

For every pair k, m enclosing i, j



RNAfold

pair probability

with windows of length l starting at u , mean pair probability π_{ij}^l

$$\pi_{ij}^l = \underbrace{\frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \frac{z_{1,i-1}^{u,l} \zeta_{i,j}^{u,l} z_{j+1,n}^{u,l}}{z_{1,n}^{u,l}}}_{\pi_{ij}^{*l}} + \frac{1}{l - (j - i) + 1} \sum_{u=j-l}^i \sum_{k < i} \sum_{m > j} p_{km}^{u,l} \equiv_{ij, km}$$

probability of pair i, j enclosed by k, m (0 if $k < u$)



pair probability

$$\begin{aligned}
 \pi_{ij}^l &= \pi_{ij}^{*l} + \frac{1}{l-(j-i)+1} \sum_{u=j-l}^i \sum_{k<i} \sum_{m>j} p_{km}^{u,l} \Xi_{ij,km} \\
 &= \pi_{ij}^{*l} + \sum_{k=j-l}^{i-1} \sum_{m=j+1}^{i+l} \sum_{u=m-l}^k \frac{p_{km}^{u,l} \Xi_{ij,km}}{l-(j-i)+1} \\
 &= \pi_{ij}^{*l} + \sum_{k=j-l}^{i-1} \sum_{m=j+1}^{i+l} \frac{l-(k-m)+1}{l-(j-i)+1} \pi_{km}^l \Xi_{ij,km} .
 \end{aligned}$$

This can also be computed in $\mathcal{O}(l^2 n)$

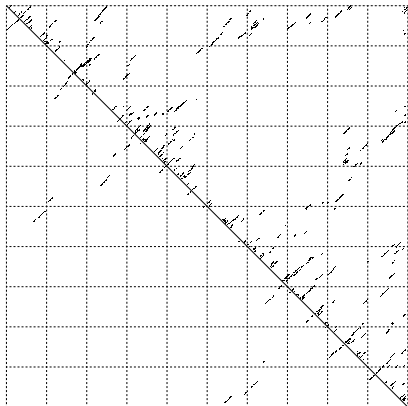
We can compute $s = 1$ as fast as $s = l$



RNAplfold

Output

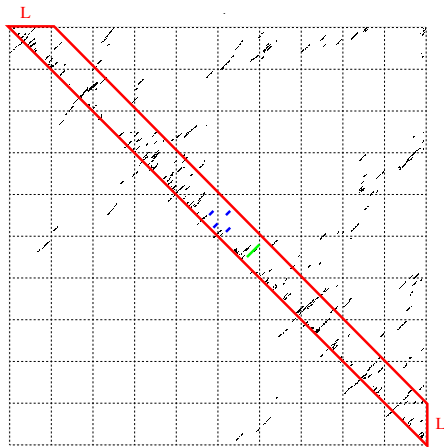
RNAfold -p Output



RNAplfold

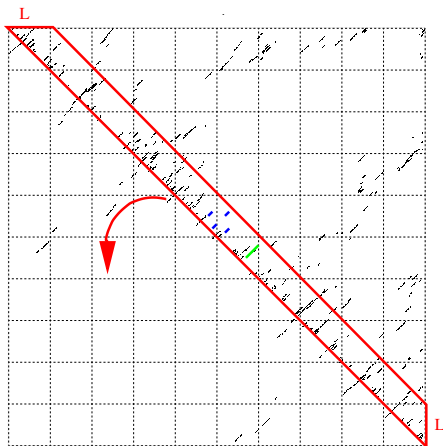
Output

Information of pfold



RNAfold Output

Cut it out/turn it round

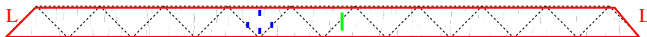




RNAplfold

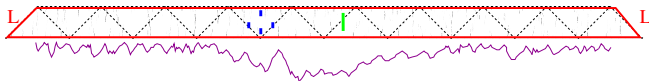
Output

Cut it out/turn it round



Conclusion

- Tool for fast computation of pair probabilities
- Can maybe be combined with genome browsers
- Problem: probabilities don't add up to 1
- Some measure of relevance of the structures is needed
- We plan to integrate energy z-score information





Local Alifold

Alifold

Alifold predicts consensus structure of aligned sequences



Local Alifold

Alifold

Alifold predicts consensus structure of aligned sequences

- Functional RNAs have conserved structures
- Consistent mutations e.g. A-U to G-U
- Compensatory Mutations e.g. G-C to A-U

RNAalifold uses consistent and compensatory mutations as well as thermodynamic energy parameters to evaluate the energy of a structure.



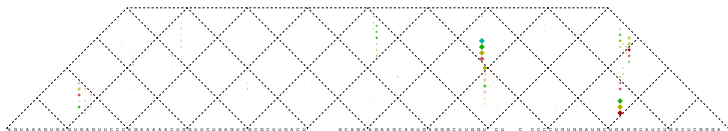
Local Alifold

- Algorithmically, the only difference between RNAalifold and RNAfold are the energy evaluation functions.
- It is straightforward to write local variants of RNAalifold
- We did local variants of mfe and partition function Alifold



Alifold Local Output

alidotL.ps





I want to thank

Peter Stadler
Ivo Hofacker
Ulli Mückstein
Hakim Tafer
Stefan Washietl
..and the audience

Something else you always wanted to have from ViennaRNA or
to implement yourself but just did not get around to do it?