Pseudoknot Prediction

- like Zuker algorithm, but without restriction to nested structures
- no method for arbitrary pseudoknot available (NP-hard)

  BUT: pseudoknot base-pair maximization NOT NP-hard!

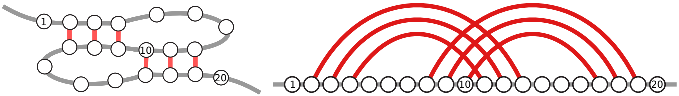
  Lyngso and Pedersen. RECOMB, 2000: NP-hard in “nearest neighbor model” by reduction to 3SAT.

  implies NP-hard for loop-based models

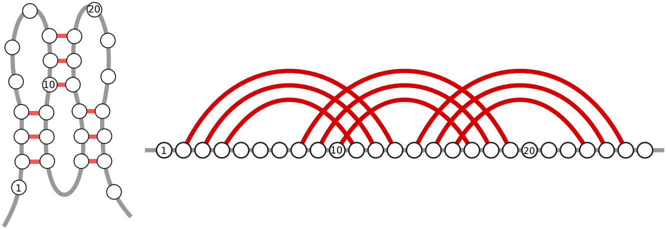
⇒ restrict complexity of pseudoknots (then, solve efficiently)

Pseudoknot Types

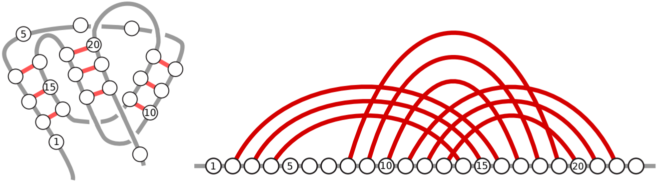
- Simple, H-type



- Kissing Hairpin



- Three-knot



Pseudoknot Prediction

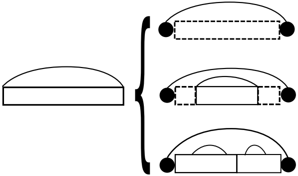
- algorithms for several restricted classes of pseudoknots:

class	R&G	A/U	L&P	D&P	CCJ	R&E
time	$O(n^4)$	$O(n^4)/O(n^5)$	$O(n^5)$	$O(n^5)$	$O(n^5)$	$O(n^6)$
space	$O(n^2)$	$O(n^3)/O(n^3)$	$O(n^3)$	$O(n^4)$	$O(n^4)$	$O(n^4)$

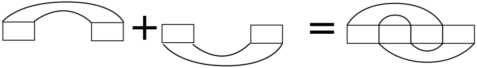
- R&G (Reeder-Giegerich) is most restricted class of pseudoknots → fastest algorithm (PKnotsRG)
- R&E (Rivas-Eddy) is most general class of pseudoknots → slowest algorithm

Pseudoknot Prediction: General Idea

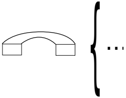
Like Zuker algorithm cases for different loops:



additional cases for pseudoknots using fragments with gaps:



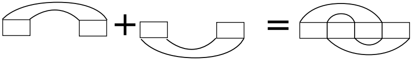
separate recursion for gapped fragments necessary:



- in practice many different types of gapped fragments necessary (e.g. with/without base pair outside/around the gap)
- exact recursion is different for each of the algorithms

Only One Gap: R&E

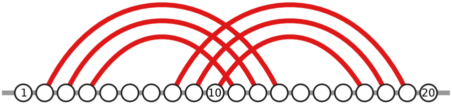
Idea of R&E:

- all you can do with restriction "only one gap";
  for example: 
- loop-based energy model ⇒ combinatorial blow-up
- energy parameters for pk-loops?
- all cases with "only one gap" fragments
  ⇒ specific computational complexity

⇒ R&E is the most complex, still "reasonably" efficient PK-prediction algorithm based on DP

Efficiently Decomposable Pseudoknots

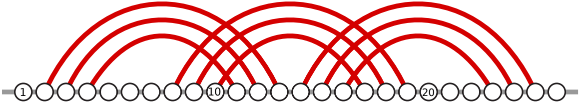
- Simple, H-type (Akutsu (A/U), $O(n^4)/O(n^3)$)



- Kissing Hairpin (Chen, Condon, Jabbari (CCJ), $O(n^5)/O(n^4)$)
- Three-knot (Rivas-Eddy $O(n^6)/O(n^4)$)
- Closed Five-chain ($O(n^?)$)

Efficiently Decomposable Pseudoknots

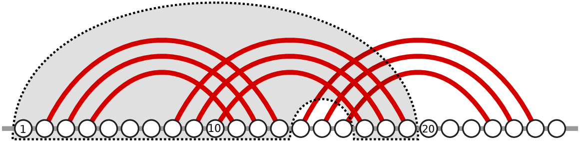
- Simple, H-type (Akutsu (A/U), $O(n^4)/O(n^3)$)
- Kissing Hairpin (Chen, Condon, Jabbari (CCJ), $O(n^5)/O(n^4)$)



- Three-knot (Rivas-Eddy $O(n^6)/O(n^4)$)
- Closed Five-chain ($O(n^?)$)

Efficiently Decomposable Pseudoknots

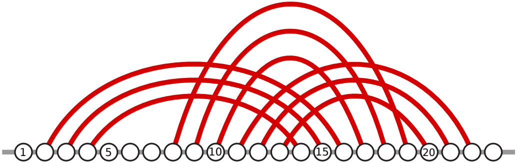
- Simple, H-type (Akutsu (A/U), $O(n^4)/O(n^3)$)
- Kissing Hairpin (Chen, Condon, Jabbari (CCJ), $O(n^5)/O(n^4)$)



- Three-knot (Rivas-Eddy $O(n^6)/O(n^4)$)
- Closed Five-chain ($O(n^?)$)

Efficiently Decomposable Pseudoknots

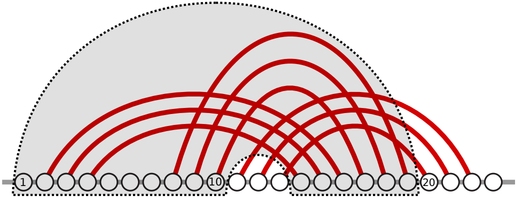
- Simple, H-type (Akutsu (A/U), $O(n^4)/O(n^3)$)
- Kissing Hairpin (Chen, Condon, Jabbari (CCJ), $O(n^5)/O(n^4)$)
- Three-knot (Rivas-Eddy $O(n^6)/O(n^4)$)



- Closed Five-chain ($O(n^?)$)

Efficiently Decomposable Pseudoknots

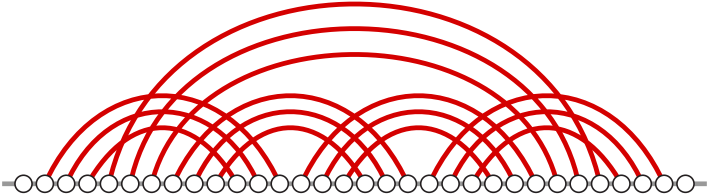
- Simple, H-type (Akutsu (A/U), $O(n^4)/O(n^3)$)
- Kissing Hairpin (Chen, Condon, Jabbari (CCJ), $O(n^5)/O(n^4)$)
- Three-knot (Rivas-Eddy $O(n^6)/O(n^4)$)



- Closed Five-chain ($O(n^?)$)

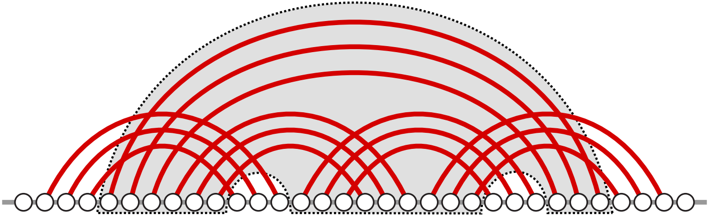
Efficiently Decomposable Pseudoknots

- Simple, H-type (Akutsu (A/U), $O(n^4)/O(n^3)$)
- Kissing Hairpin (Chen, Condon, Jabbari (CCJ), $O(n^5)/O(n^4)$)
- Three-knot (Rivas-Eddy $O(n^6)/O(n^4)$)
- Closed Five-chain ($O(n^?)$)



Efficiently Decomposable Pseudoknots

- Simple, H-type (Akutsu (A/U), $O(n^4)/O(n^3)$)
- Kissing Hairpin (Chen, Condon, Jabbari (CCJ), $O(n^5)/O(n^4)$)
- Three-knot (Rivas-Eddy $O(n^6)/O(n^4)$)
- Closed Five-chain ($O(n^?)$)



Very Efficient Pseudoknot Prediction: PKnotsRG

PKnotsRG is restricted to **canonical pseudoknots**

A canonical stem with outermost base pair (i, j) consists of the base pairs $(i + k, j - k)$, $k \geq 0$ such that for all $0 \leq k' \leq k$ $(i + k', j - k')$ is a valid Watson Crick base pair.
A canonical pseudoknot consists of two crossing canonical stems.

Example
Canonical pseudoknot with outermost base pairs $(i, j), (i', j')$

```
GAGACACGAGCUAUUGCGGAUCGUAGCUUAGCUCGUUCCCGAUCAGUGC
.....i............i'...............j.......j'....
```

Very Efficient Pseudoknot Prediction: PKnotsRG

PKnotsRG is restricted to **canonical pseudoknots**

A canonical stem with outermost base pair (i, j) consists of the base pairs $(i + k, j - k)$, $k \geq 0$ such that for all $0 \leq k' \leq k$ $(i + k', j - k')$ is a valid Watson Crick base pair.
A canonical pseudoknot consists of two crossing canonical stems.

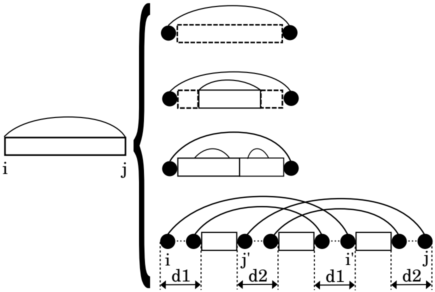
Example
Canonical pseudoknot with outermost base pairs $(i, j), (i', j')$

```
.....((((((((.....[[[[[.....))))))))...]]]]].....
GAGACACGAGCUAUUGCGGAUCGUAGCUUAGCUCGUUCCCGAUCAGUGC
.....i............i’...............j.......j’....
```

Very Efficient Pseudoknot Prediction: PKnotsRG

- canonical pseudoknots are likely to occur (stable)
- a sequence contains only $O(n^4)$ canonical pseudoknots
- limitations: only 2 stems, no bulges,...

PKnotsRG = Zuker recursion with one additional pseudoknot case



for all i' and j'
for d1, d2 such that the
stems are cannonical

$\longrightarrow O(n^4)$ time, $O(n^2)$ space.

Efficient (DP) Pseudoknot Prediction: Literature

- E. Rivas, S. R. Eddy, A dynamic programming algorithm for RNA structure prediction including pseudoknots,JMB 1999
- Rune B. Lyngso, Christian N. S. Pedersen, Pseudoknots in RNA Secondary Structures, RECOMB 2000
- T. Akutsu, Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots,DAM 2000
- Robert M. Dirks, Niles A. Pierce, A partition function algorithm for nucleic acid secondary structure including pseudoknots, JCC 2003
- Jens Reeder, Robert Giegerich,Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics, PNAS 2004
- Anne Condon, Beth Davy, Baharak Rastegari, Shelly Zhao, Finbarr Tarrant,Classifying RNA pseudoknotted structures, TCS 2004
- Ho-Lin Chen, Anne Condon, Hosna Jabbari ,An O(n(5)) Algorithm for MFE Prediction of Kissing Hairpins and 4-Chains in Nucleic Acids, JCB 2009

Heuristic Pseudoknot-Prediction

- ILM (Iterated Loop Matching)

  Ruan,J., Stormo,G. and Zhang,W. An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. Bioinformatics, 2004.

  - iterates Nussinov-like algorithm: “hierarchical folding”
  - least commitment strategy: keep only most reliable stem

- HotKnots

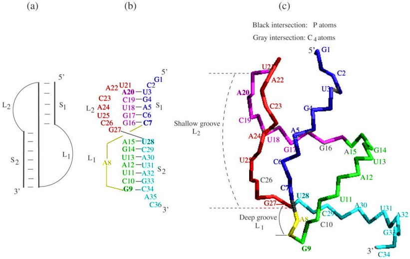
  Jihong Ren, Baharak Rastegari, Anne Condon, and Holger H. Hoos. Hotknots: Heuristic prediction of RNA secondary structures including pseudoknots. RNA 2005

  - Select “hot spots” = simple secondary structure elements with good energy
  - Iteratively add elements to final structure

Why Heuristics Pseudoknot Prediction?

- Speed: DP-Algos for most general cases are expensive
- Accuracy: can all effect be covered in the loop-based model? For example simple H-type pseudoknots:

Song Cao and Shi-Jie Chen* Predicting RNA pseudoknot folding thermodynamics



(a) Loops L1 and L2 span the deep narrow (major) and the shallow wide (minor) grooves, respectively. (b) gene 32 mRNA pseudoknot of bacteriophage T2 and (c) corresponding atomic structure from NMR structure (PDB ID: 2TPK).