Efficient pseudoknot partition functions:

Results and (still some) open problems

Sebastian Will

École Polytechnique - Institute Polytechnique de Paris

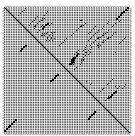
RNA@Benasque'24

Non-crossing RNA partition functions

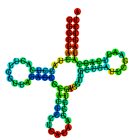
MFE prediction $\min_S E(S)$ $O(n^3)/O(n^2)$ [*Zuker*]

⇓ DISAMBIGUATION

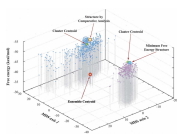
Partition Function $\sum_S \exp(-E(S)/RT)$ $O(n^3)/O(n^2)$ [*McCaskill*]



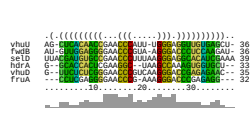
Base pair probabilities



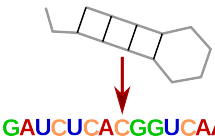
Local reliabilities



Structure sampling, Centroids, ...



RNA Alignment à la LocARNA

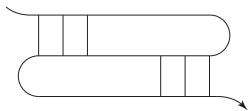


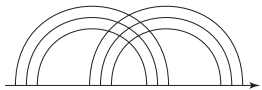
RNA Design: Pr, Ensemble defect, ...

Zuker, Stiegler 1981; Zuker, Sankoff 1984; McCaskill, Biopolymers 1990

Dirks&Pierce and the path to PK partition functions

- Simple pseudoknots

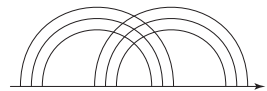




Dirks, Pierce. A partition function algorithm [...] including pseudoknots. JCC 2003

Dirks&Pierce and the path to PK partition functions

- Simple pseudoknots

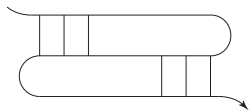


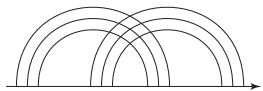
- Unambiguous decomposition *grammar*; 4D entries $\rightarrow O(n^5)$ / $O(n^4)$

Dirks, Pierce. A partition function algorithm [...] including pseudoknots. JCC 2003

Dirks&Pierce and the path to PK partition functions

- Simple pseudoknots



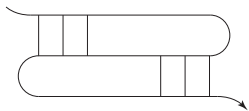


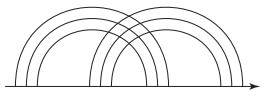
- Unambiguous decomposition *grammar*; 4D entries $\rightarrow O(n^5)$ / $O(n^4)$
- Partition function: "McCaskill" for simple PK

Dirks, Pierce. A partition function algorithm [...] including pseudoknots. JCC 2003

Dirks&Pierce and the path to PK partition functions

- Simple pseudoknots





- Unambiguous decomposition *grammar*; 4D entries $\rightarrow O(n^5)$ / $O(n^4)$
- Partition function: "McCaskill" for simple PK

  $\ldots +$ also MFE computation—same decomposition with *algebra exchange*

Dirks, Pierce. A partition function algorithm [...] including pseudoknots. JCC 2003

Dirks&Pierce and the path to PK partition functions

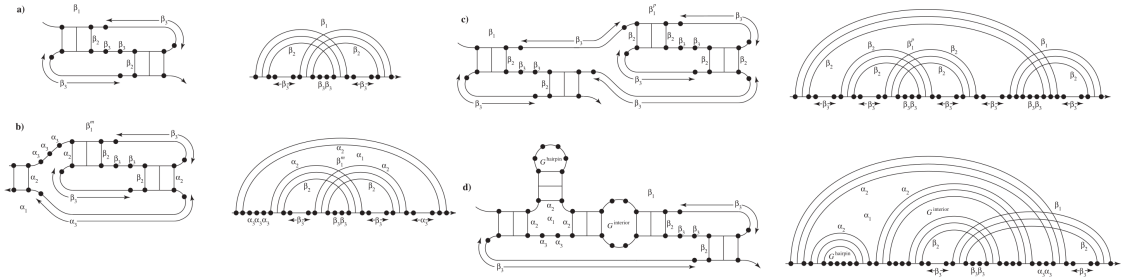
- Simple pseudoknots

- Unambiguous decomposition *grammar*; 4D entries $\rightarrow O(n^5)$ / $O(n^4)$
- Partition function: "McCaskill" for simple PK

  $\ldots +$ also MFE computation—same decomposition with *algebra exchange*
- Implementation in NUPACK

Dirks, Pierce. A partition function algorithm [...] including pseudoknots. JCC 2003

Dirks&Pierce / Hotknots 2.0 energy model

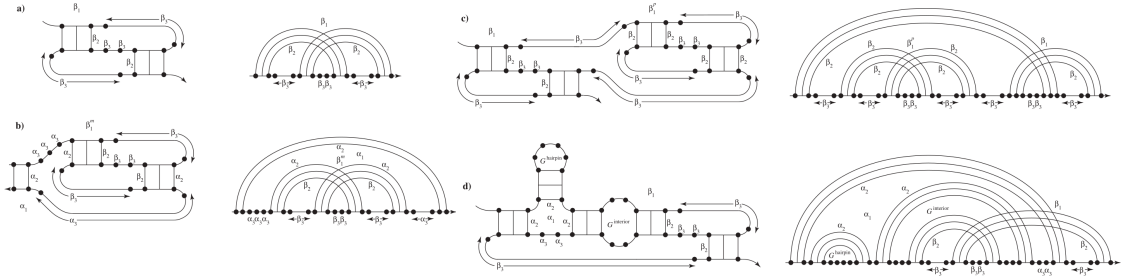


- recursive PK
- Turner NN + additional parameters: external/internal pseudoknot base pairs, bases in pseudoloops...
- Improved by Andronescu et al., 2010 → Hotknots 2.0

Dirks, Pierce. A partition function algorithm [...] including pseudoknots. JCC 2003

Andronescu et al. Improved free energy parameters ..., RNA 2010.

Dirks&Pierce / Hotknots 2.0 energy model

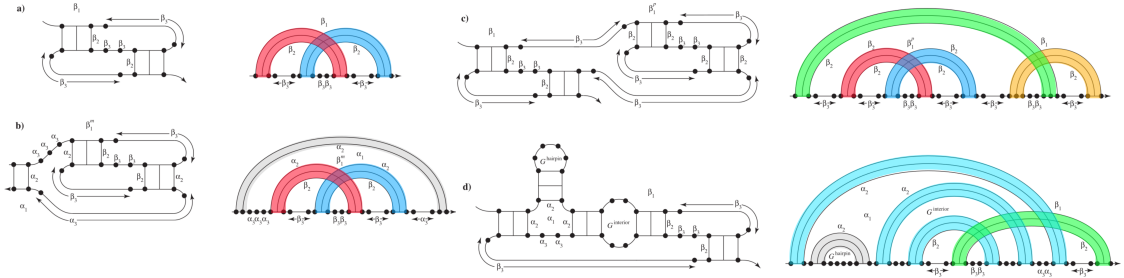


- recursive PK
- Turner NN + additional parameters: external/internal pseudoknot base pairs, bases in pseudoloops...
- Improved by Andronescu et al., 2010 → Hotknots 2.0
- general PK prediction in this energy model: NP-hard [Akutsu, 2000; Lyngsø&Pedersen, 2000]

Dirks, Pierce. A partition function algorithm [...] including pseudoknots. JCC 2003

Andronescu et al. Improved free energy parameters . . . , RNA 2010.

Dirks&Pierce / Hotknots 2.0 energy model

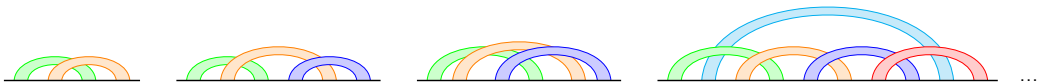


- recursive PK
- Turner NN + additional parameters: external/internal pseudoknot base pairs, bases in pseudoloops...
- Improved by Andronescu et al., 2010 → Hotknots 2.0
- general PK prediction in this energy model: NP-hard [Akutsu, 2000; Lyngsø&Pedersen, 2000]
- *bands*: mutually nested base pairs, crossed by the other base pairs in the same way

Dirks, Pierce. A partition function algorithm [...] including pseudoknots. JCC 2003

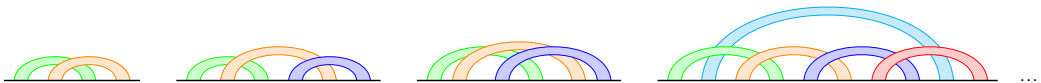
Andronescu et al. Improved free energy parameters . . . , RNA 2010.

Beyond simple pseudoknots (overview)



- Partition function for CCJ pseudoknots (or Rivas/Eddy pseudoknots)
  Pseudoknot classes motivated by computational complexity; "gap grammars"
- Partition functions for Fatgraphs/Shadows
  Controlled pseudoknot classes defined by sets of band configurations
- Hierarchically constrained partition function (HFold $\rightarrow$ CParty)
  High efficiency due to **hierarchical folding** *and* **structure class restriction**

Beyond simple pseudoknots (overview)



- **Partition function for CCJ pseudoknots (or Rivas/Eddy pseudoknots)**
  Pseudoknot classes motivated by computational complexity; "gap grammars"

- Partition functions for Fatgraphs/Shadows
  Controlled pseudoknot classes defined by sets of band configurations

- Hierarchically constrained partition function (HFold $\rightarrow$ CParty)
  High efficiency due to **hierarchical folding** *and* **structure class restriction**

CCJ partition function

- CCJ: $O(n^5)$ KHPs and 4-Chains minimization of DP/Hotknots 2.0 energy

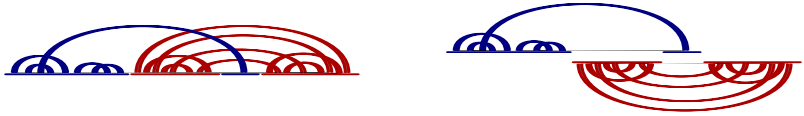
- DP algo idea: 

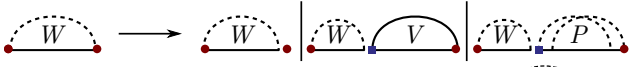
Chen, Condon, Jabbari. An $O(n^5)$ algorithm for KHP and 4-chain. JCB 2009.

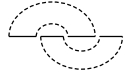
Jabbari et al., Knotty: prediction of complex RNA pseudoknot structures. Bioinformatics, 2018.

CCJ partition function

- CCJ: $O(n^5)$ KHPs and 4-Chains minimization of DP/Hotknots 2.0 energy



- DP algo idea: 

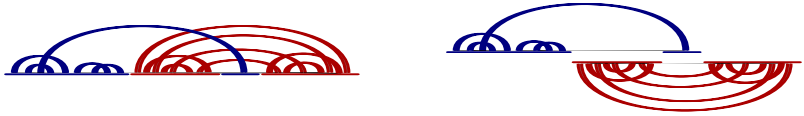
  - Compose CCJ PKs from two "1 gap" TGB fragments 

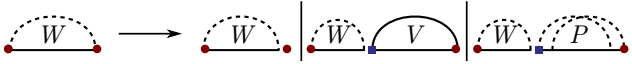
Chen, Condon, Jabbari. An $O(n^5)$ algorithm for KHP and 4-chain. JCB 2009.

Jabbari et al., Knotty: prediction of complex RNA pseudoknot structures. Bioinformatics, 2018.

CCJ partition function

- CCJ: $O(n^5)$ KHPs and 4-Chains minimization of DP/Hotknots 2.0 energy



- DP algo idea: 

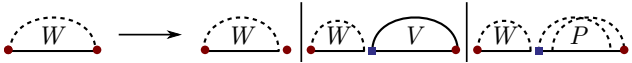
  - Compose CCJ PKs from two "1 gap" TGB fragments 
  - in TGB, alternate decomposition between Three Groups of Bands L,R,O/M

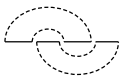
"interleaved bands"

Chen, Condon, Jabbari. An $O(n^5)$ algorithm for KHP and 4-chain. JCB 2009.

Jabbari et al., Knotty: prediction of complex RNA pseudoknot structures. Bioinformatics, 2018.

CCJ partition function

- CCJ: $O(n^5)$ KHPs and 4-Chains minimization of DP/Hotknots 2.0 energy
- DP algo idea: 

  - Compose CCJ PKs from two "1 gap" TGB fragments 
  - in TGB, alternate decomposition between Three Groups of Bands L,R,O/M

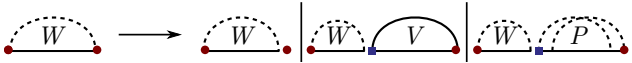
"interleaved bands"

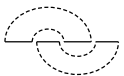
- Major challenge: **disambiguation**

Chen, Condon, Jabbari. An $O(n^5)$ algorithm for KHP and 4-chain. JCB 2009.

Jabbari et al., Knotty: prediction of complex RNA pseudoknot structures. Bioinformatics, 2018.

CCJ partition function

- CCJ: $O(n^5)$ KHPs and 4-Chains minimization of DP/Hotknots 2.0 energy
- DP algo idea: 

  - Compose CCJ PKs from two "1 gap" TGB fragments 
  - in TGB, alternate decomposition between Three Groups of Bands L,R,O/M

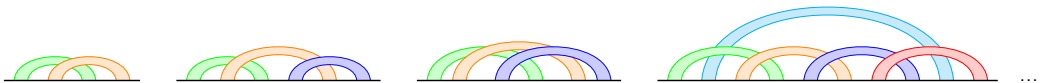
"interleaved bands"

- Major challenge: **disambiguation** $\Rightarrow$ **decomposition into restricted TGB**

Chen, Condon, Jabbari. An $O(n^5)$ algorithm for KHP and 4-chain. JCB 2009.

Jabbari et al., Knotty: prediction of complex RNA pseudoknot structures. Bioinformatics, 2018.

Beyond simple pseudoknots (overview)



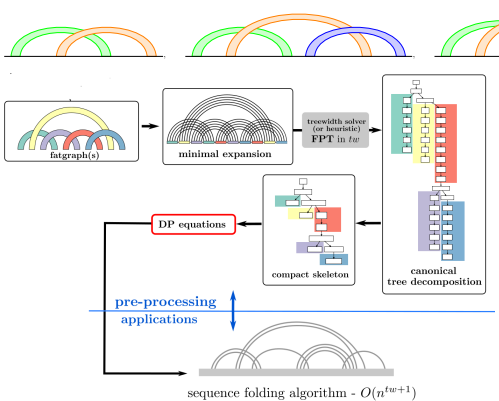
- Partition function for CCJ pseudoknots (or Rivas/Eddy pseudoknots)
  Pseudoknot classes motivated by computational complexity; "gap grammars"

- **Partition functions for Fatgraphs/Shadows**
  Controlled pseudoknot classes defined by sets of band configurations

- Hierarchically constrained partition function (HFold $\rightarrow$ CParty)
  High efficiency due to **hierarchical folding** *and* **structure class restriction**

Automatically generated algos based on fatgraphs

Recursive PK of controlled "band" configurations / topology

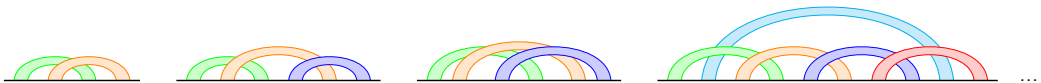


- Gfold: mfe + partition function for 1-structures
  manual rewriting of grammar rules per fatgraph
  e.g. $P_{KHP} \rightarrow A_1 IB_1 IA_2 IC_1 IB_2 IC_2$ "$O(n^{12})$"
- AutoDP automates grammar rewriting
  by tree decomposition
- in progress: full energy function, disambiguation, algebra exchange

Reidys et al., Topology and prediction of RNA pseudoknots. Bionformatics, 2011.

Marchand et al., Automated design of DP schemes for RNA folding with pseudoknots. ALMOB, 2023.

Beyond simple pseudoknots (overview)

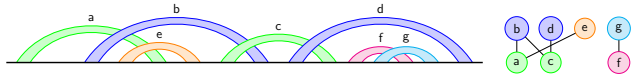


- Partition function for CCJ pseudoknots (or Rivas/Eddy pseudoknots)
  Pseudoknot classes motivated by computational complexity; "gap grammars"

- Partition functions for Fatgraphs/Shadows
  Controlled pseudoknot classes defined by sets of band configurations

- **Hierarchically constrained partition function (HFold → CParty)**
  High efficiency due to **hierarchical folding** *and* **structure class restriction**

Hierarchically constrained partition function (CParty)

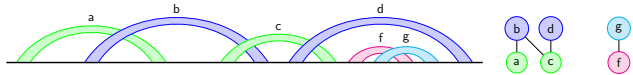


- **HFold**: Hotknots 2.0 MFE of bi-secondary "densitity-2" structures

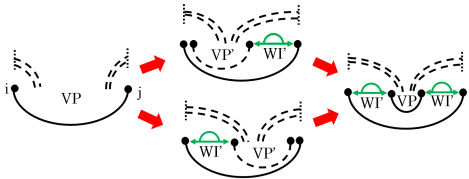
Trinity et al., CParty: Hierarchically Constrained Partition Function of RNA Pseudoknots, Submitted.

Jabbari et al., RNA Secondary Structure Prediction Using Hierarchical Folding. JCB, 2008.

Hierarchically constrained partition function (CParty)



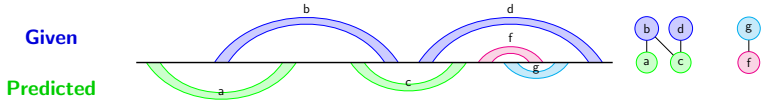
- **HFold**: Hotknots 2.0 MFE of bi-secondary "densitity-2" structures
- The ambiguity of HFold and its disambiguation



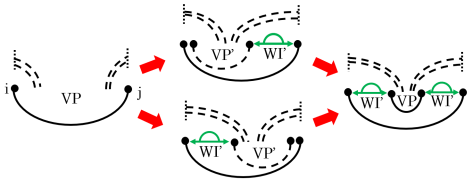
Trinity et al., CParty: Hierarchically Constrained Partition Function of RNA Pseudoknots, Submitted.

Jabbari et al., RNA Secondary Structure Prediction Using Hierarchical Folding. JCB, 2008.

Hierarchically constrained partition function (CParty)



- **HFold**: Hotknots 2.0 MFE of bi-secondary "densitity-2" structures
- The ambiguity of HFold and its disambiguation

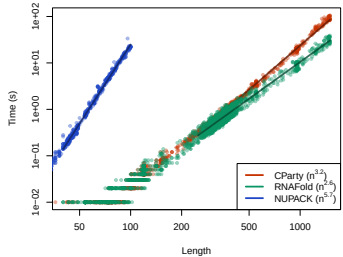


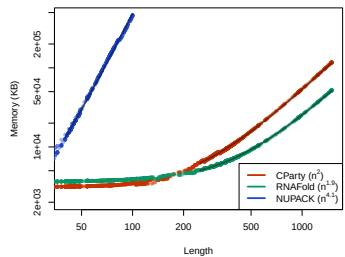
Trinity et al., CParty: Hierarchically Constrained Partition Function of RNA Pseudoknots, Submitted.

Jabbari et al., RNA Secondary Structure Prediction Using Hierarchical Folding. JCB, 2008.

CParty Empirical Results

- CParty: $O(n^3)/O(n^2)$

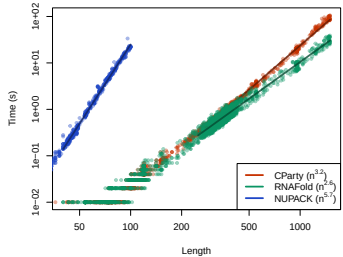


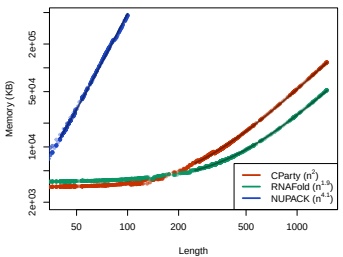


Trinity et al., CParty: Hierarchically Constrained Partition Function of RNA Pseudoknots, Submitted.

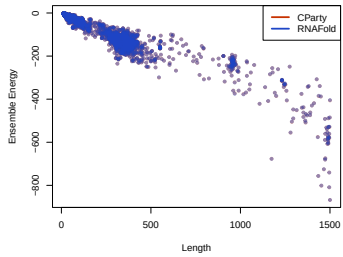
CParty Empirical Results

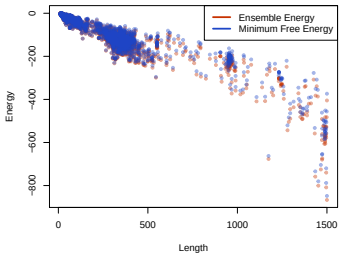
- CParty: $O(n^3)/O(n^2)$





- Empirical Validation





Trinity et al., CParty: Hierarchically Constrained Partition Function of RNA Pseudoknots, Submitted.

Conclusions

- Algorithm + Implementation of hierarchical partition function

  bioRχiv doi:10.1101/2023.05.16.541023  github:HosnaJabbari/CParty

- CCJ partition function full energy model: *in progress*
  - challenge: disambiguation (!) → restrictions of parts of PK

- Automatically generated PF algos based on fatgraphs: *in progress*
  - less complex disambiguation compared to CCJ
  - ongoing: implementation with ViennaRNA grammar extensions + full energy model

Conclusions

- Algorithm + Implementation of hierarchical partition function

  bioRχiv doi:10.1101/2023.05.16.541023 github:HosnaJabbari/CParty

- CCJ partition function full energy model: *in progress*
  - challenge: disambiguation (!) → restrictions of parts of PK

- Automatically generated PF algos based on fatgraphs: *in progress*
  - less complex disambiguation compared to CCJ
  - ongoing: implementation with ViennaRNA grammar extensions + full energy model

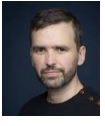

Hosna Jabbari


Ulrike Stege


Luke Trinity


Mateo Gray


Bertrand Marchand


Yann Ponty


Sarah Berkemer