

Disruptive Base-Pairs in RFAM alignments

By Patrick-Pascal Koller

Supervised by Hua-Ting Yao and Ivo Hofacker

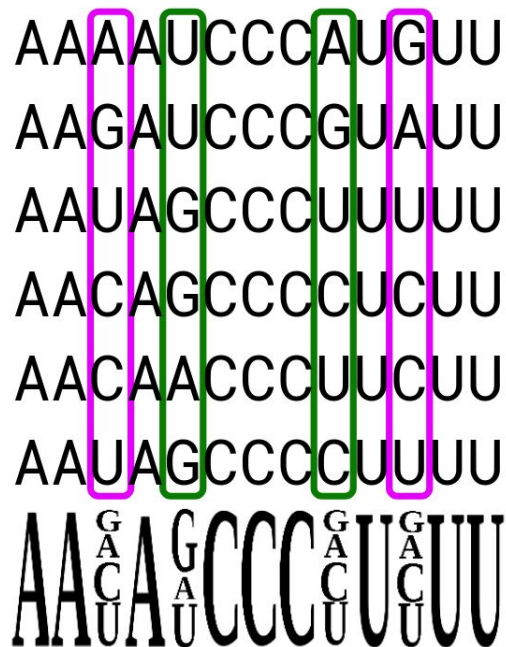


universität
wien

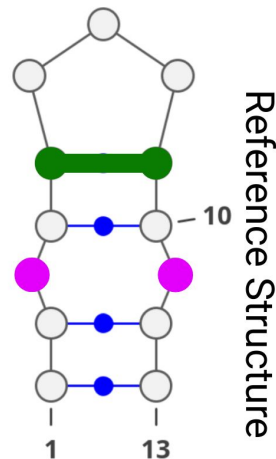
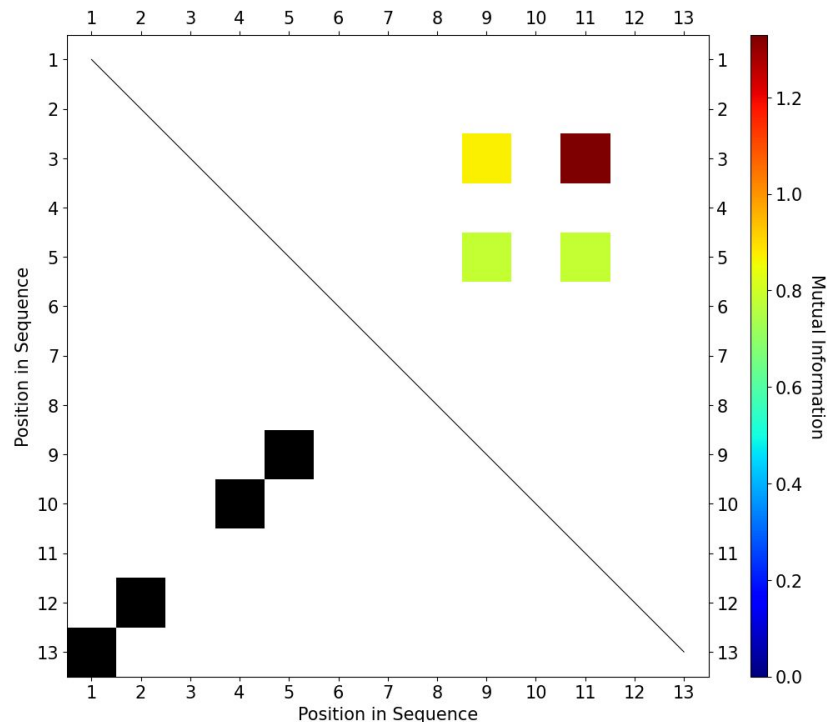
tbi

What are disruptive base-pairs (DBPs)?

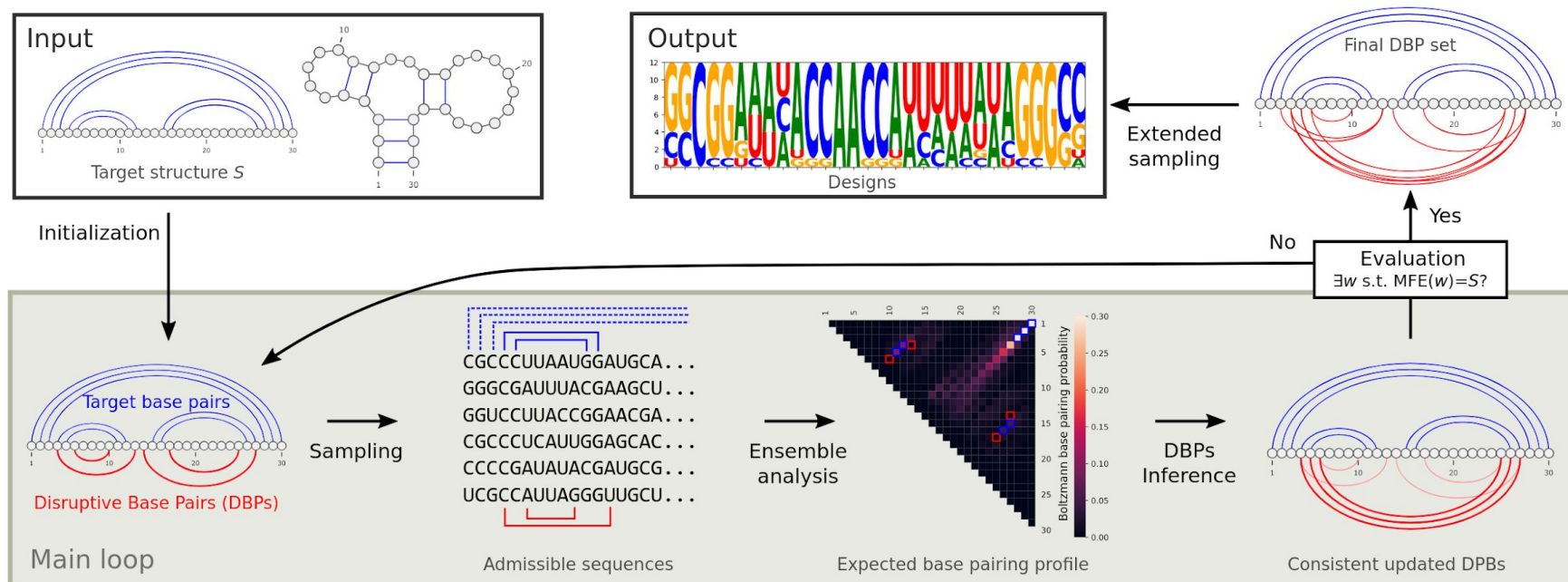
((. ((. . .)) .))



Mutual Information (top) vs. Consensus structure (bottom)



What are disruptive base-pairs (DBPs)?

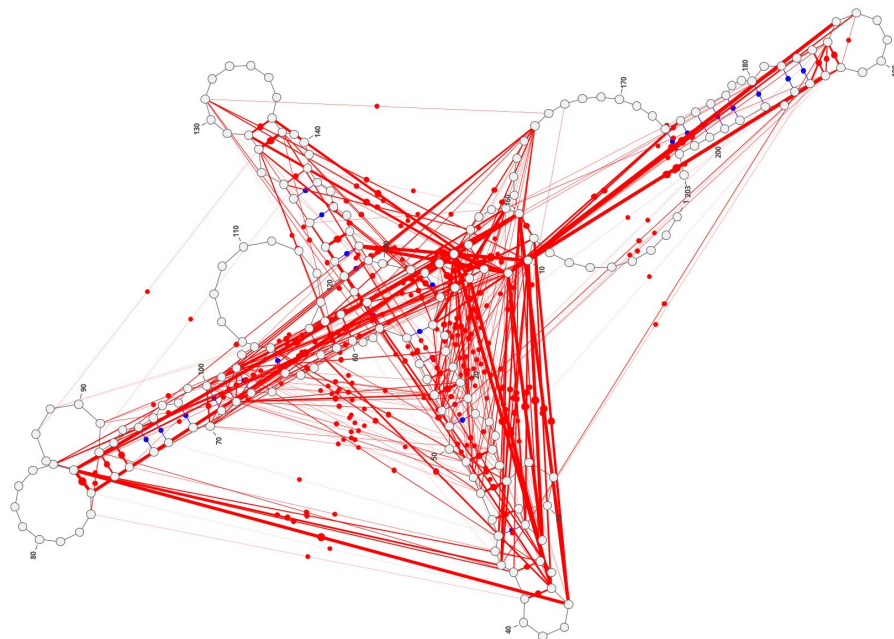
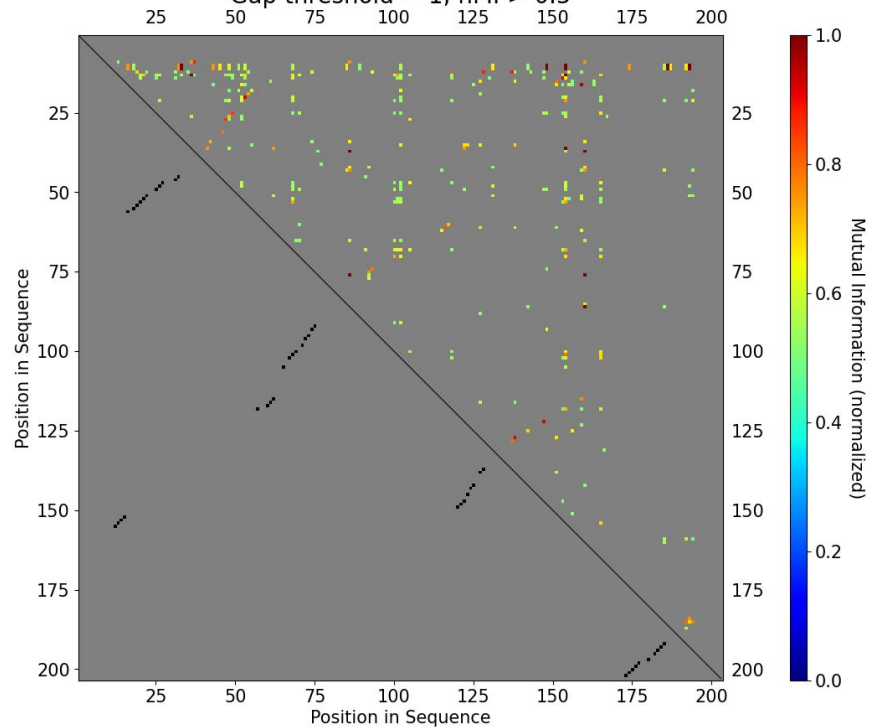


Can DBPs be found in nature? Mutual information

Mutual Information (Normalized, top) vs. Consensus structure (bottom)

RFAM family : RF00003

Gap threshold = 1, $nMI > 0.5$



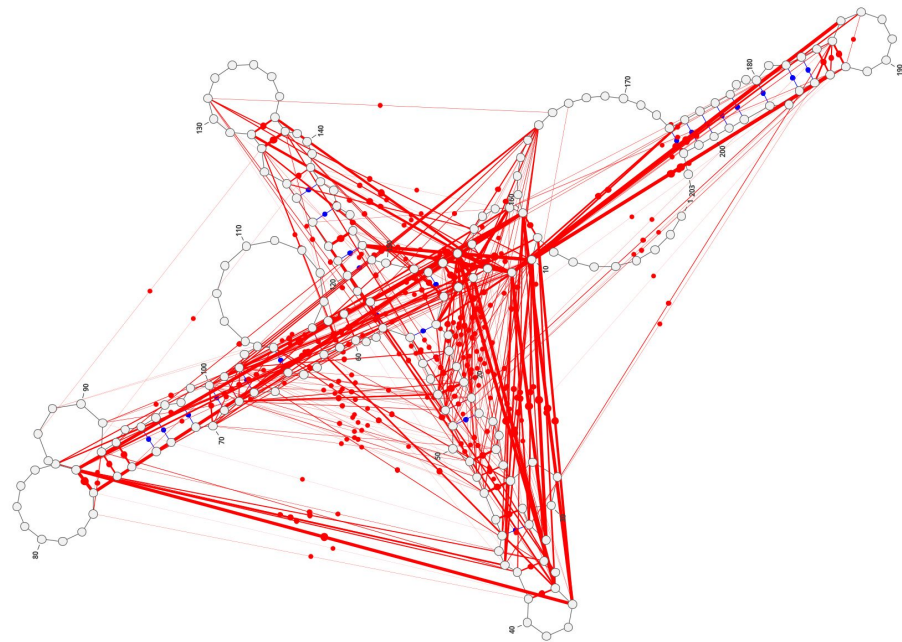
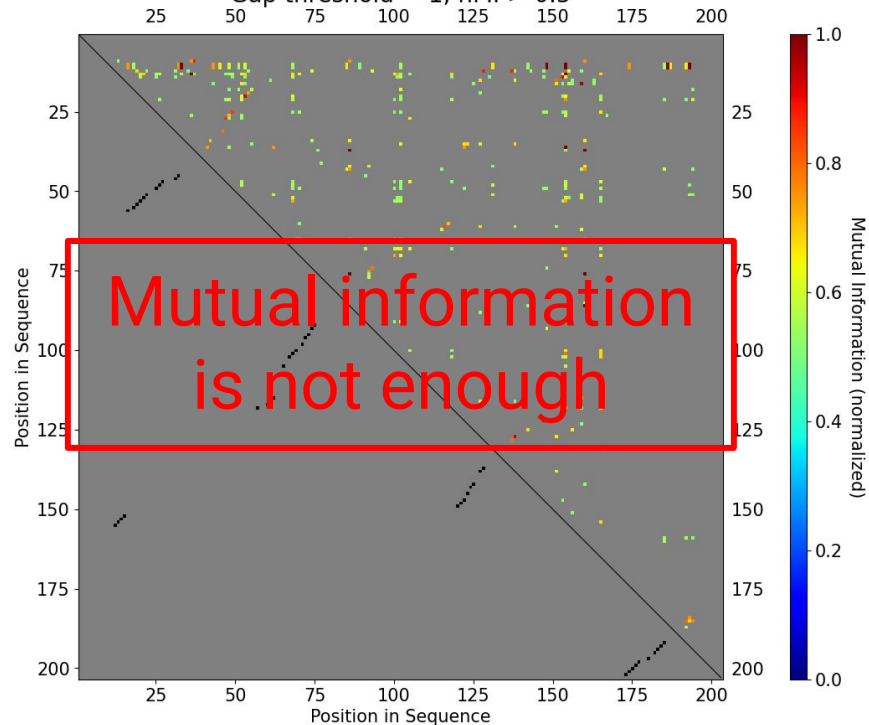
- Mutual information -> Measures structural conservation
- Normalized to (0, 1)

Can DBPs be found in nature? Mutual information

Mutual Information (Normalized, top) vs. Consensus structure (bottom)

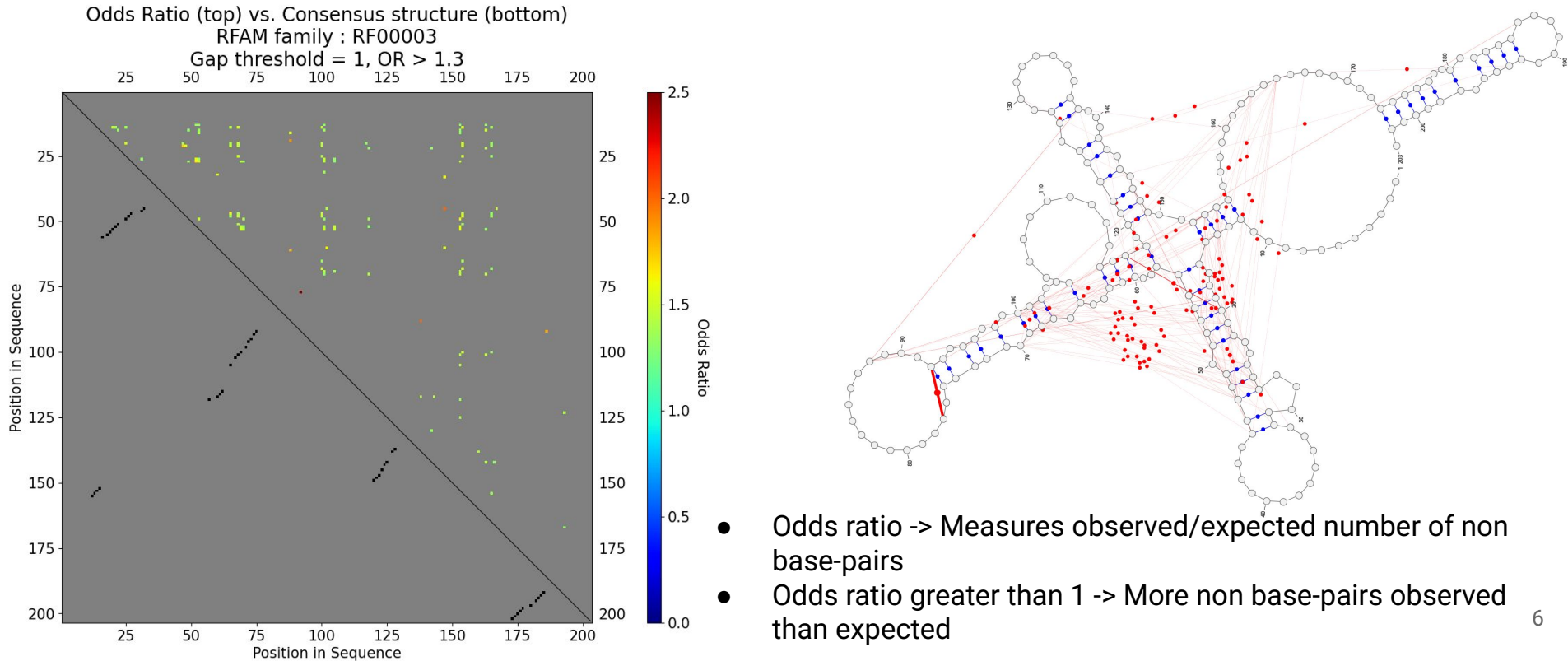
RFAM family : RF00003

Gap threshold = 1, $nMI > 0.5$



- Mutual information -> Measures structural conservation
- Normalized to (0, 1)

Can DBPs be found in nature? Odds ratio

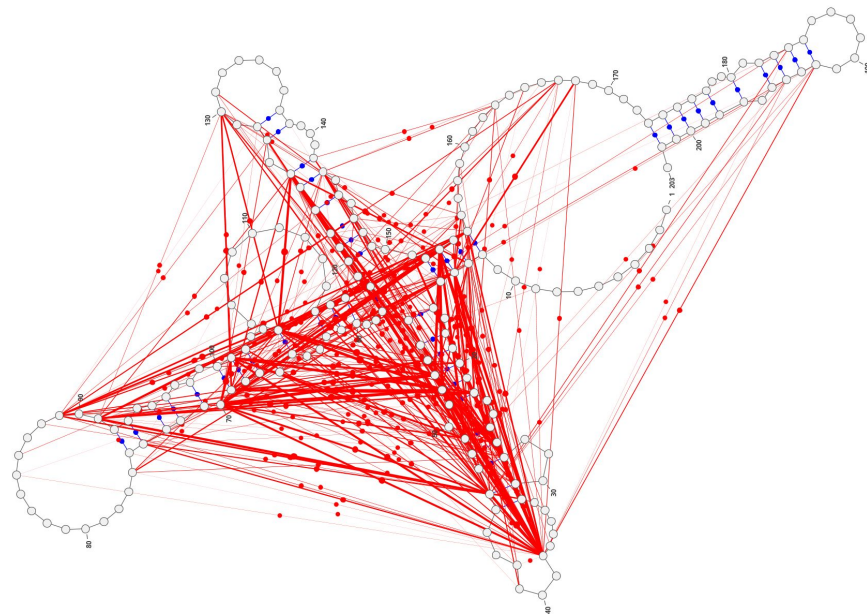
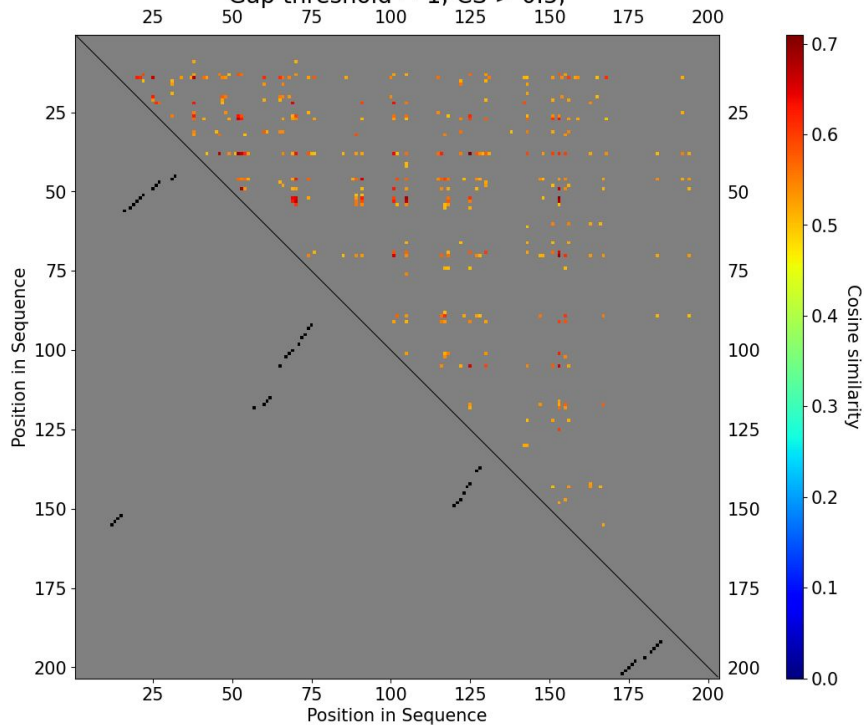


Can DBPs be found in nature? Cosine Similarity

Cosine similarity (top) vs. Consensus structure (bottom)

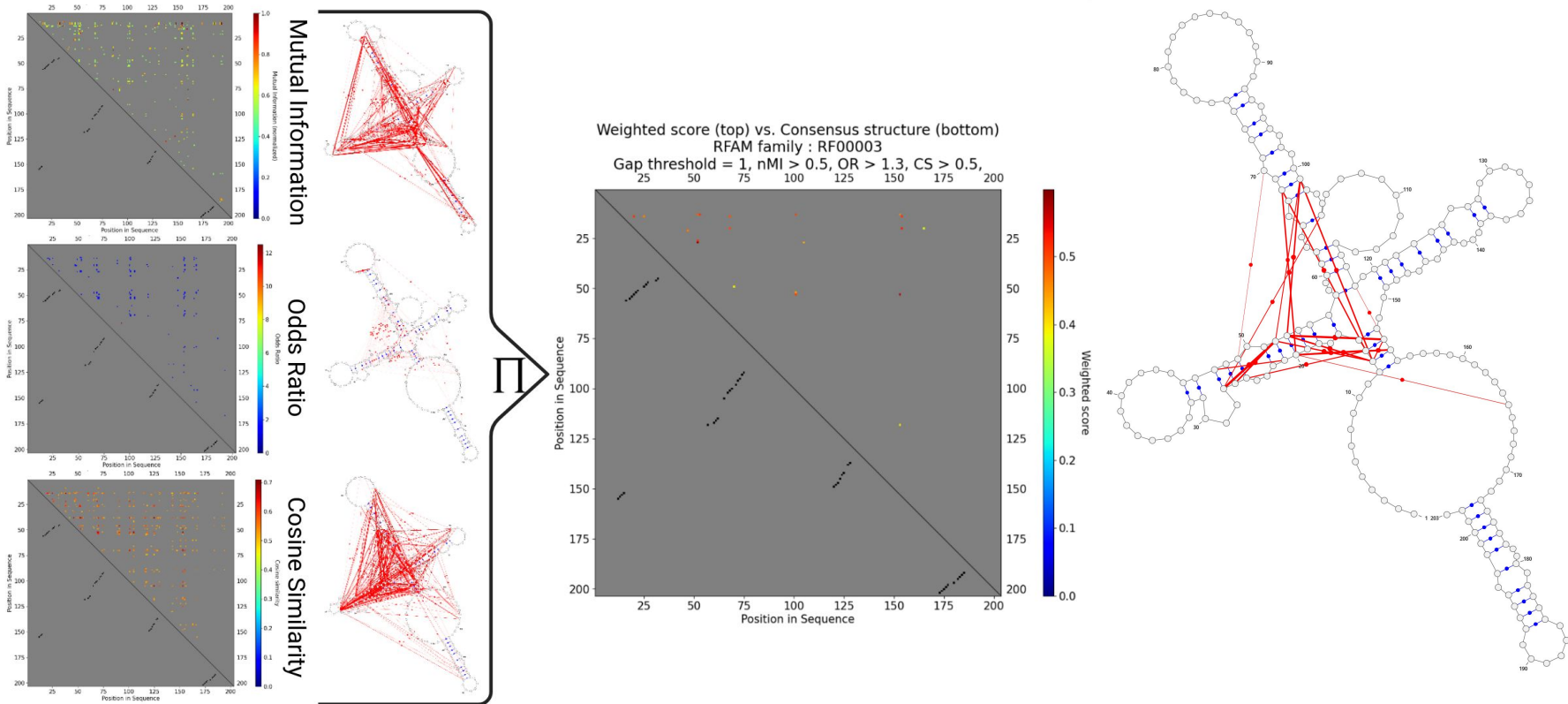
RFAM family : RF00003

Gap threshold = 1, CS > 0.5,



- Cosine similarity -> Tests if frequencies for all 10 non base-pairs (AA, AG..) are uniformly distributed
- More uniform distribution -> Cosine similarity closer to 1

How can DBPs be discerned from canonical BPs?

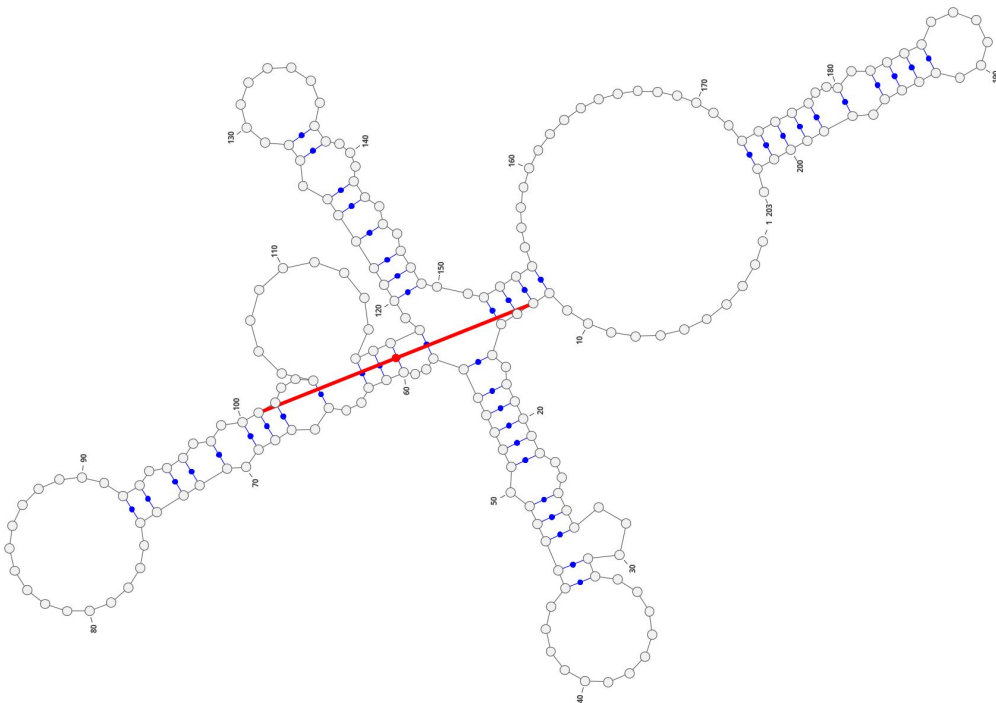


Under very strict conditions, can we still find DBPs?

Weighted score (top) vs. Consensus structure (bottom)

RFAM family : RF00003

Allowed canonical pairs = 0, Gap threshold = 1, nMI > 0.5, OR > 1.3, CS > 0.5,



Base-pair position: (13, 101)

Weighted score ≈ 0.5

Mutual information (normalized) ≈ 0.73

Odds ratio ≈ 1.31

Cosine similarity ≈ 0.52

Composition:

(C, C): 52

(A, G): 22

(C, U): 22

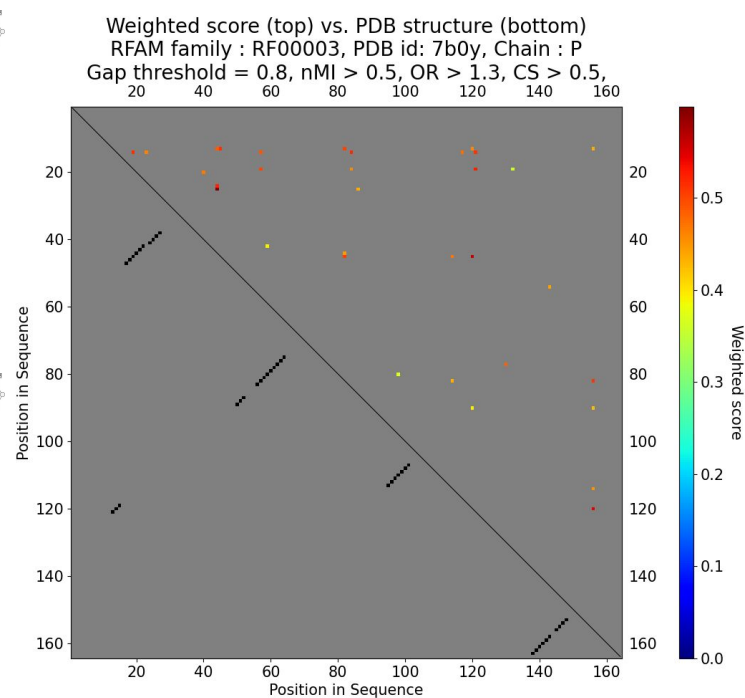
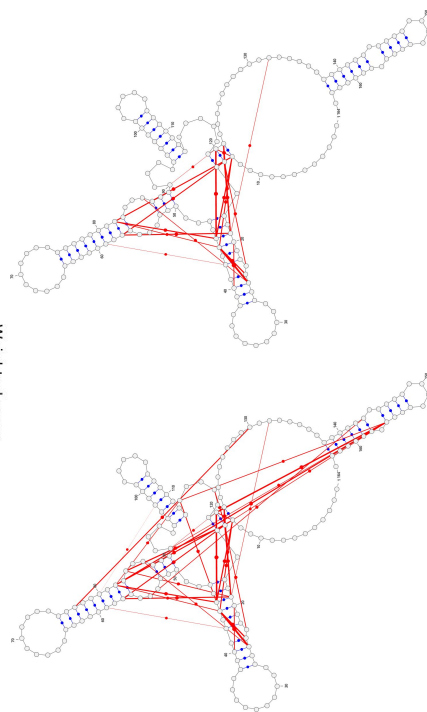
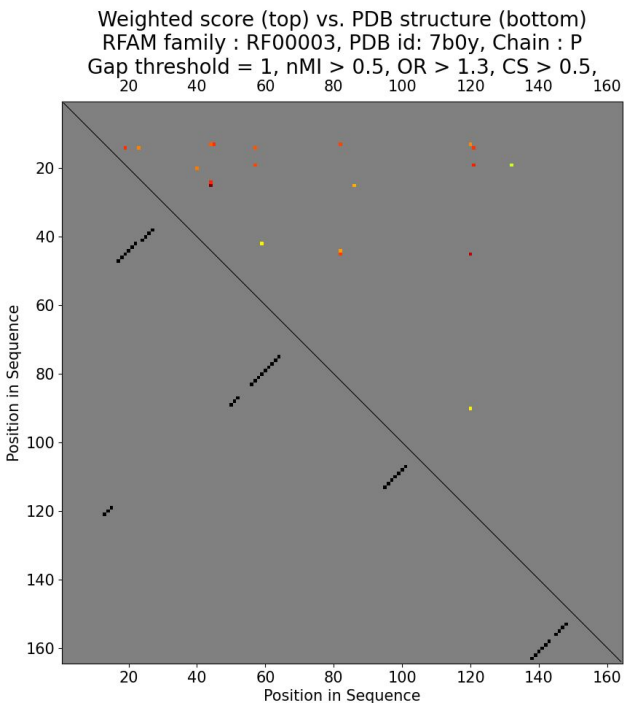
(A, C): 1

(U, U): 1

(U, C): 1

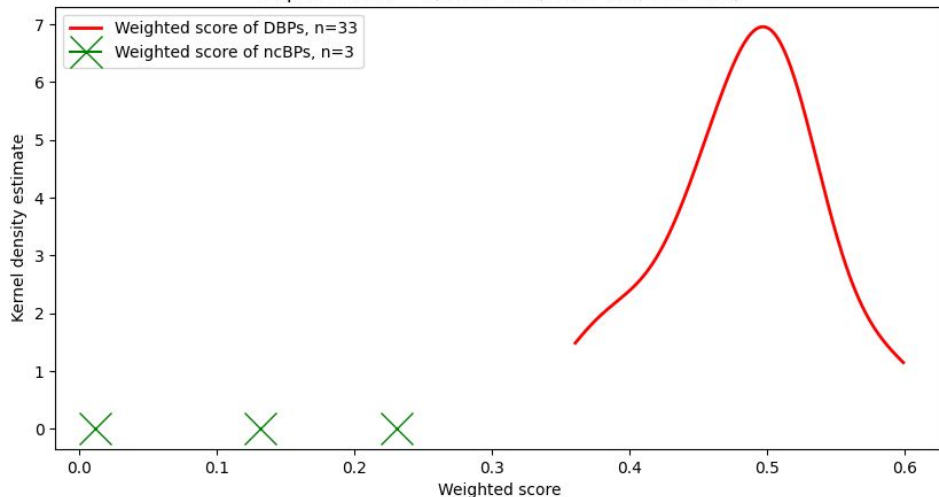
(G, G): 1

Effect of threshold reduction on DBP detection?

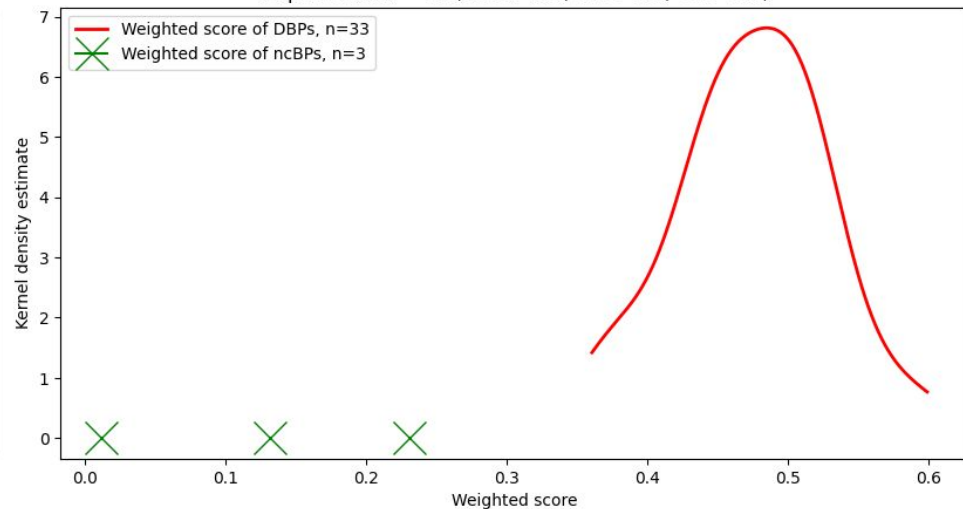


Is there a correlation between ncBPs and DBPs?

Comparison of weighted scores between DBPs and ncBPs
RFAM family : RF00003, PDB id: 7b0y, Chain : P
Gap threshold = 1, nMI > 0.5, OR > 1.3, CS > 0.5,

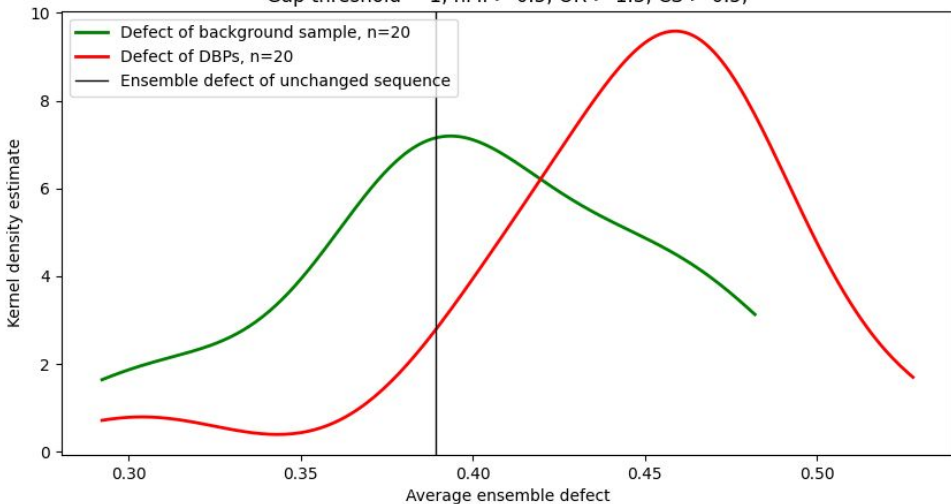


Comparison of weighted scores between DBPs and ncBPs
RFAM family : RF00003, PDB id: 7b0y, Chain : P
Gap threshold = 0.8, nMI > 0.5, OR > 1.3, CS > 0.5,

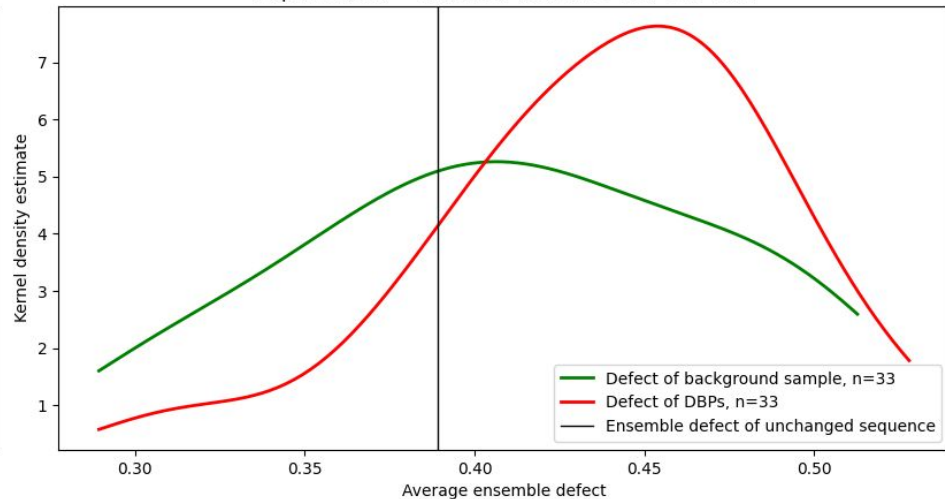


Do DBPs alter the ensemble defect greatly when forcing them to pair?

Averaged ensemble defect of DBPs and sampled background distribution
RFAM family : RF00003, PDB id: 7b0y, Chain : P
Gap threshold = 1, nMI > 0.5, OR > 1.3, CS > 0.5,



Averaged ensemble defect of DBPs and sampled background distribution
RFAM family : RF00003, PDB id: 7b0y, Chain : P
Gap threshold = 0.8, nMI > 0.5, OR > 1.3, CS > 0.5,



Future Outlook

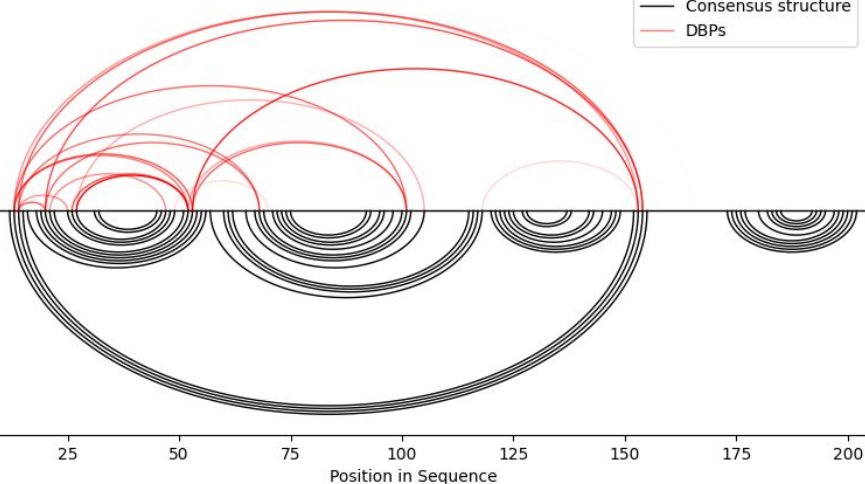
- Comparative analysis between different RNA families
- Further evaluation of DBP compositions
- DBPs may be needed for structure formation, e.g. Multi-loops:

Positions of disruptive base-pairs in relation to consensus structure

RFAM family : RF00003

Gap threshold = 1, nMI > 0.5, OR > 1.3, CS > 0.5,

— Consensus structure
— DBPs

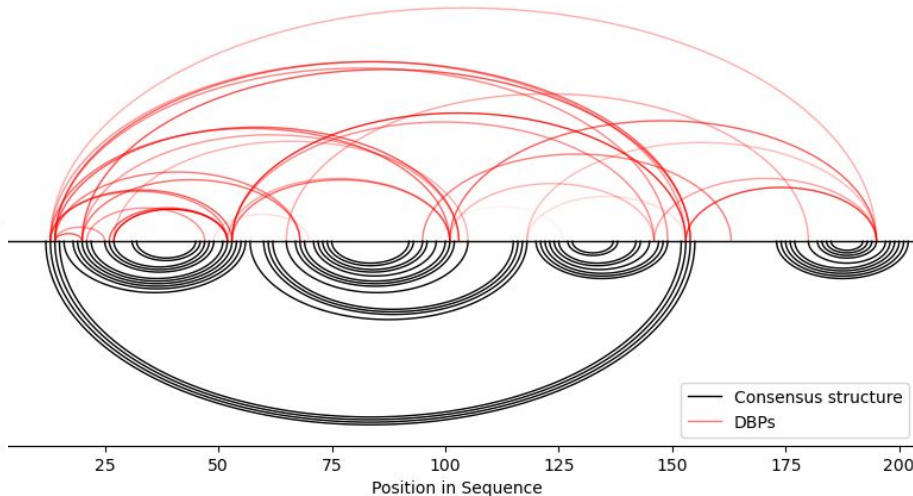


Positions of disruptive base-pairs in relation to consensus structure

RFAM family : RF00003

Gap threshold = 0.8, nMI > 0.5, OR > 1.3, CS > 0.5,

— Consensus structure
— DBPs



Conclusion

Disruptive base-pairs...

- ...can be found in nature-based multi-sequence alignments
- ...they do not coincide with non-canonical base-pairs
- ...have starting and ending positions primarily in duplexes
- ...are primarily located at the start or end of a substructure
- ...on average increase the ensemble defect

