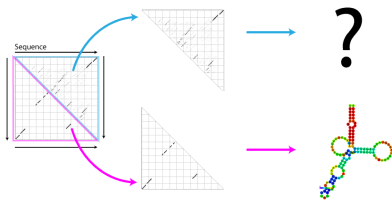


Visualizing RNA Secondary Structure Base Pair Binding Probabilities using Nested Concave Hulls

J. Sansen and R. Bourqui and P. Thebault and J. Allali and D. Auber
jsansen@labri.fr
LaBRI - Université de Bordeaux

Abstract—The challenge 1 of the BIOVIS 2015 design contest consists in designing an intuitive visual depiction of base pairs binding probabilities for secondary structure of ncRNA. Our representation depicts the potential nucleotide pairs binding using nested concave hulls over the computed MFE ncRNA secondary structure. Thus, it allows to identify regions with a high level of uncertainty in the MFE computation but also the structures which seem to match to reality.



1. From the MFE base-pair bindings and ncRNA sequence to Secondary structure [1]

2. Data Processing: Select sequences of matching pairs

Select **MFE-rejected sequence** of pairs binding using 2 parameters:

mPairs : minimum numbers of pairs in each sequence

nbLevels : numbers of levels of probability required.

Using the parameter *b* we define a **threshold *t*** which determine the **minimum base pair probability value authorized** in the sequence.

$$t = (\max P - \min P) / \text{nbLevels}$$

We loop *nbLevels* times through the matrix **increasing *t*** to pick smaller and smaller sets with a **range of probability more and more restrained**.

Additionally, we can **tune** the threshold in order to **filter out** of the selectable sets the ones with **negligible matching probabilities**.

3. Design: Using Nested Concave Hulls to reflect the uncertainty of pairs binding

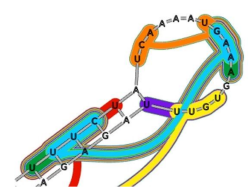
Potential base pairs binding are **emphasized** with the addition of **nested concave hulls**[2] which put forward the subsequences.

This technique makes possible to **order and nest hulls** by size of sets. Even sets with **low pairing probabilities** can be **depicted** and identified while putting forward **the most likely sequences**.

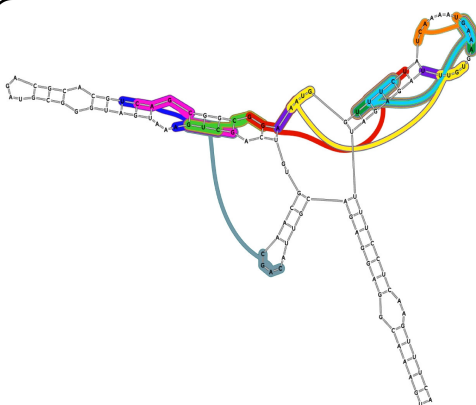
The **more** a sub-sequence is wrapped **into hulls**, the more its **probability is important**.

The **two hulls** of a base pairs sequence are **connected** with a **link**.

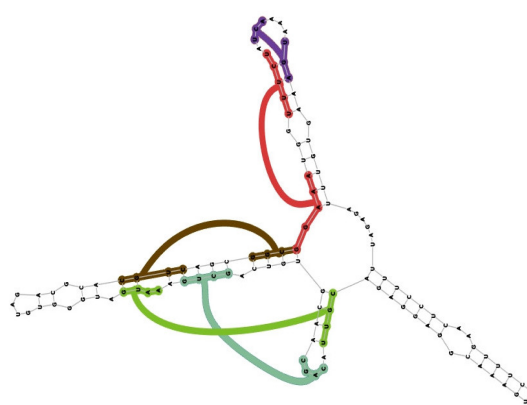
Using a bezier **curve bending** algorithm[3] we **ease the identification** of pairwise elements by **reducing clutter** and **links crossings**.



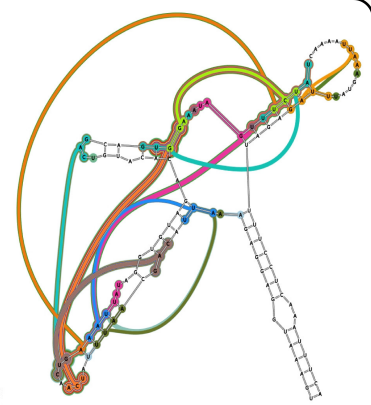
4: Nested Concave Hulls onto loop reflect the uncertainty of pair binding



a) Human



b) Denisovan



c) Ancestral Chimp

5. MFE-structure of ncRNA with potential pairs binding probabilities. From left to right: human, denisovan and ancestral chimp. Stems with multiple hulls indicate a high uncertainty while regions without hulls indicate a close-to-reality prediction.

Results and Conclusion

Effective and intuitive depiction of a ncRNA secondary structure **showing** the most **important potential** base-pairs **binding** rejected during the MFE-structure computation.

Our method **emphasizes** the most **probable** pair **bindings** and the **influence** they may have onto the predicted MFE-structure. The use of **nested concave hulls** produces a pleasant depiction of the potential base pairs binding which reveals the **uncertainty** in the **predicted MFE-structure** and the **region** of the structure that should remain **stable** with different base-pairs binding.

Multiple hulls over the same regions **indicate** a high level of **uncertainty** and a few or **no hulls** reflect the **strength** of the prediction for these regions.



REFERENCES:

- [1] D. Auber, M. Delest, J.-P. Domenger, and S. Dulucq. Efficient drawing of rna secondary structure. J. Graph Algorithms Appl., 2006.
- [2] A. Lambert, F. Queyroi, and R. Bourqui. Visualizing patterns in node-link diagrams. In Information Visualisation (IV), 2012.
- [3] A. Lambert, R. Bourqui, and D. Auber. Winding roads: Routing edges into bundles. Computer Graphics Forum, 2010.