RNA Structure Prediction & Design of Protein/Nucleic Acid Complexes

Ambika Acharya, Julia Wang

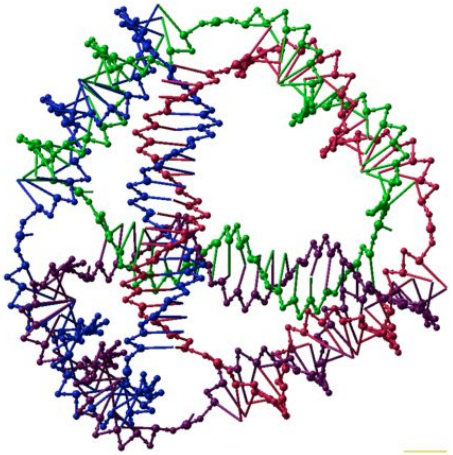
Computational design of co-assembling protein-DNA nanowires

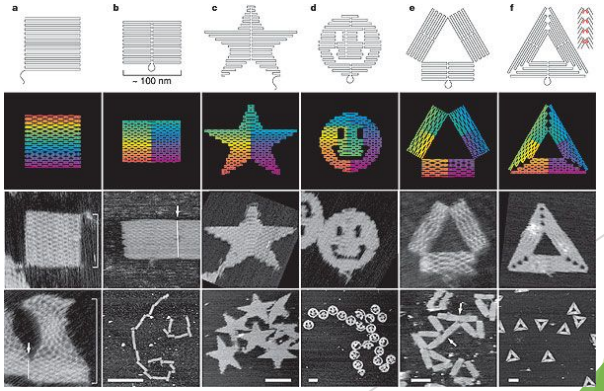
Yun Mou, Jiun-Yann Yu, Timothy M. Wannier, Chin-Lin Guo, & Stephen L. Mayo

California Institute of Technology, September 2015

DNA Nanotechnology

- Strong, Predictable Structure
- Self-Assembly
- Characterization
- Applications



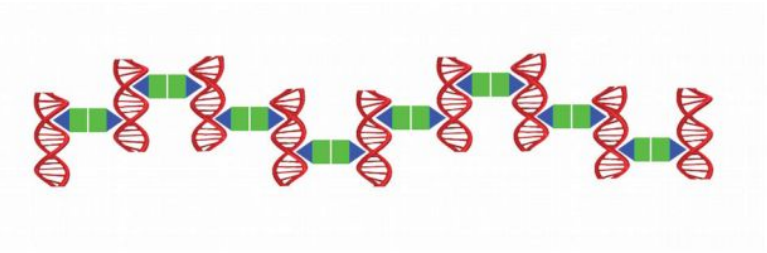


Hybrid protein-DNA assemblies

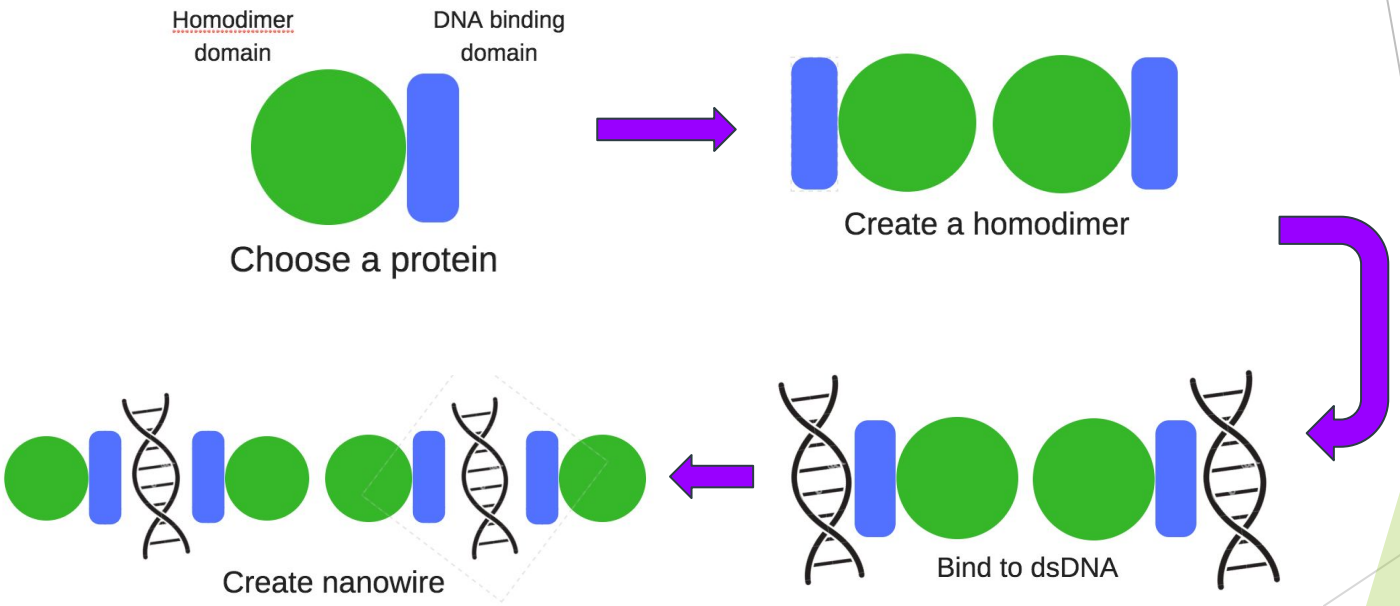
- Applications
- Previous Techniques
  - DNA scaffold + chemical conjugation
  - Many Limitations

Non-covalent co-assemblies

- Goal protein-DNA “wire”
  - self-assembly
- Method
  - Protein homodimer
  - Binds to dsDNA

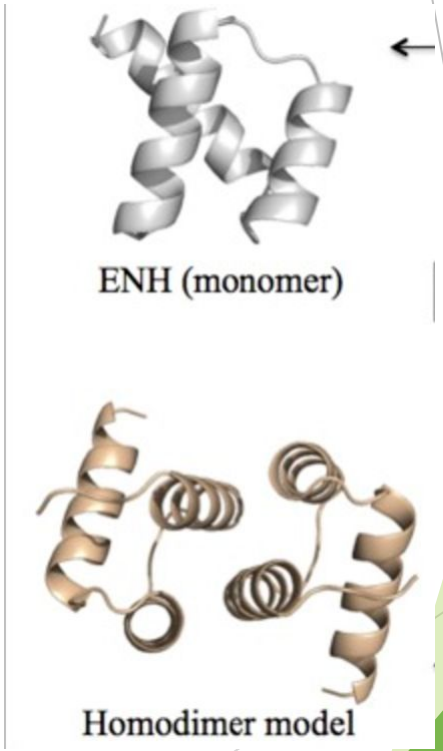


Model


Homodimer domain
DNA binding domain
Choose a protein
Create a homodimer
Bind to dsDNA
Create nanowire

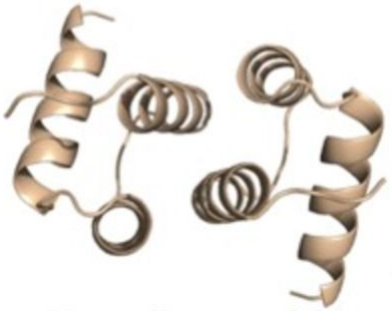
Choosing a Protein

- Engrailed homeodomain (ENH)
- Binds dsDNA tightly and specifically
- Highly studied
- 3 Helix Structure



Create a Homodimer

- Protein-docking algorithm to find homodimer
  - ENH as scaffold
  - Fast Fourier Transform -- surface complementarity
  - Symmetry reduces search space
- Clustering and visual inspection
  - Top 200 models by structural similarity
- Computational methods to stabilize side-chain interactions



Homodimer model

dualENH

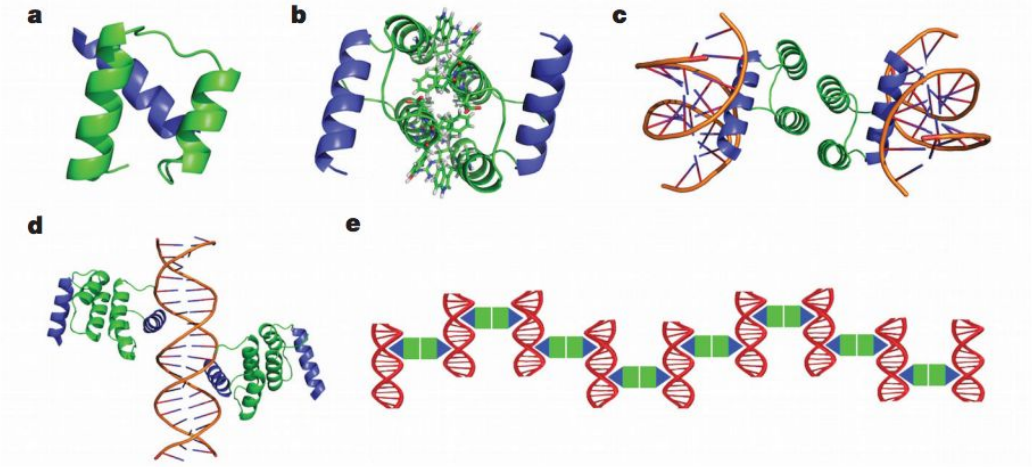


Figure 1 from Paper

dualENH Self-assembly

- Combined dualENH with dsDNA
- Observed self-assembly using fluorescence microscopy
- Nanoparticles formed immediately

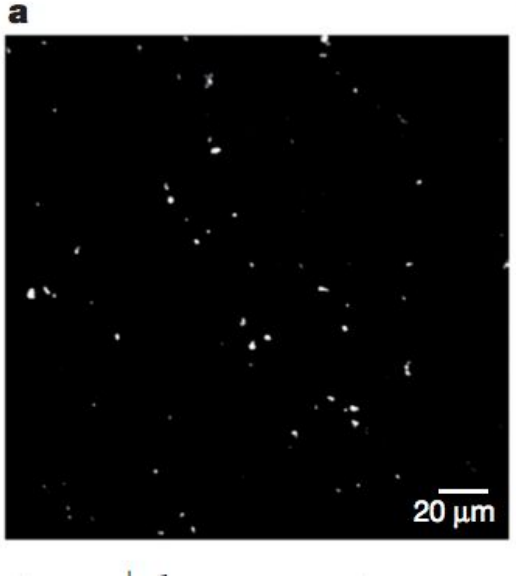


Figure 2 from Paper

Linear protein-DNA nanowire

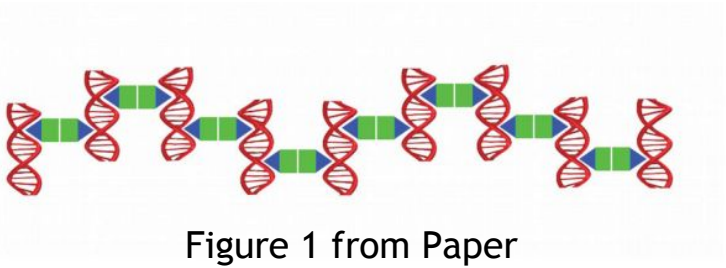


Figure 1 from Paper

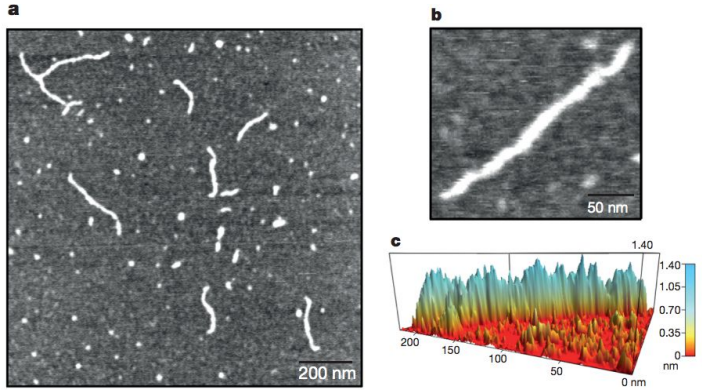


Figure 2 from Paper

Multiple Configurations

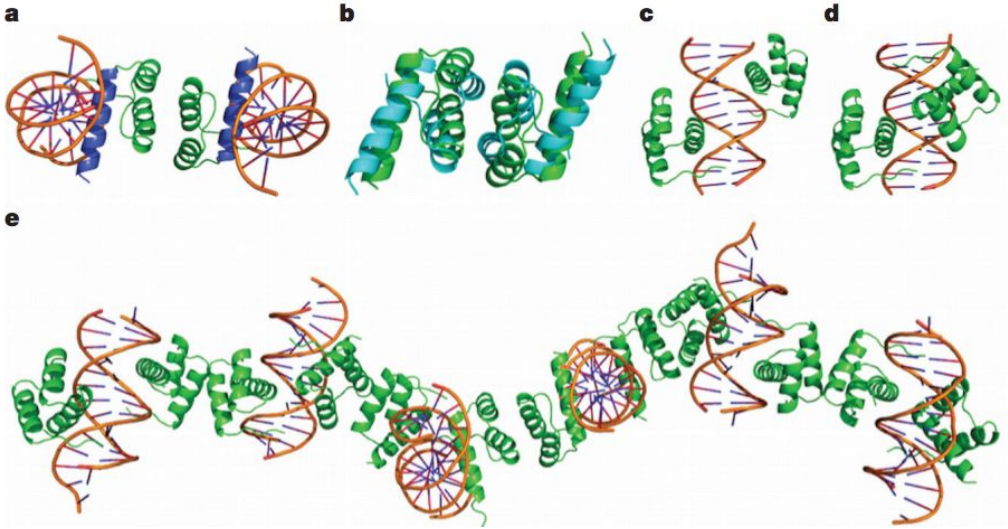


Figure 4 from Paper

Summary and Future Work

- Created a protein-DNA nanowire that self-assemblies solely on non-covalent interactions
- Used previously known computational methods to create a protein homodimer
- Unexpected result: multiple configurations!
- Next Steps
  - DNA origami & aptamers
  - DualENH fused to peptide tags

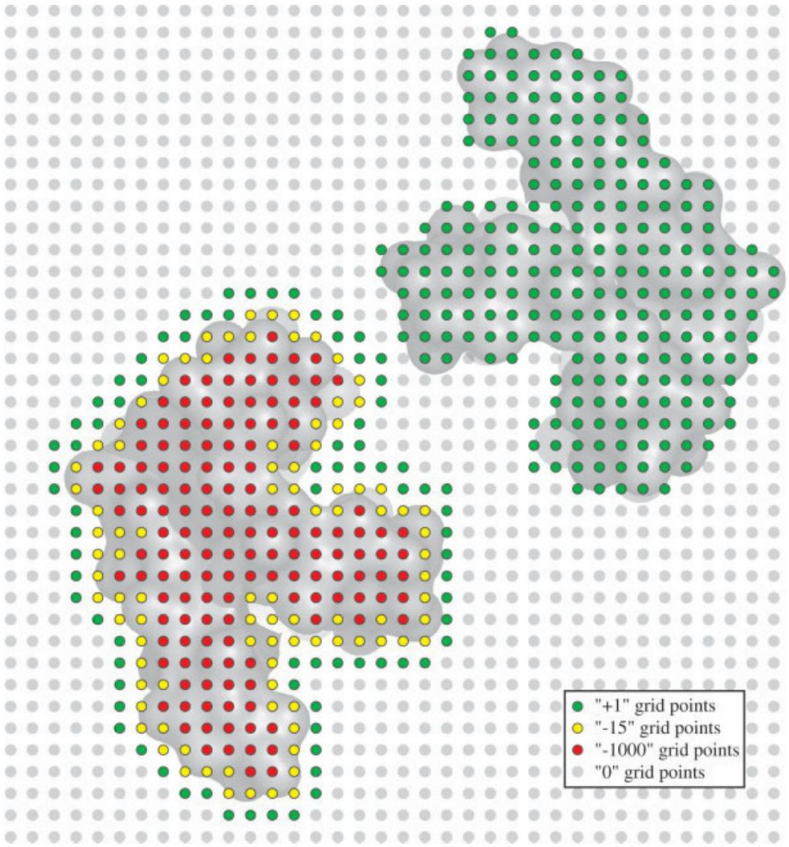
Strengths

- Composition of Paper
  - Clearly written
- Breakthrough in the creation of protein-DNA self-assembly
- Techniques could be applied broadly

Limitations

- Composition of Paper
  - Little focus on computation
- How does this improve hybrid assemblies?
- Multiple Configurations
  - Affect Structure
- Generalizability

Questions?


"+1" grid points
"-15" grid points
"-1000" grid points
"0" grid points

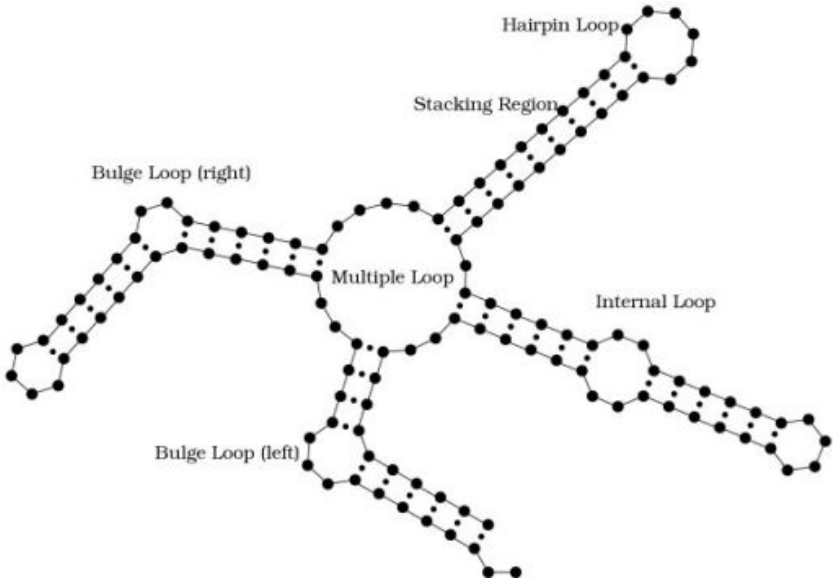
Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots

Christine E. Hajdin[a], Stanislav Bellaousov[b], Wayne Huggins[a], Christopher W. Leonard[a] , David H. Mathews[b], and Kevin M. Weeks[a]

a Department of Chemistry, University of North Carolina, Chapel Hill, NC 27599-3290
b Department of Biochemistry and Biophysics, and Center for RNA Biology, University of Rochester Medical Center, Rochester, NY 14642

RNA Structure Prediction

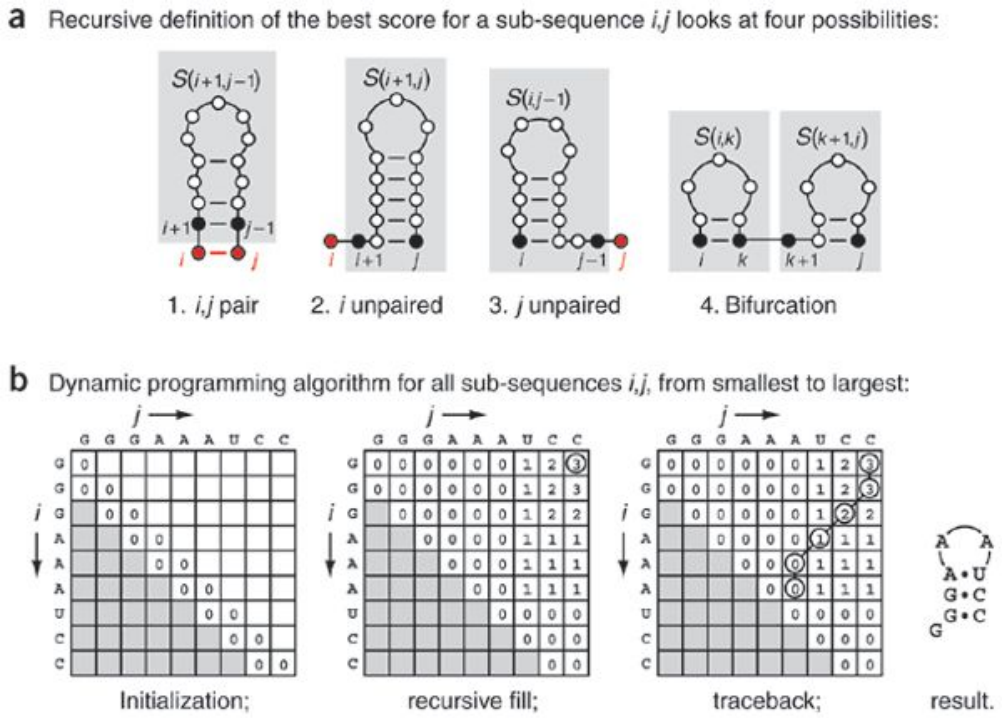
- **Motivation:** Essential to understand RNA's ability to form stable secondary structures (important for gene expression)



NIH

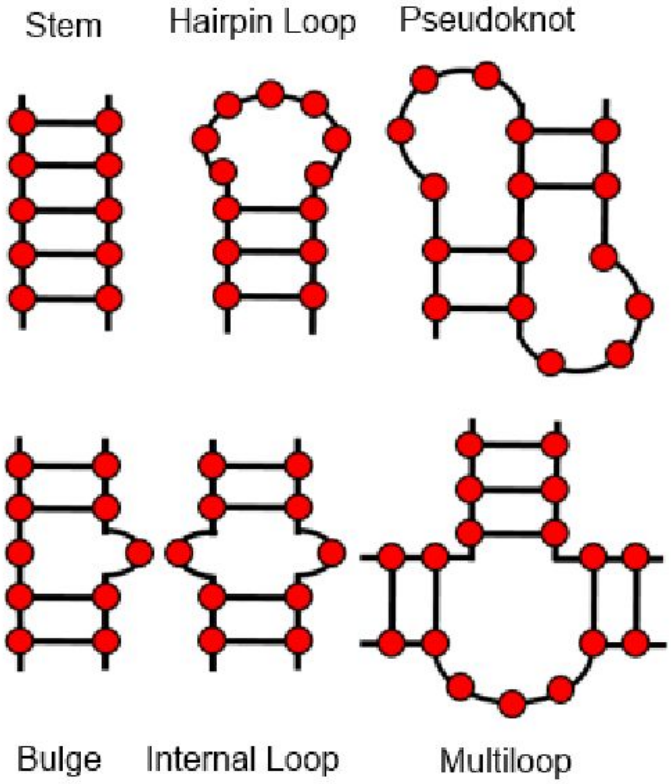
RNA Structure Prediction

- Can be computationally challenging for complex structures
- Can we also predict more complex structures, **namely pseudoknots?**



Nature

What is a pseudoknot?



Research Gate

Why do we care about pseudoknots?

- Often occur in regions of RNA which are essential to function
  - Large catalytic RNA’s
  - Most **riboswitches**
    - Regions of mRNA which regulate gene expression and are essential for ligand binding
  - Inside regulatory elements of **viruses**, which they use to break down a host cell’s metabolism

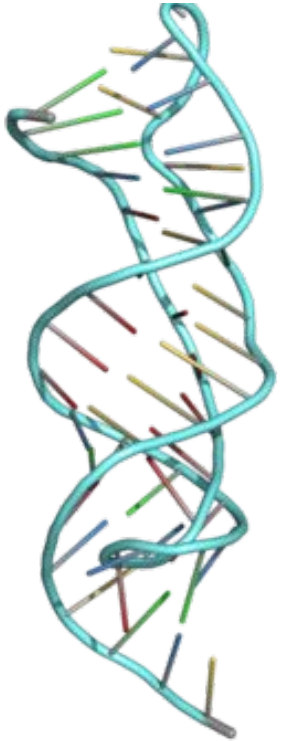


Figure from paper

RNA Prediction with Pseudoknots

- Pseudoknots are often left out of RNA prediction algorithms
  - Challenging to incorporate pseudoknots into algorithms currently used because they don’t easily fall into the DP framework, can we use lowest free energy models?
  - Finding the lowest free energy structure with pseudoknots has been shown to be an **NP-Complete** problem (meaning it cannot be solved in polynomial time as a function of its length)
  - When used, they tend to increase false positives, which require a lot more time to analyze

RNA Prediction with Pseudoknots

- Pseudoknot prediction is challenging
  - Energy models extrapolate from experimental data, and there are few containing pseudoknots
  - The stability of pseudoknots is not fully understood, making it hard to generate energy models

Previous Work using SHAPE (Selective 2′-hydroxyl acylation analyzed by primer extension)

- Probing technique used to determine stability of local nucleotides
  - Measures nucleotide flexibility, and is inversely correlated to base-pairing
    - I.e. the higher the SHAPE reactivity, the less likely it is to pair
  - Find SHAPE reactivities and use these as free energy terms and add onto the DP algorithms in place

$$\Delta G^{\circ}_{\mathrm{SHAPE}} = m\ ln[\mathrm{SHAPE} + 1] + b.$$

- Doesn’t take pseudoknots into account

Using SHAPE to make ShapeKnots

- Modify the equation to include the entropic likelihood of pseudoknots being formed
- Use the idea that energetically favorable pseudoknots have small numbers of the following:
  - Single stranded nucleotides
  - In-line helices
  - Nested helices

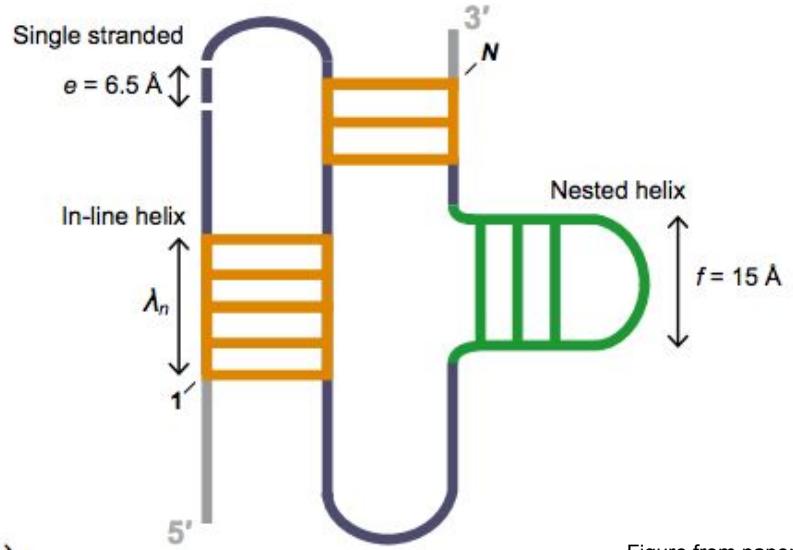


Figure from paper

$$\Delta G^{\circ}_{\mathrm{PK}} = P1\ ln\left(e^2\ SS + f^2\ NE\right) + P2\ ln\ \Sigma IL(n)\left(\lambda_n^2\right),$$

Experiments

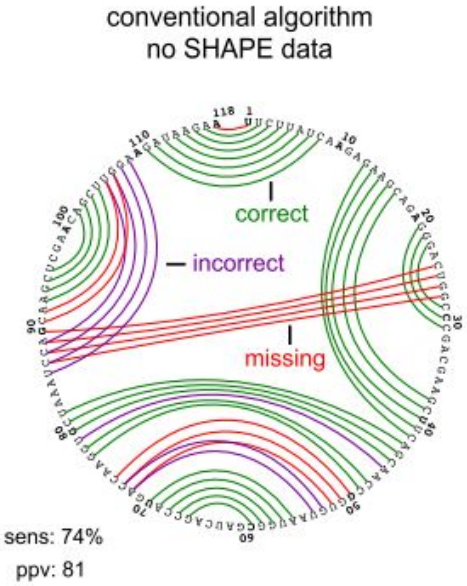
- Training set of 16 examples (pseudoknotted and non-pseudoknotted)
  - Riboswitches, long RNA strands, RNA’s which were poorly predicted by previous algorithms

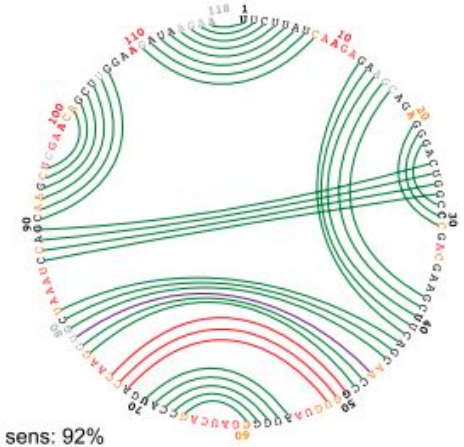
- Test Set of 6 examples (pseudoknotted and non-pseudoknotted)

Evaluation Metrics

- Sensitivity
  - “Fraction of base pairs in the the accepted structure predicted correctly”
- Positive Predictive Value (PPV)
  - “Fraction of predicted pairs that occur in the accepted structure”

Results





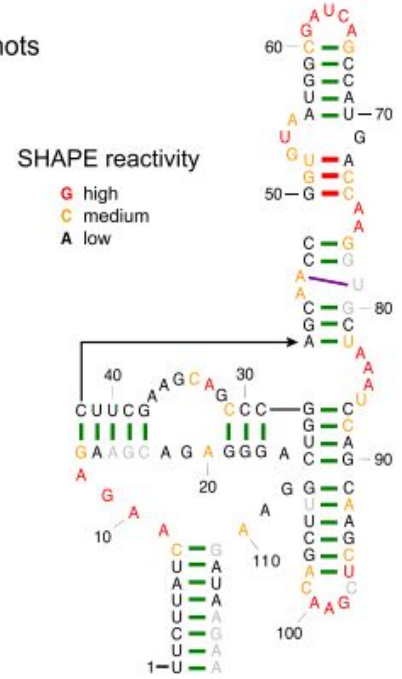


Figure from paper

Results

93% average sensitivity when using ShapeKnots (up from ~72%)

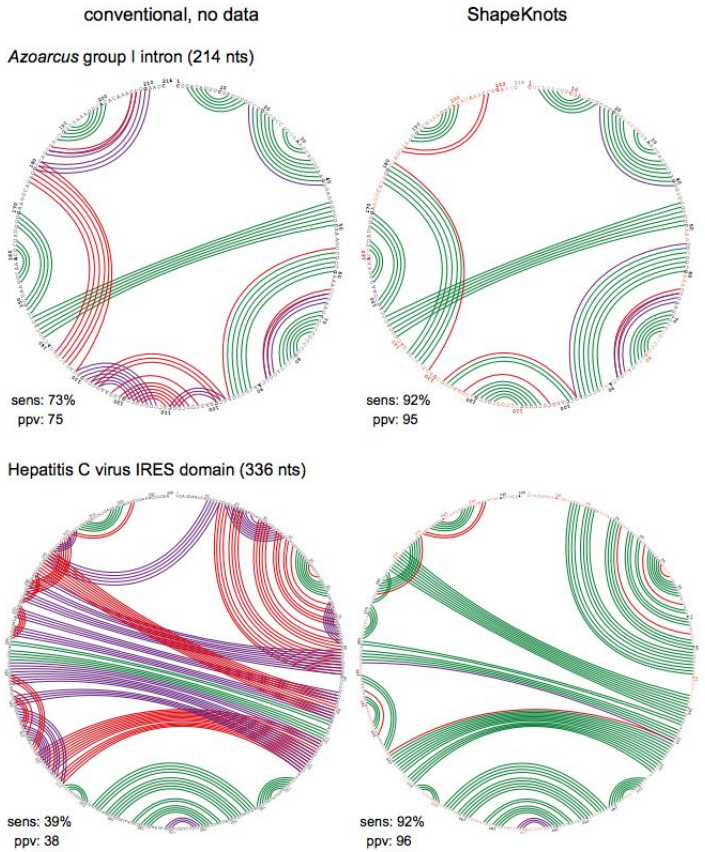


Fig. 4. Prediction summaries for two large, pseudoknot-containing RNAs. Structural annotations are as described in Fig. 2.

Figure from paper

Strengths

- Innovative approach on incorporating pseudoknots, hasn’t been done by other papers
- Looked at a diverse set of RNA with complex structures and evaluate specific examples in each category
- Discuss the challenges in RNA folding and how they affect results

Critiques

- Paper was intended for an expert in the field; didn’t give a lot of background on RNA prediction nor on pseudoknots.

- How can we scale these methods?

- Lack of future work they tend to do
    - Only looked at a small subset of examples, how do they plan on expanding this, will their methods work for other types of RNA structures?

Questions?

Thank you!

Experiments

- Short Pseudoknotted RNAs
- Large, Complex RNAs
- RNAs with Difficult to Predict Pseudoknots
- RNAs That Do Not Adopt Their Accepted Structures