Approaching 3D RNA Structure Prediction

Peter Kerpedjiev

TBI, University of Vienna

February 18, 2012

How do we predict 3D Structures

How do we predict 3D Structures

Start with a structure and repeat the following lots of times.

How do we predict 3D Structures

Start with a structure and repeat the following lots of times.

1. Change the structure slightly

How do we predict 3D Structures

Start with a structure and repeat the following lots of times.

1. Change the structure slightly
2. Evaluate its quality (Energy)

How do we predict 3D Structures

Start with a structure and repeat the following lots of times.

1. Change the structure slightly
2. Evaluate its quality (Energy)
3. Decide if we like it (Metropolis criterion)
   - If yes, keep it
   - If no, reverse the change from step 1
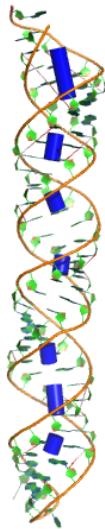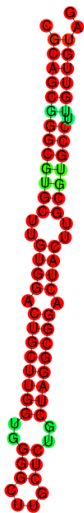4. Go back to step 1

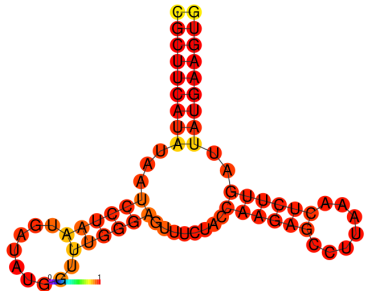# Exercise 1

What does this RNA look like in 3D?
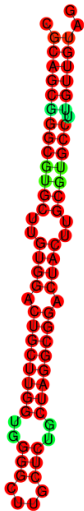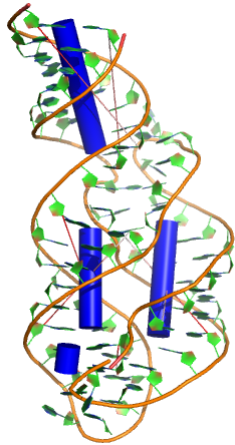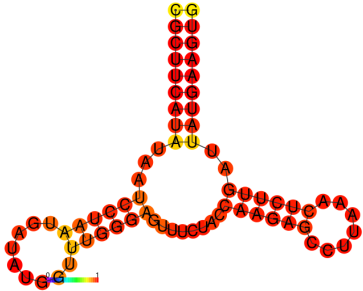
Exercise 1

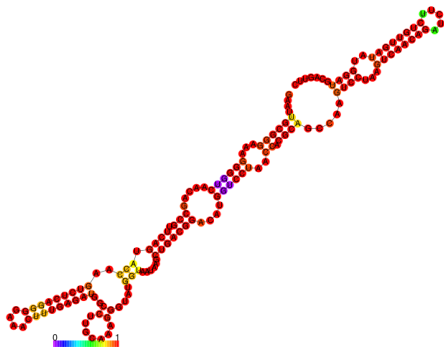What does this RNA look like in 3D?

Exercise 2

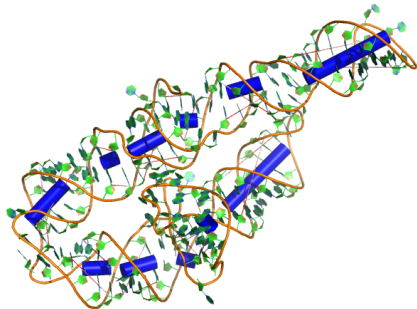What about this one?

Exercise 2

What about this one?

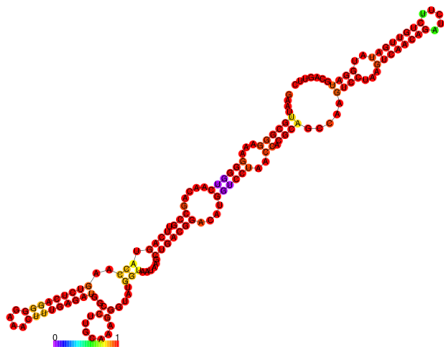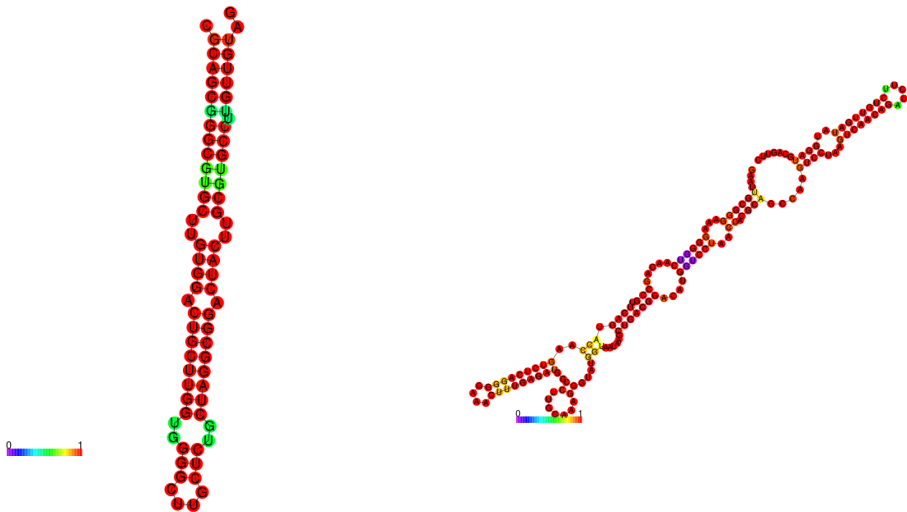# Exercise 3

And finally... this one?

Exercise 3

And finally... this one?
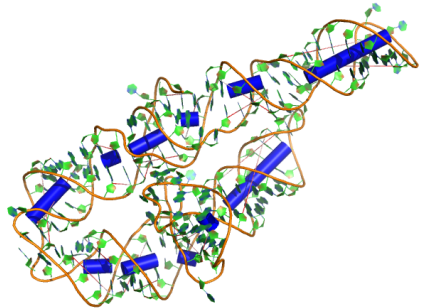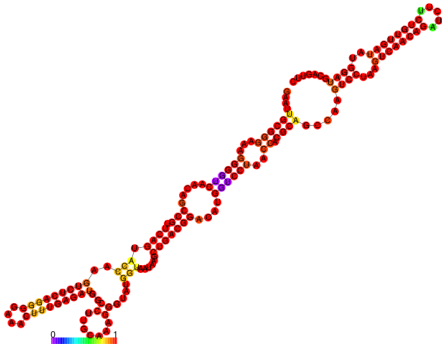
Key Features

Why do these two molecules fold so differently?

Key Features

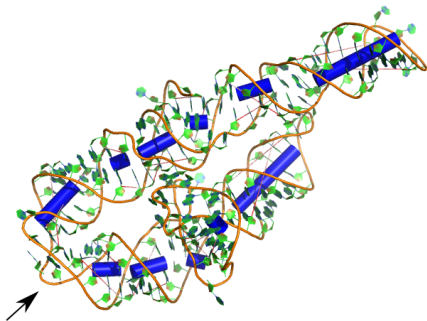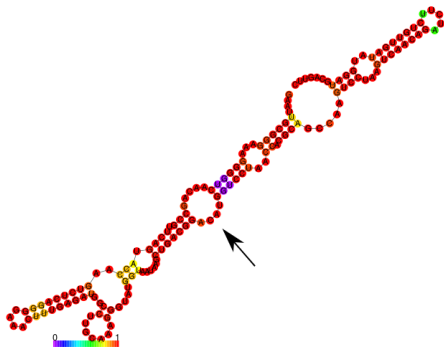Small elements lead to large changes in the 3D structure

-
-

Key Features

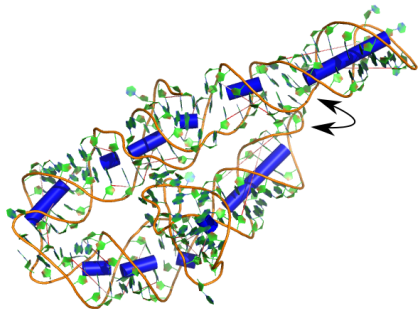Small elements lead to large changes in the 3D structure

- Bending
-

Key Features

Small elements lead to large changes in the 3D structure

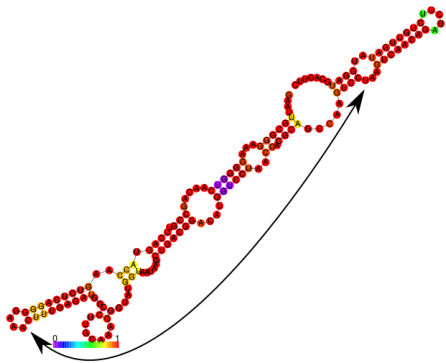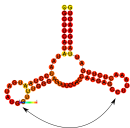- Bending
- Long range interactions

Are These Features Useful in 3D Modelling

Long range interactions inferred from mutate and map experiments [1]

Secondary Structure

Native Structure (1Y26)

[1] Kladwang et al. - Nature Chemistry - 2011 - A two-dimensional mutate-and-map strategy for non-coding RNA structure

Are These Features Useful in 3D Modelling

Long range interactions inferred from mutate and map experiments [1]

Secondary Structure

Native Structure (1Y26)

No long-range constraints (21.6 Å RMSD)

[1]Kladwang et al. - Nature Chemistry - 2011 - A two-dimensional mutate-and-map strategy for non-coding RNA structure
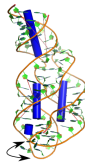
Are These Features Useful in 3D Modelling

Long range interactions inferred from mutate and map experiments [1]
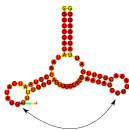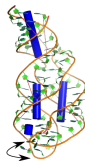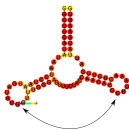

Secondary Structure

Native Structure (1Y26)

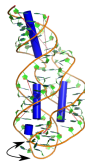No long-range constraints (21.6 Å RMSD)
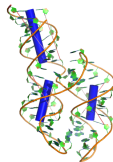
Long-range constraints (8.3 Å RMSD)

[1] Kladwang et al. - Nature Chemistry - 2011 - A two-dimensional mutate-and-map strategy for non-coding RNA structure
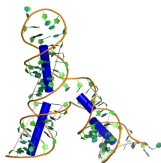
What about the bends?
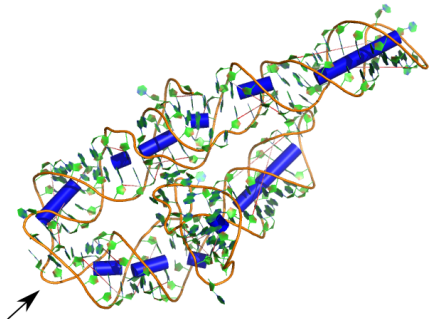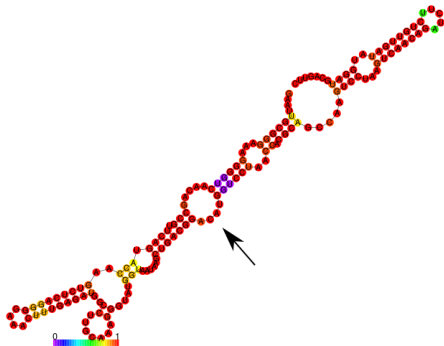
Are These Features Useful in 3D Modelling

What about the bends?
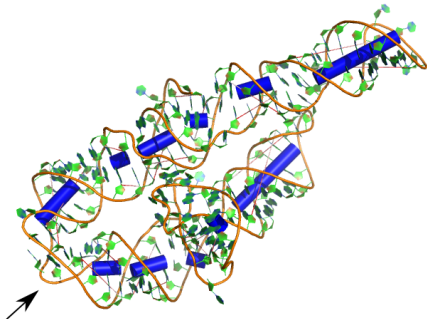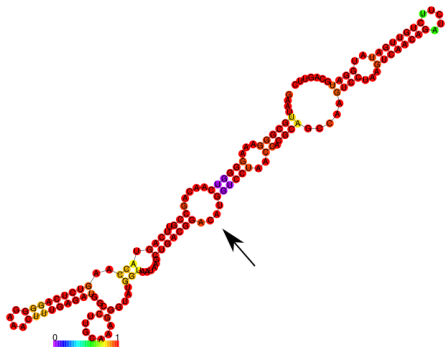
- Where are they likely to occur?

Are These Features Useful in 3D Modelling

What about the bends?
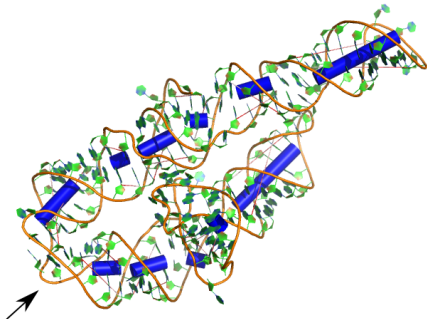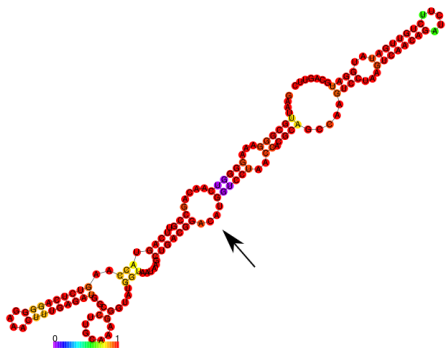
- Where are they likely to occur?
- How do we characterize them?

Bend/Kink Modelling

Where are bends/kinks likely to occur?

How do we characterize them?

Bend/Kink Modelling
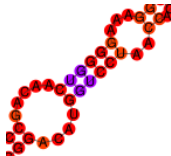
Where are bends/kinks likely to
occur?

- In bulge regions

How do we characterize them?

Bend/Kink Modelling

Where are bends/kinks likely to
occur?

- In bulge regions
- In branching regions

How do we characterize them?
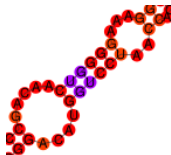
Bend/Kink Modelling

Where are bends/kinks likely to occur?

- In bulge regions
- In branching regions

How do we characterize them?

- Distance Constraints
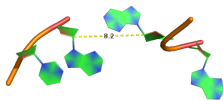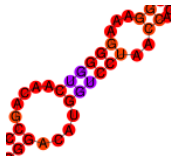
Bend/Kink Modelling

Where are bends/kinks likely to occur?

- In bulge regions
- In branching regions

How do we characterize them?

- Distance Constraints
- Non-canonical Base Pairs
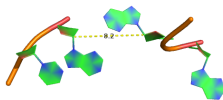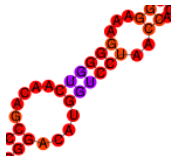
What is a distance constraint?

What is a distance constraint?

- It's a modification of the energy function.
- High energy when two atoms are not the ideal distance apart
- Low energy when the are the ideal distance

Constraint Distance

Energy Contribution
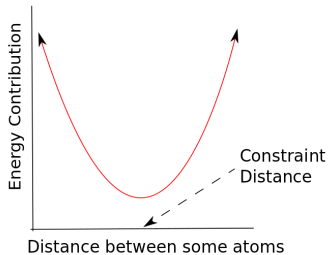
Distance between some atoms

Problem

Which non-canonical base pairs do we include? How exactly do we include them?

Solution

Find a common constraint among all particular base pair types. For example, the distribution of distances between particular nucleotide atoms.
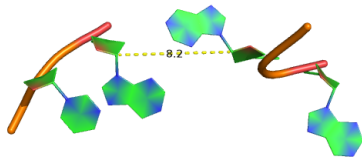
Non-canonical Base Pair Distance Distribution

Type	μ	σ	%
CG Ww/Ww	10.60	0.17	45%
AU Ww/Ww	10.45	0.22	43%
GU Ww/Ww	10.41	0.19	5%
AG Hh/Ss	9.37	0.27	5%
AG Ss/Ss	8.17	0.17	2%
AU Hh/Ws	9.60	0.15	2%

The distribution of C1*-C1* distances in various base pair types (as annotated by MC-Annotate):

Non-canonical Base Pair Distance Distribution

Type	μ	σ	%
CG Ww/Ww	10.60	0.17	45%
AU Ww/Ww	10.45	0.22	43%
GU Ww/Ww	10.41	0.19	5%
AG Hh/Ss	9.37	0.27	5%
AG Ss/Ss	8.17	0.17	2%
AU Hh/Ws	9.60	0.15	2%

The distribution of C1*-C1* distances in various base pair types (as annotated by MC-Annotate):
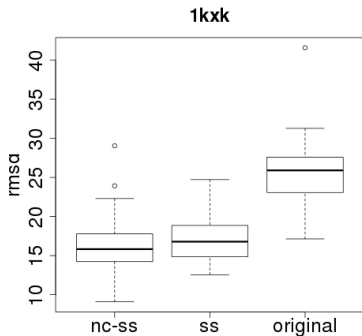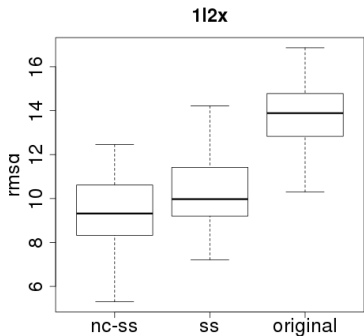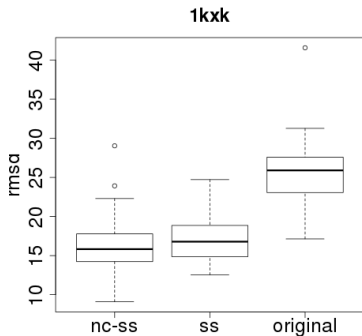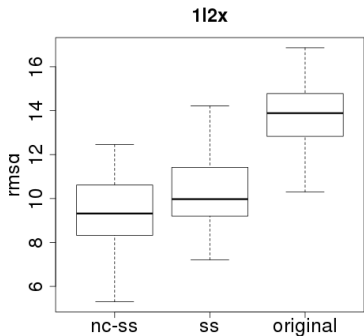
Results

Adding the non-canonical base pair constraints lowers the mean and minimum rmsd of sampled structures.

Results

Adding the non-canonical base pair constraints lowers the mean and minimum rmsd of sampled structures.

Next: Use predicted non-canonical interactions, along with a more diverse test set.

Results

Using real non-canonical base pair constraints improves structure prediction.

1L2X

Information Provided	Minimum RMSD
Sequence	10.3
+ Secondary Structure	7.2
+ Noncanonical Bases	5.3

1KXK

Information Provided	Minimum RMSD
Sequence	17.1
+ Secondary Structure	12.5
+ Noncanonical Bases	9.1

Secondary Structure Prediction Failure
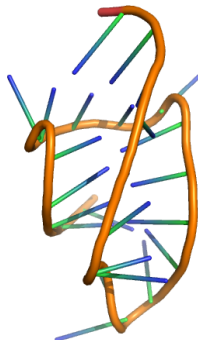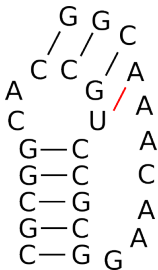
Q: Why is the red base pair predicted but not actually present?

Secondary Structure Prediction Failure
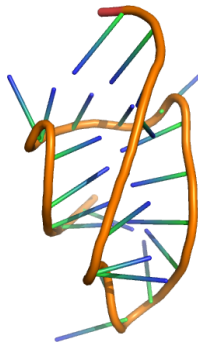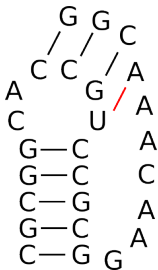
Q: Why is the red base pair predicted but not actually present?

A: Steric hinderance and non-canonical interactions.

Possible Solution: Modified Constraints As Bonuses

- Give bonuses for correct base pairs.

- Don't penalize incorrect base pairs.

- Potentially vary bonus as a function of base pair probability.
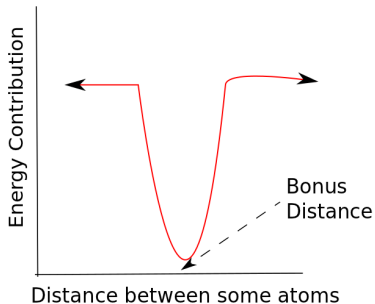
Application

Real
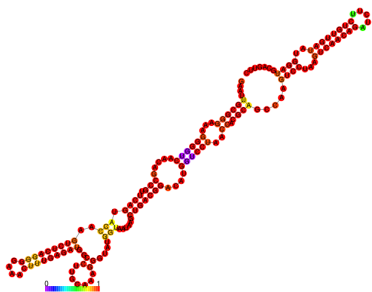
Predicted

Ideally, the major energy contribution will be from the large groups of high confidence base pair prediction bonuses.

Current Methods

MC-Fold | **MC-Sym**: Sequence $\rightarrow$ Secondary Structure $\rightarrow$ 3D Structure (using nucleic cyclic motifs) [2]

Current Methods

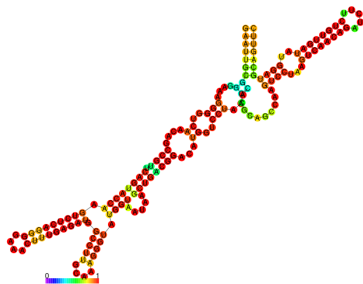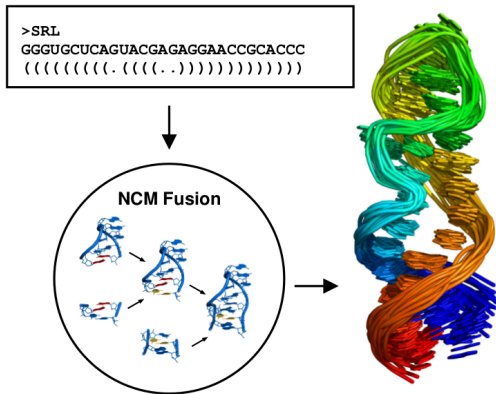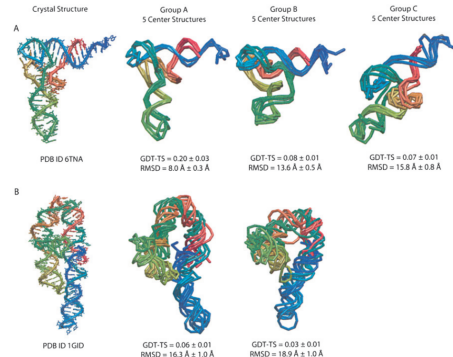**NAST**: Nucleic Acid Simulation Tool [3]

- Coarse grained model using knowledge-based potential
- Works reasonably well when long range tertiary interactions are known

[3]Jonikas - 2009 - Bioinformatics - Coarse-grained modelling of large RNA molecules with knowledge-based potentials and structural filters

Current Methods

FARNA / FARFAR: **F**ragment **A**ssembly of **RNA** / **F**ragment **A**ssembly of **R**NA with **F**ull-**A**tom **R**efinement

- Based on the popular Rosetta protein modelling tool
- Very good at predicting small RNA structures
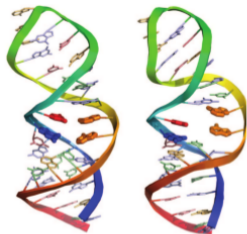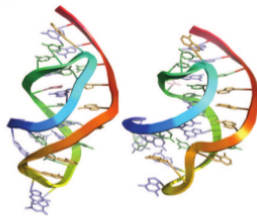
1ESY   1L2X

Summary

- RNA Structure Prediction tools exist

Summary

- RNA Structure Prediction tools exist
- None of them are very good

Summary

- RNA Structure Prediction tools exist
- None of them are very good
- Non-canonical constraints improve structure prediction

Summary

- RNA Structure Prediction tools exist
- None of them are very good
- Non-canonical constraints improve structure prediction
- Sub optimal base pair bonuses will be tried in the near future

Acknowledgements

You!

- Ivo Hofacker
- Xtof
- Everybody at the TBI