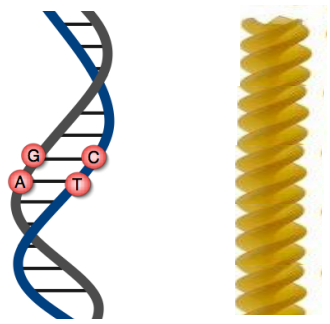
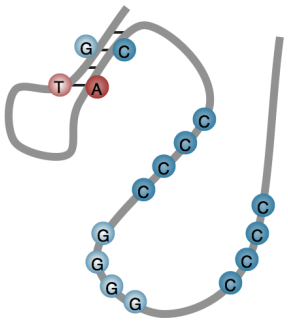
Nucleic Acid Secondary Structure Prediction and Design

Anne Condon
The Department of Computer Science
U. British Columbia

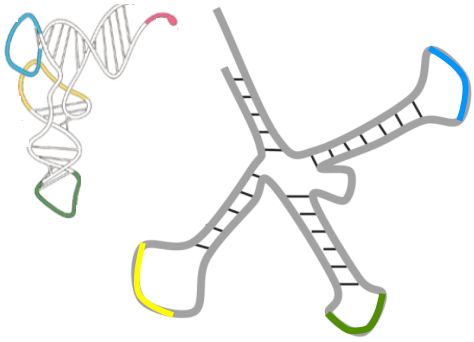
nucleic acid secondary structure



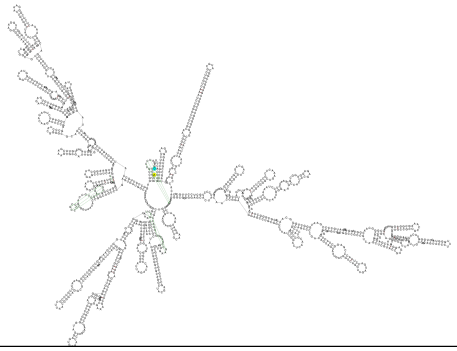
nucleic acid secondary structure



nucleic acid secondary structure

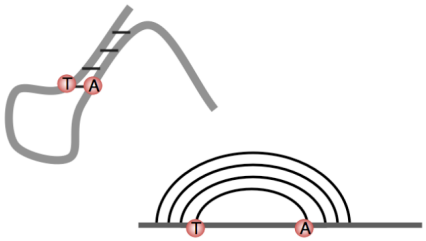


nucleic acid secondary structure



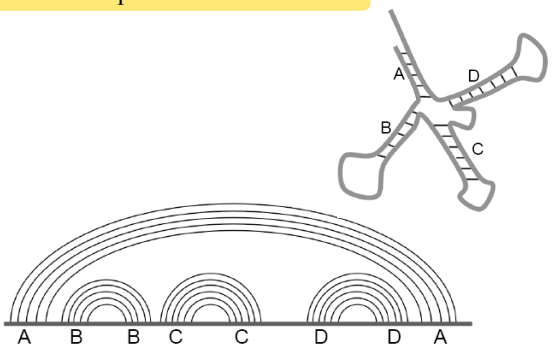
nucleic acid secondary structure

structure representation



nucleic acid secondary structure

structure representation: exercise



why predict secondary structure?

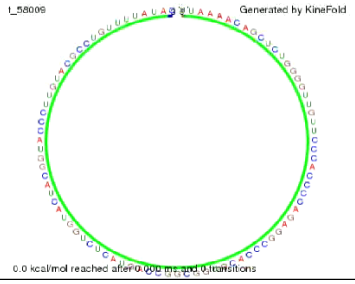
- to determine RNA 3D structure and ultimately function
- to establish phylogenetic relationships among organisms, via better RNA sequence alignment
- to understand gene regulation
- … and much more

research question

- given a sequence, *predict* its secondary structure

predicting structure

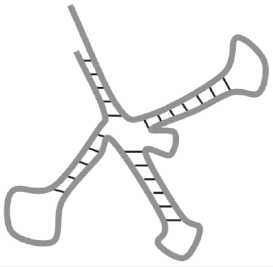
- structure formation is a (thermodynamically driven) process, whereby the strand seeks a minimum free energy configuration



predicting structure

simple energy model

- each *base pair* contributes a (negative) score
- sum the scores to get the total energy



predicting structure

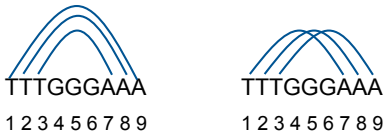
precise problem formulation

- given an RNA sequence, find the *minimum free energy (MFE)* structure, relative to the simple energy model

predicting structure

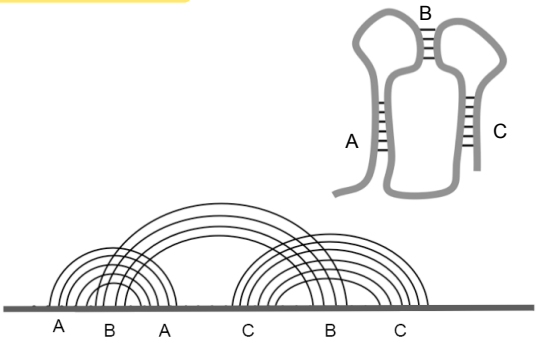
exercise

- draw two MFE secondary structures for the sequence



predicting structure

a pseudoknot



predicting structure

revised problem formulation

- given an RNA sequence, find the *minimum free energy (MFE)* ***pseudoknot free structure***, relative to the simple energy model
- pseudoknot free: no arcs cross in the arc diagram representation of the structure

predicting structure

a "brute force" algorithm

- enumerate all possible pseudoknot free secondary structures for the given input sequence
- for each, calculate its free energy
- keep track of structure with lowest free energy enumerated, so far, and output this at the end
- exercise: can you find a lower bound on the number of distinct pseudoknot free structures a sequence could have?

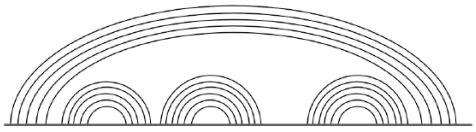
predicting structure

a "brute force" algorithm

- the number of pseudoknot free secondary structures consistent with an RNA or DNA sequence may be exponential in its length ☹

predicting structure

other ideas?



predicting structure

dynamic programming

- let W(*i*,*j*) be the energy of the MFE pseudoknot free structure from position *i* to position *j* of the sequence
- express W(*i*,*j*) in terms of "smaller" W()'s

predicting structure

dynamic programming

$$W(i,j) = \min \begin{cases} -1 + W(i+1,j-1), & \text{if i and j can pair} \\ \min_{i \le k < j} \{ W(i,k) + W(k+1,j) \}, & \text{otherwise} \end{cases}$$

predicting structure

dynamic programming

if i < j:

$$W(i,j) = \min \begin{cases} -1 + W(i+1,j-1), & \text{if i and j can pair} \\ \min_{i \le k < j} \{ W(i,k) + W(k+1,j) \}, & \text{otherwise} \end{cases}$$

$W(i,j) = 0$ if $i = j$

MFE approach: applying the recurrences

dynamic programming

- input: AGCUU

i \ j	1	2	3	4	5
1	0	0	1	2	2
2		0	1	1	1
3			0	0	0
4				0	0
5					0

W(*i,j*)

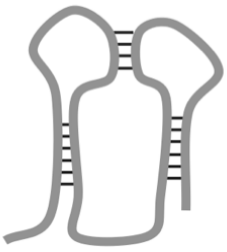
what RNA research do I do?

…in collaboration with many others at UBC, U. Rochester, Caltech, and elsewhere

what RNA research do I do?

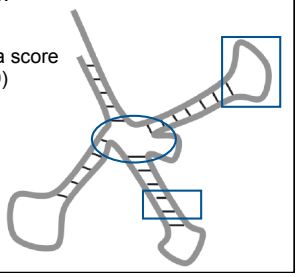
structure prediction

- design algorithms for prediction of certain types of pseudoknotted structures



what RNA research do I do?

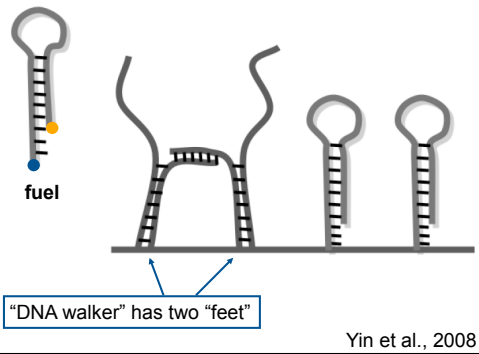
- apply machine learning to improve "loop based" energy models:
  - each *loop* contributes a score (Mathews et al., 1999)
  - sum the scores to get the total energy



what RNA research do I do?

- predict *folding pathways* and their properties and design folding pathways (or show that these problems are NP-hard)

what RNA research do I do?



Yin et al., 2008

the possibilities are endless…