RNA Design with Backbone k-Tree Model

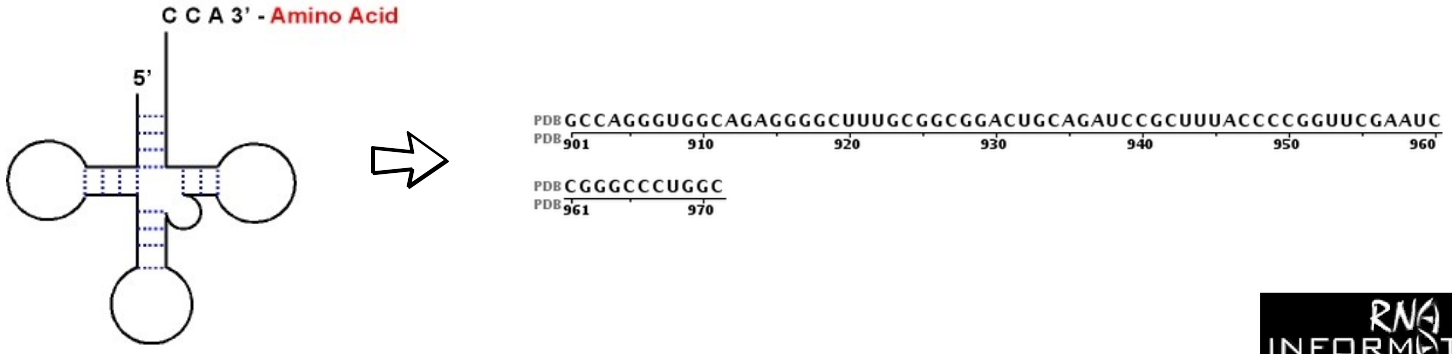
RNA Informatics Laboratory
The University of Georgia

RNA INFORMATICS

THE UNIVERSITY OF GEORGIA

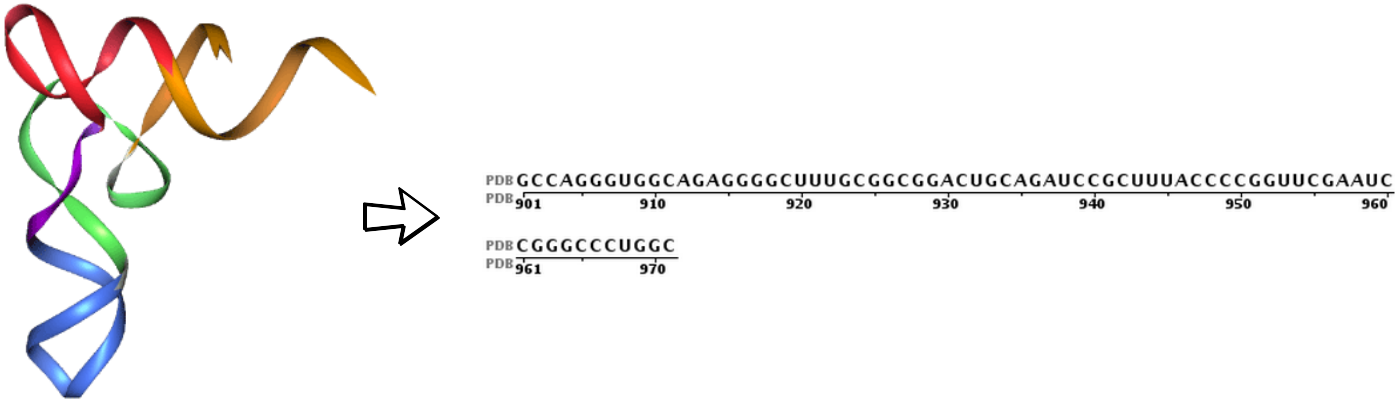
RNA Design/Inverse Folding

- Given a target structure, of determining one or more RNA sequences that fold into the structure.
- RNA engineering is offering new routes for deciphering genetic regulation and for developing therapeutic and bioengineering applications.
- Numerous tools for design of RNA secondary structures.
  - INFO-RNA(dynamic programming),
  - RNAinverse(ViennaRNA package),
  - RNA designer(stochastic local search), etc.



Our problem

- Input: coordinates of the backbone of a target RNA 3D structure.
- Output: one or more sequences that fold into the target structure.



RNA-Redesign

- The only web server for designing 3D structure.
- To find sequence variants compatible with a given backbone structure.
- Input is a pdb file with backbone and sugar heavy atoms.
- Rosetta packer algorithm, to sample the base identity and side-chain torsions, is well-developed for protein design.
- Optimized based on 3 routines: glycosidic torsion angles, placement of 2'-OH hydrogen, side-chain conformation and identity.
- Minutes for 20 nts.

RNA-Redesign Results

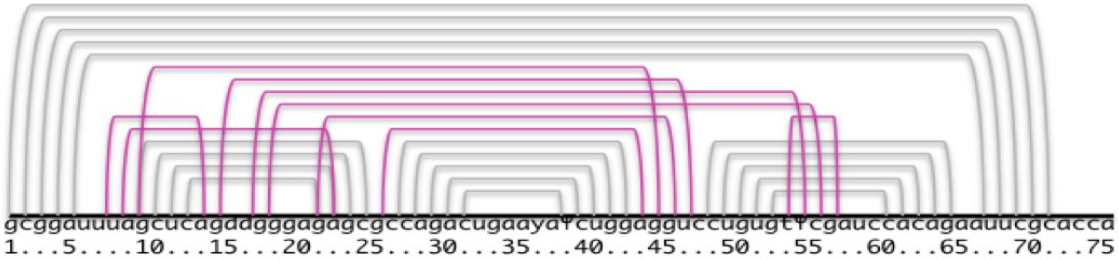
- Singal recognition particle domain IV RNA (PDB ID: 1LNT).
  - One of its conserved A-C base pair could be mutated to an A-G pair, which yields a lower Rosetta score.

- J4/5 motif from the P4-P6 domain of gourp I Tetrahymena ribozyme (PDB ID: 1GID), which forms contacts in the core.
  - A conserved G-U pair above an unpaired A and two sheared A-A pairs. Three pairs form contacts between P4-P6 domain and the full ribozyme's catalyic core.
  - Mutations of A-A pairs to A-U and A-C pairs, suggesting that A-A pairs are not necessary for the structure, but for the structure roles.

- A GGAA tetraloop. Several mutations do not affect the overall folding.

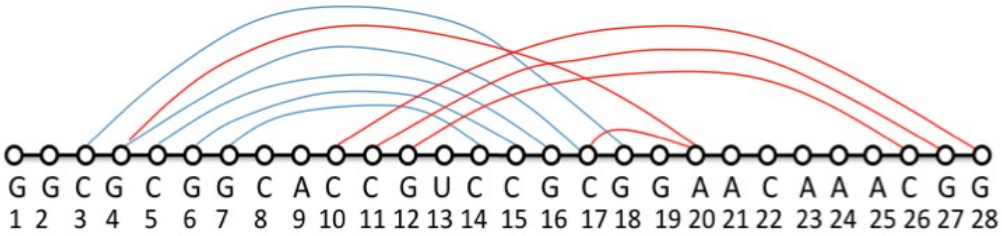
BkTree-Design

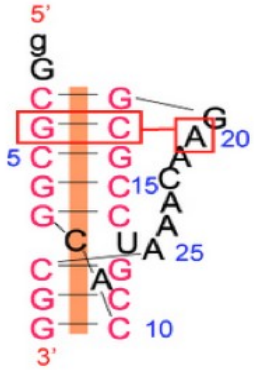
- Preprocessing input coordinates to generate a backbone graph:
  - To determine a non-backbone edge: choose a cutoff C1-C1 distance: 17.5 angstrom(avg+2*std)
  - Merge close non-backbone edges, e.g., in 2DU3,
  - (0 63), (0 64), (0 65), (0 67), (0 69), (0 70) become (0, 63); (8 12), (8 13), (8 14), (8 15), (8 19), (8 20), (8 21), (8 22), (8 23), (8 24), (8 25), (8 26), (8 41), (8 42), (8 43), (8 44), (8 45), (8 46), (8 47), (8 57), (8 58) become (8 13), (8 20), (8 45), (8 57).

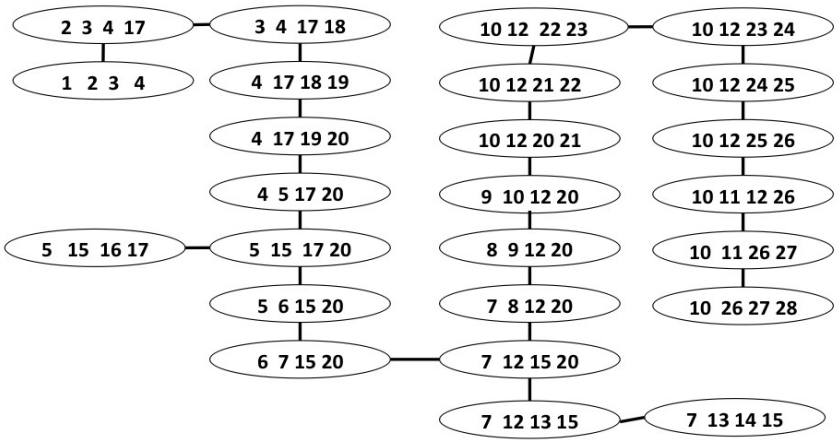


BkTree-Design

- Run BkTree program to generate a backbone 3-tree for the backbone graph.







RNA INFORMATICS
THE UNIVERSITY OF GEORGIA

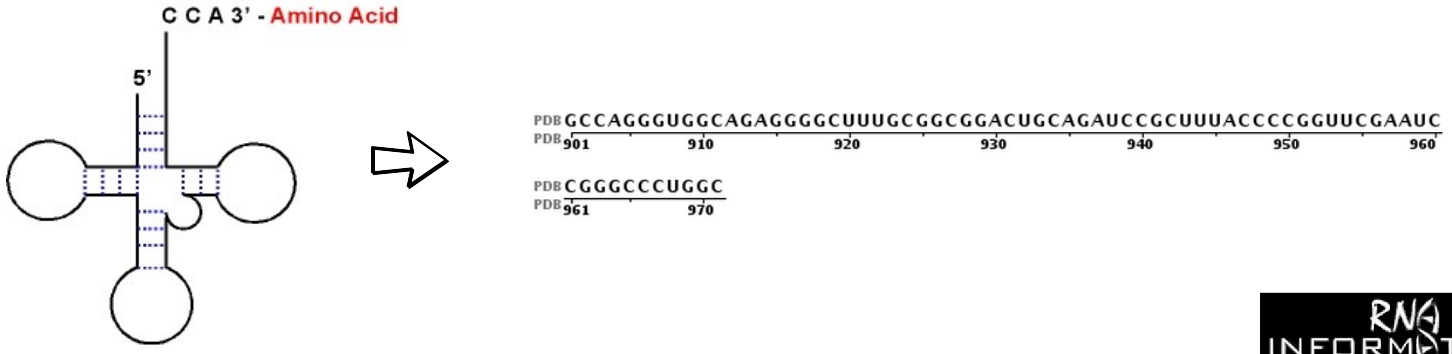
BkTree-Design

- Build a pattern database for every 4 nucleotides (form a 4-clique of a backbone 3-tree) from resolved 3D structures of RNAs.

- [10, 10, 10, 19, 20, 5][4, 4, 1]5 :
  UGGG['', '', '', '', '', 's35'] 1 ;
  AGUC['', '', '', '', '', 'perp'] 1 ;
  GACG['', 'cWW', 's55', '', '', 's35'] 1 ;
  CGCC['cWW', '', '', '', '', 's35'] 1 ;
  AAAG['', '', '', '', '', 's35'] 1 ;

RNA Design/Inverse Folding

- Given a target structure, of determining one or more RNA sequences that fold into the structure.
- RNA engineering is offering new routes for deciphering genetic regulation and for developing therapeutic and bioengineering applications.
- Numerous tools for design of RNA secondary structures.
  - INFO-RNA(dynamic programming),
  - RNAinverse(ViennaRNA package),
  - RNA designer(stochastic local search), etc.



BkTree-Design

- DP algorithm on the backbone 3-tree to calculate the best assembling of patterns of all 4-cliques in the 3-tree.
- Not only predict a sequence, but also predict a set of interactions.
- Produce a constant number of best sequences.

Problems

- Some 4-cliques cannot find the exact matching pattern.
  - One solution is to allow some similar patterns to match the 4-cliques.
- How to score each pattern? Geometric alignment of the target and its candidates?
- How to evaluate the predicted sequences?