

## Lecture 21, Phylogeny & RNA: Pfold (cont.)

12/6;

No note taker today. In a nutshell, Pfold uses a stochastic context free grammar (SCFG) to model (pseudoknot-free) RNA 2ary structure. E.g. the rule  $F \rightarrow dFd \mid X$  generates a nested set of base pairs (the d's), with whatever X generates at the end of that helix. Probabilities on the rules provide a prior on structures; e.g. average length of helices, etc. In Pfold, the terminal symbols (d above) are not individual letters, but *columns* in an alignment, and the probabilities of those columns come from the given phylogenetic tree and Felsenstein's algorithm. All this gives a generative framework for RNA 2ary structure alignments. As usual, we turn this into algorithms for scoring an alignment and predicting its structure by a Viterbi-like approach: among all parses consistent with the given input alignment (tree, evolutionary rate matrix, etc.) find the the one of maximum posterior probability. This can be done by a generalization of the Cocke-Kasami-Younger context-free grammar recognition algorithm. Validation presented in the papers is a bit thin, but empirically the method works well.