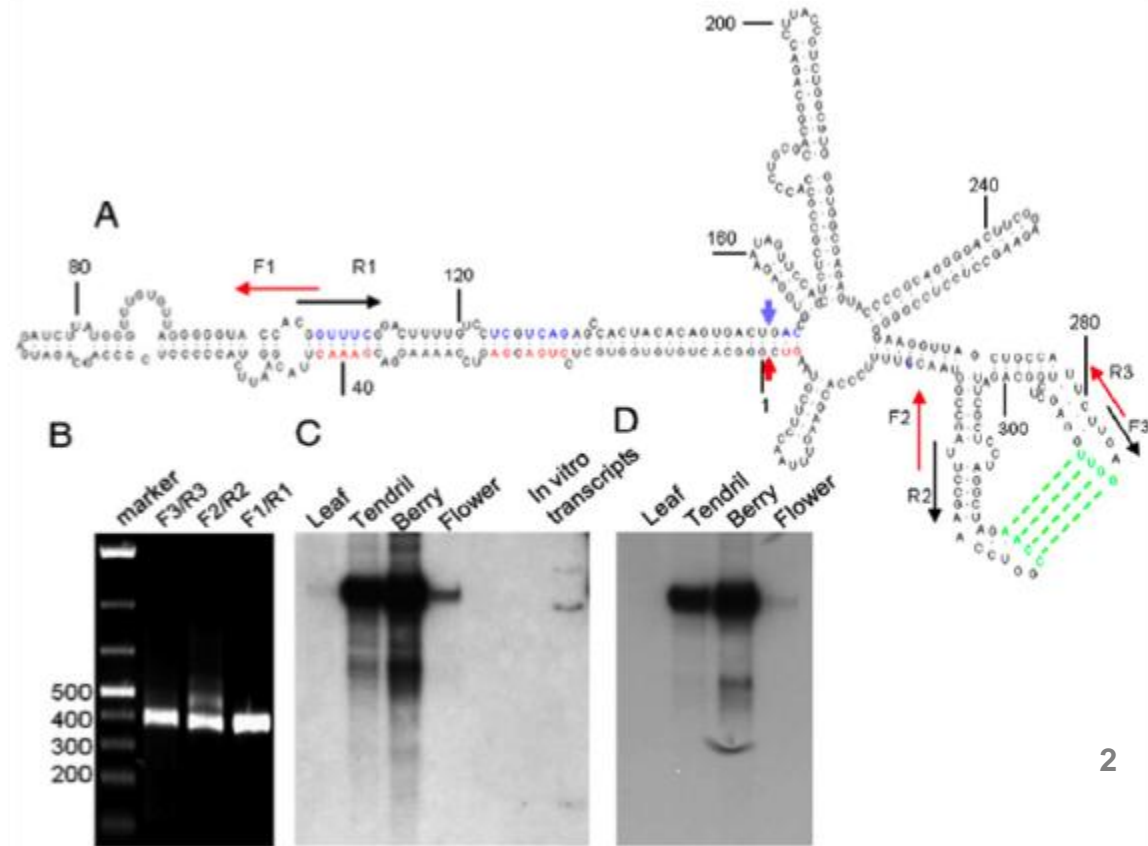


RNA二级结构预测

375nt 的环形类病毒GHVd 的 RNA二级结构预测和RT-PCR、Northern blot检测结果

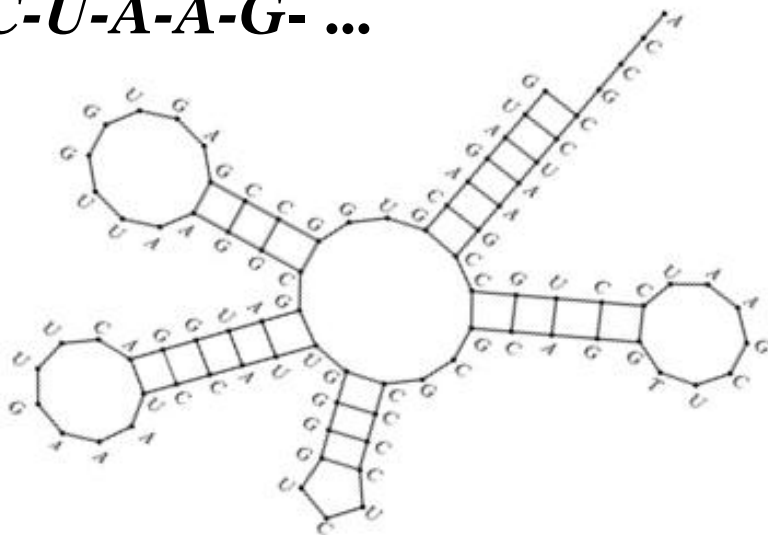


RNA二级结构

一级结构： 由字母 *A, C, G, U* 标记的核苷酸构成的一条链.

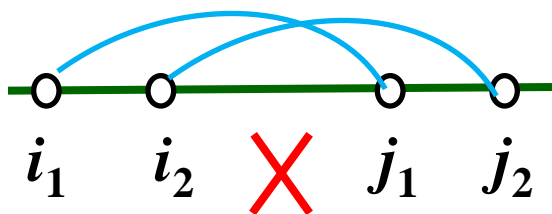
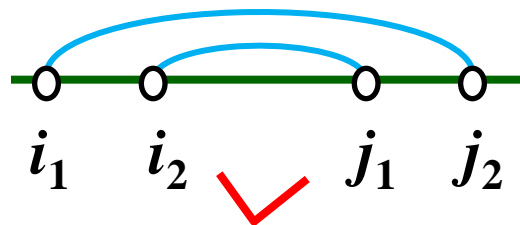
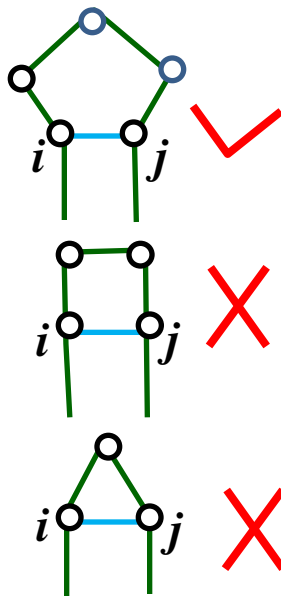
实例： *A-C-C-G-C-C-U-A-A-G-C-C-G-U-C-C-U-A-A-G- ...*

二级结构：
核苷酸相互
匹配构成的
平面结构

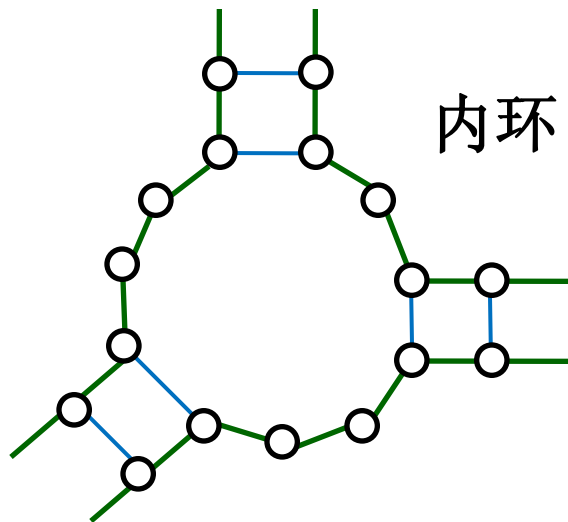
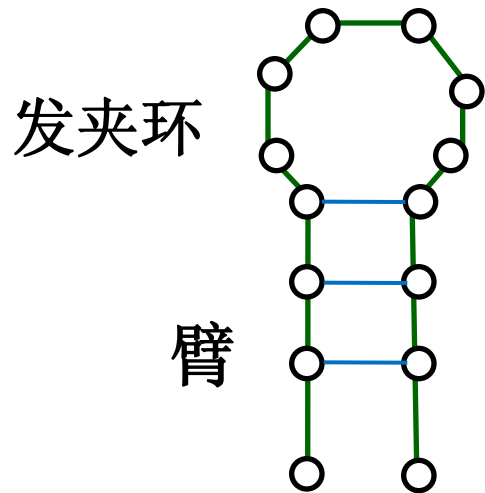


匹配原则

- 配对 $U-A$, $C-G$
- 末端不出现“尖角”，位置 $i-j$ 配对，则 $i \leq j-4$
- 每个核苷酸只能参加一个配对
- 不允许交叉，即如果位置 i_1, i_2, j_1, j_2 满足 $i_1 < i_2 < j_1 < j_2$ ，不允许 i_1-j_1, i_2-j_2 配对，但可以允许 i_1-j_2, i_2-j_1 配对。



匹配的结构



RNA二级结构问题

给定RNA的一条链（一级结构），预测它的可能的稳定的二级结构

稳定二级结构满足的条件

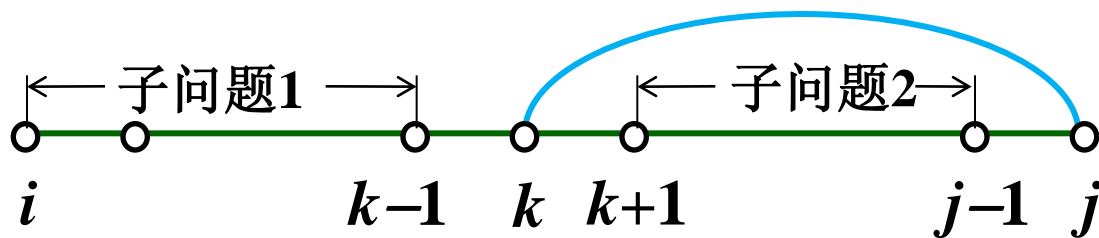
生物学条件：具有最小自由能

简化条件：具有最多的匹配对数

问题：给定RNA链，求具有最多匹配对数的二级结构，即最优结构。

建模

- 子问题界定：前边界 i , 后边界 j
- 若 j 与 k (所有可能) 位置匹配, 归约为
子问题 1: i 到 $k-1$ 的链
子问题 2: $k+1$ 到 $j-1$ 的链



- 若 j 不参与匹配, 则原问题归约为 i 到 $j-1$ 的子问题

优化函数的递推方程

令 $C[i,j]$ 是序列 $S[i..j]$ 的最大匹配对数

$$C[i,j] = \max\{C[i,j-1],$$
$$\max_{i \leq k \leq j-4} \{1 + C[i,k-1] + C[k+1,j-1]\}\}$$

$$1 \leq i, \quad j \leq n, \quad j - i \geq 4$$

$$C[i,j] = 0 \quad j - i < 4$$

满足优化原则

计算顺序：按照子问题长度计算

计算复杂度分析

子问题个数: i, j 对的组合有 $O(n^2)$ 个

对于给定的 i 和 j , j 需要考察与所有可能的 k 是否匹配, 其中 $i \leq k \leq j-4$, 需要 $O(n)$ 时间.

算法时间复杂度是 $O(n^3)$.

小结

- 划分子问题，确定子问题边界 i, j 与归约方法.
- 定义优化函数,列递推方程和初值.
- 自底向上计算，设计备忘录 (表格)
- 设立标记函数，记下最优划分位置
- 时间复杂度估计