

毕业论文的摘要

RNA 作为一种重要的生物大分子，在生命活动中具有关键性的作用。自 2000 年以来有关 RNA 的研究已经有 6 年被 *Science* 杂志列入世界十大科技进展，这充分说明了 RNA 的研究在当前生物学领域的核心地位。由于 RNA 的结构决定其功能，而 RNA 分子由于其本身的特性很难由实验方法测定其结构，于是面对海量的 RNA 序列，通过计算方法获得 RNA 的二级和三级结构就成为一项重要的任务。

自从 20 世纪 70 年代以来，RNA 二级结构预测就一直是计算生物学领域研究的热点问题。基于自由能最小原理来预测 RNA 二级结构的方法是目前被广泛接受和采用的方法。自 1981 年 Zuker 提出第一个自由能最小算法以来，随着自由能模型和参数，以及优化算法的改进，RNA 二级结构预测已经有了很大发展。但这个问题至今仍没有得到令人满意的解决，主要存在两个方面的问题：1) 在计算方面，算法复杂度高。特别是对于含假结的 RNA 二级结构预测，计算其最小自由能的问题已被证明是 NP 难问题，因而限制了算法在长序列 RNA 上的应用；2) 在自由能模型方面，目前普遍使用的最小自由能规则本身也存在缺陷，而且由于 RNA 二级结构折叠自身的动力学因素的影响，使得真正的 RNA 结构往往并不是给定自由能模型下全局最优的结构。

本文在分析了 RNA 二级结构预测目前所存在的问题和难点的基础上，提出了用缩小搜索空间来提高 RNA 二级结构预测准确度的新思路。在此基础上我们设计和开发了含假结的 RNA 二级结构预测软件—FlexStem。

FlexStem 仍然是基于自由能最小原理，它采用了目前被广泛使用的标准自由能模型和具有较强描述能力的假结模型。和传统的基于最小自由能原理的 RNA 二级结构预测算法不同的是，FlexStem 以一种模拟 RNA 二级结构形成的方式——茎叠加的方式来预测含假结的 RNA 二级结构，并通过启发式策略来有效地缩减茎的搜索空间以达到提高预测准确度的目的。这个启发式策略包含两个主要组成部分：其一是极大茎策略。我们采用极大茎来作为 RNA 二级结构的基本组成单元并在茎叠加过程中通过极大茎融合策略来不断累积茎。这一方式可以有效减少候选茎的数量从而大大减小茎组合空间。其二是在极大茎策略的基础上，我们通过对真实 RNA 二级结构的实验和分析进一步归纳出了茎叠加规律。基于这一规律，我们在不断累积茎的过程中，每次只需要在当前很小的候选茎范围中进行选择，而且在多数情况下我们只需要选择当前最优茎即可。我们通过极大茎策略和茎叠加规律构造出了一个被大大缩小的茎搜索空间，并且实验还证明了大多数情况下，真实的 RNA 二级结构就出现在这个局部搜索空间中。基于这些启发式策略我们开发了能够预测含假结的 RNA 二级结构的软件—FlexStem 并在目前通用的数据集上对它进行了全面测试。实验结果证明，FlexStem 对于含假结的 RNA 二级结构的预测性能要明显高于 PKNOTS, PknotsRG, HotKnots 以及 ILM 等目前国际上最著名的几种软件。

相比 RNA 二级结构预测，RNA 三级结构预测是一个更为困难的问题，目前国际还没有真正实用的算法或者系统。本文探索性地提出了用 **Threading** 方法来预测 RNA 三级结构的思想、建立了初步的整数规划模型，并对当前 RNA 三级结构预测面临的主要困难以及可能的解决方法进行了讨论。