

# RNA 分子折叠的统计力学\*

陈世杰<sup>†</sup> 谭志杰 曹 松 张文炳

( 密苏里大学物理与天文系和生物化学系 哥伦比亚 密苏里 65211 美国 )

**摘要** RNA 分子折叠对决定分子( 基因 ) 水平上的生命活动是至关重要的. 本文对 RNA 分子折叠( 特别是二级结构的折叠 ) 的热力学和动力学性质作一简单回顾与介绍. 因为静电相互作用在 RNA 折叠过程中的特殊重要作用, 我们对 RNA( 与 DNA ) 折叠的电相互作用单独作较为详细的讨论.

**关键词** RNA 折叠, 热力学, 动力学, 静电相互作用

## The Statistical Mechanics of RNA Folding

Shi-Jie CHEN<sup>†</sup> Zhijie TAN Song CAO Wenbing ZHANG

( Department of Physics and Astronomy and Department of Biochemistry, University of Missouri - Columbia, MO 65211, U. S. A. )

**Abstract** RNA folding is critical for the biological activities at the molecular( genetic ) level. In this article, we briefly review and introduce the thermodynamic and kinetic properties of RNA folding( especially the secondary structure folding ). Due to the particular importance of the electrostatic interactions in RNA folding, we also discuss the electrostatic interactions in RNA( DNA ) folding in a separate section.

**Keywords** RNA folding, thermodynamics, kinetics, electrostatic interactions

### 1 引言

RNA 是核糖核酸分子的简写. RNA 和 DNA、蛋白质一起构成了最基本的三大类生物分子. RNA 分子执行着从遗传信息的传递、储存到对蛋白质合成和其他生化过程的基因调控和催化的生物功能. RNA 分子的生物功能是直接与其结构和结构的变化相关的, 所以对 RNA 分子结构及其折叠过程的热力学与动力学的定量预测与分析对其功能的研究是至关重要的.

当一个 RNA 分子在细胞中合成后, 他最终将从一个相对展开的形态折叠成一个能执行其特定生物功能的非常紧凑的三维结构( 天然结构 ). 从物理上说, 这个结构正是对给定碱基序列下自由能最低的分 子构型. RNA 分子的结构与其折叠的物理化学性质是由其碱基序列决定的. 从碱基序列预测其自由能最低的结构即通常意义上的所谓的 RNA 折叠问

题. 然而, RNA 分子的功能往往是通过结构的变化来完成的<sup>[1]</sup>, 所以仅仅对 RNA 的天然结构的预测本身不足以给出我们想知道的 RNA 生物功能的微观机制. 我们还需要知道 RNA 结构的热力学稳定性、结构涨落、各种可能结构出现的几率, 以及 RNA 结构变化与折叠过程的动力学性质, 例如: 在结构变化过程中是否出现中间态, 中间态的稳定性, 结构变化的速度、决定速度的瓶颈过程( 过渡态 ), 以及结构变化所历经的路径等. 因为 RNA 分子的可能构型及其溶液环境构成了一个很大的统计系综, 所以上述 RNA 热力学与动力学性质将由系统的统计力学性质决定.

对 RNA 分子的结构及其热力学和动力学性质的定量预测与分析具有重要的实际意义. 首先, 通过

\* 美国 National Institutes of Health ( NIH/NIGMS ) 的资助( Grant No. R01 - GM063732 to S - J C ).

2005 - 07 - 11 收到初稿 2005 - 09 - 01 修回

<sup>†</sup> 通讯联系人. Email :Chenshi@ missouri. edu

对 RNA 折叠和结构变化的物理规律的了解,我们有可能用计算机沿着折叠的物理路径(而不是通过无规的随机搜索)来找出天然结构。这样可以大大缩短计算时间并可能找到随机搜索达不到的结构。另外,对 RNA 分子结构及其热力学和动力学性质的预测将直接有助于了解 RNA 生化过程的微观机制和可能由 RNA 分子的“错误”折叠所导致的疾病。如果我们能干涉或阻止某些致病 RNA 的折叠从而抑制其相关的生化反应,那么我们就可能有效地从基因水平上治疗疾病。所以, RNA 分子正越来越成为一些制药公司针对许多重大疾病(如癌症)的药物设计的重点。而我们对 RNA 折叠的预测能力无疑将大大加快药物的开发过程。

## 2 关于 RNA 分子折叠问题的一些简单介绍

### 2.1 RNA 结构

与蛋白质一样, RNA 是一种生物高分子。其高分子链中的每个单元被称为核苷酸。每个核苷酸由一个磷酸基(phosphate)、核糖(sugar)和碱基(base)组成(见图 1)。磷酸基和核糖的重复单元构成了 RNA 的主链,而碱基则构成 RNA 的侧链。RNA 的碱基有四种:腺嘌呤(A)、鸟嘌呤(G)、尿嘧啶(U)和胞嘧啶(C)。一个 RNA 分子的基因序列也就是其侧链上四种碱基的排列顺序。不同的碱基顺序对应着不同的 RNA 分子并决定了其相应的结构与功能。生物高分子与人工合成的化工材料高分子(如聚乙烯)最大区别可能在于:生物高分子有其特定的碱基(基因)序列而人工合成的化工材料高分子则没有这种复杂的序列。正是这种特定的碱基顺序决定了生物分子各种令人难以想象的奇异的生物功能,而一般的人工合成的化工高分子则只能作为一种材料。

RNA 分子中的腺嘌呤(A)和尿嘧啶(U)以及鸟嘌呤(G)和胞嘧啶(C)可分别配对形成氢键(A—U, G—C)。对一给定 RNA 碱基序列,不同的碱基之间可按照不同的组合进行配对从而形成各种不同的 RNA 结构。这种由碱基配对形成的结构叫 RNA 二级结构。如图 1 所示,基本的二级结构单元有两类:由连续的碱基对形成的螺旋(helix)以及由不含碱基对的区域形成的环(loop)。RNA 最终需要形成稳定的紧凑的三维结构。在 RNA 折叠过程

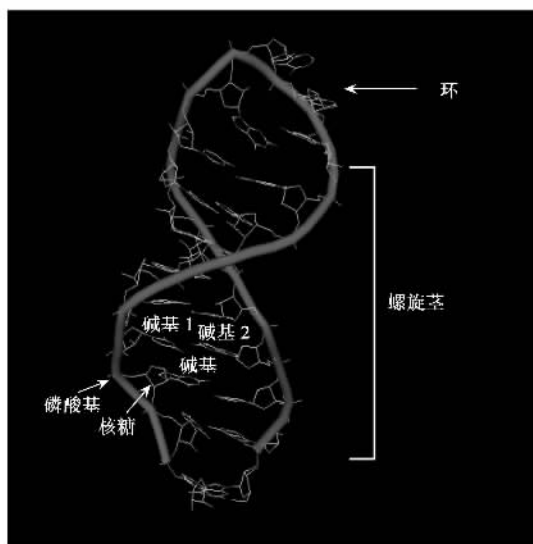


图 1 RNA 发卡结构(一种典型的 RNA 二级结构)由两部分组成:碱基配对区构成的螺旋茎(helical stem)以及由不含碱基对的区域形成的环(loop)。RNA 主链由磷酸基和核糖组成,侧链由碱基组成。碱基之间可通过氢键配对(如 base1—base2 之间)。相邻碱基对可形成碱基堆集力(如 base—base1 之间和 base—base2 之间)。

中,二级结构中的不同的螺旋和环相互作用而靠近进一步折叠聚集成一个致密的结构。这样的结构叫三级结构。在现行生物研究文献中,对 RNA 三级结构存在的不同的定义。一种比较常见的定义是通过碱基配对的高分子图来表述的<sup>[2-4]</sup>。在图 2 所示, RNA 分子主链由一直线表示,配对的碱基之间则用一曲线相连。如果高分子图存在交叉连接的曲线(碱基对),那么相应的结构则属于三级结构,否则为二级结构。根据这样的定义, RNA 胰结(见图 2)则属于三级结构<sup>[5,6]</sup>。由于 RNA 三级结构的复杂性,尽管近来人们对其折叠进行了大量的实验研究,但是对其物理与化学性质(例如能量参数等)的了解仍处在非常初级的阶段。本文将着重讨论 RNA 二级结构的折叠问题。

### 2.2 RNA 分子内的相互作用

RNA 分子的折叠和结构变换是通过核苷酸之间不同的配对来实现的,所以导致 RNA 折叠并稳定其结构的力来自于氢键之类的非共价键力<sup>[6]</sup>。具体来说, RNA 分子中存在下列几类非共价键力:静电力,原子间的色散力和范德瓦尔斯力,氢键和相邻碱基对间的堆集力(见图 1)。RNA 分子主链上的每个磷酸基带有一个电子单位的负电荷,所以 RNA 分子主链带有很强的负电荷。除此之外,主链和侧链上的一些基团可带有极化偶极子或感应极化偶极子。

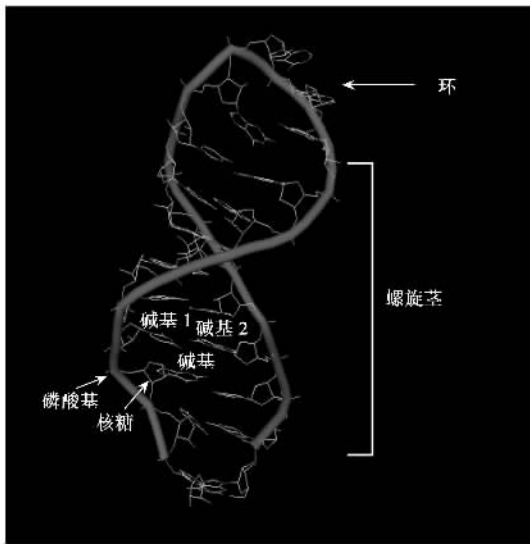


图2 高分子图及对应结构示意图。直线(粗)代表RNA主链,曲线(细)代表碱基配对。任何两根曲线之间存在三种关系:相互嵌套(nested)、互不相关(unrelated)、相互交叉(crossing linked)。如果高分子图存在交叉连接的曲线(碱基对),那么相应的结构则属于三级结构,否则为二级结构。根据这样的定义,RNA 假结(pseudoknot)则属于三级结构

这些电荷—电荷,电荷—偶极子,偶极子—偶极子之间的相互作用导致了RNA分子内的静电相互作用。对RNA二级结构来说,稳定其结构的最主要的力来自于碱基对之间的堆集力。碱基对之间的堆集力是由相邻配对碱基上不同基团的极化而导致的静电力以及色散力等引起的。堆集力将驱使相邻的碱基对形成有序的(螺旋)结构,以尽可能达到相邻碱基对之间的最大程度的堆集。但是,在另一方面,一个有序的(更受约束的)RNA分子结构的形成一定伴随着相应构形熵的减小。例如,环的形成伴随着熵的减小从而导致自由能的上升(如图3中两个环的自由能分别为 $+0.8 \text{ kcal/mol}$ 和 $+5.9 \text{ kcal/mol}$ )。熵的减少将阻止RNA的折叠<sup>[7]</sup>。以上两种有利和不利于RNA分子折叠的力的竞争结果决定了RNA分子的最终稳定结构分布及其热力学性质。RNA(及其溶液环境)作为一个统计系统,其结构与热力学和动力学性质是由温度等外界条件及RNA本身的碱基顺序决定的。所以温度的变化或碱基顺序的突变经常可导致RNA结构及其稳定性的变化<sup>[7-10]</sup>。

## 2.3 RNA分子折叠的一些特点

### 2.3.1 静电相互作用在RNA折叠中起很关键的作用 因为RNA分子主链上的每个磷酸基带有一个

电子单位的负电荷,所以RNA分子带有很强的负电荷。RNA分子的折叠需要溶液中的正离子(如 $\text{Na}^+$ 和 $\text{Mg}^{2+}$ )以(部分)中和磷酸基上的负电荷和(部分)屏蔽RNA主链上负电荷之间强烈的排斥作用<sup>[11,12]</sup>。特别是某些高价离子(如 $\text{Mg}^{2+}$ )对RNA的三级结构的形成是不可缺少的。这些离子在RNA结构中的一些特定部位(往往是一些电势较负的“口袋”部位)与RNA结合,其产生的力对稳定RNA紧凑的三级结构起到特定的作用。这些特定RNA结构中的“口袋”部位往往对离子的尺寸、水合性以及电子云的分布有其特异的选择性,而且除了静电作用以外,这种离子与RNA特定部位的结合一般来说还伴随着离子和RNA的溶剂化/去溶剂化的过程<sup>[13]</sup>。所以对这种特定的离子结合的分析不仅需要考虑到静电相互作用,还要考虑由于溶剂化和去溶剂化过程导致的自由能的变化。

### 2.3.2 RNA分子折叠过程中经常伴随着亚稳态结构的形成

与RNA分子折叠相比较,蛋白质分子折叠的主要驱动力来自氨基酸的亲水/疏水性,即疏水的氨基酸基团形成一个致密的疏水核心以最大可能地避免与水的接触,而亲水的基团则分布在疏水核心的表面<sup>[14-17]</sup>。显然,疏水核心的形成需要多个(沿着蛋白质主链)远程氨基酸基团之间的多体协作。理论分析发现天然蛋白质的氨基酸序列的最低自由能结构(天然结构;基态)和其他各种可能的结构(激发态)之间往往存在一个较大的能量间隙。大量实验和理论分析证实了蛋白质折叠有很强的协同性,即在展开的未折叠态与其致密折叠态之间不存在其他稳定分布的中间态。然而,RNA折叠与蛋白质的折叠则很不相同,RNA折叠往往不存在这种协同性。无论在热力学平衡态的(熔化)过程<sup>[8,9]</sup>还是非平衡的动力学过程中<sup>[18-23]</sup>,RNA往往需要经历(亚稳的)中间态。这可能是因为RNA中的碱基堆集力是由近邻的碱基对之间的相互作用决定的,所以是稳定RNA二级结构的力是短程(近邻)相关的力。因此从热力学角度看,碱基堆集力不足以引起不同碱基对之间的长程协同性。从动力学的角度看,因为碱基堆集力是一种相对来说比较强的力(每个碱基堆集稳定性 $\Delta G \sim 1-2 \text{ kcal/mol}$ ,而对蛋白质,整个分子的稳定性只有 $\Delta G \sim 10 \text{ kcal/mol}$ ),所以RNA二级结构很容易被稳定。在RNA折叠过程中,一个相对稳定的结构一旦形成就很难被破坏,这样的(亚稳)

结构就有可能成为一个可观测到的中间态。

### 2.3.3 二级与三级结构的比较

正是因为 RNA 二级结构本身是很稳定的,对大多数 RNA 分子来说,往往二级结构(螺旋、环)先形成,然后二级结构的不同部分再折叠形成紧凑的三级结构。人们目前对稳定三级结构的力还没有很好的定量了解,但许多实验证实了三级结构的力比二级结构的力要弱。在 RNA 融化实验中,三级结构往往先于二级结构融化,所以作为对 RNA 折叠一种简单化的近似,就可以把二级结构的折叠脱离三级结构单独处理<sup>[24]</sup>。事实上, RNA 分子的许多生物功能正是通过其二级结构的重组来实现的。但对蛋白质,仅仅靠其二级结构(α螺旋,β折叠等)本身是无法形成疏水核心的,所以脱开三级结构的蛋白质二级结构不稳定的。正因如此,脱开三级结构而只研究蛋白质二级结构本身是没有多大意义的。

在以下几节中,我们将分别就 RNA 分子折叠的热力学,动力学进行讨论。由于静电相互作用在 RNA 分子折叠中的独特的重要性,我们将单独为此进行较为详细的讨论。

## 3 RNA 分子的热力学性质

### 3.1 独立结构单元模型(最近邻模型)

RNA 分子的自由能来源于堆积力和环的贡献。对于堆积力,有来自两方面的贡献:碱基对之间堆积力(吸引力)倾向于形成碱基对堆积以减低系统的能量(焓),另外一方面来自碱基对堆积的熵的贡献,他倾向于打开碱基对,以增加系统的熵。而对于环,主要来自于熵的贡献,他同样倾向于打开碱基对,以增加系统的熵。

如果我们忽略 RNA 各个结构单元之间的相互作用,那么 RNA 二级结构的总自由能等于各个单独结构单元(碱基对堆积和环)自由能之和。由于一个结构单元是由一对最近邻碱基对决定的(除分叉环外,见以下讨论),所以该模型被称为最近邻(NN)模型。最近邻(NN)模型被广泛用于计算 RNA 分子的热力学性质。在图三中我们给出了一个如何计算 RNA 结构片段自由能的例子。该结构的总自由能(相对于完全打开结构的自由能)ΔG 可以通过各个结构单位的自由能相加得到;

ΔG < 0 表示所示结构相对于所有碱基打开的

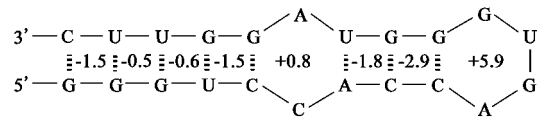


图3 一个 RNA 分子各个部分在 1M NaCl 溶液条件下的自由能参数(单位: kcal/mol),这些能参数来源于文献<sup>[27]</sup>。

自由态有更低的自由能。

RNA 分子的统计热力学性质由其配分函数 Z 决定:

$$Z = \sum e^{-E/k_B T}$$

其中  $\Sigma$  表示对所有可能构象求和, E 为构象的能量。如果直接通过简单地罗列所有构象的去计算 Z,所需时间将随着 RNA 分子的长度增加而呈指数增长,所以这种方法对于长的 RNA 分子是不现实的。为了快速计算 RNA 分子的配分函数,McCaskill 发展了一种动态的算法<sup>[26,27]</sup>。区别于穷举方法,在动态算法中,计算 RNA 分子的配分函数从比较短的部分开始,逐步增加碱基数目,短的 RNA 分子的配分函数被存储下来为下一步计算所用。利用动态算法,对于计算长度为 L 的 RNA 分子所需的时间正比于 ( $L^3$ )。McCaskill 的算法能较快速给出 RNA 二级结构的配分函数。然而 McCaskill 算法基于最近邻模型,他忽略了各个结构单元之间的相互作用(如排斥体积效应,即每个空间位置不能同时被两个 RNA 核苷酸单元或原子占有)。忽略排斥体积有时会导致较大的误差,例如,一个长度为 58 的 RNA 分子,如果忽略各个结构单元的排斥体积,可能会对整个系统的熵导致 40% 的误差<sup>[2]</sup>。另外,对于多分叉环结构,即与多个螺旋茎相连的分叉的环,McCaskill 假定熵和环的大小呈线性依赖关系,这个假设是非物理的,不正确的。

### 3.2 (结构单元间)近邻相互作用模型

如果采用高分子图表述 RNA 二级结构中碱基配对的情况, RNA 分子的配分函数可通过对各种可能的高分子图(而不是 RNA 分子构象本身)求和得到<sup>[34]</sup>:

$$Z = \sum_{\text{Graph}} \Omega e^{-E/k_B T}$$

上式中的  $\Omega$  是满足给定高分子图的 RNA 构象数目,通过高分子图形计算配分函数的好处在于对 RNA 二级结构高分子图形的求和可通过动态算法非常快速的进行。整个理论的关键在能否用解析的方法对任一给定图形快速地计算  $\Omega$ 。

为了有效地考虑不同结构单元的相互排斥体

积, Chen 和 Dill 以及后来 Zhang 和 Chen 发展了一种晶格模型去研究 RNA 分子的热力学性质<sup>[2-4]</sup>. 在晶格模型中, 每个格点代表一个 RNA 分子的碱基单元, 他们通过晶格中的网格线连在一起, 而这些网格线则等价于 RNA 分子的主链. 通过晶格模型, Chen 和 Dill 以及 Zhang 和 Chen 发展了一种矩阵相乘的方法来计算  $\Omega$ <sup>[4-6]</sup>. 下面我们来简单介绍一下这种方法.

根据碱基配对的情况, 如图 4(a) 所示, 我们可以把一个 RNA 结构分解成不同的单元. 根据最近邻和次近邻核苷酸单元在接口处空间构形的不同, 在三维晶格空间, 我们可以定义六种类型的接口, 如图 4(b) 所示. 为了描述这六种接口的相互关联, 我们进一步引进矩阵  $Y$ , 他定义了两个近邻结构片段的相互排斥情况. 例如, 如图 4(c) 所示, 类型为 2 的接口和类型为 4 的接口相互排斥, 所以  $Y_{24} = Y_{42} = 0$ . 类型为 2 的接口和类型为 3 的接口相互兼容并通过转动可以有三种连接法, 所以  $Y_{23} = Y_{32} = 3$ .

同时对每个结构单元我们可以引进结构矩阵  $S$ , 某矩阵元  $S_{mn}$  表示入口类型为  $n$ , 出口类型为  $m$  的构象数目. 从  $S$  和  $Y$  矩阵, 我们可以计算某一个含有  $N$  个结构元的整个高分子图对应的 RNA 构象数目  $\Omega$ :

$$S = S^{(N)} \left[ \prod_{j=1}^{(N-1)} Y \cdot S^{(j)} \right],$$

$$\Omega = U \cdot S \cdot U^+, \quad U = [1, 1, 1, 1, 1, 1]$$

上式中  $S^{(j)}$  是第  $j$  个结构元的结构矩阵. 用矩阵相乘的方法去计算某一特定高分子图所对应 RNA 的构象数目, 是非常快速有效的. 通过这种方法, 把大的 RNA 分子结构根据碱基配对的位置分成许多结构元, 因此我们所需要计算的只是结构元的  $S$  矩阵. 在实际计算中, 由于结构元的简单性, 我们可以事先把各种可能结构元的  $S$  矩阵计算并列表储存.

因为考虑了不同结构单元的相互排斥体积, 上述模型能较为准确地计算 RNA 的构象熵及其构象的统计(特别是对含有分叉的 RNA 环). 这种方法可以用来更准确地预测 RNA 分子的热力学性质, 在文献 [7] 中, Chen 和 Dill 计算了 6 种 RNA 分子的热容曲线, 发现计算结果和实验结果吻合. 另外, 该方法还可进一步扩展来处理有外力情况下 RNA 分子的热力学性质<sup>[28, 29]</sup>.

### 3.3 RNA 分子的自由能曲面图

为了更好地理解 RNA 分子的稳定性, 我们需要

计算出某个温度下 RNA 分子的自由能曲面图<sup>[7]</sup>. 所谓的自由能曲面就是把自由能作为结构的函数构造出的曲面. 通过自由能曲面图, 我们可以得到整个分子构象分布的全貌. 在自由能曲面图上, 极小值对应着稳定的结构. 整个自由能曲面上自由能最低的结构即是 RNA 分子最稳定的结构(天然结构). 为方便讨论, 我们定义出现在天然结构中的碱基对为天然碱基对, 其他的为非天然正常碱基对. 对给定一个高分子图, 其对应的 RNA 结构中的天然和非天然碱基对的数目也就确定了. 对所有含有  $n$  个天然碱基对和  $nn$  个非天然碱基对的高分子图求和, 我们便可得到相应的配分函数  $Z(n, nn)$  和自由能(曲面)函数  $F(n, nn)$ :

$$F(n, nn) = -K_B T \ln Z(n, nn)$$

自由能函数  $F(n, nn)$  给出了含有  $n$  个天然碱基对和  $nn$  个非天然碱基对的所有结构集合态的稳定性. 与总自由能  $F = \sum_{n, nn} F(n, nn)$  相比, 自由能曲面  $F(n, nn)$  非常直观地揭示了自由能和结构  $(n, nn)$  的关系. 从而便于我们能直接从自由能曲面的形状及其变化来得到结构分布及结构变化的信息.

在图 5 中, 我们给出了 *E. coli* 23S rRNA 片段<sup>[9]</sup> 自由能曲面及对应的折叠路径. 随着温度的升高, 该分子呈现出非常复杂的折叠路径. 在  $T = 30^\circ\text{C}$ , 我们发现一个非天然稳定结构 ( $Z$ ) 与天然结构 ( $N$ ) 共存, 我们称  $Z$  为非天然结构是因为  $Z$  含有许多非天然的碱基对. 而且, 该分子似乎存在两条完全不同的并行的 RNA 折叠路径, 一条含有非天然结构中间态, 另一条则没有. 而对另外有些 RNA 分子, 例如一种突变  $\alpha$  mRNA 片段<sup>[8]</sup>, 自由能曲面及对应的折叠路径要简单得多. 随着温度的升高, 突变  $\alpha$  mRNA 片段的自然结构逐步经历几个只含天然碱基对的中间态. 并且只存在一条路径部分. 在整个过程中没有出现稳定的非天然碱基对. 对结构变化的过程的了解有助于我们对 RNA 分子性能的设计和改造. 例如, 对  $\alpha$  mRNA 片段, 因为只有一条路径, 所以阻断其中间态的形成就可能阻止其结构的变化. 而对 *E. coli* 23S rRNA, 阻断一条路径可能会导致 RNA 结构以更大的可能性通过另一条与之平行的路径变化.

从图 5 我们可以看出, RNA 分子的自由能曲面高低不平, 对应着多个局域的亚稳态结构(即面上的诸个低点). 这些亚稳态结构的存在可能和 RNA 分子生物功能有关, 有些 RNA 分子需要通过

不同结构之间的跃迁来执行多种不同的生物功能。这种复杂的多个局域的亚稳态结构,可能是一种自然界生物进化的选择结果。

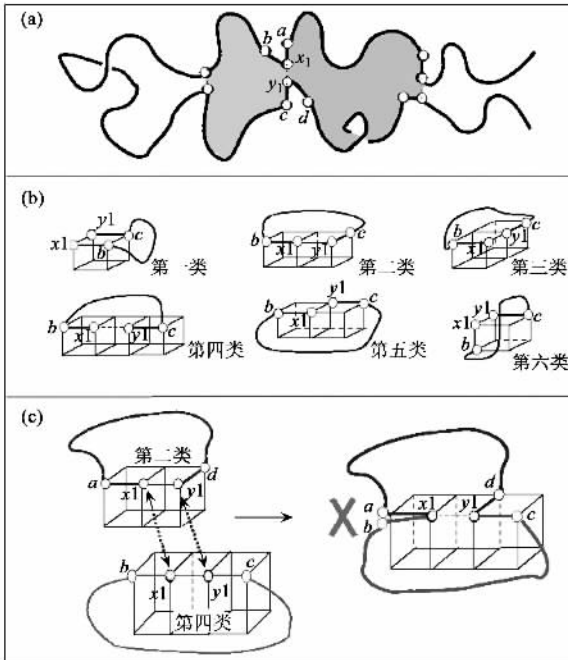


图4 (a) RNA 二级结构可以按照碱基对分成不同的结构单元。每个结构单元内部不含任何碱基对。(b) 在三维晶格模型中,按照接口处碱基对的构形,我们把结构单元可以分成六类。(c) 整个 RNA 分子的构形可由各个结构单元通过接口处碱基对接而成。根据接口处碱基对的构形是否匹配,有些拼接可行而有些则不可行。例如第二类与第四类接口因为排斥体积作用(图中 *a* 和 *b* 碰撞)而无法对接

## 4 核酸分子折叠中的静电作用

核酸分子,包括 RNA 和 DNA,带有高密度的负电荷。这些电荷主要分布在核酸分子主链(磷酸基)上,因而核酸分子主链上存在非常强的静电排斥作用。因此,溶液中的离子对核酸分子结构有非常重要的影响。溶液离子被吸引并束缚在磷酸基周围屏蔽核酸分子上的库仑排斥作用,这样核酸分子的不同部分才能彼此靠近并折叠形成致密的天然结构。所以,理解离子与核酸分子的静电作用对预测核酸折叠的结构以及折叠的热力学性质和动力学过程非常重要。然而,因为核酸分子溶液是一个典型的复杂多体系统,这里涉及到离子与离子,核酸与离子,核酸与核酸,以及核酸与溶液和离子与溶液间相互作用是极其复杂的<sup>[13,30]</sup>。由于历史原因,无论是实验还是理论,对 DNA 的静电性质的研究要比 RNA 更广

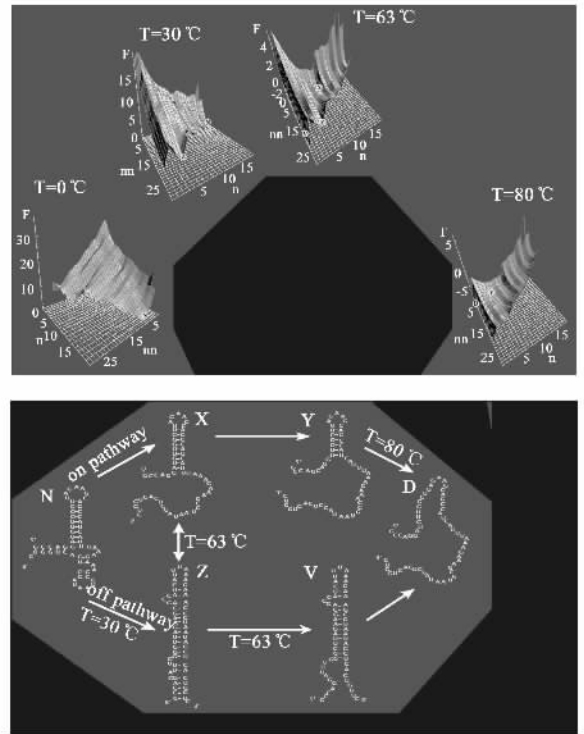


图5 突变 *E. Coli* 23S rRNA 片段的自由能曲面与折叠路径<sup>[7]</sup>

泛。因为 RNA 和 DNA 有着极相似的静电性质,所以在本节中我们将同时讨论关于 DNA 和 RNA 静电相互作用的一些主要理论和实验结果。

### 4.1 理论方法

由于溶液离子对核酸结构的重要性,从 20 世纪 60 年代到现在,此领域的研究一直广泛受到关注。下面,我们介绍一下此领域的主要理论和最近进展,以及结合相关实验介绍这些理论的应用。这里主要介绍离子凝聚(counterion condensation, CC)理论,泊松—玻尔兹曼(Poisson - Boltzmann, PB)理论和紧束缚离子(tightly bound ion, 简称 TBI)理论。

#### 4.1.1 离子凝聚(CC)理论

离子凝聚(CC)理论<sup>[31]</sup>是由 Manning 上个世纪 60 年代末发展起来的,此理论的提出是为了描述离子与 DNA 的相互作用。CC 理论是一个二态模型:一阳离子可均匀分布在凝聚态区域(即靠近核酸表面的一定区域)或自由态区域。当一个正离子从自由态转变为凝聚态时,其减小的熵将被系统静电能的降低而部分抵消,两者间的竞争导致了最终平衡态的离子分布。CC 理论形式简单并在一定程度获得了成功<sup>[32]</sup>。但是此理论有以下缺点:不能处理离子分

布的涨落和离子-离子间的关联,同时他也无法处理核酸分子的具体形状,所以 CC 理论不是一个精确的理论。

#### 4.1.2 泊松-玻尔兹曼(PB)理论

泊松-玻尔兹曼(PB)理论<sup>[33-36]</sup>的历史可以追溯到 20 世纪初 Gouy-Chapman 的电双层理论和 Debye-Huckel 的离子溶液理论。这两种理论是 PB 理论应用于不同具体对象的形式:Gouy-Chapman 理论是 PB 对于电双层问题的一维简化;而 Debye-Huckel 理论是 PB 理论对于离子溶液球对称的一维简化及线性化。PB 方程的导出比较简单:媒质中的静电场分布遵循如下泊松方程,

$$\nabla \cdot \varepsilon \nabla \psi = -4\pi(\rho_f + \sum_a z_a e c_a),$$

上式中  $\psi$  是平均电势,  $\rho_f$  是固定电荷的密度。溶液中的游离离子电荷的分布遵循玻尔兹曼分布:

$$c_a = c_a^0 e^{-ze\psi/k_B T}.$$

上述两个方程结合起来,就是泊松-玻尔兹曼(PB)方程。PB 方程还要满足边界条件:在无穷远处,  $\psi \rightarrow 0$ ;在分子和溶液的交界处,  $\psi$  和  $\partial\psi/\partial n$  都必须保持连续。

PB 理论的物理非常简洁,方程的求解也非常方便,所有的物理量都可以方便的计算,而且可以应用于具有非常复杂形状的生物分子,所以 PB 理论被广泛应用于生物分子电相互作用能的计算。然而, PB 理论用了一个很重要的近似,那就是平均场近似,即离子间的关联作用被忽略了。PB 方程实际上是关于平均场势  $\psi$  的一个自治方程。然而对于高价(如  $MgCl_2$ ) 盐溶液,离子间的关联作用非常强烈,所以 PB 理论不能定量的预测 RNA(或其他生物分子)在高价盐溶液的热力学性质。然而,这些高价的阳离子对于很多生命反应过程非常重要,包括这里谈到的核酸折叠问题。

#### 4.1.3 紧束缚离子(TBI)理论

为了能够处理对于核酸折叠非常重要的高价阳离子溶液,必须要考虑离子间的关联相互作用。Tan 和 Chen 近来发展了紧束缚离子(TBI)理论<sup>[37]</sup>。此理论的出发点是把强关联的离子和弱关联的离子分开进行分别处理。离子间有两种关联作用:库伦作用和排斥体积相互作用。两种强关联作用分别由以下条件给出:

$$\frac{(ze)^2}{\varepsilon k_B T a_{ws}} \geq \Gamma_c, 2a_{ws} \geq (2r_c + \Delta r),$$

上式中  $\Gamma_c$  为离子系统汽-液相变的临界值,  $a_{ws}$  是 Wigner-Seitz 半径,  $r_c$  为离子半径,  $\Delta r$  为离子从平衡位置热涨落的距离,他的临界值可以由 Linderman 熔化理论给出。按照以上条件,溶液空间划分为紧束缚(等于号强关联)区域和游离(等于号弱关联)区域(见图 6)。游离离子可用 PB 方程来描述。而对于紧束缚离子,我们考虑他们在各个磷酸基周围的不同分配模式,对每一模式我们精确地计算在紧束缚区内所有电荷(包括磷酸基的负电荷)之间的相互作用。这样一来我们既考虑了紧束缚离子在核酸分子附近分布的涨落,又考虑了紧束缚区内所有电荷之间的强关联效应。而且, TBI 理论可以考虑核酸分子结构的细节。分子结构的细节往往在溶液离子与核酸分子间的相互作用中扮演着至关重要的角色。

对低浓度单价离子,因其关联效应很弱,所以 TBI 和 PB 给出相同的结果。但对于双价或更高价的离子溶液, TBI 理论可以给出 PB 更精确的结果。例如,因为 PB 的平均场无法考虑紧束缚离子的单个分布模式,特别是那些稳定的模式,所以 PB 往往低估高价离子溶液中核酸分子的稳定性,而 TBI 理论却能给出较为准确的结果。

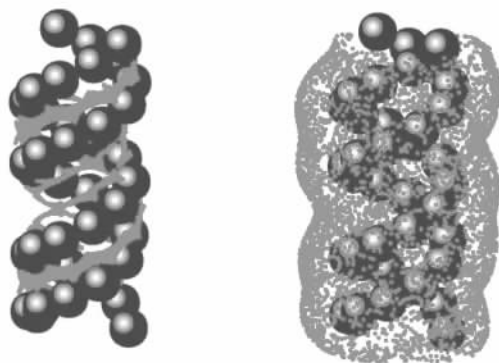


图 6 紧束缚离子(TBI)理论中 B-DNA 周围的紧束缚离子区域。这里 DNA 长度为 20-bp,离子浓度为 0.1M;左图为 NaCl 溶液,右图为  $MgCl_2$  溶液。由于镁离子比钠离子的静电关联要强得多,所以其相应的紧束缚离子区域也大得多

#### 4.1.4 其他的多电极溶液理论及计算机模拟

除了上述介绍的静电溶液理论外,人们还发展了其他的一些理论。比如,有不同精确度近似的积分方程理论<sup>[38]</sup>,基于 Kirkwood 或 BBGY 方程的改进 PB 理论<sup>[39]</sup>等等。从某种程度上,这些理论比前面所介绍的理论更为精确,但是,由于这些理论非常复杂,计算复杂度非常之高,在目前计算条件下,很难

应用于具有复杂形状的生物大分子。

#### 4.2 核酸分子折叠过程中的电相互作用

在核酸分子折叠过程中,二级结构因强烈的碱基对以及堆集力的作用而可以稳定地存在。这些伸展的带负电的二级结构单元,例如螺旋茎和环,最终要折叠形成致密的天然结构。他们究竟是如何克服相互间的静电排斥力而稳定地聚集在一起?这仍然是一个尚未完全解决的具有挑战性的问题。

对于 DNA,三级致密结构的形成称为 DNA 凝聚(DNA condensation)<sup>[30]</sup>和 DNA 聚集(DNA aggregation,多条 DNA 分子聚集在一起形成致密结构)<sup>[40]</sup>。Bloomfield 等系统地研究了不同价数的离子对 DNA 凝聚的影响,他们发现只有高价离子才能导致 DNA 的凝聚,具体来讲,+3 或 +3 价以上的阳离子可以引起 DNA 凝聚,而 +2 价离子(比如  $Mg^{2+}$ )可以在特殊的条件导致 DNA 凝聚,比如加入酒精<sup>[30]</sup>。近年来在实验发现,+2 价离子可以导致夹在两层脂膜中的 DNA 聚集,实验用到的多种 +2 价离子皆可导致 DNA 聚集<sup>[41]</sup>。对于其他带有电荷的生物大分子,人们也广泛研究了离子诱导的聚集现象,这里包括 Tang 等人 and Wong 等人对 F-actin, bacteriophages fd 等分子的研究<sup>[42,43]</sup>。以上的实验表明,高价离子所特有的性质与带电生物分子的聚集(或凝聚)有着直接联系。因此,高价离子间的强关联作用被认为是导致 DNA 凝聚的主要原因<sup>[30]</sup>。

对于 RNA 分子,高价离子(比如  $Mg^{2+}$ )在 RNA 三级结构形成中起着不可替代的作用。Woodson 和 Thirumalai 等系统地研究了离子价数以及离子大小对 RNA 折叠结构以及折叠动力学的影响<sup>[19]</sup>,实验表明,高价离子和小尺寸离子对 RNA 折叠更有效,为了解释这些实验现象,他们提出一个简单的物理模型:RNA 三级结构形成被认为来自于离子与磷酸基形成的偶极-偶极(凝聚离子的关联)吸引相互作用,而高价离子产生的吸引要强得多。Herschlag 等人也探究了 RNA 三级结构折叠的驱动力<sup>[44]</sup>,他们不认为驱动力来自于凝聚离子关联导致的吸引作用,理由是高浓度的单价离子(如 NaCl)也可以导致 RNA 三级结构折叠,尽管这样折叠成的三级结构只是一种与天然结构相似的致密结构;他们认为,在 RNA 分子折叠的初期,高价离子(比如  $Mg^{2+}$ )和高浓度单价离子一样,只是对带电的核酸分子提供高效的静电屏蔽(或电荷中和),从而使 RNA 二级结构单元能够在溶液中较自由地折叠,使不同的二级

结构单元接触而形成三级结构。

由于核酸三级结构的重要性,除了上述介绍的一些实验研究外,人们在这方面还开展了大量的理论工作,包括对核酸螺旋茎之间静电力的研究以及离子在三级结构中特定的稳定作用的研究等。

Draper 等人利用 PB 理论系统的研究了 RNA 三级结构中离子(特别是  $Mg^{2+}$ )的作用,他们计算了自由能等热力学量,并得到一些与实验一致的结果<sup>[13]</sup>。但因为 PB 理论忽略了离子间的关联作用,所以 PB 对高价离子的预测结果只能是定性的而且无法解释许多重要的物理机制。比如,大量的实验和计算机模拟表明,核酸分子的螺旋茎间存在由于高价离子关联而导致的吸引作用,而 PB 理论对不同价数的离子溶液都永远只能给出排斥作用。

为解释(带负电的)核酸分子螺旋茎间的吸引作用,Gelbart 等人用分子动力学方法计算了两个平行带电分子间的电相互作用力<sup>[45]</sup>,他们模拟表明,对于高电荷密度的分子,高价离子可以导致分子间产生吸引作用。他们通过对分子相互吸引状态下离子分布的分析得出,吸引是由于 DNA 靠近时强烈的库仑相互作用导致了类似 Wigner 晶体结构的形成。Stevens 利用分子动力学研究了离子溶液中带电分子的聚集行为<sup>[46]</sup>,模拟结果显示,带电分子在 +2 价离子溶液中形成致密的扎堆结构,而单价离子不能导致此结构的形成。Lyubartsev 等人利用蒙特卡罗法,系统计算了多根 DNA 系统的渗透压,模拟得出,+2 价离子可以导致 DNA 之间产生吸引作用,而单价离子不能,吸引力的力程为 26—34Å(DNA 轴间距)。他们还得到,离子浓度、离子大小对吸引力的强度、力程有重要影响<sup>[47]</sup>。最近,Gelbart 等人用分子动力学计算了两根非平行(有一定夹角)带电分子间的静电力,结果显示,随着高价离子浓度的增加,溶液中带电分子的结构会发生相变,分子间夹角由相互垂直变为相互平行,因为离子浓度低时,相互作用为排斥力,而相互垂直情形下排斥力最小,当离子浓度高时,分子间作用为吸引力,分子相互平行吸引力最强<sup>[48]</sup>。

在 CC 理论框架下,Ray 和 Manning 研究两根平行的 DNA 螺旋间的吸引作用<sup>[49]</sup>。他们计算了系统自由能随距离的变化关系,发现两根 DNA 间存在吸引作用。他们还发现,这种吸引来自于两根 DNA 靠近时凝聚态离子的自由体积的增大,从而凝聚离子的熵增大而导致自由能降低,因而产生吸引。对于 NaCl 等单价离子溶液,此理论预测核酸-核酸分子



间存在强的吸引力. 显然, 这种吸引力产生的机制与大量实验和计算机模拟并不相符. Rose 等也应用 CC 理论分析了两根核酸片段的相对平衡方位, 并估算了他们间电相互作用自由能<sup>[50]</sup>. 显然, 这种将无限长的 CC 理论来估算短的核酸片段很不严格, 而且 CC 理论本身并不是严格的理论.

针对核酸分子螺旋茎间吸引力的问题, 人们发展了大量的物理模型来解释这种现象, 这些模型主要基于形成 Wigner 晶体的思想<sup>[51-54]</sup>: 由于强的电相互作用, 当分子靠近时, 结合于核酸分子表面附近的离子会通过自组织形成有序结构, 而这种有序结构导致核酸分子间等效吸引力以降低系统能量(对于高价离子, 这种自组织尤其更为重要). 随着温度的升高, 这种有序结构逐渐遭到破坏, 从而导致吸引力逐渐减弱. 在高温极限下, 离子完全无序, 螺旋茎之间互相排斥.

尽管大量的实验和计算机模拟证实了核酸分子间吸引力的存在, 而且人们提出了大量的物理理论来解释这些现象, 然而, 这些物理理论过于简化, 条件过于理想, 比如, 很多理论没有考虑溶液的盐浓度, 这些理论的对象也很简单, 比如无限大带电平板、无限长带电圆柱等, 过于简化的理论可能可以给人们以物理机制的理解, 但是不能给出定量甚至半定量的结果, 而这些定量结果正是我们预测核酸三级结构所需要的.

## 5 RNA 折叠的动力学性质

RNA 分子折叠速度经常比所参加的生化反应速度慢. 在这种情况下, RNA 分子的结构(分布)尚未达到平衡态, 所以其性质不是由平衡态热力学决定的, 而是由折叠过程的动力学决定的. 由此可见, 了解 RNA 二级和三级结构的折叠机制(例如折叠路径、折叠中间态)对于理解 RNA 以及相关的药物设计非常重要. 由于篇幅限制, 在此我们对 RNA 分子折叠的动力学只能做一些极其初略的介绍.

RNA 折叠和反折叠的动力学测量始于 20 世纪 70 年代通过电弧放电使溶液温度瞬间升高, 从而测量其光吸收随时间的变化<sup>[55-60]</sup>. 因为温度升高, 一部分要熔化, 而打开碱基对时, 光吸收会增加, 这样通过光吸收的变化可以得到打开和未打开的浓度变化. 未打开(未熔化)分子的浓度呈指数变化:  $\Delta[U] = \Delta[U]_0 e^{-t/\tau}$ , 其中  $\Delta[U]_0$  为温度跳跃之前和之后平衡态的浓度差,  $t$  为时间,  $\tau$  为弛豫时间,

对于单分子反应,  $\tau^{-1} = k_1 + k_{-1}$ , 其中  $k_1$  为折叠率,  $k_{-1}$  为反折叠(打开)率, 同时根据平衡常数  $K = k_1/k_{-1}$ , 可以得到折叠率和反折叠率.

这些早期的工作主要是针对一些简单的发夹结构(见图 1)的折叠的动力学的研究. 对于在这些早期实验中的发夹结构来说, 其折叠的瓶颈在于环的形成. 一旦发夹结构的环形成了, 其螺旋茎的形成是很快的. 这种折叠过程可用“拉链”模型来描述: 折叠和反折叠可描述为沿着发夹结构的螺旋茎合上或打开. 大多数 RNA 发夹折叠的时间为微秒级别. 近几年来, 随着技术的发展和 RNA 分子的结构及其在疾病的机理和治疗方面的重要性的进一步认识, RNA 的动力学研究开始活跃起来. 其中主要包含两类: 一类是关于 RNA 发夹的动力学, 另一类是关于大的 RNA 分子的动力学.

RNA 发夹结构是构成 RNA 三维结构的基本单元, 是形成 RNA 其他复杂结构的先驱. 了解发夹结构如何折叠是理解其他复杂结构的基础. 同时, RNA 发挥其功能时, 常常伴随着 RNA 发夹的折叠与打开. 近来关于 RNA (DNA) 等发夹结构的动力学研究活跃起来, 采用的方法主要有用激光快速升温测量光吸收和荧光法<sup>[61-64]</sup>. 越来越多的实验发现 RNA (DNA) 发夹结构虽简单, 但其动力学却远比以往人们所想象的复杂: 例如, 对不同的碱基序列和温度, 在折叠过程中可能出现非天然的结构, 其折叠过程的激活能可正可负.

实验上对于大的 RNA 的动力学研究主要集中在二级结构如何相互作用形成三级结构. 实验上已发展了一系列方法: oligonucleotide hybridization<sup>[65]</sup>, UV cross-linking<sup>[66]</sup>, time resolved hydroxyl radical footprinting<sup>[67]</sup>, 小角 x-ray 散射<sup>[68]</sup>, 单分子荧光光谱及单分子力拉伸<sup>[69-72]</sup>. 这种大的 RNA 分子的折叠通常伴随着特定的正离子( $Mg^{2+}$ )的参与. 其折叠通常分为早期和后期两个阶段. 在折叠开始时, 由于 RNA 分子的不同片段的强烈的静电排斥, RNA 分子为展开的结构. 当加入正离子时, 静电排斥被正离子屏蔽, RNA 很快塌缩成紧凑的结构. 这种早期过程形成的结构有两类<sup>[73]</sup>: 一类是与蛋白质折叠过程中形成的熔解团(molten globules)相似的非特定的结构; 另一类则是由于特定的 RNA 内部不同核苷酸之间或 RNA-离子之间相互作用而形成的一些局域结构. 这种早期结构对折叠的影响非常大. 如果这种早期结构在后期折叠过程中必须打开, 那么这种结构就成为动力学陷阱而导致折叠速度的降低. 后期折

叠过程(三级结构的形成)存在三种限制折叠率的过程<sup>[73,74]</sup>(1)构象搜索(2)与特定金属离子的结合(3)跳出动力学陷阱。第一类情况类似于RNA搜索成核构象,一旦稳定的临界成核结构的形成了,其余部分则可以在成核结构的基础上很快地成长;第二类情况是由于有些金属离子必须与RNA结构的特定区域结合;第三类情况是由于RNA必须跳出早期过程形成的陷阱以便形成正确的天然结构。对于一个序列,可能有多条路径折叠。一部分快速折叠成功,而另一部分则陷在陷阱。这种机制通常描述为动力学分配机制<sup>[75]</sup>。

随着单分子技术的进展,用单分子研究生物分子的折叠包括单分子荧光(FRET,荧光共振能量传递)和力学拉伸<sup>[68-72]</sup>。在通常系综测量时,我们得到的是一个对各个可能构象及其所有可能折叠路径平均的结果。而单分子实验则不同。采用单分子的一个优势是在不改变溶液条件的情况下进行多次折叠与反折叠,每次可以得到该单个路径的折叠和反折叠率。用外力拉伸打开折叠的一个优势就是伸长可以作为一种自然的反应坐标(而不是对伸长求平均),有助于直接从微观结构了解其动力学机理。

理论上研究RNA动力学的方法主要包括蒙特卡罗模拟<sup>[76-79]</sup>,分子动力学模拟<sup>[80]</sup>,统计力学理论模型<sup>[80]</sup> Monte Carlo把RNA的折叠过程当成一个随机过程。Flamm<sup>[77]</sup>等人采用的是两个态之间的跃迁率为Kawaski的模型,两态之间可增加,减少或扩散一个碱基对。Isambert<sup>[78]</sup>等人采用的是动力学Monte Carlo,每两个态之间形成或打开一个螺旋茎。

基于统计力学模型的方法也被用来研究RNA折叠的动力学<sup>[81]</sup>。这些方法基于描述折叠过程中不同构象间跃迁的主方程:

$$\frac{dp_i}{dt} = \sum_{j=1}^{\Omega} (k_{j \rightarrow i} p_j - k_{i \rightarrow j} p_i)$$

其中 $p_i$ 为态 $i$ 的几率, $k_{i \rightarrow j}$ 为相应的变化率。对RNA二级结构来说,由于RNA最近邻碱基对之间的相互作用,两个相邻的碱基对形成一个稳定的堆积。所以我们可以把形成或去掉一个堆积(或形成堆积的碱基对)定义为一个基本的动力学步骤(即主方程中的 $i \rightarrow j$ 或 $j \rightarrow i$ )。由于形成碱基对时熵减小,而打开堆积时熵上升,因此我们假定形成和打开一个碱基对堆积的势垒分别来自减小的熵( $\Delta S$ )和增加的焓( $\Delta H$ )。基于这样的假定,合上( $k_+$ )和打开( $k_-$ )一个碱基对堆积的反应率为:

$$k_+ = k_0 e^{-\Delta S/k_B} \quad k_- = k_0 e^{-\Delta H/k_B}$$

常数 $k_0$ 可根据与一系列实验的对比得到。根据得到的反应率矩阵( $k_{ij}$ )的本征值和本征矢可得到折叠率和反折叠率以及限制折叠率和反折叠率的瓶颈<sup>[82]</sup>。这种方法包含全部的构象,可得到精确的解。但是,该方法只能给出对各种微观跃迁过程平均的结果,无法给出微观的折叠路径。而且由于构象数目随链的长度指数增长,该方法只能用于短分子。

如果在折叠过程中,存在一个或几个瓶颈过程,那么在通过这些瓶颈过程之前,分子构象被滞留较长时间,从而使得这些构象间能有足够长的时间相互转换并达到平衡,那么我们就可以把这些构象定义为一个达到(局域)平衡的集团。这样一来,我们就可以把巨大的构象系综约化为一些集团。通过对集团间跃迁的细致研究,我们就可得到分子折叠的微观动力学机制,包括折叠路径<sup>[83]</sup>。这种方法称为动力学集团法。假定 $U$ 和 $N$ 是两个集团,那么两个集团之间的转变率可写为:

$$k_{U \rightarrow N} = \sum [U_i] k_{U_i \rightarrow N_i}; \quad k_{N \rightarrow U} = \sum [N_i] k_{N_i \rightarrow U_i}$$

其中 $[U_i]$ 和 $[N_i]$ 为构象 $U_i$ 和构象 $N_i$ 在相应的集团 $U$ 和 $N$ 中所占的几率。因为各集团内部已达到(局域)平衡, $[U_i]$ 和 $[N_i]$ 可根据玻尔兹曼分布得到。 $k_{U_i \rightarrow N_i}/k_{N_i \rightarrow U_i}$ 为 $U_i$ 和 $N_i$ 之间的转变率。 $k_{U_i \rightarrow N_i}$ 和 $k_{N_i \rightarrow U_i}$ 根据其不同过程(合上还是打开一个碱基对堆积)由 $K_+$ 或 $K_-$ 得到。通过动力学集团法,我们可以得到通过任何一条微观折叠路径 $U_i \rightarrow N_i$ 的几率:

$$f_i = [U_i] k_{U_i \rightarrow N_i} / k_{U \rightarrow N}$$

而整个动力学过程的最可几的路径就是 $f_i$ 最大的那条。由于 $[U_i]$ 和 $k_{U_i \rightarrow N_i}$ 与温度和分子序列有关,因此最可几的路径也与温度和分子序列有关。RNA分子的折叠动力学依赖于分子序列,对不同的分子序列的理论研究得到了以下一些结论:

(1) RNA二级结构的折叠通常在微秒量级内完成,但RNA三级结构的折叠一般比较慢(几秒到几个小时)。RNA二级结构的折叠可能并不存在对蛋白质折叠的一种理论猜测,即好的(快速并且有较高协同性的)折叠分子在其基态和第一激发态之间存在大的能隙<sup>[76]</sup>。

(2) 折叠过程中的路径形成动力学分叉以及动力学分配机制<sup>[77]</sup>。

(3) RNA的折叠与序列和温度有关。在折叠过程中存在非天然构象(陷阱)的中间态导致了快速折叠路径与慢速折叠路径的竞争。非天然构象中间态的存在以及不同路径间的竞争显然与序列和温

度有关,从而使 RNA 的折叠在不同的序列和温度有关可表现为不同的特征:从高协作性到低协作性甚至玻璃化转变,激活能可从负变到正,等等<sup>[81—83]</sup>。

相对于蛋白质折叠来说, RNA 折叠是一个新的领域. 目前该领域面临的最大的挑战是如何从物理原理来预测三级结构及其热力学和动力学性质. RNA 折叠的复杂性很大程度上来自其中极其复杂的电相互作用. 但是随着实验技术和计算机技术的不断发展和更精确的关于 RNA 折叠的物理模型的建立, RNA 分子折叠问题一定会逐步得到解决.

### 参 考 文 献

- [ 1 ] Staley J P, Guthrie C. *Cell* 1998, 92 :315
- [ 2 ] Chen S J, Dill K A. *J. Chem. Phys.*, 1995, 103 :5802
- [ 3 ] Chen S J, Dill K A. *J. Chem. Phys.*, 1998, 109 :4602
- [ 4 ] Zhang W, Chen S J. *J. Chem. Phys.*, 2001, 114 :7669
- [ 5 ] Tinoco I Jr, Bustamante C. *J. Mol. Biol.*, 1999, 293 :271
- [ 6 ] Burkard M E, Turner D H, Tinoco I Jr. 1999. The interactions that shape RNA structure. In: *The RNA world* (2nd ed.). Gesteland R F, Cech T R, Atkins J F. Cold Spring Laboratory Press. 233
- [ 7 ] Chen S J, Dill K A. *Proc. Natl. Acad. Sci. USA*, 2002, 97 :646
- [ 8 ] Gluick T C, Draper D E. *J. Mol. Biol.*, 1994, 241 :246
- [ 9 ] Laing L G, Gluick T C, Draper D E. *J. Mol. Biol.*, 1994, 237 :577
- [ 10 ] Matveev S V, Filimonov V V, Privalov P L. *J. Mol. Biol.*, 1982, 16 :990
- [ 11 ] Heilman - Miller S L, Thirumalai D, Woodson S A. *J. Mol. Biol.*, 2001, 306 :1157
- [ 12 ] Heilman - Miller S L, Pan J, Thirumalai D *et al.* *J. Mol. Biol.*, 2001, 309 :57
- [ 13 ] Draper D E, Grilley D, Soto A M. *Annu. Rev. Biophys. Biomol. Struct.*, 2005, 34 :221
- [ 14 ] Dill K A, Chan H S. *Nature Struct. Biol.*, 1997, 4 :10
- [ 15 ] Wolynes P G, Onuchic J N, Thirumalai D. *Science*, 1995, 267 :1619
- [ 16 ] Sali A, Shakhnovich EI, Karplus M. *Nature*, 1994, 369 :248
- [ 17 ] Dill K A. *Protein Science*, 1999, 8 :1166
- [ 18 ] Russell R, Zhuang X, Babcock H P *et al.* *Proc. Natl. Acad. Sci. USA*, 2002, 99 :155
- [ 19 ] Woodson S A. *Curr. Opin. Chem. Biol.*, 2005, 9 :104
- [ 20 ] Brenowitz M, Chance M R, Dhavan G *et al.* *Curr. Opin. Struct. Biol.*, 2002, 12 :648
- [ 21 ] Sosnick T R, Pan T. *Curr. Opin. Struct. Biol.*, 2003, 13 :309
- [ 22 ] Treiber D K, Williamson J R. *Curr. Opin. Struct. Biol.*, 2001, 11 :309
- [ 23 ] Batey R T, Doudna J A. *Nature. Struct. Biol.*, 1998, 5 :337
- [ 24 ] Zarrinkar P P, Williamson J R. *Science*, 1994, 265 :918
- [ 25 ] Serra M J, Turner D H. *Methods Enzymol.*, 1995, 259 :242
- [ 26 ] McCaskill J S. *Biopolymers*, 1999, 29 :1105
- [ 27 ] Bundschuh R, Hwa T. *Phys. Rev. Lett.*, 1999, 83 :1479
- [ 28 ] Liu F, Dai LR, Ou - Yang Z C. *J. Chem. Phys.*, 2003, 119 ( 15 ):8112
- [ 29 ] Dai LR, Liu F, Ou - Yang Z C. *J. Chem. Phys.*, 2003, 119 ( 15 ):8124
- [ 30 ] Bloomfield V A. *Biopolymers*, 1997, 44 :269
- [ 31 ] Manning G S. *Q. Rev. Biophys.*, 1978, 2 :179
- [ 32 ] Record M T Jr, Lohman T M. *Biopolymers*, 1978, 17 :159
- [ 33 ] Sharp K A, Honig B. *J. Phys. Chem.*, 1990, 94 :7684
- [ 34 ] Davis M E, McCammon J A. *J. Comput. Chem.*, 1991, 12 :909
- [ 35 ] You T J, Harvey S C. *J. Comput. Chem.*, 1993, 14 :484
- [ 36 ] Baker N A, Sept D, Joseph S *et al.* *Proc. Natl. Acad. Sci. USA*, 2000, 98 :10037
- [ 37 ] Tan Z J, Chen S J. *J. Chem. Phys.*, 2005, 122 :044903
- [ 38 ] Vlachy V. *Annu. Rev. Phys. Chem.*, 1999, 50 :145
- [ 39 ] Carnie S L, Torrie G M. *Adv. Chem. Phys.*, 1984, 56 :141
- [ 40 ] Pelta J, Livolant F, Sikorav J L. *J. Biol. Chem.*, 1996, 271 :5656
- [ 41 ] Koltover I, Wagner K, Safinya C R. *Proc. Natl. Acad. Sci. USA*, 2000, 97 :14046
- [ 42 ] Tang J X, Janmey P A, Lyubartsev A *et al.* *Biophys. J.*, 2002, 83 :566
- [ 43 ] Angelini T E, Liang H, Wriggers W *et al.* *Proc. Natl. Acad. Sci. USA*, 2003, 100 :8634
- [ 44 ] Bai Y, Das R, Millett I S *et al.* *Proc. Natl. Acad. Sci. USA*, 2005, 102 :1035
- [ 45 ] Gronbech-Jensen N, Mashl R J, Bruinsma R F *et al.* *Phys. Rev. Lett.* 1997, 78 :2477.
- [ 46 ] Stevens M J. *Phys. Rev. Lett.*, 1999, 82 :101
- [ 47 ] Lyubartsev A P, Nordenskiold L. *J. Phys. Chem.*, 1995, 99 :10373
- [ 48 ] Lee K C, Borukhov I, Gelbart W M *et al.* *Phys. Rev. Lett.*, 2004, 93 :128101
- [ 49 ] Ray J, Manning G S. *Langmuir*, 1994, 10 :2450
- [ 50 ] Murthy V L, Rose G D. *Biochemistry*, 2000, 39 :14365
- [ 51 ] Rouzina I, Bloomfield V A. *J. Phys. Chem.*, 1996, 100 :9977
- [ 52 ] Ha B Y, Liu A J. *Phys. Rev. Lett.*, 1997, 79 :1289
- [ 53 ] Shklovskii B I. *Phys. Rev. Lett.*, 1999, 82 :3268
- [ 54 ] Levin Y, Arenzon J J, Stilck J F. *Phys. Rev. Lett.*, 1999, 83 :2680
- [ 55 ] Craig M E, Crothers D M, Doty P. *J. Mol. Biol.*, 1971, 62 :383
- [ 56 ] Porschke D. *Biophys. Chem.*, 1974, 2 :97
- [ 57 ] Porschke D. *Biophys. Chem.*, 1974, 1 :381
- [ 58 ] Porschke D. *Biophys. Chem.*, 1974, 2 :83
- [ 59 ] Ravetch J, Gralla J, Crothers D M. *Nucleic Acids Res.*, 1974, 1 :109

- [ 60 ] Uhlenbeck O C , Martin F H , Doty P. J. Mol. Biol. , 1971 , 57 : 217
- [ 61 ] Ansari A , Kuznetsov S V , Shen Y. Proc. Natl. Acad. Sci. USA , 2001 , 98 : 7771
- [ 62 ] Bonnet G , Krichevsky O , Libchaber A. Proc. Natl. Acad. Sci. USA , 1998 , 95 : 8602
- [ 63 ] Wallace M I , Ying L , Balasubramanian S *et al.* Proc. Natl. Acad. Sci. USA , 2001 , 98 : 5584
- [ 64 ] Wallace M I , Ying L , Balasubramanian S *et al.* J. Phys. Chem. B , 2001 , 104 : 11551
- [ 65 ] Zarrinkar P P , Williamson J R. Natu. Struct. Biol. , 1996 , 3 : 432
- [ 66 ] Downs W D , Cech T R. RNA , 1996 , 2 : 718
- [ 67 ] Sclavi B , Sullivan M , Chance M R *et al.* Science , 1998 , 279 : 1940
- [ 68 ] Russel R , Millet I S , Tate M W *et al.* Proc. Natl. Acad. Sci. USA , 2002 , 99 : 4266
- [ 69 ] Bokinsky G , Rueda D , Mirsa V K *et al.* Proc. Natl. Acad. Sci. USA , 2003 , 100 : 9302
- [ 70 ] Russel R , Zhuang X , Babcock H P *et al.* Proc. Natl. Acad. Sci. USA , 2002 , 99 : 155
- [ 71 ] Zhuang X , Bartley L E , Babcock H P *et al.* Science , 2000 , 288 : 2048
- [ 72 ] Liphardt J , Onoa B , Smith S B *et al.* Science , 2001 , 292 : 733
- [ 73 ] Sosnick T R , Pan T. Curr. Opin Struct Biol. , 2003 , 13 : 309
- [ 74 ] Triber D K , William J R. Curr. Opin Struct Biol. , 1999 , 9 : 339
- [ 75 ] Thirumalai D , Lee N , Woodson S A. Annu. Rev. Phys. Chem. , 2001 , 52 : 751
- [ 76 ] Flamm C , Fontana W , Hofacker I L *et al.* RNA , 2000 , 6 : 232
- [ 77 ] Isambert H , Siglla E D. Proc. Natl. Acad. Sci. USA , 2000 , 97 : 6515
- [ 78 ] Gullywaev A P , Batenburg F H , Pleij C W. J. Mol. Biol. , 1995 , 250 : 37
- [ 79 ] Galzitskaya O , Finkelstein A V. Fold Des. , 1998 , 3 : 69
- [ 80 ] Sorin E J , Engelhardt M A , Herschlag D *et al.* J. Mol. Biol. , 2002 , 317 : 493
- [ 81 ] Zhang W B , Chen S J. Proc. Natl. Acad. Sci. USA , 2002 , 99 : 1931
- [ 82 ] Zhang W B , Chen S J. J. Chem. Phys. , 2003 , 118 : 3413
- [ 83 ] Zhang W B , Chen S J. J. Chem. Phys. , 2003 , 119 : 8716