

蛋白质与DNA序列上特异和非特异位点结合的研究现况^{*}

刘雪梅^{1,2}, 谢汇章^{1,2}, 艾保全¹, 王先菊³, 刘良钢¹

(1. 中山大学物理系, 广东 广州 510275;

2. 华南理工大学物理系; 广东 广州 510640;

3. 华南师范大学信息光电子科技学院光子中医实验室; 广东 广州 510630)

摘要: 介绍了蛋白质和DNA序列上特异和非特异位点结合的研究状况, 并详细介绍了 Michael Slutsky 和 Maria Barbi 等人提出的有关蛋白质和DNA序列结合的理论模型, 同时对这几种理论模型提出新的展望。

关键词: 特异结合; 非特异结合; 结合能; 阈值

中图分类号: Q274 **文献标识码:** A **文章编号:** B147 (2005) S1-0154-04

蛋白质与DNA的结合已有了广泛的研究, 其中一种重要的研究方法是考虑蛋白质和DNA序列上特异位点和非特异位点的结合方式。蛋白质沿着DNA分子滑动时, 可先随机的结合到DNA的非特异位点, 确定其目标位点, 然后通过沿着DNA的一维扩散搜寻到特异位点并结合。这是由于DNA中的非特异位点可捕获蛋白质并把它们转换到特异位点^[1], 蛋白质和DNA的非特异位点的相互作用在与特异位点的调节和结合过程中起着非常重要的协调作用。蛋白质如何结合到DNA的非特异位点, 又怎样加快了与特异位点的结合? 最近无论是在实验还是理论方面都有了很大的发展。但是仍存在几个方面的问题。

(1) 蛋白质在特异位点和非特异位点是如何切换的。

(2) 还没有一套精确的统计规律来描述蛋白质搜寻特异位点的过程。

(3) 蛋白质和DNA结合过程中, 一方面要求结合速度很快, 另一方面在结合位点要求结合物具有一定的稳定性, 如何解决这种矛盾。

自从1967年 Ptashne^[2] 首次提出蛋白质可结合到DNA序列的特异位点, 就引出一个问题, 蛋白质如何在无数交替的序列中找到它们的目标位点, 而且要求速度很快。随后在1970年, Riggs 等人^[3] 报道了大肠杆菌阻遏蛋白找到其目标位点的速度比受三维扩散控制的搜寻速度还要快1 000倍, 从而引起了许多专家的注意。Von Hippel 等人^[4] 在1989

年提出假设, 蛋白质可沿着DNA链的任何位点结合然后通过扩散过程转换到特异位点。与此同时实验方面也取得了很大的成果, 例如 Kabata 等人^[5] 在1993年用荧光显微镜观察了标有荧光的RNAP (RNA聚合酶) 和噬菌体-DNA之间的相互作用。实验中看到RNAP分子的一个片段偏离了总体流动方向, 并沿着DNA分子的延伸部分移动, 从而揭示了RNAP可沿着DNA的非特异位点滑动。Craig 等人^[6] 在1998年的一篇文章中报道了在大肠杆菌(E. Coli)中, 可用一个单独的酶去完成转录, 即大肠杆菌RNAP, 并提出大肠杆菌RNAP及其它蛋白质在与DNA的特异位点结合时所面临的一个共同的问题是在有许多非特异位点的DNA中如何发现特异位点, 并去越过这一过程中所遇到的障碍。Fichert 等人^[7] 在1999年提出“节间的转移”模型, 来解释这一问题, 即蛋白质可通过一个中间媒介(它可在短时间结合到两个位点)从一个位点转移到另一个位点。可惜这种思想只能应用到两种DNA结合位点, 即乳糖阻遏子和核酸内切酶。

Michael Slutsky 和 Maria Barbi 等人近期在这方面做了大量的研究, 并建立了几种理论模型, 在此作一详细介绍。

1 “快速而稳定”的结合机制

这一机制是由 Michael 等人^[9] 在近期提出的。

在识别过程中, 一个蛋白质能在一段很长的DNA分子上找到它的目标位点并牢固地结合。

* 收稿日期: 2004-11-08

基金项目: 国家自然科学基金资助项目(10275099)

作者简介: 刘雪梅, (1975年生), 女, 讲师, 博士研究生; 通讯联系人: 刘良钢; E-mail: stdp05@zsu.edu.cn

Michael Slutsky 等人的目标是寻找出一种机制, 即能快速搜寻 ($1 \sim 10 \text{ s}$), 又能牢固结合 ($K_d = 10^{-15} - 10^{-8} \text{ M}$)).

1.1 结合能

当蛋白质沿着 DNA 扩散时, 蛋白质在遇到每个结合位点 r 处经历了一个结合势能 $U(r)$ 。蛋白质和 DNA 的相互作用分为两部分—特异和非特异位点的结合, 有如下关系。

$$U = U(r = r_i, \cdots r_{i+l-1}) + U_w \tag{1}$$

其中, 非特异结合势能 U_w 取决于蛋白质和 DNA 的相互作用而不是蛋白质将要结合的 DNA 序列, 例如蛋白质与 DNA 磷酸骨架的相互作用, 特异位点的结合势能对 DNA 序列的依赖性很强。如图 1^[9]。

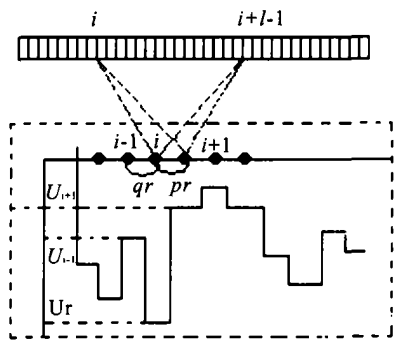


图 1 势能模型
Fig 1 The potential model

如图 1 为一段长为 l 的 DNA 分子, 这段 DNA 分子可被看成是一个一维的 $\{r_i\}$ 排列位点, 从位点 r_i 到位点 r_{i+l-1} 的碱基对组成了一定的结合序列。从任意一个位点 r_i , 跃迁到位点 r_{i+1} 的概率为 p_i , 跃迁到位点 r_{i-1} 的概率为 q_i 。 p_i 和 q_i 取决于蛋白质和 DNA 在位点 r_i 处特异性结合势能 U_i 和在位点 r_{i+1} 或 r_{i-1} 处的势能 U_{i+1} 或 U_{i-1} 。并满足如下形式:

$$p_i \propto e^{-\beta(U_{i+1} - U_i)} \tag{2}$$

$$q_i \propto e^{-\beta(U_{i-1} - U_i)} \tag{3}$$

其中 $\beta = (k_B T)^{-1}$, k_B 是玻尔兹曼常数, T 是绝对温度, 从而定义了一个与位点有关的随机行走跃迁概率。同时得出如下三点结论:

- (1) 当蛋白质和 DNA 的结合能大于时, 蛋白质沿着 DNA 的扩散变得非常慢, 蛋白质的扩散一般不能超过几个碱基对, 整个结合过程很慢。
- (2) 假如结合过程是由 1 维和 3 维扩散组成, 蛋白质与 DNA 序列上非特异位点的结合在协调这

两种扩散的平衡中起着重要的作用。蛋白质与非特异位点的最佳结合能可使整个结合速度达到最大, 虽然比单独的 1 维和 3 维扩散的都快, 但是如果结合能大于 $2k_B T$ 时, 1 维和 3 维的最佳组合不能够加快蛋白质的搜寻速度。

(3) 矛盾的是, 蛋白质与 DNA 的结合体在目标位点的稳定性要求该处的结合势能相对大于 $k_B T$, 然而在这样一个能量下通过 1 维或 3 维的快速搜寻是不能达到的。

1.2 蛋白质在特异和非特异位点的切换

蛋白质在与 DNA 序列的结合过程中, 一方面要求以最快的速度找到目标位点, 既要提高其搜寻速度, 另一方面又要求在结合位点, 蛋白质和 DNA 的结合物具有一定的稳定性, BWH^[4] (Berg, Winter and von Hippel) 在 1989 曾预言“搜寻速度—稳定性”是自相矛盾的, 他们认为在“特异”和“非特异”结合模型之间可能存在某些类型的构象改变来切换模型。在非特异模型中, 蛋白质只是在 DNA 的一个等势能面 ($U=0$) 上“滑动”, 只有在“特异”模型中才与位点结合 ($E \gg k_B T$), 蛋白质在非特异结合模型中是不能够意识到它所要结合的 DNA 序列。因此, 它应该在结合模型中转换, 搜索到特异位点。这种模型很自然地引出一个问题就是有关构象改变的性质。假如蛋白质在 DNA 序列中搜寻到每个特异位点, 它将在裸露的位点上停留数小时。假如有一种方法可直接探测到 DNA 中的特异位点, 即只有那些具有高势能的特异位点, 那么搜寻时间将会大幅度下降。蛋白质与那些势能相对弱的特异位点的相互作用 (例如平滑面 ($E > k_B T$)) 对 DNA 的扩散特征及总的搜寻时间的影响并不明显。但是, 在这个平面, 联系实际特异结合能量分布 ($E \sim 5 - 6 k_B T$), 这些特异位点将是两种模型中最强的点, 蛋白质的构象一般在这些位点发生改变。

他们由此建立了两种蛋白质—DNA 蹬结合模型: 搜寻模型和识别模型。在搜寻模型中, 蛋白质的构象只允许与那些结合能相对弱的特异位点 ($E \sim 1.0 \cdots 1.5 k_B T$) 进行相互作用, 在识别模型中, 蛋白质的构象改变结束, 与 DNA 的相互作用非常强 ($E \geq 5 k_B T$)。这两种能量分布有很强的相关性, 因此在搜寻模型中, 具有最低结合能的位点 ($E \leq 5 k_B T$) 可能对应识别模型中最强的位点, 即特异位点。当蛋白质在 DNA 上结合能较低的位点被捕捉

时, 这两种模型将发生切换。

2 与 DNA 序列排列有关的扩散模型

Maria Barbi 等人^[10]介绍了蛋白质在 DNA 分子上滑动并搜寻目标位点时的概率模型, 这种模型可解释蛋白质和 DNA 结合时由于受 DNA 序列顺序的影响而可能出现的问题。它主要针对 T7RNA 聚合酶。

碱基对之间是通过氢键相连的, DNA 和蛋白质也是通过氢键相互作用的, 且氢键的长度可达到 35 nm。这个值大约为两个碱基对之间氢键的长度。因此, RNAP 与 DNA 相互作用时可直接从一个位点转换到另一个位点, 在这个过程中不需要活化能量。相反地, 假如 RNAP 从一个位点转换到另一个位点需要部分地或全部地毁掉氢键, 它需要克服一个附加的活化障碍。也就是说, RNAP 的内部具有一定的弹性使得构象容易发生改变。这种改变取决于结合物的稳定性程度。例如, 如果发现太多的失配, RNAP 就有可能从“阅读”模型到“滑动”模型, 在这种情况下, 不能形成氢键, 在这种状况下, 存在几个状态的模型。蛋白质在跃迁时需要克服一个阈值为 E_t 的能量障碍。假如总的跃迁能量 $E(n)$ 超过了阈值能量 $E(t)$, 系统虽然经历了一系列不同的状态, 能量始终为 E_{st} , 这时 RNAP 可自由滑动。

如图 2^[10], 其中点线为阈值分界线, 图 I 中阈值为 0, 图 II 中阈值为最大值, 图 III 和图 IV 中阈值分解线为中间的点线。在图 IV 中阈值线以上的值都为常数 E_{st} 。即实线部分这时 $E(n) = E_{st}$ 。

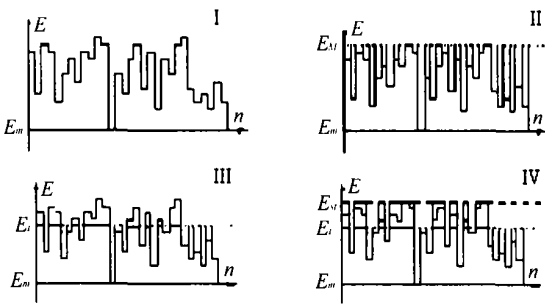


图 2 四种变量模型

Fig 2 A schematic picture of the four considered variants of the model

(1) 阈值为 0 时的模型 (如图 2, I)。

这时 RNAP 不需要破坏掉氢键就能直接从一个位点转换到另一个位点。在这种状态下, 蛋白质从

DNA 位点 n 转换到点 $n' = n \pm 1$ 的能量差 $\Delta E_{n \rightarrow n'}$ 为

$$\Delta E_{n \rightarrow n'} = \max [E(n') - E(n), 0] \quad (4)$$

当 $E(n') - E(n)$ 为负值时, $\Delta E_{n \rightarrow n'}$ 等于 0。

(2) 阈值为最大时的模型 (图 2, II)。

为了转换到另一个位点, RNAP 必须破坏掉所有氢键并经历一个“全部失配”的状态, 在这种状态下

$$\Delta E_{n \rightarrow n'} = E_M - E(n) \quad (5)$$

其中 $E_M = \max [E(n)]$ (6)

(3) 阈值介于零和最大时的模型 (图 2, III)。

为了转换到这一位点, RNAP 必须破坏掉所有氢键并经历一个由 E_t 决定的中间“0”状态。在这种状态下

$$\Delta E_{n \rightarrow n'} = \max [E_t - E(n), E(n') - E(n), 0] \quad (7)$$

模型 I 和模型 II 是模型 III 的两个特例, 分别对应阈值为 0 和最大值时。

(4) 阈值的两态模型 (图 2, IV)。

在图 IV 中, E_t 为分界处, 其中 $E(n) < E_t$ 定义为“阅读”域, $E(n) > E_t$ 定义为“滑动”域。在“阅读”域, 蛋白质如图 III, 需要克服阈值, 但在“滑动”域, 它不需要克服阈值, 即不需要破坏掉氢键而自由的滑动, 而能量始终为 $E_M = [E(n)]$ 。能量如下分布:

$$E(n) = \begin{cases} E(n) & E(n) < E_t \\ E_{st} & E(n) > E_t \end{cases} \quad (8)$$

$\Delta E_{n \rightarrow n'}$ 与模型 III 的定义形式相同。

如果发现太多失配, 蛋白质和 DNA 的特异位点的结合可通过一个构象改变“切换掉”。这时 RNAP 更大程度上被认为是在一个“滑动”模型下进行, 这种状态下氢键没有被激活 n 与 n' 之间的跃迁率 $r_{n \rightarrow n'}$ 与 $\exp(-\Delta E_{n \rightarrow n'} / k_B T)$ 成比例, 当 $n' = n \pm 1$ 时, 这个模型包括一个零概率 (也就是 RNAP 停留在某个位置上), 因此一套完整的跃迁率定义如下:

$$\begin{cases} r_{n \rightarrow n'} = 1/2 \exp(-\Delta E_{n \rightarrow n'} / k_B T) & n' = n \pm 1 \\ r_{n \rightarrow n'} = 1 - r_{n \rightarrow n+1} - r_{n \rightarrow n-1} \end{cases} \quad (9)$$

在平滑的结合能分布下 ($\Delta E_{n \rightarrow n'} = 0$) 所有的概率为 1/2。

Maria Barbi 等人的不足是只针对一种 RNA 聚合酶, 而对其他蛋白质并不通用。

3 结 语

Michael Slutsky 等人进一步完善了蛋白质和

DNA 相互作用的定量描述, 另外从能量的角度分析了蛋白质在特异位点的跃迁概率, 用构象改变来解释蛋白质在特异和非特异位点的切换。这为蛋白质在特异和非特异位点相互交替的 DNA 序列中如何滑动和结合有了理论上的完善, 而且与实验数据也相吻合。但还是缺乏蛋白质和非特异位点结合的实验数据, 尤其是有关结构方面的数据, 这些问题仍然处于一种理论阶段, 还需进一步讨论。

Maria Barbi 等人介绍了 RNAP 在 DNA 上滑动时的一个简单的概率模型, 并考虑了 DNA 的排列顺序(特异和非特异位点的排列)影响到 RNAP 与 DNA 的相互作用。提出了蛋白质在 DNA 上滑动的四种可能的概率模型, 用概率统计分析解释了蛋白质在与 DNA 序列的排列有关的位点上如何识别特异位点, 并且通过区别“阅读”域与“滑动”域解释了由于没有氢键的形成而导致 RNAP 可在一个有效的固定的势能面上自由滑动(模型 IV)。这种理论可充分的解释蛋白质在 DNA 特异和非特异位点上的滑动和搜寻, 但其局限是只考虑了 T7 RNA 聚合酶, 对其他蛋白质还需进一步探讨。

参考文献:

[1] SHIMAMOTO N. One— dimensional diffusion of proteins

along DNA [J] . Biol Chem, 1999, 274: 15293.
[2] PIASHNE M. Specific binding of phage repressor [J] . Nature, 1967, 214: 232.
[3] RIGGS A D, BOURGEOIS S, COHN M. The lac repressor — operator interaction[J] . Mol Biol, 1970, 53: 401.
[4] VON HIPPEL P H, BERG O G. Facilitated target location in biological systems [J] . Biol Chem, 1989, 264: 675.
[5] KABATA H O. Visualization of single molecules of RNA polymerase sliding along DNA[J] . Science, 1993, 262: 1561.
[6] CRAIG M L. DNA footprints of the two kinetically significant intermediates in formation of an RNA polymerase — promoter open complex: evidence that interactions with start site and downstream DNA induce sequential conformational changes in polymerase and DNA[J] . Mol Biol, 1998, 283: 741.
[7] FICKERT R, MÜLLER HILL B. How lac repressor finds lac operator in vitro[J] . Mol Biol, 1992, 226: 59.
[8] ROBIJIN F, BRUINSMA. Physics of protein — DNA interaction[J] . Physica A, 2002, 313: 211.
[9] MICHAEL SLUTSKY, LEONID A, MIRNY. How does a protein find its site on DNA[C] . Q— bio . BM/0402005 3 Feb 2004.
[10] MARIR BARBI. A model of sequence dependent protein diffusion along DNA[C] . arXiv: physics/ 0112007 12 Jul 2002.

The Development Situation of Protein Binding Specific or Non-specific Sites on DNA Sequence

LIU Xue-mei^{1,2}, XIE Hui-zhang^{1,2}, AI Bao-quan¹, WANG xian-ju³, LIU Liang-gang¹

- (1. Department of Physics, Zhongshan University, Guangzhou 510275, China;
2. Department of Physics, South China University of Technology, Guangzhou 510640, China;
3. Lab of photonic TCM, School for Information and Optoelectronic Science and Engineering, South China Normal University, Guangzhou 510631, China)

Abstract: The developement of protein binding with specific and non-specific sites on DNA sequence is introduced, the lakely work of Michael Slutsky and Maria Barbi of theoretical models of profein binding with specific and non-specific sifes on DNA sequend is also discribed, and the new prospect of these models is presented.

Key words: specific binding; non-specific binding; binding energy; threshold