

DNA 超螺旋构象的理论研究

郝 鸣 鸿

(山西大学分子科学研究所, 太原)

提 要

本文总结了近年来描述 DNA 超螺旋结构的基本关系式和各种参量, 综述了计算超螺旋的几何参数, 定量分析 DNA 分子与蛋白质结合时的特殊构象, 以及基于物理模型解释和预见精确的超螺旋空间构象等方面的理论研究进展。

关键词 旋转数, 扭曲数, 表面缠绕, 弹性模型, 计算机模拟

一、引 言

DNA 分子最常见的二级结构是由两条互补的聚核苷酸链相互缠绕而形成的双股螺旋。当双螺旋的两端受限制而无法自由转动时, 例如在环状分子或 DNA 与蛋白质的络合物中, 细长柔韧的 DNA 就常常形成特殊的高级结构, 即所谓的超螺旋 (Supercoil 或 Superhelix)^[1]。

超螺旋的性质对于 DNA 分子的生物功能具有重要的影响。这是由于所有质粒和原核生物物的 DNA 都是环形的分子, 而真核生物的染色质则是 DNA 与蛋白质的络合物。超螺旋在这些 DNA 中是普遍存在的。DNA 的复制和转录通常要求模板双螺旋的部分解绕。由于超螺旋能够帮助或阻碍双螺旋的解绕, 它直接影响 DNA 的遗传过程^[2]。超螺旋的存在也对 DNA 分子的局部结构产生张力, 它能推动双螺旋的局部构象发生变化, 如形成 Z-DNA、发夹和十字花构象等^[3]。这些局部结构的变化也影响着 DNA 的生物功能。

DNA 超螺旋结构的某些性质通过理论研究已经得到了深入的理解。这一领域的研究取得了相当的进展, 某些结果也在指导实验工作中发挥着作用。这是当前在核酸研究领域相当

活跃的一个方面。本文拟对 DNA 超螺旋的理论研究及进展作一综述。

二、基本参数及相互关系式

环 DNA 的基本特征之一是双螺旋的两股链相互纠缠在一起; 只要环不断开, 两股链就无法完全分开^[3]。环 DNA 中两条链之间的缠绕数是一整数, 称为连环数 (Linking Number, Lk) 它可以解析地表达为:

$$Lk = \frac{1}{4\pi} \oint_{c_1} \oint_{c_2} \frac{[d\mathbf{r}_1 \times d\mathbf{r}_2] \cdot (\mathbf{r}_1 - \mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|^3} \quad (1)$$

式中 $\mathbf{r}_1$ 和 $\mathbf{r}_2$ 分别是两条闭合链 (曲线) c_1 和 c_2 的矢量函数表达式。大量的实验已经证明天然环 DNA 的连环数一般不等于相同长度而完全松弛后的环 DNA 的连环数, 这个差值称为连环数差, 简记为 ΔLk 。现在已知大多数天然环 DNA 的 ΔLk 都为负值^[4]。连环数差对环 DNA 的影响与超螺旋密度 σ 有关, $\sigma = \Delta Lk / Lk^0$, 它是连环数差与相同长度 DNA 的自然连环数的比值^[1]。

定量地研究连环数差对 DNA 结构的影响需要定义双螺旋结构的局部旋转率 (Rate of Twist)。最简单的情况是一个细杆的局部旋转率 ω 。设 $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$ 为沿着细杆轴线的流动坐

标系, e_1 位于杆轴线的切线方向, 则细杆的横截面的局部旋转率定义为^[4]:

$$\omega = e_3 \cdot de_2 = -e_2 \cdot de_3 \quad (2)$$

上述定义可以推广到双螺旋的局部结构中去。在描述 DNA 结构时, 常常用螺旋重复数 h 。在某一局部范围内, h 等于 ω 由零积分到 2π 角度时所经过的碱基对数目。双螺旋的局部旋转率沿着分子总长度的积分称为旋转数 (Twisting Number, Tw)^[5]

$$Tw = \frac{1}{2\pi} \oint \omega dl \quad (3)$$

当整个环 DNA 位于一个平面内时, 旋转数等于连环数。这反映了 Tw 和 Lk 都与双螺旋的缠绕联系着。但是在具有三维结构的超螺旋中, Lk 一般不等于 Tw 。Fuller 严格地定义了 Lk 和 Tw 的差别^[6]; 二者的差值称为扭曲数 (Writhe, Wr)

$$Wr = Lk - Tw \quad (4)$$

Crick 也对上述关系式作过详细地研究^[3]。 Wr 是一条空间曲线, 如双螺旋的轴线的函数。因为它与超螺旋的扭曲程度和手性直接相联系, Wr 也被不精确地称为“超螺旋数”。(4) 式是环 DNA 普遍遵守的基本关系式。它的重要意义在于将两个描述超螺旋的几何参数 Tw 和 Wr 与一个拓扑数 Lk 联系了起来, 为定量分析超螺旋的结构提供了一个基础。近年来关于 DNA 超螺旋的理论分析在某种程度上都与 (4) 式联系着^[4,5,7-13,15,19,20]。

三、超螺旋几何参数的计算

扭曲数 Wr 和旋转数 Tw 是两个决定超螺旋几何形状的重要参数, 它们的应用依赖于是否能从理论上将其方便地计算出来。所以, 近年来人们发展了数种计算 Wr 和 Tw 的方法。

计算 Wr 的第一个方法是将一个空间曲线沿着不同的方向平面投影, 然后计算这些投影曲线的方向扭曲数^[6]。原空间曲线的扭曲数 Wr 是所有方向扭曲数的平均值。但是在实际计算中这种方法是很不方便的, 需要取很大数目的投影方向才能得到 Wr 的可靠数值。Le-

vitt 修改了这种方法^[7]。他将给定的链分子按其化学结构单元进行分割。设每一单元用一虚键矢量 l_i 来表示。连接两个矢量 l_i 和 l_j 的中心作矢径 r_{ij} , 将 l_i 和 l_j 再投影到以 $|r_{ij}|$ 为半径的球面上。以 l_i 和 l_j 为边的平行四边形在球面上截得面积元 ΔS , 对应于立体角 $\Delta\Omega_{ij} = \frac{\Delta S}{|r_{ij}|^2}$ 。该曲线的扭曲数则为:

$$Wr = \sum_i \sum_j \Delta\Omega_{ij} \quad (5)$$

根据 (5) 式计算 Wr 可以免去对 Wr 没有贡献的那些方向求和, 又可按分子的化学结构来决定求和项数及计算精确度, 具有计算上的优越性。

Fuller 指出若将一条闭合空间曲线上每一点的单位切线矢量平移到坐标原点, 这些矢量的端点将在单位球面上划一闭合曲线。令该曲线包围的球面面积为 A , 则原空间曲线的扭曲数与 $A/2\pi$ 只相差一个整数。因为 Wr 在一定区域上是连续的, 所以利用上述几何性质很容易计算一些规则超螺旋的扭曲数^[8,9]。Fuller 又证明了两条曲线之间的扭曲数差可以用下式计算^[8]:

$$Wr_2 - Wr_1 = \frac{1}{2\pi} \int \frac{(t_2 \times t_1) \cdot \frac{d}{ds}(t_2 - t_1)}{1 + t_2 \cdot t_1} ds \quad (6)$$

其中 t_1 和 t_2 分别为两条曲线的单位切线矢量, s 为曲线参数方程中的参数。

曲线的扭曲数还可以用高斯积分计算^[10]:

$$Wr = \frac{1}{4\pi} \iint \frac{[t(l_1) \times t(l_2)] \cdot [r(l_1) - r(l_2)]}{|r(l_1) - r(l_2)|^3} dl_1 dl_2 \quad (7)$$

式中 $r(l)$ 为曲线的矢量表达式, $t(l)$ 为曲线的切线矢量, l 为弧长参量。(7) 式一般需用数值积分计算。许多典型超螺旋的扭曲数都已用该法计算过^[11]。

DNA 双螺旋有时能够局部地分离; DNA 转录时的中间体以及在某些特殊的 DNA 中还存在着三股螺旋。分析这类 DNA 的局部旋转

率和总旋转数具有重要的理论和实际意义。White 和 Bauer 最近在这方面的研究中取得了一定的进展。考虑图 1 中的两条曲线 A 和 C^[9]。假设 A 与 C 是对应的,即在 A 上任取一点

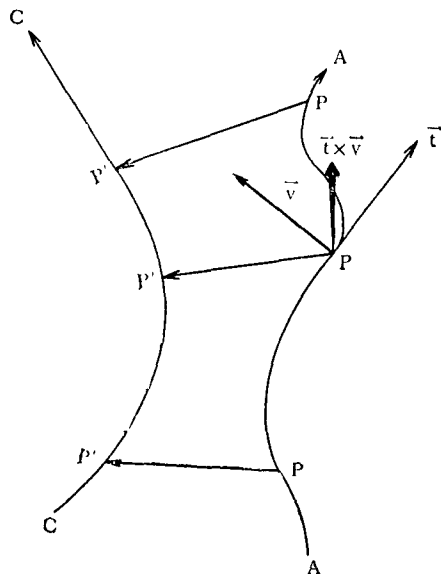


图 1 根据曲线 A 和 C 之间的对应矢量计算其旋转数^[9]

t 为曲线 A 在 P 点的单位切线矢量, pp' 为曲线 A 和 C 之间的对应矢量, v 位于 t 和 pp' 决定的平面内并垂直于 t

P, 则在 C 上对应一点 P'。矢量 $pp' = p - p'$ 称为 A 与 C 的对应矢量。设在 P 点曲线 A 的单位切线矢量为 t , 在矢量 pp' 和 t 定义的平面内可找到一垂直于 t 并与 pp' 成锐角的单位矢量 v 。曲线 C 相对于 A 在 P 点的旋转率可定义为 $\omega = (t \times v) \cdot dv$, C 相对于 A 的总旋转数则可计算为:

$$Tw(C, A) = \frac{1}{2\pi} \int (t \times v) \cdot dv \quad (8)$$

值得注意的是一般 $Tw(C, A) \neq Tw(A, C)$ 。

进一步来分析三股螺旋之间的缠绕问题^[12]。设 c_1 、 c_2 和 c_3 为三条分离的空间曲线, c_1 和 c_2 相对于 c_3 的旋转数分别为 $Tw(c_1, c_3)$ 和 $Tw(c_2, c_3)$ 。一个重要的问题是 $Tw(c_1, c_3) - Tw(c_2, c_3)$ 与三条曲线之间几何形状的关系。类似于计算两条曲线之间的旋转率, 在曲线 c_3 上任意一点 P 可找到曲线 c_1 和 c_2 分别

于 c_3 的对应矢量 v_{13} 和 v_{23} 。根据上述的定义可使 v_{13} 和 v_{23} 分别垂直于在 P 点曲线 c_3 的切线矢量。因此 v_{13} 和 v_{23} 都位于 c_3 的法平面内。令 v_{13} 和 v_{23} 之间的夹角为 ϕ , 可以证明^[12]:

$$Tw(c_1, c_3) - Tw(c_2, c_3) = \frac{1}{2\pi} \int d\phi = (\phi_i - \phi_f)/2\pi \quad (9)$$

式中 ϕ_i 和 ϕ_f 分别为夹角 ϕ 在起点和终点的数值。根据以上结果, 当 DNA 发生局部解旋时, 若其起点和终点的构型被固定不变, 则总旋转数将保持不变。这是直观想象上难以预计的结果。

当给定分子长度和连环数差时, 超螺旋的实际构象是由 ΔLk 在 Wr 和 Tw 之间的分配所决定的。环 DNA 的连环数差可以由实验方法精确测定, 但 Wr 和 Tw 一般确不能够测定。关于超螺旋几何参数的理论计算方法已被应用于分析一定的连环数差在环 DNA 的 Wr 和 Tw 之间分配的问题^[10, 13]。

四、缠绕在蛋白质表面上的 DNA 超螺旋构象

当 DNA 分子缠绕在蛋白质表面上时, 其局部几何结构要发生变化。此时双螺旋的构象能够相对于蛋白质的表面来加以描述。核酸酶 (DNaseI) 切割 DNA 的实验为这种描述方法提供了实验基础。DNaseI 非选择地切割 DNA 的磷酸二脂键。当 DNA 缠绕在蛋白质表面上时, 朝向蛋白质的那些磷酸二脂键受到保护作用因而具有最小的被切割几率。当沿着双螺旋的一股描述其被 DNaseI 的切割几率时, 得到的是一条振荡曲线。凭着直觉, 并且也在一段时间内, 这样测得的曲线的振荡周期被作为缠绕在蛋白质表面上的 DNA 的螺旋重复数。最近, White 等的理论分析澄清了这个问题^[15]。

DNA 在蛋白质表面缠绕的几何关系可用图 2 来描述。White 等指出核酸酶的切割几率曲线的振荡频率实际上是对应矢量 pp' 相对于表面法矢量 n_p 的旋转速率。如果 n_p 的方向是固定的, 例如表面是平面时, 实验测定的振荡周

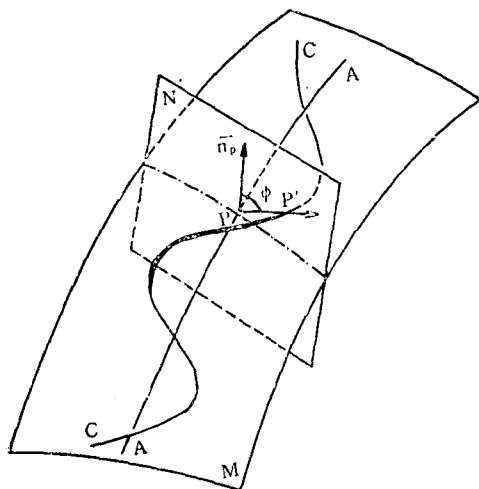


图2 DNA 双螺旋在一个表面上的缠绕^[13]

设 A 为 DNA 的轴线, C 为双螺旋的一股, M 为一个曲面, 例如蛋白质的表面。图中曲线 A 完全位于曲面 M 上, n_p 为曲线 A 的 P 点上曲面 M 的法向矢量, pp' 为曲线 A 与 C 之间的对应矢量, ϕ 是 pp' 与 n_p 之间的夹角

期确实等于 DNA 的螺旋重复数。但若 n_p 的方向在曲线 A 上处处变化时, 实验测定的振荡周期就不再简单地等于在蛋白质表面上的 DNA 螺旋重复数了。法线 n_p 的端点描述一条连续的曲线; 设这条曲线相对于曲线 A (DNA 的轴线) 的旋转数为 S_{Tw} , 而它们之间的连环数 (称为表面连环数) 为 S_{Lk} 。若实验测得的 DNA 被切割的几率曲线振动了 Φ 次, 则可证明曲线 C 相对于曲线 A 的旋转数为 $Tw(C, A) = S_{Tw} + \Phi$ 。再利用公式(4), 即得下面关系式:

$$Lk = S_{Tw} + \Phi + Wr = S_{Lk} + \Phi \quad (10)$$

上式的重要意义在于 Φ 是一个实验可测数, S_{Lk} 能够从已知的蛋白质表面形状计算出来, 因此通过(10)式可以计算在 DNA 与蛋白质络合物中 DNA 的连环数。

例如染色质就是 DNA 与蛋白质的络合物。图 3 是一个微型染色质模型。它的结构是由一系列称为核小体的单元所组成。在每一个核小体中, 大约 140 个碱基对的 DNA 以左手超螺旋绕了大约 1.8 圈。关于染色质结构, 曾经有一个所谓的“连环数佯谬”^[24]。因为染色质环 DNA 的连环数差完全是由表面连环数所决

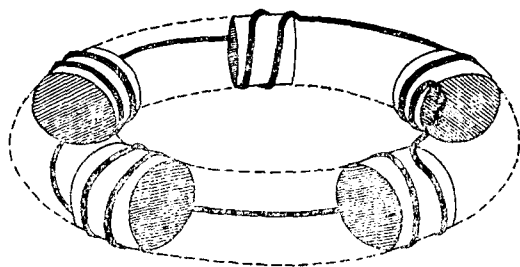


图3 微型染色质模型

每一个圆柱体表示一个核小体 (Nucleosome)。在每一个核小体中, DNA 以左手超螺旋缠绕了大约 1.8 圈

定的, 而根据图 3 可以估计出每一个核小体对表面连环数的贡献为 -1.8 , 所以这类 DNA 的连环数差应为 $-1.8m$, 假定 m 为核小体的总数。但是实验确发现, 在溶液中脱掉蛋白质后测得的 DNA 连环数差只有 $-1.1m$ 。因为在脱蛋白质过程中环 DNA 并没有开环, 上述实验结果与环 DNA 连环数差不变的规则是相抵触的。解决这个矛盾可以利用 DNA 在蛋白质表面上缠绕的理论分析^[15]。

设在溶液中环 DNA 的超螺旋密度为 σ , DNA 的自然螺旋重复数为 h_0 , 而在染色质中 DNA 的螺旋重复数为 h 。根据 (10) 式以及 h 和 σ 的定义可以推导出下面的关系式^[15]:

$$h = h_0 / (\sigma - S_{Lk} / Lk^0 + 1) \quad (11)$$

以 SV40 病毒 DNA 为例: 它有 5243 个碱基对, 含有 26 ± 2 个核小体单元, 实验测得在溶液中该 DNA 脱去蛋白质后的超螺旋密度 $\sigma = -0.053$ 。取在溶液中 DNA 的自然螺旋重复数为 $h_0 = 10.6 \pm 0.1$, 则 $Lk^0 = 5243 / 10.6$ 。将这些数据代入(11)式即可算得 $h = 10.17$ 。其结论是在核小体中 DNA 的螺旋重复数 h 比 h_0 为小, 只有 10.17。这个结果与已知的核小体结构数据相符合。因此连环数佯谬并非真正的矛盾, 只是与蛋白质结合时 DNA 结构发生变化的结果。

五、DNA 超螺旋构象的弹性力学模型

目前已知的解释 DNA 超螺旋构象的理论大多基于下面的简化模型^[9,16]: DNA 被近似为

细长的弹性杆，具有圆形的横截面和各向同性的力学性质，在未畸变状态下杆是笔直的并具有均一的局部旋转率。在超螺旋结构中杆发生了形变：同时存在着弯曲和扭转。根据经典弹性力学理论，超螺旋形状的杆的总弹性能可表示为：

$$E = \frac{A}{2} \int \kappa^2 dl + \frac{C}{2} \int (\omega - \omega_0)^2 dl \quad (12)$$

式中 A 和 C 分别为杆的弯曲和扭转力常数， κ 为杆轴线的曲率， ω 和 ω_0 分别是杆横截面的实际和自然旋转率。上述模型应该看做是对实际 DNA 分子的一级近似。它基本上忽略了 DNA 的详细分子结构，并假定分子只是在局部上发生了很小的畸变（这不妨碍很长的分子在整体构象上发生很大的形变），因此对能量的主要贡献是形变能的线性项。

根据弹性力学理论，在杆的构型达到平衡状态时，作用在杆的每一个横截面上的力和力矩必须与杆形变产生的内张力相平衡。平衡条件可以用微分方程来描述；在适当的边界条件下，求解这种方程即可得到杆的平衡构型。例如，设 θ ， ϕ 和 ψ 是将杆的每一个横截面上的局部坐标系变换到固定坐标系中的三个欧拉角，根据杆的平衡条件可以导出下列微分方程^[17]：

$$\frac{d\phi}{dl} = \frac{p - qu}{A(1 - u^2)} \quad (13a)$$

$$\left(\frac{du}{dl}\right)^2 = \frac{2}{A} (e - Nu)(1 - u^2) - \frac{(p - qu)^2}{A^2} \quad (13b)$$

式中 A 的定义同(12)式， $u = \cos\theta$ ， p ， q ， e 和 N 是由边值条件所决定的常数。关于 ψ 的方程与杆的空间构型无关所以被忽略了。由(13)式而解得 θ 和 ϕ 随着杆长度变化的函数称为杆的指示线，它们决定了杆轴线在每一点的单位切线矢量的方向。对指示线积分即得到杆的空间构型。Benham 以及后来多个作者将上述分析方法应用于 DNA 超螺旋的平衡构象的问题^[16-19]，即将 DNA 分子处理为一个细长的弹性杆。这些作者的结果证明了根据一级弹性力

学近似能够预见 DNA 超螺旋构象的存在。例如当杆的两端被限制时其平衡的杆轴线空间构象为螺旋状，而环形杆的平衡构象类似于环形的螺旋圈。这种处理方法目前已知的问题有：(1) 满足环合条件的一般解难以得到；(2) 当考虑实际 DNA 中存在的链间相互作用时方程变得过分复杂而无法求解。

研究超螺旋构象的另一种方法是极小化超螺旋的总势能。极小化势能函数(12)式可以分两步进行：首先在一给定空间构象的条件下极小化扭转能部分(ΔLk 固定)，这一步处理使势能函数简化为^[9,61]：

$$E = \frac{A}{2} \int \kappa^2 dl + \frac{2\pi^2 C}{L} (\Delta Lk - Wr)^2 \quad (14)$$

式中 L 是分子的总长度；再进一步极小化(14)式即得到超螺旋的平衡构象。但(14)式是超螺旋构象的泛函，直接极小化是非常困难的。Fuller 曾对这个问题作过定性的讨论^[6]。最近有人用变分法研究了上述问题^[9,19]。他们将超螺旋构象表示为螺旋几何参数，例如螺旋半径、圈数以及螺距角的函数，变化这些几何参数而使总能量变为最小。这种方法在研究一类给定的构象时是有用的。但是由于可能选取的变分函数的限制，得到的变分结果是比较局限的。

Hao 和 Olson 最近提出了一个新的数值

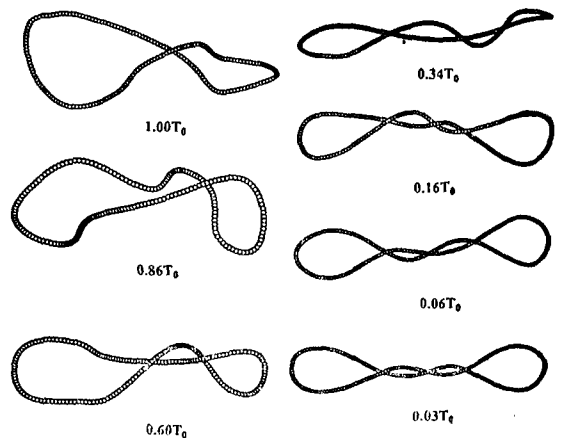


图4 计算机模拟的 DNA 超螺旋构象随着温度变化而变化的过程^[20]

T_0 为起始模拟温度， T_0 乘一个系数是降低的温度。连环数差固定在 4

方法^[20]。他们用 B-Spline 函数来表示 DNA 超螺旋的空间轨迹。这种方法的优点是可以将超螺旋的形变能以及链间的范德华作用能表示为坐标的显函数。求解能量最低构象的步骤如下：从一个随机的超螺旋构象出发，在某一给定温度下用 Monte Carlo 方法模拟 DNA 分子的热运动；通过逐渐降低模拟的温度，超螺旋构象最终被冻结在整体构象的能量最小值点上。这种极小化方法又称为“模拟退火” (Simulated Annealing)。图 4 是在计算机模拟过程

时环 DNA 的能量最低构象。通过上述处理方法建立了如图 5 所示的环 DNA 连环数差与超螺旋构象的定性关系：当 $\Delta Lk = 0$ 或 1 时，最稳定的构象是平面圆形(图 5b 和 c)；当 $\Delta Lk = 2$ 时，最稳定的构象是“8”字形 (图 5d)；当 $\Delta Lk \geq 3$ 时，最稳定的构象是一系列返回式超螺旋结构(图 5e—g)，其螺旋圈数随 ΔLk 线性增加。上述结果与在电子显微镜下观察到的环 DNA 超螺旋构象是定性一致。这些结果证明了关于 DNA 超螺旋的弹性理论至少在一级近似下是正确的。

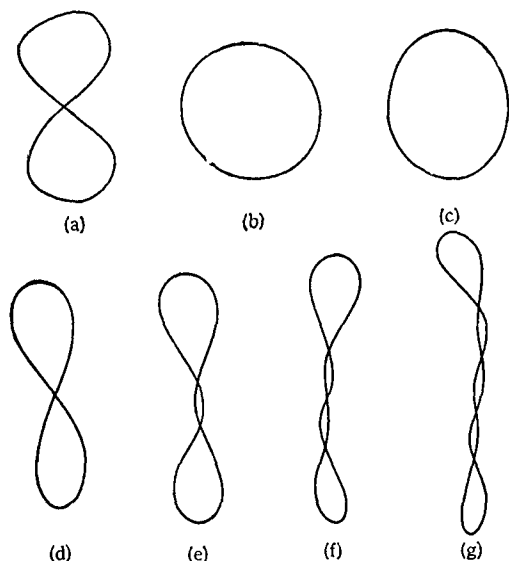


图 5 理论预见的在不同连环数差时 DNA 超螺旋的最稳定构象

(a) 为任意选择的起始构象，其余均为从(a)出发经过计算机模拟而得到的能量最低构象。(b)、(c)，(d)、(e)、(f) 和 (g) 分别为 $\Delta Lk = 0, 1, 2, 3, 4$ 和 5 的情况

中，一个连环数差为 4 的环 DNA 从高温时的随机形状，随着温度的降低逐渐接近于返回式超螺旋构象的过程。最终的结构是 $\Delta Lk = 4$

参 考 文 献

- 1 Wang J C In: Semlyen J ed, *Cyclic Polymers*, Elsevier Applied Science Publishers LTD, 1986: 320--360
- 2 Wu H -Y et al. *Cell*, 1988; 53: 433
- 3 Crick F. *Proc Natl Acad Sci USA*, 1976; 73: 2639
- 4 Pohl W. *Math Intelligencer*, 1980; 2: 20
- 5 White J, Bauer W J. *J Mol Biol*, 1986; 189: 324
- 6 Fuller F. *Proc Natl Acad Sci USA*, 1971; 68: 815
- 7 Levitt M. *J Mol Biol*, 1983; 170: 723
- 8 Fuller F. *Proc Natl Acad Sci USA*, 1978; 75: 3557
- 9 Tanaka F, Takahashi H. *J Chem Phys*, 1986; 83: 6017
- 10 Vologodskii A et al. *Nature*, 1979; 280: 294
- 11 Hao M -H, Olson W K. *Biopolymers*, 1989; 28: 373
- 12 White J et al. *Proc Natl Acad Sci USA*, 1988; 85: 772
- 13 Levene S, Crothers D. *J Mol Biol*, 1986; 189: 73
- 14 Travers A et al. *Phil Trans R Soc Lond*, 1987; B317: 537
- 15 White J et al. *Science*, 1988; 241: 323
- 16 Benham C. *Proc Natl Acad Sci USA*, 1977; 74: 2397
- 17 Le Bret M. *Biopolymers*, 1984; 23: 1835
- 18 Benham C. *Biopolymers*, 1979; 18: 609
- 19 Tsura H, Wadati M. *Biopolymers*, 1986; 25: 2083
- 20 Hao M -H, Olson W K. *Macromolecules*, 1989; 22: 3292

[本文于1989年3月31日收到]

(上接第 197 页)

- 17 Lanyi J K, Schobert B. *Biochemistry*, 1983; 22: 2763
- 18 Stoeckenius W, Bogomolni R A. *Ann Rev Biochem*, 1982; 52: 587

- 19 Bogomolni R, Spudich J L. *Biophys J*, 1987; 52: 1071
- 20 Wolff E K et al. *Proc Natl Acad Sci USA*, 1986; 83: 7272

[本文于1989年4月14日收到]