

## 综述与专论

# DNA 双螺旋的精细结构

张 春 霆

(天津大学物理系)

### 提 要

70年代末,由于DNA寡聚体(短的DNA片段)单晶X射线衍射实验的结果,发现双螺旋参数随着序列的不同而起伏变化着——双螺旋的精细结构。本文综述了这方面的进展情况,重点介绍Calladine的解释;由DNA序列预测螺旋参数变化的Dickerson规则。其生物学意义在于:DNA碱基的序列信息可能贮存在双螺旋的局部精细结构之中。

**关键词** 螺旋扭角, Dickerson 求和函数, Tung-Harvey公式, 氢键网

DNA 双螺旋的精细结构,这是什么意思?当前进展情况如何?研究双螺旋的精细结构有何生物学意义?这是本文欲简要回答的三个问题。

众所周知, Watson-Crick 的双螺旋模型是根据 Franklin 和 Wilkins 所拍摄的纤维DNA的X光衍射照片提出来的。由于这种衍射技术的有限的解像力,因此只能得到双螺旋结构的平均信息。70年代末,由于人工合成技术的提高,人们可以合成寡核苷酸链(短的DNA片段),提纯并使之形成单晶体,从而可采用单晶的X光衍射技术。这种方法大大提高了解像力,结果发现DNA双螺旋结构,不论是A型或B型的,甚至左旋的Z-DNA都是不均匀的。就是说,双螺旋的一系列螺旋参数随着碱基序列的不同而在一定范围内起伏变化。这种现象就称为双螺旋的精细结构。它是对均匀双螺旋结构的一种否定,也是认识上的一种深化。

由于技术上的突破,各种长度不同、序列不同的寡核苷酸单晶的X光衍射结果纷纷出现,积累了相当丰富的数据。于是人们就试图对这

些结果加以解释。其中, Calladine 的解释被认为是最合理的。另外,一个重要问题是:任意给出一种碱基序列,能否用某种规则预测螺旋参数的变化规律,答案是肯定的。本文将重点介绍 Dickerson 规则。

正如 Dickerson<sup>[1]</sup>最近所指出的:“双螺旋DNA分子晶体的X射线分析使人们认识到,碱基顺序信息可能贮存在螺旋的局部结构内”。进一步的研究表明,各种酶或蛋白质特异地识别某DNA片段,是与该段双螺旋的精细结构息息相关的。由此可见,研究双螺旋的精细结构在生物学上是相当重要的。

本文将对以上提及的问题加以介绍。读者亦可参见有关综述文章<sup>[1,2]</sup>。

### 一、双螺旋的精细结构与 Calladine 假设

如上所述, DNA 双螺旋的精细结构系指螺旋参数随碱基序列而变化的现象。用于研究精细结构的螺旋参数很多,下面介绍其中比较重要的四个参数。

1. 螺旋扭角 (Helix twist angle), 符号为

tg。定义：一个碱基对中 C1'-C1' 连线与相邻的另一碱基对中的对应连线，投影到与螺旋轴垂直的平面上的夹角。对 B-DNA，纤维衍射的结果约为 36°；单晶衍射的结果表明，它可在 27°—40° 的范围内变化。

2. 碱基转角 (Roll angle of base-pairs), 符号为  $\theta_R$ 。定义：每一碱基对的平均平面可绕 C6-C8 (嘧啶-嘌呤) 轴线转动。两个相邻碱基对的平均平面之间的夹角称为碱基转角。同时规定：当此角开口朝向小沟时, 转角为正; 若开口朝向大沟, 则转角为负。对于 B-DNA, 此角平均值约为  $-1.1^\circ$ 。

3. 主链扭角 (Main-chain torsion angle), 符号为  $\delta$ 。定义：根据 IUB-IUPAC 所建议的术语和符号, 沿 5'→3' 方向从一个磷原子到下一个相邻的磷原子之间的六个扭角分别以  $\alpha, \beta, \gamma, \delta, \epsilon$  和  $\zeta$  表示之。这里所说的扭角是  $\delta$ , 系指该链上的角 C5'-C4'-C3'-O3'。在与其互补的链上也有对应的角。通常研究同一碱基对中这两个对应角之差  $\Delta\delta$ 。对于 B-DNA,  $\Delta\delta$  的平均值为零。

4. 螺旋桨式扭角 (Propeller twist angle), 符号为 pr. tw.。定义：在一碱基对中的每一碱基平面可绕它们的长轴 (大体上靠近 C6-C8 线) 旋转, 这两个平面之间的夹角称为螺旋桨式扭角。同时规定：沿碱基对长轴两个方向看, 靠近观察者那一碱基平面若按顺时针方向旋转, 则螺旋桨式扭角为正; 反之为负。

前两个螺旋参数, 反映了双螺旋中从一个碱基对到另一个相邻的碱基对之间的结构情况; 后两个参数则反映了单个碱基对本身的结构特征。还有另外一些参数用于双螺旋的精细结构研究, 这里不再赘述。所有这些参数中, 显然螺旋扭角 tg 是最重要的。本文将主要讨论这个参数。图 1 是 B-DNA 十二聚体 CGCGAATTCGCG 中参数 tg 随着碱基序列而变化的情况<sup>[1]</sup>。

除了 tg 以外, 对于其它螺旋参数也有类似现象, 只是其变化情况不同而已。另外, A-DNA 和 Z-DNA 寡聚体单晶的实验结果也显示了

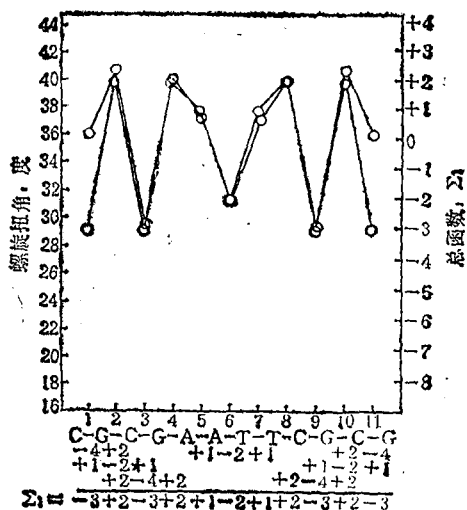


图 1 B-DNA 双螺旋十二聚体螺旋扭角的测量值 (粗线) 和用 Dickerson 求和规则  $\Sigma 1$  所得到的预测值 (细线) 随序列的变化

图的下部数字计算说明了应用规则  $\Sigma 1$  的具体过程

相似的现象。这种双螺旋的精细结构主要有两个显著的特点：1. 螺旋参数不再像 Watson-Crick 经典双螺旋模型那样保持不变；2. 碱基顺序不同其变化情况也不同。

不久前 Calladine 提出一种假设来解释上述现象<sup>[3]</sup>。他首先注意到这样一种事实：嘌呤是双环的；而嘧啶则是单环的。因而嘌呤的尺寸比嘧啶大。所以在双螺旋结构中，嘌呤环伸长越过了螺旋轴“跑”到嘧啶一侧。同样，对侧的嘌呤环也可以伸长到这一侧来。图 2 给出了这种情况的示意图。对 B-DNA 来说，如果螺旋桨式扭角为 0，那么所有碱基对将相互

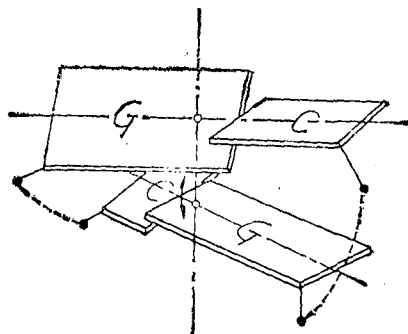


图 2 两个相邻碱基对 (嘧啶-嘌呤) 中交叉链的嘌呤-嘌呤之间的立体障碍示意图

图中箭头表示了这种障碍的具体位置

无妨碍地逐层堆积起来。但实验结果显示,这个角度,对 B-DNA 来说具有正  $10^{\circ}$ — $20^{\circ}$  的数值。这将使 B-DNA 的堆积更有效。但这样一来,两个都伸展到对方一侧的相邻的嘌呤环将会发生“冲突”,被称为立体化学障碍,它的本质是一种分子间的范德瓦尔斯力。图 2 中的箭头显示了这种“冲突”。如果我们研究的是一种 Y-R 序列(Y 代表嘧啶, C 或 T; R 代表嘌呤, G 或 A),则冲突发生在小沟一侧;如果是 R-Y 序列,则冲突发生在大沟一侧。对于 R-R 或 Y-Y 序列,不存在这种立体化学障碍。

为了缓解这种立体化学障碍, Calladine 假设,整个双螺旋结构将作出适当的调整。下面四种变化都是可能的: 1. 减少局部的螺旋扭角  $\tau g$ ; 2. 在冲突发生的那一侧使碱基转角  $\theta_R$  增大; 3. 使一个碱基对或相邻的两个碱基对中的嘌呤环沿它们的长轴向两边滑动离开一些,这将使参数  $\Delta\delta$  发生变化; 4. 减少一个或两个相邻碱基对中的螺旋桨式扭角。

正是这四种变化导致了双螺旋的精细结构。Calladine 又进一步利用弹性束力学原理来分析数据,发现这种交叉链的嘌呤-嘌呤立体障碍在小沟一侧比在大沟一侧严重。粗略地说,前者的严重程度约比后者大一倍。

这种对精细结构现象发生原因的解釋,当前被认为是比较合理的。上述 Calladine 的工作,通常被称为 Calladine 假设或原理。

## 二、Dickerson 规则及其性质

根据 Calladine 原理可知,螺旋参数随序列而变化,归根到底是由序列本身决定的。那么,应该有一种规则,根据它就可以由碱基序列预测各种螺旋参数的变化。Dickerson 规则就是第一种比较成功的规则<sup>[4]</sup>。从 Calladine 假设出发, Dickerson 定义了四个求和函数  $\Sigma 1$ — $\Sigma 4$ , 可分别用来预测上述四个螺旋参数的变化。这里重点介绍求和函数  $\Sigma 1$ , 它预测参数  $\tau g$  的变化。

求和函数  $\Sigma 1$  每次取四个碱基序列加以考察。如果它是 X-R-Y-X (X 代表任何碱基),

则在 X-R 之间填上一整数 +1, R-Y 之间填上 -2, Y-X 之间填上 +1。如果它是 X-Y-R-X, 则分别填上 +2, -4, +2。对于任意给定的碱基序列,从头开始,从左往右,每四个碱基为一组应用上述规则。然后步进式地往右推移: 即先考察碱基 1—4; 次考察 2—5; 再次考察 3—6, …, 等等, 依此类推。最后,把同一列的整数加起来即可。图 1 下部的数字即应用求和函数或求和规则  $\Sigma 1$  的实例。读者只须略加考察即可明白。这件工作易于编写程序交由计算机去完成。最后得到一个整数序列, 逐个乘上一个所谓求和函数单位, 即成为所预测的螺旋扭角序列。遗憾的是, 这个求和函数单位并非普适常数, 对部分 B-DNA 序列来说, 它取值  $2.1^{\circ}$ 。值得指出的是, 求和规则  $\Sigma 1$  对 A-DNA 也是成立的, 只是其求和函数单位取值不一样。其它三个求和函数与此类似, 这里不再赘述。

图 1 中的细线折线给出了求和函数  $\Sigma 1$  的预测结果, 可见与实际观测值符合, 两者的相关系数达到了 0.994。求和函数  $\Sigma 2$  则差一些,  $\Sigma 3$  和  $\Sigma 4$  则更差。可见,  $\Sigma 1$ — $\Sigma 4$  都是近似规则。

Dickerson 求和规则是根据寡聚体单晶的衍射结果总结出来的, 那么, 对于生物体内的实际 DNA 序列, 这种规则是否仍然成立或近似成立? 这是令人感兴趣的问题。近年来, 应用高解像力的核磁共振谱方法可以研究水溶液中 DNA 片段的精细结构<sup>[5]</sup>。其中某些实验结果支持 Dickerson 规则对溶液中的 DNA 序列仍然成立的观点<sup>[5]</sup>; 但另一些则不支持<sup>[6]</sup>。因此, 这个问题仍未解决。

最近, 我们研究了用 Dickerson 求和规则所得到的这些螺旋参数的性质。发现它们的方差存在着上确界。以螺旋扭角  $\tau g$  为例, 它的方差不大于  $4.625^{\circ}$ <sup>[7]</sup>。这表明它的均方差  $\leq \sqrt{4.625} \times 2.1^{\circ} \cong 4.5^{\circ}$ , 其中  $2.1^{\circ}$  是  $\Sigma 1$  函数的单位 (B-DNA)。即, 平均偏离不超过  $4.5^{\circ}/36^{\circ} = 12.5\%$ 。

Dickerson 规则是一种近似的规则。有人对它作了一些改进<sup>[8,9]</sup>。

### 三、生物学意义

Nussinov 曾经用 Dickerson 规则研究了 60 种原核生物的 DNA 序列, 结果发现, 大的构象起伏刚好位于蛋白质或酶特异识别的某些 DNA 位点附近<sup>[10]</sup>。早就发现, 蛋白质或酶可以特异地识别某些 DNA 片段, 而不能识别另一些。这是为什么? 现在, 从 DNA 双螺旋的精细结构中, 可望找到答案。众所周知, 在原核生物 DNA 转录过程中, 有一个保守的 DNA 片段, 即所谓 Pribnow 盒——TATAATG 序列, 被认为与 RNA 聚合酶和 DNA 的相互识别有关。在真核生物的转录过程中也有类似的 Hogness 盒——TATAAA (AG) 序列。如果我们把 Dickerson 规则应用到这两个特殊序列中, 就会发现螺旋扭角的最大值与最小值都出现在这两个序列之中, 而且两者紧相邻着! 可以想像, 这一局部的双螺旋结构起伏最大, 因而易于为酶所识别。如果按“钥匙与锁”模型来考察这一识别, 那么这段 DNA 序列就是一把很特殊的“钥匙”或“锁”。一般认为, 在基因调控中, 阻遏物或类似的控制蛋白质中的氨基酸可

与 DNA 双螺旋中互补碱基对边上的氮、氧原子形成氢键。既然如此, 很难想像双螺旋结构的局部起伏会影响这种氢键的形成。氢键的形成需要一定的距离和方位, 因此, DNA 双螺旋的局部起伏也许起着精细调节 DNA 与酶两者之间相互识别尺寸的作用。总之, 正如 Dickerson 所指出的那样: DNA 序列的信息可能贮存在它的局部精细结构之中。若果真如此, 那么 DNA 双螺旋的精细结构对生物学的重要意义, 将是不言而喻的了。

### 参 考 文 献

- 1 Dickerson R E, *Scientific American*, 1983; 249: 6
- 2 Shakked Z *et al. Progress in Biophysics & Mol Biol*, 1986; 47: 159
- 3 Calladine C R. *J Mol Biol*, 1982; 161: 343
- 4 Dickerson R E. *J Mol Biol*, 1983; 166: 419
- 5 Patel D J *et al. Proc Natl Acad Sci USA*, 1983; 80: 3908
- 6 Clore G M *et al. J Mol Biol*, 1985; 185: 219
- 7 Zhang Chunting (张春霆) *et al. Int J Biol Macromol*, 1989; 11: 9
- 8 Tung C S *et al. J Biol Chem*, 1986; 216: 3700
- 9 Kabsch W *et al. Nucl Acids Res*, 1982; 10: 1097
- 10 Nussinov R. *J Theor Biol*. 1985; 115: 197

[本文于1988年8月8日收到]

(上接第9页)

广谱促生长剂, 开发 hGRF 产品具有事半功倍的效果。此外, 由于 hGRF 是释放 GH 的一种诱导剂, 其有效剂量比 GH 少 100 倍以上, 用少量的 hGRF 即可达到 GH 的效果。目前, 我国正在开展对 GRF 的研究, 随着对 GRF 各方面研究的不断深入, 开发和应用 GRF 产品将是指日可待的。

### 参 考 文 献

- 1 Reichlin S. *Endocrinol*, 1961; 69: 225
- 2 Deuben R and Meites J. *Endocrinol*, 1964; 74: 408
- 3 Guillemin R *et al. Science*, 1982; 218: 585
- 4 Rivier J *et al. Nature*, 1982; 300: 276
- 5 Wehrenberg W *et al. Hormone Res*, 1986; 24: 82
- 6 Mayo K *et al. Proc Natl Acad Sci USA*, 1986; 82: 63
- 7 Ling N *et al. Proc Natl Acad Sci USA*, 1984; 81: 4302

- 8 Ling N *et al. Ann Rev Biochem*, 1985; 54: 403
- 9 Bohlen P *et al. Biochem Biophys Res Commun*, 1983; 116: 726
- 10 Esch F *et al. Biochem Biophys Res Commun*, 1983; 117: 772
- 11 Brazeau P *et al. Biochem Biophys Res Commun*, 1983; 125: 606
- 12 Bohlen P *et al. Biochem Biophys Res Commun*, 1984; 125: 1005
- 13 Baird A *et al. Neuroendocrinol*, 1985; 42: 273
- 14 Ling N *et al. Biochem Biophys Res Commun*, 1984; 123: 854
- 15 Kempe T *et al. Bio Tech*, 1986; 4: 565
- 16 Jamila A *et al. Gene*, 1987; 53: 219
- 17 Coude F *et al. Trends in Biotech*, 1984; 2: 83
- 18 Engels J W *et al. Protein Engineering*. 1987; 1: 3
- 19 Anba J *et al. Gene*, 1987; 53: 219
- 20 Thorner M *et al. Hormone Res*, 1986; 24: 91
- 21 Arther M *et al. Metabolic diseases and endocrine function*. Academic Press, 1985: 185

[本文于1988年10月18日收到]