

DNA碱基质子化学位移受所在序列的五联体性质影响

余多慰, 袁生, 金晓燕

(南京师范大学生命科学学院, 江苏省分子医学技术重点实验室, 南京 210097)

摘要: 根据生物大分子核磁共振数据库 (BMRB) 内 16 条单链 DNA 序列中的碱基特征质子的化学位移信息, 分析结果表明, 五联体 (pentaplet) 是到目前为止可以由实验数据证明的、决定中部碱基质子化学位移水平的基本单位, 即 DNA 碱基质子的 NMR 化学位移受所在五联体序列的控制。以五联体中部是 T 碱基为例, 来自化学位移的证据符合来自量子力学计算所得“5' 嘧啶 - 嘌呤比 5' 嘌呤 - 嘧啶的顺序更稳定”的论断, 表现为 5' 嘧啶 - 嘌呤侧翼顺序导致的中部 T 碱基质子化学位移, 比 5' 嘌呤 - 嘧啶顺序 δ 值小 0.089。对于中部碱基质子化学位移, 5' 侧翼二联体效应与 3' 侧翼二联体效应明显不同。5' 侧翼序列对五联体中部碱基质子化学位移的影响从大到小, 与 5' 序列的色散力排列顺序更相关。氢谱上 A H8、A H2、G H8、T H6、C H6 的位移分布顺序, 与从头计算 (ab initio) 和 $\delta + \text{HMO}_N$ 二种伴氢碳原子净电荷计算结果最为接近, 相关性好。与 ab initio 法得到的氢原子净电荷相关性不好。二翼碱基可以对五联体中心碱基的非交换质子在 8.5 Å 的距离上产生影响, 这是对 NMR 偶极作用距离极限的突破。DNA 的质子次级化学位移不是像蛋白质那样由氢键起主导作用。以上分析为建立双链 DNA 碱基质子化学位移理论预测公式提供了依据。

关键词: NMR; 质子; 化学位移; 五联体; DNA

中图分类号: O482.53

0 引言

生物大分子中的每种氢质子, 在核磁共振 (NMR) 测试时, 都有特定的化学位移。在 DNA 序列为 20 个碱基以上时, ^1H 的信号峰有大量的重叠, 使用 2D-NOESY 也难以全部分辨。使用异核多维 NMR 方法, 则因同位素标记以及测试条件的复杂化等问题, 使高残基数的生物大分子结构解析存在困难。因此有必要从其他角度来考虑 DNA 质子化学位移产生的机理, 以期提高各种质子化学位移的理论预测准确性。

核磁共振次级化学位移 (即由空间结构带来的化学位移变化) 和蛋白质二级结构的关系国内外均有分析, 认为氢键效应可能是造成次级化学位移规律变化的主要因素。K. Wüthrich 等曾研究了蛋白质 αH 和 NH 的次级化学位移, 与质子和氢键受体羰基氧的距离关系, 并总结出经验式^[1-3]。

可以认为 DNA 中不同的氢质子化学位移不同, 这是因为每个质子所处的化学环境不一样, 也就是多聚核苷酸中的氢质子会在单核苷酸氢质子的基准值上有变量存在。关于碱基质子化学位移的增量, 1983 年就建立有估算公式 (针对单链 RNA 和 DNA 寡聚体 70°C 下随机卷曲状态)^[4,5]。van de

Ven 等^[6]做了 500 个三联体 (triplet) 和 120 个末端二联体 (terminal doublet) 的分析, 发现双螺旋 DNA 中碱基质子化学位移有“三联体”性, 即某碱基质子的位移值依赖于相邻的 5' 和 3' 是什么碱基。1997 年, Wijmenga 等^[7]也证实了这个性质。Pu-N-Pu 和 Py-N-Py 三联体中的 N 分别在 δ_{ref} (基准化学位移值) 的基础上向低场移动平均 $\delta 0.4 \sim 0.5$ 和 $\delta 0.15 \sim 0.2$ 。这些工作是为了帮助指认二维 NOESY 波谱, 不过特异的三联体和末端二联体的数据没能全部给出, 不能直接加以运用。

Altona 等^[8]收集了 1982~1997 年的双螺旋 B-DNA 的数据, 得到三联体和末端二联体分析结果。开始把目光从“三联体”向外延伸, 但由于长期以来在 NMR 技术中对长程作用极限认识的限制, 该文献中只有本文提出的“五联体”概念的雏形, 描述的语言不确定, 结论基本上是否定的。认为 Second-order effects 缺乏屏蔽作用。

收稿日期: 2006-10-08

基金项目: 国家自然科学基金面上项目 (30170005) 及江苏省高校自然科学基金 (03KJB 180064)

通讯作者: 余多慰, 电话: (025)85891865,

E-mail: yuduowei@njnu.edu.cn

Cromsigt^[9]分析了 RNA 化学位移, Lam 等^[10]分析了单链 DNA, Wijmenga 等^[7]分析了双链 DNA, 他们都提出了预测化学位移的计算公式。尤其是 Lam 等的分析结果初步承认了单链 DNA 中存在有目标碱基前后各二个碱基的屏蔽作用, 并提出 $\delta_{\text{pred}} = \delta_{\text{triplet}} - \Delta_2^{5T} - \Delta_2^{3T} + \Delta_2^{5N} + \Delta_2^{3N}$ 公式, 但其中还存在一些问题, 没有区分二末端 (next nearest neighbor) 碱基的效应。

本文分析了生物大分子核磁共振数据库 (Biological Magnetic Resonance Data Bank, BMRB) 收录的、在同一条件下测试的 16 条 DNA 序列中的五联体 (pentaplet) 中心碱基特征质子的化学位移, 探讨不同 DNA 序列、屏蔽距离以及和碱基堆积相关的各种作用力 (包括了长距离屏蔽环境), 对碱基特征质子化学位移的影响, 并提出修改公式的依据, 以建立双链 DNA 五联体化学位移预测公式。

1 材料与方 法

1.1 实验数据

在 BMRB 数据库 (Wisconsin 大学 Madison 分校生物化学系, 2006 年 3 月) 的资料中, 含 16 个单链 DNA 样品 (表 1) 的质子化学位移数据。序列原本是为分析三联体而设计的, DNA 序列没有

Table 1 Sequence of single strands DNA sample

Accession ID	System name	¹ H number	Sequence
5634	ss1	92	TCCCTTAGTAGATGCTT
5635	ss2	91	TGCCTCTGTTGATACTT
5636	ss3	92	TACCTGTGCAATTCTT
5637	ss4	91	TTCCTATGTGAATCGTT
5638	ss5	90	TCGCTTTGTAAATGGTT
5639	ss6	91	TGGCTCCGTTAATAGTT
5640	ss7	90	TAGCTGCGTCTATTGTT
5641	ss8	91	TTGCTACGTGTATCATT
5642	ss9	92	TCACTTCGTATATGATT
5643	ss10	91	TGACTCGGTTTATAAATT
5644	ss11	92	TAACTGGGTCCATTATT
5645	ss12	91	TTACTAGGTGCATCTTT
5646	ss13	90	TCTCTTGGTACATGTTT
5647	ss14	91	TGTCTCAGTTCATATTT
5648	ss15	90	TATCTGAGTCGATTTTT
5649	ss16	91	TTTCTAAGTGGATCCTT

反向重复性, 各三联体之间以 T 相分割, 因此 T 出现的频率最大 (47.059%), 五联体类型最完整, 其他三种碱基出现的频率均只有 17.647%。所以在做五联体分析时, 将中心碱基为 T 的序列作为研究对象。另外三种碱基的部分分析数据仅做为参考。

1.2 分析方法

从上述序列中获得全部可利用的五联体资料, 包括具有空位末端效应 (5' 或 3') 的序列在内, 中部碱基为 T 的五联体 89 个, A44 个, G44 个, C44 个, 共 221 个五联体数据。在分析中部碱基一侧的二联体序列特异效应时, 另一侧的二联体以 NN 表示, 序列均以 5'→3' 表示, 中部碱基为 T 的五联体类型各自有 4~10 个数据供平均求值, 标准差在 δ 0.076~0.118 之间。

2 结果与讨论

碱基质子化学位移主要与环流效应、磁各向异性、极化效应有关^[9,11], 但此前的研究还没有考虑单链上的 DNA 序列、相邻碱基的堆积能、色散力等因素与碱基质子化学位移是否有关。NMR 偶极作用距离极限 (长程屏蔽效应) 可以达到什么程度的研究还很少。本研究对这些因素进行探讨。

对 BMRB 数据库中 DNA 碱基特征质子的化学位移, 按照 16 条 DNA 序列中的五联体中心碱基所处序列环境, 进行了全面归类。五联体类型极限, 如果暂不涉及 5' 或 3' “末端” 的占位效应, 比如 “末端磨损 (end fraying) 效应”, 应为 4⁵ 种, 但 BMRB 至今所含实际测试过的 DNA 序列数据未达此极限。虽然序列尚不齐全, 有待 NMR 研究对象的扩大来加以充实, 但现有资料按照本文方法简并归类, 已经可以反映出一些基本规律。

本文揭示五联体是目前可以用实验数据证实的、决定中部碱基质子化学位移的基本单位, 5' 二联体序列和 3' 二联体序列对于中部碱基的化学位移所发挥的影响是不同的, 并非等值, 这与 Lam 等预测单链 DNA 碱基化学位移的计算公式中的要素不同。

2.1 碱基质子化学位移与 DNA 单链上相邻碱基堆积和序列的关系

2.1.1 5' 侧翼序列的影响

仅以五联体中部为碱基 T 的数据群为本文论述对象, 首先分析 5' 序列的影响。先分析 5' 嘌呤 -

嘧啶与 5' 嘧啶 - 嘌呤这二种类型。共收集了全部在数据库中存在的五联体数据, n 均 ≥ 4 。堆积能数据取自文献^[2]。结果见表 2。基本规律为: 5' 嘌

呤 - 嘧啶的堆积能值均高于 5' 嘧啶 - 嘌呤, 不稳定, 导致化学位移低场化 (位移加大)。

Table 2 Correlation of 5' doublets sequence and stacking energy

Pentaplet	Mean value* (δ)	Diversity in group ($\Delta\delta$)	Stacking energy (kJ·mol ⁻¹)	Diversity in group (kJ·mol ⁻¹)
<u>G</u> TNN	7.468±0.112		<u>GT</u> -27.50	
<u>T</u> GNN	7.394±0.107	0.074	<u>TG</u> -43.99	16.49
<u>A</u> TNN	7.511±0.090		<u>AT</u> -15.99	
<u>T</u> A ₂ NN	7.412±0.102	0.099	<u>TA</u> -27.50	11.51
<u>A</u> C ₂ NN	7.499±0.116		<u>AC</u> -27.50	
<u>C</u> A ₂ NN	7.398±0.081	0.101	<u>CA</u> -43.99	16.49
<u>G</u> C ₂ NN	7.513±0.091		<u>GC</u> -40.56	
<u>C</u> G ₂ NN	7.431±0.090	0.082	<u>CG</u> -61.07	20.51
<u>C</u> T ₂ NN	7.486±0.108		<u>CT</u> -41.06	
<u>T</u> C ₂ NN	7.540±0.099	0.054	<u>TC</u> -28.38	12.68
<u>T</u> T ₂ NN	7.539±0.104		<u>TT</u> -22.48	
<u>C</u> C ₂ NN	7.533±0.097	0.006	<u>CC</u> -34.57	12.09
<u>G</u> A ₂ NN	7.352±0.084		<u>GA</u> -28.38	
<u>A</u> G ₂ NN	7.412±0.102	0.060	<u>AG</u> -41.06	12.68
<u>A</u> A ₂ NN	7.337±0.106		<u>AA</u> -22.48	
<u>G</u> G ₂ NN	7.406±0.098	0.069	<u>GG</u> -34.57	12.09

*Only central residue is T in pentaplets

本文分析的是 H8 和 H6 质子。如果在规则的双螺旋中, 考虑到堆积的空间排布类型, 这些数值移动能相当好地被理解。能量分析所得结论为“5' 嘧啶 - 嘌呤比 5' 嘌呤 - 嘧啶的顺序更稳定”^[8], 本文分析结论与其符合, 表现在对中部碱基化学位移的影响上, 是 5' 嘧啶 - 嘌呤二联体位移影响 (δ 7.409±0.017) 比 5' 嘌呤 - 嘧啶 (δ 7.498±0.021) 小 δ 0.089, 即导致中部碱基质子的化学位移值小 (位于高场方向)。

再分析 5' 嘧啶 - 嘧啶与 5' 嘌呤 - 嘌呤这二种类型。由表 2 中数据可见, 堆积能数值有对应性规律。比如, CT 和 AG 堆积能相同, 不便分析 5' 嘧啶 - 嘧啶和 5' 嘌呤 - 嘌呤之间关系, 但 5' 嘧啶 - 嘧啶的化学位移影响 (δ 7.525±0.026) 普遍强于 5' 嘌呤 - 嘌呤 (δ 7.377±0.038), 即 5' 嘧啶二聚体的影响强于嘌呤二聚体。嘧啶同聚二联体之间偏差小 (δ 7.536±0.004); 而嘌呤同聚二联体的影响之间偏差大 (δ 7.372±0.049), 与嘌呤异聚二联体相似 (δ 7.382±0.042)。

其次, 5' 嘌呤 - 嘧啶二联体位移影响 (δ 7.498±0.021) 多数比 5' 嘧啶 - 嘧啶 (δ 7.525±0.026) 少, 但明显比 5' 嘌呤 - 嘌呤 (δ 7.377±0.038) 多。5' 嘧啶 - 嘌呤二联体位移影响 (δ 7.409±0.017) 也是明显比 5' 嘧啶 - 嘧啶 (δ 7.525±0.026) 少, 但也同样是比 5' 嘌呤 - 嘌呤的影响 (δ 7.377±0.038) 大。5' 二联体影响的顺序为 5' 嘧啶 - 嘧啶 > 5' 嘌呤 - 嘧啶 > 5' 嘧啶 - 嘌呤 > 5' 嘌呤 - 嘌呤。在能量分析中被证明是稳定的 5' 堆积结构, 其对碱基非交换质子的化学位移影响小。

图 1 中可见三种关系类型: 1) 嘧啶 - 嘌呤、嘌呤 - 嘧啶二联体, 位移变化均与堆积能变化一致。2) 嘧啶 - 嘧啶, 位移变化与堆积能变化一致, 但分成二类: 异二联体中 TC/CT 与嘧啶 - 嘌呤、嘌呤 - 嘧啶二联体堆积能变化一致, 而 CT/TC (如图 1 中所示) 则不同; 嘧啶同二联体也存在类似变化。3) 嘌呤 - 嘌呤, 位移变化与堆积能变化正相反。

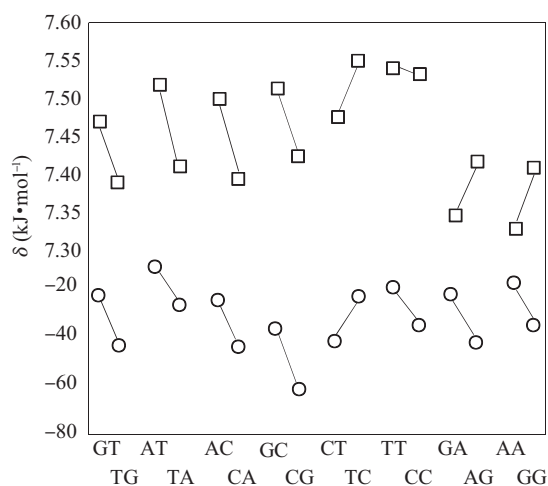


Fig.1 The chemical shift of central residue proton with influence of 5' doublets sequence. \square : 5' doublet; \circ : Stacking energy

2.1.2 3' 侧翼序列的影响

还是以中部碱基为 T 的五联体做代表性分析,共收集了全部在数据库中存在的五联体数据, n 均 ≥ 4 。

表 3 反映的基本规律为: 嘌呤 - 嘧啶 3' 的堆积能值均高于嘧啶 - 嘌呤 3', 不稳定, 但也导致化学位移高场化 (位移减少)。这个现象与 5' 二联体的情况相反。

尽管由 5' \rightarrow 3' 特定二联体堆积能值没有变, 但特定二核苷酸位于中部碱基的下游, 对化学位移均值的影响就有和位于上游相反的表现。5' 效应与 3'

效应明显不同 (图 2)。这个结论有助于正确预测 DNA 化学位移。

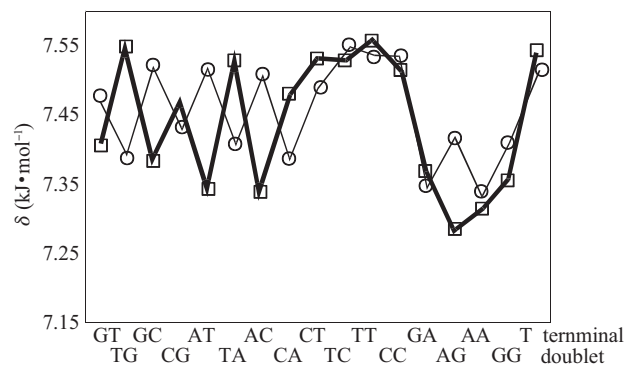


Fig.2 Correlation of the chemical shift of central residue proton and flanking doublets sequence. \circ : 5' doublet; \square : 3' doublet

表 3 反映的基本规律还有: 堆积能值有对应性, 比如, CT 和 AG 堆积能相同, 不便分析嘧啶 - 嘌呤 3' 和嘌呤 - 嘌呤 3' 之间堆积能与化学位移的关系, 但嘧啶 - 嘧啶 3' 的化学位移影响 ($\delta 7.532 \pm 0.017$) 普遍高于嘌呤 - 嘌呤 3' ($\delta 7.329 \pm 0.037$)。这一点与 5' 分析结论一致。二联体序列性质对中部碱基化学位移的影响, 除可从以下数据判别以外, 还可观察数据偏差的信息, 嘧啶同聚二联体之间偏差 ($\delta 7.534 \pm 0.029$) 与嘌呤同聚二联体之间偏差相同 ($\delta 7.334 \pm 0.029$), 这一点与 5' 分析结果不一致。嘧啶异聚二联体之间偏差最小 ($\delta 7.531 \pm 0.001$), 而嘌呤异聚二联体之间偏差最大

Table 3 Correlation of 3' doublets sequence and stacking energy

Pentaplet	Mean value* (δ)	Diversity in group ($\Delta\delta$)	Stacking energy ($\text{kJ}\cdot\text{mol}^{-1}$)	Diversity in group ($\text{kJ}\cdot\text{mol}^{-1}$)
<u>NNTCG</u>	7.468 ± 0.110		<u>CG</u> -61.07	
<u>NNTGC</u>	7.388 ± 0.084	0.080	<u>GC</u> -40.56	20.51
<u>NNTTA</u>	7.530 ± 0.062		<u>TA</u> -27.50	
<u>NNTAT</u>	7.346 ± 0.056	0.184	<u>AT</u> -15.99	11.51
<u>NNTCA</u>	7.476 ± 0.080		<u>CA</u> -43.99	
<u>NNTAC</u>	7.338 ± 0.080	0.138	<u>AC</u> -27.50	16.49
<u>NNTTG</u>	7.544 ± 0.065		<u>TG</u> -43.99	
<u>NNTGT</u>	7.406 ± 0.056	0.138	<u>GT</u> -27.50	16.49
<u>NNTTC</u>	7.530 ± 0.077		<u>TC</u> -28.38	
<u>NNTCT</u>	7.531 ± 0.077	0.001	<u>CT</u> -41.06	12.68
<u>NNTTT</u>	7.554 ± 0.061		<u>TT</u> -22.48	
<u>NNTCC</u>	7.513 ± 0.084	0.041	<u>CC</u> -34.57	12.09
<u>NNTGA</u>	7.365 ± 0.052		<u>GA</u> -28.38	
<u>NNTAG</u>	7.285 ± 0.075	0.080	<u>AG</u> -41.06	12.68
<u>NNTGG</u>	7.354 ± 0.085		<u>GG</u> -34.57	
<u>NNTAA</u>	7.313 ± 0.053	0.041	<u>AA</u> -22.48	12.09

*Only central residue is T in pentaplets

($\delta 7.325 \pm 0.057$)。化学位移影响的顺序是嘧啶 - 嘧啶 3' > 嘧啶 - 嘌呤 3' > 嘌呤 - 嘧啶 > 3' 嘌呤 - 嘌呤 3'。这中间二项的顺序与 5' 二联体不同, 但本质都

是接近中部碱基的, 如果是嘧啶, 后者对中部碱基质子化学位移的影响就大。5' 二联体与 3' 二联体效应比较见表 4。

Table 4 Correlation of 5' doublets and 3' doublets for the chemical shift of central residue proton

Pentaplet	Mean value* (δ)	Pentaplet	Mean value* (δ)
<u>G</u> TNN	7.468±0.112	NN <u>T</u> GT	7.406±0.056
<u>T</u> GTNN	7.394±0.107	NN <u>T</u> TG	7.544±0.065
<u>A</u> TTNN	7.511±0.090	NN <u>T</u> AT	7.346±0.056
<u>T</u> ATNN	7.412±0.102	NN <u>T</u> TA	7.530±0.062
<u>A</u> CTNN	7.499±0.116	NN <u>T</u> AC	7.338±0.080
<u>C</u> ATNN	7.398±0.081	NN <u>T</u> CA	7.476±0.080
<u>G</u> CTNN	7.513±0.091	NN <u>T</u> GC	7.388±0.084
<u>C</u> GTNN	7.431±0.090	NN <u>T</u> CG	7.468±0.110
<u>C</u> TTNN	7.486±0.108	NN <u>T</u> CT	7.531±0.077
<u>T</u> CTNN	7.540±0.099	NN <u>T</u> TC	7.530±0.077
<u>T</u> TTNN	7.539±0.104	NN <u>T</u> TT	7.554±0.061
<u>C</u> CTNN	7.533±0.097	NN <u>T</u> CC	7.513±0.084
<u>G</u> ATNN	7.352±0.084	NN <u>T</u> GA	7.365±0.052
<u>A</u> GTNN	7.412±0.102	NN <u>T</u> AG	7.285±0.075
<u>A</u> ATNN	7.337±0.106	NN <u>T</u> AA	7.313±0.053
<u>G</u> GTNN	7.406±0.098	NN <u>T</u> GG	7.354±0.085

*Only central residue is T in pentaplets

表 5 中数据显示, 除 T 以外的三种碱基也基本显示了同样的倾向。除 T 以外, 有平均值的序

列类型, $n=5$ 。其他类型中仅有一个分析序列, 所以列出的不是平均值, 没有标准差。

Table 5 Correlation of 5' doublets and 3' doublets for the chemical shift of four base proton at central residue site (\bullet)

Pentaplet	Value (δ)			
	T	C	A	G
<u>G</u> T•NN	7.468±0.112	7.679±0.107	8.220±0.112	7.864±0.092
<u>T</u> G•NN	7.394±0.107	7.653±0.094	8.207±0.102	7.865±0.083
<u>A</u> T•NN	7.511±0.090	7.694±0.091	8.237±0.097	7.868±0.079
<u>T</u> A•NN	7.412±0.102	7.631±0.116	8.181±0.095	7.838±0.076
<u>A</u> C•NN	7.499±0.116	7.797	8.311	7.917
<u>C</u> A•NN	7.398±0.081	7.689	8.269	7.977
<u>G</u> C•NN	7.513±0.091	7.804	8.324	7.923
<u>C</u> G•NN	7.431±0.090	7.738	8.297	7.938
NN• <u>G</u> T	7.406±0.056	7.603±0.039	8.128±0.030	7.770±0.012
NN• <u>T</u> G	7.544±0.065	7.732±0.066	8.277±0.051	7.895±0.032
NN• <u>A</u> T	7.346±0.056	7.550±0.046	8.115±0.036	7.803±0.045
NN• <u>T</u> A	7.530±0.062	7.746±0.047	8.293±0.034	7.921±0.020
NN• <u>A</u> C	7.338±0.080	7.594	8.143	7.837
NN• <u>C</u> A	7.476±0.080	7.718	8.272	7.918
NN• <u>G</u> C	7.388±0.084	7.630	8.163	7.823
NN• <u>C</u> G	7.468±0.110	7.728	8.310	7.926

由表 3~5 中数据可知, 同样的嘌呤 - 嘧啶或嘧啶 - 嘌呤二联体序列, 位于中部碱基的不同方向上, 可以发挥完全相反的作用。这与 DNA 磷酸核糖主干链的旋转有关。

按照 Lam 公式^[10], δ_2^{5N} 和 δ_2^{3N} 的值相等, 从表 4 内容可知, 这不符合实际。以末端是 G 的情况为例, 7.468 ± 0.112 、 7.544 ± 0.065 、 7.513 ± 0.091 、 7.468 ± 0.110 、 7.352 ± 0.084 、 7.285 ± 0.075 、 7.406 ± 0.098 、 7.354 ± 0.085 这八组数据, 从中扣除三联体的数值 (N_{i-1} 取均值), 再扣除该公式的常数 0.076 (δ_2^{5T} 和 δ_2^{3T}), 不能得到 Lam 公式中 δ 0.013 的校正正值。

2.2 碱基伴氢碳原子净电荷和碱基氢原子净电荷的影响

在单链 DNA 中碱基之间的堆积力、静电力、色散力、极化力、排斥能诸因素中, 次级化学位移受何种因素的影响, 何种因素的影响最大, 这是下面要分析的内容。不考虑 DNA 序列因素, 仅分析原子类型, 发现与碱基质子直接相连的碳原子净电荷的数值分布, 与质子类型化学位移分布顺序相关性好。不同碱基质子本身的净电荷, 则相关性很差。

表 6 中 δ +HMO_N 法, 指 Huckel 分子轨道法, 这是一个经验的处理 π 电子体系的理论近似方法,

Table 6 The net charge value of base C-atom with H

Atom	δ +HMO _N	ab initio
A C8	+0.317	+0.025
A C2	+0.313	+0.018
G C8	+0.239	-0.000
U C6	+0.192	-0.057
C C6	+0.259	-0.018
C C5	-0.155	-0.300

主要用于平面共轭分子 (如碱基) 体系。ab initio 法指从头计算法。从氢谱上看, A H8、A H2、G

H8、T H6、C H6 的位移分布顺序, 与从头计算 (ab initio) 和 δ +HMO_N 二种计算结果^[13,14]最为接近, 相关性好。

从 NMR 氢谱上看, A H8、A H2、G H8、T H6、C H6 的位移分布顺序, 与从头计算 (ab initio) 得到的氢原子净电荷相关性不好 (表 7)。在化学位移影响因素中“环流效应”的研究较多, 通过上述碱基伴氢碳原子净电荷的分析, 可以认为它是环流效应的重要组成部分。而碱基氢原子净电荷与环流效应的关系, 则需重新考虑。

Table 7 The net charge value of base H-atom by ab initio

Atom	eV
A H8	+0.22482
A H2	+0.22182
G H8	+0.21306
T H6	+0.23578
C H6	+0.24122

2.3 堆积能测试存在不同结果

显然, 在用 eV 单位来表示堆积能和氢键能时, 应该与 $\text{kJ}\cdot\text{mol}^{-1}$ 一样 (是换算关系)。由表 8 中内容可以整理堆积能 (eV 单位) 由大到小为, $\text{GC} > \text{TG} (\text{CA}) > \text{TC} (\text{GA}) > \text{GG} (\text{CC}) > \text{AA} (\text{TT}) > \text{AT}$ 。从中不能归纳出嘌呤 - 嘧啶与嘧啶 - 嘌呤的顺序之间对比有何简单关系。由 Tm 的顺序可见 82°C 在序列中排列地位异常, 究其原因, 是氢键能偏大 (-0.484) 所致。使用 $\text{kJ}\cdot\text{mol}^{-1}$ 单位数据^[2] (见最右边一列) 对比后, 可见 $-34.57 \text{ kJ}\cdot\text{mol}^{-1}$ 在数值排列中也异常, 因而可以从另一角度证明这个 $\text{kJ}\cdot\text{mol}^{-1}$ 单位测试, 同样也包含有纵向碱基 (堆积) 和横向碱基 (配对) 二方面的作用。值得思考的是, 在本文分析单链 DNA 时, 得到的规律恰恰是与双链 DNA 能量分析相符的。

这里所用堆积能 (eV) 数据 (表 8) 与表 2 和表 3 中的数据不同, 在表 2 和表 3 中 GG 是

Table 8 Compared with activity energy and melt temporary of doublets sequence

Stacked dimers	Stacking energy (A) (eV)	H-bond energy (B) (eV)	A+B (eV)	Tm $^\circ\text{C}$	Stacking energy ($\text{kJ}\cdot\text{mol}^{-1}$)
(G-C)•(G-C)	-0.201	-0.484	-0.685	95	-40.56
(T-G)•(C-A)	-0.163	-0.280	-0.443	73	-43.99
(T-C)•(G-A)	-0.142	-0.280	-0.422	67	-28.38
(G)•(C)	-0.134	-0.484	-0.618	82	-34.57
(A)•(T)	-0.131	-0.076	-0.208	48	-22.48
(A-T)•(A-T)	-0.121	-0.076	-0.197	39	-15.99

-34.57 kJ·mol⁻¹, GC 是 -40.56, TG -43.99, TC -28.38, AA -22.48, AT -15.99 kJ·mol⁻¹[12]。与表 7 内二组数据相比较, 可见差异。这说明 eV 单位的测量数据[13]与能量测量的数据[12]不同, 不是换算的问题, 而是二次测量之间的条件差异所致。因此所反映的碱基间相互作用情况也就有所不同。在化学位移计算中, 依据哪种数据更为合适? 通过下一节的分析可知结果。

2.4 堆积能与碱基重叠平衡距离的关系

以 CG 和 GC 为例 (表 9), 嘧啶 - 嘌呤的距离比嘌呤 - 嘧啶短 0.048 nm, 堆积能相差 0.035 eV, 这可能是 5' 嘧啶 - 嘌呤更稳定乃至使五联体中部碱基质子化学位移少的重要原因。表 9 中 CG、CA、TG 碱基间距均比 GC 距离短, 但在平衡距离下的堆积能并不都是更负, 仅 CG 有更负的现象。不清楚是否因该文献中堆积能 (eV) 的测试条件不同而造成无规律可循。使用 kJ·mol⁻¹ 单位的实验结果[12] (见最右边一列), 则能从中看出堆积能与碱基重叠平衡距离还是有规律可循, 仅 GA 和 AA 二种嘌呤 - 嘌呤存在例外。而这一点, 在前面的分析 (2.2 节以及图 2) 中也已出现, 有一致性。

Table 9 Compared with balance distance and stacked energy of doublets sequence

Stacked dimers	Balance distance (nm)	Stacking energy at BD* (eV)	Stacking energy (kJ·mol ⁻¹)
CG	0.280	-0.182	-61.07
CA	0.305	-0.087	-43.99
TG	0.315	-0.091	-43.99
GA	0.325	-0.115	-28.38
GC	0.328	-0.147	-40.56
AA	0.328	-0.082	-22.48
GG	0.352	-0.010	-34.57
CC	0.368	-0.020	-34.57
TT	0.368	-0.028	-22.48

* BD means balance distance

2.5 色散力的影响

色散力 (E_{disp}) 与堆积力 (E_{HF}) 数据来自文献 [14], 堆积能数据来自文献 [12]。对比于堆积力 (能), 5' 侧翼二联体序列对五联体中部碱基化学位移的影响 (见表 3) 从大到小, 与 5' 序列的色散力 (E_{disp}) 排列顺序 (表 10) 更相关, 比如: TC > TT > TA > TG (δ 7.540, δ 7.539, δ 7.412, δ 7.394), 以及 GC > GT > GA/GG (δ 7.513, δ 7.468,

δ 7.352/ δ 7.406)。

Table 10 Compared with dispersion energy and stacked energy of doublets sequence

Stacked dimers	E_{disp}	E_{HF}	Stacking energy	Chemical shift*
TC	-31.21	6.95	-28.38	7.540
GA	-30.21	-0.08	-28.38	7.352
TT	-29.62	16.99	-22.48	7.539
GG	-29.62	18.16	-34.57	7.406
TA	-27.99	8.28	-27.50	7.412
TG	-26.41	8.58	-43.99	7.394
GC	-24.99	4.78	-40.56	7.513
GT	-23.77	9.83	-27.50	7.468

*Only central residue is T in pentaplets

2.6 距离的影响

van de Ven 等^[9]认为, K Wüthrich 建立的二维 NMR 交叉峰指认的步进方法, 就反映了 DNA “序列” 具有影响 NMR 信号的性质。步进方法包含二种距离影响因素的作用: 第一, 与 J 偶合有关的, 在大多数情况下, 是局限于三个或者更少的化学键之内的影响范围尺寸 (through-bond 理论); 第二, 涉及偶极反应的 4.5 Å 的最大影响范围尺寸 (through-space 理论)。碱基序列也是局部构象的决定因子。van de Ven 等因此认为三核苷酸作为碱基序列决定因子的基本单位是最好的选择, 可见他们已经从中部碱基两侧碱基的共同影响来考虑, 但他们否认 “更长范围效应” 的存在, 认为三核苷酸是可提供有效数据、以使统计学有关结论成为根据的最大的单位 (it is the largest unit for which sufficient data are available to warrant statistically relevant conclusions)^[6]。这个观点在 21 世纪初已经被改变, Altona 等^[8]在 2000 年指出 “也许还有 next nearest neighbors” 的作用。Lam 等^[10]在化学位移预测公式中进一步把 “next nearest neighbors” 做为了参数项。

本文认为, DNA 单链的相邻碱基非交换质子之间的距离, 由于碱基之间堆积力等约束因素的存在, 并不随着 OPO 骨干链的柔性而剧烈变化, 而且在二级结构出现后, 相邻碱基距离更受限制。五联体实际上在 B-DNA 中只有半圈, 但围绕着双螺旋轴心, 在轴向上, 最外侧碱基与中部碱基之间仅 6.8 Å 的距离, 但在环绕轴心的旋转后, 五联体最外侧碱基对中心碱基的特定非交换质子的间隔位置, 可以在 8.5 Å 的距离上产生影响。

以往的“through space”理论仅限于 4.5 Å 的距离，在此距离尚有弱的核 Overhauser 效应 (NOE)。现在已经有 8 Å 距离的研究^[15]。本文提出的 8.5 Å 距离认识，尚需要量子力学分析的进一步支持。

3 结 论

本文认为，DNA 的质子次级化学位移不是像蛋白质那样由氢键起主导作用。双链 DNA 质子化学位移预测公式，除保留参考（基准）化学位移成分 δ_{ref} ，以及环流效应 δ_{rc} 、磁各向异性 δ_{ma} 、极化（电场）效应 δ_{E} 、氢键作用 δ_{CT} 、 δ_{ib} (χ 键角为 240 度时由自己碱基诱导的化学位移效应) 等 δ_{conf} (与构象有关) 参数项外， δ_{intrin} (原意是“与构象无关”) 中还应包括五联体的序列组成效应、序列方向性影响效应、碱基堆积力效应、色散力效应的参数项。DNA 的质子化学位移计算值 δ_{calc} 应由下列公式表达：

$$\delta_{\text{calc}} = \delta_{\text{ref}} + \delta_{\text{rc}} + \delta_{\text{ma}} + \delta_{\text{E}} + \delta_{\text{CT}} + \delta_{\text{ib}} + \delta_{\text{intrin}}$$

总之，DNA 中五联体序列对中部碱基的质子化学位移产生作用是客观存在的事实。对“through space”理论的深入研究，有助于揭示化学位移产生的完整机制。

参考文献:

- [1] 李临生. 化学位移的规律和本质. 波谱学杂志, 1997,14: 257~260
- [2] 胡红雨, 鲁子贤, 杜雨苍. 核磁共振次级化学位移和蛋白质二级结构. 波谱学杂志, 1995,12:29~38
- [3] Wishart DS, Sykes BD, Richards FM. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy.

Biochemistry, 1992,31:1647~1651

- [4] Bell RA, Alkema D, Coddington JM, Hader PA, Hughes DW, Neilson T. Prediction of ¹H NMR chemical shifts of DNA oligomers. *Nucleic Acids Res*, 1983,11:1143~1149
- [5] Bell RA, Everett JR, Hughes DW, Coddington JM, Alkema D, Hader PA, Neilson T. Parameters for the calculation of proton NMR chemical shifts of oligoribonucleotides. *J Biomol Struct Dyn*, 1985,2:693~707
- [6] van de Ven FJM, Hilbers CW. Resonance assignments of non-exchangeable protons in B type DNA oligomers. *Nucleic Acids Research*, 1988,16:5713~5726
- [7] Wijmenga SS, Kruijthof M, Hilbers CW. Automated combined assignment of NOESY spectra and three-dimensional. *Journal of Biomolecular NMR*, 1997,10:337~350
- [8] Altona C, Faber H, Hoekzema AJAW. Double-helical DNA ¹H chemical shifts: an accurate and balanced predictive empirical scheme. *Magnetic Resonance in Chemistry*, 2000,38:95~107
- [9] Cromsigt JAMTC, Hilbers CW, Wijmenga SS. Prediction of proton chemical shifts in RNA. *Journal of biomolecular NMR*, 2001,21:11~29
- [10] Lam SL, Ip LN, Cui XD, Ho CN. Random coil proton chemical shifts of deoxyribonucleic acids. *Journal of Biomolecular NMR*, 2002,24:329~337
- [11] 姜 凌, 毛希安. 化学位移估算研究 ATP 构象随溶液 pH 值的变化. 物理化学学报, 2002,18(9):796~801
- [12] Ornstein RL, Rein R, Breen DL, MacElroy RD. An optimized potential function for the calculation of nucleic acid interaction energies 1Base stacking. *Biopolymers*, 1978, 17:2341~2360
- [13] 刘次全. 量子生物学及其应用. 北京: 高等教育出版社, 1990. 126,132~135
- [14] 刘次全, 白春礼, 张 静. 结构分子生物学. 北京: 高等教育出版社, 1997. 131
- [15] Fayon F, Massiot D. Through-space contributions to two-dimensional double-quantum *J* correlation NMR spectra of magic-angle-spinning solids. *J Chem Phys*, 2005,122:194313~194326

CHEMICAL SHIFT OF CENTRAL BASE AROMATIC PROTON DETERMINED BY PENTAPLET EFFECT IN SINGLE STRAND DNA

YU Duo-wei, YUAN Sheng, JIN Xiao-yan

(College of Life Science, Nanjing Normal University, Nanjing 210046, China)

Abstract: With analysis of NMR in BMRB 16 strands ss-DNA, the pentaplet concept can be determined for chemical shift prediction of the central base protons in pentaplet. The relations between the secondary chemical shift and some factor (for example, stack force, hydrophobic force, hydrogen bond force, London force, distance effect of next-nearest neighbors base et al.) have been detected, and the results proved that 5' next-nearest neighbors effect is different from 3' next-nearest neighbors when the chemical shift forming of characteristic protons of central base in the pentaplet. The 8.5 Å effective distance of “through-space” theory was supposed. The analysis throws new light on the mechanism involved predictive empirical scheme of base protons chemical shift in ds-DNA.

Key Words: NMR; Proton; Chemical shift; Pentaplet; DNA

This work was supported by grants from The National Natural Science Foundation of China (30170005) and The University Science Technology Research Project of Jiangsu Province (03KJB 180064)

Received: Oct 8, 2006

Corresponding author: YU Duo-wei, Tel: +86(25)85891865, E-mail: yuduowei@njnu.edu.cn