

1003-7713/2005/03-325-6

密度泛函方法研究气相胞嘧啶的互变异构化*

孙巨龙, 杨光辉, 何国钟, 韩克利**

(中国科学院大连化学物理研究所, 分子反应动力学国家重点实验室, 大连 116023)

摘要: 分别采用 7 种基组、3 种理论方法对胞嘧啶异构体 Cyt1 的结构进行优化, 通过与 Cyt1 的实验结果进行比较, 选取了适合研究胞嘧啶分子的 B3LYP/6-311 + G** 方法. 用该方法对胞嘧啶分子的 8 种异构体构型进行了充分优化, 研究了其中能量较低的 6 种胞嘧啶异构体的互变异构化过程. 对于得到的所有优化构型都进行了频率分析. 对于基态构型, 所有的频率都是正的; 对于过渡态构型, 只有一个虚频. 同时, 做了详尽的内禀反应坐标计算, 以保证所得到的过渡态连接相应的始末异构体. 所有给出的能量都已做了零点能校正. 理论研究结果可以对已有的实验结果给予合理解释.

关键词: 密度泛函理论; 胞嘧啶; 异构化

中图分类号: O641 文献标识码: A

DFT Study on Tautomerism of Gaseous Cytosine*

Sun Julong, Yang Guanghui, He Guozhong, Han Keli**

(State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023)

Abstract By comparing the ready experimental results with the calculated results obtained at 7 basis sets and 3 theoretical levels, respectively, for the tautomeric form Cyt1 of cytosine, the relatively accurate B3LYP/6-311 + G** theoretical method to study the tautomerism of cytosine was chosen. The ground-state structures of 8 tautomers of cytosine were fully optimized at B3LYP/6-311 + G** level, and the tautomerism of 6 relatively stable tautomers of cytosine was studied. The frequency analysis was performed on all the optimized structures. For the ground-state geometries, all the calculated frequencies are real; for the transition-state geometries, there is only one imaginary frequency for each structure. Detailed Intrinsic Reaction Coordination calculations were carried out to guarantee the optimized transition-state structures being connected to the related tautomers. All the energies given include the zero-point energy corrections. The theoretical results can give a reasonable interpretation for the experimental results.

Key words Density Functional Theory, Cytosine, Tautomerism.

1 引言

胞嘧啶(Cytosine)以及其它 4 种自然界存在的碱基分子(腺嘌呤、鸟嘌呤、尿嘧啶、胸腺嘧啶), 是构成重要生命大分子——核酸(Nucleic Acid)的基

本单元. 无论是高等生物、低等生物, 还是最简单的病毒, 它们都有共同的分子基础, 这就是核酸和蛋白质两类基本的生命大分子. 核酸对生物演化起着决定性作用, 它是生物遗传变异的物质基础, 决定着生物的发展方向. 研究核酸的基本组成单元——碱基

* Project supported by the National Basic Research Program of China (G1999075302).

** Corresponding author, Email: klhan@dicp.ac.cn Received 19 March 2004; in final form 9 June 2004.

分子的性质,可以为人们在分子水平上了解和认识生命大分子,提供更加丰富的信息^[1]. 碱基分子自身结构及碱基分子结合形成的碱基对,决定着 RNA (核糖核酸) 和 DNA (脱氧核糖核酸) 的结构和功能. 比如,在 DNA 双螺旋结构中,胞嘧啶分子通常以“正常的”(normal)氨-酮式(amino-keto)异构体 Cyt1 (见图 1) 形式,通过氢键与其它碱基分子结合形成碱基对,即标准碱基对(standard base pairs). 如果碱基分子不是以正常异构体形式参与碱基配对,将形成非标准碱基对(non-standard base pairs),这会导

致基因码的改变,生物体就要发生变异^[2].

实验研究表明^[3-6],碱基分子普遍存在互变异构现象. 有人指出^[7],对于含量低于 0.1% ~ 1% 的异构体,在实验上(比如光谱方法)很难检测到. 然而,在碱基对的复制过程中,异构体的含量在 10^{-8} ~ 10^{-11} 即会发生自发变异^[8],也就是说,导致遗传变异的异构体的浓度比实验可检测到的最低浓度还要低若干个数量级. 因此,用实验方法直接检测导致碱基对错配的微量异构体形式、研究异构体的相互转化机理是困难的. 而理论研究则不受此限制.

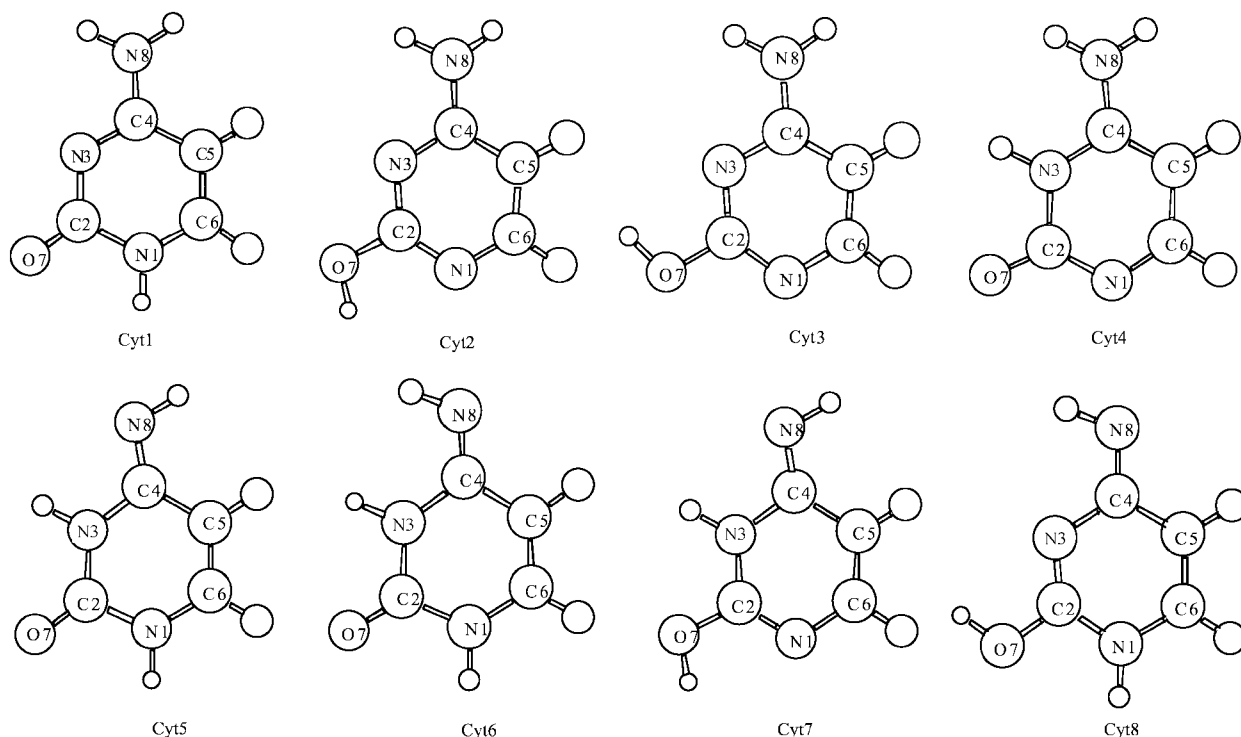


图 1 B3LYP/6-311 + G** 方法计算得到的 8 种胞嘧啶异构体基态构型

Fig. 1 Calculated ground-state geometries of 8 tautomers of cytosine at B3LYP/6-311 + G** level

胞嘧啶分子存在多种异构体形式,这些异构体的区别在于 H 原子位置的不同,它或者连在环上的氮原子上,或者连在环外的氧原子上,或者位于氨基团中(见图 1). 很多理论工作把胞嘧啶分子作为研究对象. 由于所采用的方法不同,基组大小不一样,所得到的计算结果也不可避免地存在着差异. 一般认为考虑了电子相关效应的 MP_n 方法,其计算精度是比较高的. 但由于该方法对计算机资源有较高要求,而且计算很费机时,因此它并不是首选的方法. 而密度泛函理论(DFT),由于它包含了交换能项和部分电子相关能,被认为是能很好地用来描述有机分子的特性^[9-12],特别是含有单、双键的环状有机

分子^[13]. 在本工作中,我们首先比较了不同理论方法、不同大小基组对胞嘧啶异构体 Cyt1 (见图 1) 的计算结果,并与 Cyt1 的实验数据进行比较,最后选取了适合研究胞嘧啶分子的理论方法——B3LYP/6-311 + G**. 用该方法优化了 8 种胞嘧啶异构体,研究了其中 6 个能量较低异构体的互变异构化过程,并做了详尽的内禀反应坐标(IRC)计算,以保证所得到的过渡态连接相应的异构体.

2 方 法

理论上确定异构体的相对稳定度,是通过比较各个异构体能量的方法得到的,包括计算零点振动

能的差别. 从数值计算角度讲, 这是一个计算量非常大的工作. 因为这项工作是要在两个大的、而且几乎相等的量之间计算其差值. 另外, 计算互变异构体之间的相对稳定度, 需要对电子相关效应给予足够考虑, 因为电子相关效应能够适当描述具有单、双键混合结构这样的碱基分子. 为了获得与实验数据符合得更好的计算结果, 必须使用包含电子相关效应的更高水平的理论方法来进行构型优化. 包含电子相关效应的高水平从头算 (*ab initio*) 方法的计算量是非常大的, 因此, 所计算体系的大小受到严格限制. 密度泛函方法的计算量比高水平从头算方法的计算量小许多, 而且该方法也考虑了电子相关效应. 密度泛函方法不仅考虑了大多数动力学以及非动力学电子相关效应, 而且仍然保留了独立粒子模型的简洁性. 这对于研究胞嘧啶这样的分子是非常必要的.

基函数的选择是影响计算结果好坏的重要因素之一. 在用密度泛函理论研究胞嘧啶异构化现象之前, 我们首先考察了基组效应. 在表 1 中, 我们分别采用 7 种基组在 B3LYP 水平下, 对胞嘧啶异构体 Cyt1 进行构型优化, 表 1 中最后一列是实验测得的 Cyt1 的结构参数^[14]. 从表 1 可以看出, 从 6-31G 到 6-31 + G**, 随着基组的增大, 计算结果与实验结果越来越接近; 但基组继续增大, 这种逼近效果却不明显, 甚至计算结果反而不如 6-31 + G** 的结果, 如 6-311G 和 6-311G** 这两组基函数的结果. 直到基组增大到 6-311 + G** 时, 其计算结果才与基组 6-31 + G** 的结果相近, 都与实验结果最接近. 因为采用 6-311 + G** 基组计算能量的精度, 比采用 6-31 + G** 基组计算能量的精度高, 所以在下面的胞嘧啶异构化研究中我们都采用 6-311 + G** 基组.

表 1 B3LYP 方法在 7 种基组下对 Cyt1 构型的计算结果
Table 1 Calculated structural parameters (Å and °) for Cyt1 using 7 basis sets at B3LYP level

Parameter	6-31G	6-31G*	6-31G**	6-31 + G**	6-311G	6-311G**	6-311 + G**	Exp. [14]
$r(\text{N1C2})$	1.436	1.430	1.430	1.427	1.432	1.430	1.428	1.392
$r(\text{C2N3})$	1.381	1.373	1.373	1.371	1.381	1.370	1.369	1.358
$r(\text{N3C4})$	1.333	1.319	1.320	1.321	1.333	1.316	1.317	1.339
$r(\text{C4C5})$	1.443	1.441	1.441	1.442	1.442	1.440	1.440	1.433
$r(\text{C5C6})$	1.363	1.359	1.359	1.361	1.360	1.356	1.356	1.357
$r(\text{C6N1})$	1.365	1.355	1.355	1.356	1.365	1.353	1.354	1.360
$r(\text{C2O7})$	1.244	1.220	1.220	1.224	1.243	1.213	1.216	1.237
$r(\text{C4N8})$	1.362	1.366	1.363	1.361	1.361	1.360	1.360	1.324
$\angle(\text{N1C2N3})$	116.0	116.0	116.1	116.3	115.9	115.9	116.1	118.6
$\angle(\text{C2N3C4})$	120.8	120.3	120.3	120.4	120.8	120.6	120.5	120.5
$\angle(\text{N3C4C5})$	123.3	124.2	124.2	123.9	123.2	124.1	124.0	121.5
$\angle(\text{C4C5C6})$	116.6	116.0	116.1	116.1	116.7	116.1	116.1	117.0
$\angle(\text{C5C6N1})$	120.1	119.9	119.9	120.0	120.1	120.0	120.0	121.2
$\angle(\text{C6N1C2})$	123.1	123.5	123.5	123.3	123.2	123.4	123.3	121.2
$\angle(\text{N1C2O7})$	118.5	118.2	118.2	118.2	118.5	118.3	118.3	118.9
$\angle(\text{N3C4N8})$	116.9	116.9	116.9	117.0	116.8	117.0	117.0	118.3

前面的分析表明, 密度泛函理论适合研究胞嘧啶这样的碱基分子. 为了进一步检验密度泛函方法的可靠性, 我们采用上面选取的基组 6-311 + G**, 考察了 3 种不同理论方法 HF、B3LYP、MP₂ 对 Cyt1 构型的优化结果 (见表 2). 为了便于比较, 表 2 中最后一列数据也是同一个实验测得的 Cyt1 的结构参

数^[14]. 从表 2 给出的结果可以看出, HF 方法计算的结果与实验数据偏离较大. B3LYP 和 MP₂ 这两种方法对 Cyt1 构型优化的结果相近, 都与实验结果符合得较好, 而 B3LYP 的计算结果能稍好一些, 特别是键长优化的结果更接近实验值. 通过以上的比较分析可以看出, B3LYP/6-311 + G** 方法对胞嘧啶

异构体 Cyt1 的构型优化,与实验结果最接近,是研究胞嘧啶分子的理想方法.因此,在所有计算中,我们都采用 B3LYP/6-311 + G** 方法.

表2 HF、B3LYP 和 MP₂ 在 6-311 + G** 下对 Cyt1 构型的计算结果(键长: Å, 键角:(°))

Table 2 Calculated structural parameters (bond length in Å and bond angle in(°)) for Cyt1

Parameter	HF/ 6-311 + G**	B3LYP/ 6-311 + G**	MP2/ 6-311 + G**	Expt. [14]
r(N1C2)	1.402	1.428	1.419	1.392
r(C2N3)	1.361	1.369	1.381	1.358
r(N3C4)	1.295	1.317	1.316	1.339
r(C4C5)	1.445	1.440	1.439	1.433
r(C5C6)	1.338	1.356	1.362	1.357
r(C6N1)	1.347	1.354	1.358	1.360
r(C2O7)	1.192	1.216	1.219	1.237
r(C4N8)	1.346	1.360	1.373	1.324
∠(N1C2N3)	116.5	116.1	116.1	118.6
∠(C2N3C4)	120.5	120.5	119.9	120.5
∠(N3C4C5)	123.9	124.0	124.5	121.5
∠(C4C5C6)	115.6	116.1	116.0	117.0
∠(C5C6N1)	120.4	120.0	119.6	121.2
∠(C6N1C2)	123.1	123.3	123.9	121.2
∠(N1C2O7)	118.4	118.3	118.9	118.9
∠(N3C4N8)	117.7	117.0	117.0	118.3

本工作中的所有计算都是采用高斯 98 程序包 (Gaussian 98) 软件 (A. 7) 完成的. 因为带有氨基的分子一般是非平面的, 胞嘧啶分子也不能例外. 为了使计算结果具有普遍性, 我们在优化胞嘧啶异构体的初始构型时并未将其限制在一个平面内.

3 结果与讨论

在 B3LYP/6-311 + G** 水平下, 我们对 8 种胞嘧啶异构体 (Cyt1 ~ Cyt8, 见图 1) 的基态构型进行了充分优化, 同时进行了频率计算, 所有频率都是正的 (all frequencies real), 保证了优化构型是势能面上的能量极小点 (local minima), 并对基态能量进行了零点能校正 (zero-point energy corrections). 计算结果列于表 3. 根据我们的计算, 胞嘧啶最稳定的异构体是氨-酮式异构体 Cyt1, 其次是两个氨-烯醇式

(amino-enol) 异构体 Cyt2 和 Cyt3, 它们与 Cyt1 的能量差分别是 5.0 和 7.9 kJ/mol. 另一个氨-酮式异构体 Cyt4, 其能量比 Cyt1 高 29.2 kJ/mol. 两个亚氨-酮式 (imino-keto) 异构体 Cyt5 和 Cyt6 的能量都比 Cyt4 低, 但比 Cyt2 和 Cyt3 高, 它们与最稳定的 Cyt1 能量相差分别为 9.2 和 15.9 kJ/mol. 而两个亚氨-烯醇式 (imino-enol) 异构体 Cyt7 和 Cyt8, 在这 8 个异构体中能量是最高的, 它们与 Cyt1 的能量差分别是 58.9 和 79.0 kJ/mol. 因此, 在下面研究胞嘧啶互变异构化过程时, 没有考虑异构体 Cyt7 和 Cyt8.

表3 B3LYP/6-311 + G** 方法计算得到的胞嘧啶 8 种异构体和 5 个过渡态构型的能量

Table 3 Calculated energies for 8 tautomers and 5 transition-state geometries of cytosine at B3LYP/6-311 + G** level

Compound	Energy/Hartree
Cyt1	-394.9551769
Cyt2	-394.9533061
Cyt3	-394.9520696
Cyt4	-394.9440696
Cyt5	-394.9517405
Cyt6	-394.9490955
Cyt7	-394.8425915
Cyt8	-394.8347944
TS12	-394.9006168
TS23	-394.9394799
TS34	-394.8950470
TS15	-394.8914084
TS56	-394.9170599

如前所述, 在胞嘧啶分子中, 由于其中一个 H 原子位置不同, 导致胞嘧啶存在多种异构体形式. 当然, 也存在两个 H 原子同时变换位置的可能性, 但从能量角度考虑, 这种变换的可能性极小. 因此, 我们从胞嘧啶异构体 Cyt1 出发, 研究了包括 Cyt1 在内的 6 种胞嘧啶异构体的互变异构化过程. 图 2 是我们找到的连接胞嘧啶异构体 Cyt1 ~ Cyt6 的 5 个过渡态构型. TS12 表示连接稳定构型 Cyt1 和 Cyt2 的过渡态构型, 以此类推, TS23、TS34、TS15、TS56 分别表示连接 Cyt2 和 Cyt3、Cyt3 和 Cyt4、Cyt1 和 Cyt5、Cyt5 和 Cyt6 的过渡态. 优化使用的基组同优化基态构型的基组相同, 都是 6-311 + G**. 为了保

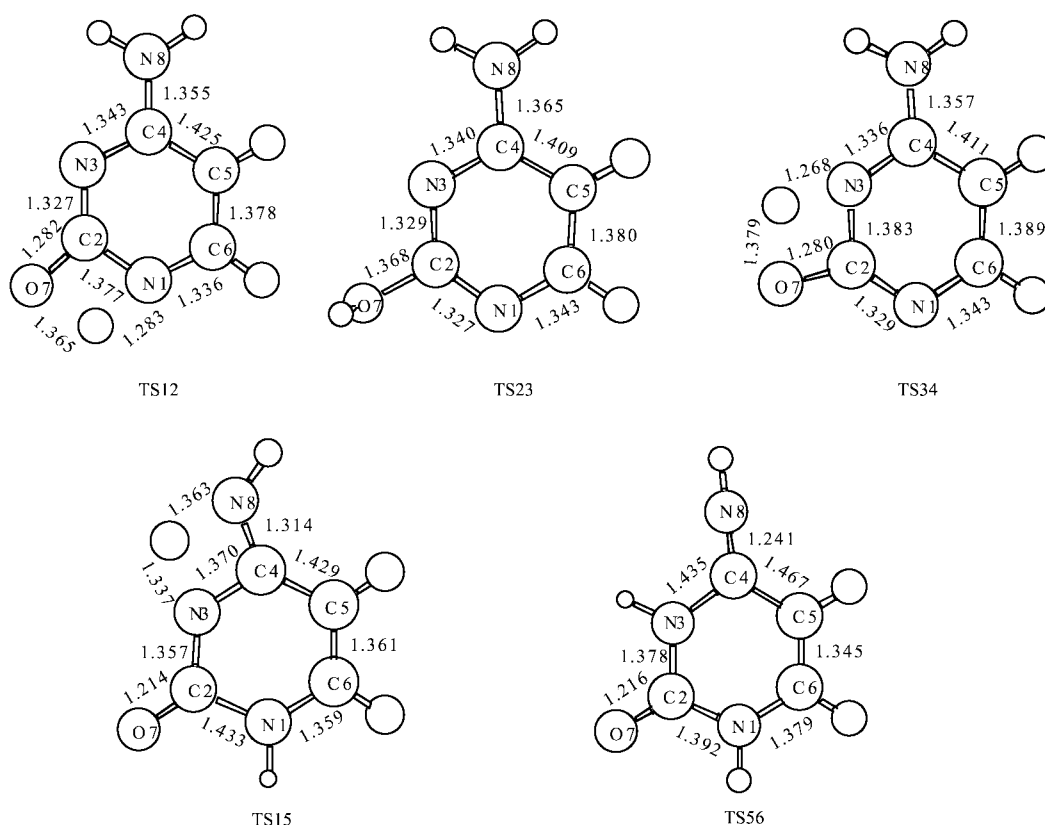


图2 B3LYP/6-311 + G** 方法计算得到的 5 个胞嘧啶过渡态构型

Fig. 2 Calculated 5 transition-state geometries (distances in Å) of cytosine at B3LYP/6-311 + G** level

证我们所找出的过渡态构型确实是连接相应的异构体,我们做了详尽的内禀反应坐标(IRC)计算,如图3所示.结合图1和图2给出的优化构型,可以看出,在Cyt1向Cyt2的异构化过程中,连接在N1上的H原子与N1原子分开,向O7原子方向迁移,停留在N1和O7之间,形成过渡态TS12.而过渡态构型TS12中的其它结构参数与稳定构型Cyt1或Cyt2相比变化不大.由Cyt1向Cyt2转化,需越过的能垒高度为142.9 kJ/mol;由Cyt2向Cyt1转化,需越过的能垒高度为138.3 kJ/mol.在两个氨-烯醇式异构体Cyt2和Cyt3的互变异构化过程中,O7-H键绕C2-O7键旋转约 90° ,形成过渡态TS23.在Cyt2和Cyt3的相互转化过程中,只涉及O7上的H原子的转动,没有化学键的断裂和生成,所以Cyt2和Cyt3之间的能垒较低,分别为36.3和33.0 kJ/mol.在氨-烯醇式异构体Cyt3向氨-酮式异构体Cyt4转化过程中,O7上的H原子向N3方向迁移,位于O7和N3两个原子之间时,形成过渡态TS34,正向和反向异构化的能垒高度分别是149.6和128.7 kJ/mol.

由于胞嘧啶的亚氨-烯醇式异构体(Cyt7、Cyt8)

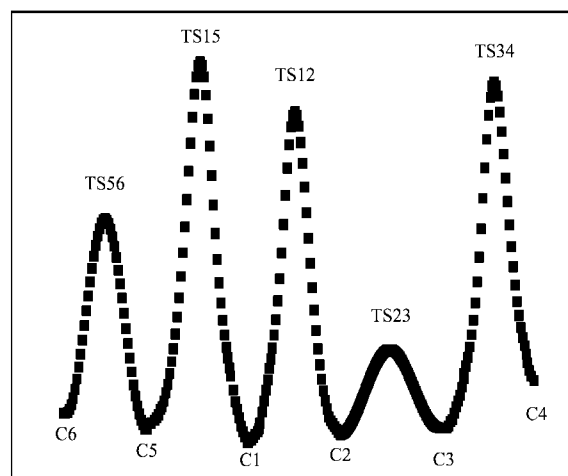


图3 B3LYP/6-311 + G** 方法对胞嘧啶异构化的内禀反应坐标计算结果

Fig. 3 Calculated Intrinsic Reaction Coordinate for tautomerism of cytosine at B3LYP/6-311 + G** level

自身能量较高,而且在胞嘧啶的氨式和亚氨式异构体的互变异构化过程中,两个或两个以上H原子同时变换位置的可能性极小,因此,我们只考察了

氨-酮式异构体(Cyt1)和亚氨-酮式异构体(Cyt5、Cyt6)之间的相互异构化过程.在Cyt1向Cyt5的异构化过程中,C4-N8键向N3原子倾斜,N8原子的一个H原子向N3原子方向迁移,位于N3和N8原子之间,形成过渡态TS15.在Cyt1、Cyt5和TS15三个构型中,除了H原子的位置不同外,键角N3C4N8的变化比较明显.Cyt1异构化为Cyt5需要克服的能垒高度为167.2 kJ/mol,反方向异构化的能垒高度为158.4 kJ/mol.在亚氨-酮式异构体Cyt5和Cyt6之间的互变异构化过程中,由于C4=N8键为双键,C4-N8键无法带着N8上的H原子转动,因此位于N8侧向的H原子近似在一平面内向N8原子上方迁移,到达N8原子上方时形成过渡态TS56.Cyt5向Cyt6异构化的能垒高度为91.1 kJ/mol,反方向异构化的能垒高度为84.0 kJ/mol.与氨-烯醇式异构体Cyt2、Cyt3之间的异构化相比,亚氨-酮式异构体Cyt5、Cyt6之间的互变异构化的能垒要高些,这是因为虽然这两类异构化都不涉及化学键重组,但Cyt2、Cyt3之间的异构化是由C2-O7键的转动实现的,而在Cyt5、Cyt6中,C4=O8键为双键,无法转动,其异构化是由N8原子上的H原子迁移实现的.

实验研究表明^[3-6],在常温常压下,胞嘧啶分子以氨-酮式异构体Cyt1形式存在;在水溶液中,胞嘧啶分子主要以氨-酮式异构体Cyt1形式存在,同时存在少量Cyt4,实验测得在水溶液中Cyt1约是Cyt4的800倍;在气相实验中发现,气相状态胞嘧啶以氨-酮式异构体Cyt1及氨-烯醇式异构体Cyt2和Cyt3形式存在,同时,不能排除存在少量亚氨-酮式异构体Cyt5的可能性,但无任何迹象表明有亚氨-烯醇式异构体的存在.分析实验结果可以发现,胞嘧啶无论在气相、液相还是固相,都存在氨-酮式异构体Cyt1,说明Cyt1应该是胞嘧啶分子最稳定的异构形式之一.我们的计算表明,Cyt1是8种异构体中能量最低的,理论预测与实验事实相符.在气相条件下,胞嘧啶存在Cyt1、Cyt2、Cyt3异构体形式,也可能有Cyt5存在,这一实验事实也可以由我们研究胞嘧啶异构化的结果予以解释.计算表明,Cyt1向Cyt2异构化需越过的能垒(142.9 kJ/mol)比Cyt1向Cyt5异构化需越过的能垒(167.2 kJ/mol)小,所

以尽管异构体Cyt5也很稳定,能量与Cyt2差不多,但Cyt1的异构化主要还是生成Cyt2.而Cyt2和Cyt3之间的能垒很低,Cyt2很容易异构化生成Cyt3.异构体Cyt1、Cyt2、Cyt3能量相近,都很稳定,所以在气相实验中观察到的胞嘧啶的异构体主要是Cyt1、Cyt2、Cyt3这三种形式.

参 考 文 献

- [1] Les' A, Adamowicz L, Bartlett R J. *J. Phys. Chem.*, 1989, **93**: 4001
- [2] Szczepaniak K, Szczesniak M. *J. Mol. Struct.*, 1987, **156**: 29
- [3] Radchenko E D, Sheina G G, Smorygo N A, Blagoi YU P. *J. Mol. Struct.*, 1984, **116**: 387
- [4] Szczesniak M, Szczepaniak K, Kwiatkowski J S, KuBulat K, Person W. B. *J. Am. Chem. Soc.*, 1988, **110**: 8319
- [5] Nowak M J, Lapinski L, Fulara J. *Spectrochimica Acta*, 1989, **45A**: 229
- [6] Brown R D, Godfrey P D, McNaughton D, Pierlot A P. *J. Am. Chem. Soc.*, 1989, **111**: 2308
- [7] Szczepaniak K, Shugar D. *Int. J. Quantum Chem.*, 1981, **20**: 573
- [8] Scanlan M J, Hillier I H. *J. Am. Chem. Soc.*, 1984, **106**: 3737
- [9] Zhang C (张翀), Bu Y X (步宇翔). *Chin. J. Chem. Phys.* (化学物理学报), 2001, **14**: 75
- [10] Li J (李江), Zhou X G (周晓国), Pei L S (裴林森), Chen C X (陈从香), Yu S Q (俞书勤), Ma X X (马兴孝). *Chin. J. Chem. Phys.* (化学物理学报), 2002, **15**: 224
- [11] Pei K M (裴克梅), Li Y M (李益民), Kong X L (孔祥蕾), Li H Y (李海洋). *Chin. J. Chem. Phys.* (化学物理学报), 2003, **16**: 251
- [12] Jin Y L (金艳玲), Mu X L (牟晓兰), Zhang B L (张柏林), Wang X Y (王秀岩). *Chin. J. Chem. Phys.* (化学物理学报), 2003, **16**: 445
- [13] Kwiatkowski J S, Bartlett R J, Person W B. *J. Am. Chem. Soc.*, 1988, **110**: 2353
- [14] Rich A, Voet D. *Prog. Nucleic Acid Res. Mol. Biol.*, 1970, **10**: 196