

模拟退火方法研究线性肽构象与环化的关系*

邓巧临 来鲁华 韩玉真 苗振伟 季爱雪 徐筱杰

(北京大学化学系, 北京 100871)

摘要 采用蒙特卡罗模拟退火方法对一系列线性五肽进行了构象分析, 探索了线性肽的优势构象与其环化产率之间的关系. 构象分析所得到的预测结果与实验结果吻合.

关键词: 蒙特卡罗模拟退火, 构象分析, 优势构象, 环化

1 前言

人们对分子构象的研究着重于寻找分子的优势构象, 即低能量的构象. 这是分子存在的最可几方式, 也常常被认为与分子起作用时的存在状态有关^[1]. 因此, 进行构象分析的重点是寻找分子的低能量构象. 多肽分子是柔性分子, 它们的势能函数主要是分子中可旋转二面角的函数. 这是一个多极值函数, 局部势阱极多, 如果采用传统的最小化方法难以解决构象分析问题. 模拟退火方法^[2-4]是寻找多极值函数最优值的有效算法, 它采用蒙特卡罗 Metropolis 算法^[5]来搜索构型空间. 体系不仅用能量作为参量, 同时也把温度作为参量. 模拟退火方法是对柔性分子进行构象分析^[6-9]的一个可行的方法. 因此我们将该方法应用于多肽分子的构象分析.

同一个环肽可以由一系列线性肽环化而成, 氨基酸顺序不同, 环化的难易程度也不同. 影响环化产率的因素很多, 如所采用的环化方法, 所用的溶剂等等. 特别是线性肽的构象这个内在的因素是不容忽视的. 在本文中, 我们采用了模拟退火方法分析系列线性肽的优势构象, 探讨了线性肽的环化难易程度与构象的关系.

2 方法

模拟退火 (simulated annealing) 的概念是分别由 Kirkpatrick、Gelatt、Vecchi 三人 (1982)^[2] 和 Cerny^[3](1985) 独立提出的.

蒙特卡罗模拟退火过程可分为二个步骤: (1) 首先对体系在一个较高的初始温度 T_0 下进行“加热”, 使之“熔化”; (2) 体系在熔化之后, 逐渐降温.

在任一温度 T_k 时, 体系的构象为 x_i 相应的能量为

$$E_i = E(x_i) = E(x_{i1}, x_{i2}, \dots, x_{in})$$

1992-04-06 收到初稿, 1993-06-21 收到修改稿. 联系人: 徐筱杰. *863 计划资助项目

然后, 给构象一个随机变化 Δx , 则新构象为 $x_{i+1}=x_i+\Delta x$, 相应的能量为

$$E_{i+1} = E(x_{i+1}) = E(x_{(i+1)1}, x_{(i+1)2}, \dots, x_{(i+1)n})$$

体系变化前后的能量差为

$$\Delta E = E_{i+1} - E_i$$

这时采用 Metropolis 算法来判断是否接受这种变化, 具体判别标准如下:

如果 $\Delta E \leq 0$, 则接受变化, 新的构象作为下一步的起点;

如果 $\Delta E > 0$, 则构象被接受的几率为 $p(\Delta E) = \exp(-\Delta E/k_B T)$, 即在 (0,1) 间均匀分布的随机数中取一数 R , 让它与 $p(\Delta E)$ 相比较, 如果 $R < p(\Delta E)$, 则接受新构象, 否则拒绝新的构象, 仍将旧的构象作为下一步的起点;

如此循环往复, 就可以模拟出体系在温度 T_k 时构象的波兹曼分布. 在每一个温度下都如此循环, 直至温度从 T_0 降至足够低.

从上述的判据可以明显地看出, 模拟退火方法与传统的分子力学方法的不同之处, 即它不仅接受能量下降的变化, 同时也按一定的几率来接受能量上升的变化. 正是这点的不同使它优于传统的分子力学方法.

在模拟退火方法中, 退火参数的设置是影响退火效率的关键^[4]. 比较重要的参数有: 起始温度, 每一温度下构象变化的次数, 决定降温快慢的降温因子和终止温度. 对于体系来讲, 起始温度必须足够高, 扩大体系的运动范围, 保证体系“熔化”, 这样体系才会在更大的空间中运动, 也才能跨越更大的势垒. 而每一温度下构象变化的次数也必须足够大, 否则无法正确地模拟体系在该温度下构象的波兹曼平衡. 降温因子必须适当, 以保证体系温度缓慢下降以求体系达到稳定态, 否则急冷会导致体系达到亚稳态. 最后一个重要因素是终止温度必须足够低, 保证体系在该温度下“冻结”, 即构象稳定在一个特定形式, 而不再任意变化. 合适的退火进程可以使体系逐渐地变化, 最终达到整体最优构象或接近最小值的低能量构象. 通常对于一个体系的研究, 仅作一次模拟退火是不够的, 需要进行一组计算, 来寻找分子的稳定构象. 从一组的結果可以分析分子的构象分布, 还可以反映出分子的柔性大小.

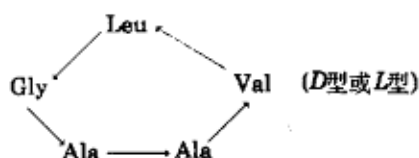
我们采用了 ECEPP/2 力场 (H. A. Scheraga, 1974, 1975, 1983)^[10-12] 编制了适用于多肽分子的程序 SAPEP. ECEPP/2 力场适宜于多肽分子的计算. 总能量由静电能, 非键能 (包括氢键能) 和扭角能三项组成.

目前程序的最大适用范围是不超过 300 个原子且二面角变量不超过 40 个的体系. 对于一般的多肽分子, 可以预测十几肽的构象. 程序对于多肽分子构象分析不限制是 L-型氨基酸, 还是 D-型氨基酸, 只要初始构型确定了 L-型或 D-型, 则程序在计算中不会改变其手性. 程序可以在 Micro VAX-II -VMS 系统或 Silicon Graphics IRIS 4D/70GT 系统上运行.

3 结果与讨论

为了研究线性肽环化为环肽的合成问题, 季爱雪等人曾设计并合成了一系列线性五肽, 并且进行了环化, 环肽由五个氨基酸组成, 其结构式如下. 实际上这是两个环肽, 一

个全部由 *L*-型氨基酸残基组成, 另一个含有一个 *D*-型氨基酸. 环肽中五个肽键均可切断, 则对应于每一个环五肽都可以由五个不同的氨基酸顺序排布的线性肽环化而成. 线性肽的氨基酸顺序标号如下:



1D	Gly-Ala-Ala-DVal-Leu	1L	Gly-Ala-Ala-Val-Leu
2D	Ala-Ala-DVal-Leu-Gly	2L	Ala-Ala-Val-Leu-Gly
3D	Ala-DVal-Leu-Gly-Ala	3L	Ala-Val-Leu-Gly-Ala
4D	DVal-Leu-Gly-Ala-Ala	4L	Val-Leu-Gly-Ala-Ala
5D	Leu-Gly-Ala-Ala-DVal	5L	Leu-Gly-Ala-Ala-Val

此中凡是含 *D*-型氨基酸的线性肽环化为同一个环五肽, 不含 *D*-型氨基酸的线性肽环化为另一个环五肽, 用不同方法进行环化得到的产率列于表 1 中.

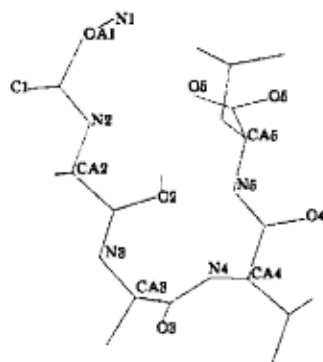


图 1 线性肽 1D 的优势构象示意图
Fig.1 Favorable conformation of peptide 1D

从表 1 可看出, 虽然产物是同一个环五肽, 但作为反应物的线性肽中氨基酸顺序不同, 环化产率相差得很大. 构象是线性肽的本质结构特征, 很可能在环化中起了重要作用. 为此, 我们选取了有产率数值的线性肽进行构象分析, 寻找优势构象, 试图探索构象与产率的关系.

构象分析的结果显示, 这些线性肽的优势构象中都存在转角结构, 氮端和羧端相互靠近, 倾向于形成环状结构. 图 1 是线性肽 1D 的优势构象的示意图, 为清楚起见, 图中只显示了非氢原子.

表 1 线性肽环化产率

Table 1 Yields of cyclization of the linear peptides

线性肽	1D	2D	3D	4D	5D
DPPA 法	0.73	0.184	/	0.161	0.12
迭氮法	0.465	0.37	/	0.21	0.16
线性肽	1L	2L	3L	4L	5L
DPPA 法	0.289	0.04	/	微量	微量
迭氮法	0.363	0.28	/	/	0.11

考虑构象特点, 我们选择线性肽氮端的 N 原子和羧端 C 原子之间距离 d 作为研究对象, 定义 d_{\min} 表示最低能量构象的距离值, d_{wavg} 表示各构象两端距离的加权平均值 ($d_{\text{wavg}} = d_i p_i$, 式中 p_i 是构象 i 出现的几率). 多肽是柔性分子, 因此可能存在多个低能量的构象, 而各个构象由于其能量的高低不同, 因而出现几率 p_i 不同. d_{wavg} 是各个构象的端基距离 d 的几率加权平均值, 比 d_{\min} 更有效地反映出多肽分子的整体结构特征. 为了便于寻找结构和产率的相关性, 表 2 按多肽环化产率下降的次序列出了距离值.

表 2 线性肽的产率和两端距离值

Table 2 Yields of cyclization and distance of terminals of linear peptides

线性肽	产率	$d_{\min}/\text{\AA}$	$d_{\text{wavg}}/\text{\AA}$
1D	0.73 0.465	3.88	3.77
2D	0.184 0.37	3.79	3.99
4D	0.161 0.21	3.85	3.98
5D	0.12 0.16	5.32	4.26
1L	0.289 0.363	4.49	4.47
2L	0.04 0.28	3.95	4.00
5L	微量 0.11	6.68	6.33

表 2 中的数据表明产率与构象间确有一定关系. 1D 产率最高, 其 d_{wavg} 最小, d_{\min} 次小, 说明 N 端和 C 端距离很近, 容易环化. 这是由于这个肽的 N 端为甘氨酸, 空阻小. 产率最低的 5L, 其 N 端和 C 端距离明显大于其它几个肽, 最难环化. 其产率低的原因是两端均是带大侧链的残基, 空阻较大, 使得两端不能充分接近, 而不利于环化.

表 2 的上部分是含 D-型氨基酸的线性肽, 其环化产率依次为 $1D > 2D > 4D > 5D$. 构象分析的结果显示产率应为 $1D > 2D \approx 4D > 5D$, 两者在顺序上是完全一致的. 表 2 的下部分是全 L-型的线性肽, 环化产率 1L 较高, 2L 居中, 5L 较低. 构象分析的结果在次序上有所不同, 2L 较高, 1L 居中, 5L 仍较低, 1L 和 2L 的次序颠倒. 仔细观察 2L 的产率可以发现用两种不同方法环化, 其产率相差很大, 这反映了在这个线性肽的环化中影响产率的因素较复杂, 因此造成了预测与实验的不符.

总体上看, 存在一个大致的趋势, 即产率高的线性肽其优势构象的两端间距小, 产率低的线性肽两端间距大, 产率居中的线性肽两端间距也居中. 此外, 从表 2 中结果还可知, 凡含 D-型氨基酸都比其相应的全 L-型线性肽两端间距小, 即在结构上更倾向于成环, 这与环化产率的高低规律是一致的.

参 考 文 献

- 1 Levitt M. *Ann. Rev. Biophys. Bioeng.*, **1982**, **11**: 251
- 2 Kirkpatrick S, Gelatt C D, Vecchi M P. *Science*, **1983**, **220**: 671
- 3 Cerry V J. *Opt. Theo. Appl.*, **1985**, **45**: 41
- 4 Arts E H L, *Simulated Annealing and Boltzmann Machines: A Stochastic Approach to Combinatorial Optimization and Neural Computing*. Chichester: Wiley, 1989, 1-114

- 5 Metropolis N, Rosenbluth A W, Rosenbluth M N, Teller A H, *et al.* *J. Chem. Phys.*, **1953**, **21**: 1087
- 6 Wilson S R, Cui W, Moskowitz J W, *et al.* *Tetrahedron Lett.*, **1988**, **29**: 4373
- 7 Moskowitz J W, Schemidt K E, Wilson S R, Cui W. *Int. J. Quant. Chem.*, **1988**, **22**: 611
- 8 Kawai H, Kikuchi T, Okamoto Y. *Protein Engineering*, **1989**, **3**: 85
- 9 Pannetier J, Bassas-Alsina J, Rodriguez-Carvajal J, Cargnaert V. *Nature*, **1990**, **346**: 343
- 10 Momany F A, Carruthers L M, McGuire R F, Scheraga H A. *J. Phys. Chem.*, **1974**, **78**: 1595
- 11 Momany F A, Carruthers L M, Scheraga H A. *J. Phys. Chem.*, **1975**, **79**: 2361
- 12 Nemethy G, Pottle M S, Scheraga H A. *J. Phys. Chem.*, **1983**, **87**: 1883

Application of Simulated Annealing: Study on Relationship between Conformation of Linear Peptides and Their Cyclization Yields

Deng Qiaolin Lai Luhua Han Yuzhen Miao Zhenwei Ji Aixue Xu Xiaojie
(Department of Chemistry, Peking University, Beijing 100871)

Abstract Simulated annealing is a promising new method for conformational analysis of flexible molecules. It differentiates from the conventional molecular mechanics optimization in that it not only accepts the changes that the energy decreases, but also accepts some uphill changes. So it is possible to make transitions out of the local minima and reach the global minimum of the system. The program SAPEP has been made using ECEPP/2 force field for conformational analysis of peptides. A set of linear peptides which was cyclized into the same cyclopentapeptide was analysed using SAPEP in order to explore the relationship between the conformation and yield of cyclization. The results show that relationship exists between the conformation and yield.

Keywords: Monte-Carlo simulated annealing, Conformational analysis, Favorable conformation, Cyclization