嘉南藥理科技大學專題研究計畫成果報告

以圖形理論推導烷類同分異構物的立體結構

計畫類別:■個別型計畫
計畫編號:CNMI93-03
執行期間:93年1月1日至93年12月31日
カロは日、オリ
計畫主持人:劉育寰
共同主持人:
計畫參與人員:

執行單位:資管系

中華民國 九十三 年 二 月 二十九 日

嘉南藥理科技大學專題研究計畫成果報告 以圖形理論推導烷類同分異構物的立體結構

計畫編號: CNMI93-03 執行期限: 93 年 1 月 1 日至 93 年 12 月 31 日 主持人:劉育寰 嘉南藥理科技大學 資訊管理系

中文摘要

在先前研究中已發展出一個新的演算法可 以自動求出 10 個碳數以下的烷類同分異構物, 而此演算法是先建立所有支鏈的碳可以擺放位 置的取樣空間(Sampling Space),求出所有組合, 然後再做對稱性與重複性的消除。但該計畫並無 法處理支鏈碳數為 3 以上的狀況,例如在碳數為 10 取出 3 個碳及取出 5 個碳的情況下會有支鏈 碳數為 3 的狀況發生。本計畫則擬以圖形理論 (Graph Theory)來解決這個問題。

在先前研究中,計算取樣空間的組合時所採 用的是"排列的演算法",故需要多花費比較多 的時間。尤其是取樣空間比較多的時候,有時需 要14!次的計算。以 Pentium 4 2.4 GHz 的個人電 腦計算總碳數 10 的同分異構物所發的時間約為 10 分鐘。在本研究中擬將"排列"的演算法改 成"組合"的演算法。

關鍵詞:同分異構物、烷類、Java 3D、分子模型

Abstract

In our previous study, a new algorithm has been developed for automatically finding the isomers of decane. Our previous algorithm is based on the sampling space where the branching carbon can be placed on. Find out all the combination of the sampling space except the symmetry and redundancy. However, our previous algorithm can not handle the case which have more than three branching carbons, e.g., decane has such situation when main chain contains 3 or 5 carbons. In this study, the graph theory will be included to overcome this problem.

In previous study, it is very time consuming to calculate the combination of the sampling space by means of "permutation algorithm". Sometimes, it needs 14! times calculation, it almost spends more than 10 minutes by Pentium 4 2.4GHz PC. In this study, the "combination algorithm" is used instead of "permutation algorithm".

Keywords: Isomers, Alkane, Java 3D, Molecular Modeling.

Introduction

本計畫主要針對先前研究的一些方法提出 改善:

- 1. 將"排列演算法"改為"組合演算法"。
- 修改顯示的畫面。加入鍵結,及選擇觀 察任一個同分異構物的選項。
- 3. 建立支鏈有3個碳的取樣空間。
- 以圖形理論的多形(isomorphism)來消除 重複性。
- 以下則針對這些主題詳細敘述其作法。

Method

將"排列演算法"改爲"組合演算法"

在先前的研究中以"排列演算法"來計算取 樣空間,所以有時需有14!次的計算,相當費時。 將其改為"組合演算法"計算次數將可降低為 2^{14} 次以內。組合演算法是將一個集合(如{1, 2, ..., n})以"字典編纂"順序(lexicographic order)選出 r 個元素,稱之為 r-組合(r-combinations)。 $a_1a_2...a_n$ 的下一個組合可由下列方法求得:首先選出最後 一個元素 a_i , 使的 a_i 不等於 n-r+i; 然後 a_i 以 a_i +1 取代, a_j 以 a_i +j-i+1 取代, j=i+1, i+2, ..., r。其演 算法如下^[1,2]:

Algorithm: Generating the next r-combination in lexicographic order.

procedure next r-conbination($\{a_1, a_2, ..., a_r\}$: proper subset of $\{1, 2, ..., n\}$ not equal to $\{n-r+1, ..., n\}$ with $a_1 < a_2 < ... < a_r$)

 $\begin{array}{l} i{:=}r \\ \textbf{while} \ a_i{=}n{-}r{+}i \\ i{:=}i{-}1 \\ a_i{:=}a_i{+}1 \\ \textbf{for} \ j{:=}i{+}1 \ \textbf{to} \ r \\ a_j{:=}a_i{+}j{-}i \end{array}$

修改顯示的畫面

先前研究中所開發的軟體並沒有將鍵結的 部分描繪出來,所以比較分辨不出來其立體關 係。在本研究中,以Java 3D的 Cylinder 幾何物 件來描繪鍵結的部分。由於在建構 Java 3D 中的 Cylinder 幾何物件時只能指定半徑及長度,至於 旋轉的角度需要另外指定。因此,定義了下列幾 個參數來建構 Cylinder 物件以表示 C-H 及 C-C 鍵結。

如同先前計畫所述,假設座標原點在碳的中 心點,且H1在y-軸上,H2在y-z平面上(如圖 1所示),求得4個氫的座標如公式(1)~(4)。假設 表示鍵結的Cylinder物件之半徑為r/15,長度為 r,且對所有的鍵結的長度及半徑都一樣。由於 所建構出來的Cylinder物件都以原點為中心,y-軸為方向,所以對每個鍵結須有不同的位移量及 旋轉角度。表格1為各鍵結的位移及旋轉量。

因為描繪時是以 C 及 4 個氫為單位,所以 假設 C-H 的鍵結長度為 r,那麼 C-C 的鍵結長度 就應為 2r。如此,在描繪基本單元時,只要將碳 的位置沿著鍵結方向位移鍵結長度的 2 倍,便可 將 C-C 鍵節連結在一起。

表格 1 各鍵結的位移及旋轉量,其中x表示外積。

鍵結	位移	角度	旋轉軸
C-H1	(0.0, r/2, 0.0)		to an and the second
C-H2	(0.0, -h/2, -a/2)	-109.5°	x-軸
C-H3	$(\sqrt{3}a/4, -h/2, a/4)$	109.5°	C-H1 × C-H3
C-H4	$\left(-\sqrt{3}a/4, -h/2, a/4\right)$	109.5°	C-H1 × C-H4

$$H1 = (0.0, r, 0.0)$$
(1)

$$H2 = (0.0, -h, -a)$$
(2)

$$H3 = \left(\frac{\sqrt{3}}{2}a, -h, \frac{a}{2}\right)$$
(3)

$$H4 = \left(-\frac{\sqrt{3}}{2}a, -h, \frac{a}{2}\right)$$
(4)

 $\ddagger r = 1.1 \text{ Å}, \ \theta = 109.5^{\circ} - 90^{\circ}, \ h = r \sin \theta,$ $a = r \cos \theta \circ$



圖 1 碳與氫的相對立體座標。假設座標原點在碳的中心

點,且H1在y-軸上,H2在y-z平面上。

建立支鏈有3個碳的取樣空間

在先前研究中提到^[3],若從 CNum 個碳中取 出 MaxC 個碳做為支鏈,那麼主鏈上的碳數為 SC = CNum - MaxC。主鏈上第 *i* 個碳可以接的 支鏈碳數原則上不可以超過它到主鏈端點的碳 數,且非端點主鏈上的碳都有 2 個鍵結可以接支 鏈的碳,所以主鏈上第 *i* 個碳可以接的支鏈碳數 為 $2 \times \min(i-1, MaxC - i)$,故主鏈碳數為 SC時, 總 共 可 以 接 的 支 鏈 碳 數 為 $C = \sum_{i=1}^{SC} 2 \times \min(i-1, MaxC - i)$,找出一個最大的 MaxC 使得 MaxC $\leq C$,該 MaxC 即為總碳數為 CNum 時可取出的最多碳數。表格 2 顯示 SC 與 CNum 之間的關係,在"結構"欄位中的"〇"表示 在該位置上可以接一個支鏈的碳。

但是如今要考慮支鏈有3個碳的情況,就必 須修改表格2的結構。底下是幾個主要的考量:

- 在支鏈的末端上不接超過1個碳,因為 那會造成支鏈過長,而會與主鏈數+1有 多型圖(isomorphism),如圖2所示。圖 中"◎"表示支鏈末端。
- 在非末端的支鏈上最多可以再接 2 個 碳,如圖 3 所示。其中"●"或"0"表示 第1層有分支的支鏈。第2層以上的支 鏈分之在本研究中先當成特例來處理。
- 當支鏈的碳數超過 3 個(含)以上,才有 可能接到圖 3 中的●位置上,但支鏈上 的碳數剛好只有 3 個時,可能會與主鏈 數+1 有多型圖,如圖 3 所示。

由上述幾個考量點得知,同一個支鏈上最多 可接的碳數可以為4,所以修正表格2的結構及 總碳數,如紅色部分所示。而總碳數、最少主鏈 碳數、可取出碳數NC及主鏈碳數SC之關係亦 修正如表格3所示。

表格 2 先前計畫中提到:主鏈碳數為 SC,最多會有多少 個總碳數 CNum。^[3]

S C	結構	CNum
3	$ \begin{array}{c} \bigcirc \\ \\ -C - C - C - \\ \\ \bigcirc \end{array} $	4~5
4	$ \begin{array}{cccc} $	5~8







圖 4 在非末端支鏈上的碳數剛好只有 3 個時,可能會與 主鏈數+1 有多型圖(isomorphism)。

表格 3 總碳數、最少主鏈碳數、可取出碳數 NC 及主鏈碳數 SC 之關係。

	CNum	MSC	SC	NC	<i>F1</i>	F2	F3	<i>F4</i>
			9	1	1			
			0	2		1		
			8	2	2			
				3			1	
			7		1	1		
			10.00		3			
				4				1
		5	6		1		1	
	10					2		
1		A			2	1		
					4			
					1			1
		1.4	- 1	§	2		1	
			5	5		1	1	
			5	5	1	2		
				-	3	1		
l				12	5			
		11	10	1	1			
		1	0	2		1		
1	1 T	-	9	2	2			
			8				1	
1				3	1	1		
					3			
								1
			7	4	1		1	
						2		
					4			
					1			1
	11	5				1	1	
	11	5	6	5	2		1	
			-		1	2		
					5			
						1		1
					2			1
							2	
					1	1	1	
			5	6	3		1	
						3		
					2	2		
					4	1		
					6			
	12	5	11	1	1			
			10	2		1		
			10		2			
			9	3			1	

圖 3 在非末端的支鏈上最多可以再接 2 個碳。

CNum	MSC	SC	NC	<i>F1</i>	<i>F2</i>	F3	<i>F4</i>		
				1	1				
				3					
							1		
				1		1			
		8	4		2				
				2	1				
				4					
				1			1		
					1	1			
		7	~	2		1			
		/	5	1	2				
				3	1				
				5					
					1		1		
				2			1		
						2			
				1	1	1			
		6	6	6	3		1		
							3		
				2	2				
				4	1	-	1997 - C.		
				6	and the second second		and the second second		
					and a start of the second	1	1	1	
				1	1		1		
			7	3		and an	1		
				1		2			
		-			2	1			
		5		2	1	1		1	
				4	15	1			
				1	3				
				3	2			1	
		- 10-	16	5	1	No. of Concession, Name			
L				5	1	100	1	J	

以圖形理論的多型(isomorphism)來消除重複性

只有在主鏈碳數 SC 大於等於 5 且支鏈碳數 大於等於 3 才有可能出現多型圖。圖形的多型可 以由幾個不變性(invariance)來做判斷:^[4-7]

- 1. 圖形的總節點(vertex)數要一致
- 2. 圖形的連結(edge)數要一致
- 3. 相同維度(degree)的節點被歸為同一類 別,而圖形的類別總數要一致
- 4. 圖形的總節點維度要一致

其中第1點是一定會符合的,因爲總結點數 就是總碳數,同分異構物本來就是在總碳數相同 的情況下才會有的。而第2點以一定會符合,因 爲目前考慮的是烷類結構碳跟碳之間只有單鍵 且沒有環狀結構,所以連結數一定會等於總碳數 減1。

所謂節點的維度是指某個節點有幾個連結。以直鏈的結構為例,兩個端點的碳之維度為 1,其餘的碳之維度為2。若有之鏈存在,則連 接支鏈的碳之維度至少是3。 所以先根據表格 3 這些參數所建構的取樣 空間,排列出所有可能的組合,並先做重複性與 對稱性的消除,然後計算它們的總節點維度與類 別總數。再對 F3 及 F4 參數不等於 0 的幾個結構 去搜尋並消除多型圖。如此便完成所有同分異構 物的建立。

Results

本計畫所建立的圖形介面如圖 5 所示。其 中最上方"C Number"為碳的個數,在本計畫中先 探討碳數為 10~12;左下方"Total Com #"為所有 的同分異構物數目;右下方的"Cur Com #"為目 前所顯示的同分異構物編號,可以直接輸入所欲 觀察的編號;最下方的"Rx"與"Rz"可用來改變視 角;中間區域則為同分異構物的立體圖形顯示 區。



圖 5 本計畫所設計的圖形介面,其中最上方"C Number" 爲碳的個數;左下方"Total Com #"為所有的同分異 構物數目;右下方的"Cur Com #"為目前所顯示的同 分異構物編號,可以直接輸入所欲觀察的編號;最 下方的"Rx"與"Rz"可用來改變視角。

支鏈碳數為三的幾個例子,如圖 6 所示。 從圖中可以很清楚的看出斐個分子的空間位置 關係。





Conclusion

Reference

- 1. Kenneth H. Rosen, *Discrete Mathematics and Its Applications*, 2nd edition (NY: McGraw-Hill, 1991), pp. 293-297.
- 2. Michael Gilleland, Merriam Park Software, *Combination Generator*, <u>http://www.merriampark.com/comb.htm</u>.
- 3. 劉育寰, 烷類同分異構物立體結構之建立與顯示, 嘉 南學報, pp. 19-32, 2004.
- 4. K. Riaz, M.S.H. Khiyal and M. Arshad, 2005. "Matrix equality: an application of graph isomorphism," *Information Technology Journal*, 4(1), pp. 6-10.
- 5. K. Riaz, M.S.H. Khiyal and M. Arshad, 2003. "An improved algorithm to discover graph isomorphism," *Proceedings INMIC 2003, 7th International Multi Topic Conference*, 8-9 December, pp. 396-401.
- 6. Danail Bonchev, Dennis H. Rouvray, and D. H. Rouvray, *Chemical Graph Theory: Introduction and Fundamentals (Mathematical Chemistry)*, Taylor & Francis (December 1, 1991).
- D. H. Rouvray, Dennis H. Rouvray, Danail Bonchev, *Chemical Topology: Introduction and Fundamentals* (*Mathematical Chemistry*), Gordon & Breach Science Publishers (April 1, 1999).

