

# 有机化合物分子式的组合方式和 分布规律的研究及其应用

丛 浦 珠

(中国医学科学院药用植物研究所 北京 100094)

[摘要] 多年的观察和研究初步证实,自然界有机化合物可能存在的合理分子式在种类和数量上是有限和可知的,对分子量 1000 以下的分子式范围提出了符合实际的界定方法和计算式。据此已经人工书写或计算机打印出分子量 16~800、只含 C、H、O、N 的分子式 203470 个,初步估算分子量到 1000 时共有分子式 474000 个。若以分子量为横坐标,分子式中的碳原子数为纵坐标作图,除个别外,这些分子式均分布在坐标系的一个三角形区域内,因而称其为三角形分布。但这只是自然界有机化合物分子式的一部分,全部则可能为流线形分布,三角形为流线形的头部。有机化合物分子式的三角形分布为有机化学家提供了有机分子式的预测、仪器测试结果的评判以及积极参与分子式的决定的主动能力,一改从前只能接受仪器测试结果的被动局面,从而加快了分子结构测定的速度。有机化合物分子式范围的界定和分子式的预测也为今后分子结构的界定和预测打下了初步基础。

关键词:分子式预测 三角形分布 应用

## 1 引言

有机化合物分子结构测定的首要事项是分子式的确定。在有机质谱出现之前,分子式的测定主要由微量元素分析来完成,但是由于各种主客观因素造成的误差把分子式弄错屡见不鲜,而且元素所给数据只是各元素在分子式中的简单比,真正的分子式可能是这个简单比各项的 1 倍到数倍。有机质谱出现后问题基本解决,但也常出现分析值与分子量之间的矛盾,常常是改变原子数目以符合分子量。高分辨质谱出现后,计算机可以根据测得的精确分子量拼出分子式。但是大量事实证明,高分辨质谱也常因仪器性能不佳给出误差较大的精确分子量,误差范围缩小时所给分子式不是真正的分子式,误差范围放大时又给出数个分子式供挑选。正确的分子式误差大,不正确的分子式误差反而小,使缺乏经验的人不知所措。最关键的问题是在以上情况下人们只能被动地接受测试结果,难以发挥判

1994 年 4 月 14 日收

断和决定分子式的能动作用,预测分子式则更不可能。

把问题的范围扩展一下,人们能否知道自然界可能存在的合理分子式?其形成和存在是否具有某种规律?其种类和数量是无限的、任意和不可知的,还是有一定范围、有限的、非任意的和可知的?根据我们多年的观察和研究认为,地球的自然界在某种稳定条件下是有限的,继言之,自然界的有机化合物及其分子式也必然是有限的。这个看法不仅是哲学推理,而且有一定事实作为根据,例如我们已经对小分子量的分子式范围提出了符合实际的界定方案,提出了界定计算式,编写了程序,人工书写或计算机打印出分子量 16~800、只含 C、H、O、N 的全部分子式 203470 个。举其中几例:分子量 16 时只有 1 个分子式,100 时只有 8 个分子式,200 时 27 个,300 时 72 个,400 时 147 个,500 时 269 个,600 时 442 个,700 时 677 个,800 时 978 个,900 时 1363 个以及 1000 时 1834 个等等。每 100 中的分子式数目见表 1。

表 1 Table 1 Range of molecular weight and number of molecular formulas

range of molecular weight	number of molecular fomulas	ratio between latter and former term
16~100	164	
101~200	1285	7.84
201~300	4302	3.35
301~400	10094	2.35
401~500	19676	1.95
501~600	33943	1.73
601~700	53807	1.59
701~800	80199	1.49
amount	203470	—

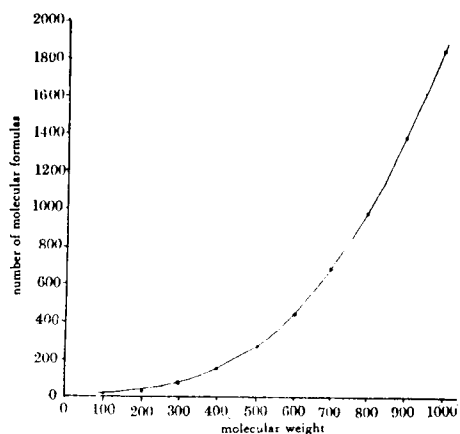


图 1 fig. 1 relation between number of molecular formulas and molecular weight

由上述和表 1 可见,随着分子量的增大,可能存在的合理分子式数目和种类也随之增加。分子量越大,分子式中各元素原子数目越多,排列组合也越多。若以分子量为横坐标、分子式数目为纵坐标制图,在坐标系中可得一园滑上升曲线(图 1)。

由表 1 第 3 栏还可看出,分子式数目与分子量的关系开始增加很快,以后趋于缓慢,接近于分子量每增加 100,分子式数目增加 0.5 倍。增加的绝对值则很大,分子量 16~800 有 203470 个分子式,而 801~1000 将再增加约 270000 个,总计约 474000 个。

## 2 分子式范围的界定

有机界主要由碳氢二元素构成,其他早已称之为杂原子的元素如氧、氮、硫、磷、氟、氯、溴、碘、镁、铜、硅、锗、铁、钴等都是配角。在天然有机化合物中,碳、氢、氧、氮是最重要的组成元素。氢与碳的搭配最高值即饱和值为  $C_nH_{2n+2}$ ,最低值则至今未定。但由多快

和稠环芳香体系观察,一般不低于 $\frac{1}{2}n$ ,这些高度不饱和化合物的分子量均较小,多在 200 以下。氧与碳搭配时在绝大多数情况下,最高值与碳原子数相等,普遍存在的、构成天然有机界的基本化合物的六碳单糖的分子式 $C_6H_{12}O_6$ 就是例证。所以氧的最高值定为 $O_n$ 较为合理,而最低值已知为零。但仍有个别的小分子如甲酸( $CH_2O_2$ )和草酸( $C_2H_2O_4$ )等,氧原子数为碳原子数的 2 倍,其分子量多在 100 以下,这些例外可如下述对待。氮是一个特殊元素,它的电子数和化合价为奇数,而原子量为偶数,如把此三项均为奇数的卤素、磷、氢等称为奇数元素和把三项均为偶数的碳、氧、硫等称为偶数元素,则只好把氮称为奇偶元素。正是由于氮的奇偶性,才使有机化合物的分子量产生了所谓氮规则,即分子量为偶数时,分子式必不含氮或含偶数个氮原子;分子量为奇数时,分子式必含氮且只含奇数氮原子(分子量超过 1000 时,各原子量尾数之和可能使之进位 1 或更多,运用氮规则不计尾数进位)。另外,氮原子在参与分子式构成时,绝大多数总要带有同原子数的氢原子,在计算分子式中的氢和不饱和度时须加注意,即必须先减去氮原子及同数目的氢原子,再如不含氮分子式一样处理。氮与碳的搭配有如氧的情形,不过这里是以氨基酸或肽为根据比较符合实际。构成肽的最小氨基酸甘氨酸的分子量为 75,含 1 个氮,或者说分子量 100 时氮原子一般不超过 2。例如全由甘氨酸构成的分子量 1044 的十八肽就有 18 个氮原子,所以,设定氮的最高值按分子量计,每增加 100 就增加 2 个氮原子,能够覆盖绝大多数天然含氮化合物的分子式,亦即分子量 100 以内有 2 个氮原子,101~200 有 4 个,401~500 有 10 个,901~1000 有 20 个,而最低值也是零。尽管最高值放宽很多,与氧的情形也一样,也有个别、特别是分子量 100 以下的小分子的含氮量超过了最高值,具体对待也见下述。总之,对于只含碳、氢、氧、氮的绝大多数天然有机分子来说,分子式范围的界定共有 6 项,其中 3 项即氢的最高值和氧、氮的最低值早已界定,另外 3 项即氢的最低值和氧、氮的最高值原未界定,本文根据构成有机界基本化合物的实际状况提出了初步界定方案。只有首先把分子式的范围界定,才有可能把绝大多数可能存在的天然有机分子式书写或打印出来,以便为结构推定打下基础。

设碳原子数为  $n$  和氮原子数为  $n'$ ,则天然有机化合物分子式范围的界定方案为:

1. 氢原子数:  $\frac{1}{2}n+n' \sim 2n+2n+n'$

2. 氧原子数:  $0 \sim n$

3. 氮原子数: 分子量 16~100,  $n'=0 \sim 2$ ; 101~200,  $n'=0 \sim 4$ ; 201~300,  $n'=6 \dots\dots$   
401~500,  $n'=0 \sim 10 \dots\dots$  901~1000,  $n'=0 \sim 20$

设  $M$  表示分子量,则界定通式为为:

$$C_n H_{n/2+n'} \sim 2n+2+n' O_0 \sim n N n' = M$$

### 3 分子式的分布

根据以上分子式范围界定通式书写或打印出的分子式,发现分子式的分布有一定规律性,即随着分子量的增大,分子式中的碳原子数也增加,其中最高碳原子数增加较快,最低碳原子数增加缓慢,两者之差越来越大,但均呈直线关系。若以分子量为横坐标和以分

分子式中的碳原子数为纵坐标,在坐标系中即出现了近于原点相交的两条射线,上线代表分子式中碳原子数的最高值,因而称为上限;下线代表分子式中碳原子数的最低值,因而称为下限。若截止到任一分子量,就构成一三角形(图2中的B),绝大多数有机分子式都分布在三角形B中。在A区内若有分子式出现,应是极个别更不饱和的碳氢或碳氢氮分子式,而个别氧化或氮化程度更高的分子式将出现在分子量很小的C区。分子式的这种分布现象,我们称为三角形分布。图2是根据只含碳、氢、氧的分子式绘制的,含氮分子式同样呈现这样的分布。

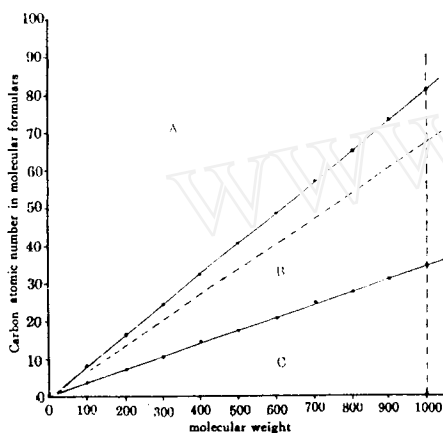


图2 fig. 2 Triangle distribution of molecular formulas

实际上三角形B还有一定厚度,在上下限处,每种分子量各只有1个分子式,是三角形的最薄处。若由上下限向三角形B的中部移去,每种碳原子数下的分子式种类将逐渐增加,在三角形中线和上限之间同碳数分子式种类最多,分子量越大种类越多。若把每种分子量下同碳数分子式最多的各点连结起来,就构成了三角形B内的虚线D,在虚线D附近分子式种类最多,为三角形的最厚处。因此,分子式的分布又呈立体状态,很似具有流线形鱼身的横切面,这个横切面也呈流线形。

分子式数目随分子量的增大而增加并非无限进行下去,亦即三角形B的上下射线并非无限直线延长。有迹象表明,氢的最低值逐渐大于 $\frac{1}{2}n$ ,而氧的最高值将逐渐变得小于 $n$ ,后者可由葡萄糖形成淀粉或由氨基酸形成肽的过程看出。在这些过程中,每增加1个葡萄糖或1个氨基酸都至少失去1分子水,或者说减少1个氧,而不是如现在的界定方案那样,碳氧数目始终保持相等。氢的增加和氧的减少,必将使上下射线都向三角形B的内部弯曲,最终在分子量数百万处有可能再次接近或相交,交点可能落在虚线D附近,那里的分子式可能只有1或几个。因此,全部分子式分布的正面图形也可能呈流线形,现在的三角形为流线形的头部,射线的再次交点为流线形的尾部。至于分子量1000以上的大分子分子式范围如何界定,只能由以后的研究解决。本文只讨论小分子分子式范围的界定,并把重点放在只含碳、氢、氧、氮4种常见元素构成的分子式方面,含其它元素的分子式推算只作简单说明,质谱中的碎片离子式也有涉及。

#### 4 分子式的计算

以下具体说明给定分子量分子式的计算。首先要计算出上下限处的分子式。把有关元素的原子量代入分子式范围界定通式中,再把改变后的算式变为计算上下限式中碳原子数 $n_{\text{上}}$ 和 $n_{\text{下}}$ 的方程式:

1. 上限分子式中的碳原子数:

因上限式中不含氧以及  $C_nH_{n/2} + n'Nn' = M$  和  $12n + \frac{1}{2}n + n' + 14n' = M$ ,

故上限分子式中的碳原子数  $n_{\pm} = \frac{M-15n'}{12.5}$  (遇小数弃去) (1)

2. 下限分子式中的碳原子数:

因下限式中含氧最多以及  $C_nH_{2n+2} + n'O_nNn' = M$  和  $12n + 2n + 2 + n' + 16n + 14n' = M$ ,

故下限分子式中的碳原子数  $n_{\mp} = \frac{M-2-15n'}{30}$  (遇小数大多进 1, 个别小分子进 2 或进 3) (2)

上下限式中的碳原子数  $n_{\pm}$  和  $n_{\mp}$  计算出来之后, 便可计算上下限分子式, 首先是由分子量减去碳或碳和氮的质量, 余下的质量在上限式中即为氢的原子数; 在下限式中则为氢氧共有, 可试减  $O_n$ , 再余者即为氢原子数, 若无余数或不足, 则改氧原子数小于  $n$  或改  $n_{\mp}$  进 1 为进 2 或进 3, 以求氢原子数符合界定值  $\frac{1}{2}n + n' \sim 2n + 2 + n'$ . 这是因为某些小分子的质量调剂范围有限, 每调剂 1 个原子, 质量变动相对很大, 加上所有下限式都是按  $O_n$  计算, 所以少数情况下会出现不合理的下限式。

上下限分子式产生后, 就可自上限式开始向下渐次减少碳原子, 增加氢原子和氧原子 (氮原子数已经固定), 以得到二式之间的其它各式。具体调剂步骤如下: (1) 首先减 1 个碳原子, 增加 12 个氢原子, 直至接近或达到饱和; (2) 然后减 16 个氢原子, 增加 1 个氧原子, 直至氢原子接近或达到  $\frac{1}{2}n$  或  $\frac{1}{2}n + n'$ ; (3) 重复前两项, 若不能重复, 则可同时减 1 碳 4 氢, 增加 1 个氧; (4) 若同时减碳减氢后又能重复 1、2 两项, 则继续重复, 否则可重复第 3 项。

依此调剂直到氧原子数接近碳原子数, 下一分子式应为下限, 否则必然是计算错误或遗漏, 遇此情形应重新审核或计算各式分子量, 务求与下限式吻合。现举例计算如下:

4.1 分子量 100 时有几个分子式?

4.1.1 不含氮分子式 ( $n' = 0$ ): 上限式中的碳原子数按式 (1) 计算  $n_{\pm} = \frac{100}{12.5} = 8$ ,  $100 - 12 \times 8 = 4$ , 上限式为  $C_8H_4$ 。下限式中的碳原子数按式 (2) 计算为  $\frac{100-2}{30} = 3.27$ ,  $n_{\mp} = 4$ ,  $100 - 12 \times 4 = 53$ , 可调剂为 3 个氧和 4 个氢, 下限式为  $C_4H_4O_3$ 。上下限式之间的其它分子式依次为  $C_7H_{16}$ ,  $C_6H_{12}O$  和  $C_5H_8O_2$ , 再向下恰好是下限式, 因而书写无误, 共 5 个不含氮式。

4.1.2 若分子中还含氮, 根据界定方案绝大多数分子量 100 的化合物只能含 2 氮 (含 4 氮的特殊情形见下述), 所以上限式的碳原子数计算为  $\frac{100-15 \times 2}{12.5} = 5.6$ ,  $n_{\pm} = 5$ ,  $100 - 12 \times 5 - 14 \times 2 = 12$ , 上限式为  $C_5H_{12}N_2$ 。下限式中的碳计算为  $\frac{100-15 \times 2-2}{30} = 2.27$ ,  $n_{\mp} = 3$ ,  $100 - 12 \times 3 - 14 \times 2 = 36$ , 可调剂为 2 氧 4 氢, 下限式为  $C_3H_4O_2N_2$ 。上下限式之间只有  $C_4H_8ON_2$ , 共 3 个含氮式, 连同不含氮者, 分子量 100 时共有 8 个分子式。

4.2 分子量 253 时有几个分子式?

这里分子量为奇数, 根据氮规则不可有不含氮分子式, 且只能有含 1、3 和 5 个氮 3 种情形, 分子式分别如下:

4.2.1 含1个氮:上限式中的碳计算为 $\frac{253-15}{12.5}=19.04$ , $n_{\text{F}}=19$ ,上限式为 $C_{19}H_{11}N$ 。下限式的碳计算为 $\frac{253-15-2}{30}=7.87$ , $n_{\text{F}}=8$ ,下限式为 $C_8H_{15}O_8N$ 。二式之间的其它式为 $C_{18}H_{23}N$ , $C_{17}H_{35}N$ , $C_{17}H_{19}ON$ , $C_{16}H_{31}ON$ , $C_{16}H_{15}O_2N$ , $C_{15}H_{27}O_2N$ , $C_{15}H_{11}O_3N$ , $C_{14}H_{23}O_3N$ , $C_{13}H_{19}O_4N$ , $C_{12}H_{15}O_5N$ , $C_{11}H_{11}O_6N$ , $C_{10}H_{23}O_6N$ , $C_{10}H_7O_7N$ ,和 $C_9H_{19}O_7N$ 。

4.2.2 含3个氮:上限式中的碳计算为 $\frac{253-15 \times 3}{12.5}=16.64$ , $n_{\text{F}}=16$ ,上限式为 $C_{16}H_{19}N_3$ 。下限式中的碳计算为 $\frac{253-15 \times 3-2}{30}=6.87$ , $n_{\text{F}}=7$ ,下限式为 $C_7H_{15}O_7N_3$ 。二式间的其它式为 $C_{15}H_{31}N_3$ , $C_{15}H_{15}ON_3$ , $C_{14}H_{27}ON_3$ , $C_{13}H_{11}O_4N_3$ , $C_{13}H_{23}O_2N_3$ , $C_{12}H_{19}O_3N_3$ , $C_{11}H_{15}O_4N_3$ , $C_{10}H_{11}O_5N_3$ , $C_9H_{23}O_5N_3$ 和 $C_8H_{19}O_6N_3$ 。

4.2.3 含5个氮:上限式中的碳计算为 $\frac{253-15 \times 5}{12.5}=14.24$ , $n_{\text{F}}=14$ ,上限式为 $C_{14}H_{15}N_5$ 。下限式中的碳计算为 $\frac{253-15 \times 5-2}{30}=5.87$ , $n_{\text{F}}=6$ ,下限式为 $C_6H_{15}O_6N_5$ 。二式间的其它式为 $C_{13}H_{27}N_5$ , $C_{12}H_{23}ON_5$ , $C_{11}H_{19}O_2N_5$ , $C_{10}H_{15}O_3N_5$ , $C_9H_{11}O_4N_5$ , $C_8H_{23}O_4N_5$ 和 $C_7H_{19}O_5N_5$ 。

以上3项总计为37个分子式。

4.3 分子量1000时有几个不含氮分子式?含20个氮原子时的上下限分子式如何?

按上法同样计算,不含氮的上限式为 $C_{80}H_{40}$ ,下限式为 $C_{34}H_{48}O_{34}$ ,二式间还有225个分子式,共227个。含20个氮的上限式为 $C_{56}H_{48}N_{20}$ ,下限式为 $C_{24}H_{48}O_{24}N_{20}$ 。

4.4 分子量456的三萜有几个可能的分子式?

如上法计算上限式为 $C_{36}H_{24}$ ,下限式为 $C_{16}H_8O_{16}$ 。此二式显然与三萜的情况不符,因此应从二式之间的其它分子式中挑选出合理者。其它式由上限式起依次为 $C_{35}H_{36}$ , $C_{35}H_{20}O$ , $C_{34}H_{48}$ , $C_{34}H_{32}O$ , $C_{33}H_{60}$ , $C_{33}H_{44}O$ , $C_{33}H_{28}O_2$ , $C_{32}H_{56}O$ , $C_{32}H_{40}O_2$ , $C_{32}H_{24}O_3$ , $C_{31}H_{52}O_2$ , $C_{31}H_{36}O_3$ , $C_{31}H_{20}O_4$ , $C_{30}H_{48}O_3$ , $C_{30}H_{32}O_4$ 和 $C_{30}H_{16}O_5$ 。到此为止,因为以下就是 $C_{29}$ 等少于30个碳的分子式了,也与绝大多数三萜不符。如何由这16个式中选出合理者,主要应根据化合物的理化性质和光谱特征,例如是烃、是醇、醛酮、羧酸、还是酯等。由质谱是否失水、一氧化碳或羧基以及由裂解方式和碎片离子等推测最高含氧量。由核磁决定碳氢原子数、双键、羰基及骨架外基团如甲氧基、乙酰氧基,当归酰氧基及糖等。最后还可以计算不饱和度,这是1个极其重要和有效的方法,由此可以排除很多不合理者,例如计算以上各式的不饱和度后,只有 $C_{31}H_{52}O_2$ 和 $C_{30}H_{48}O_3$ 二式符合绝大多数分子量456三萜的情况,前者的不饱和度为6,可能含1甲氧基、1羟基和1个双键,或含1甲氧基和1个羰基;后式的为7,可能含1羟基、1羧基和1个双键,或含1羰基和1羧基。此4化合物可能都是五环三萜,根据光谱数据不仅分子式很易决定,结构也近于明朗。

若使用计算机,可按上述分子式范围界定通式及氮规则等编制程序,以获得任意分子量的全部合理分子式,不仅快速且可防止错误和遗漏,再根据其它信息正确选择具体化合物的分子式。

## 5 应用实例

所谓应用就是如何从众多分子式中选出具体的化合物的分子式,现举例说明如下:

5.1 由番荔枝得 1 个分子量为 281 的含氮化合物<sup>[1]</sup>,质谱显示能连续失去 4 个 28 或 29 基团,据经验这些基团应是 CO、CHO 或 CH<sub>2</sub>NH。另外也显示化合物只含 1 个氮,因此失去的 4 个基团可能是:(1) 3 个含氧基团、1 个含氮基团;(2) 4 者都是含氧基团。该化合物的高分辨质谱很长时间不能测定,CH<sub>2</sub>NH 含否不能证实,这关系到最高含氧量。化合物开始的纯度又较差,核磁出现了不成比例的脂肪氢,因而认为质谱能失去 CH<sub>2</sub>NH,分子式中有 3 个氧。分子式的计算是:281-16×3-14-1=218,这是碳氢的质量,218÷12.5=17.44, n<sub>C</sub>=17,有碳氢式 C<sub>17</sub>H<sub>14</sub>和 C<sub>16</sub>H<sub>26</sub>,原式应为 C<sub>17</sub>H<sub>15</sub>O<sub>3</sub>N 或 C<sub>16</sub>H<sub>27</sub>O<sub>3</sub>N,不饱和度分别为 11 和 4,后式与质谱不符,因为质谱显示其为高度不饱和的阿朴啡类,再由红外和核磁分析提出了初步结构 a。纯度提高后脂肪氢消失,质谱就不可能失去 CH<sub>2</sub>NH,因而应含 4 个氧,重新计算分子式为 281-16×4-14-1=202,202÷12.5=16.16, n<sub>C</sub>=16,有碳氢式 C<sub>16</sub>H<sub>10</sub>和 C<sub>15</sub>H<sub>22</sub>,原式应为 C<sub>16</sub>H<sub>11</sub>O<sub>4</sub>N 或 C<sub>15</sub>H<sub>23</sub>O<sub>4</sub>N。不饱和度分别为 12 和 5,后式与质谱不符,理由同前。同样由红外和核磁很易提出了结构 b,且解释了为什么在质谱中没有出现由结构 a 应产生的强 M-1 离子 m/z280。显然结构 b 接近于真实,分子式应为 C<sub>16</sub>H<sub>11</sub>O<sub>4</sub>N。



5.2 由番荔枝得另一化合物分子量为 465<sup>[2]</sup>,红外和核磁显示分子中含 3 个甲氧基,3 个羟基和 1 个羰基,共有 7 个氧原子,质谱推断只有 1 个氮,这些信息已经足够。如上例一样只计算碳氢即可:465-16×7-14-1=338,338÷12.5=27.04, n<sub>C</sub>=27,碳氢式有 C<sub>27</sub>H<sub>14</sub>, C<sub>26</sub>H<sub>26</sub>, C<sub>25</sub>H<sub>38</sub>和 C<sub>24</sub>H<sub>50</sub>,不饱和度分别为 21,14,7 和 0,光谱分析第二式相符,因而分子式很易确定为 C<sub>26</sub>H<sub>27</sub>O<sub>7</sub>N,与高分辨质谱数据完全一致。

由上 2 例可知,用本文的方法推测具体化合物的分子式时,要尽可能由各种光谱收集碳、氢、氧、氮,特别是后二者的最高含量。计算不饱和度尤其重要和有效,它可把分子式的选择范围缩得很小,直至 1 个而无他选择。运用本法还可辨认和排除杂质对核磁氢谱和碳谱的干扰,或证明氢谱和碳谱数据。基各种光谱数据准确齐全,用本法推测分子式的同时,分子结构也常露端倪,两者同时进行比只顾解析光谱推导结构要周到迅速得多。

## 6 含其它杂原子的分子式

有机界尚有许多含其它杂原子的化合物,分子式的计算很简单,只须减去其它杂原子并依其化合价补足氢原子,然后按本文方法计算即可。补充氢原子是为了正确计算上下限式和不饱和度及防止分子式的遗漏。计算完成后再减去补进的氢加进杂原子。有时不加氢亦可,例如含硫者只先减去硫,这与含氧情形类似,也可另立通式。现举 1 例:1 个含一氯的化合物,分子量 192<sup>[3]</sup>。分子式计算为 192-35+1=158,158÷12.5=12.6, n<sub>C</sub>=12,

上限无氯式为  $C_{12}H_{14}$ 。  $192 - 35 + 1 - 2 = 156$ ,  $156 \div 30 = 5.2$ ,  $n_F = 6$ , 下限无氯式为  $C_6H_6O_5$ 。 两式间的其它式为  $C_{11}H_{10}O$ ,  $C_{10}H_{22}O$ ,  $C_{10}H_6O_2$ ,  $C_9H_{18}O_2$ ,  $C_8H_{14}O_3$  和  $C_7H_{10}O_4$ 。 由于质谱中只能失去两分子  $CO$ 。 再无其它含氧基团的丢失, 所以只能选择  $C_{10}H_6O_2$  和  $C_9H_{18}O_2$  两式, 不饱和度计算前式为 8, 后式为 1, 前式与核磁相符。 恢复成原式应为  $C_{10}H_5O_2Cl$ , 其结构也很快确定为 2-氯代萘。

## 7 质谱中的离子式

质谱中碎片离子式也可用本文方法判断, 但首先要弄清两点: (1) 分子中是否含氮, 以判断碎片离子是否也含氮和含几个氮; (2) 碎片离子在结构上是小分子还是自由基, 后者在计算时须加 1 氢以补成小分子。

举前述分子量 465 化合物<sup>[2]</sup>为例, 质谱中的中强离子  $m/z$  判断其为含有分子中仅有的 1 个氮原子和结构上为自由基的离子, 且含有芳香环。 计算方法为  $152 - 15 + 1 = 138$ ,  $138 \div 12.5 = 11.04$ ,  $n_F = 11$ , 上限无氮式为  $C_{11}H_6$ 。 下限无氮式计算为  $n_F = 5$ , 须改为 7 (进 3) 得下限无氮式  $C_7H_6O_3$ 。 二式间的其它式为  $C_{10}H_{18}$ ,  $C_9H_{14}O$  和  $C_8H_{10}O_2$ 。 五式的不饱和度分别为 9, 2, 3, 4 和 5, 第四式与上述数据相符, 恢复成原式只加氮即为结构上是自由基的离子式  $C_8H_{10}O_2N$ , 高分辨质谱证实了此式。

## 8 特殊例外的分子式

特殊例外的分子式是指氢的最低值和氧、氮的最高值超出本文规定的界定值、分子量多在 100 以下的少数小分子的分子式。 解决这个问题并不困难, 只要放宽界定范围即可计算。

8.1 氢原子数小于  $\frac{1}{2}n$  例如  $\frac{1}{4}n$ : 上限式计算可改为  $n_F = \frac{M - 15n'}{12.25}$ , 下限式计算则仍如前。 例如分子量 100 的不含氮上限式仍为  $C_8H_4$ , 但分子量 1000 的不含氮上限式则由  $C_{80}H_{40}$  变为  $C_{81}H_{28}$ 。

8.2 氧原子数大于  $n$  例如  $2n$ : 上限计算式不变, 下限式计算则要改为  $n_F = \frac{M - 15n' - 2}{46}$ 。

例如草酸的分子式计算:  $\frac{90 - 2}{46} = 1.91$ ,  $n_F = 2$ , 下限式即草酸的分子式为  $C_2H_2O_4$ 。 同样计算也得到甲酸的分子式  $CH_2O_2$ 。

8.3 氮原子数在分子量 100 以下超过 2 个: 例如分子量 100 时有 4 个氮, 仍可用本文方法计算得上限式  $C_3H_8N_4$ , 下限式为  $C_2H_4ON_4$ 。 氮原子如此之多大概要自行连接而没有多带氢原子进入分子式, 因此, 若用式(1)和式(2)计算得不到合理的上下限式, 也可另立通式  $n_F = \frac{M - 14n'}{12.5}$  和  $n_F = \frac{M - 14n' - 2}{30}$  更为适宜。 例如氰化氢(HCN)、胍( $CH_5N_3$ )和三氮唑( $C_2H_3N_3$ )的分子式的均由上限式计算而得(胍用式(1)计算亦可), 嘌呤的分子式( $C_5H_4N_4$ )也须如此计算。

## 9 结语和讨论

9.1 有机化合物的分子式范围可以界定。 由任何给定的分子量能够计算出所有可能存在



的合理分子式。这项研究不仅在认识论上,而且在实践中的均具有重要意义。本项研究为基础有机化学增添了新的内容,也为有机化学家提供了主观判断能力,一改从前在确定有机分子式时只能接受仪器测试结果的被动局面。

9.2 由于地球自然界以及有机界的有限性,有机化合物的分子式也必然是有限的。有机界的有限性表征之一就是本文提出的有机物分子式的三角形分布(由局部看)和可能存在的流线形分布(由整体看)。

9.3 人的主观判断不能代替客观的测试,判断仍须测试来证实,而测试也须人来评判。只有主客观统一了,结论才能作出。

9.4 既然有机分子式可以预言,有机化合物的分子结构也应当能够预言,前者为后者打下了基础,后者将更是一项巨大的工程。不过在计算机快速发展的今后年代,实现这个理想极有可能。若果然如此,必将大大加快人类向自然索取的进程。

## 10 致射

本文得到北京大学邢其毅先生的热情指导及何美玉先生的热情关怀,上海市在美国普度大学就读的诸钢同志帮助编制了计算机程序,作者在此表示衷心谢意。

## 参 考 文 献

- 1 廖永红等. 待发表
- 2 杨小江等. 番荔枝化学成分研究. 药学学报, 1992; 27(3): 185-190
- 3 张锦哲教授一筛选样品的储存分解物

# Study and Application on the Compositive Pattern and the Distribution of Molecular Formulas of Organic Compounds

Cong Puzhu

(Institute of Medicinal Plant Development,  
Chinese Academy of Medical Sciences, Beijing 100094, China)

Received 1994-04-14

## Abstract

Limitation and knowability on possible and rational molecular formulas of natural organic compounds are preliminarily discovered in kinds and quantities based on our researches for years. The delimitate schemes and calculating formulas corresponding to ob-

jective realities for the range of molecular formulas below 1000 of molecular weight are tentatively proposed. 203470 molecular formulas which contain only C,H,O,N four elements and molecular weight is from 16 to 800 have been written manually or printed by computer and about 474000 molecular formulas for molecular weight up to 1000 are evaluated. A triangle area which the most of molecular formulas is distributed in it comes out in a coordinate system if the abscissa stands for molecular weight and the ordinate for carbon atomic number in molecular formulas. The triangle, as a head, however, will be a part of whole molecular formula distribution, perhaps as a streamline. Organic chemists will have the initiative in evaluation of experiment data from instruments, calculation of molecular formula according to the delimitate formulas and final determination of them, change the former passive position just accepting experiment data and raise the efficiency of organic molecular structure elucidation if the technique is adopted. The primary foundation for future delimitate and calculation of organic molecular structure will also be laid on the range delimitate and calculation of molecular formulas.

**Key Words:** molecular formula calculation, triangle distribution, application