

# 信息论与药学科学

何 绍 雄

(天津药物研究所)

当代广义信息论向药学科学的渗透, 已为药学科学提供了新的研究方法, 推动药学科学的研究向纵深发展。本文拟就信息论在药学科学中某些主要应用进展进行概括的讨论, 为二者进一步结合提供线索。

## 生物活性的分子结构特征

任一化合物在其原子核周围空间分布的电子, 可以视为包含着信息的电子概率场形式, 而作为化合物的药物分子与受体相互作用时, 可看作是所携带的信息向接受者(受体)馈送, 当某些信息被受体译出时, 便导致生物反应的开始。据此, Kier<sup>[1]</sup> 提出应用 Shannon 方程可求得作为分子信息量的度量: 分子负熵, 以此来译制支配生物活性的结构信息。分子负熵的计算法, 以丙烷为例, 其每个原子的负熵  $i$  据 Shannon 方程应为

$$i = -K \sum_j P_j \log P_j \quad (1)$$

(1) 式中  $K$  是由对数的底所决定的常数,  $i$  指某种组合。丙烷分子共有 11 个原子, 其分子分解为拓扑等价原子, 共有四种集合, 其原子相重数分别为 6, 2, 2 和 1, 则其随机选择机率分别为  $6/11$ ,  $2/11$ ,  $2/11$  和  $1/11$ , 代入 (1) 式:

$$i = -\frac{6}{11} \log \frac{6}{11} - \frac{2}{11} \log \frac{2}{11} - \frac{2}{11} \log \frac{2}{11} - \frac{1}{11} \log \frac{1}{11} = 0.507 \quad (2)$$

因丙烷分子有 11 个原子, 故其分子负熵  $I$  为:

$$I = i \times 11 = 5.582 \quad (3)$$

作者将此种分子信息量作为参数研究了一些化合物的构效关系, 如不同醇类对酯酶、脂肪氧化酶的抑制作用、32 个非特异性麻醉剂的活

性研究等, 发现具有良好的相关性。作者由此认为, 在生物活性是非特异性的情况下, 分子信息量作为一种结构指数可将分子中影响生物活性的突出结构的特征译制出来, 并按照分子的非特异生物效价的大小, 正确地依对称性将分子排成顺序, 并可从分子水平上对作用机制进行解释。

由于蛋白质可看作为包含着各种氨基酸符号的消息, 因此可用类似上述的方法计算出某种蛋白质中每一个氨基酸的平均负熵  $I_a$ , 然后再计算其最大负熵  $I_{\max}$ , 并求其比值  $I_a/I_{\max}$ 。Branson 计算了胰岛素、胃蛋白酶、丝蛋白等 26 种蛋白质的  $I_a/I_{\max}$ , 发现每一氨基酸的负熵似与分子大小无关, 在大多数活体组织中  $I_a/I_{\max} \geq 0.7$ 。作者估计只有氨基酸数目在 400—500 个以上所组成的蛋白质, 且其  $I_a/I_{\max} > 0.5$  时, 才有生理功能;  $I_a/I_{\max}$  越大, 其生理功能越显著。

## 药物设计与模式识别

模式识别 (Pattern Recognition) 是信息处理发展到智能阶段的产物。六十年代末, 模式识别法已应用到药物研究之中, 七十年代, 该法被引入药物设计研究。由于研究药物活性和化学结构的定量规律发展十分迅速, 定量构效关系的数学模型日益增多, 多种结构信息参数不断出现, 如何从中提出描述药物生物活性的结构特征, 如何从众多的结构信息中提取关键性的代码, 以及如何区分结构类同或不同的药物的药理活性, 都是新药设计中重要的课题。模式识别法正是一种可能客观地、方便地解决这个问题方法。

模式识别法包括三个相互联系, 而又互有

区别的过程:即数据生成、模式分析和模式分类。在药物设计中,药物原始结构信息常表现为数字形式,可直接用矢量表示,以便于计算机处理,故数据生成已基本解决。模式分析是对模式(药物)的数据进行加工,特征即为表征一个药物(模式)的测度。在药物设计中,通常有如下的分子特征可供选择:如理化参数描述符,分子片断描述符,位置描述符,环境描述符,几何描述符,对症描述符,分子连接性指数代码,质谱质荷比代码等。由这些特征所构成的特征空间,其维数即为所选取的特征数。故任一模式均可由特征空间的矢量表示:

$$X_i = (X_{i1}, X_{i2}, \dots, X_{in})^T, \\ i = 1, 2, \dots, m \quad (4)$$

(4)式中 $n$ 为特征空间的维数, $m$ 为模式数。通常,需要使每个特征的均值为0,标准差为1,即标准化;对存在已知类别的训练集系统而言,有时还需对标准化了的数据依其区分类别的相对作用的不同而进行加权处理。由于药物分子特征常为高维空间,故需用非迭代型和迭代型(或两者兼用)技术将其压缩(映照)为二维或三维空间,以便能进行直观分析。模式分类是根据模式分析所得的信息,决定一个“学习”或“训练”的过程,从而制定出判断标准,对待测药物(模式)进行分类。一般采用判别分类,类聚群分析等多元统计方法。

近些年来,用模式识别法对药物进行分类设计的研究论文时有报道。Cammarata曾用模式识别法对13个降压药,43个抗组胺药,抗胆碱能药,抗精神病药等进行分类,得到合理的解释;该作者又对39个 $\alpha$ 激动剂、 $\beta$ 激动剂以及中枢神经系统药物进行分类,并用因子分析法作为模式分析中高维数据的压缩工具获得良好的分类结果<sup>[2]</sup>。Chu等对138个化合物的抗癌活性进行预测,结果表明106个无活性,32个有活性;用此法对另外24个化合物进行考察,也获得满意结果<sup>[3]</sup>。

## 药物分析过程的最优化

分析化学的目的可认为是获取物质的信

息,而分析过程则是获取物质化学组成信息的过程,因之,不难理解,信息论对以分析化学为基础学科的药物分析具有重要意义。

将信息量的概念应用于药物分析过程的最优化已取得相当进展。所谓分析过程的最优化包括分析方法的最优选择及分析条件的最优选择。前者是从为满足某一分析目的的可用的分析方法中选择最合适的方法,后者是选择最合适的分析条件而使所用分析方法得到的结果最适合于分析目的。由于分析过程所获得的样品信息的大小与分析方法和条件有关,因此信息论的方法可以成为实现最优化的工具。

七十年代初,Massart采用Shannon方程得到的信息量作为评价薄层层析(TLC)分离效能的定量指标,其后,作者又借用数值分类法,以欧氏距离为分类依据,并用信息作为选取标准,实现了TLC不同展开剂系统的最优组合<sup>[4]</sup>,现已将这种数值分类-信息选取法成功地用于22种磺胺类药物的56种TLC展开溶剂系统的最优化选择上<sup>[5]</sup>。我国学者安登魁等提出了系统差值信息量分类法,他们指出,在TLC中,所期望的是所选取的最优组合的各系统分离的化合物彼此不同(即互相补偿),同时对于各自都能分离的化合物其 $\Delta RF$ 值差距越大越好。显然,若两系统分离特性近似,则对应 $RF$ 值差值的几率分布较集中,由Shannon方程计算所得的信息量较小;如若两系统分离特性不同,则对应 $RF$ 值差值的几率分布较分散,计算所得的信息量较大。这样,就可以系统 $RF$ 值差值的信息量作为系统相似性的度量;如将 $RF$ 值差值信息量大的某些系统组合起来,就能得到最优组合。作者已将此法用于26种食用染料对10种TLC展开溶剂和22种磺胺药物对56种TLC展开溶剂的最优选择上,后者所得结果与数值分类-信息选取法基本一致<sup>[6]</sup>。

Dupuis等<sup>[7]</sup>则根据气相层析保留系数提供的信息实现了最优层析柱或最优柱组合的选择。

药物分析采用模式分析的方法可以鉴别或

预测样品某些不能通过简单的测量所直接观察到的性质,如对化合物分子结构的解析等。例如模式识别中的学习机法使得低分辨质谱可用来确定分子式的化学问题<sup>[8]</sup>,经检验该法确定分子式的可靠程度达98%以上。Ziemer等<sup>[9]</sup>用“酶水解-气相色谱-质谱”法以鉴别多肽中氨基酸的序列,并运用了计算机化的模式识别法,从而形成一种有效的快速联机分析法。对86个五氟丙酰二肽甲酯类的分类结果表明,N端氨基酸检验准确率为100%,C端为97%。还有人将模式识别应用于单官能团的气液色谱中,通过预处理得到保留系数的修正参数和骨架数,从而获得结构信息。

某些诊断疾病的化验方法,实际上也是一种分析方法,因此,同样可以用信息论的方法来对之评价。模式识别法现已用于临床常规化验方法的选择,此外,计算信息量的大小还可用于判断某种化验结果与疾病之间关系密切的程度,由此以提高化验诊断的准确率。

### 药品质量控制<sup>[10]</sup>

药品的质量标准是根据所含组分允许的极限含量而给出的该组分含量的先验估计值来作为判断药品是否合格的标准。此时,药物分析的目的在于确定所测组分含量是否在规定的极限之内。

通常允许极限是根据经验按下式确定的, $X_L$ 和 $X_U$ 分别代表允许区间的低限和高限:

$$X_L = \bar{x} - K\sigma F \quad (5)$$

$$X_U = \bar{x} + K\sigma F \quad (6)$$

式中, $\bar{x}$ 是药品组分含量的要求值; $\sigma F$ 表示含量变化幅度的标准差; $K$ 为允许系数,决定允许区间的宽度。这样,占有产品100r%( $r < 1$ )的组分含量 $\mu$ 以 $\alpha$ 的几率在区间 $(X_L, X_U)$ 之内,区间外的几率为 $(1 - \alpha)$ ,在实际工作中,一般 $\alpha$ 值选在 $0.90 \leq \alpha \leq 0.99$ 。

如分析所得组分含量为 $X \in \langle X_L, X_U \rangle$ ,则其几率分布为:

$$P_0(X) = \begin{cases} \frac{\alpha}{X_U - X_L} & X \in \langle X_L, X_U \rangle \\ \frac{1 - \alpha}{100 - (X_U - X_L)} & X \notin \langle X_L, X_U \rangle, 0 \leq X \leq 100 \end{cases} \quad (7)$$

若测定组分的含量在允许极限之内(即药品含量合格),信息量为:

$$I(P, P_0)_\alpha = \ln \frac{X_U - X_L}{\alpha \sigma \sqrt{2\pi e}} \quad \mu \in \langle X_L, X_U \rangle \quad (8)$$

若含量超出允许极限,药品不合格时,信息量为:

$$I(P, P_0)_{1-\alpha} = \ln \frac{100 - (X_U - X_L)}{(1 - \alpha) \sigma \sqrt{2\pi e}} \quad \mu \notin \langle X_L, X_U \rangle \quad (9)$$

由于 $0.90 \leq \alpha \leq 0.99$ ,且 $(X_U - X_L) < 100$ ,所以 $I(P, P_0)_{1-\alpha}$ 总比 $I(P, P_0)_\alpha$ 为大。显然,在实际工作中,发现不合格的药品远比合格药品为重要。

生命过程依靠调节和控制来维持其统一性和相对稳定性,而信息的传递正是实现调节和控制的必要条件,因此,信息论的概念和方法对生命科学的研究具有重要意义。药物是人们用以调节控制病理过程,使之恢复正常的重要手段之一,因此也不难想见,信息论在药学科学中同样具有重要的作用。从以上远非全面的介绍可见,信息论已经在药物研究的许多领域中发挥了重要作用;可以设想,它还将在更多的领域中发挥作用,例如不仅考虑血药浓度监测的信息而且同时考虑时间药理学的信息的给药方案最优化的拟订;不仅提取结构-活性特征而且同时提取结构-选择性分布特征的对所设计的药物的判别分类等,都是值得探讨的领域。可以预见,随着药学科学定量化过程的加速,信息论与药学科学之间的进一步渗透结合必将迅速发展,而药学科学终将成为广泛应用信息论的又一广阔领域。

### 参 考 文 献

- [1] Kier, L. B.: *J. Pharm. Sci.*, **69**, 807, 1980.
- [2] Cammarate, A.: *ibid.*, **66**, 304, 1977.

- [3] Chu, K. C. et al.: *J. Med. Chem.*, 18, 593, 1975.  
 [4] Massart, D. L. et al.: *Anal. Chem.*, 46, 1988, 1974.  
 [5] De Clercq, H. et al.: *ibid.*, 66, 1269, 1977.  
 [6] 安登魁等:《南京药学院学报》, (1), 1;(2), 14, 1982.

- [7] Dupuis, F. et al.: *Anal. Chem.*, 47, 379, 1975.  
 [8] Jurs, P. C. et al.: *ibid.*, 41, 21, 1969.  
 [9] Zierner, J. N. et al.: *ibid.*, 51, 1732, 1976.  
 [10] Eckschlager, K. et al.: *Information Theory as Applied to Chemical Analysis*, 1979.

[本文于 1984 年 8 月 25 日收到]

## 人血浆 $\alpha_2$ -巨球蛋白的研究概况

彭启明 张淑琴\*

(中国医科大学, 沈阳)

$\alpha_2$ -巨球蛋白 ( $\alpha_2$ -Macroglobulin, 简称  $\alpha_2$ -M) 是人体中重要的血浆蛋白之一, 能与多种蛋白水解酶反应, 是血浆中含量最多的蛋白酶抑制剂。 $\alpha_2$ -M 通过对蛋白酶水解活性的调节而参与凝血、纤溶、激肽释放、炎症及免疫等一系列生理及病理过程。因此,  $\alpha_2$ -M 的研究对某些疾病发病机制的了解具有重要的意义。

1953 年, Jacobsson 首先报道了  $\alpha_2$ -M<sup>[1]</sup>。此后对其组成、结构、理化性质及生物活性等方面的研究都有较详细的报道和论述<sup>[2-5]</sup>。国内这方面的报道尚未见到。

### 一、 $\alpha_2$ -M 的组成、结构及理化性质

$\alpha_2$ -M 是一种高分子量的糖蛋白, 含糖量约为 8—11%, 已经证实糖单位中包含有半乳糖、甘露糖、N-乙酰葡萄糖胺、N-乙酰神经氨酸、岩藻糖等。 $\alpha_2$ -M 的多糖部分都连接在对碱稳定的天冬酰胺的酰胺键上<sup>[2]</sup>。有关  $\alpha_2$ -M 的氨基酸组成已有报道<sup>[2,3,17]</sup>。人血浆  $\alpha_2$ -M 中每克分子还含有 4—8 克原子锌, 结合在  $\alpha_2$ -M 上的锌约占血浆总锌的 20%。这样多的锌对  $\alpha_2$ -M 的生理功能起什么作用还不清楚。有人认为可能没有重要作用<sup>[4]</sup>。用超速离心法测出人  $\alpha_2$ -M 的分子量为 725,000 左右<sup>[3]</sup>。等电点为 5.0—5.5。其等电点受唾液酸含量的影响; 唾液酸含量低则等电点偏高<sup>[6]</sup>。从电泳中发现人血浆中  $\alpha_2$ -M 以两种形式存在: 泳动慢的区带, 称为 S- $\alpha_2$ M;

泳动快的区带, 称为 F- $\alpha_2$ M。实验证明, S- $\alpha_2$ M 是存在于血液中未与蛋白酶结合的天然  $\alpha_2$ -M, 是  $\alpha_2$ M 的活性形式。而 F- $\alpha_2$ M 则是天然  $\alpha_2$ -M 与蛋白酶或胺类结合后形成的复合物, 不再具有反应活性, 能从血循环中被迅速清除<sup>[7]</sup>。 $\alpha_2$ -M 分子是含有四个相同亚基的四聚体, 其中每两个亚基通过二硫键连成二聚体, 两个二聚体又通过非共价键结合成  $\alpha_2$ -M 分子<sup>[6]</sup>。

$\alpha_2$ -M 性质比较稳定。实验证明, 在相同条件下比  $\alpha_1$  蛋白酶抑制剂稳定得多。 $\alpha_2$ -M 对酸和热敏感, pH4 以下易失活, 在 pH 5.0—8.4 范围内比较稳定。一般贮存在 4℃; 冻干后在 -70℃ 可以存放一年。反复冻融则易失活<sup>[3,8,9]</sup>。

### 二、 $\alpha_2$ -M 与蛋白酶的反应特性和机制

Barrett 等首先系统地阐明了  $\alpha_2$ -M 与蛋白酶反应的独特性质。指出  $\alpha_2$ -M 仅能与蛋白水解酶中的内肽酶反应(其中包括丝氨酸类、巯基类、羧基类、金属蛋白酶类全部四类内肽酶)。对外肽酶则毫无作用。而且它只能与有活性的内肽酶反应, 与灭活的酶或酶原都不能反应。这种反应是不可逆的, 即  $\alpha_2$ -M 与蛋白酶一旦结合就不能再产生游离的酶和天然的  $\alpha_2$ -M。虽然, 用 3M 的 NaSCN 处理后能使被  $\alpha_2$ -M 结合的酶解离出来, 并恢复 5—10% 的活力, 但

\* 沈阳医学专科学校