

配基和受体相互作用的数学描述及其参数的计算机辨识

钱国正、杨义力 (第二军医大学, 上海 200433)

摘要 为使受体研究中参数估计更加客观和准确, 本文提出以 Feldman 方程的一种显式作为配基-受体反应系统的数学模型, 并给出用一族曲线逐个拟合估计参数的数学方法及相应算法。利用这算法就可确定一个未知的配基-受体反应系统的定性特征及有关的所有系统参数。

关键词 受体; 数学模型; 配位体; 计算机辨识; 曲线拟合

配基和受体的反应是双分子反应, 即

1986年2月13日收稿 1986年7月14日修回

$$P + Q = PQ,$$

当这反应达到平衡时, 平衡常数 K 可表示为:

$$K = [PQ]/[P][Q] \quad [1]$$

式中 $[P]$ 、 $[Q]$ 和 $[PQ]$ 分别为平衡时游离态 P 和 Q 的浓度及 PQ 的浓度。如系统中有多种配基和多种受体, 相互作用就很复杂。一些学者⁽¹⁾对这些系统的数学描述和参数求法进行了研究。在这些研究的基础上, 我们提出更加适用计算机使用的数学描述和参数辨识方法。

数学模型 对于多配基和多受体反应系

统, Feldman 得出下列数学模型⁽⁴⁾:

$$p_i = [P_i] + \sum_{j=1}^m \frac{K_{ij}[P_i]q_j}{1 + \sum_{s=1}^n K_{sj}[P_s]} \quad i = 1, 2, \dots, n \quad [2]$$

这是一族非线性方程, 只要给定各配基的总浓度 p_i , 各受体总浓度 q_j 及平衡常数矩阵 K_{ij} , 就可解出各配基的游离态浓度, 进而求出 $[Q_i]$ 以及 $[P_i Q_i]$ 等参数。式中 n 是配基种数, m 是受体种数。

在实际中常常需通过实验数据, 辨识出系统参数 K_{ij} 和 q_j , 若已知是一种配基和一种受体的作用, 则可以用经典的 Scatchard 作图法估计 q 和 K 。此外, 也提出过一些近似作图和计算方法⁽⁵⁾, 但都不理想, 并只能用到一种配基和二种受体的情况。八十年代初, Munson 等提出直接用方程[2]来估计参数的方法, 据简介, 算法复杂, 只能在小型计算机上运行⁽¹⁾。

我们不采用式[2], 而进行下列变换, 令

$$R_i = \frac{p_i - [P_i]}{[P_i]} \quad i = 1, 2, \dots, n \quad [3]$$

将式[3]代入式[2]得本方法依据的数学模型:

$$R_i = \sum_{j=1}^m \frac{K_{ij} q_j}{1 + \sum_{s=1}^n K_{sj} [P_s]} \quad i = 1, 2, \dots, n \quad [4]$$

这是一个显性方程, 下面将看到它对参数估计的多种递归算法特别合适。 $[P_i]$ 和 R_i 在实验中均可直接得到的。

非线性曲线拟合估计参数的数学方法 假设 m 、 n 已知, 式[4]就是一个确定的显性代数表达式, 能直接用最小二乘法进行曲线拟合。由于初值选取不当或其它因素, 很难保证在高斯-牛顿方法中收敛, 故采用修正高斯解法的 Marquardt 方法⁽⁶⁾, 即在其系数矩阵对角线元素上都加因子 $d(d \geq 0)$ 。当 $d = 0$ 时, 则退化为高斯解法。我们还使 d 能根据计算情况智能性改变取值, 这样即避免了不收敛又加速迭代。

一般经典的迭代终止条件为修正值足够小, 我们则采取曲线拟合数据的残差平方和足够小这个更加严格的终止条件, 使终止时的参数值更接近最佳, 具体终止条件为

$$\left| \frac{Q^{(n+1)} - Q^{(n)}}{Q^{(n)}} \right| \leq \epsilon$$

式中 $Q^{(n)}$ 表示第 n 次迭代的残差平方和, $Q^{(n+1)}$ 表示第 $(n+1)$ 次迭代的残差平方和, ϵ 为预先指定的允许误差。

一族曲线逐个拟合数据的数学方法 数学模型式[4]是一族曲线, 对未知的配基-受体反应系统, 可依次取 n 、 m , 采用上述方法逐个地去拟合实验数据, 再按一定准则挑选一个拟合优度最好的表达式作该系统的数学模型。从数学上可以证明式[4]能确保逐个拟合过程能递归进行。因为每次非线性曲线拟合中最主要的运算是求函数对参数的偏导数值、函数值和每次迭代的残差平方和。以 $n = 1$ 为例, 这时式[4]具有下列形式:

$$R = \sum_{j=1}^m \frac{K_j q_j}{1 + K_j [P]} \quad [5]$$

当 m 增加时, 从式中可清楚看出变量组 K_m 、 q_m 是分离增加的, 求偏导数也可分离完成, 所以很易递归, 偏导数计算公式为

$$\frac{\partial R}{\partial K_j} = \frac{q_j}{(1 + K_j [P])^2} \quad j = 1, 2, \dots, m \quad [6]$$

$$\frac{\partial R}{\partial q_j} = \frac{K_j}{1 + K_j [P]} \quad j = 1, 2, \dots, m \quad [7]$$

因式[5]为显性连加形式, 随 m 增加求函数值和残差平方和只是增加累加项, 可递归完成。

增加系统模型的参数是为改善拟合优度, 为不使参数数盲目增加, 必须有一个判断曲线拟合优度是否满意的准则。这可用“额外平方和”原则, 建立一种假设检验方法来确定增加参数所引起的拟合优度改善是否比原来参数显著。设原参数时拟合残差平方和为 SS_1 , 增加一级参数后拟合残差平方和为 SS_2 , 因残差平方和是符合 χ^2 分布的, 显然可构造 F 统计量如下:

$$F = \frac{(SS_1 - SS_2) / (df_1 - df_2)}{SS_2 / df_2} \quad [8]$$

式中 df_1 、 df_2 分别为 SS_1 、 SS_2 符合 χ^2 分布时的自由度(等于数据对数减去参数个数)。对于给定的显著性水平 α , 可计算临界值 $F_\alpha(df_1 - df_2, df_2)$, 如果 $F > F_\alpha$, 则这次拟合的拟合优度改善是显著的, 承认这次拟合有效, 再使 m 增 1, 进行下一次曲线拟合。重复这过程直至满足 $F \leq F_\alpha$, 因为这条件表明曲线拟合优度的改善并不显著, 放弃这次尝试。以上一次曲线为最终结果。

初值选取也必须自动递归完成, 我们是这样取的: $m = 1$ 时用线性回归求 K, q , 并不需初值。 $m = 2$ 时从 Scatchard 图上可看出用一条直线($m = 1$)去拟合一支双曲线($m = 2$), 其 X 轴截距 q 约接近双曲线的一条渐近线的 X 轴截距 q_2 , 而其负斜率 K 接近双曲线的另一条渐近线的负斜率 K_1 , 另二个参数 q_1, K_2 分别小于 q, K , 选初值为 $K_1^{(0)} = K$; $q_1^{(0)} = q/c$; $K_2^{(0)} = K/c$; $q_2^{(0)} = q(c > 1)$ 。类推第 m 次拟合时前 $m - 2$ 组参数初值取参数本身值, 用上述同样原则确定 $K_{m-1}^{(0)}, q_{m-1}^{(0)}, K_m^{(0)}, q_m^{(0)}$ 。因初值选取偏差引起的负参数值和残差不收敛均用 Marquardt 参数 d 的自动取值纠正。

计算机算法 综上所论, 下面给出 $n = 1$ 时用一族曲线逐个拟合实验数据判别反应系统特征并辨识其最佳参数的算法($n > 1$ 时算法也是类似的)。

1. 确定样本数 L 和显著性水平 α 。
2. $m = 1$ 用一元线性回归得 K, q , 置 $m = 2$ 。
3. 计算初始值或人工干预输入初始值。
4. 依式[5]、[6]、[7]计算函数值、偏导数值, 建立修正值的正规方程并使之标准化。
5. 选取 Marquardt 参量 d , 解正规方程求得修正值。
6. 原值加修正值为估计值, 判它们是否为非负? 否则修改选取 d 的参数 β 并转到 5。
7. 用估计值求拟合残差平方和 Q_2 , 判是

否收敛? 否则修改 β 并转到 5。

8. 判本次迭代是否满足 $(Q_1 - Q_2)/Q_1 \leq \epsilon$? 否则以此估计值为原值转到 4。

9. 将 Q_2 作为 SS_2 , 按式[8]计算 F 值。

10. 在给定 α 下计算临界值 $F_\alpha(df_1 - df_2, df_2)$

11. 判 $F \leq F_\alpha$? 否则 m 增加 1, 转到 3。

12. 以第 $m - 1$ 次为最终结果, 输出最终估计的参数值 $K_1, K_2, \dots, K_{m-1}, q_1, q_2, \dots, q_{m-1}$ 及拟合优度指标(即拟合残差平方和)等。

结果比较 上述算法已在微型计算机 TRS-80 上用 BASIC 语言实现, 并在我校的配基-受体研究中实际应用。在此我们将采取某些论文中的实例加以比较。下文中以 $10^{-9} L/mol$ 、 $10^9 mol/L$ 为 K, q 的单位, 以 S 表示参数值代入我们模型拟合数据的残差平方和, 以%计。罗荣生⁽⁶⁾以孕酮受体互相作用实验为例。因原数据较少分布又不均匀, 我们在其拟合的曲线上多取一点来弥补这一缺陷, 显然这对作图法和以图为基础的计算毫无影响。罗荣生通过作图计算得 $K_1 = 1.4$, $K_2 = 0.03$, $q_1 = 0.145$, $q_2 = 2.20$ 而 $S = 0.0148$ (原数据时为 0.0135)。采用近似解法得 $K_1 = 1.0$, $K_2 = 0.033$, $q_1 = 0.203$, $q_2 = 2.20$ 而 $S = 0.188$ (原数据时为 0.181)。我们的结果为 $K_1 = 1.52$, $K_2 = 0.0300$, $q_1 = 0.132$, $q_2 = 2.38$, $S = 0.00935$, 正如罗荣生在论文中叙述的, 作图计算法优于近似解法而计算机数值计算法又优于作图计算法。

Munson⁽¹⁾也给出一组模拟数据, 并提供产生这些数据的真实系统的参数 $K_1 = 2.3$, $K_2 = 0.11$, $q_1 = 0.11$, $q_2 = 0.55$, 计算得 $S = 0.503$ 。用 Munson 算法结果为 $K_1 = 2.1$, $K_2 = 0.074$, $q_1 = 0.14$, $q_2 = 0.62$, 计算得 $S = 0.614$ 。我们的结果为 $K_1 = 2.63$, $K_2 = 0.072$, $q_1 = 0.114$, $q_2 = 0.64$, $S = 0.000677$ 。因 Munson 模型在 Feldman 公式中增加了二项修正参数, 和我们模型不尽相同, 难于精确比较, 但从两算法的 S 的数量级上差异可看出我们算法在估计参数

的精度上优于 Munson 算法。

讨 论

本方法用确切的数学模型通过数值计算和判别拟合优度改善的 F 检验方法去最佳拟合实验数据，避免经典的近似方法可能引入的谬误。较之 Munson 等人的方法，本方法除参数估计精度高外还有以下优点：第一，算法规模小，具有很强的递归功能便于在微机上实现。第二，无论初值选取、d 参量修正、非负参数值限制和收敛保证均在计算过程中智能性自动实现，这样使最终模型确定和参数辨识一次自动完成。而 Munson 算法在这些步骤上不是不予考虑就是采取人工干预，使计算无法自动完成还常需经验。

Munson 在最近发表的论文⁽³⁾中提及，他们的算法有些改进，也考虑了初值选取，但没有给出方法。McPherson 也曾将 Munson 的原程序改在微机上运行⁽²⁾，但只是把大程序分块分时在微机上运行而已。

应该注意到，Feldman 推导数学模型的前提是不考虑协同作用，而协同作用在生物体内是存在的，负协同效应可能和多受体情况相混

淆，这当然不是本方法所能解决的，有待于对协同作用的反应系统的数学描述的深入研究。

致谢 本文得到徐仁宝教授的指导和帮助。

参 考 文 献

- 1 Munson PJ, Rodbard D. Ligand, a versatile computerized approach for characterization of ligand-binding system. *Anal Biochem* 1980; 107 : 220
- 2 McPherson GA. Analysis of radioligand binding experiment on a microcomputer system. *Trends in Pharmacol Sci* 1983; 4 : 369
- 3 Munson PJ, Rodbard D. Computerized analysis of ligand binding data: basic principles and recent developments. In: Rodbard D, Forti G, eds. *Computer in endocrinology*. NY: Raven Press, 1984 : 117-44
- 4 Feldman HA. Mathematical theory of complex ligand-binding system at equilibrium: some methods for parameter fitting. *Anal Biochem* 1972; 48 : 317
- 5 罗荣生. 激素-受体分子相互作用的 Scatchard 图数学分析. 生物化学与生物物理进展 1980; (3) : 40
- 6 刘德贵、费景高、于泳江、李广元. FORTRAN 算法汇编(第二分册). 第 1 版. 北京: 国防工业出版社, 1983 : 129-34

Acta Pharmacologica Sinica 1987 Jul; 8 (4) : 297-300

A mathematical description of ligand-receptor interaction and computer identification of its parameters

QIAN Guo-Zheng, YANG Yi-Li (The Second Military Medical College, Shanghai 200433)

ABSTRACT For the purpose of making the parameter estimation more reasonable and accurate in the receptor research, one form for the Feldman's formula is given as the mathematical model of ligand-receptor interaction, and the corresponding microcomputer algorithm is presented in this paper. By using this program, the character of an

unknown ligand-receptor interaction system and all the parameters in the model describing this system can be identified.

KEY WORDS receptors; mathematical models; ligands; computer identification; curve fitting