

简 报

计算蛋白质可及性的一种新方法

温元凯 戴 亚 孟庆涛 赵 辉

(中国科学技术大学量子生物化学实验室, 合肥)

关于蛋白质的静态可及性, 许多作者作过不少的专门研究和探讨^[1-8], 我们在前人工作的基础上, 探讨了蛋白质中原子的环境和对溶剂的暴露, 球状蛋白质中可解离基的外围特征, 以及蛋白质分子面积。在这

里, 我们主要讨论蛋白质的可及表面积与分子量之间的函数关系, 建立一个简单的表达式。

为了叙述的方便, 我们采用下列符号:

A_T , 表示非折叠蛋白质可能的可及表面积;

表 1 15 种蛋白质的可及表面面积和埋藏面积 ($\text{\AA}^2$)

蛋 白 质	分子量	残基取和理论值			Cyrns Chothia 的结果			本文的结果		
		A_T	A_S	A_B	A_T	A_S	A_B	A_T	A_S	A_B
胰岛素*	5,787	8,380	3,440	4,940	8,333 (-0.6)	3,103 (-9.8)	5,230 (5.8)	8,533 (1.8)	3,447 (0.2)	5,086 (3.0)
红氧还蛋白†	6,108	8,670	3,320	5,350	8,795 (1.4)	3,259 (-1.8)	5,536 (3.5)	8,995 (3.7)	3,587 (8.0)	5,480 (1.1)
猪胰蛋白酶抑制剂	6,158	9,440	3,700	5,740	8,868 (-6.0)	3,330 (-10.0)	5,583 (-2.7)	9,068 (-3.9)	3,610 (-2.4)	5,458 (-4.9)
高位离子蛋白	8,976	13,380	4,600	8,780	12,925 (-3.4)	4,591 (-0.2)	8,334 (-5.0)	13,125 (-1.9)	4,755 (3.4)	8,370 (-4.7)
钙结合蛋白	11,489	16,740	5,740	11,000	16,544 (-1.2)	5,653 (-1.5)	10,890 (-1.0)	16,744 (0.02)	5,678 (-1.1)	11,066 (0.6)
核糖核酸酶 S	13,690	19,840	6,550	13,290	19,714 (-0.6)	6,513 (-0.6)	13,201 (-0.7)	19,914 (0.4)	6,432 (-1.8)	13,482 (1.4)
溶菌酶	14,814	20,560	6,480	14,080	21,332 (3.4)	6,908 (6.6)	14,424 (2.4)	21,532 (4.7)	6,801 (5.0)	14,732 (4.6)
葡萄球菌核酸酶††	15,973	23,800	7,840	15,960	23,001 (-3.4)	7,305 (-6.8)	15,696 (-1.7)	23,201 (-2.5)	7,172 (-8.5)	16,029 (0.4)
木瓜蛋白酶	23,423	33,850	9,210	24,640	33,730 (-0.4)	9,363 (1.7)	24,367 (-1.1)	33,930 (0.2)	9,375 (1.8)	24,555 (-0.3)
伴刀豆球蛋白 A*	25,578	37,030	10,230	26,800	36,832 (-0.5)	9,797 (-4.2)	27,035 (0.9)	37,032 (0.005)	9,966 (-2.6)	27,066 (1.0)
枯草杆菌蛋白酶	27,534	40,090	9,790	30,300	39,649 (-1.1)	10,129 (3.5)	29,520 (-2.6)	39,849 (-0.6)	10,488 (7.1)	29,361 (-3.1)
嗜热菌蛋白酶	34,334	49,410	11,570	37,840	49,441 (0.06)	10,824 (-6.4)	38,617 (2.1)	49,041 (0.06)	12,215 (5.6)	37,426 (-0.9)
羧肽酶-A	34,409	49,870	11,260	38,610	49,549 (-0.7)	10,828 (-3.8)	38,721 (0.3)	49,749 (-0.2)	12,233 (8.6)	37,516 (-2.8)
乳酸脱氢酶*	36,470	53,080	17,060	36,020	52,517 (-1.1)	10,895 (-36.1)	41,622 (15.6)	52,717 (-0.6)	12,733 (-25.4)	39,984 (11.0)
α -胰凝乳蛋白酶	24,608	35,900	10,130	25,700	35,436 (-1.3)	9,611 (-5.1)	25,825 (0.5)	35,636 (-0.7)	9,702 (-4.2)	25,934 (0.9)
平均偏差百分数					1.68	6.54	3.06	1.42	5.73	2.71

面积用 $\text{\AA}^2$ 表示。圆括号中数字表示相对误差。

* 计算在孤立的单体上进行。† 由 X-射线分析得到的序列分子量。

†† 计算不包括末端残基 143—149, 其残基在 Fourier 射线中见不到。

A_S , 表示折叠的天然蛋白质可及表面面积;

A_B , 表示蛋白质折叠时被埋藏的可及表面面积,

$$A_B = A_T - A_S;$$

F_{AS} , 表示蛋白质埋藏面积所占的比例, $F_{AS} = A_B/A_T$;

E_{HP} , 表示疏水自由能。

A_T 与蛋白质分子量 M 的关系 C. Chothia^[4] 给出了一个简单的关系式

$$A_T = 1.44M \quad (1)$$

由上式得到的结果, 偏差较大, 且多为负偏差, 见表 1。本文根据误差理论将上式校正为

$$A_T = 1.44M + 200 \quad (2)$$

对一个球状物体来说, 根据表面 A 和体积 V 公式

$$A = 4\pi R^2 \quad (3)$$

$$V = \frac{4}{3}\pi R^3 \quad (4)$$

其中 R 是球半径。

得 A 与 $V^{2/3}$ 成比例。我们不妨假定分子的密度相同, 换言之, 单位质量的体积 $\bar{V}$ 相同, 于是

$$V = \frac{4}{3}\pi R^3 = M\bar{V}/N \quad (4')$$

式中 M 为分子量, N 为 Avogadro 常数。于是得到 A 与 $M^{2/3}$ 成比例。本文以 A_S 对 $M^{2/3}$ 作图, 如图 1。发现两者有较好的线性关系, 数学处理的结果是

$$A_S = 11.9M^{2/3} - 400 \quad (5)$$

于是得到

$$A_B = A_T - A_S = 1.44M - 11.9M^{2/3} + 600 \quad (6)$$

$$E_{HP} = 24A_B = 34.56M - 285.6M^{2/3} + 14400 \quad (7)$$

用上述公式, 我们计算了 15 种蛋白质的可及表面面积和埋藏面积, 从所得结果的偏差来看, 我们的结果比前人的要好些, 见表 1。

从这里我们可以看出, 蛋白质的可及表面面积、埋藏面积以及疏水自由能均可以通过简单的方程与分子

量联系起来, 这就使我们能够通过生物大分子的分子量去描述某些结构特征。

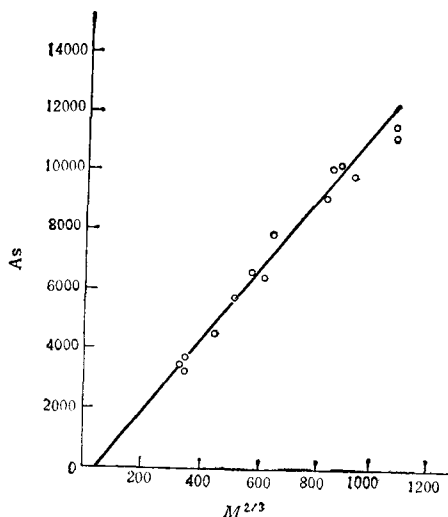


图 1 $A_S-M^{2/3}$ 图

参 考 文 献

- [1] Lee, B. and Richards, F. M.: *J. Mol. Biol.*, 1971, 55, 379.
- [2] Shrake, A. and Rupley, J. A.: *J. Mol. Biol.*, 1973, 79, 351.
- [3] Chothia, C.: *Nature* (London), 1974, 248, 338.
- [4] Chothia, C.: *J. Mol. Biol.*, 1976, 105, 1.
- [5] Richards, F. M.: *Annu. Rev. Biophys. Bioeng.*, 1978, 6, 151.
- [6] Connolly, M. L.: *QCPE. Bull.*, 1981, 1, 75.
- [7] Connolly, M. L.: *Science*, 1983, 221, 709.
- [8] Richmond, T. L.: *J. Mol. Biol.*, 1984, 178, 63.

[本文于 1987 年 4 月 25 日收到]

科技消息

第一届分子生物物理学术讨论会征文通知

由中国生物物理学会分子生物物理专业委员会主办的第一届分子生物物理学术讨论会, 拟定于一九八八年十二月十二日至十五日在福建省厦门市召开。

本届学术讨论会以生物大分子的结构、运动、功能、改造为中心议题, 其主要内容包括: (一) 蛋白质、核酸等生物大分子的结构与功能研究; (二) 生物大分子的动力学研究; (三) 蛋白质工程及有关问题; (四) 生

物大分子构象理论与分子设计等分子生物物理问题。

论文摘要请于九月三十日前寄到北京 349 信箱生物物理学会吕克定同志处。每篇摘要限 A₄ 复印纸一页, 请用中文打印。上空 3cm, 下和左右空 2.5cm 边框。有关其它事宜请与吕克定同志联系。

[中国生物物理学会办公室]