

蛋白质一级结构的“远缘”比较

王克夷 冯佑民

(中国科学院上海生物化学研究所)

提 要

由于核苷酸顺序测定方法的突破,致使蛋白质(尤其是分子量大、低含量的蛋白质)的一级结构测定大为简便,越来越多的蛋白质的一级结构已被测定。与此同时,电子计算机技术发展迅速,并广泛地应用于生物科学。这两者的结合,开拓了蛋白质一级结构比较研究的领域,在一些尚未想到会有关系的蛋白质一级结构之间也发现有同源性。这种比较研究可称为“远缘”比较,其结果是建立了蛋白质一级结构的“家谱”。本文例举七个类型蛋白质一级结构的“远缘”比较,并对“远缘”比较给予蛋白质结构和功能关系,分子进化的研究提供的启迪作一概要的讨论。

七十年代中期, Sanger 等建立了测定核苷酸顺序的方法,从而为测定生物体内含量低、分子量大的蛋白质的一级结构提供了有力的手段。近十年来,越来越多的蛋白质借助核酸法测定其一级结构。已知一级结构的蛋白质的数目和氨基酸排列顺序已构成庞大的数据库。

电子计算机技术的发展和运用,使人们可以运用电子计算机比较蛋白质的一级结构^[1]。在一个蛋白质的一级结构或其 cDNA 的顺序被测定后,立即可和数据库中的信息进行比较。这样的比较研究不局限于同种蛋白质或功能相近的蛋白质,为蛋白质一级结构的比较研究展示出更广阔的天地。有些科学家致力于这种比较研究^[2,3]。如果称不同来源的同种蛋白质的比较研究为“同宗”比较,一些功能相近的比较为“近亲”比较,则借助电子计算机,对看来毫不相关的蛋白质进行的比较可称为“远缘”比较。一些“远缘”相关的蛋白质组成相应的超级家族。

一、“远缘”比较的若干例子

(1) 1967 年 Brew 等发现牛乳白蛋白和

鸡蛋清中的溶菌酶,两者的一级结构有约 50% 的同源性^[4]。溶菌酶结合部位中某些残基,在乳白蛋白也可找到对应者。从来源看,它们可谓“远缘”。但它们还不是典型的“远缘”相关的例子。因为两者在功能上已知有相似之处。乳白蛋白是乳糖合成酶的一个亚基,溶菌酶则是内切糖苷水解酶,它们都和糖类的合成或降解代谢有关。

(2) 至今所知的最大蛋白质超级家族之一是 Serpins^[5]。它们中最先发现的成员是血清中丝氨酸蛋白酶抑制剂 (serine proteinase inhibitors), serpin 则是英文词首字母组成的新词。这些抑制剂的特点是分子量约 4 万,多数已被鉴定是糖蛋白,活性中心都在分子的 C 端附近。活性中心 $P_1-P'_1$ 为 X-S, S 为丝氨酸, X 因其作用的酶而异。这些成员的活性丝氨酸 C 端一侧近 30 个残基约有 1/3 是保守的。

有些蛋白质,尽管它们并不是抑制剂,但从结构来看,也属于这一超级家族。鸡蛋清中卵白蛋白和抗凝血酶 III 以及 α_1 蛋白酶抑制剂之间的同源性分别为 31% 和 24%,而上述两个抑制剂之间的同源性却是 28%。这些同源的残

基分散在整个分子中。至今仍未测得卵白蛋白具有蛋白酶抑制剂的活性。

血管紧张素原也是 Serpin 家族的一员。它所释放的血管紧张素 I 和 II，则在分子 N 端附近。其一级结构和 α_1 -蛋白酶抑制剂、卵白蛋白、抗凝血酶 III 的同源性分别为 23.0%、20.9% 和 17.8%。

甲状腺素结合蛋白的一级结构最近被测定^[6]，和 α_1 -抗胰凝乳蛋白酶以及 α_1 -蛋白酶抑制剂的同源性高达 58% 和 53%，但与抗凝血酶 III 仅 27% 同源性。血清中另一个能与甲状腺素以及胡萝卜素结合的前白蛋白，虽然和上述蛋白有相似的结合活性，但结构却不同，它和一些多肽激素有同源性^[3]。

更令人惊奇的是大麦内胚乳中的蛋白质 Z，是一个分子量为 4.3 万的贮藏蛋白，但也是 Serpin 家族的成员。

(3) α_2 巨球蛋白^[7]、C₃ 原^[8]和 C₄ 原^[7]三者分别由 1451、1641 和 1722 个残基构成。它们都具有分子内环状硫酯键。这类蛋白质很少，因此很自然地被归在一个家族中。分子内环状硫酯键都是由半胱氨酰-甘氨酰-谷氨酰-谷氨酰胺酰这四肽首尾二残基的巯基和 γ 羧基缩合而成。尽管它们属于同一家族，但目前所知的功能却相反。 α_2 巨球蛋白是广谱蛋白酶抑制剂，而 C₃ 原和 C₄ 原则是补体系统中两个成员，激活后具有蛋白酶活性。就整个分子而言，C₃ 原和 C₄ 原， α_2 巨球蛋白分别有 29% 和 23.5%

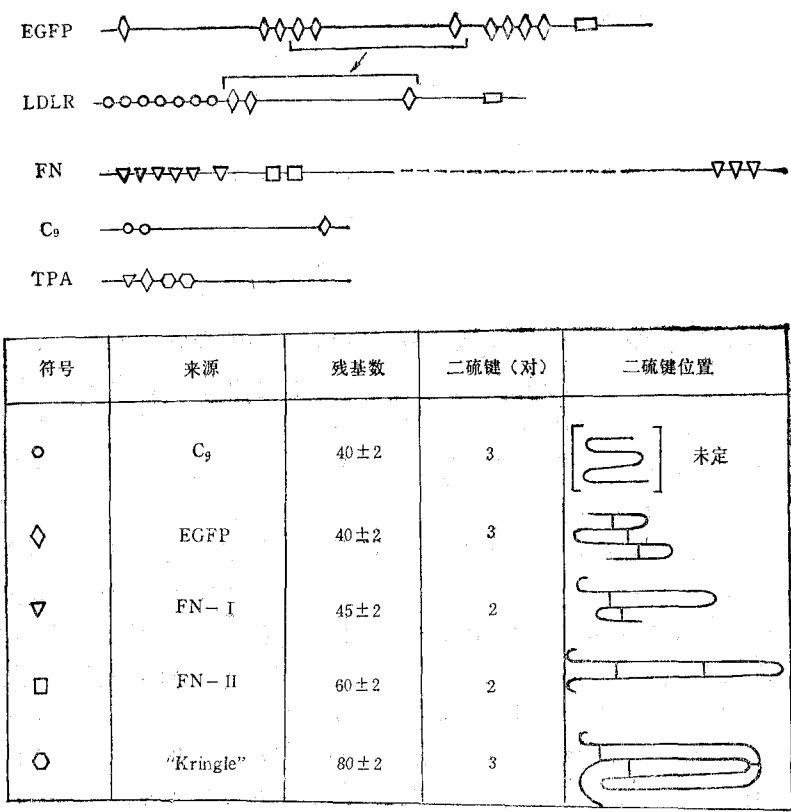


图 1 若干镶嵌蛋白的结构比较^[27]

图中缩写 EGFP: 表皮生长因子前体; LDLR: 低密度脂蛋白受体; FN: 纤维粘连蛋白; C₃: 补体第九成分; TPA: 组织溶血纤蛋白酶原激活剂; ms: 跨越膜的肽段

的同源性。C₃ 原和 α_2 巨球蛋白分别有八个相似的肽段，这些肽段中相同和相似的残基，则为 32% 至 40%。这八个肽段为一些大小不等的

肽段所隔开。

(4) 近几年来一些肝细胞表面的糖结合蛋白的一级结构已被测定^[9]。虽然它们的糖结合

专一性不尽相同,但它们之间有相当的同源性,尤其是C端半分子更相似,被认为是糖结合部位。

肺泡表面活性物质是类脂和糖蛋白的复合物。经“远缘”比较,发现人和狗肺泡表面活性物质中分子量约3.4万的脱辅基蛋白和大鼠肝表面的甘露糖结合蛋白在一级结构上有惊人的相似性^[9]。整个分子可分为三部分,N端起始的十至二十个残基不相同。接着是由G-X-Y构成的重复结构,G为甘氨酸,X和Y为其他氨基酸,但Y经常是脯氨酸或羟脯氨酸,是典型的胶原中的结构单元。C端半分子则和糖结合部位相似。这一发现可认为是典型“远缘”比较的收获,因为在这以前根本不知道,也不会想到肺泡表面活性物质中3.4万脱辅基蛋白会和糖结合蛋白有同源性。

人淋巴细胞表面IgE Fc区的受体,N端是跨越质膜的肽段,而C端半分子也和上述动物凝集素有30%的同源性^[10]。

(5) 表皮生长因子(EGF)是一个53肽。但用核酸法测得其前体由1276个残基构成,约EGF的20倍。其结构和血清中低密度脂蛋白的受体的结构很相似(见图1)^[2,11]。这两个分子在C端附近有一段高度疏水的肽段。两者都有在分子中重复出现多次的、半胱氨酸含量较高的肽段,约40个残基组成。EGFP和LDLR,两者的中间部分(y和y')又有高度同源性。正中间约40%同源,而N端和C端两侧同源性分别为30%和27%。

(6) 1983年测定了血小板衍生的生长因子2,由70个残基组成,其中61个与Simian肉瘤病毒的肿瘤基因v-sis表达产物同源^[3]。以后又发现,鸟类红血球母细胞瘤病毒的VerB产物中间部分约300个残基中仅18个残基和EGF受体的相应部位不同,而且60个残基在src家族中是保守的^[12]。胰岛素具有生长因子的活性,其受体的β亚基与肿瘤基因src家族表达产物在一级结构上非常相似,有一肽段的同源性高达70%。而且这些膜蛋白的细胞浆内的部分都有催化酪氨酸残基磷酸化的活

性^[13]。细胞浆内的G蛋白也是生长因子作用的重要环节,肿瘤基因表达产物p20ras和G蛋白在一级结构上也同源^[14]。

(7) 甲状腺球蛋白有两个亚基,每个亚基由2750个残基构成。这亚基是迄今所知最大的亚基^[15]。它们是甲状腺素的前体。整个分子可分为四部分,其中三部分都有重复结构。只有C端约600个残基,既无重复结构,和其他三部分也无同源性。有趣的是,这一部分竟和多种乙酰胆碱酯酶中的一种有28%同源性^[16]。两者中有一片段的同源性高达60%,且多数半胱氨酸的位置也很保守。只是乙酰胆碱酯酶活性丝氨酸,在甲状腺球蛋白中没有对应者。

二、“远缘”比较的启迪

(1) “远缘”比较可提示蛋白质可能具有未知的功能

很多蛋白质在结构上常常可以分为若干结构域,不同结构域承担着不同的功能。“远缘”比较发现一些功能不同的蛋白质具有相似的结构,这提示了这些蛋白质也可能具有多个结构域,并对应着不同的功能。这些相似的结构域可能承担着尚未为人们所知晓的功能。

Serpin这一超级家族中很多成员至今仍未见其抑制活性,但它们可能是某些活性物质的前体。卵白蛋白在四十年代就发现,经枯草杆菌酶作用后,除去一段六肽即变成片清蛋白。但至今仍未发现酶切后的两个产物有任何生物活性。血管紧张素原可释放出活性的血管紧张素,但其C端残余部分有无活性,还不知道。然而,血清中低分子量的激肽原,既可释放出有活性的舒缓激肽,分子的剩余部分则又是巯基酶的抑制剂。因此,Serpin家族诸成员是否既是一些活性小肽的前体,又是抑制剂呢?

EGFP和LDLR的同源性提示EGFP是某些活性物质受体的可能性。小鼠肾脏等组织中都发现有EGFP,但都不分泌EGF,这事实也支持上述假设。但究竟是不是受体?是什么的受体?还有待进一步研究。

(2) “远缘”比较可为研究蛋白质的未知功

能提供线索

最近发现大鼠血清中甘露糖结合蛋白部分一级结构和肝脏表面甘露糖结合蛋白相似,并且还能引发补体反应^[47]。再者淋巴细胞表面 IgE F_c 的受体也和这些动物凝集素有同源性。这些事实提示这一家族蛋白质,包括肺泡表面活性物质中 3.4 万的脱辅基蛋白,都可能参与动物的非特异免疫防卫反应,与入侵异物清除有关。

肝脏表面甘露糖结合蛋白没有跨膜的疏水肽段,但有胶原的特征重复结构。这意味着肝脏表面甘露糖结合蛋白,和其有相似结构的肺泡表面活性物质中 3.4 万脱辅基蛋白是通过和纤维粘连蛋白中胶原结合部位的作用,再粘着到细胞表面。

甲状腺球蛋白分子极大,如果仅释放四个甲状腺素,似乎不符合生物体内“经济”的原则,可能还有目前还不知晓的功能。它 C 端近 600 个残基和乙酰胆碱酯酶有同源性给予一定的启示。这两个分子分别是激素的前体和神经递质的降解者,因此它们结构上的相似部分很可能反映出两者和其相关细胞的作用有共性。

不同来源肿瘤基因的表达产物与细胞生长有关的一些活性分子有同源性。这表明生长因子的作用原理和癌变的原因可能有关。为此有人认为,生长因子及其相关分子在错误时间或错误地点发生作用就会导致癌变^[48]。

(3) “远缘”比较对生物大分子进化研究的启示

Doolittle 在比较脊椎动物蛋白质家谱关系时,提出了融合蛋白 (Fusion Protein) 和镶嵌蛋白 (Mosaic Protein) 这两个名词^[2]。我们不仅可以从结构上,而且可以从分子进化角度来理解这两名词。该文指出:补体蛋白 C_{1q} 是胶原基因和一个尚未鉴定基因的融合产物,故而有胶原样结构;一个活性和钙有关的蛋白酶是组织蛋白酶基因和钙调蛋白基因的融合产物。就这意义而言,肺泡表面活性物质中 3.4 万脱辅基蛋白则是胶原蛋白基因和糖结合蛋白基因融合后的表达产物。甲状腺球蛋白 C 端 600 个

残基,很可能是在进化时融合了类似于乙酰胆碱酯酶基因而获得的。镶嵌蛋白的突出例子是 TPA (见图 1)。这分子中明显地可以见到很多存在于其他蛋白质中的片段:有 EGFP 中的重复结构,有纤维粘连蛋白中的 II 型结构单元,还有一些存在于血清蛋白酶前体中的“Kringle”片段。这些例子说明蛋白质分子进化时,不仅有基因突变,基因重复,还有基因融合以及多种外显子的搅拌混杂 (Shuffling)。“远缘”比较可以揭示有关分子在进化时的亲缘关系。

目前很多蛋白质的结构来自核酸结构的测定,因此直接比较基因结构,又可能得到更多信息。例如,从蛋白质结构比较,仅看到血管紧张素原和 α_1 蛋白酶抑制剂的同源性为 23%。然而在它们的基因中还有十个内含子的分布也相同,这也有力地支持两者属于同一家族^[49]。

多结构域,多功能蛋白质决不是结构和功能简单地加和,一定有其内在有机的联系。因此蛋白质在进化时,也绝不是基因简单的重复和融合,也不是外显子机械地剪接拼凑。控制基因和外显子有机联系的因素又是什么?这是一个值得深入研究的课题。

总之,蛋白质一级结构和基因结构的“远缘”比较必将为我们提供更多、有些甚至是意想不到的结果,从而极大地丰富我们蛋白质结构和功能关系,以及蛋白质和基因分子进化的知识。

参 考 文 献

- [1] Dayhoff, G. A. et al.: *Methods of Enzymol.*, 1983, 91, 524.
- [2] Doolittle, R. F.: *Trends in Biochem. Sciences*, 1985, 10, 233.
- [3] Doolittle, R. F.: *Science*, 1983, 221, 275.
- [4] Hill, R. L. et al.: *Adv. Enzymol.*, 1975, 43, 411.
- [5] Carrell, R. W. et al.: *Proteinase Inhibitors*, eds. Barrett, A. J. et al. Elsevier, Amsterdam, 1986, 403.
- [6] Flink, I. L. et al.: *Proc. Natl. Acad. Sci. U. S. A.*, 1986, 83, 709.
- [7] Sottrup-Jensen, L. et al.: *Proc. Natl. Acad. Sci. U. S. A.*, 1985, 82, 9.
- [8] de Bruijn, M. H. L. et al.: *Proc. Natl. Acad. Sci. U. S. A.*, 1985, 82, 708.

(下转第 192 页)

性,并使其对复合的气味刺激也具有敏感性。此外,局部中间神经元还起着巨型球状体与其他球状体之间的联系作用(图4)。局部中间神经元的背景放电对输出中间神经元有抑制作用^[48],因而这种自发放电频率的模式是将嗅觉信息从触角叶转换到前脑的重要而经济的方式。所以,中脑的中间神经元不仅起一个会聚信息的作用,而且还将信息进行了转换加工。

中脑输出中间神经元具有广泛的辐散特性。经过中脑加工的嗅觉信息在前脑有广泛的分布,这种性质决定了前脑的信息整合进入了更高水平。Light^[48]提出,前脑的功能可能有三种:1)对真实环境的性外激素(比如一定方向的气流所含的气味物质)进行抽象与重建,需要多种感觉通道,这些通道的信息在前脑得到整合;2)由于先期刺激的持续作用,前脑某些神经元回路对多感觉通道传来的一系列刺激可同时编码,产生新的适宜输出,因而前脑可对系列的环境事件进行反应。3)进入前脑广泛分布的信息可能转换成运动指令,并编码运动方向。

这样,似乎传入通路的整合有一确定过程,然而我们对中脑的神经元回路及其在前脑的投射还知之甚少^[47],因而,昆虫嗅觉的中枢过程研究依然是初步的。

六、结 语

近十年来神经科学研究领域内已采用几项新技术,一是重组DNA,单克隆抗体技术及原位杂交技术在神经系统研究中的应用;二是采用膜片钳制技术研究离子通道机制;三是神经细胞标记与示踪方法的革新等。这几项技术在昆虫性外激素嗅觉过程的研究中,还尚未被广

泛采用。随着这些技术的广泛应用,将使我们昆虫嗅觉的神经生物学过程有更深入的了解。

参 考 文 献

- [1] Mustaparta, H.: *Chemistry Ecology of Insect* (Bell W. J. and Cardre, R. T. eds), Chapman and Hall Ltd., London, 1984, 37—70.
- [2] Schneider, D.: *Science*, 1969, **163**, 1031.
- [3] Boeckh, J.: *Insect Communication* (Lewis, T. ed.), Academic Press, Inc. London, 1984, 83—102.
- [4] Kaissling, K. E.: *Chemical Ecology: Odour Communication in Animals* (Ritter, F. J. ed.), Elsevier North-Holland Biomed. Press, 1979, 43—56.
- [5] Bestmann, H. J. et al.: *Z. Naturforsch.*, 1978, **42c**, 435.
- [6] 吴才宏等:《科学通报》,1989,Vol. **34**, No. 15.
- [7] Zack, C.: *Sensory Adaptation in the Sex Pheromone Receptor Cell of Saturniid Moths*, Ludwig Maximilians-Universität, München, 1979, 1—99.
- [8] Minks, et al.: *J. Insect Physiol.*, 1974, **20**, 1659.
- [9] Roelofs, W. L. et al.: *J. Insect Physiol.*, 1971, **17**, 1969.
- [10] Miller, J. R. et al.: *J. Chem. Ecol.*, 1978, **4**, 187.
- [11] Roelofs, W. L. et al.: *Ann. Rev. Entomol.*, 1977, **22**, 377.
- [12] Kochansky, J. et al.: *J. Insect Physiol.*, 1975, **121**, 1977.
- [13] Kaissling, K. E.: *Int. Symp. Olfaction and Taste 5* (LeMagnen, J. et al. eds) Information Retrieval, London, 1977, 9—16.
- [14] Bestmann, H. J. et al.: *Z. Naturforsch.*, 1987, **42c**, 631.
- [15] Priesner, E.: *Chemical Ecology: Odour Communication in Animals* (Ritter, F. J. ed.), Elsevier North-Holl and Biomed. Press., 1979, 60.
- [16] Silverstein, R. M.: *Insect Communication* (Lewis, T. ed.), Academic Press, Inc. London, 1984, 105—119.
- [17] Boeckh, et al.: *J. Insect Physiol.*, 1984, **30**, 15.
- [18] Light, D. M.: *Mechanisms in Insect Olfaction* (Payne, T. L. et al. eds), Clarendon Press, Oxford, 1986, 287—302.
- [19] Kaissling, K. E.: *Ann. Rev. Neurosci.*, 1986, **9**, 121.
- [20] Steinbrecht, R. A. et al.: *Cell Tissue Research*, 1984, **235**, 25—45.

[本文于1988年1月29日收到]

(上接第179页)

- [9] Drickamer, K. et al.: *J. Biol. Chem.*, 1986, **261**, 6878.
- [10] Ikuta, K. et al.: *Proc. Natl. Acad. Sci. U. S. A.*, 1987, **84**, 819.
- [11] Yamamoto, T. et al.: *Cell*, 1984, **39**, 27.
- [12] Ullrich, A. et al.: *Nature*, 1984, **309**, 418.
- [13] Ullrich, A. et al.: *Nature*, 1985, **313**, 756.
- [14] Itoh, H. et al.: *Proc. Natl. Acad. Sci. U. S. A.*, 1986,

83, 3776.

- [15] Mercken, L. et al.: *Nature*, 1985, **316**, 647.
- [16] Ikeda, K. et al.: *J. Biol. Chem.*, 1987, **262**, 7451.
- [17] Schumacher, M. et al.: *Nature*, 1986, **319**, 407.
- [18] Massague, J.: *Trends in Biochem. Sci.*, 1985, **10**, 237.
- [19] Tanaka, T. et al.: *J. Biol. Chem.*, 1984, **259**, 8063.

[本文于1988年2月4日收到]