

DOI: 10.13376/j.cblls/2015104

文章编号: 1004-0374(2015)06-0755-06

胰岛素结构的序列分析(节译)

F. 桑格(Frederick Sanger)

中国科学院生物化学与细胞生物学研究所 刘望夷译



F. 桑格(F. Sanger) (1918—2013)
英国科学家, 曾于1958年和1980年
两次获Nobel化学奖。

译文简介: F. Sanger 于1955年完成胰岛素全部氨基酸序列分析, 1958年获诺贝尔化学奖, 又于1978年发表双脱氧DNA序列测定法, 1980年再次获诺贝尔化学奖。本译文节译自Sanger的回忆录(Sanger F. Sequences, sequences and sequences. Ann Rev Biochem, 1988, 57:1-28)中测定胰岛素序列的一部分内容。

F. Sanger 于1943年获得博士学位(25岁)后就到剑桥大学A.C. Chibnall教授领导的生物化学实验室工作。Chibnall的研究课题就是分析胰岛素的氨基酸组成及其化学结构。一开始, Chibnall就建议Sanger测定胰岛素的自由氨基(实际上就是N-末端氨基酸)。此后, Sanger就一直从事胰岛素氨基酸的序列分析, 直到完成这个蛋白质的全部氨基酸序列的测定和其中三对二硫键的定位。Sanger非常感激Chibnall引导他一开始工作就着手研究胰岛素。他认为没有Chibnall的建议, 他可能会继续从事他的博士论文——赖氨酸代谢方面的研究。他研究胰岛素N-端氨基酸的课题实际上是Chibnall给他提出的, 发表文章时Chibnall让Sanger单独署名, 而Sanger认为文章上Chibnall应该署名。这样的例子国内少见。

在研究胰岛素氨基酸序列中, α -氨基的显色标记是重要的一步。Sanger在这个问题上花费了很多时间。另外, 肽段的分离也是非常关键的一步。肽段分离不好就无法测定氨基酸序列。在20世纪40年代中期Martin和Sygne就成功地发展了纸层析和柱层析技术。这一点Sanger感到非常幸运。用这些方法解决了小肽和大肽的分离, 也完成了测定肽段中的氨基酸序列。

测定胰岛素两条肽链的序列, Sanger的合作者主要是两个人, 即奥地利人Tuppy和澳大利亚人Thompson。Sanger对他们两个都很满意。到1953年胰岛素A、B链的氨基酸序列测定就靠他们两个人基本上完成了。后来另有三位合作者与Sanger一起又花费了两年时间解决了难度较大的二硫键定位问题。从1943年算起到1955年测定出胰岛素全部氨基酸序列历时12年, 工作人员先后只有6人, 发表与此项工作直接有关的文章仅11篇。

Sanger等开展的这项工作主要是在地下室中完成的。与他的实验室为邻的是饲养大鼠的动物房。可想而知, 地下走廊中散发出难以忍受的臭味。但是, Sanger觉得在这间实验室度过的时间是他最愉快的年华。他与同事们合作得很好, 关系融洽。这时他们完成了一项有重大科学意义的工作。Sanger也遇到过不高兴的事。他与生化界一位大人物有点分歧。欲知详情如何, 且看译文分解。

1943年,我拿到博士学位时,Neuberger已经离开实验室。接着我得到机会到A.C. Chibnall教授实验室工作。他的研究组得到医学研究委员会(Medical Research Council, MRC)基金的资助。当时,Chibnall刚继任Hopkins的席位,被任命为生物化学教授。他和他的几位同事正忙于从伦敦皇家学院搬到剑桥。那时他们的主要研究兴趣是分析蛋白质,特别是胰岛素的氨基酸组成。他选择胰岛素作研究材料是因为它的医学重要性,而且在当时是少数几种能够买到的纯化的蛋白质。选择它似乎未必是因为它的分子量小(M_r 6 000),因为当时并不知道它这样小。当然,分子量小对我们今后的工作是非常幸运的。在当时,氨基酸分析是一件艰难辛苦的工作。很多例子说明要得到准确的结果是困难的。Chibnall组的氨基酸分析大概在当时是发表出的最好结果。

当时,Bergmann和Nieman的理论特别流行。他们认为,蛋白质中特定的氨基酸残基沿多肽链有规律地相间排列着,而且每种氨基酸的含量有规则地按公式: $2^m \times 3^n$ 的方式表现出来。这种理论的一个有益效果就是引起了人们对分析氨基酸的兴趣。人们试图按照这个公式表述他们的实验结果,由于缺少精确的实验方法,很少数据能够适用于这个公式。这是完全可以理解的,因为实际上不存在这种规律。Chibnall实验室的足够精确的实验结果使得他们成为怀疑这个公式的第一个研究组。

这里的工作氛围对我是有益的。我开始的实验就是这个研究组工作的合理继续。事实上就是Chibnall建议我做胰岛素工作的。如果没有他的建议,我可能还会继续我的赖氨酸代谢研究。Chibnall等发现,胰岛素的自由氨基数目比计算出来的赖氨酸含量还要多。他们就认为多余的自由氨基是由于出现了 α -氨基基团。这说明肽链比较短,因此,特别适于化学研究。Chibnall建议,要我探索定量测定并鉴定这些多出的自由 α -氨基究竟在哪些氨基酸上。这就是我做胰岛素工作如何开始的。事实上,胰岛素的工作可能开始得更早些。此前,Neuberger曾经开始做过一些初步的胰岛素末端分析工作。这大概是他的兴趣和经验传授给我了吧。毫无疑问,当时和以后相当一段时间,我就开始并一直在做胰岛素全序列的测定工作。仅在最后阶段做某些实验,如定位二硫键和一个酰胺基团时,我才真正感到有了压力。

根据我和别人过去写的综述文章记载,我那时发展了一种使用二硝基氟苯(FDNB)的化学试剂测定蛋白质的末端基团^[1]。虽然这是事实,后来查看了我的旧笔记本,我发现事情并非那么简单。当时已经有几种标记自由氨基的方法,其中仅有一种方法是鉴定N-末端氨基酸的。早在1935年,Jensen和Evans就用苯异腈酸酯(phenylisocyanate)检测出胰岛素的N-末端是苯丙氨酸。他们使用的这种试剂比Edamn试剂,苯异硫腈酸酯(phenylisothiocyanate)的活泼性在某种程度上要差一些。我认为,在任何蛋白质中,苯丙氨酸可能都是第一个被检测出的末端氨基酸。而我们需要的一种能普遍使用的试剂应该可以在温和条件下与氨基反应形成的衍生物对酸水解是稳定的。而且,使用比较简单的方法能很容易分离和鉴定这种氨基酸衍生物也是很重要的。为了达到这个目的,我们很不安地使用了一种新发现的分配层析(Partition Chromatography)法。这种方法比以前使用过的任何方法都好。G.R. Tristram组已经用这种层析法很好地分离了乙酰氨基酸。

除了异腈酸酯外,可以使用的试剂还有磺酰氯(sulfonyl chloride)及其衍生物苯磺酰氯(benzene sulfonyl chloride)。后者使用较多,市场上可以买到。但其产物的溶解度似乎不那么理想。我试着使用的第一种这类试剂是甲基磺酰氯(methane sulfonyl chloride)。选择它的主要原因是我希望甲基磺酰氨基酸的层析行为与乙酰氨基酸衍生物相似。前面提过已经有人很好地做过对后者的层析分离。但是,我却没有得到任何清晰的结果。接着我们转向苯的衍生物,即2,4-二硝基氯苯(2,4 dinitrochlorobenzene)。Abderhalden和Stix于1923年曾经研究过这种衍生物。但是,它仅在高温条件下才能发生反应。在这种条件下,多肽键会有某种程度的水解。我用这种试剂做出了一些二硝基苯(DNP)氨基酸,发现这种氨基酸衍生物在硅胶分配层析柱上分离得很好。一个很大的优点是这些DNP-氨基酸衍生物呈现黄色。那时,所说的层析(Chromatography)真的成了“有色”层析(“Chroma” tography)了。不需要可靠的部分收集器就可以直接在层析柱上看到黄色条带。很明显,氯化物不够活泼。相对来说,氟化物或者硝基化合物的活性是比较理想的。我也曾做了1,2,4-三硝基苯(1,2,4 trinitrobenzene)。后来Chibnall对我说,在附近化学实验室工作的B.C. Saunders博士非常友好地愿意提供给我一些氟化合

物。在冷的条件下，这些试剂与氨基酸的反应是令人非常满意的。于是，我们就有一个很好的提供二硝基氟苯 (2,4-dinitrofluorobenzene, FDNB) 的地方了，以后的工作中经常使用它。可是，我的第一个实验结果是非常扫兴的。我试着用 FDNB 与甘氨酸反应，把产物上柱分离时，不幸的是产物不是预期的一个黄色条带——DNP-甘氨酸，而是两个产量差不多的清晰条带。第二个条带似乎就是含有两个二硝基苯基团的甘氨酸 (diNDP-glycine)。就是说一个 α -N 原子上有两个 DNP 基团。我克服了相当多的困难，总算解决了这个问题。事实上，甘氨酸是个特例，其他氨基酸并没有太多的第二种产物。把磷酸缓冲液换成碳酸缓冲液就完全消除了这个副产品。

我用这个方法分析胰岛素的 N 末端氨基酸，得到了 DNP-苯丙氨酸和 DNP-甘氨酸^[1]。这样看来，如果每个胰岛素分子量是 Mr 6 000，应仅有一个 N 末端氨基酸衍生物。而那时我们认为胰岛素的分子量是 Mr 12 000。这就意味着胰岛素中有两类多肽链。它们通过胱氨酸的二硫键连接在一起。用过甲酸能够破坏这种二硫键，从而有可能将相应的两条肽链分离开来^[2-3]。

就在这时 (1947 年) 我有机会到瑞典的乌普萨拉 (Uppsala) 访问 A. Tiselius 实验室。Tiselius 是蛋白质研究领域享有盛名的一位科学家。因为他发明了分析电泳仪。这种仪器与他的名字连在一起，得到了广泛应用。我前面指出，能用氧化方法破坏连接两条链的二硫键，并试图找到一种方法将肽链分开。当时真没有好的方法能分开这样大小的肽链。Tiselius 的同事们正在使用一种将蛋白质吸附在活性碳柱上的层析系统。看来这种方法似乎有点希望。有人建议我到乌普萨拉花点时间使用这种方法试试看结果如何。对我来说，这似乎是一个激动人心机不可失的好机会。特别是战后的英国有点不景气，百业待举。比较来说瑞典似乎是个世外桃源。

但是，后来我对活性碳柱层析的实验不是很满意。检测层析柱底部出现的条带是一种根据折射系数老的成熟方法。这个实验室的习惯一般是请有熟练技术的助手进行实际操作的。一位女技术员在观察层析柱的技术方面得心应手。我就请她帮我观察柱上的条带。她连续做了几次实验。有一次她观察到 4 个条带。对此，我非常激动。因为有一个时期，我们认为胰岛素的分子量是 12 000，有 4 个肽链。

我将这个结果告诉了 Tiselius。他立即建议以他和我的名义给 *Nature* 杂志送去一封信 (文章)，其中包括我在剑桥已经做出的初步胰岛素氧化工作。我当时感到颇为震惊，因为他对此项工作没有做过任何贡献，我从来没有看到他在实验室里工作过。我巧妙地设法说服他，我还没有准备好发表胰岛素的氧化工作。作为他的客人，而且是一个职位很低的工作人员，我没有生硬地顶撞他，只同意发表一篇我们约定的研究结果^[4]。幸运的是，发表的这篇是我的一篇不怎么出名的文章。这真是一篇令我感到羞耻的文章。(译注：因为后来证明，这篇文章的内容是错误的)

Tiselius 确实是一位很友善、有魅力的人。我当然感谢他过去和后来给我的许多支持。但是，通过这件事使我认识到与 Chibnall 合作是多么的幸运。他允许以我一个人的名义发表胰岛素 N-端氨基酸的文章，而正是他启动了这项工作。他有理由把他的名字写在上面。

访问瑞典还有一个收获是认识了 R.L.M. (Dick) Syngé (1952 年诺贝尔化学奖得主)。当时他在那里工作，并且向我介绍了淀粉区带电泳。这种方法是以后更加成功地做纸电泳或凝胶电泳的先驱。

完全酸水解胰岛素，再用 DNP 标记，可以得到各种 DNP-氨基酸。如果酸水解条件降到部分水解胰岛素，可以得到一些 DNP-肽。研究胰岛素的 DNP-肽，我能够鉴定靠近 N 端 4~5 个氨基酸短的序列^[5]。实际上，从胰岛素中只能得到两种 N 端 DNP-氨基酸序列。老实说我当时就觉得胰岛素中肯定有两条多肽链。更重要的是，这可能说明使用酸部分水解法可以测定任何蛋白质的特定序列。这些结果排除了怀疑蛋白质是单一结构的化合物，仅用有机化学方法就可以解决蛋白质结构问题。后来不论是研究蛋白质还是核酸，在大多数工作中，部分水解是测序工作的一种普遍使用的方法。仅在近晚期的 DNA 测序工作中不使用这种方法。

我的胰岛素 N 端氨基酸的第一篇文章曾广泛被别人引用^[1]，而我本人对我的多肽文章更加激动^[5]。回想起来，我感到这篇多肽文章确实具有更加伟大的意义。这是第一次阐明蛋白质分子中排列着各种不同的氨基酸序列。由此，我们可以明确得出结论：蛋白质不是简单的单一化合物。这篇多肽文章是我自己的首创，而第一篇文章则更多地依赖于 Chibnall 的鼓励。

除了我的两位老师：Neuberger 和 Chibnall，我的第一位合作者是 Rodney Porter。虽然他的年龄比我大点，他可是我的正式入室弟子（博士研究生）。这是因为战争年代他在军队服役，而我一直在实验室工作。Rodney 不是那种容易驾驭的人，但他是一位很好的合作者。我觉得，我从他那里学到的东西比他从我这里学到的更多一些。他教我的主要是对待研究工作要持乐观态度。他这样开诚布公、心表如一、推心置腹，大家都感到愉快。他来我这里时，我们已经开展了 DNP 的方法，正在把它扩大到多肽和其他蛋白质研究上。Rodney 曾对 γ -球蛋白 (γ -globulin) 发生兴趣，正想研究它的末端基团。我劝告他说，这种蛋白质是一种复杂的混合物，不值得探索它的末端结构。但是，我没能阻止他。后来，大家都感到很惊奇，他却得到了有见识的简单结果。他对事业表现出“运用之妙，存乎一心”，从那时起，他就与 γ -球蛋白结下了不解之缘，对科学作出了有价值的贡献。

这期间我们暂时栖身在地下室一间实验室里。因为靠近饲养实验大鼠的动物房，尽管我们处于这种气味不愉快的环境里，但这可能是我工作过的最高兴的一个实验室。我们与 Kenneth Bailey 和他的博士研究生 S.V.(Sam) Perry 合用这间实验室。Sam 在军队中服过役，还在利物浦大学 Porter 那里呆过。于是，Bailey 和我这两个有后备席位并且严谨的科学家与两个刚从军队转业下来的野孩子泡在一起。我们当然知道，他们经常会高声歌唱。Sam 偏爱唱的歌曲“巨大的山峰”是模仿 Paul Roberson 的腔调。而 Rodney 则喜欢唱“无人知晓我见过的麻烦”。这是一种乐不可支的和谐氛围，同时我们也完成了许多工作。不料祸从天降，一个烧瓶突然在 Rodney 眼前爆炸并损伤了他的角膜。虽然没有完全失明，可惜他那只眼睛再没有很好的视力了。

虽然在以前工作的基础上 DNP 方法有些改进，而且重要的工作都是用它完成的，DNP 方法仍有不少缺点。最严重不足之处是在酸性条件下，DNP-氨基酸不很稳定，还必须使用校正系数。不同的氨基酸有不同的校正系数。有些氨基酸（比如甘氨酸）的系数是相当大的。事实上，在测定血红蛋白的末端氨基酸中这个问题就产生很大的误差。因此，我仍然对发展新方法很感兴趣。磺酰胺 (sulfonamide, SO_2NH) 键比 DNP 更加稳定，所以使用磺酰氯可能是一种改进。我也曾不安地试图探索颜色较深的试

剂来提高灵敏度。我花了不少时间在一种试剂上，就是二甲氨基叠氮苯磺酰氯 (dimethylaminophenylazobenzene sulfonyl chloride) 或称甲基橙酰氯 (helianthy chloride)。甲基橙-氨基酸在层析图谱上呈现很鲜艳的深红色带。这种甲基橙-氨基酸对酸似乎很稳定。但是，当我们把这种试剂用于蛋白质时，并没有得到很好的结果。这大概是因为叠氮化合物产生某种没有预料的副反应。还有一个不满意的理由，与我们同实验室的同事的生物制品都变成了很光亮的红色。最后我们放弃了这种试剂。后来，Hartley 和 Gray 使用其他的磺酰氯却获得成功。他们的试剂是二甲氨基萘磺酰氯 (dimethylaminonaphthalene sulfonyl chloride, *i.e.* dansyl chloride)。Dansyl-氨基酸的稳定性非常令人满意。强荧光标记的氨基酸很容易检测，并且大大提高了灵敏度。

DNP 法建立之后，使用相当广泛。但是，后来完全被其他新方法特别是 Edman 方法取而代之。这种新方法的优点是，它不仅测定 N-末端氨基酸，而且可以扩大到逐步降解来测定多肽链的序列。有趣的是，Edman 与 DNP 两种方法的建立和发展几乎是同时进行的，但是早期很少使用前者。其主要原因大概是 DNP 氨基酸可见的黄色容易观察和操作，而且那时可靠的部分收集器和微量测定方法尚不够过关。

从测定蛋白质全部序列的角度考虑，DNP 法是很有限的。因为这种方法仅能分析 N-末端少数几个氨基酸序列。要测定蛋白质的全部序列必须先部分水解蛋白。这里的主要技术问题是分离开水解产生的复杂多肽混合物。在全部序列分析工作中，分离多肽链是一件关键性的工作。序列测定的进展的确依赖于多肽链分离方法的进展。在测定胰岛素序列工作中，我们再次感到非常幸运。因为 Martin 和他的同事们正在建立纸层析方法，并且已经用于测定了一个五肽——短杆菌肽 -S (gramicidins-S) 的序列。这种分离小肽的方法远远优于以前任何成功分离其他化合物的方法。这似乎出现了奇迹，当时很难对付的多肽产物却在一张简单的滤纸上彼此分离得非常清晰。使用这个方法，我们能够从酸水解产物中分离出小肽，从中可以经过片段重叠法而推断出较大的肽^[6-7]的序列。然而单独使用酸水解法，我们还不能测出胰岛素的全序列，因为我们必须得到更长的多肽。因此，我们需要一种更加特殊的水解方法。很明显，这就是蛋白水解酶。这个时候，

- [3] Sanger F. Fractionation of oxidized insulin. *Biochem J*, 1949, 44: 126-8
- [4] Tiselius A, Sanger F. Adsorption analysis of oxidized insulin. *Nature*, 1947, 160: 433-34
- [5] Sanger F. The terminal peptides of insulin. *Biochem J*, 1949, 45: 563-74
- [6] Sanger F, Tuppy H. The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysis. *Biochem J*, 1951, 49: 463-81
- [7] Sanger F, Thompson EOP. The amino-acid sequence in the glycyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. *Biochem J*, 1953, 53: 353-66
- [8] Sanger F, Tuppy H. The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysis. *Biochem J*, 1951, 49: 481-90.
- [9] Sanger F, Thompson EOP. The amino-acid sequence of the glycyl of insulin. 2. The investigation of peptides from enzymatic hydrolysates. *Biochem J*, 1953, 53 : 366-74
- [10] Sanger F, Thompson EOP. The insertion of a dipeptide sequence during hydrolysis in dilute acid. *Biochim Biophys Acta*, 1952, 9: 225-6
- [11] Ryle AP, Sanger F. Disulphide interchange reactions. *Biochem J*, 1955, 60: 535-40
- [12] Ryle AP, Sanger F, Smith LF, et al. The disulphide bonds of insulin. *Biochem J*, 1955, 60: 541-56