

未来生物学原理

[英] H. 哈里斯 著

科学出版社

内 容 简 介

本书介绍了人类生化遗传学的基本概念,描述了人类遗传生化的多样性、个体间的生化差异等基本问题。从分子水平上论述了先天性代谢差错、血型抗原的差异、血红蛋白病、酶与蛋白质的多态性、基因突变与遗传病等遗传学和医学中的一些重要问题。

可供遗传学、生物化学和医学等科研工作者及有关大专院校师生参考。

Harry Harris

THE PRINCIPLES OF HUMAN BIOCHEMICAL GENETICS

2nd. REVISED & ENLARGED ED.

North-Holland Publishing Company, 1975

人 类 生 化 遗 传 学 原 理

〔英〕H. 哈里斯 著

沈若谦 薛京伦 许宝孝 译

项 维 校

责任编辑 刘 安

科 学 出 版 社 出 版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1981 年 10 月第 一 版 开本: 787×1092 1/16

1981 年 10 月第一次印刷 印张: 19 1/2

印数: 0001—5,900 字数: 448,000

统一书号: 13031·1722

本社书号: 2349·13-10

定 价: 3.00 元

序

本书汇集了作者在高尔顿实验室 (Galton laboratory) 所作的一系列讲演, 这些讲演不只为了人类遗传学的学生和研究工作者, 也是为生物化学、生物学、医学的学生以及有关领域中感兴趣的研究工作者而作的。作者打算解释人类生化遗传学中近代观点的基本概念; 描述特有的遗传生化多样性的概貌——最近研究表明, 这是人类群体的一个特征; 同时还将指出, 详尽地分析人类各个体间遗传的生化差异, 不仅能阐明遗传学的基本问题, 而且还能阐明医学、以及人类生物学中常碰到的一些基本问题。

正好在十年前, 我写过一篇关于这一课题的报告 (“人类生化遗传学” Cambridge University Press, 1959), 其中大多数资料在当时是有用的, 这一点从时间上来看, 是合乎逻辑的。此后, 这一领域的研究得到了近乎爆炸性的发展, 因此在写现在这本书时, 如果只是对前版加以修订, 很明显不能满足要求。如果要清楚地介绍当前的知识和概念, 许许多多的进展要求有一种截然不同的安排。这不仅是因为前书所涉及的专题在今天已获得更为丰富的知识, 而且还由于在该课题中开辟了整个的新领域, 这种情况在几年前几乎是不能想象的。这些进展的一个重要结果, 是使整个课题成为一个比较统一的整体。过去一个时期内看起来是极不相同并毫无联系的现象——如先天性代谢差错、血型抗原的差异、血红蛋白病、酶与蛋白质的多态性, 现在已可以在统一的理论上考虑它们之间的相互关系, 这在以前几乎是不可能的。这无疑给我们创造了一个尽量以更为系统的分析方法来阐述这一课题的机会。因而现在这本书在写作和安排上均与原先的著作大不相同, 许多讨论的材料当然也是新的。

写这类教科书的一个困难, 是决定用什么样的例子来说明有争议的各种观点, 并且应详尽到什么程度。如果总字数要保持不多不少, 同时又不致因叙述性材料过多而使论点模糊不清, 还必须决定哪些内容应舍弃。常使我感到为难的是, 这类书又能用作参考文献的来源。我打算对书中许多未详述的材料, 通过列出主要的参考文献来解决, 并把许多材料整理成表格或索引的形式, 使大家能方便地选取需要的参考文献。由于这一课题的范围目前是如此之广, 论文数量如此之多, 又分布在各种各样的杂志上, 所以许多论题的参考文献就必然要省略。但是我希望所包括的参考文献能足以使本书既可作为简明阐述这一课题主要原理的入门课本, 又可成为进一步阅读专著的指南。

哈 里 斯

1969 年 7 月于伦敦大学学院高尔顿实验室

第 二 版 序

在过去五年中,人类生化遗传学的工作在继续迅速地发展。因此对该教科书修订再版时,补充了许多新的材料,但本书的基本安排仍按原样。

哈 里 斯

1974年3月于伦敦大学学院高尔顿实验室

目 录

第一章 基因突变和单个氨基酸取代	1
1.1 引言: 基因、DNA 和蛋白质	1
1.2 血红蛋白变种	3
1.3 异常血红蛋白的结构	5
1.4 单个氨基酸取代和遗传密码	7
1.5 单个氨基酸取代的效应	11
1.5.1 镰形细胞病	11
1.5.2 遗传性高铁血红蛋白血症	13
1.5.3 不稳定的血红蛋白	15
1.5.4 改变氧亲和力	16
1.6 一条多肽链上二个各别的氨基酸取代	18
第二章 一个基因——一条多肽链	19
2.1 杂合子中的“杂种”蛋白质	19
2.1.1 血红蛋白	19
2.1.2 杂合子中的“杂种”酶	20
2.1.3 杂合子中对称的及不对称的电泳图谱	23
2.2 多基因座位	26
2.2.1 血红蛋白座位	26
2.2.2 乳酸脱氢酶	33
2.2.3 葡萄糖磷酸变位酶	38
2.3 基因和同功酶	44
2.3.1 多座位决定的同功酶	45
2.3.2 一个座位上的复等位性	49
2.3.3 次级同功酶	49
2.4 酶的亚单位结构	53
第三章 重复、缺失、不等交换、链的延伸和其它重排	57
3.1 结合珠蛋白变种	57
3.2 基因重复和蛋白质进化	63
3.3 不等交换	65
3.3.1 Lepore 血红蛋白	65
3.3.2 Hb Kenya: γ - β 并合产物	68
3.3.3 不等交换是产生更多结合珠蛋白变种的原因	69
3.4 缺失	71
3.5 链的延伸	72
3.6 免疫球蛋白——以上法则的一个例外	74
3.6.1 免疫球蛋白的结构	75
3.6.2 免疫球蛋白基因	77

第四章 影响蛋白质合成速率的基因突变	80
4.1 酶和蛋白质合成的遗传调节	80
4.2 “结构”基因和“调节”基因	81
4.3 结构与速率的关系	83
4.4 蛋白质合成速率的遗传性缺陷——地中海贫血症	84
4.4.1 β -地中海贫血	85
4.4.2 胎儿血红蛋白的遗传性保留(非洲型)	87
4.4.3 影响 β -、 δ -和 γ -链合成的其它异常基因	88
4.4.4 α -地中海贫血	89
第五章 酶量的变异和酶质的变异	92
5.1 酶活力量的变异	92
5.2 血清胆碱酯酶	93
5.2.1 “非典型”血清胆碱酯酶	94
5.2.2 “静止的”等位基因	96
5.2.3 “抗氟化物”的血清胆碱酯酶	98
5.2.4 引起活力“连续”变异的复等位性	98
5.2.5 第二个座位—— E_2	100
5.3 葡萄糖-6-磷酸脱氢酶 (G-6-PD)	101
5.3.1 葡萄糖-6-磷酸脱氢酶缺乏	101
5.3.2 常见的黑人和地中海 G-6-PD 变种: Gd B、Gd A、Gd A-和 Gd 地中海型	103
5.3.3 其它 G-6-PD 变种	106
5.3.4 G-6-PD 和莱昂 (Lyon) 假说	108
5.4 红细胞酸性磷酸酯酶	112
5.4.1 电泳变种	112
5.4.2 数量差异	115
第六章 先天性代谢差错	117
6.1 Garrod 和“先天性代谢差错”的概念	117
6.2 苯酮尿症	119
6.3 半乳糖血症	122
6.4 同功酶缺乏症和组织差异	123
6.4.1 遗传性果糖不耐症的醛缩酶缺乏	124
6.4.2 丙酮酸激酶缺乏症	126
6.5 部分酶缺乏及其代谢后果。尿素循环酶	127
6.6 糖原病	130
6.6.1 糖原合成的缺陷	132
6.6.2 糖原流通的缺陷	133
6.7 溶酶体体积聚病	134
6.7.1 粘多糖症	135
6.7.2 神经鞘脂类沉积症	137
6.7.3 组织变异	138
6.7.4 遗传异质性	139
6.7.5 组织培养研究	142
6.8 杂合子	144

6.8.1	杂合子中部分酶的缺乏	144
6.8.2	急性间歇性血卟啉症: 一种“显性”先天性代谢差错	147
6.8.3	X连锁先天性代谢差错中的杂合子	149
6.9	主动运转系统的缺陷	150
6.9.1	胱氨酸尿症	150
6.9.2	其它氨基酸的运转缺陷	152
6.10	药物代谢的“先天性差错”	154
6.10.1	酶缺乏和药理学上的异常	154
6.10.2	异烟肼失活	155
6.11	酶活力提高的“先天性差错”	157
第七章	血型物质	159
7.1	ABO 血型	159
7.2	“分泌型”和“H”座位	164
7.3	“刘易斯 (Lewis)” 或 “Le” 座位	166
7.4	血型特异糖蛋白的生物合成途径	168
第八章	人类群体中酶和蛋白质的多样性	171
8.1	“常见”和“罕见”等位基因	171
8.1.1	出现多少等位基因	171
8.1.2	血红蛋白	173
8.1.3	葡萄糖-6-磷酸脱氢酶 (G-6-PD)	178
8.1.4	结合珠蛋白变种	178
8.1.5	葡萄糖磷酸变位酶	180
8.2	多态现象	181
8.2.1	酶的调查	181
8.2.2	杂合性的平均度	182
8.2.3	个体的特殊性	183
8.3	罕见的等位基因	185
8.4	电泳检查不出的等位基因	189
8.5	决定先天性代谢差错的等位基因的发生率	191
8.6	等位基因多样性的原因	192
8.6.1	自然选择和随机的遗传漂变	192
8.6.2	分子进化	194
8.6.3	“中性”基因假说: 功能当量问题	198
8.6.4	随机漂变和差别选择之间的争论: 其它的一些问题	200
8.7	等位基因变异范围: 结语	205
第九章	基因突变和遗传病	207
9.1	遗传病的分子病理学	207
9.1.1	在 DNA 水平上确定突变	208
9.1.2	蛋白质变种的分子性质	209
9.1.3	代谢和临床的结果	210
9.2	显性和隐性	211
9.3	遗传病的杂合性	212

9.3.1 一个座位上的复等位性引起的杂合性	213
9.3.2 不同基因座位上的突变引起的杂合性	215
9.3.3 先天性高铁血红蛋白血症: 一个典型例子	216
9.3.4 遗传杂合性在治疗上的意义	217
9.3.5 遗传病变异的其它原因	218
9.4 遗传和环境	220
附录1 缺乏特异性酶引起的病症(先天性代谢差错)	224
A 1.1 碳水化合物代谢病	224
A 1.2 溶酶体酶的缺乏	229
A 1.3 氨基酸代谢的失调	232
A 1.4 混杂性的失调	237
附录2 酶和蛋白质的多态现象	247
A 2.1 酶的多态现象	247
A 2.2 “蛋白质”的多态现象	248
参考文献	250

第一章 基因突变和单个氨基酸取代

1.1 引言:基因、DNA 和蛋白质

人是各种各样的。他们在正常体格、生理和智力等属性上,都彼此不同。他们在是否会生特殊的疾病或其它异常方面亦彼此各异。这些差异,部分是由于他们生活的环境条件不同而引起的。但也取决于其先天的差异。事实上,除了一卵双生的双胞胎外,不可能存在二个遗传素质完全相同的个体。从分子水平来分析这些遗传学上决定性的差异的性质和效应,就是人类生化遗传学的主要内容。

经典遗传学的概念认为,遗传的基本生物学单位是基因,它必须具备三个基本性质。在细胞以及整个机体中,基因必须具有特定的功能。基因必须能够精确地自我复制,从而使其功能特异性一代代保存下去。最后,虽在一般情况下基因是一个极为稳定的实体,但必须会发生突然变化或突变,结果出现一个在功能上与原来不同的新单位或等位基因,并以新的形式自我复制。已经知道:在染色体中,这些单位是如何呈直线排列的,而且每个基因有它特殊的位置或座位;它们是如何通过卵和精子从亲体传递到后代去的,所以它们通常是成对的,其中一个来自一个亲体,另一个来自另一个亲体;以及如何由于以前世代中发生的突变,使一个基因的复等位形式占据一个特殊的基因座位,结果一个自然群体中每个个体根据他们从亲体随机获得的等位基因的特异性而在特征上彼此不同。

四个重要的进展使人们有可能开始从分子水平了解遗传多样性的本质。第一个是发现了赋予基因特性的特殊化学物质是脱氧核糖核酸——DNA。第二个是阐明了 DNA 的分子结构。第三个是认识到 DNA 在生物细胞中的主要生化作用是指导酶和其它蛋白质的合成。第四个是解释了遗传密码。这是核酸结构和蛋白质结构相互关系的细节。

1953 年 Watson 和 Crick 首次阐明了 DNA 分子结构的主要特征,同时指出如何用他们提出的结构来解释遗传物质的三个基本属性:基因特异性、基因复制和基因突变。

DNA 分子是由二条很长的多核苷酸链组成,两条链围绕着一个共同的轴互相盘绕而形成一个双螺旋。每一条链的骨架由磷酸和糖(脱氧核糖)基有规则地交替构成。每个糖基上连着一个指向链内侧的含氮碱基。它可以是下列四种不同类型中的一种:腺嘌呤或鸟嘌呤(都是嘌呤),胸腺嘧啶或胞嘧啶(都是嘧啶)。两条链通过同一平面上突出的碱基对之间的氢键连在一起。因此整个结构可以说象一个螺旋形楼梯,碱基对相当于阶梯。碱基配对有一定限制。每对碱基中的一个碱基一定是嘌呤,另一个一定是嘧啶,可能的组合只有两种:腺嘌呤和胸腺嘧啶,鸟嘌呤和胞嘧啶。一定的碱基对总是两种当中的一种。例如:在两条链中都可能含有胸腺嘧啶,但配对时,在另一条链上的配偶必定是腺嘌呤。

可以认为一个基因是含有几百或几千个碱基对的一段 DNA。虽然构成双螺旋的二条链的磷酸——脱氧核糖骨架是十分规则的,但碱基对的序列可以是任意的。因而可能有许许多多的排列方式,每一基因可有其特异的结构,由此而产生其功能的特异性。一个

特定基因中,碱基对的精确序列好象以密码形式携带着一份特异的遗传信息。

由于一个碱基的性质决定了与之配对的另一个碱基的性质,所以构成分子的两条多核苷酸链,尽管在质上是不同的,但彼此是严格互补的。一条链上的碱基序列决定了另一链的碱基序列。双链解开和分离,从能利用的核苷酸库中在每条链上重新形成与它相对应的链,就可以进行复制。这样每一条链就可成为形成另一条链的模板,结果可从一个分子产生二个精确无误的复制品,每一个的碱基序列与原来的一模一样。

基因突变可看成是使一个特定基因的碱基对序列发生改变的某种事件的结果。许多突变或许大多数突变只不过是序列上某处的一个碱基改变成另一个碱基。但另一些突变则包含更为剧烈的变化,如序列中部分的重复或缺失,或某些其它类型的重组。一般来说,新的基因结构一旦形成,就通过常规的 DNA 复制过程,在随后的细胞分裂中保留下来。

许许多多不同的酶及其它蛋白质,是在生物细胞中合成的。它们各有其特殊的性质与功能,一起决定和控制着为物种与个体所特有的复杂的代谢和发育过程的型式。蛋白质是由一条或几条多肽链组成,多肽链又是由氨基酸通过肽键以特异的线性序列连接起来的长链构成。存在着 20 种不同的氨基酸,典型的多肽链具有长达 100—500 个氨基酸的序列,因此对于 DNA 来说,可能的结构数目是非常庞大的。此外,任一特定蛋白质的三维空间排列、特性和功能活性,最终由其多肽链组份中氨基酸的精确序列来决定。

关于蛋白质结构与 DNA 结构有关的基本观点是:一特定基因中碱基对的序列,决定了相应多肽链中氨基酸的序列。因此一个个体所能制造的所有酶和蛋白质的结构与性质,是由其基因的碱基对序列所决定的。

遗传密码的细节——即碱基序列与氨基酸序列的相互关系——通过对微生物的实验研究大致上已解决。毫无疑问,其主要特征也适用于包括人在内的高等生物。每个氨基酸是由三个碱基组成的一个序列决定的。碱基三联体连续地排列着,而且不重迭。也就是说,在决定某一氨基酸的一个三联体之后,紧接着另一个决定下一个氨基酸的三联体,如此等等。所以这两个序列是线性相关的。DNA 链四种特殊的碱基可有 64 种不同的三联体序列,其中 61 种分别决定二十种不同的氨基酸,因此一种氨基酸可为二个或更多个碱基三联体编码(见图 1.5)。还有三个所谓“无意义”的三联体,它们并不决定氨基酸,而只表示链的终止。

基因 DNA 中的碱基序列转译成多肽链中相应的氨基酸序列,这一系列过程是复杂的,其中涉及到作为中间体的某些类型的核糖核酸(RNA)分子。第一步是 DNA 的二条多核苷酸链的分离,其中之一可作为模板从可利用的核糖核苷酸合成一条互补的 RNA 链。在这一过程中,也是按照与 DNA 同样的碱基配对法则进行的,只是在 RNA 中,以尿嘧啶代替了胸腺嘧啶,并与腺嘌呤配对。这样就形成一条带有与 DNA 链相同遗传信息的 RNA 链,但它是与互补的碱基序列编码的。这条 RNA 链称为信使 RNA(或称 mRNA),然后与 DNA 分开,从细胞核到达细胞质中的核糖体上,这里是蛋白质合成的场所。附着在核糖体上的 mRNA,就作为形成多肽链的模板。氨基酸连接在另一种称为转移 RNA(或称 tRNA)分子上,来到 mRNA 模板上。tRNA 分子较小(约 80 个核苷酸)并以一系列不同的分子类型出现。每一种 tRNA 对一种特定氨基酸是特异的,氨基酸可连接在该 tRNA 分子的一端;而且每一个 tRNA 的多核苷酸序列中,都含有一个特殊的碱基三联体,与 mRNA 中决定该氨基酸编码的碱基三联体互补。由此,需要连结的氨基酸就放在

正确的位置上,从而合成由 mRNA 密码序列所决定的多肽链。

多肽链是循序制成的,从氨基端开始,每次连接上一个氨基酸。导致多肽合成的这一过程的主要特征是: 序列中每个氨基酸是由 DNA、mRNA 和 tRNA 这三类分子中的一个三核苷酸或碱基三联体所指定。mRNA 中的三联体与 DNA 中的三联体互补,同时也与 tRNA 中的三联体互补,因而虽然实际的碱基并不相同,但决定的却是同一种氨基酸。

这样,编码于基因中的遗传信息,可看成是一种蓝图,它决定每个个体制造的所有酶和其它蛋白质的结构。但是基因不仅决定蛋白质的结构,而且显然也参与蛋白质合成的调节。调节功能中,这些分子相互关系的性质,还不太清楚,并尚未提出令人满意的一般性学说,至少对多细胞生物来说是如此(见第四章)。然而,可以这样说: 每个个体的酶与蛋白质的组成,从直接的意义上讲,必然是其遗传结构的反映。而且还可这样预料: 物种中各个体间的遗传差异,无论是表现在正常的体格、生理或智力特征上的差异,还是表现在某种异常发育上的差异,都可能是酶或蛋白质合成有差异的结果。

1.2 血红蛋白变种

一个基因突变会导致合成一种改变了的蛋白质,第一个直接证据来自镰形细胞病或镰形细胞贫血症患者血红蛋白的研究。

许多年前已经知道,某些个体的红细胞具有氧分压变化时形状会发生可逆性变化的特性 (Herrick, 1910; Hahn 和 Gillespie, 1927)。氧合时,这些细胞象正常个体的红细胞一样呈双凹圆盘状。但脱氧后,就变长,呈镰刀形。红血球呈现这种所谓“镰形化现象”的人,大多数是很健康的。他们被称为具有镰形细胞性状。但有些人的“镰形化”伴随着发生严重的特殊贫血症——在童年与青春期是致命的。这种病症一般称为镰形细胞病。“镰形化现象”在非洲中部很普遍,许多地区的群体中有 20% 以上的人具有镰形细胞性状,预测其中有一定数量的人(1—2%) 早年就夭折于镰形细胞贫血症。生活于世界其它地区,例如美国的黑人中,这种病也是屡见不鲜的。但在人类的其它群体中甚为罕见,甚至是绝无仅有的。

这种特性是遗传的。1949 年, Neel 和 Beet 分别研究指出: 只要假设具有镰形细胞性状和镰形细胞贫血症的人, 分别是一条常染色体上的一个特殊的异常基因的杂合个体和纯合个体, 就可以简单地解释他们的系谱。如果个体从一个亲体得到异常基因而从另一个亲体得到正常的等位基因, 则预期这些人会具有镰形细胞性状; 而从两个亲体均得到异常基因的人, 就会出现镰形细胞贫血症。根据这一假设, 可以预测一个正常的人和一个具有镰形细胞性状的人婚配后, 平均有一半孩子是正常的, 还有一半具有镰形细胞性状。如果二个具镰形细胞性状的人婚配, 则会有三种后代: 正常孩子、具有镰形细胞性状的孩子和患镰形细胞贫血症的孩子, 他们大致上按孟德尔式比率 1:2:1 发生。进而还可推测出: 镰形细胞贫血症患者的双亲均应出现镰形化现象。家系资料 (Neel, 1951) 列出的结果, 与这些及其它预测非常一致, 以后的研究也充分证实了这一假设。

差不多与此同时, Pauling 及其同事们 (1949) 获得了决定性的发现, 他们指出: 镰形细胞贫血症患者红细胞中的血红蛋白与正常人红细胞中的血红蛋白有质的不同。他们证

明了这两种蛋白质的物理性质不同,所以可能结构上也不同,因为电泳可把两种蛋白质分离开来。他们进一步证明具镰形细胞性状的个体的红细胞,既含有正常的血红蛋白(HbA),也含有异常的镰形细胞血红蛋白(HbS)。由此看来,个体的遗传组成与所合成的血红蛋白之间有直接的相关。镰形细胞基因的纯合个体,形成 HbS; 正常等位基因的纯合个体形成 HbA; 从一个亲体得到镰形细胞基因,而从另一个亲体得到正常等位基因的杂合个体,会形成两种类型的血红蛋白。

不久,发现了另一种异常的血红蛋白,叫做 HbC (Itano 和 Neel, 1950)。从家系研究来看,它与 HbS 一样,也是由单个基因决定的。这个基因和其它的正常等位基因的杂合子的红细胞中形成两种血红蛋白 HbA 和 HbC。他们很健康,并被称之为具有血红蛋白 C 性状。这一基因的纯合子,形成 HbC 而没有 HbA, 并有温和或轻度的贫血症。这种病症称为血红蛋白 C 病,在临床上远不如镰形细胞病严重。还有第三种严重程度居中的病症,称为镰形细胞——血红蛋白 C 病,患者红细胞中既有 HbS 又有 HbC。家系研究表明,他们是决定 HbC 的基因和镰形细胞基因的杂合体,他们从一个亲体获得一个异常基因,而从另一亲体获得另一个异常基因。用电泳可方便地区别开这些彼此不同的病症。图 1.1 所示是其典型的分离。

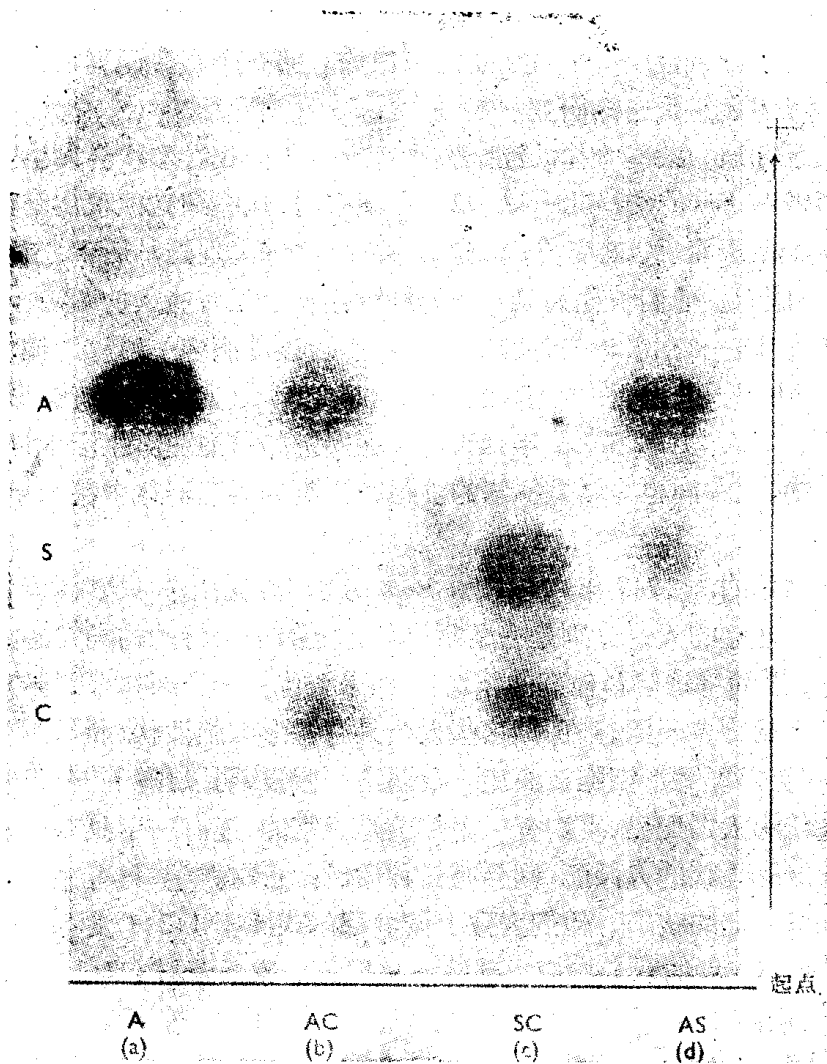


图 1.1 血红蛋白电泳(a)正常成人,(b)血红蛋白 C 性状 (c) 镰形细胞——血红蛋白 C 病和(d)镰形细胞性状。(电泳在淀粉凝胶上进行, pH8.6)

患镰形细胞——血红蛋白C病的病人中,未发现 HbA。这一事实暗示着: 决定 HbS 和 HbC 的基因是等位的,当这两个基因都存在时,形成 HbA 所必需的正常等位基因就不可能存在。有关一个亲体患镰形细胞——血红蛋白C病,而另一个亲体是正常的家系研究结果,证明这种看法是对的。其后代或是具血红蛋白C性状 (AC),或具镰形细胞性状 (AS),但没有镰形细胞——血红蛋白C病 (SC) 的患者或具有正常血红蛋白(A)者。正如后面将看到的,完全不同的证据也支持这一结论,即决定 HbS 和 HbC 的基因,事实上是等位基因。

这些异常的发现,使人们开始寻找其它的血红蛋白变种,其后几年对各种各样的不同贫血症进行了研究。在许多群体中随机选择个体,并对各个体的血红蛋白进行了广泛的电泳调查。通过这些研究,现已鉴定出人类中存在一百三十多种由遗传决定的不同的血红蛋白变种 (Lehmann 和 Carrell, 1969; Stamatoyannopoulos, 1972)。其中大多数是极为罕见的,并和正常的血红蛋白一起,以杂合状态出现。但是有些变种象镰形细胞血红蛋白,在世界某些地区有相当的发生率。例如 HbE 在东南亚许多群体中频率相当高, HbC 在西非常见,而 HbD Punjab 在印度西北部有相当高的频率。

许多具有一个异常血红蛋白的基因和正常等位基因的杂合个体,是很健康的。但也有重要的例外,那就是有些杂合状态个体也会有一种特殊的慢性溶血病或其它异常 (见 1.5)。

这些决定变异血红蛋白的不同突变基因中,有可能研究其纯合状态的只有几个。有些病例,象镰形细胞病,慢性的溶血性贫血是其标准症状。但显然也有另一些例子: 决定某种血红蛋白变种基因的纯合子,可能很健康 (例如 HbG Accra, Edington 和 Lehmann, 1954)。

慢性溶血病也可能是某些基因杂合状态的特征,如镰形细胞——血红蛋白C病,其中存在着二个不同的异常等位基因。对于某一种变异血红蛋白基因是杂合的,同时对于血红蛋白合成有特殊缺陷的地中海贫血基因(将在第四章讨论),也是杂合的个体,往往也有慢性贫血的特征。这样,以前不能明显区分的各种不同的血液异常,现在可根据血红蛋白形成时特异的异常来辨别。

1.3 异常血红蛋白的结构

随着许多遗传决定的血红蛋白变种的发现,出现了关于这些蛋白质在分子结构上与正常血红蛋白有何确切不同之处的问题。

蛋白质是由一条或几条多肽链组成,每条多肽链又是由通过肽键连结并排成一定序列的氨基酸组成。某一特定蛋白质分子结构中含二条或二条以上多肽链,它们的序列可能是相同也可能是不同的。血红蛋白A由四条多肽链组成。其中有两种不同类型,每一种具有一特异的氨基酸序列,分子中每种类型有二条。这些不同的多肽链叫做 α -和 β -链,因此血红蛋白A分子的结构为 $\alpha_2\beta_2$ 。 α -链含有 141 个氨基酸残基, β -链含有 146 个氨基酸残基,它们的精确序列已经搞清 (Braunitzer 等 1964) (见图 2.8)。

蛋白质中每条多肽链是以特异方式盘旋和折叠着,故整个分子具有复杂的三维排列。空间构型由每一多肽链所决定,故由此形成整个蛋白质分子的三维排列,可以认为主要是

由组成蛋白质的多肽链中的氨基酸序列(所谓初级结构)决定的。血红蛋白的四条链盘旋和折叠的详细方式,以及它们互相配合形成球蛋白分子的情况,已由 Perutz 及其同事们(1968)用 X 射线晶体分析加以阐明。

许多蛋白质上还有一些附加的基团,它们一般都较小,而且不是由氨基酸组成的,这些基团叫做辅基。它们在整个分子功能上起着很特殊的作用。血红蛋白上的辅基是血红素。它是一种卟啉环结构,中央有一个铁原子。血红蛋白分子中有四个血红素基团,每一个是通过铁原子和每条多肽链中特定的组氨酸残基之间的键,而与一条多肽链相连。在血红蛋白的三维模型中,血红素基团是位于相应多肽链折叠而成的四个分开的小袋中。当氧与血红蛋白结合时,就结合在血红素的铁原子上,同时整个分子的三维构象发生微妙的变化。

HbS 与 HbA 的电泳特性不同的事实暗示这两种分子在结构上是不同的,不久就证明,其差异并不在血红素基团上,而在于蛋白质本身。Ingram (1957, 1958) 阐明了这种差异的本质:他发现在 β -多肽链的氨基酸序列中一个特定位置上,在 HbA 中是一个谷氨酸残基,而在 HbS 中则是缬氨酸残基。

最初用于显示蛋白质中这种差异的方法,就是现在一般所说的“指纹法”,它已广泛应用于研究许多不同蛋白质及其遗传的变种的初级结构。蛋白质先用胰蛋白酶或胰凝乳蛋白酶这些特异的蛋白酶来消化,多肽链会在许多不同的点上断裂,形成不同的较小的肽的混合物。然后把这些肽放在滤纸上,先用电泳再行层析以进行二向的分离。由此产生的肽“斑点”的图谱,一般是该特定蛋白质所特有的。Ingram 发现,血红蛋白经胰蛋白酶(它会使赖氨酰或精氨酰肽键断裂)消化后,产生一个复杂的肽图谱(图 1.2)。由 HbA 所得的肽图谱,大多数与 HbS 是一样的。但是 HbA 中有一个肽是 HbS 图谱中没有的,而 HbS 中也有一个肽是 HbA 中所没有的。将这二个肽分离出来,并测定了其中的氨基酸序列。它们各含 8 个氨基酸残基,其序列基本一致,只是 HbA 肽中从链的氨基端算起的第六个

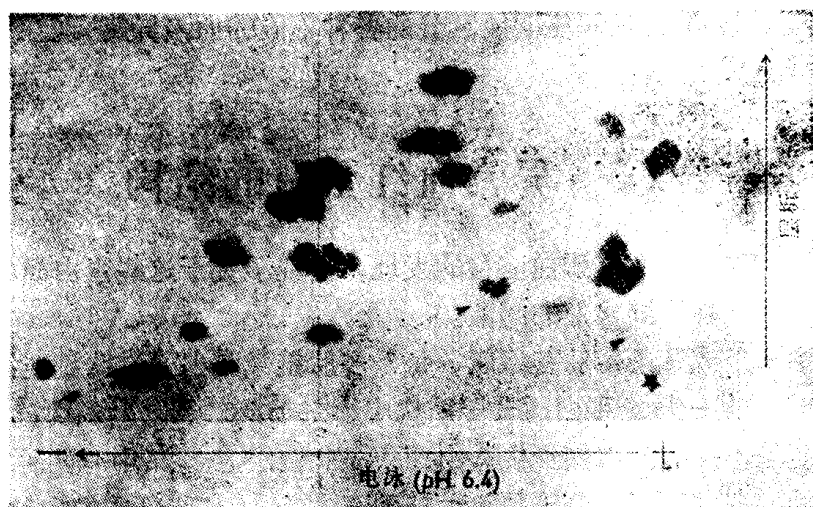


图 1.2 HbS 的“指纹”。经酶消化的血红蛋白的肽在滤纸上先用电泳分离(水平向),后进行层析(垂直向)而分离开来。然后用茚三酮显示出肽“斑点”。

- * 为肽混合物在滤纸片上的点样位置。
 - 1 HbA 中存在而 HbS 中没有的肽的位置。
 - 2 HbS 中存在而 HbA 中没有的肽的位置。
- (引自 Lehmann 和 Huntsman, 1966)

残基是谷氨酸,而在 HbS 中则是缬氨酸(图 1.3)。后来证明从这两种血红蛋白中得到的其它肽,彼此都一样。同时也发现 HbA 和 HbS 中所不同的肽,差异就在于 β -链氨基端的序列。由此得出结论:这两种血红蛋白的差异只在于 β -链氨基端算起的第 6 位上的一个氨基酸残基,这两种血红蛋白分子中 β -链其余部分的序列及整个 α -链的序列都是相同的。习惯上一条肽链上氨基酸残基位置,是从氨基端算起。因此 HbA 的结构就可写成 $\alpha_2\beta_2$,而 HbS 的结构可写成 $\alpha_2\beta_2^{6\text{ 谷}\rightarrow\text{缬}}$ 。

	1	2	3	4	5	6	7	8
HbA	Val.	His.	Leu.	Thr.	Pro.	Glu.	Glu.	Lys.↑.....
HbS	Val.	His.	Leu.	Thr.	Pro.	Val	Glu.	Lys.↑.....

图 1.3 在 HbA 的胰蛋白酶水解物中有的而 HbS 中没有的肽的氨基酸序列,和在 HbS 的胰蛋白酶水解物中有的而 HbA 中没有的肽的氨基酸序列。该序列表示这两种血红蛋白 β -多肽链中前八个氨基酸序列。它们仅在第六位上不同。箭头表示胰蛋白酶水解断裂的位点。(氨基酸缩写见图 1.5)

HbS 是第一个由遗传决定的蛋白质变种的例子,它的结构特异性已被精确证明。所得结果中值得注意的是它的简明性。单个基因的差异,如镰形细胞基因和它的正常等位基因之间的差异,大概是一次突变的结果。这是遗传变异的最小单位。它显然导致特定蛋白质初级结构中最小单位的差异,即一个氨基酸残基为另一个所取代。以后的工作证明这是一个普遍的法则。Stamatoyannopoulos (1972) 列举了一百二十多种因一个氨基酸取代,而与正常血红蛋白不同的变种类型,取代发生于 α -链或 β -链上的某一位置上(图 1.4 表示一些 β -链的取代)。在人和其它物种中,也可发现其它各种各样改变了的蛋白质同样可归因于单个基因突变。

1.4 单个氨基酸取代和遗传密码

图 1.5 列示了 mRNA 中可能存在的 64 种碱基三联体,它们与 DNA 中 64 个三联体当然是互补的,并在旁边列出了在多肽合成中由之决定的各种氨基酸。这就是所谓的遗传密码,可能所有物种都是一样的。许多氨基酸由二种不同的三联体决定,有一些甚至由四种或六种来决定。还有三种所谓“无意义”的三联体,它们在多肽链中的位置是决定多肽链合成的终止。

根据碱基三联体或密码子来测定不同血红蛋白变种的单个氨基酸取代,是具有启发性的。例如 HbA 突变为 HbC 引起了 β -链第 6 位上谷氨酸残基为赖氨酸所取代。从图 1.5 可见,编码谷氨酸的核苷酸三联体是 GAA 或 GAG,而编码赖氨酸的核苷酸三联体是 AAA 或 AAG。因此,突变只需要三联体中一个碱基的变化(即 GAA $\rightarrow$ AAA,或 GAG $\rightarrow$ AAG)。异常血红蛋白中发现的其它单个氨基酸取代也是如此。这证明了在每一例子中,单个突变只需要特定三联体中一个碱基的变化。在研究其它蛋白质和其它完全不同的物种中引起单个氨基酸取代的突变时,也可得出类似的结论。总之,现在大量的资料有力地支持了如下观点:即大多数基因突变,仅仅是一个特定基因里全部 DNA 碱基序列中一个碱基的变化。

人类血红蛋白的正常 β -多肽链含有 146 个氨基酸残基,据推测可能是由相应基因中

1 Val:	51 Pro:	101 Glu:
2 His: <u>Tyr</u>	52 Asp: <u>Asn</u>	102 Asn: <u>Thr^b, Lys</u>
3 Leu:	53 Ala:	103 Phe:
4 Thr:	54 Val:	104 Arg:
5 Pro:	55 Met:	105 Leu:
6 Glu: <u>Val, Lys, Ala</u>	56 Gly: <u>Asp</u>	106 Leu:
7 Glu: <u>Gly, Lys</u>	57 Asn:	107 Gly:
8 Lys:	58 Pro: <u>Arg</u>	108 Asn: <u>Asp^b</u>
9 Ser: <u>Cys</u>	59 Lys: <u>Glu, Thr</u>	109 Val:
10 Ala:	60 Val:	110 Leu:
11 Val:	61 Lys: <u>Asn, Glu</u>	111 Val: <u>Phe^b</u>
12 Thr:	62 Ala:	112 Cys:
13 Ala:	63 His: <u>Tyr^c, Arg^{a,b}</u>	113 Val: <u>Glu</u>
14 Leu: <u>Arg</u>	64 Gly:	114 Leu:
15 Try:	65 Lys:	115 Ala:
16 Gly: <u>Asp, Arg</u>	66 Lys: <u>Glu^a</u>	116 His:
17 Lys: <u>Glu</u>	67 Val: <u>Glu^c, Asp^{a,b}, Ala^a</u>	117 His: <u>Arg</u>
18 Val:	68 Leu:	118 Phe:
19 Asn:	69 Gly: <u>Asp^{a,b}</u>	119 Gly:
20 Val: <u>Met^b</u>	70 Ala: <u>Asp^{a,b}</u>	120 Lys: <u>Glu</u>
21 Asp:	71 Phe: <u>Ser^a</u>	121 Glu: <u>Lys, Gln</u>
22 Glu: <u>Lys, Ala, Gly</u>	72 Ser:	122 Phe:
23 Val:	73 Asp: <u>Asn^{a,b}</u>	123 Thr:
24 Gly: <u>Arg^a, Val^a</u>	74 Gly: <u>Asp^{a,b}</u>	124 Pro: <u>Arg^a</u>
25 Gly: <u>Arg^b</u>	75 Leu:	125 Pro:
26 Glu: <u>Lys</u>	76 Ala:	126 Val: <u>Glu</u>
27 Ala:	77 His: <u>Asp</u>	127 Gln:
28 Leu: <u>Pro^a</u>	78 Leu:	128 Ala:
29 Gly:	79 Asp: <u>Asn, Gly</u>	129 Ala: <u>Asp^a</u>
30 Arg: <u>Ser^a</u>	80 Asn: <u>Lys</u>	130 Tyr: <u>Asp^a</u>
31 Leu:	81 Leu:	131 Gln:
32 Leu:	82 Lys:	132 Lys: <u>Gln</u>
33 Val:	83 Gly: <u>Cys</u>	133 Val:
34 Val:	84 Thr:	134 Val:
35 Tyr: <u>Phe^a</u>	85 Phe:	135 Ala:
36 Pro:	86 Ala:	136 Gly: <u>Asp</u>
37 Try: <u>Ser^b</u>	87 Thr: <u>Lys</u>	137 Val:
38 Thr:	88 Leu: <u>Pro^a, Arg^a</u>	138 Ala:
39 Gln:	89 Ser:	139 Asn:
40 Arg:	90 Glu: <u>Lys^a</u>	140 Ala:
41 Phe:	91 Leu: <u>Pro^a</u>	141 Leu: <u>Arg^a</u>
42 Phe: <u>Ser^{a,b}, Leu^{a,b}</u>	92 His: <u>Tyr^c</u>	142 Ala:
43 Glu: <u>Ala</u>	93 Cys:	143 His:
44 Ser:	94 Asp: <u>Asn</u>	144 Lys:
45 Phe:	95 Lys: <u>Asp, Glu</u>	145 Tyr: <u>His^b, Cys^b</u>
46 Gly: <u>Glu</u>	96 Leu:	146 His: <u>Asp</u>
47 Asp: <u>Asn</u>	97 His: <u>Gln^b</u>	
48 Leu:	98 Val: <u>Met^{a,b}</u>	
49 Ser:	99 Asp: <u>His^b, Asn^b, Tyr^b</u>	
50 Thr:	100 Pro:	

图 1.4 84 种 β -链血红蛋白变种中的单个氨基酸取代。以 McKusick (1971) 和 Stamatoyannopoulos (1972) 所列的表为基础。每项的左边是正常血红蛋白 β -链氨基酸序列的总数码, 右边表示在不同位置上的各种取代。氨基酸的缩写, 列在图 1.5 中。

a) 不稳定血红蛋白; b) 改变了氧亲和力; c) 高铁血红蛋白贫血症。

一段 438 个碱基对的特异序列表达的 (146 $\times$ 3)。正常基因突变为决定 HbC 基因, 结果造成多肽链中第 6 位上的一个氨基酸的取代。所以它大概发生于相应的一段 DNA 中第六个碱基三联体上, 从密码上我们可推测: 该突变实际上就是整个序列中第十六个碱基的变化。同样, 我们也可推测: 正常基因突变为镰形细胞基因, 涉及同一个三联体中毗邻碱基的变化 (GAA $\rightarrow$ GUA, 或 GAG $\rightarrow$ GUG)。这相当于整个序列中第十七个碱基的变化。大多数其它的变异血红蛋白中氨基酸取代的本质也已确定, 从它们也可获得类似的精确推断。

如图 1.6 所示, 在遗传密码图中, 用一个箭头来表示一种已知蛋白质变种, 例如血红

		第二碱基					
		U	C	A	G		
第一碱基	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA term. UAG term.	UGU } Cys UGC } UGA term. UGG Try	U	
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	C	
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	A	
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } CAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	G	
						U	

图 1.5 遗传密码。

碱基:

U: 尿嘧啶, C: 胞嘧啶, A: 腺嘌呤, G: 鸟嘌呤。

氨基酸:

Ala: 丙氨酸	Gly: 甘氨酸	Pro: 脯氨酸
Arg: 精氨酸	His: 组氨酸	Ser: 丝氨酸
Asn: 天冬酰胺	Ile: 异亮氨酸	Thr: 苏氨酸
Asp: 天冬氨酸	Leu: 亮氨酸	Try: 色氨酸
Cys: 胱氨酸	Lys: 赖氨酸	Tyr: 酪氨酸
Gln: 谷氨酰胺	Met: 蛋氨酸	Val: 缬氨酸
Glu: 谷氨酸	Phe: 苯丙氨酸	

term= “无意义”三联体-链的终止。

要详细查阅有关密码由来的参考文献,可见 Cold Spring Harbor Symp. Quant. Biol. 31 (1966); Crick (1967); Woese (1967) 和 Yčas (1969)。

蛋白变种中各种单个氨基酸的取代情况,是很方便的 (Crick, 1967)。如果取代是由于在三联体的第一位上一个碱基变化,则表示变化的箭头是垂直方向的,并开始与结束在图中方格内同样的相对位置上。如果变化发生在第二碱基,则箭头是水平方向的。如果变化是在第三个碱基,则箭头呈垂直方向,但开始与结束均在同一方格内。因而如果一个氨基酸取代最低限度只需要密码三联体中一个碱基变化,箭头或呈水平向或呈垂直向,在三种情况下看来均是如此。对角线向的箭头暗示至少有二个碱基发生变化,如果氨基酸的变化是非常任意与随机的,预期可在相当多的例子中看到这种情况。

可以看到大多数箭头集中在图上特定区域。这种现象是由于发现大多数变异血红蛋白时所采用的方法而引起的。之所以能发现它们,主要是由于电泳性质上与正常血红蛋白 A 有差异,因而氨基酸的变化可能也涉及电荷的变化。图中带电氨基酸的部分是用深色表示,而其余部分为中性氨基酸。从发现的方式可以预期,大部分箭头是开始或结束于深色部分 (84 种中的 61 种)。寻找血红蛋白变种时,广泛使用电泳,这主要是由于技术上的方便。由此可以预料:一定还有许多其它的血血红蛋白变种,因不发生电荷变化,而尚未发现。表示这类突变的箭头,大概会出现在图 1.6 中的空白部分。随着寻找变种时其它技