



## [12] 发明专利申请公开说明书

[21]申请号 94193870.0

[51]Int.Cl<sup>6</sup>

A61K 39/00

[43]公开日 1996年10月16日

[22]申请日 94.10.20

[30]优先权

[32]93.10.21 [33]US [31]08 / 140,797

[86]国际申请 PCT / US94 / 12028 94.10.20

[87]国际公布 WO95 / 11040 英 95.4.27

[85]进入国家阶段日期 96.4.22

[71]申请人 纽约血库公司

地址 美国纽约

[72]发明人 A·O·波戈

A·查德利

[74]专利代理机构 中国国际贸易促进委员会专利商  
标事务所

代理人 杜京英

A61K 39 / 015 A61K 45 / 00

C07K 14 / 435 C07K 14 / 47

C07K 14 / 705 C12N 15 / 12

权利要求书 2 页 说明书 35 页 附图页数 7 页

[54]发明名称 DUFFY血型抗原的克隆

[57]摘要

分离了 gpD 蛋白,即 Duffy 血型抗原系统的主要亚单位。gpD 蛋白含有间日疟原虫进入红细胞并引起疟疾的受体。gpD 和人及兔的白细胞介素-8 受体有明显的序列同源性,因此 gpD 蛋白有可能是一类新的趋化细胞因子受体。gpD 蛋白 cDNA 和人海马 cDNA 克隆 HHCMF86 有准完全同源性,因此, gpD 蛋白质或一同源蛋白可能以一神经肽受体存在于脑中。gpD 蛋白存在于所有的红细胞祖细胞中,可能是细胞增殖和 / 或分化的受体。gpD 蛋白 cDNA 在人肾中识别大小和骨髓中一样的 mRNA。由于肾不是促红细胞生成器官,也没有成为促红细胞生成器官的潜能,这种推定的趋化物受体可能具有必需的肾功能。gpD 蛋白在预防疟疾和调节促红细胞生成、神经和肾功能上有治疗价值并可以和生理可接受的稀释剂结合产生适于这些目的的治疗剂。同 gpD 蛋白含有受体的部分相应的多肽也已被合成。此类多肽有与 gpD 蛋白相同的治疗作用。gpD 蛋白和这些多肽还可以用于产生治疗剂如抗体、互补多肽等,它们对治疗疟疾及调节必需的促红细胞生成、神经和肾功能也是有用的。

# 权 利 要 求 书

---

1. 分离或合成的编码 gpD 蛋白的 DNA。
2. 根据权利要求 1 的分离或合成的 DNA, 其具有 SEQ ID No: 1 所示的核苷酸序列。
3. 分离或合成的未变性的 gpD 蛋白。
4. 含有 SEQ ID No:14 中所示的氨基酸序列的肽。
5. 根据权利要求 4 的肽, 其含有 SEQ ID No:8 所示的氨基酸序列。
6. 根据权利要求 4 的肽, 其含有 SEQ ID No:9 所示的氨基酸序列。
7. 根据权利要求 4 的肽, 其含有 SEQ ID No:10 所示的氨基酸序列。
8. 根据权利要求 4 的肽, 其含有 SEQ ID No:14 所示的氨基酸序列。
9. 用于保护温血动物抵抗疟疾感染的疫苗, 其包含对此有效量的权利要求 3 的分离或合成 gpD 蛋白。
10. 用于保护温血动物抵抗疟疾感染的疫苗, 其包含与生理可接受稀释剂混合的对此有效量的权利要求 4 的肽。
11. 根据权利要求 10 的疫苗, 其中该肽具有 SEQ ID No:8 所示的氨基酸序列。
12. 根据权利要求 10 的疫苗, 其中该肽具有 SEQ ID No:9 所示的氨基酸序列。

13. 根据权利要求 10 的疫苗, 其中该肽具有 SEQ ID No:10 所示的氨基酸序列。

14. 根据权利要求 10 的疫苗, 其中该肽具有 SEQ ID No:14 所示的氨基酸序列。

15. 保护温血动物免于疟疾感染的方法, 包括给予所述动物对此有效量的权利要求 3 的分离或合成 gpD 蛋白。

16. 保护温血动物免于疟疾感染的方法, 包括给予所述动物对此有效量的权利要求 4 的肽。

17. 根据权利要求 16 的方法, 其中该肽具有 SEQ ID No:8 所示的氨基酸序列。

18. 根据权利要求 16 的方法, 其中该肽具有 SEQ ID No:9 所示的氨基酸序列。

19. 根据权利要求 16 的方法, 其中该肽具有 SEQ ID No:10 所示的氨基酸序列。

20. 根据权利要求 16 的方法, 其中该肽具有 SEQ ID No:14 所示的氨基酸序列。

# 说 明 书

## DUFFY 血型抗原的克隆

本发明涉及 Duffy 血型抗原的主要亚单位 gpD 蛋白及其在检测和治疗疟疾中的应用。

疟疾是人类最流行的传染疾病。其广泛的地理分布及该传染严重的病理结果使疟疾成为许多发展中国家主要的医疗和经济负担。

有几种不同类型的疟疾,其中一种由寄生虫间日疟原虫(*Plasmodium vivax*)引起,它能攻击敏感个人的血红细胞。对于间日疟原虫特别令人感兴趣的一个遗传特性就是缺少由血型系统编码被称为 Duffy(F. B. Livingston, "The Duffy Blood Groups, Vivax Malaria and Malaria Sections in Human Populations: Review," *Human Biol*, 56, 413, (1984))的抗原。已经显示血红细胞缺少 Duffy 基因产物的人对间日疟原虫的侵入是不敏感的,这是由于 Duffy 分子能作为该寄生虫的受体。(L. H. Miller, H. J. Mason, D. F. Clyde and M. H. Mc Ginnis, "The Resistanse Factor to Plasmodium Vivax in Blacks, The Duffy Blood Group Genotype (a - b -)", *N. Eng. J. Med*, 295, 302 (1976))。

疟疾寄生虫是由按蚊属的几种吸血雌性虫进行宿主至宿主间传播的。间日疟原虫生活周期的性阶段发生在蚊子中,导致孢子的产生。这些孢子进入新的宿主后寄居于肝脏的实质细胞中并进行无性增殖,最终导致肝细胞的破裂并把无性形式(裂殖子)释放到血液中。在那里裂殖子以一种几乎同步的方式活跃地进入血红细胞,

并由于间日疟原虫细胞生长和分裂的速率基本一致,被感染的红细胞同时达到使它们破裂的寄生虫滴度阶段。这就产生每 48 小时发热的典型周期,由此得名为间日疟疾。

间日疟原虫感染在不治疗的情况下可持续长达五年的时间。间日疟原虫寄生物血症相对低,主要因为寄生虫喜欢存在于外周血液中的少数幼红细胞或网织红细胞中。

对间日疟原虫的免疫性本质上通常只是部分的,这就使得发生重复感染,并独立发展致使寄生虫释放的周期发生重叠,导致在较短的周期中出现发热。间日疟原虫和其它疟疾疟原虫一样(M. Hommel, "Antigenic Variation in Malaria Parasites," *Immunology Today*, 6, 28, (1985)), 显示相当的抗原多样性和变异性,虽然最近显示存在寄生虫不同分离物共有的间日疟原虫子孢子抗原组份(F. Zavala, A. Masuda, P. M. Graves, V. Nussenzweig and R. Nussenzweig, "Ubiquity of the Repetitive Epitope of the Cs Protein in Different Isolates of Human Malaria Parasites," *J. Immunol.*, 135, 2790, (1985))。

考虑到间日疟原虫分离物间抗原性差别的来源及它们对于免疫接种的后果,间日疟原虫不同种的裂殖子具有进入血红细胞的相同受体即 Duffy 分子(Miller et al, *N. Engl. J. Med.*, supra)是很重要的。此外,不管其改变其它抗原分子的能力,寄生虫识别分子,即结合 Duffy 分子的分子必须保持一致,因为正是它和恒定受体间的互补性使得裂殖子进入红细胞,由此导致感染的继续。这种分子的配体专一性的变化引起寄生虫感染能力的丧失,因为间日疟原虫裂殖子好象不能利用人其它的血红细胞受体进入体内,这可以由 Duffy 阴性红细胞的抗性看出。

Duffy 血型系统含有两种主要的抗原  $Fy^a$  和  $Fy^b$ , 分别由  $Fy^a$  和  $Fy^b$  等位基因产生。抗血清抗  $-Fy^a$  和抗  $Fy^{-b}$  定义了四种表现型,  $Fy(a+b-)$ ,  $Fy(a-b+)$ ,  $Fy(a+b+)$  和  $Fy(a-b-)$ 。W. L. Marsh, Crit. Rev. Clin. Lab. Sci., 5, 387, (1975)。没有一种抗血清能聚合 Duffy· $Fy(a-b-)$  细胞, 此细胞类型为黑人中主要的表现型。定义为其它表现型的抗血清,  $Fy_3$ ,  $Fy_4$  和  $Fy_5$  非常稀少。一鼠的单克隆抗体, 抗  $-Fy_6$ , 定义了存在于所有 Duffy 阳性细胞中的新的 Duffy 抗原决定簇, 但它不存于  $Fy(a-b-)$  细胞中。M. E. Nichols, P. Rubinstein, J. Barnwell, S. R. de Cordoba, 和 R. E. Rosenfield, J. Exp. Med., 166, 776(1987)。具有  $Fy(a-b-)$  红细胞的黑人不被间日疟原虫感染。这些细胞对诺尔斯氏疟原虫 (*P. knowlesi*) 的侵入也具有抗性。诺尔斯氏疟原虫是一种入侵  $Fy(a+b-)$  和  $Fy(a-b+)$  人红细胞的猴寄生虫。L. H. Miller, S. J. Mason, J. A. Dvorak, M. H. McGinniss 和 K. I. Rothman, Science, 189, 561(1985)。因此这些寄生虫入侵红细胞的受体与 Duffy 血型系统相关。

如下的氨基酸在别处描述中可能由以下 3 或 1 个字母的代码表示。

<u>氨基酸</u>	<u>3-字母代码</u>	<u>1-字母代码</u>
丙氨酸	Ala	A
精氨酸	Arg	R
天冬酰胺	Asn	N
天冬氨酸	Asp	D
半胱氨酸	Cys	C
谷氨酰胺	Gln	Q

谷氨酸	Glu	E
甘氨酸	Gly	G
组氨酸	His	H
异亮氨酸	Ile	I
亮氨酸	Leu	L
赖氨酸	Lys	K
甲硫氨酸	Met	M
苯丙氨酸	Phe	F
脯氨酸	Pro	P
丝氨酸	Ser	S
苏氨酸	Thr	T
色氨酸	Trp	W
酪氨酸	Tyr	Y
缬氨酸	Val	V

利用一抗 Fy6 单克隆抗体, 现已发展了一种从人红细胞中纯化 Duffy 抗原的程序。Duffy 抗原看来是由不同亚单位组成的多体红细胞膜蛋白。一种命名为 gpD 分子量为 35 - 45kDa 的糖蛋白是此蛋白复合物的主要亚基, 并且具有由抗 Fy<sup>a</sup>、抗 Fy<sup>b</sup> 和抗 - Fy6 抗体确定的抗原决定簇。在分子水平对这一新的蛋白质的描述对发现其在红细胞膜上的功能、了解寄生虫—红细胞识别的过程和解开寄生虫入侵的分子机制是重要的。本发明涉及一编码 gpD 蛋白质的 mRNA 的分离、序列分析及组织表达。

gpD 蛋白和人及兔的白细胞介素 - 8 受体有显著的序列同源

性,因此 gpD 蛋白极有可能是一类新的趋化细胞因子受体。gpD 蛋白 cDNA 和一人海马 cDNA 克隆 HHCMF86 有准全同源性,因此很有可能 gpD 蛋白或一同源蛋白以一种神经肽受体存在于大脑中。gpD 存在于所有的红细胞祖细胞中,并存在作为细胞增殖和/或分化受体的可能性。gpD 蛋白 cDNA 能识别人肾中和骨髓中同样大小的 mRNA。由于肾不是促细胞生成器官也没有成为促细胞生成器官的潜能,所以有可能这一推定的趋化物受体具有必需的肾功能。

gpD 蛋白在预防疟疾和调节必须的红细胞、神经和肾功能中有医疗价值,并可以和生理可接受的稀释剂结合产生适合于这些目的的治疗剂。

与 gpD 蛋白中含受体的部分相应的多肽也已被合成。这些多肽具有和 gpD 蛋白本身相同的治疗应用效果,并且和 gpD 蛋白情况一样,合成的多肽可以和生理可接受的稀释剂结合产生抗疟疾的疫苗或对调节必须的红细胞、神经和肾功能有用的治疗剂。

gpD 蛋白和相应于 gpD 蛋白一部分的合成肽在生产治疗剂如抗体、互补多肽及以 gpD 蛋白或合成多肽的三级结构为模型的药物中也有用处,这些蛋白和肽在治疗疟疾、调节必须的红细胞、神经和肾功能中也有治疗价值。

现在将参照附图详细描述本发明,其中:

图 1a 是两个最长的 gpD 蛋白 cDNA 克隆的图解和部分限制性内切酶图谱。图 1b 是编码 gpD 蛋白的组分的 Fyb71-81 cDNA 克隆的核苷酸和氨基酸序列。

图 2a 是 gpD 蛋白序列的亲水性图。图 2b 是一个 gpD 蛋白膜



定向的推定模型。

图 3 是以 Fyb71 或 Fyb81 插入片段为探针的 Northern(图 3a) 和 Southern 印迹(图 3b)。

图 4 是以 Fyb81 克隆的插入片段为探针对从人组织获得的 poly(A)<sup>+</sup> RNA 的 Northern 印迹分析。

图 5 是显示 gpD 蛋白和海马 cDNA 克隆 HHCMF86 间 DNA 序列同源性的图表。

已分离到编码 Duffy 血型抗原系统主要亚单位 gpD 蛋白的四个 cDNA 克隆。已从这四个 cDNA 克隆测定了编码 gpD 蛋白的结构基因的核苷酸序列。序列显示于图 1 中并被指定为 SEQ ID No:1。由于遗传密码的简并性,可能存在其它编码相同 gpD 蛋白的天然 DNA 序列。因此本发明扩延到这样的其它天然 DNA 序列和具有 SEQ ID No:1 中显示的相同 DNA 序列的合成 DNA 序列及其它这样的天然 DNA 序列。可由任何常规的方法合成 DNA。

图 1 也显示了 gpD 蛋白的氨基酸序列。如关于结构基因 DNA 序列的情况一样,本发明扩展到从天然来源分离的 gpD 蛋白或由化学合成制备的 gpD。可由任何常规方法进行化学合成。

图 1 中,氨基酸残基编号于左侧;核苷酸位置编号于右侧。相应于预测氨基酸的多肽位置由单实线显示。两个潜在的糖结合位点的天冬酰胺残基由向上箭头显示。第三个糖基化位点,第 37 位的天冬酰胺,由于其后是天冬氨酸而不可能存在。参见 R. D. Marshall, *Annu. Rev. Biochem.*, 41, 673(1972)。位于 5'末端下的双线是被用来引物延伸 5'末端的序列,而在 3'末端是添加 poly(A)的共有序列。

gpD 蛋白质是一高度疏水的膜内糖蛋白,有九个推定的穿膜  $\alpha$

-螺旋。相关基因存在于 Duffy 阳性和阴性个体中,但 Duffy 阴性个体的骨髓不合成 gpD 专一性 mRNA。在成人肾、脾和胎肝中,这种 mRNA 和 gpD mRNA 大小一样;但在脑中 mRNA 要长的多。已被表征的克隆将提供待研究的元素:(i)gpD 基因的结构组份,(ii)gpD 蛋白在人骨髓和其它组织中的生物合成和表达,(iii)这种可能存在于其它细胞类型中并可能作为趋化因子受体的新的红细胞膜蛋白的结构—功能,及(iv)gpD 蛋白作为间日疟原虫裂殖子入侵受体的作用。

P. Rubinstein 等的美国专利 5, 101, 017, 已公开了一种针对于 gpD 蛋白的单克隆抗体(以下称为“Rubinstein 抗体”)。Rubinstein 抗体通过有效地阻断间日疟原虫疟疾寄生虫的靶分子从而阻止间日疟原虫疟疾寄生虫侵入人红细胞。有可能 Rubinstein 抗体有一个和间日疟原虫上配体位点立体化学相同的结合位点,并引发抗独特型抗体与寄生虫反应。由于这些性能,Rubinstein 抗体可用于如间日疟原虫感染的免疫诊断、诱导抗独特型反应(如上所述,这产生对抗这些寄生虫的保护作用),和在体内直接阻断寄生虫的红细胞受体。Rubinstein 抗体的这些和其它应用的细节在该专利中已述,其整个内容并入本文作为参考。

Rubinstein 抗体基本上按照 G. Kohler 和 C. Milstein (Nature, 256, 495(1975))的方法制备。以 Fy(a + b +)型人红细胞免疫小鼠,取出脾并与鼠骨髓瘤细胞系 P3/NSO - Ag4 - 1(NS - O)(从 ATCC 获得(American Type Culture Collection, Rockville, Maryland))的悬浮液杂交,测定杂交瘤分泌结合人红细胞的抗体。其中一孔发现含结合 Fy(a + b -)和 Fy(a - b +)型红细胞的抗体,但不能结合 Fy(a

-b-)型的红细胞。收集该孔中的细胞物,稀释并克隆。发现其中一个克隆能分泌 Rubinstein 抗体。

如上所述, Rubinstein 抗体对 gpD 蛋白是专一性的,现已被克隆。因此,本发明的 gpD 蛋白对制备与 Rubinstein 抗体有相同专一性的单克隆抗体也是有用的。制备这些抗体的方法基本是 Rubinstein 抗体采用的相同方法,只是以 gpD 蛋白免疫而不是以人红细胞免疫鼠。

此外, gpD 蛋白的 N-末端(胞外)区域已被确认参予疟疾寄生虫和红细胞的相互作用。目前正在进行以合成多肽确认参予相互作用的精确氨基酸残基的工作。以下的多肽已发现在 ELISA 测试中与 Rubinstein 抗体结合:

(1) MASSGYVLQAELSPSTENSSQLDFEDVWNSSYGVNDSFPDGDYDAN  
LEAAAPCHSCNLLDDSA LPF, 其被称为 SEQ ID No: 8;

(2) MASSGYVLQAELSPSTENSSQLDFEDVWNSSYGVNDSFPDGDYD,  
其被称为 SEQ ID No: 9; 和

(3) AELSPSTENSSQLDFEDVWNSSYGVNDSFPDGDYD, 其被称为  
SEQ ID No: 10.

以下多肽在 ELISA 测试中与 Rubinstein 抗体结合:

(4) DFEDVWNSSYGVNDSFPDGDYD, 其被称为 SEQ  
ID No: 11;

(5) ANLEAAAPCHSCNLLDDSAALPF, 其被称为 SEQ ID No: 12; 和

(6) AELSPSTENSSQL, 其被称为 SEQ ID No: 13.

多肽(3)结合 Rusinstein 抗体而多肽(4)和(6)不结合这一事实暗示多肽(6)的 C-末端和多肽(4)N-末端间的连接对结合是重要的。被指定为 SEQ ID No: 14 的氨基酸序列 AELSPSTENSSQLDFEDVWNSS, 可能含有 Rubinstein 抗体的表位。因此本发明也扩展至含 SEQ ID No: 14 中描述的氨基酸序列的多肽。

指定为 SEQ ID No. 8、SEQ ID No. 9、SEQ ID No. 10 和 SEQ ID No. 14 的肽以及含 SEQ ID No. 14 的肽在体内与寄生虫结合, 因此可用作抗疟疾疫苗的免疫原。因而本发明还涉及这种疫苗以及保护温血动物(特别是人)抗间日疟原虫感染的方法, 该方法包括给予这种动物有效量的所述免疫原。因为合成肽一般讲免疫原性差, 最好将肽与一载体偶联, 例如蛋白载体如破伤风类毒素、钥孔血蓝素(KLH)等, 或脂质载体, 如 R. Neurath 等在美国专利 4, 847, 080 和 5, 204, 096(其公开内容并入本文作为参考)或 T. Hoppe 在美国专利 5, 019, 383(其公开内容并入本文作为参考)中介绍的那些。

本发明的多肽可从天然来源中形成, 例如通过 gpD 蛋白质的蛋白酶裂解, 或通过化学合成。多肽的化学合成及多肽与载体的连接的细节可以在美国专利 4, 847, 080 和 5, 204, 096 中找到。

在本发明的疫苗中,本发明的多肽通常和一种生理可接受的稀释剂(介质)如含一种佐剂的磷酸缓冲盐水一起存在。一般说来,生理可接受稀释剂中多肽的量为每剂量大约1微克到1毫克。

利用 gpD 蛋白以相似量在相似的稀释剂中也可以形成合适的疫苗,虽然对大多数应用来说使用合成多肽更实际一些。

在每一种情况下,本发明的疫苗可通过皮下、静脉、皮内或肌肉内注射使用。而优选的途径依赖特异的疫苗,肌肉内注射一般将更合适。使用的频率将根据疫苗而异。

已经知道在所有疟原虫疟疾蛋白中确认的结合区域都是同源的,如 J. H. Adams 等, Proc. Natl. Acad. Sci. USA, 89, 7085(1992)。因此 gpD 蛋白和本发明的多肽结合并妨碍所有疟原虫与红细胞的结合。这就意味着本发明的疫苗一般对抵抗所有的疟疾都有作用,因此本发明扩展到一种保护温血动物抵抗由任何疟原虫所引起的疟疾的方法。

gpD 蛋白和红细胞上的白细胞介素-8受体有显著的同源性。这与最近的一篇报道相一致,该报道提出 Duffy 血型抗原和红细胞趋化因子受体是同一种蛋白。R. Horuk 等, "A Receptor for the Malaria Parasite Plasmodium vivax: The Erythrocyte Chemokine Receptor", Science, 261, 1182(1993)。红细胞受体明显不同于中性白细胞上的 IL-8 受体、IL-8RA 和 IL-8RB。红细胞受体结合一个家族的趋化和炎症前可溶性多肽,包括 IL-8、黑素瘤生长刺激活性(MGSA)、单核细胞趋化蛋白 1(MCP-1)及激活时调节的正常 T 表达和分泌蛋白(RANTES)。引入 gpD 蛋白(或本发明的合成多肽)干扰这些蛋白与红细胞受体的正常结合,因此对调节分泌这些蛋白

的生理学效应有用。例如已经假定红细胞受体作为某些炎症介质(包括 IL-8)的净化剂。gpD 蛋白的引入(或本发明的合成多肽)因此将加强 IL-8 的清除,由此减轻任何 IL-8 引起的炎症。为此目的,上述的本发明疫苗适于作为一种治疗剂。

gpD 蛋白与一人海马 cDNA 克隆 HHCMF86 也显示出显著的同源性,因此极有可能 gpD 蛋白或另一同源蛋白以一种神经肽受体存在于脑中。gpD 蛋白在人肾中识别一和骨髓中同样大小的 mRNA,因此, gpD 蛋白或一种同源蛋白在调节肾功能中起某种作用。发明的治疗剂将相应地用于这些神经和肾功能的调节上。

与 gpD 蛋白或本发明的合成多肽互补的蛋白质如 gpD 特异性抗体,将阻断天然受体,因此也将具有上述的治疗作用。在这些互补蛋白质的制备中, gpD 蛋白或本发明的合成多肽的使用有明显的价值。

本发明将根据下述非限制性实施例作进一步描述。

#### 实施例 1:gpD 蛋白的部分氨基酸序列分析

红细胞(Fy(a-b+))用冷的磷酸缓冲盐溶液(PBS)(pH7.4)洗三次,重新悬浮于同样溶液中,与浓度为每 ml 压紧的红细胞 10 微克的 Rubinstein 抗体 4℃ 连续混合过夜。(此浓度以放射碘标记抗体确定,超过这一浓度需要饱和 Duffy 抗原位点)。未结合的抗体以冷 PBS 洗涤红细胞而去除。以 20 倍体积含 1mM 苯甲基磺酰氟和每毫升 100 激肽释放酶失活单位 Trasylol™(抑肽酶)的冷 5mM 磷酸钠缓冲液(pH7.4)通过低渗裂解制备红细胞影。然后彻底洗红细胞影直至其变为淡粉红色。以 43,000xg 离心 30 分钟收集红细胞影,

倒掉上清,沉淀悬于 50mM Hepes - NaOH, pH8.0, 1mM 苯甲磺酰氟, 每毫升 100 个激肽释放失活单位的 Trasylol 中, 在  $-20^{\circ}\text{C}$  冰冻。

冰冻的红细胞影随后融化, 43,000xg 离心 30 分钟。沉淀重新悬浮于 50mM Hepes - NaOH, pH8.0, 1mM 苯甲磺酰氟、每毫升 100 个激肽释放酶失活单位 Trasylol<sup>TM</sup> 中, 体积为压紧红细胞最初体积的 3 倍。加入 Triton X - 100<sup>TM</sup> (无过氧化物) 去污剂至最终浓度为 1%, 溶液在室温温和混合 1 小时。43,000xg 离心 30 分钟去除外壳。上清液在氮压下以 PMY10 滤膜 (Amicon Corp.) 通过 Amicon 浓缩器浓缩 10 倍。

加 0.1 倍体积、10 倍于正常浓度的 PBS 溶液于去污剂提取液中。去污剂提取物然后和偶联了抗-鼠 IgG 的 Sepharose 4B<sup>TM</sup> 珠在室温下温育 1 小时。Sepharose 4B<sup>TM</sup> 珠与去污剂提取物的比例为 1:100 (v/v)。通过离心去除抗-鼠 IgG Sepharose 珠, 用含有 PBS 和 0.5% Triton X - 100 的溶液洗涤, 珠与洗涤溶液的比例为 1:20 (v/v)。洗涤在室温下进行, 重复三次。通过把珠子在含有 62.5mM Tris - HCl (pH6.8), 0.5% 十二烷基硫酸钠 (SDS) 的溶液中温育进行洗脱, 珠和洗脱溶液的比例为 1:2 (v/v)。温育在  $65^{\circ}\text{C}$  进行 10 分钟并重复三次。洗脱物用 PMY10<sup>TM</sup> 滤膜在 Amicon 浓缩器中在氮压下浓缩。

根据 U. K. Laemmli, Nature, 227, 680 (1970), 在 0.1% SDS 存在的条件下进行聚丙烯酰胺凝胶电泳, 但作以下修改: 丙烯酰胺的浓度为 10%, 聚合反应过夜破坏氧化剂, 在上面槽中加入 0.1mM 的巯基乙酸盐。向亲和纯化物的浓缩溶液中加入以下化合物: 尿素至 4M, SDS 至 2%, 和  $\beta$ -巯基乙醇至 5%。电泳后凝胶在 10% 异戊醇和

5%醋酸中固定30分钟,并用0.002%考马斯亮蓝R-250染色直至看见标准蛋白带。相应于34-46kDa间和96kDa以上的区域被切下,用5%醋酸脱色几次,并用双蒸水洗涤。胶片于-20℃保存或立即使用。

切成4×4mm立方体的胶片放入一Elutrap™装置(Schleicher and Schuell)的洗脱室中,用50mM碳酸氢铵、0.1% SDS溶液在100伏(稳压)洗脱过夜。加入新鲜的50mM碳酸氢铵、0.1% SDS溶液,再继续电洗脱6—8小时。洗脱物用Centricon™微浓缩器(Amicon Corp.)浓缩。

纯化的蛋白质按如下被烷基化并用溴化氰(CNBr)裂解:纯化的蛋白质用冷丙酮在1mM HCl存在条件下-20℃沉淀两小时。沉淀用100%冷丙酮洗涤,室温蒸发至干并溶于加有0.5% SDS的0.1M Tris-HCl (pH8.0)中。向溶液中加入固体DTT使其终浓度为10mg/ml,溶液在85℃还原两小时。向溶液中加入十分之一体积的2.68M碘乙酰胺,试管用氮气吹洗,于室温下在黑暗中温育30分钟。温育后,加入固体DTT至10mg/ml,对加有0.5% SDS的0.1M Tris-HCl (pH8.0)透析过夜。蛋白用丙酮按如上述沉淀并风干。将沉淀溶于96μl 70%甲酸和4μl溶于70%甲酸的1M CNBr溶液中,黑暗中在室温温育48小时。将酸蒸发干,用水洗涤沉淀,蒸发至干、反复几次。消化的蛋白或进行高压液相(HPLC)或进行聚丙烯酰胺凝胶分级分离。

通过利用邻苯二甲醛(OPA)阻断试剂(见, A. W. Brauer 等, Anal Biochem., 137, 134(1983))对未分级分离的CNBr消化产物测序而获得Pe1 (SEQ ID No:2)多肽。Pe5 (SEQ ID No:6)多肽是从



经过三层 SDS—PAGE 系统 (参见, H. Shagger et al., *Anal. Biochem.*, 168, 368(1987)) 的 CNBr 消化产物中分离得非常好的唯一片段 (~4KDa) 的部分序列。跑完电泳后, 将此肽片段电印迹至 ProBlott™ (Applied Biosystems) 并测序 (见, N. LeGendre et al., *A Practical Guide to Protein and Peptide Purification for Microsequencing*, P. T. Matsudaira, ed., (Academic Press, New York), 49~69(1989))。另一份用胃蛋白酶 (50/1 比率) 37℃ 消化过夜, 应用 Vydac C-18 柱经反相 HPLC 分离这些片段。Pe2 (SEQ ID No:3), Pe3 (SEQ ID No:4), Pe4 (SEQ ID No:5) 和 Pe6 (SEQ ID No:7) 肽作为单峰从反相 HPLC 柱上洗脱, 为经反相 HPLC 产生的少数几种胃蛋白酶肽, 对它们进行测序。根据厂商的说明使用 Applied Biosystems Protein/Peptide Sequencer™ 470 或 477 型。以 100/1 比率, 4℃ 30 分钟或 60 分钟进行胃蛋白酶消化没有产生大的肽。

#### 实施例 2: 引物设计和聚合酶链式反应 (PCR)

Pe5 (SEQ ID No:6) 为最有希望来制备用以筛选 gpD 蛋白克隆的探针。Pe2 (SEQ ID No:3), Pe3 (SEQ ID No:4), Pe4 (SEQ ID No:5) 和 Pe6 (SEQ ID No:7) 肽太短不适于 PCR 扩增, Pe1 (SEQ ID No:2) 肽较大, 但它有三个不明确的残基。

引物 (每个含 23 个碱基) 的核苷酸序列由 Pe5 (SEQ ID No:6) (见图 1b) 的 N-末端和 C-末端氨基酸序列推演而来。因为 Pe5 (SEQ ID No:6) 肽是通过 CNBr 裂解而产生的, 所以 N-末端包含一甲硫氨酸从而将肽的长度增加至 24 个残基。按照 R. Lathe 在 *J. Mol. Biol.*, 183, 111(1985) 中描述的密码子优先原则选择碱基, 该文

献内容并入本文作为参考;在除 3'端外简并超过三倍的位置引入脱氧肌苷(I)。

化学合成了两个的针对 N-末端氨基酸(引物 A)和 C-末端氨基酸(引物 B)的引物,并用于从混合的 Fy(a-b+)个体的人骨髓 mRNA 中扩增 Pe5(SEQ ID No:6)的编码序列。引物 A(正义的)针对于残基 245 至 252(见图 1b),并由 12 重简并度的 5'-ATGAAX-ATYXTITGGGCITGGTT(其中 I=脱氧肌苷;X=C 或 T;和 Y=C, T 或 A)组成。引物 B(反义的)针对于残基 261—268(见图 1b),并由 32 重简并度的 5'-ACIAGMAAMTCIAGICCIAMNAC(其中 I=脱氧肌苷;M=A 或 G;和 N=G, A, T, 或 C)组成。

应用来自 BRL(Bethesda, Maryland)的预扩增试剂盒并以寡聚 dT 作为引物从 Fy(a-b+)表型的 mRNA 中合成 cDNA 的第一条链。为了酶法扩增,cDNA、引物 A、引物 B 和 Taq 聚合酶(Stratagene)被孵育于一 Perkin-Elmer DNA 热循环仪上。将预期大小(72bp)的扩增产物亚克隆到 pBluescript-SK 载体(Stratagene)上。插入片段的推演氨基酸序列与 Pe5(SEQ ID No:6)肽(见图 1b)的序列匹配。根据肽 Pe5(SEQ ID No:6)的序列 WFIFWWPH,化学合成了寡聚核苷酸 TGGTTTATTTTCTGGTGGCCTCAT(SEQ ID No:16),通过 T<sub>4</sub> 多聚核苷酸激酶(New England Biolabs)在 5'末端用 <sup>32</sup>P 标记,并用作标针以筛选人骨髓 cDNA 文库(见下文)。该具有针对氨基酸 251 至 258 的密码习惯的 24 碱基寡核苷酸成功地鉴定出了真正的 gpD 蛋白 cDNA 克隆。

实施例 3:人 mRNA 和 DNA 的分离

Poly(A)<sup>+</sup> RNA 提取如下:洗涤人骨髓抽吸物,并在 5%β-巯基乙醇加 6M 硫氰酸胍,25mM 柠檬酸钠(pH7.0),50mM 乙二胺四乙酸(EDTA)的溶液中裂解细胞,使胍终浓度为 5M。溶液通过一 25G 皮下注射针头以剪切 DNA,并加入 2%的 Sarkosyl。溶液在一 5.7M CsCl,50mM EDTA(pH7.0)垫层上在 SW41 转子中以 32krpm 20℃ 离心 18 小时。沉淀用 6M 盐酸胍洗涤,最后用干冰预冷的无水乙醇洗涤。沉淀重新悬浮于焦碳酸二乙酯处理的水中,调整至 0.3M 乙酸钠(pH5.2),并用乙醇沉淀。沉淀重新悬浮于蛋白酶 K 降解缓冲液中并于 37℃ 降解 2 小时,酚-氯仿抽提并用乙醇沉淀。将沉淀溶解于水中,调整至 1×DNase 降解缓冲液,并用不含 RNase 的 DNase (BRL) 处理。使用 Invitrogen 的 mRNA 提取试剂盒 FAST TRACK™,根据厂商的操作流程提取 PolyA<sup>+</sup> RNA。源于白种成人肝、脾、肾、脑和胎肝,及红白血病细胞 K562 的 mRNA 从 CLONTECH LABORATORIES 获得。通过用 0.83% NH<sub>4</sub>Cl pH7.4 裂解全血中血红细胞,接着应用由 T. Maniatis, E. F. Fritsh 和 J. Sambrook 在 Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Lab., Cold Spring Harbor, N. Y.)(1982)中所述的 DNA 提取标准流程(其全部内容特此引用作为参考),从四种 Duffy 表型的外周血白细胞中得到 DNA。

#### 实施例 4: gpD 蛋白 cDNA 克隆的核苷酸序列

一种从混合的 Fy(a-b+) 个体 mRNA 构建的非扩增人骨髓 cDNA 文库,用 24 碱基探针进行筛选。从  $1.9 \times 10^6$  重组 λZAP II™ 噬菌体中,筛选出 4 个阳性克隆并测序。所有克隆具有重叠序列但

没有覆盖 gpD cDNA 的全长。Fyb81 和 Fyb71 是最长的, 1085bp 的 Fyb81 是唯一含有最终的 5' 末端的克隆, 1083bp 的 Fyb71 包括核苷酸 185 位至 poly(A)<sup>+</sup> 尾。989bp 的 Fyb31 和 726bp 的 Fyb82 分别包括从核苷酸 275 和 527 位至 Poly(A) 尾。Fyb81 和其它任何克隆组合, 产生 gpD 蛋白的全长 cDNA。图 1a 表示了两个最长克隆的重叠和组合。

接合的 Fyb71-81 克隆预示了一起始于 176 位和终止于 1192 位的开放读码框架, 其编码 339 个氨基酸残基的多肽(图 1b)。应用 BLAST<sup>TM</sup> 网络服务系统, 在 NcBI 上进行 GenBank<sup>TM</sup> 序列检索 (release 77), 显示与人和兔白细胞介素-8 受体的明显的蛋白序列同源性, 并且与人海马 cDNA 克隆 HHCMF86(见下文) 具有准全核苷酸序列同源性。反义引物的延伸产物(从 57 至 80 位, 图 1b), 形成一与 Fyb81 克隆 5' 末端的预期大小完全吻合的 80 个核苷酸的序列(未显示)。起始密码子在 176~178 位上, 而没有嵌入在与哺乳动物翻译起始最常相关的序列中, 见 M. Kozak, *Nucleic Acids Res.*, 87, 8125(1987)。但是, 根据下列原因可以推测它就是真正的起始密码子: (i) 在 5' 末端, 它是唯一的 ATG 密码子; 和 (ii) 从第一个甲硫氨酸开始, 由重组克隆所编码的多肽具有与去糖基化的 gpD 蛋白相同的分子量。见, A. Chaudhuri 和 A. O. Pogo (in press) 的 *BloodCell Biochemistry*. eds. J. P. Cartron and P. Rouger. (Plenum Press, New York) Vol 6。在 3' 末端, 克隆 Fyb71-81 含有通用的 Poly(A) 加尾信号 AATTAAA(图 1b)。

两个克隆具有很好的核苷酸序列匹配, 除了在 5' 末端几个碱基取代产生 6 个不同的预期氨基酸。这些差异不是测序错误, 因为两

条 DNA 链经多次测序。它们是蛋白异源性的结果, 因为该 cDNA 文库是从几个 Fy(a-b+) 个体的 mRNA 构建的。

为了确定 Fy71-81 具有针对于 gpD 蛋白的编码序列, 将翻译的序列与从六种肽 Pe1~Pe6 得到的部分氨基酸序列资料进行比较。预期的氨基酸序列各部分与用 OPA 试剂测序的 Pe1 (SEQ ID No:2) 肽, 与用反相 HPLC 分离的四种肽 (Pe2 (SEQ ID No:3), Pe3 (SEQ ID No:4), Pe4 (SEQ ID No:5), 和 Pe6 (SEQ ID No:7) 肽), 和与用 SDS-PAGE 分离的 Pe5 (SEQ ID No:6) 肽相吻合。但是, 在全部的 62 个氨基酸中有 2 个不吻合。因此, 通过密码子序列分析在 92 和 327 位的残基为色氨酸, 但通过氨基酸序列测定, 它们分别是异亮氨酸和精氨酸。因为色氨酸是一种很不稳定的残基, 这一差异可能是氨基酸序列分析中的技术问题。另一方面, 可能是由于 gpD 蛋白的异源性。

Fyb71-81 编码 gpD 蛋白的其它证据可来自 Northern 印迹和 ELISA 分析。Fyb81 在 Duffy 阴性个体中没有检测到任何 mRNA, 而在 Duffy 阳性个体中可以检测到 1~1.27kb 代表全长 gpD mRNA 的转录本 (图 3a)。抗-Fy6 抗体可与一从 Fyb71-81 克隆可以预示的 35 个氨基酸的合成肽 (9-44 残基, 见图 1b) 反应, (未显示)。Fy(a-b-) 表型缺少 gpD 蛋白特定的 mRNA (见下文) 和抗-Fy6 与由 gpD cDNA 得到的肽的反应, 强烈地说明该分离克隆为真正的 Duffy 克隆。

#### 实施例 5: gpD 蛋白的氨基酸序列和膜拓扑结构

Fyb71-81 克隆的预期翻译产物, 为一等电点 5.65 及分子量为

Mr35,733 的酸性蛋白。该蛋白仅在羧末端带有两个用于天冬酰胺残基 N-糖基化的潜在的规范序列。见 R. D. Marshall, *Annu. Rev. Biochem.*, 41, 673 (1972)。这与 N-糖苷酶 F 消化增加 gpD 在 SDS-PAGE 上的迁移率和 N-乙酰葡糖胺的化学检测的早期研究相一致。见 A. Chaudhuri 和 A. O. Pogo, *supra*; M. J. A. Tanner, D. J. Anstee, G. Mallison, K. Ridgwell, P. G. Mantin, N. D. Aveni, 和 S. F. Parsons, *Carbohydr. Res.*, 178, 203 (1988); 和 K. Wasniowska, P. Eichenberger, F. Kugele, 和 T. J. Hadley, *Biochem. Biophys. Res. Commun.*, 192, 366 (1993)。

应用 Engelman 等 (*Ann. Rev. Biophys. Chem.* 15, 321 (1986)) 的亲水图从序列资料中预测跨膜螺旋的位置, 并且一 20 残基扫描窗显示出大部分蛋白镶嵌于膜中 (图 2)。可以预测 9 个跨膜  $\alpha$ -螺旋, N-末端 66 个残基的亲水结构域, C 末端 25 个氨基酸的亲水结构域, 及短的突出的亲水性连接区段。一对螺旋 D 和 E 由于十分靠近, 所以它们排列成一对反平行螺旋。gpD 蛋白拓扑结构图示于图 2。

由 Hartman 等 (*Proc. Natl. Acad. Sci. USA*, 86, 5786 (1989)) 提出的电荷差规则, 可预测 N 末端位于胞外侧, 蛋白的 C 末端位于膜的细胞质一侧。N-末端两个潜在 N-糖基化位点的发现证实了 N-末端的预测。而且, 抗-Fy6 与由该结构域推演而来的合成肽的反应也从试验上确定了其胞外位置, 因为该抗体与红细胞结合。膜插入的信号锚序列可能位于 N-末端区域之后的第一个  $\alpha$ -螺旋中, 见 H. P. Wessels 和 M. Spies, *Cell*, 55, 61 (1988); 和 G. Blobel, *Proc. Natl. Acad. Sci. USA*, 77, 1496 (1980)。根据上述, 该蛋白深埋在膜中, 并在残基 314 开始存在于膜的胞质侧 (图 2)。螺旋、亲水性连接

区段和 C-末端片段的位置的拓扑结构预测,应该通过直接的生化和免疫化学分析加以证实。

Duffy gpD 蛋白深埋于膜中,如同条带 3 的膜相关片段(见 D. Jay, *Annu. Rev. Biochem.*, 55, 511 (1986))、人 Rh 血型多肽(见, B. Chérif-Zahar et al., *Proc. Natl. Acad. Sci. USA*, 87, 6243 (1990); 及 N.D. Avent et al., *Biochem. J.*, 271, 821 (1990)), 细菌视紫红质(见 P. Carlton et al., *EMBO J.*, 4, 1593 (1985)) 和 Lipophilin (见 W. Stoffel et al. *Hoppe-Seyler' Z. Physiol. Chem.*, 364, 1455 (1983)) 一样。gpD 蛋白与白细胞介素 -8 受体的显著同源性是非常严格的。见 W. E. Holmes et al., *Science*, 253, 1278 (1991); 和 P. M. Murphy et al., *Science*, 253, 1280 (1991)。如果 gpD 蛋白结合趋化因子并具有激活信号传导级联系统的能力,这使得 gpD 蛋白成为新的一类炎症前介质。因此,白细胞内不存在 gpD 蛋白,因为抗纯化、失活的 gpD 蛋白的兔多克隆抗体(抗-gpD)与红细胞及其前体反应,而不与任何白细胞反应(未发表的结果)。

#### 实施例 6: RNA 印迹分析(Northern)

对 Poly(A)<sup>+</sup> RNA 进行甲醛/琼脂糖凝胶电泳并转移到 Hybond<sup>TM</sup>N<sup>+</sup> 尼龙膜 (Amersham Corp.) 上。根据厂商说明,将膜在 QuickHyb<sup>TM</sup> (Stratagene) 上杂交并洗膜。

在 Northern 印迹分析中, Fyb71 或 Fyb81 克隆在三个 Duffy 阳性表型的骨髓中检测到 ~1.27kb mRNA, 但在 Fy(a-b-) 表型个体中并未检测到(图 3a)。在 Duffy 阴性个体中缺少 gpD mRNA 与缺乏 gpD 蛋白相一致。抗-gpD 抗体不与任何 Fy(a+b-) 红细胞的红

细胞膜蛋白反应(未显示)。Duffy 阴性个体不表达 gpD 蛋白,因为它们不能合成 Duffy 特异性 mRNA。

在图 3a 中,泳道 1 含有  $10\mu\text{g}$  Fy(a-b-)mRNA,泳道 2 和 3 分别含有  $5\mu\text{g}$  Fy(a+b-)mRNA 和 Ry(a-b+)mRNA,及泳道 4 含有  $2\mu\text{g}$  Fy(a+b+)mRNA。它们于 2% 变性琼脂糖凝胶上分离、印迹、杂交并在  $-80^{\circ}\text{C}$  放射自显影 72 小时。RNA 大小标准显示为:人 28S(5.1kb)和 18S(2.0kb)rRNA,及 1.35kb GIBCOBRL 标准(LIFE TECHNOLOGIES),其用于计算 gpD mRNA 的大小。底部的肌动蛋白探针作为样品上样量的对照。在 Poly(A)<sup>+</sup>组和肌动蛋白探针中存在两种 rRNA 显示出 RNA 的完整性。

#### 实施例 7:DNA 印迹分析(Southern)

根据供应商(New England Biolabs)所建议的条件进行所有的限制性酶消化。降解的 DNA 在 0.8% 的琼脂糖凝胶上进行大小分离并按 Northern 分析中所述方法印迹。根据厂商说明,在 Quick-Hyb<sup>TM</sup>溶液中  $68^{\circ}\text{C}$  杂交 1 小时。

在 Southern 印迹分析中,Fyb71 或 Fyb81 探针与 Duffy 阳性和阴性个体的 DNA 杂交(图 3b)。它们确定为在 BamHI 中的 6.5kb 单一带,在 EcoRI 中的 12kb 和 2kb 两条带,及在 PstI 降解的 DNA 中为 3.5kb 和 1.4Kb 两条带。这些发现与 Fyb71 和 Fyb81 的限制性酶谱相一致并显示出为单拷贝基因。Duffy 阳性和阴性个体的基因中结构差异的确定,应能阐明在阴性个体中 gpD 基因阻遏的机制。在其它系统中描述的功能性沉默子元件可以在 Fy(a-b-)个体的红细胞中选择性地抑制 gpD 基因的转录。见 L. Li, T. Suzuki,



N. Mori, 和 P. Greengard, Proc. Natl. Acad. Sci. USA, 90, 1460 (1993)。Duffy 系统与 ABO(F. Yamamoto et al, Nature, 345, 229 (1990))和 Kell 系统不同,这些系统中不表达血型决定簇的个体中仍可发现 mRNA。

在图 3b 中,每一泳道中含  $10\mu\text{g}$  降解的 DNA;泳道 1-4 含有 Fy(a-b-)DNA,泳道 5-8 含 Fy(a+b-)DNA,泳道 9 至 12 含 Fy(a-b+)DNA。酶降解如下:泳道 1,5 和 9 为 BamHI;泳道 2,6 和 10 为 EcoRI;泳道 3,7,和 11 为 HinfI;泳道 4,8 和 12 为 PstI。它们于 0.8% 琼脂糖凝胶上分离、印迹、杂交并在  $-80^{\circ}\text{C}$  放射自显影 7 天。根据 GIBCOBRL DNA 标准的位置计算大小。

如图 4 所示,可在成人脾和肾、胎肝中发现 1.27Kb mRNA,但在成人肝和 K562 红白血病细胞中未发现。用  $\beta$ -球蛋白探针杂交在骨髓和胎肝中显示出很强的信号;成人脾中显示微弱信号,但成人肝、脑和肾中没有信号显示(未表示)。胎肝中存在 gpD mRNA 是预料之中的,因为胎肝是促红细胞生成器官。在人脑中,可检测到一 8.5kb 强带和 2.2kb 弱带。这存在许多令人感兴趣的可能性。这表明在脑中存在 Duffy 相关蛋白。而且,这一见解由 Fyb71-81 克隆与人海马 cDNA 克隆 HHCMF86 之间的准全同源性所支持,后者最近被鉴定(见 M. D. Adams 等, Nature, 355, 632(1992))并定为 SEQ ID No:15。但是,8.5kb 脑 mRNA 不可能编码 Duffy 蛋白而带有长的 5'和 3'非翻译序列。该脑 mRNA 可能编码了与 gpD 蛋白具有广泛同源性的较大的蛋白。这些种类的 mRNA 与 gpD 特定的 mRNA 之间的同源性仍有待于通过序列分析加以证实;但是,这一发现有力地表明在肾、非生血性脾细胞和可能在脑中产生 gpD 蛋白或一相似

的蛋白。

图 5 代表显示 gpD 蛋白和人海马 cDNA 克隆 HHCMF86 的 DNA 序列同源性的图表。该 HHCMF86 cDNA 克隆来自一两岁女性高加索人(Adams et al. Nature, 355, 632(1992))。在 HHCMF86 中存在几个未确定的碱基,而且这两个克隆具有相同的直到 gpD 蛋白 623 位为止的 ORF(在 HHCMF86 中为 296)。HHCMF86 在此位置后有一个额外的腺嘌呤,在 ORF 中产生移码。该额外的腺嘌呤很可能是 HHCMF86 cDNA 的测序错误。

在图 4 中,泳道 1, 3, 5 和 7 分别含有 2 $\mu$ g Fy(a - b +)骨髓、胎肝、成人脾和红白血病(K562)mRNA。泳道 2, 4 和 6 分别含有 7 $\mu$ g 全脑、成人肝和成人肾 mRNA。它们在 1.5% 变性琼脂糖凝胶上分离并于 -80 $^{\circ}$ C 放射自显影 5 天。

#### 实施例 8: 人骨髓 cDNA 文库的构建和筛选

几个 Fy(a - b +) 个体 mRNA 的混合物, BRL Superscript Choice<sup>TM</sup> System 和作为引物的寡聚 dT 用以制备 cDNA。将该 cDNA 连接到  $\lambda$ ZAP II<sup>TM</sup> 载体上并用 Gigapack Gold<sup>TM</sup> (Stratagene) 抽提物包装。用上述的<sup>32</sup>P 标记的探针筛选约  $1.9 \times 10^6$  个非扩增 cDNA 克隆。根据厂商的操作流程通过质粒拯救方法分离插入到 pBluescript 的 cDNA。应用载体引物对 DNA 双链进行测序; 根据转录本的已测序的区域设计引物。

#### 实施例 9: 引物延伸

应用预扩增试剂盒 (Pre-amplification Kit) (BRL), 从编码链(图

1b)57 至 80 位的<sup>32</sup>P 标记的 24 核苷酸的反义引物在 Fy(a - b + ) mRNA 上延伸。产物在 6% 的测序胶上分离。M13 序列梯用于确定产物的大小。

应认识到本说明书和权利要求只是说明性质的而非限制性的, 而且在不背离本发明的实质和范围的情况下可进行各种修饰和改变。

## 序列表

### (1)一般信息:

(i) 申请人: Pogo, Angel Oscar; Chaudhuri, Asok.

(ii) 发明题目: DUFFY 血型抗原的克隆

(iii) 序列数: 16

(iv) 联系地址:

(A) 收信人: Sprung Horn Kramer & Woods

(B) 街道: 660 White Plains Road

(C) 城市: Tarry town

(D) 州: New York

(E) 国家: USA

(F) 邮政编码: 10591—5144

(v) 计算机可读形式:

(A) 媒介类型: 软盘, 3.5 英寸, 800kb 贮存

(B) 计算机: Apple Macintosh

(C) 操作系统: System 7.0

(D) 软件: WordPerfect

(vi) 本申请资料

(A) 申请号: 未指定

(B) 申请日: 1994 年 10 月 20 日

(C) 分类: 未指定

(vii) 优先申请资料

(A) 申请号: US 08/140,797

(B) 申请日: 1993 年 10 月 21 日

(viii) 律师/代理人资料:

(A) 姓名: Kurt G. Briscoe

(B) 注册号: 33, 141

(C) 参考/卷号: NYBC 265 - KGB

(ix) 电话联系资料:

(A) 电话: (914) 332 - 1700

(B) 传真: (914) 332 - 1844

(2) SEQ ID No: 1 的信息

(i) 序列特征:

(A) 长度: 1267 核苷酸

(B) 类型: 核酸

(C) 链型: 双链

(D) 拓扑结构: 线型

(xi) 序列描述: SEQ ID No: 1

GGCTTCCCCA GGACTGTTCC TGCTCCGGCT CTCAGGCTC	40
CCTGCTTTGT CCTTTTCCAC TGTCCGCACT GCATCTGACT	80
CCTGCAGAGA CCTTGTTCTC CCACCCGACC TTCCTCTCTG	120
TCCTCCCCTC CCACCTGCCC CTCAGTTCCC AGGAGACTCT	160
TCCGGTGTAATCTCTG ATG GCC TCC TCT GGG TAT GTC CTC	199
Met Ala Ser Ser Gly Tyr Val Leu	
CAG GCG GAG CTC TCC CCC TCA ACT GAG AAC TCA AGT CAG	238
Gln Ala Glu Leu Ser Pro Ser Thr Glu Asn Ser Ser Gln	
CTG GAC TTC GAA GAT GTA TGG AAT TCT TCC TAT GGT GTG	277
Leu Asp Phe Glu Asp Val Trp Asn Ser Ser Tyr Gly Val	
AAT GAT TCC TTC CCA GAT GGA GAC TAT GAT GCC AAC CTG	316
Asn Asp Ser Phe Pro Asp Gly Asp Tyr Asp Ala Asn Leu	

GAA	GCA	GCT	GCC	CCC	TGC	CAC	TCC	TGT	AAC	CTG	CTG	GAT	355
Glu	Ala	Ala	Ala	Pro	Cys	Asn	Ser	Cys	Asn	Leu	Leu	Asp	
GAC	TCT	GCA	CTG	CCC	TTC	TTC	ATC	CTC	ACC	AGT	GTC	CTG	394
Asp	Ser	Ala	Leu	Pro	Phe	Phe	Ile	Leu	Thr	Ser	Val	Leu	
GGT	ATC	CTA	GCT	AGC	AGC	ACT	GTC	CTC	TTC	ATG	CTT	TTC	433
Gly	Ile	Leu	Ala	Ser	Ser	Thr	Val	Leu	Phe	Met	Leu	Phe	
AGA	CCT	CTC	TTC	CGC	TGG	CAG	CTC	TGC	CCT	GGC	TGG	CCT	472
Arg	Pro	Leu	Phe	Arg	Trp	Gln	Leu	Cys	Pro	Gly	Trp	Pro	
GTC	CTG	GCA	CAG	CTG	GCT	GTG	GGC	AGT	GCC	CTC	TTC	AGC	511
Val	Leu	Ala	Gln	Leu	Ala	Val	Gly	Ser	Ala	Leu	Phe	Ser	
ATT	GTG	GTG	CCC	GTC	TTG	GCC	CCA	GGG	CTA	GGT	AGC	ACT	550
Ile	Val	Val	Pro	Val	Leu	Ala	Pro	Gly	Leu	Gly	Ser	Thr	
CGC	AGC	TCT	GCC	CTG	TGT	AGC	CTG	GGC	TAC	TGT	GTC	TGG	589
Arg	Ser	Ser	Ala	Leu	Cys	Ser	Leu	Gly	Tyr	Cys	Val	Trp	
TAT	GGC	TCA	GCC	TTT	GCC	CAG	GCT	TTG	CTG	CTA	GGG	TGC	628
Tyr	Gly	Ser	Ala	Phe	Ala	Gln	Ala	Leu	Leu	Leu	Gly	Cys	
CAT	GCC	TCC	CTG	GGC	CAC	AGA	CTG	GGT	GCA	GGC	CAG	GTC	667
Asn	Ala	Ser	Leu	Gly	Asn	Arg	Leu	Gly	Ala	Gly	Gln	Val	
CCA	GGC	CTC	ACC	CTG	GGG	CTC	ACT	GTG	GGA	ATT	TGG	GGA	706
Pro	Gly	Leu	Thr	Leu	Gly	Leu	Thr	Val	Gly	Ile	Trp	Gly	
GTG	GCT	GCC	CTA	CTG	ACA	CTG	CCT	GTC	ACC	CTG	GCC	AGT	745
Val	Ala	Ala	Leu	Leu	Thr	Leu	Pro	Val	Thr	Leu	Ala	Ser	
GGT	GCT	TCT	GGT	GGA	CTC	TGC	ACC	CTG	ATA	TAC	AGC	ACG	784
Gly	Ala	Ser	Gly	Gly	Leu	Cys	Thr	Leu	Ile	Tyr	Ser	Thr	
GAG	CTG	AAG	GCT	TTG	CAG	GCC	ACA	CAC	ACT	GTA	GCC	TGT	823
Lys	Leu	Lys	Ala	Leu	Gln	Ala	Thr	Asn	Thr	Val	Ala	Cys	
CTT	GCC	ATC	TTT	GTC	TTG	TTG	CCA	TTG	GGT	TTG	TTT	GGA	862
Leu	Ala	Ile	Phe	Val	Leu	Leu	Pro	Leu	Gly	Leu	Phe	Gly	
GCC	AAG	GGG	CTG	AAG	AAG	GCA	TTG	GGT	ATG	GGG	CCA	GGC	901
Ala	Lys	Gly	Leu	Lys	Lys	Ala	Leu	Gly	Met	Gly	Phe	Gly	
CCC	TGG	ATG	AAT	ATC	CTG	TGG	GCC	TGG	TTT	ATT	TTC	TGG	940
Pro	Trp	Met	Asn	Ile	Leu	Trp	Ala	Trp	Phe	Ile	Phe	Trp	
TGG	CCT	CAT	GGG	GTG	GTT	CTA	GGA	CTG	GAT	TTC	CTG	GTG	979
Trp	Pro	Asn	Gly	Val	Val	Leu	Gly	Leu	Asp	Phe	Leu	Val	

AGG TCC AAG CTG TTG CTG TTG TCA ACA TGT CTG GCC CAG Arg Ser Lys Leu Leu Leu Leu Ser Thr Cys Leu Ala Gln	1018
CAG GCT CTG GAC CTG CTG CTG AAC CTG GCA GAA GCC CTG Gln Ala Leu Asp Leu Leu Leu Met Leu Ala Glu Ala Leu	1057
GCA ATT TTG CAC TGT GTG GCT ACG CCC CTG CTC CTC GCC Ala Ile Leu Asn Cys Val Ala Thr Pro Leu Leu Leu Ala	1096
CTA TTC TGC CAC CAG GCC ACC CGC ACC CTC TTG CCC TCT Leu Phe Cys Lys Gln Ala Thr Arg Thr Leu Leu Pro Ser	1135
CTG CCC CTC CCT GAA GGA TGG TCT TCT CAT CTG GAC ACC Leu Pro Leu Pro Glu Gly Trp Ser Ser Asn Leu Asp Thr	1174
CTT GGA AGC AAA TCC TAGTTCTCTT CCCACCTGTC AACCTGAATT Leu Gly Ser Lys Ser	1219
AAAGTCTACA CTGCCTTTGT GAAAAAAAAA AAAAAAAAAA	1259
AAAAAAAAA	1267

(2)SEQ ID No:2 信息

(i)序列特征

(A)长度:48 个核苷酸

(B)类型:核酸

(C)链型:单链

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:2

CCTCTCTTCC GCTGGCAGCT CTGCCCTGGC TGGCCTGTCC	40
TGGCACAG	48

(2)SEQ ID No:3 信息

(i)序列特征

(A)长度:15个核苷酸

(B)类型:核酸

(C)链型:单链

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:3

TTCAGCATTG TGGTG

(2)SEQ ID No:4 的信息

(i)序列特征:

(A)长度:15个核苷酸

(B)类型:核酸

(C)链型:单链

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:4

TTTGCCCAGG CTTTG

(2)SEQ ID NO:5 的信息

(i)序列特征

(A)长度:9个核苷酸

(B)类型:核酸

(C)链型:单链

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:5

GTGGGAATT

(2)SEQ ID No:6 的信息

(i)序列特征



(A)长度:72 个核苷酸

(B)类型:核酸

(C)链型:单链

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:6:

ATGAATATCC TGTGGGCCTG GTTTATTTTC TGGTGGCCTC

40

CTCATGGGGT TCTAGGACTG GATTCCTGG TG

72

(2)SEQ ID NO:7 的信息:

(i)序列特征:

(A)长度:27 个核苷酸

(B)类型:核酸

(C)链型:单链

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:7:

CCCTCTCTGC CCCTCCCTGA AGGATGG

(2)SEQ ID No:8 的信息:

(i)序列特征

(A)长度:66 个氨基酸

(B)类型:氨基酸

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:8:

```

Met Ala Ser Ser Gly Tyr Val Leu Gln Ala Glu Leu Ser Pro Ser
          5                      10                      15
Thr Glu Asn Ser Ser Gln Leu Asp Phe Glu Asp Val Trp Asn Ser
          20                      25                      30
Ser Tyr Gly Val Asn Asp Ser Phe Pro Asp Gly Asp Tyr Asp Ala
          35                      40                      45
Asn Leu Glu Ala Ala Ala Pro Cys His Ser Cys Asn Leu Leu Asp
          50                      55                      60

Asp Ser Ala Leu Pro Phe
          65

```

(2) SEQ ID No:9 的信息:

(i) 序列特征

(A) 长度: 44 个氨基酸

(B) 类型: 氨基酸

(D) 拓扑结构: 线型

(xi) 序列描述: SEQ ID No:9:

```

Met Ala Ser Ser Gly Tyr Val Leu Gln Ala Glu Leu Ser Pro Ser
          5                      10                      15
Thr Glu Asn Ser Ser Gln Leu Asp Phe Glu Asp Val Trp Asn Ser
          20                      25                      30
Ser Tyr Gly Val Asn Asp Ser Phe Pro Asp Gly Asp Tyr Asp
          35                      40

```

(2)SEQ ID No:10 的信息:

(i)序列特征

(A)长度:35 个氨基酸

(B)类型:氨基酸

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:10:

```
Ala Glu Leu Ser Pro Ser Thr Glu Asn Ser Ser Gln Leu Asp Phe
                    5                      10                      15
Glu Asp Val Trp Asn Ser Ser Tyr Gly Val Asn Asp Ser Phe Pro
                    20                      25                      30
Asp Gly Asp Tyr Asp
                    35
```

(2)SEQ ID No:11 的信息:

(i)序列特征:

(A)长度:22 个氨基酸

(B)类型:氨基酸

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:11:

```
Asp Phe Glu Asp Val Trp Asn Ser Ser Tyr Gly Val Asn Asp Ser
                    5                      10                      15
Phe Pro Asp Gly Asp Tyr Asp
                    20
```

(2)SEQ ID No:12 的信息

(i)序列特征

(A)长度:22 个氨基酸

(B)类型:氨基酸

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:12:

```
Ala Asn Leu Glu Ala Ala Ala Pro Cys His Ser Cys Asn Leu Leu
      5                               10                               15
Asp Asp Ser Ala Leu Pro Phe
      20
```

(2)SEQ ID No:13 的信息:

(i)序列特征:

(A)长度:13 个氨基酸

(B)类型:氨基酸

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:13:

```
Ala Glu Leu Ser Pro Ser Thr Glu Asn Ser Ser Gln Leu
      5                               10
```

(2)SEQ ID No:14 的信息:

(i)序列特征:

(A)长度:22 个氨基酸

(B)类型:氨基酸

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:14:

Ala Glu Leu Ser Pro Ser Thr Glu Asn Ser Ser Gln Leu Asp Phe  
5 10 15  
Glu Asp Val Trp Asn Ser Ser  
20

(2)SEQ ID No:15 的信息:

(i)序列特征:

(A)长度:328 个核苷酸

(B)类型:核酸

(C)链型:单链

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:15:

CCACTCCTGT AACCTGCTGG ATGACTCTGC ACTGCCCTTC	40
TTCATCCTCA CCAGTGTCCT GGGTATCCTA GCTAGCAGCA	80
CTGTCCTCTT CATGCTTTTN AGACCTCTCT TCCGCTGGCA	120
GCTCTGCCCT GGCTGGCCTG TCCTGGCACA GCTGGCTGTG	160
GGCAGTGCCC TCTTCAGCAT TGTGGTGCCC GTTTTGGCCC	200
CAGGGCTAGG TAGCACTCGC AGCTCTGCCC TGTGTAGCCT	240
GGGCTACTGT GTCTGGTATG GCTCAGCCTT TGNCCAGGCT	280
TTGCTGCTAA GGGTGCCATG CCTCCCTGGG NCACAGACTG	320
GGTGCAGG	328

(2)SEQ ID No:16 的信息:

(i)序列特征:

(A)长度:24 个核苷酸

(B)类型:核酸

(C)链型:单链

(D)拓扑结构:线型

(xi)序列描述:SEQ ID No:16:

TGGTTTATTT TCTGGTGGCC TCAT

说明书附图

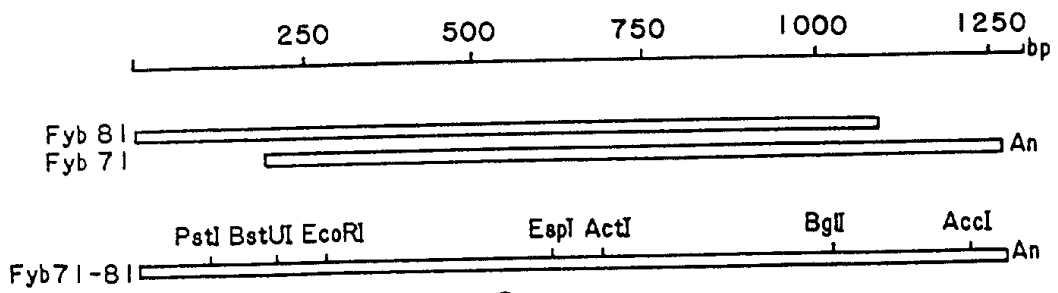


图 1A

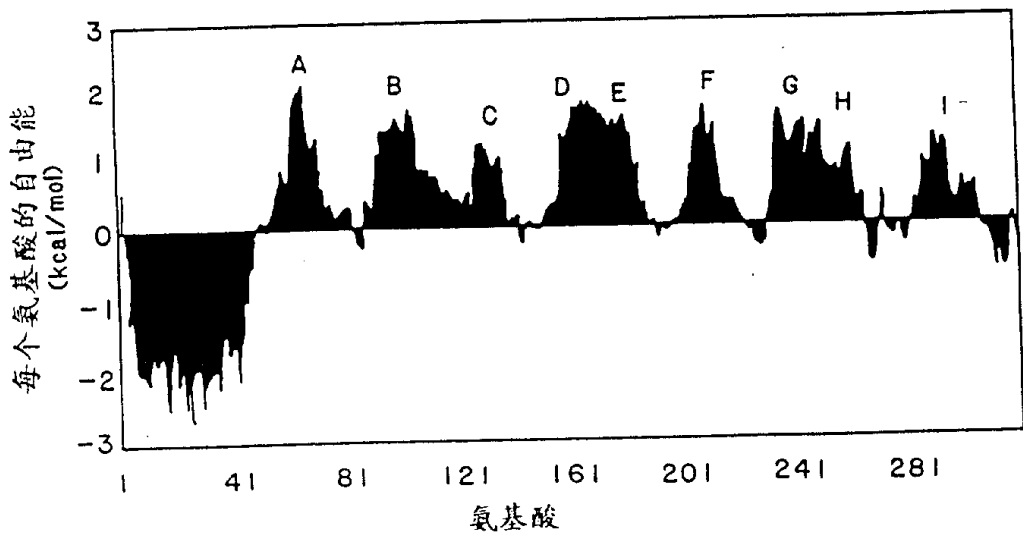


图 2A

55 G GCT TCC CCA GGA CTG TTC CTG CTC CGG CTC TTC AGG CTC CCT GCT TTG TCC TTT  
 112 TCC ACT GTC CGC ACT GCA TCI GAC ICC TGC AGA GAC CTT GTT CTC CCA CCC GAC CTT  
 169 CCT CTC TGT CCT CCC CTC CCA CCT GCC CCT CAG TTC CCA GGA GAC TCT TCC GGT GTA  
 226 ACT CTG ATG GCC TCC TCT GGG TAT GTC CTC CAG GCG GAG CTC TCC CCC TCA ACT GAG  
 283 AAC TCA AGT CAG CTG GAC TTC GAA GAT GTA TGG AAT TCT TCC TAT GGT GTG AAT GAT  
 340 TCC TTC CCA GAT GGA GAC TAT GAT GCC AAC CTG GAA GCA GCT GCC CCC TGC CAC TCC  
 397 TGT AAC CTG CTG GAT GAC TCT GCA CTG CCC TTC TTC ATC CTC ACC AGT GTC CTG GGT  
 454 ATC CTA GCT AGC AGC ACT GTC CTC TTC ATG CTT TTC AGA CCT CTC TTC CGC TGG CAG  
 511 CTC TGC CCT GGC TGG CCT GTC TGG CCT GTC GCA CAG CTG GCT GTG GGC AGT GCC CTC TTC AGC  
 568 ATT GTG GTG CCC GTC TTG GCC CCA GGG CTA GGT AGC ACT CGC AGC TCT GCC CTG TCT  
 625 AGC CTG GGC TAC TGT GTC TGG TAT GGC TCA GCC TTT GCC CAG GCT TTG CTG CTA GGG  
 1 GCT TCC CCA GGA CTG TTC CTG CTC CGG CTC TTC AGG CTC CCT GCT TTG TCC TTT  
 18 TCC ACT GTC CGC ACT GCA TCI GAC ICC TGC AGA GAC CTT GTT CTC CCA CCC GAC CTT  
 37 CCT CTC TGT CCT CCC CTC CCA CCT GCC CCT CAG TTC CCA GGA GAC TCT TCC GGT GTA  
 56 ACT CTG ATG GCC TCC TCT GGG TAT GTC CTC CAG GCG GAG CTC TCC CCC TCA ACT GAG  
 75 AAC TCA AGT CAG CTG GAC TTC GAA GAT GTA TGG AAT TCT TCC TAT GGT GTG AAT GAT  
 94 TCC TTC CCA GAT GGA GAC TAT GAT GCC AAC CTG GAA GCA GCT GCC CCC TGC CAC TCC  
 113 TGT AAC CTG CTG GAT GAC TCT GCA CTG CCC TTC TTC ATC CTC ACC AGT GTC CTG GGT  
 132 ATC CTA GCT AGC AGC ACT GTC CTC TTC ATG CTT TTC AGA CCT CTC TTC CGC TGG CAG  
 Pe 1  
 Pe 2  
 Pe 3

IB-1



151 TGC CAT GCC TCC CTG GGC CAC AGA CTG GGT GCA GGC CAG GTC CCA GGC CTC ACC CTG L 682  
 C H A S L G G H R L G A G G Q V P G L T L  
 170 GGG CTC ACT GTG GGA ATT TGG GGA GTG GCT GCC CTA CTG ACA CTG CCT GTC ACC CTG L 739  
 G L T V G L W G G A A L L T L P V T L  
 189 GCC AGT GGT GCT TCT GGT GGA CTC TGC ACC CTG ATA TAC AGC ACG GAG CTG AAG GCT A 796  
 A S G A S G G G L C T T L I Y S T E L K A  
 208 TTG CAG GCC ACA CAC ACT GTA GCC TGT CTT GCC ATC TTT GTC TTG TGG CCA TTG GGT L 853  
 L Q A T H T V A C L A I F V L L P L G  
 227 TTG TTT GGA GCC AAG GGG CTG AAG GCA TTG GGT ATG GGG CCA GGC CCC TGG ATG L 910  
 L F G A K G L K A L G M G P G P W M  
 246 AAT ATC CTG TGG GCC TGG TTT ATT TTC TGG TGG CCT CAT GGG GTG GTT CTA GGA CTG L 967  
 H I L W A W F I F W W P H G V V L G L  
 265 GAT TTC CTG GTG AGG TCC AAG CTG TTG CTG TTG TCA ACA TGT CTG GCC CAG CAG GCT L 1024  
 D F L V R S K L L L L L S T C L A Q Q A  
 284 CTG GAC CTG CTG AAC CTG GCA GAA GCC CTG GCA ATT TTG CAC TGT GTC GCT ACG L 1081  
 L D L L L N L A E A L A I L L H C V A T  
 303 CCC CTG CTC CTC GCC CTA TTC TGC CAC CAG GCC ACC CGC ACC CTC TTG CCC TCT CTG L 1138  
 P L L L A L F C H Q A T R T L L P S L  
 322 CCC CTC CCT GAA GGA TGG TCT TCT CAT CTG GAC ACC CTT GGA AGC AAA TCC TAG TTC L 1195  
 P L P E G W S S H L L D T L G S K S  
 TCT TCC CAC CTG TCA ACC TGA ATT AAA GTC TAC ACT GCC TTT GTG AAA AAA AAA 1252  
 AAA AAA AAA AAA AAA 1267

IB-2

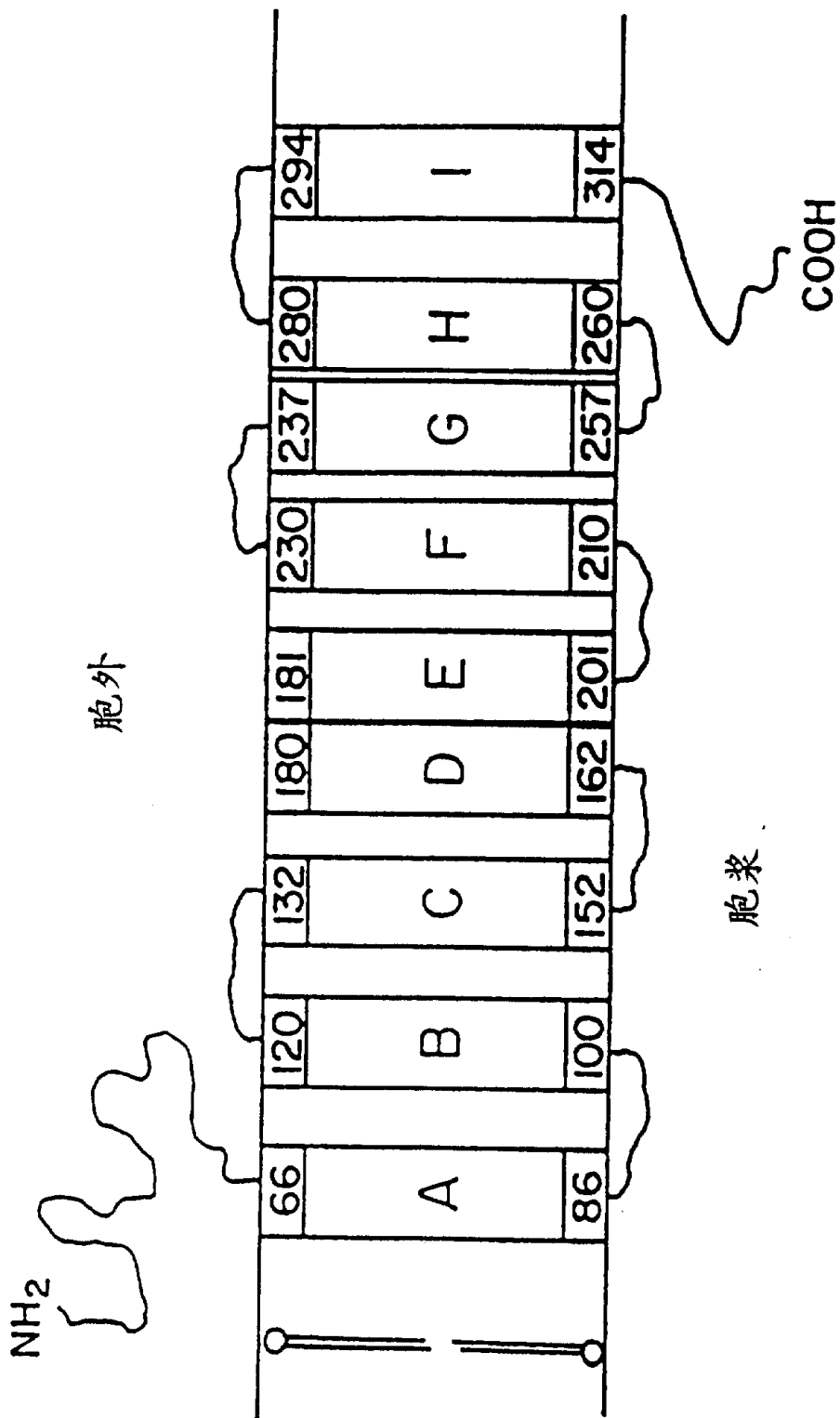
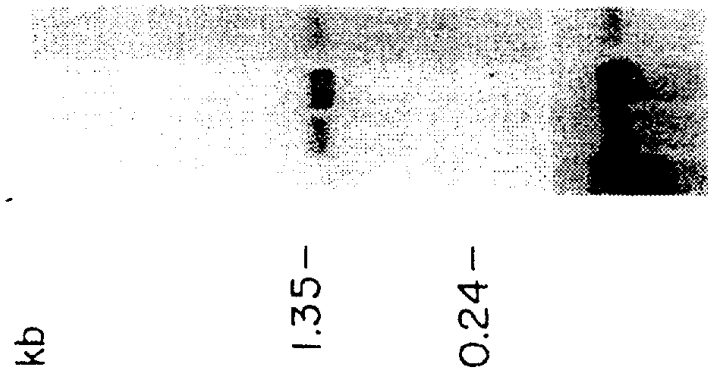


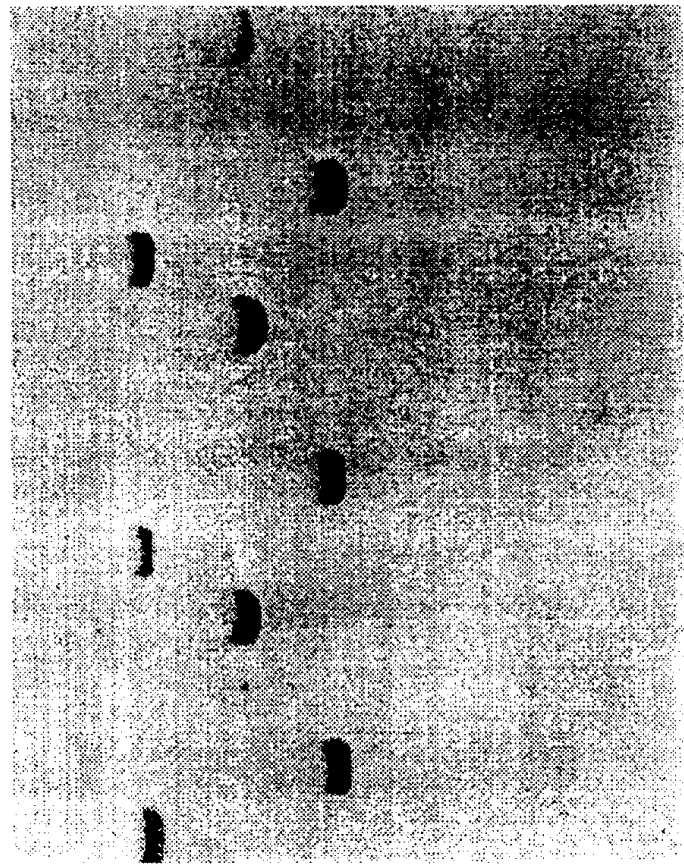
图 2B



1 2 3 4

图3A

kb  
12.0—  
6.5—  
3.5—  
2.0—  
1.4—  
0.5—  
肌动蛋白



1 2 3 4 5 6 7 8 9 10 11 12

图3B

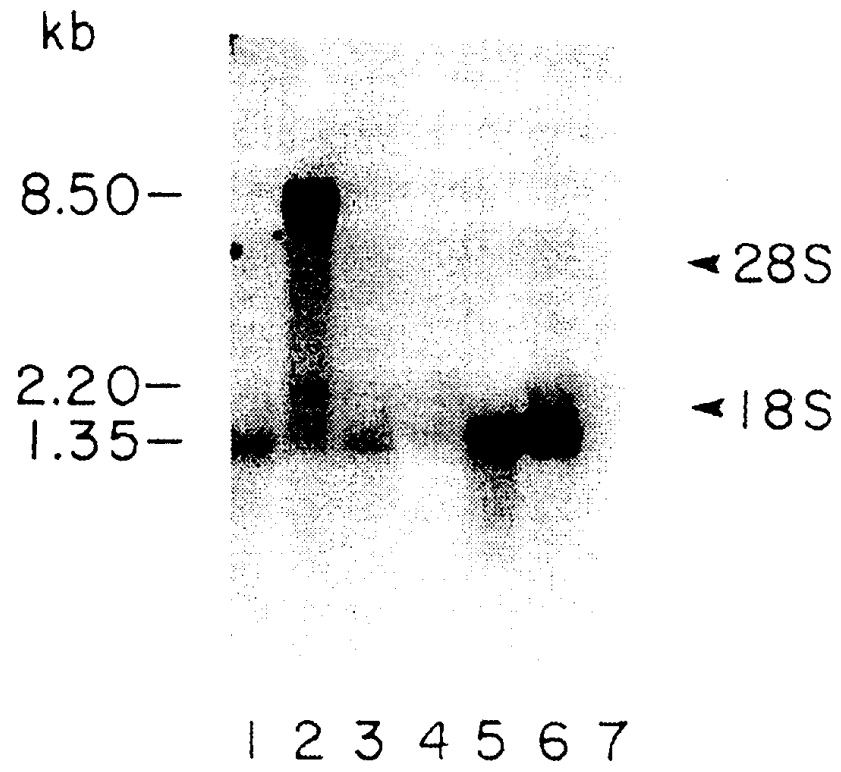


图. 4

```

GPD 蛋白      334      343      353      363      373      383      393
               CCACTCCTGTAACCTGCTGGATGACTCTGCACTGCCCTTCTTCATCCTCACCAGTGTCT
HHCMF86      1         10        20        30        40        50        60
               CCACTCCTGTAACCTGCTGGATGACTCTGCACTGCCCTTCTTCATCCTCACCAGTGTCT

GPD 蛋白      394      403      413      423      433      443      453
               GGGTATCCTAGCTAGCAGCACTGTCTCTTCATGCTTTTCAGACCTCTTCCGCTGGCA
HHCMF86      61        70        80        90        100       110       120
               GGGTATCCTAGCTAGCAGCACTGTCTCTTCATGCTTTNAGACCTCTTCCGCTGGCA

GPD 蛋白      454      463      473      483      493      503      513
               GCTCTGCCCTGGCTGGCTGTCTGGCACAGCTGGCTGTGGGCAGTGCCTCTTCAGCAT
HHCMF86      121       130       140       150       160       170       180
               GCTCTGCCCTGGCTGGCTGTCTGGCACAGCTGGCTGTGGGCAGTGCCTCTTCAGCAT

GPD 蛋白      514      523      533      543      553      563      573
               TGTGGTGCCCGTCTTGGCCCCAGGGCTAGGTAGCACTCGCAGCTCTGCCCTGTGTAGCCT
HHCMF86      181       190       200       210       220       230       240
               TGTGGTGCCCGTTTTGGCCCCAGGGCTAGGTAGCACTCGCAGCTCTGCCCTGTGTAGCCT

GPD 蛋白      574      583      593      603      613      623      633
               GGCTACTGTGTCTGGTATGGCTCAGCCTTTGCCAGGCTTTGCTGCTA-GGGTGCCATG
HHCMF86      241       250       260       270       280       290       300
               GGGCTACTGTGTCTGGTATGGCTCAGCCTTTGNCCAGGCTTTGCTGCTAAGGGTGCCATG

GPD 蛋白      634      643      653      663      673      683      693
               CCTCCCTGGGCCACAGACTGGGTGCAGGCCAGGTCCCAGGCCTCACCTGGGGCTCACTG
HHCMF86      301       310       320       330       340       350       360
               CCTCCCTGGGNACAGACTGGGTGCAGG.....

GPD 蛋白      694
               TGGGAATT
HHCMF86      361

```

匹配百分率

总窗 87% ( 323/ 368 )  
 连线窗 98% ( 323/ 328 )

图 5