

1470

17 Scholtz J M, Baldwin R L. Annu Rev Biophys Biomol

Struct, 1992; 21: 95

18 Vasquez M, Pincus M A, Scheraga H A, Biopolymers, 1987; 26: 351

19 DeGrado W F. Science, 1990; 250: 646

# 肽库及其在分子识别中的应用

齐 杰 鲁治滨 王玉宏 李 惟

(吉林大学分子生物学系, 长春 130023)

**摘要** 肽库 (peptide library) 是利用基因克隆技术将合成的一组寡核苷酸混合物 (小肽基因混合物) 克隆至线性噬菌体基因组中, 使之以融合蛋白的形式在噬菌体的外壳蛋白 (Ⅲ 或 Ⅳ) 的氨基端表达。肽库技术可以应用于与分子识别有关的许多领域, 如: 药物设计, 疫苗设计, 酶的抑制剂的筛选, 蛋白质间的相互作用的研究等等。这是一项新兴的具有很高的实用价值和理论价值的技术。文中综述了其产生、发展和未来的应用前景。

**关键词** 分子识别, 肽库, 基因克隆, 线性噬菌体

肽库是大量某一长度 (如六肽) 的短肽的集合, 它包括了该长度的短肽的各种可能序列或其中的绝大部分, 是近几年发展起来的一种新技术, 其原理很简单 (如图 1 所示)。显然该项技术具有重要的理论价值和实践上的应用价值, 可以用来研究蛋白质间的相互识别<sup>[1-4]</sup>、蛋白质的折叠及空间构象预测<sup>[5,6]</sup>, 及肽与有机物 (如染料、亲和素) 间的特异相互作用<sup>[2,7,8]</sup>, 还可以用来研究酶与底物的结合、抗原-抗体的相互作用、激素与受体的结合, 以及建立在这些信息基础上的实际应用。如疫苗设计、基因定位、小分子药物设计等。由于肽库技术具有如此重要的理论价值和实际应用前景, 因此, 目前它的发展是很迅速的。这方面的报道在 *Nature*、*Science* 上多次出现过。它将给蛋白质、核酸和药物设计等领域带来一场革命。

肽库技术的形成基于两方面的进展。早在 1986 年, Geysen<sup>[9]</sup>认为: a. 含有关键残基的短肽能够模拟蛋白质上的决定簇。b. 多数情况下, 几个关键残基与它的结合分子间形成的非共价键构成了全部结合能的绝大部分。也就是说, 蛋白质之间的相互作用或识别是通过局部肽段间的相互作用来实现的。上述这两点为肽库的提出奠定了理论基础。1988 年 Stephen 和 George<sup>[10]</sup>建立了一种新的表达载体: Anti-body-selectable filamentous fd phage vector, 它可将外源短肽表达并伸到噬菌体的表面。可以用亲和纯化法筛选表达特异肽序列的噬菌体。通过测定噬菌体的 DNA 序列, 就可知道所表达的肽段的氨基酸序列。这一进展为肽库的提出提供了技术保障。

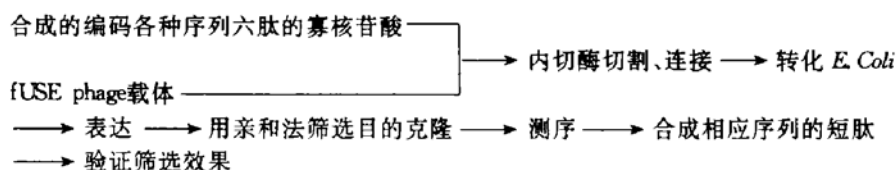


图 1 肽库技术流程图

受 Geysen 等研究小组的工作的启发, Smith, McCutchan 和 dela Cruz 认为利用新的表达载体和筛选技术, 能够从给定长度的含有各种可能序列的肽中筛选出决定簇中的关键残基 (minotopes). 实践证明, 他们的这种想法是可行的.

## 1 肽库的构建

肽库的构建有两种途径: 一种是有有机合成法<sup>[11]</sup>, 一种是基因工程法<sup>[1,12]</sup>. 所谓有机合成法, 就是直接利用固相肽合成技术, 合成含有各种可能序列的短肽. 方法如下:

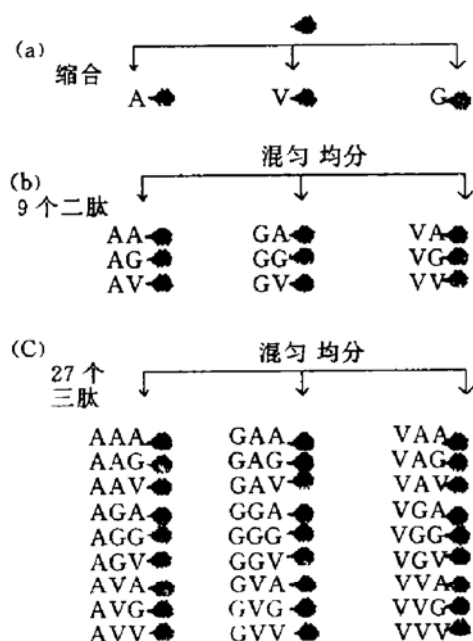


图 2 肽库合成和筛分示意图

表格为三肽库的固相“均分合成”方法的简单说明. 它采用标准的 Fmoc 或 Boc 固相肽合成系统, 含丙氨酸 (A), 甘氨酸 (G) 和缬氨酸 (V). 缩合后, 3 个反应器中的载体被取出, 相互混匀, 重新分装在 3 个反应器中, 进行下一步缩合反应. 经过三步缩合反应后, 27 种 ( $3^3$ ) 全部可能序列的三肽都被合成了.

这样就可以保证各种序列的多肽等几率出现. 每个载体上只含有一种序列的短肽. 其优点是构建方法简单、迅速, 可以引入 D-氨基酸和非天然氨基酸及特异的二级结构短肽, 如环

肽. 库的大小理论上不受限制. 其缺点是: 不能扩增, 筛选较麻烦, 需进行荧光标记或 ELISA, 在显微镜下操作工作量比较大.

这里我们重点谈一下基因工程方法, 因为这种方法是一种创新, 具有很大潜力. 采用这种方法建库的人比较多, 尽管目前该方法还有一些局限性.

基因工程方法是将编码各种序列特定长度的短肽的目的 DNA 克隆进表达载体 (fUSE phage)<sup>[13]</sup>, 与噬菌体的外壳蛋白融合表达. 一个噬菌体上含有一种序列的肽. 所用的噬菌体为 M13 系列中的 fd 噬菌体. 它是一种单链噬菌体, 侵染雄性大肠杆菌. 侵染后, 宿主菌并不死亡, 只不过生长速度变慢, 并向溶液中大量分泌噬菌体颗粒. 我们实验室使用的载体 fUSE 具有四环素抗性, 可以像质粒一样使用. 这种载体在设计时加入了内切酶位点, 同时进行了移码突变, 使载体本身失去侵染能力, 外源片段的插入能够回复其侵染能力, 这样建库时就减少了载体的背景干扰. 目前, 用来作外源肽的载体的蛋白有两种, 一种是微量蛋白 p III, 一种是主要外壳蛋白 p VIII. 两种蛋白均含有

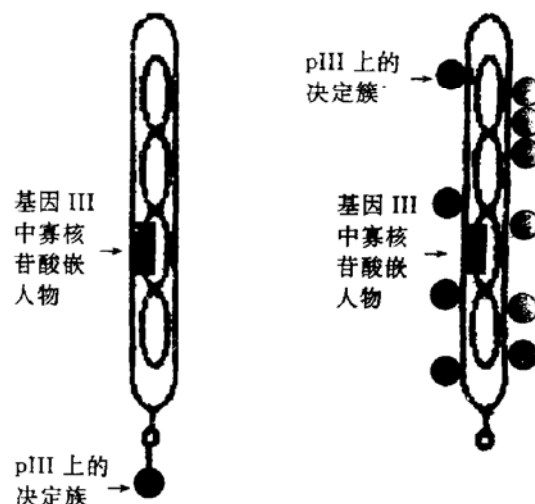


图 3 线性噬菌体表达的融合蛋白示意图

p III 只显示了一个拷贝, 而实际上要有 4—5 个拷贝.

一信号肽, 引导它们到细菌的内膜. 在此处, 被信号肽酶剪切, 形成成熟蛋白, 当噬菌体分泌

时形成外壳. pⅦ蛋白形成噬菌体的躯体, pⅢ蛋白形成噬菌体的尾部, 有 4—5 个拷贝.

pⅢ与噬菌体吸附 *E. coli* 有关, 含 406 个氨基酸, 折叠成两个结构域. 羧基端与噬菌体外壳蛋白作用, 也参与噬菌体的组装. 氨基端伸出噬菌体外, 参与噬菌体对 *E. coli* 的侵染. 外源肽插入 pⅢ的氨基端, 当插入肽段较短时, 并不影响噬菌体的感染性. 如果想表达较大肽段 (Fv, Ab, 酶等), 需有辅助噬菌体或含有基因Ⅲ的质粒.

pⅦ参与噬菌体的组装, 含有 50 个氨基酸, 根据氨基酸的组成为 3 个区域. N 端为酸性氨基酸区, C 端为碱性氨基酸区 (有 4 个 Lys), 中间为疏水区 (21—39). N 端和疏水区与噬菌体的组装有关. 外源肽段插入 pⅦ的 N 端, 破坏它的组装, 使用时需有辅助噬菌体或含有基因Ⅶ的质粒的存在.

无论是表达在 pⅢ上还是 pⅦ上, 建库的方法是一样的. 有两种路线:

路线 1<sup>[1]</sup> (图 4):

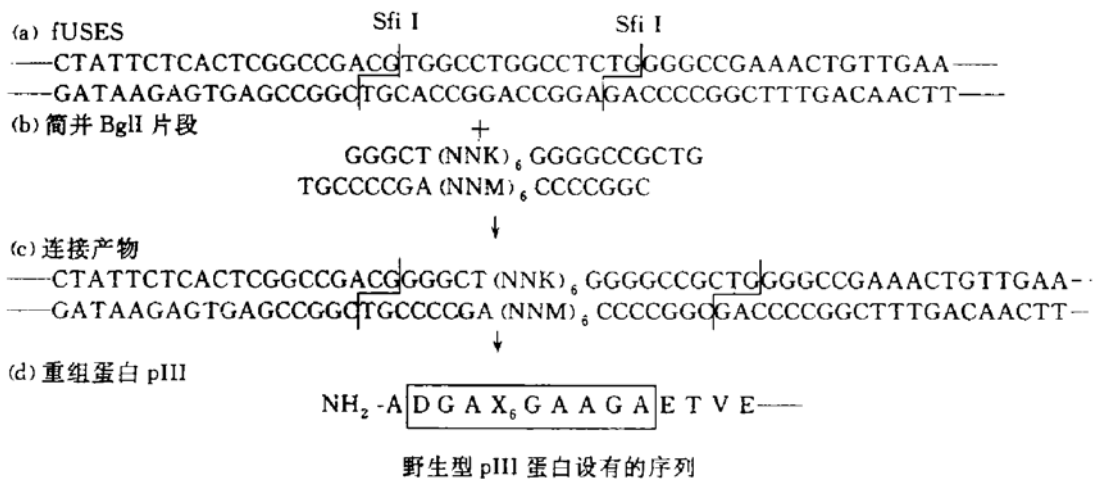


图 4 含有编码肽决定簇的寡核苷酸嵌入物的重组噬菌体基因组库的构建

(a) fUSE5 载体 Sfi I 酶切后, 与含简并密码 (NNK) 的双链寡核苷酸 (b) 连接生成重组基因组片段 (c), (d) 为表达的重组尾部蛋白 pⅢ的 N 端氨基酸序列.

路线 2<sup>[2]</sup> (图 5):

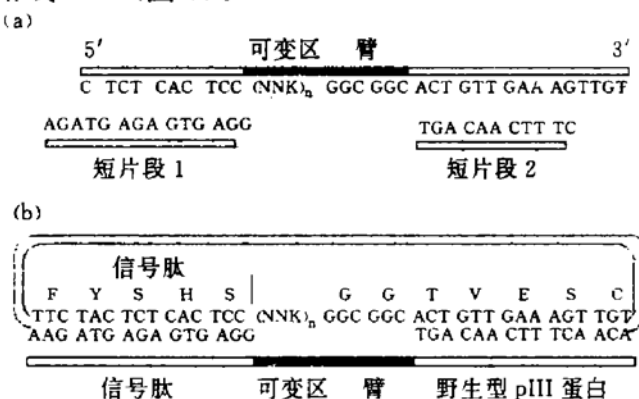


图 5 寡核苷酸库的构建

(a) 寡核苷酸与两个 DNA 短片段杂交形成和 BstXI 两个酶切位点相匹配的粘性末端, (b) 载体 fAFF1 含有两个非互补的 BstXI 位点, 中间有一段 31 个碱基的片段, 除去该片段后, 可定向连接相应的外源片段.

M 为等摩尔的 A 和 C, N 为等摩尔的 G、A、T 和 C、K 为等摩尔的 G 和 T. 因此, NNK 代表 32 种三联码的等量混合物, 编码所有 20 种氨基酸和琥珀型终止子.

路线 1 中合成寡核苷酸单链后, 利用 PCR 扩增形成带限制内切酶切点的双链 DNA, 之后, 再酶切、克隆.

路线 2 中合成寡核苷酸单链后, 克隆形成带有部分单链 DNA 的重组子, 转入大肠杆菌后, 由大肠杆菌将豁口填满.

转化率与肽库的好坏很有关系, 对于六肽库, 有  $0.64 \times 10^8$  ( $20^6$ ) 种不同的序列, 即约  $10^9$  ( $32^6$ ) 种不同的密码子组合. 为了包含所有的序列, 必需有  $10^9$  以上的转化子.

## 2 肽库中序列的多样性

NNK 代表简并密码子结构, 产生 32 种密码子, 其中, 12 种氨基酸各一个密码子, 5 种氨基酸各两个密码子, 3 种氨基酸各三个密码子, 此外还含一个琥珀型终止子. 由于各种氨基酸的密码个数不同, 各种序列的肽的比例并不相同. 例如, NNK 中有三种 Ser 密码子, 只有一种 Glu 密码子. 这意味 (Ser)<sub>6</sub> 的数量为 (Glu)<sub>6</sub> 的 729 (3<sup>6</sup>) 倍. 由于这种情况, 3×10<sup>8</sup> 的六肽库只能包含全部 0.64×10<sup>8</sup> 种六肽的 75%, 如果每个氨基酸只用一种密码子, 同样的肽库能含有全部六肽的 97%. 因此, 如果建长肽库, 最好用核苷酸三联码, 一个氨基酸一个密码子. 因为每增加一个氨基酸, 这种分布的不均衡将扩大三倍.

序列分布的不均衡性还与生物选择有关. Steven 等人<sup>[12]</sup>报导, 随机取出 52 个转化克隆进行测序, 发现密码子的不同位置核苷酸的分布如表 1.

表 1 随机克隆中核苷酸的分布

| 碱基 | 每种碱基出现的频率 |    |    |
|----|-----------|----|----|
|    | 1         | 2  | 3  |
| G  | 31        | 27 | 59 |
| A  | 22        | 22 | <1 |
| T  | 25        | 26 | 39 |
| C  | 22        | 24 | 1  |

最后, 序列分布的不均衡性还与化学法合成寡核苷酸的过程有关, 因为种种核苷酸的反应活性不同.

## 3 肽库质量的检测

肽库的好坏标准主要看两点: a. 对抗原-抗体来说, 能否筛出与抗原有同源性的保守序列. b. 含有筛出的保守序列的肽能否竞争抗原-抗体的作用. 目前, 有四个小组进行过这方面的研究, 多用 Mab. 因为这些 Mab 的抗原的序

列都是已知的, 便于比较. 表 2 是 Kit<sup>[11]</sup>的工作结果, 他筛选了抗 β-内啡肽的配体:

表 2 抗-β内啡肽配基的亲性和. 天然配基与从五肽肽库中筛选出的配基的比较

|      | 序列    | K <sub>i</sub> /nmol · L <sup>-1</sup> |
|------|-------|----------------------------------------|
| 天然配基 | YGGFL | 17.5 ± 3.2                             |
| 肽配基  | YGGLQ | 15.0 ± 1.7                             |
|      | YGGFA | 32.9 ± 2.0                             |
|      | YGGFT | 36.9 ± 7.7                             |
|      | YGGLS | 726 ± 134                              |
|      | YGALQ | 1980 ± 303                             |
|      | YGGMQ | 8780 ± 1500                            |

注: 肽配基亲和常数是通过对放射性配基竞争结合实验确定的, 以 YGGFL 为标准.

## 4 目的克隆的筛选

建成肽库后, 下一步的工作就是筛选出能够与目标分子 (如抗体, 激素等) 紧密结合的噬菌体即目的克隆. 一般经过三个循环, 称为 Biopanning. 第一个循环为粗筛, 生物素标记的抗体先与亲和素包被的全塑培养皿反应, 之后, 再与肽库作用, 洗去游离的噬菌体. 将与抗体结合的噬菌体用酸洗下来, 进行扩增. 第二和第三个循环一样, 将上一循环扩增的噬菌体与生物素标记的抗体作用, 再与亲和素包被的全塑培养皿反应, 洗去游离的噬菌体, 回收得到与目标分子 (如抗体, 激素等) 紧密结合的噬菌体. 最后, 将筛选出的噬菌体进行测序、比较, 找出保守序列. 下一步合成含有保守序列的肽, 看看是否与目标分子有亲和性, 检查筛选的效果.

## 5 肽库的应用

### 5.1 蛋白质的折叠及立体结构预测的研究

Gething<sup>[5]</sup>进行了蛋白质的折叠的研究, 她利用一个八肽库, 以 Bip (immunoglobulin heavy chain-binding protein) 为模型, 已经取得初步结果. 她希望经过进一步的研究, 能够清

楚 Bip 结合蛋白在蛋白质折叠中的作用。

蛋白质空间构象的形成起始于肽段间（一般为六肽）的相互识别，利用肽库研究肽段间相互识别的规律，为蛋白质空间构象的理论预测提供实验依据。这个工作我们正在进行。

## 5.2 分子识别

分子识别包括许多方面，这些方面的研究是目前进行的最多的。如抗原-抗体间的相互识别，激素-受体间的相互识别，蛋白质-DNA 间的识别<sup>[14]</sup>和蛋白质-非蛋白质（如亲和素）间的相互识别等。

## 5.3 在酶工程中的应用前景

目前，肽库技术在酶工程中的应用还属空白。不过，肽库技术在酶工程中的应用潜力是很大的。可以用它来筛选酶的抑制剂、激活剂、稳定剂等。与 X 衍射技术、快速反应动力学技术等结合，可以更深入地研究酶与底物作用的机制，结构与功能的关系。此外，对酶的修饰、改造及人工酶设计、蛋白质工程等均有很大的参考和指导作用。

## 5.4 其他领域中的应用前景

由于肽库技术刚刚起步，许多领域还没有开发，它的巨大应用潜力没有充分发挥，这些领域的工作急需开展。

药物设计：如受体抑制剂，激活剂的筛选、激素类似物的筛选（如脑啡肽类似物）。可以直接用筛选出的肽片段作药物，也可以再加以修饰或参考其结构合成有机药物。

疾病检测：一些疾病常常伴有特殊蛋白质的产生，如自身免疫疾病的自身抗体，爱滋病的 HIV 抗体等。可以用筛选出的肽作为代抗原进行 ELISA 检测。

疾病治疗：对于一些病理比较清楚的疾病，可以利用筛选出来的肽阻断致病的某个中间环节。爱滋病，HIV 必需侵染 CD<sub>4</sub> 细胞才能破坏人体的免疫系统，用与 HIV 有强亲和性的肽或其类似物阻断 HIV 与 CD<sub>4</sub> 细胞的结合，从而阻断 HIV 对人体的免疫系统的破坏。

疫苗设计：过去一直用灭活的病毒给人注射来获得免疫力。这种方法比较麻烦而且带有一定的危险性，用人工设计、合成的疫苗是一种必然的趋势。肽库技术是疫苗设计的有力的理论指导。

## 6 肽库技术的局限性

肽库技术还存在一些需要改善的地方。首先，目前所建的肽库只能达到 10<sup>9</sup>，构建大片断肽库很困难。其次，肽的多样性问题。第三，少数肽或者由于过于疏水，或者由于影响外膜蛋白的折叠而不能在噬菌体外正确表达。作为一种新的技术，类似的具体问题还有许多，但是，这些暂时的缺点并不能掩盖它的巨大应用潜力。

## 参 考 文 献

- 1 Jamie K S, George P S. *Science*, 1990; **249**: 386
- 2 James J D, Lucy C P, Patricia E D. *Science*, 1990; **249**: 404
- 3 Robert E H, Stephen J R, Greg W. *J Mol Biol*, 1992; **226**: 889
- 4 Luisa C, Andrea M, Roberto J. *J Mol Biol*, 1991; **222**: 301
- 5 Gething M J, Sambrook J. *Nature*, 1992; **355**: 33
- 6 Wang Y H, Lu Z B, Li W. *Chem Res in Chinese Univ*, 1994; 1
- 7 Oldenberg K R *et al.* *Proc Natl Acad Sci USA*, 1992; **89**: 5393
- 8 Scott J K, Duraikkannu L, Easley R B *et al.* *Proc Natl Acad Sci USA*, 1992; **89**: 5398
- 9 Geysen H M, Rodds S J, Mason T J. *Mol Immunol*, 1986; **23**: 709
- 10 Stephen F P, George P S. *Gene*, 1988; **73**: 305
- 11 Kit S L, Sydney E S, Evan M H *et al.* *Nature*, 1991; **354**: 82
- 12 Steven E C, Elizabeth A P, Ronald W B *et al.* *Proc Natl Acad Sci USA*, 1990; **87**: 6378
- 13 Zacher A N, Stock C A, Golden J W. *Gene*, 1980; **9**: 127
- 14 Paul S F, Andrew N L, Mark R S. *Biochem J*, 1991; **278**: 1

structure formation is the foundation of protein folding initiation. The paper presents the solution conformation of protein fragment and methods of conformation elucidation, and De novo design of peptide with secondary structure. Also, the application of these achievements in constructing the theoretical model of folding initiation and progress on research of protein folding initiation up to date are detailedly reviewed.

**Key words** framework model, secondary structure, de novo design, folding initiation

**Peptide Library and Its Applications in Molecular Recognition.** Qi Jie, Lu Zhibin, Wang Yuhong, Li Wei. (*Department of Molecular Biology, Jilin University, Changchun 130023*). *Prog. Biochem. Biophys. (China)*. 1994; 21 (6): 513—517

Peptide library is a collections of peptides. The peptides display on the N-terminal of p III or p VII coat protein of bacteriophage through gene cloning. This might be used in the fields associated with molecular recognition, such as: drug design, selection of enzymes inhibitor, vaccines selection, interaction of proteins, etc. Peptide library technique is a lately developed technique with high practical and theoretical value. Peptide library birth, development and potential applications in the future are reviewed.

**Key words** molecular recognition, peptide library, gene cloning

**The Analgesic Domain of Interleukin-2.** Xu Di, Jiang Chunlei, Zheng Zhongcheng, Fan Peifang, Sun Lanying, Liu Xinyuan, Song Chaoyou, You Zhendong, Wang Chenghai, Lu Changlin. (*Shanghai Institute of Biochemistry, Academia Sinica, Shanghai 200031*). *Prog.*

*Biochem. Biophys. (China)*. 1994; 21 (6): 518—520

Interleukin-2 (IL-2) is an important immune regulator. Recently it is found that IL-2 is also an analgesic molecule. Using the potassium iontophoresis as pain stimulus to induce tail-flick and taking the intensity of current (mA) at the moment of the response as the pain threshold (PT), it is reported that mutant 20 Leu-IL-2 (20Asp→Leu), which could not bind to the  $\beta$  subunit of IL-2 receptor and thus had no immune activity, could increase the PT of the rats, indicating that the analgesic and the immunal functions be related to two different domains in IL-2 molecule. Mutant 45Val-IL-2 (45Tyr→Val), which had immunal activity, had no analgesic effect, suggesting that the 45Tyr be crucial for the analgesic effect of IL-2 and the analgesic domain be located around the 45Tyr.

**Key words** interleukin-2, site-directed mutagenesis, analgesia, structure-function studies, central nervous system

**Characterization and Distribution of the Antigen Recognized by Breast Carcinoma McAb AF9.** Cai Guiying, Zeng Wenjie. (*Department of Biochemistry, West China University of Medical Sciences, Chengdu 610041*). *Prog. Biochem. Biophys. (China)*. 1994; 21 (6): 521—523

A preliminary study was made on the characterization and distribution in human tissues of AF9-recognized antigen. The results demonstrated that the antigen is a complex protein composed of carbohydrate, lipid and protein and susceptible to heating to 100°C. The determinant of AF9-recognized antigen was not present in the ferritin and carcinoembryonic antigen. Western blotting revealed that AF9-rec-