

# CFA/1结构基因的核苷酸序列测定及其 重组克隆的电镜观察

张兆山 李淑琴 黄翠芬

(军事医学科学院生物工程研究所, 北京 100850)

本文对编码CFA/1亚单位的核苷酸序列进行了测定。根据DNA序列推导, CFA/1的蛋白质序列由170个氨基酸组成, 其中23个氨基酸为前导序列即信号肽。发现有3个位置上的氨基酸与由蛋白质测序的结果不同。第37位氨基酸是丙氨酸而不是缬氨酸, 第76位的天冬酰胺由天冬氨酸所取代, 第97位的氨基酸不是丙氨酸而是丝氨酸。CFA/1亚单位具有非常强的疏水性, 47%的氨基酸为疏水氨基酸。

CFA/1结构基因有典型的SD序列。CFA/1启动子有-10序列TACAAT, 但没有发现-35序列。CFA/1结构基因的G/C比例占40%左右。

电镜观察结果证实, 受体菌*E. coli* C600本身没有菌毛, 然而一旦转化进CFA/1重组质粒, 在其菌体表面则生长有浓密的菌毛。

**关键词** ETEC; CFA/1; 核苷酸序列; 菌毛

肠毒素大肠杆菌 (ETEC) 感染宿主后, 粘附于小肠上皮细胞, 定居繁殖产生肠毒素而致病, 定居因子是ETEC致病的先决条件。迄今为止, 在ETEC中, 对其了解比较清楚的菌毛抗原包括定居因子抗原 I (CFA/1)<sup>[1]</sup>, 定居因子抗原 II<sup>[2]</sup> 和定居因子抗原 IV<sup>[3,4]</sup>。而且不断发现一些新型的定居因子抗原<sup>[5]</sup>。CFA/1是一种占优势血清型的定居因子, 它的基因位于60MDa的大质粒上。在大多数情况下, 该质粒同时编码对热稳定的肠毒素 (ST)<sup>[6]</sup>。丢失了编码CFA/1大质粒的ETEC不再引起腹泻<sup>[7]</sup>。动物试验证明CFA/1抗体在阻止ETEC定居及预防腹泻方面起重要作用<sup>[8]</sup>。我们在获得了高效表达的CFA/1重组克隆的基础上, 对编码CFA/1蛋白质的核苷酸序列进行了分析, 并在电镜下观察到了重组克隆所产生的菌毛。

## 材料与方 法

### (一) 菌株与质粒

受体菌 *E. coli* C600 (F<sup>-</sup>, thi-1, thr-1, leu B6, lac Y1, ton A21, sup E44)。

CFA/1重组质粒 pZLH88由本研究室构建<sup>[9]</sup>, 它含有CFA/1的结构基因和调节基因。

### (二) 培养基

LB肉汤培养基, 培养重组质粒时有50μg/ml Ap。

CFA琼脂含有下列成分: 1% 水解酪蛋白, 0.15% 酵母粉, 0.005% MgSO<sub>4</sub>·7H<sub>2</sub>O, 0.0005% MnCl<sub>2</sub> 和 1.3% 琼脂粉 (青岛水产品加工厂产品), pH7.4。

### (三) 酶试剂

DNA 序列分析试剂盒购自美国 USA

本文于1991年12月14日收到。

生化试剂公司。RNase A (Boehringer Mannheim)。

#### (四)质粒DNA的抽提

基本按照Birnboim和Doly<sup>[10]</sup>的方法抽提质粒DNA。碱变性菌液用高盐中和后,离心所得上清用RNase处理,再用等体积苯酚-氯仿抽提两次,加2倍无水乙醇沉淀DNA。从5ml培养物中得到的DNA,溶于50 $\mu$ l无菌水中,取其中15 $\mu$ l做为序列分析的模板DNA。

#### (五)DNA序列分析

采用Sanger氏双脱氧法<sup>[11]</sup>。用碱变性法使双链质粒DNA变性。15 $\mu$ l DNA样品中加入4 $\mu$ l 1mol/L NaOH/1mmol/L EDTA,室温维持5min后,加入2 $\mu$ l预冷的2mol/L NH<sub>4</sub>Ac,立即加入55 $\mu$ l无水乙醇,液氮中放置10min或更长。4 $^{\circ}$ C下,15000r/min离心15min,除去液相,用70%乙醇洗一次,真空抽干待用。

上述变性DNA样品溶于7 $\mu$ l灭菌双蒸水中,再加2 $\mu$ l序列分析缓冲液和1 $\mu$ l引物(0.1 $\mu$ g/ $\mu$ l)。65 $^{\circ}$ C保温2min,而后缓慢冷却至30 $^{\circ}$ C左右。其他步骤按试剂盒说明书操作。

反应混合物在7%聚丙烯酰胺尿素变性胶上电泳。电泳后,使X-光片在-70 $^{\circ}$ C下放射自显影12—24h。

#### (六)电镜观察

含有重组质粒pZLH88的*E. coli* C600在CFA琼脂上37 $^{\circ}$ C培养20h,刮取少许菌苔用生理盐水悬浮,涂布在用碳包被的铜网上,除去过剩的液体后,立即用1% (w/v)磷钨酸(预先用KOH调pH至6.8)进行负染。在80kV工作电压下,用电镜观察细菌菌体表面的菌毛。

## 结 果 与 讨 论

### (一)CFA/1结构基因的序列分析

pZLH88是含有CFA/1结构基因和调节基因的重组质粒。该重组质粒在*E. coli* C600中,不仅能稳定地传代,而且能有效地表达CFA/1蛋白质<sup>[9]</sup>。为了分析其结构基因与蛋白质功能之间的关系,我们首先使用根据CFA/1亚单位第1—7位氨基酸推导的寡核苷酸5'-GTAGAG-AAGAATATTACTGTA-3'作为引物,对编码N-端的核苷酸序列进行了测定;继而使用编码第62—66位氨基酸的核苷酸序列5'-CTTGGTGATACACCA-3',编码第119—123位氨基酸的寡核苷酸5'-ATTAGCGCTGCACCT-3'和5'-AGG-ATCAACACTAGC-3'(编码第9—13位氨基酸的互补核苷酸序列)作为引物,对包括启动子和终止子区域在内的CFA/1结构基因进行了序列分析,测定结果见图1。

CFA/1亚单位的蛋白质序列是由147个氨基酸组成<sup>[12]</sup>。根据核苷酸序列,我们发现CFA/1基因有一个编码170个密码子的510bp的可读框(ORF)。显然,在CFA/1亚单位前导的23个氨基酸残基是CFA/1的信号肽,在蛋白质转运过程中,在细胞质膜从CFA/1前体上切除下来。根据核苷酸序列推导出的CFA/1氨基酸序列与蛋白质测序结果<sup>[12]</sup>相比较,发现有3个位置的氨基酸有差异。第37位氨基酸是丙氨酸而不是缬氨酸,第76位的天冬酰胺被天冬氨酸所取代,而第97位的氨基酸是丝氨酸而不是丙氨酸。Karjalainen等<sup>[13]</sup>在测定CFA/1结构基因的核苷酸序列时,也发现根据核苷酸推导出的氨基酸序列,在第76位和97位与蛋白质测序结果有和本文相同的差异。因为Karjalainen等得到的CFA/1基因克隆来源于Klemm用来提取CFA/1蛋白质的相同菌株(H10407),这两位氨基酸的差异或许是由于蛋白质测序的误差所致。至于本文报道的

		-10	SD	
		ACCATGAAGGCATAGAAAAAGAGCAAGGGCTAATACAAATIAAAGCTTCCTTGATTAGTCTATATACTATAGAGGTTCTIA		81
		ATG AAA TTT AAA AAA ACT ATT GGT GCA ATG GCT CTG ACC ACA ATG TTT GTA GCA GTG AGT		141
		<u>Met Lys Phe Lys Lys Thr Ile Gly Ala Met Ala Leu Thr Thr Met Phe Val Ala Val Ser</u>		
		GCT TCA GCA GTA GAG AAA AAT ATT ACT GTA ACA GCT AGT GTT BAT CCT GCA ATT GAT CTT		201
		<u>Ala Ser Ala Val Glu Lys Asn Ile Thr Val Thr Ala Ser Val Asp Pro Ala Ile Asp Leu</u>		
		TTG CAA GCT GAT GGC AAT GCT CTG GCA TCA GCT GTA AAG TTA GCT TAT TCT CCC GCA TCA		261
		<u>Leu Gln Ala Asp Gly Asn Ala Leu Pro Ser Ala Val Lys Leu Ala Tyr Ser Pro Ala Ser</u>		
		AAA ACT. TTT GAA AGT TAC AGA GTA ATG ACT CAA GTT CAT ACA AAC GAT GCA ACT AAA AAA		321
		<u>Lys Thr Phe Glu Ser Tyr Arg Val Met Thr Gln Val His Thr Asn Ala Thr Lys Lys</u>		
		Eco RI		
		GTA ATT GTT AAA CTT GCT GAT ACA CCA CAG CTT ACA GAT GTT CTG AAT TCA ACT GTT CAA		381
		<u>Val Ile Val Lys Leu Ala Asp Thr Pro Gln Leu Thr Asp Val Leu Asn Thr Val Gln</u>		
		ATG GCT ATC AGT GTG TCA TGG GGA GGA CAA GTA TTA TCT ACA ACA GCC AAA GAA TTT GAA		441
		<u>Met Pro Ile Ser Val Ser Trp Gly Gly Gln Val Leu Ser Thr Thr Ala Lys Glu Phe Glu</u>		
		GCT GCT GCT TTG GGA TAT TCT GCA TGC GGT GTA AAT GGC GTA TCA TCT TCT CAA GAG TTA		501
		<u>Ala Ala Ala Leu Gly Tyr Ser Ala Ser Gly Val Asn Gly Val Ser Ser Ser Gln Glu Leu</u>		
		Pst I		
		GTA ATT AGC GCT GCA CCT AAA ACT GCC GGT ACC GCC GCA ACT GCA GGA AAC TAT TCA GGA		561
		<u>Val Ile Ser Ala Ala Pro Lys Thr Ala Gly Thr Ala Pro Thr Ala Gly Asn Tyr Ser Gly</u>		
		Bam HI		
		GTA GTA TCT CTT GTA ATG ACT TTG GGA TCC TGA TTGTTTATTIAAAAAGGAGGGGGTATGTTTATACAT		630
		<u>Val Val Ser Leu Val Met Thr Leu Gly Ser ***</u>		
		CCCTCCTTTTTTCATGGAAATGAATTTATATGAAGCATAAAAAAAGAATAGATTAGTCGTAGGCATTICA		700

图1 CFA/1基因的核苷酸序列

Fig.1 Nucleotide sequence of the CFA/1 gene and its flanking regions

Hydrophobic amino acids are underlined. The three shaded amino acids are the ones that differ from the protein sequence. The arrow between codons 23 and 24 indicates the cleavage site of the CFA/1 precursor. SD indicates the Shine-Delgarno sequence. \*\*\* indicates the termination codon. The arrows under the 3' end of the sequence are regions of approximate dyad symmetry that may be involved in rho-independent termination of transcription.

第37位氨基酸的差别,很可能是由于核苷酸的突变 T→C 所引起的,因为本文所用 CFA/1质粒与他们所用的菌株属于不同血清型的肠毒素大肠杆菌。

我们从图1中不难发现,CFA/1氨基酸序列的一个显著特征是具有非常强的疏水性。在前体的170个氨基酸中80个是疏水氨基酸,在CFA/1亚单位的147个氨基酸中有66个是疏水的。CFA/1亚单位正是靠它们之间的疏水相互作用,才能容易地聚在一起。

DNA序列测定结果表明,CFA/1结

构基因有典型的SD序列AGGAG,其启动子具有-10序列TACAAT,但没有发现-35序列。CFA/1启动子的这个特征似乎可以解释CFA/1结构基因象猪源ETEC的定居因子K88那样,只有在另外的调节基因产物的参与下,才能有效地表达<sup>[14]</sup>。在CFA/1基因终止密码子下游的第45—49和第64—68碱基对,是反向重复序列,其后紧接着是串联的8个'A',这是一种不依赖ρ的转录终止子的特征。

## (二)电镜观察

CFA/1蛋白质的合成和菌毛的装配是

一个十分复杂的过程,要涉及到多种基因产物的协同作用。我们在用ELISA确定了CFA/1重组质粒pZLH88能有效表达的基础上,又用电镜对含有重组质粒的受体菌进行了观察。受体菌 *E. coli* C600本身没有菌毛,一旦转化进 pZLH88后,在其菌体表面则长有浓密纤细的菌毛(图2)。ETEC的定居因子不是引起腹泻的直接原因,但却是一种很理想的免疫原。具有CFAs的菌株可在宿主体内定居繁殖,刺激抗体的产生。从理论上讲,重组菌株具有遗传背景清楚,无生物毒性及免疫原蛋白产量高等特点。因此,可望此重组克隆

株能做为预防腹泻的活菌疫苗候选株。

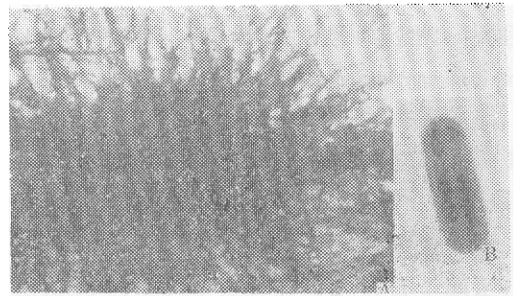


图2 CFA/1重组克隆的菌毛观察

Fig.2 Fimbriate from *E. coli* C600 with pZLH88 expressing CFA/1

A. *E. coli* C600(pZLH88) × 225000  
B. *E. coli* C600 × 6500

### 参 考 文 献

- [1] Evans, D.G. et al.: *Infect. Immun.*, 19:727, 1978.
- [2] Evans, D.G. et al.: *Eur. J. Clin. Microbiol.*, 1:178, 1982.
- [3] Thomas, L.V. et al.: *J. Gen. Microbiol.*, 131:2319, 1985.
- [4] McInnell, M.M. et al.: *Infect. Immun.*, 56:1974, 1988.
- [5] Tacket, C.O. et al.: *Infect. Immun.*, 55:1063, 1987.
- [6] Reis, M.H.L. et al.: *FEMS Microbiol. Lett.*, 22:253, 1980.
- [7] Echeverria, P. et al.: *Infect. Immun.*, 51:626, 1986.
- [8] Ahren, C. et al.: *Infect. Immun.*, 38:74, 1982.
- [9] 张兆山等: *中华微生物学和免疫学杂志*, 12:157, 1992.
- [10] Birnboim, H.C. et al.: *Nucleic Acids. Res.*, 7:513, 1979.
- [11] Sanger, F. et al.: *Proc. Natl. Acad. Sci. USA*, 74:5463, 1977.
- [12] Klemm, P.: *Eur. J. Biochem.*, 124:339, 1982.
- [13] Karjalainen, T.K. et al.: *Infect. Immun.*, 57:1126, 1989.
- [14] Mooi, F.R. et al.: Genetics and biogenesis of the K88ab and K99 fimbrial adhesins, p13-26, 1986. In D.I. Lark(ed), Protein-carbohydrate interactions. Academic Press, Inc. (London), Ltd., London.

## Nucleotide Sequence of the Colonization Factor Antigen 1 Gene of ETEC and Examination of CFA/1 Recombinant Clone by Electron Microscopy

Zhang Zhaoshan Li Shuqin Huang Cuifen

(Institute of Biotechnology, Academy of Military Medical Sciences, Beijing 100850)

The structural gene of the colonization factor antigen 1 (CFA/1) has been sequenced. The amino acid sequence of CFA/1 precursor deduced from the nucleotide sequence is composed of 170 amino acids. The first 23 amino acids are considered to be the signal peptide of the CFA/1 protein. On the basis of the nucleotide sequence, three amino acids are different from the protein sequence, amino acid 37 is an alanine instead of a valine, amino acid 76 is an aspartic acid instead of an asparagine and 97 is a serine instead of an alanine. The CFA/1 is very hydrophobic. Among the total 170 amino acids, 47% of them are hydrophobic amino acids. CFA/1 gene has a typical Shine-Dalgarno sequence AGGA. CFA/1 gene promoter, has a sequence TACAAT that was assigned as the -10 sequence, but does not have a sequence that can be designated the -35 promoter sequence. The G+C ratio of the CFA/1 gene is 40%. Examination of negatively-stained preparations of cultures by electron microscopy showed that strain *E. coli* C600 without recombinant plasmid coding for CFA/1 is non-fimbriae, but the cell carrying the plasmid has thick pili on its surface.

**Key words** ETEC; CFA/1; nucleotide sequence; fimbriate