# 圈卷产色链霉菌尼可霉素生物合成基因

# — sanH 和 sanI 的研究<sup>\*</sup>

#### 陈 蔚 田宇清 杨海花 谭华荣\*\*

(中国科学院微生物研究所 北京 100080)

摘要通过反向遗传学方法克隆到圈卷产色链霉菌尼可霉素生物合成基因簇中约7.0kb的 DNA片段。该片段除含有尼可霉素生物合成基因 sanF外,对 sanF上游约2.2kb的 Bgl [] DNA片段进行序列测定及分析表明,还含有两个完整的开放阅读框(ORF)。ORF1 由 1233 个核苷酸组成 ORF2 由 195 个核苷酸组成,它们分别编码由 410 个氨基酸残基和 64 个氨基 酸残基组成的蛋白质,依次命名为 sanH和 sanI。蛋白序列数据库比较结果表明,SanH和 SanI与浅灰链霉菌(Streptomyces griseolus)中共转录的细胞色素 P450(cytochrome P450)和铁 氧还蛋白(ferredoxin)有较高的同源性,一致性分别为 46%和 56%,相似性分别为 62%和 70%。基因功能研究表明,sanH基因的破坏虽不影响圈卷产色链霉菌产生的尼可霉素的生 物活性,但该基因可能参与了尼可霉素羟基化反应的生物合成。 关键词 圈卷产色链霉菌,尼可霉素生物合成基因,基因破坏 中图分类号 Q78 文献标识码:A 文章编号 10001-6209(2000)06-0598-04

尼可霉素属核苷肽基抗生素,包括二十多种活性单组分,其中 Nikkomycin X 和 Z 为 主要生物活性组分。尼可霉素与几丁质合成酶的天然底物 UDP-N-乙酰葡萄糖胺具有类 似的结构,而竞争性地抑制几丁质合成酶,表现出抗真菌、杀昆虫和杀螨虫等多种生物活 性<sup>11</sup>。同时,它对哺乳动物、蜜蜂、植物和细菌等无毒害作用或毒性极低且在自然界中易 被降解,而被认为是一种理想的农用抗生素。最近的研究表明它对一些危及生命的机会 致病真菌也有明显的抑制作用<sup>21</sup>,可用于人体深部真菌感染疾病的治疗。尼可霉素作为 农用及医用抗生素的开发研究正在进行中。根据标记前体掺入实验及对阻断突变株产物 的分析,表明尼可霉素的肽基和核苷两部分是分开合成,然后通过肽键相连接的。肽基中 的吡啶环及羟甲基 C 原子来自 L—赖氨酸,尼可霉素的氨基已糖醛酸部分被认为同多氧 霉素一样由来源于尿嘧啶和磷酸烯醇式丙酮酸的辛酸前体衍生而来。核苷部分的碱基分 别来源于组氨酸(Nikkomycin X)和尿嘧啶(Nikkomycin Z)<sup>31</sup>。虽然对尼可霉素生物合成 途径有了初步了解,但至今仍未研究清楚,对它的酶学反应也了解甚少,因此克隆与尼可 霉素生物合成相关的基因将有助于阐明尼可霉素的生物合成途径。

我国东北土壤中分离到的圈卷产色链霉菌 7100,是尼可霉素高产菌株。它产生的尼可霉素主要组分为 Nikkomycin X 和 Z。本研究组以圈卷产色链霉菌 7100 为研究材料,

\*\*通讯作者

作者简介 陈 蔚(1973-),女,湖北武汉人,中国科学院微生物研究所,博士生,主要从事链霉菌分子遗传学的研究

收稿日期:1999-01-10,修回日期:2000-06-0€© 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn

<sup>\*</sup> 国家自然科学基金重点资助项目(39830010)

599

通过反向遗传学方法克隆了尼可霉素生物合成基因簇中约 7.0kb 的 DNA 片段,其中 sanF基因是尼可霉素生物合成中起重要作用的基因(另文待发表)。由于抗生素生物合 成基因是紧密连锁的,故推测 sanF基因上游的基因也可能参与了尼可霉素的生物合成。 本文报道了 sanF基因上游的 sanH和 sanI两个基因的结构和功能。

## 1 材料和方法

1.1 材料

**1.1.1** 菌种和质粒: *E. coli* JM109, *E. coli* JM101, 圈卷产色链霉菌(*Streptomyces ansochromogenes*) 7100, 赤星灰霉, 质粒 Bluescript M13<sup>-</sup>, 帮助噬菌体 KO7 及用于基因破坏的质粒 pKC1139 均为本实验室保存。

**1.1.2** 培养基 :LB 培养基及 2 × YT 按文献[4] 配制。赤星灰霉生长用土豆培养基 (PDA)按文献 5 配制。链霉菌液体生长培养基(YEME),原生质体再生培养基(R2YE) 及基本培养基(MM)均按文献 6 配制。

1.1.3 酶、抗生素及化学试剂:实验所用限制性内切酶均为 Boehringer 公司产品 核酸外 切酶  $Exo \parallel$  绿豆核酸酶 ,T4 DNA 连接酶为华美生物工程公司产品。DNA 测序用 Taq-Track 测序试剂盒为 Promega 公司产品。氨苄青霉素、四环素、氯霉素为华美生物工程公 司产品 ,贮存液浓度分别为 100mg/mL、15mg/mL 和 25mg/mL ,使用终浓度分别为 100 $\mu$ g/mL、10 $\mu$ g/mL 及 25 $\mu$ g/mL。 Apramycin( 阿普霉素 )由加拿大的 Alberta 大学 B. Leskiw 博士赠送 贮存液浓度为 100mg/mL 在 YEME 和 MM 培养基中使用浓度为 5 $\mu$ g/ mL 在 R2YE 培养基中使用浓度为 50 $\mu$ g/mL。所有酶类及抗生素贮存液均于 – 20℃ 冰 箱保存。聚乙二醇( PEG )1000 购于 Merck-Schuchard 公司 ,用于链霉菌原生质体转化。 PEG 8000 购于 Sigma 公司 ,用于单链 DNA 的制备。X-gal 和 IPTG 用于大肠杆菌中阳性 转化子的选择 ,在 LB 培养基中的使用浓度为 40 $\mu$ g/mL。

1.1.4 放射性同位素及非放射性同位素试剂盒 用于 DNA 序列测定的 α-<sup>35</sup>S-dATP 购于 Amersham 公司 ,用于 DNA 探针标记的非同位素地高辛(Digoxigenin-11-dUTP)试剂盒购 于 Boehringer 公司。

1.2 方法

**1.2.1** DNA 的提取和纯化 :染色体 DNA 和质粒 DNA 的提取 ,DNA 的纯化及回收均按 文献 6 方法进行。

1.2.2 大肠杆菌感受态细胞的制备及转化:按文献 4 访法进行。

1.2.3 链霉菌原生质体的制备和转化 参照文献 6 方法进行。

1.2.4 单链 DNA 的制备按文献 7] DNA 序列测定和分析按文献 7 8 进行。

1.2.5 Southern blotting 杂交 按文献 7 方法进行。

1.2.6 尼可霉素的检测:按文献 5 访法进行。

## 2 结果和分析

2.1 目的 DNA 片段的克隆

尼可霉素是 1976 年首先从唐德链霉菌虫分离的 如此 能够重生株和不产尼可霉素的。

突变株的总细胞蛋白双向凝胶电泳发现,野生株中有 10 种蛋白(PI-P10),在不产尼可霉素的突变株中都未检测到,预示这 10 种蛋白可能与尼可霉素的生物合成直接相关。其中 6 种蛋白(P1 2 A 5 6 8)进行了 N 端氨基酸序列测定<sup>[9]</sup>。本文根据唐德链霉菌尼可霉素生物合成相关蛋白 P6 的 N 端氨基酸序列及链霉菌密码子使用的偏好性设计了一对简并引物,并通过 PCR 扩增出一个 111bp 的 DNA 片段。用非放射性同位素地高辛标记 将 该片段制备成 DNA 探针。从在大肠杆菌中构建的圈卷产色链霉菌部分基因文库中克隆 到包括 sanF 基因在内的 7.0kb KpnI 片段,重组质粒命名为 pCW7(另文发表)。pCW7 限制性酶切的琼脂糖凝胶电泳结果表明,EcoRV和 PstI 酶切后都给出了 2 条 DNA 泳带,说明分别有 2 个 EcoRV和 PstI 位点,由于载体上各有一个 EcoRV和 PstI 位点,因此 7.0kb DNA 片段中只有一个 EcoRV和 PstI 切点。BamHI和 BglII 酶切后均给出了 4 条 DNA 泳带,其中一个 BamHI 位点来自载体,故片段中有 3 个 BamHI 位点和四个 BglII 确切 pCW7 后回收了位于 sanF基因上游的 2.2kb DNA 片段。将该片段克隆至质粒 Bluescript M13<sup>-</sup> 的 BamHI 位点上,获得插入方向不同的重组质粒 pCW6 和 pCW11。





#### 2.2 DNA 片段的序列分析

将 pCW6 和 pCW11 分别进行缩小亚克隆,然后挑选长度相差 200bp 的重组质粒,转 化大肠杆菌 JM101 在帮助噬菌体 M13KO7 的作用下,制备单链 DNA。通过双脱氧链终 止法测定了 2.2kb DNA 片段的全序列(图 2)。结果表明,该序列中含有两个紧密连锁的 完整开放阅读框(ORF)(图 3),ORF1 由 1233 个核苷酸组成,ORF2 由 195 个核苷酸组 成,它们分别编码由 411 个氨基酸残基和 64 个氨基酸残基组成的蛋白质,依次命名为 sanH(Streptomyces ansochromogenes nikkomycin gene H)和 sanI(Streptomyces ansochromogenes nikkomycin gene I)。sanH 的翻译起始位点为 297 位碱基处的 ATG 终止密码子 为 1527 位碱基处的 TGA,在 ATG 上游 8 个碱基处有可能的核糖体结合位点 GGGAA; sanI 的起始密码子 GTG 位于 1546 位碱基处,距 sanH 的终止密码子 TGA 有 16 个碱基,终止密码子为 1738 位碱基处的 TGA。在 GTG 上游 8 个碱基处有可能的核糖体接合位 点 GAGGG。

蛋白序列数据库比较结果表明 SanH和 SanI 分别与许多生物中共转录的细胞色素 P450(cytochrome P450)和铁氧还蛋白(ferredoxin)有较高的同源性。其中与浅灰链霉菌 (*Streptomyces griseolus*)的细胞色素 P450 和铁氧还蛋白同源性最高,一致性分别为 46% 和 56% 相似性分别为 62%和 70%。推断 SanH 和 SanI 也是共转录的。

601

BglII

1	AGATCTCT	GCTCG	CCGA	GCAC	CCCI	TGCT	GCGG	GCGG	TCG	CGGC	GGTC	GCGA	ATGC	CGCAC	
61	CCGGTGTG	GGGGC	AGAC	GGTC	TGC	GTCTI	CGCC	GTCC	CCT	CCGA	CCCG	GAG	CAGC	AGCCG	
121	ACGCTGCT	CGGCC	TGCC	GCCGC	CTTC	CTCAC	CGA	GCGC	GGCA	TCGC	CACG	TAC.	AAAC	CTGCCC	,
181	CAGGACAT	AAGGG	TGGT	CGA	GCC	CTGC	GAT	ATC	GCG	TCGG	CAAG	ATC.	AACO	GGGTC	:
2/1	CCCCTCCC	TCCCC		'AGCC			AGGC	CGA	GAGG	AACG	GGAA	CCC.	ACCO	CATGA	7
Z 4 I	60601000	10000	0000	nuot	10110	00001	10000				RBS	OF	$F1 \rightarrow$	- M	
201	00010000		CCAC		M TO CO	COTO	2 ጥጥጥ ፖ	CCA	racc	CCGC	GAGT	GCC	CCAT	rgcaco	2
201			ICCAC	GACE			2111( 2111(	D N	<b>V V</b>	2000 P	E .	č	P N	ИH	
201	T E A			<b>U</b> 107 m m /		r s cmcco	E CDC		CCT		cac	TCC	GCA	reccce	1
301	Georgeor	GAGTE	ATCGC	GAA				F 1		CAGC	<b>D</b>	$\mathbf{v}$	R N	1 P	<i>.</i>
	PPA	E 1		ь Сара			- Y			cccc	NACO		TCG		٦
421	ACGGTCAG	GTGGC	LTGO	SCIG(	JUL .		JUAU			DCCGC D	NAGC V	T.	T. 7		<i>.</i>
	DGQ	VA	F W	Ц	v.	LK	H			, <b>R</b>		CCC			٦
481	CGCGGGGTC	AGCGC		CCCC	CTGC.	ACCCC	GCG.	rrcce	2666						1
	PRV	SZ	A D	R	L	H P	A	F. T	P G	- K		T		ч V ГССЛСС	-
541	GCGCGGCG	ACCGP	AGCG	CGTC	CGCC	GGCT	CACC	ACCC	GCG	GTCC	ATGA	TCC	ACC:	IGGACO	3
	RAA	TE	ER	v	R	RL	т	TJ	R R	S	_M	1	H J		
601	GGGACGAG	CACGO	GGGCI	[CAC	CGCA	.GGAT(	CCTG	ACCG	GCGA	GTTC	TCGC	TGC	GCCG	GATCO	j
	GDE	но	G A	н	R	R I	L	T (	G E	F	S	L	R	RI	
661	CCGCGCAG	GCGCCC	CGCG"	<b>FGTC</b>	CAGG	AGAT	CGTC	GACC	GTAG	CATC	GACO	SAGA	TGC	rggcco	ż
	A A O	R I	PR	v	Q	ΕI	v	D	r s	I	D	E	M I	ĹΑ	
721	CGCCGCÃ	CCGGC	CCGA	CCTC	GTCG	AGCA	CGTG	TCGC.	AGGC	GGTG	ICCG1	CGC	TGG	IGATC:	Ĉ.
	APO	P 2	A D	L	v	Е Н	v	S (	Q A	V V	₽	S	L	V I	
781	GCGAGCT	GCTGGG	GCGT	GCCC	CACG	AGCA	GCGC	CGTG.	ACTT	CCAC	GAGI	GGG	CGG	<b>JCATG</b>	2
	CEL	L (	ςν	P	н	ΕO	R	R 1	DE	' Н	E	W	A	G M	
8/1	TGGTCAG	TOGGT(	CGT	CTCG.	ATAC	AGGA	GCGC	GCCG	CGGC	CTCC	GAC	GCGC	TCA	ACGAC	ſ
041			c v	S	T	OE	R	A	A A	S	D	A	LI	N D	
901	TCCTCCAC		TCGT(	$\sim 2$	GĀGA	AGGA	2222	GGCG	AGCC	CGCC	GAC	GACC	TGA	TCGGC	С
90 I		λ 1	$\mathbf{r}$		E	K E	R	G	F. F	A	D	D	L	ΙG	
061					CACA		caña	ATCA	CGCI	CGAC	GĀG	ATCG	TGG	GCACC	3
901	GGCICAIC		JUAN	0000 0		ت س	VIOD	M	T L	חי	F	T	v	GΤ	-
1001			K N	CCCC	v C C C C C C C C		CACC	ACCC	incal		2270	TCGC	TTGG	GCGTG	G
1071	UCGTUAI	っしょしし	1991		.GGG(	JACCA	GACC	ACCG	100n	10111		1000	100	00010	<u> </u>
1021	3 37 34	т ·	т 17	7	~	U 0	<b>T</b>	77	2 1	vм	т	S	T.	GV	
1001	AVM	L		A	G	H Q	T	T	A I	M N		S TCC1		G V TGCTG	С
1081	A V M TGGCGCT	L GCTGG	L V AGAA	A CCCG	GAGC	H Q CACAA	<b>T</b> GGCA		A I TCGO	N M CGGC(	I CGAC	S TCC1 S	L CGC	G V TGCTG L C	С
1081	A V M TGGCGCT V A L	L GCTGG L	L V AGAA E N		GAGO E	H Q CACAA H K		T CGGA R	A I TCGO I A		I CGAC D	S ICC1 S	L ICGC S	G V TGCTG L C	C
1021 1081 1141	A V M TGGCGCT V A L CCCCGGC	L GCTGG L GATCG	L V AGAA E N AGGA	A CCCG P GATG	GAGO E CTGO	H Q CACAA H K CGCTA	GGCA A CTTC	T CGGA R AGCG	A I ATCGO I P ATCGO	N M CGGC A A FGGA	I CGAC D GAAC	S ICCI S GCG(	L ICGC S ICCG	G V TGCTG L C CCCGG A B	C G
1081 1141	A V M TGGCGCT V A L CCCCGGC P P A	L GCTGG L GATCG I	L V AGAA E N AGGA E E	A CCCG P GATG M	GAGO E CTGO L	H Q CACAA H K CGCTA R Y	T GGCA A CTTC F	T R AGCO S	A I TCGO I Z TCGO V	N M CGGC A A FGGA V E	I CGAC D GAAC N	S ICCI S GCG( A	L ICGC S CCCG P	G V TGCTG L C CCCGG A R	C G C
1021 1081 1141 1201	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC	L GCTGG L GATCG I CGAGG	L V AGAA E N AGGA E E ACAT	A CCCG P GATG M CGCG	GAGG E CTGC L GATCO	H Q CACAA H K CGCTA R Y GGCGG	GGCA A CTTC F CGTC	T CGGA R CAGCO S GACCA	A I ATCGO I P STCGO V ATCCO	N M CGGC A A FGGA V E GCAA	I CGAC D GAAC N GAAC	S S GCG A GAG	L ICGC S CCCG P GCA	G V TGCTG L C CCCGG A R TCGTC	C G G
1081 1141 1201	A V M TGGCGCTC V A L CCCCGGCC P P A TCGCCACC V A T	L GCTGG L GATCG I CGAGG E	L V AGAA E N AGGA E E ACAT D I	A CCCG P GATG M CGCG A	GAGO E GCTGO L GATCO I	H Q CACAA H K CGCTA R Y GGCGG G G	T GGCA A CTTC F CGTC V	T R AGCG S ACCA I	A 1 ATCGO I A STCGO V ATCCO I 1	N M CGGC(A A A FGGA( V E GCAA( R K	I CGAC D GAAC N GAAC N	S ICCI S GCG( A GAG( E CAC(	L CCCC P GCCA GCCA	G V TGCTG L C CCCGG A R TCGTC I V	С G G
1021 1081 1141 1201 1261	A V M TGGCGCTC V A L CCCCGGCC P P A TCGCCAC V A T TGTCCGG	L GCTGG L GATCG I CGAGG E GCTGG	L V AGAA E N AGGA E E ACAT D I CCGC	A CCCG P GATG M CGCG A CGAC	GAGO E CTGO L GATCO I CTGGO	H Q CACAA H K CGCTA R Y GGCGG G G GACGA	GGCA A CTTC F CGTC V CGAC	T R AGCG S AGCCA I GGTCI	A I TCGO TCGO TCGO V TCCO I I TCCO	N M CGGC A A IGGA IGGA SCAA R K GGCA	I CGAC D GAAC N GAAC N CCCG	S ICCI S GCGC A GAGC GACC	L CCCC S CCCCG P GCCA G CGGC	G V TGCTG L C CCCGG A R TCGTC I V TCGAC	C G G T
1081 1141 1201 1261	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G	L GCTGG L GATCG I CGAGG E GCTGG L	L V AGAA E N AGGA E E ACAT D I CCGC	A CCCG P GATG CGCG A CGCG CGAC	GAGO E CTGO L GATCO I CTGGO W	H Q CACAA H K CGCTA R Y GGCGG GGCGG GACGA D C	T GGCA A CTTC F CGTC V CGAC E	T R AGCG S ACCA I GTCI	A 1 TCG TCG TCG TCG TCC TCC F C	N M CGGC A A IGGA IGGA V E GCAA R K GGCA C H	I CGAC D GAAC N GAAC N CCCG	S ICCI S GCG( A GAG( E GAC( O CAC(	L ICGC S ICCCG P IGGCA G IGGC R	G V TGCTG L C CCCGG A R TCGTC I V TCGAC L D	C G G T
1081 1141 1201 1261 1321	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG	L GCTGG L GATCG CGAGG E GCTGG L CGGCCG	L V AGAA E N AGGA E E ACAT D I CCGC A A CGCG	A CCCG P GATG CGCG CGAC A CGAC D CCAC	GAGC E CTGC E ATCC I CTGGC W CCACC	H Q CACAA H K CGCTA R Y GGCGG GGCGG GGCGG GACGA D D GTCGC	T GGCA CTTC F CGTC V CGAC E CTTC	T R AGCG S ACCA I GTCI V CGGGI	A I ATCG I A STCG V ATCC I I STCG F C ACCG	N M CGGC A A FGGA V E GCAA R K GGCA GGCA C H GAGT	I CGAC D GAAC N GAAC N CCCG P CCAC	S S GCGC A GAGC GACC O CAG	L ICGC S CCCG P GGCA G CGGC R IGCC	G V TGCTG L C CCCGG A R TCGTC I V TCGAC L D TCGGG	C G T C
1081 1141 1201 1261 1321	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCG V S C TCGAACG F E R	L GCTGG L GATCG CGAGG E GCTGG CGGCG G	L V AGAA E N AGGA E E ACAT D I CCGC A A CGCG A R	A CCCCG P GATG CGCC A CGCC A CGAC D CCAC H	G GAGC E CTGC I CTGGC W CCACC H	H Q CACAA H K CGCTA R Y GGCGG GGCGG GACGA D D GTCGC V A	T GGCA CTTC F CGTC V CGAC E CTTC	T R CGGA S AGCG S ACCA I GTC1 V CGGG1 G G G G G G G G G G G G G G G G G	A I ATCG I A ATCC I I ATCC F C F C Y C	N M CGGCCA A A IGGAC V E GCAA GGCA GGCA GGCA GAGT GAGT GAGT C	I CGAC D GAAC N GAAC N CCCG P CCAC H	S ICCI S GCG( A GAG( GAC( O CAG)	L ICGC S CCCG P GGCA GGCC R IGCC C	G V TGCTG L C CCCGG A R TCGTC I V TCGAC L D CTCGGG L G	C G T C
1081 1141 1201 1261 1321 1381	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT	L GCTGG GATCG CGAGG E GCTGG CGGCG G GGCCC	L V AGAA E N AGGA E E ACAT D I CCGC CCGC A A CCGCG A R CGCGG A R	A CCCG P GATG CGCG A CGCC CGAC CCAC H GGAG	G GAGC E GATCO I CTGGO W CCACO H GCTGO	H Q CACAA H K CGCTA R Y GGCGG GGCGG GGCGG D D GGCGC V A GAGAT	T GGCA CTTC F CGTG CGAG CTTC F CGTG	T R R SAGCO S SACCA I G GTCI V C G G G G G T T C O	A I TCG TCG TCG TCG TCG F C TCG F C SAGA	N M CGGCO A A IGGAO GGCAA GGCAA GGCAA GGCAA GGCAA GGCAA GGCAA CGCT CGCT	I CGAC D GAAC N GAAC N CCCG P CCAC H GCTG	S GCG( A GAG( E GAC( O CAG) CAG)	L CGGC S GGCA GGCA CGGC R CGGC C CGGC C CGGC	G V TGCTG CCCCGG A R TCGTC I V TCGAC L D TCGGG L G TTCCGG	C G G T C G
1081 1141 1201 1261 1321 1381	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT Q N L	L GCTGG. L GATCG CGAGG. E GCTGG C GGCCC A	L V AGAA E N AGGA E E ACAT D I CCGC CCGC A A CCGCG A R CGCG A R V R V	A CCCCG P GATG CGCCA CCAC CCAC H GGAG	G GAGC E GATCC I GATCC I CTGGC W CCACC H GCTGC L	H Q CACAA H K CGCTA R Y GGCGG GGCGG GGCGG GGCGG O D GTCGC V A GAGAT E J	T GGCA CTTC F CGTG CGAG CTTC F CGTG	T R R GAGCG S GACCA I G GTC1 V C G G G G G T T C G F	A I NTCG( I A STCG) V NTCC( I I STCG( F C SAGA( E !	N M CGGCO A A IGGAO GCAAO GGCAA GGCAA GGCAA GGCAA GGCAA CGCT CGCT	I CGAC D GAAC SAAC N GAAC P CCCG H GCCG H GCTG	S GCGC A GAGC GAGC GACC O CAGC Q CGCC C R	L ICGC S CCCG P GGCA G CGCC R IGCC C C C C C C C C C C C C C C C C C	G V TGCTG CCCGG A R TCGTC I V TCGAC L D TCGGG L G TTCGG V P	C G G T C G
1081 1141 1201 1261 1321 1381 1441	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT Q N L GCCTGTC	L GATCG L GATCG CGAGG G GCTGG G GCCCC A GCTGG	L V AGAA E N AGGA E E ACAT D I CCGCG A A CGCGCG A R CGCGT R V CCGT	A CCCCG P GATG CGCCG A CCCAC CCAC H GGAG CCAC F GCCC	G GAGC E GATCO I CTGGO W CCACO H GCTGO L CGCCO	H Q CACAA H K CGCTA R Y GGCGG G GGCGG G GGCGG V A GAGAT E J GAGG	GGCA A CTTC CGTG CGTG CGTG CGTG CGTG V CGTG	T R R GACCA S GACCA I GGTCI V CGGGI G GTTCC F GCCGI	A I NTCG I A STCG TCG V NTCC I I TCG F C SAGA E ZAGA	N M CGGCO A A IGGAO GGCAO GGCAO GGCAO CGCTO CGCTO I L AGGAO	I CGAC D GAAC SAAC N CCCG CCAC H GCTG CCAC	S GCGC A GAGC GAGC GACC O CAGC CAGC CGCC R GCCC	L ICGC S GCCA G GCA CGCC R IGCC C CGCC R C GGGA	G V TGCTG L C CCCCGG A R TCGGC I V TCGAC L D TCCGGG L G TTCCGG V P NTCTAC	C G G T C G G
1081 1141 1201 1261 1321 1381 1441	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S C TCGAACG F E R AGAACCT Q N L GCCTGTC G L S	L GCTGG GATCG CGAGG GCTGG G GCTGG GCTGG L	L V AGAA E N AGGA E E ACAT D I CCGCG A A CGCGCG A R CGCGT R V CCCGT A V	A CCCG P GATG CGCG A CGAC A CCAC H GGAG F CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCCCG P CCCCG P CCCCG P CCCCG P CCCCG P CCCCG P CCCCG P CCCCG P CCCCG P CCCCG P CCCCG P CCCCG C C CCCG C C C C	G GAGC E GATCO I CTGGO W CCACO H GCTGO L CGCCO A	H Q CACAA H K CGCTA R Y GGCGG G G GACGA D C GACGA V A GAGAT E I GAGGA E E	T GGCA F CGTG CGTG CGTG CGTG V GCTG L	T R R GACCA S GACCA I GGTCI V CGGGI G GTTCC F GCCGI	A 1 ATCG I 4 TCG TCG V 1 ATCC I 1 TCG F 0 TACG Y 0 SAGA E 1 TACA Y 1	N M CGGCO A A IGGAO V E GGCAO GGCAO GGCAO CGCTO CGCTO I L AGGAO K D	I CGAC D GAAC N GAAC P CCCG CCAC H GCTG CCAC	S ICCI S GCGC GAGC GACC O CAGC CAGC CGCC R GCCC	L ICGC S ICCG F IGCC R IGCC C C GGG R GGGA G GGGA	G V TGCTG L C CCCCGG A R TCGTC I V TCGAC L D TTCGGG L G FTTCCGG V P TTCTAC I Y	C G G T C G G G
1081 1141 1201 1261 1321 1381 1441 1501	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA	L GCTGG GATCG CGAGG GCTGG GCTGG GGCCC A GCTGG CCGGG L CCGGGG	L V AGAA E N AGGA E E ACAT D I CCGC CCGC CCGCG CCGCG R V CCGT R V CCGT A V CCGT	A CCCG P GATG CGCG A CGAC CCAC CCAC CCAC CCAC CC	G GAGC E GATCO I CTGGO W CCACO H GCTGO L CGCCO A SAAC	H Q CACAA H K CGCTAA GGCGGG GGCGG GGCGA D C GACGA V A GAGAT E J GAGGA E E TGCTG	T GGCA CTTC F CGTG CGAG CTTC CGTG CGTG CGTG CGTG L GAGGI	T R R SAGCG SACCA I SGTCI V CGGGI G CGGI F SCCGI P AGAGC	A 1 NTCG I 4 STCG V 1 NTCC I 1 STCG F 0 SAGA Y 0 SAGA F 2 SAGA Y 1 SAGA	N M CGGCC A A IGGAC GGCAC GGCAC GGCAC GGCAC GGCAC CGCT CGCT	I CGAC D GAAC N GCCG CCAC GCTG CCAC CCAC CCAC	S ICCI S GCGC GAGC GAGC O CAGC CGCC R GCCC A GCAC	L CGCC S GCCG GGCA CGGC CGGC CGGC CGGC CG	G V TGCTG L C CCCGG A R TCGTC I V TTCGC L D TTCGGG L G TTCCGG V P CTCCCG V P CACCGC	C G T C G G G
1081 1141 1201 1261 1321 1381 1441 1501	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y	L GCTGG GATCG CGAGG CGAGG CGCCC A GCTGG GCCCC A GCTGG CCGGG R	L V AGAA E N AGGA E E ACAT D I CCGC A A CCGCG A R CCGCT R V CCGT A V CCGT A V CCGT A V	A CCCG P GATG CGAC A CGAC CCAC CCAC CCAC CCAC CC	GAGC E CTGC E CTGC CTGC C CCACC H GCTGC C CGCCC A SAAC' N	H Q CACAA H K CGCTA R Y GGCGG G GGCGG GGCGG V A GAGAI E J GAGGA E E TGCTG C Y	T GGCA CTTC CGTC CGTC CGTC CGTC CGTC CGT	T R R SAGCG SACCA I SGTCI V CGGGI G TTCC F SCCGI P AGAGC R R SCCGI	A 1 ATCG I 2 TCG V ATCC I 1 TCG F 6 ACG SAGA E 2 TACA Y 1 SAGA SAGA	N M CGGGCO A A IGGAO SGCAO SGCAO SGCAO SGCAO SGCAO SGCAO SGCAO SGCTO CGCTO SGC	I CGAC D GAAC N GAAC N CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC C CCAC	S ICCI S GCGC A GACC O CAGC CGCC R GCCC A GCCA CGCA CGCA C CGCA	L CCCC S CCCCC P GCCA CCCC CCCC CCCCC CCCCC CCCCC CCCCCCCC	G V TGCTG L C CCCCGG A R TCGTC I V TCGAC L D TCCGG L G TTCCGG V P TCTAC CACCGC I Y CACCGC	G G G G G G G G G G
1081 1141 1201 1261 1321 1381 1441 1501 1561	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGGG	L GCTGG GATCG CGAGG CGAGG CGCCG GGCCCC A GCTGG CCGGG CCGGG GGCCG GGCCG	L V AGAA E N AGGA E E ACAT D L CCGC A A CCGCG A R CCGCG A V CCGT A V CCGT A V CCGT C CCGT C CCGT C CCGT C CCGT C CCGT C C CCGT C C CCGC C C C	A CCCG P GATG CGCG CGCG CGCC H CGCAC H CGCAC H CGCAC CGCC V P CGCCC V CGCCC V CGCCC V CGCCC V CGCCC CCGCG A CCCCG A CCCCG A CCCCG A CCCCG A CCCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG C CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG A CGCCG C CGCCG A CGCCG C CGCCG A CGCCG C CGCCG C CGCCG C CGCCG C CGCCG C CGCCG C CCCCC C CCCCC C CCCCC C CCCCC C CCCCC	GAGC E GATCC L GATCC I CTGGC W CCACC H GCTGC L CGCCC A GAACT N GCGCC N	H Q CACAA H K CGCTA R Y GGCGGG G G GACGA D D C GACGA C V A GAGAI E J GAGGA E E C C C C	T GGCA A CTTC CGTC CGTC CGTC CGTC CGTC C	T R R SACCA	A 1 ATCG I 2 TCG V ATCC I 1 TCG F C ACG Y C ACG Y C ACG Y 1 ACA Y 1 ACA Y 1 ACA Y 1 ACA Y 1 ACA Y 1 ACC Y C Y C Y C Y C Y C Y C Y C Y C Y C Y	N M CGGCCA IGGAC V E GGCAC R K GGCAC F H GAGT CGCT I I L AGGA K D ACGCC ORF2 GTGG	I CGAC D GAAC N GAAC N CCC CCC CCC CCC CCC CCC CCC CCC CC	S ICCI S GCGC GAGC GACC O CGCC CGCC CGCC CGCC CG	L CGCC S GCCG GGCA GGCA CGCC CGCC CGCC CG	G V TGCTG CCCGG A R TCGTC I V TCGAC L D L G TTCGGG V P NTCTAC I Y ACCGC T A CTTCGA	C G G T C G G G C C
1081 1141 1201 1261 1321 1381 1441 1501 1561	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT C N L GCCTGTC G L S GCATCTA G I Y GACAGGG D R	L GCTGG GATGG GCTGG GCTGG GCTGG GCTGG GCTGG CCGGG CCGGG GCCGG GCCGG GCCGG	L V AGAA E N AGGA E E CCGC A A CCGCC A A CCGCC A C CCGT A V CCGT A V CCGT C C C C C C C C C	A CCCG P GATG CGCG CGCG A CGAC CGAC H GGAG CCAC H GGAG CGCC P GCCC C CGCC C CGCC C CCAC C CCAC C CCCG C CCCG A CGCCG C CGCCG A C CGCCG A C CGCCG A C CCCCG A C CGCCG A C CCCCG A C C CCCC A C C CCCC A C C C C	GAGC E GAGC E CTGC I CTGGC W CCACC H GCTGC L CGCCC A CACC C CACC C CCACC C C CCACC C C CCACC C C CCACC C C CCACC C C C CCACC C C C C C C C C C C C C C C C C C C C	H Q CACAA H K CGCTA R Y GGCGGG G G GACGA D D GACGA C V A GAGAT E I GAGGA E E C C C C C C C C C C C	T GGCA CTTC CTTC CGTC CGTC CGTC CGTC GCTC CGCC CGCC CCCC CCCC CCCC CCCC CCCC CCCCC CCCCC CCCC	T CGGA R CGGGA GACCA GACCA GACCA GACCA CGGA F GCCGA R CGGA CGAC C CGACA CGACA CACACA CCGA CCGA CCGA CCGA CCCCA CCCCCA CCCCA CCCCA CCCCCA CCCCA CCCCCA CCCCCA CCCCCA CCCCA CCCCCA CCCCCA CCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCA CCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCA CCCCCC	A I TCG I Z TCG V I C TCG F C ACG Y C ACG Y C ACG Y C C C C C C C C C C C C C	N M CGGC( A A IGGA( V E GCA( R K GGCA( R K GGCA( CGCT) I I L AGGA K CGCT) I I L AGGA K CGCT( GTGG V	$\begin{array}{c} \mathbf{I} \\ \mathbf{CGAC} \\ \mathbf{D} \\ \mathbf{GAAC} \\ \mathbf{D} \\ \mathbf{GAAC} \\ \mathbf{N} \\ \mathbf{GAAC} \\ \mathbf{N} \\ \mathbf{CCCG} \\ \mathbf{P} \\ \mathbf{CCAC} \\ \mathbf{H} \\ \mathbf{GCTG} \\ \mathbf{CGAC} \\ \mathbf{D} \\ \mathbf{CCGT} \\ \mathbf{CGCC} \\ \mathbf{A} \\ \mathbf{F} \end{array}$	S ICCI S GCG( A GAG( C C C C C C C C C C C C C C C C C C C	L CCCC S SCCCC P GCCA CCCC R CCCC CCCC CCCC CCCC CCCC CC	G V TGCTG L C CCCGG A R TTCGTC I V TTCGAC L D TTCGGG L G TTCGGG V P TTCTCGA V CACCGC T A CTTCGA	C G G T C G G G C C
1081 1141 1201 1261 1321 1381 1441 1501 1561 1621	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCG V S C TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGG D R CAGGACG	L GCTGG GATGG CGAGG C GGCGG G GGCCC A GGCGGG C CCGGG C CCGGG C C GGCGG C C C GGCGG C C C C C G C C C C G C C G C C G C C G C C G C C G C	L V AGAA E N AGGA E E CCGC A A CCGCC A A CCGCC A R CCGT A V CCGT A V CCGT A V CCGT A V CCGT A V CCGT A V CCGT A V CCGT A V CCGC C C C C C C C C C C C C C C C C C	A CCCC P GATG CGCC A CCGCC A CCGCC C GCCC C CCCC C CCCC C CCCC C CCCC C CCCC C	G GAGC E GATCC I GATCC I CTGGC W CCACC H GCCCC A GCCCC C GCCCC C GCCCC GAAC C GCCCC GAAC C GCCCC GACCC C GCCCCC GACCC C C C	H Q CACAA H K CGCTAA R Y GGCGC G G GACGA D L GTCGC V A GAGGAT E I GAGGA C C C C C C C C C C C C C C C C C C C	T GGCA CTTC CGTC CGTC CGTC CGTC CGTC CGT	T RGGA S ACCA S GCCA S GCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S CCCA S S CCCA S S CCCA S S CCCA S S CCCA S S CCCA S S CCCA S S CCCA S S CCCA S C C C S C C C S C C C S C C C S C C C S C C C S C C S C C C S C C C S C	A I TCG I A TCG V I C TCG F C ACG Y C ACG Y C ACG Y C C C C C C C C C C C C C	N M CGGCC IGGA V E GCAA GGCA GGCA GGCA GGCA GGCA GGCA CGCC I I L A A CGCC O R Z CCGG V CCGG	I CGAC D GAAC D GAAC N CCCG CCCG CCCG CCCG CCGC A CCGC A CCGC A CCGC	S ICCI S GCG( A GAC( O C GCCI C C C C	L CCGC S CCCG G G CCGGC C CCGGC C CCGGC C C C	G V TGCTG L C CCCGGG A R TTCGTC I V TTCGGG L D TTCGGG I C ACCGCC F C CCGGCC	C G G T C G G C G
1081 1141 1201 1261 1321 1381 1441 1501 1561 1621	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGGG D R CAGGACG O D	L GCTGG GATCG E CGAGG G GCCCC A GCCGGG C CCGGG C R GCCGGG C R GCCGGG C R GCCGGG C R GCCGG C C R GCCGG C C R GCCGG C C R GCCGG C C C GCCG C C C C	L V AGAA E N AGGA E E ACAT D I CCGCG A A CCGCG A V CCGCT CCCT CCCT CCCC C CCCCT CCCC C CCCCT CCCC C CCCCC C CCCCC C CCCCC C CCCCC C	A CCCG GATG CGCC CGCC CGCC CGCC GGAC CGCC CGGTG CGGTG CGGTG CCGGT CGGTG CCGGT CGGTG CCGGT CGGTG CCGGT CGGTG CCGCC CCCCC CGCC CCCC CCCC CGCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC	GAGC EGAGC EATCC I CTGGC W CCACC H CCACC H CCACC CACC CACC CACC	H Q CACAA H K GGCTAA R Y GGCGGG G G GACGA D E GTCGC V A GAGAT E I GAGGA E I GAGGA C C C C C C C C C C C C C C C C C C C	T GGA CTTC F CGTG CGTG CGTG CGTG CGTG CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTGA CGTG CGTG	T R CGGA SACCA SAC	A I ATCG TCG TCG TCG TCG F C ATCG F C ACCG F C ACCG Y C ACCG Y C ACCA Y I ACCA Y I ACCA TACA SGGC CCTG GGAT D	N MC CGGCC A A FIGGAG V E GCCAG GGCAG GGCAG GGCAG GGCAG CGCC CGCC V CGCC V CCGG P	I CGAC D CAAC D CCCG D CCCG C CCGC C CGAC C CGCC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCGC C CCCG C	S GCGCC A GAGCC GACC CCGCC A GCCA CCGCC A GCCA CCGCC A GCCA CCGCC A CCGCC A CCGCC A CCGCC A CCGCC A CCGCC A CCGCC A A A CCGCC A A A A A A A A A A A A A	L CCGC S CCCG P GGCA C GGCA C C C C C C C C C C C C C C	G V TGCTG CCCCGG A R TCGTC I V TCGAC L D TCCGG V P CACCGC T A CACCGC T A CTTCGA F D CCGGCCC R E	C G G T C G G L C G
1081 1141 1201 1261 1321 1381 1441 1501 1561 1621 1681	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G L S GCATCTA G ACAGGG D R CAGGACG Q D GCCGTCC	L GCTGG GATCG E CGAGG E GCTGG G GCCCC A GCCGG C CCGGG C CCGGG C R GCCGG C R GCCGG C R GCCGG C C C GCCGG C C C GCCGG C C C GCCGG C C C GCCG C	L V AGAA E N AGGA AGGA CCGCG A A CCGCG A R CCGCG A V CCGCG A V CCGCG C C C C C C C C C C C C C C C C	A CCCG GATG CGCC A CGCC CGC CCAC CCCAC CCCC CCC	GAGC EGACC EGATCO EGATCO EGACC EGAC EGA	H Q CACAA H K GGTAA R Y GGCGGG G G GACGA D E GTCGC V A GAGAT E J GAGAT E J GAGAT C C C C C C C C C C C C C C C C C C C	GGA GGA CGTC CGTC CGTC CGTC CGTC CGTC CG	T R CGGA SAGCA SACCA SACCA SACCA STCC SGCCA STCC SGCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCGA SCCA SCC	A L TCGG TCGG V TCCG TCGG F G ACGG Y G ACGG Y C G ACGG Y C G ACGG C T G G G C T G G G C C G G C C C G C C C C	N MC CGGCC A A FIGGAC V E GCCAC SGCAC SGCAC SGCAC SGCAC SGCAC SGCAC SGCC CGCT I I GCGC CCGC V CCGC V CCGC P ATCC	I CGAC D DAAC N CCCG CCAC CCAC CCAC CCAC CCGT CCGT CCGT	S ICC! S GCG(C A GAG(C C GCAC: Q C C C C C C C C C C C C C	L CCGC S CCCG P GCCA C CCC C C C C C C C C C C C C C C	G V TGCTG CCCCGG A R TCGTC I V TCGAC L D TCCGG V P TTCCGG C Y P TTCCGG C Y F D CCCGCC R F GGATTG	C G G T C G G G C G A
1081 1141 1201 1261 1321 1381 1441 1501 1561 1621 1681	A V M TGGCGCTC V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGGG D R CAGGACG Q D GCCGTCC A V	L GCTGG GATGG CGAGG GCTGG GCTGG GGCCC A GCTGG CCGGG C A GCCGGG C R AGGAGG E E E GCCGG C R A GCCGG C R A GCCGG C R A GCCGG C R A GCCGG C R A C GCCGG C R C C GCCG C R C C C C C C C	L V AGAA AGGA AGGA ACAT D I CCCGC A A CCGCG CGCG CGCG CCGT CCCGT	A CCCG GATG CGCC A CGCC CCAC CCAC CCAC C	G GGAGC E CCGCC L GATCC U CCACC H GCCCCC A CCCCCC A CCCCCC A CCCCCC C A A CCCCCC	H Q CACAA H K CGCTA R Y GGCGGG GGCGG GGCGG V A GAGAI E G GGCGG GCCCC GCCCC GCCCC GCCCC GCCCC C	T GGCA CCTTC CCTTC CCTC CCTC CCTC CCTC C	T R R AGCG S ACCA S G C G C G C G C G C G C G C G C G C G	A 1 TTCGC TTTCGC TTCGC TTCGC TTCGC FCC GAGAC FCCG GAGAC CCTGGC CCTGGC CCTGGC CCTGGC CCTGGCC A	N MC CGGCC A A GGCCA V E GGCA GGCA GGCA GGCA GGCA GGCA CCCC CCCC CCCC CCCC CCCC CCCC CCCC CCCC	I CGACC D GAACC N GAACC N CCAC P CCAC CCAC CCAC CCGTC CCGCC CCGT CCGCC CCCC CCGCCCC CCGCCC CCGCCC CCGCCC CCGCCC CCGCCC CCGCCCCC CCGCCCCCC	S ICCI S GCGC GAG GAG C GAC C GCA C GCCA C GCA C GCA C GCA C GCA C GCA C GCA C GCA C GCA C GCA C GCA C GCA C GCA C GCA C GCA C GCA C G GCA C G GCA C G G GCA C G G G G	L CGGC S CCCG G GGCA C CGGC C C C C C C C C C C	G V TGCTG CCCCGG A R TCGTC I V TCGAC L D TCCGGG V P TTCCGG V P TTCCGA CACCGC C Y CCGGCC R C GGATTG D * *	C G G T C G G G C G A
1081 1141 1201 1261 1321 1381 1441 1501 1561 1621 1681 1741	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V S C TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGG D R CAGGACG Q D GCCGTCC A V CCGGCCGG	L GCTGG GATGG CGAGG C GGCGG G GGCGG C GGCGG C C GGCGG C C C GGCGG C C C G G C C C C G C	L V AGAA E N AGGA E C ACAT D I CCGC CCGC A A CCGCG C CCGT C CCGC C C C C C C C C C C	A CCCG GATG GATG CGCG C CGCC CGCC CCCAC CCCCAC C	GGAGG EGGAGG L EATCO TGGC TGGC W CCACC H GCACC C GAAC' V N CCACC GAAC' V A CCTGCC C GAAC' L CGCCC GAAC' L CGCCC GAAC' L CGCCC C C C C C C C C C C C C C C C C	H Q CACAA H K CGCTAA R Y GGGCGG GGACGA D D GTCGCC V A GAGGA E E GAGGA C C GCCCC GCCCC C GCCCC C C C C C C C C	T GGCA CTTC CGTG CGTG CGTG CGTG CGTG CGT	T CGGA R SAGCA SACCA SGTCI V CGGGI C CGGI C CGGI C CGGI C CGGI C CGGI C CGGI C CGGI C CGGI C CGGI C CGGI C CGGI C CGGI C CGGI C C CGGI C C CGGI C C C C	A 1 ITCGG I 2 ITCG V TTCGC TTCGG FF C CACGG FF C CACGG FF C CACGG FF C CACGG FF C CACGG C CCTGG CCTGG CCTGG CCTGG CCCTGC CCCTGGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCTGC CCCCCCC CCCCCCCC	N MCGGCC   CGGCC GGCC   IGGAG V   EGCAAA   R K   GGCACA   S H   GGCACA   S V   CGCTC   I   LACGCC   ORF2   GTGG   V   CCGGG   V   CCGGG   V   CCGGG   TTCC	I CGACC D GAACC N CCCCG CCACC CCACC CCACC CCACC CCACC CCACC CCACC CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCC	S S S S S S S S S S S S S S	L CGCC S CCCG GGCA CGCC CGCC CGCC CGCC CG	G V TGCTG CCCGGG A R TCGGC I V TCCGGG L G TTCCGG V P TTCCG CCGGCC T P CCGGCC C F D CCGGCC C R F GGATT C CGGCC C C C C C C C C C C C C C C C C	C G G T C G G G C G A G
1081 1141 1201 1261 1321 1381 1441 1501 1561 1621 1681 1741	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGGG D R CAGGACG Q D GCCGTCC A V CCGGCGGG GTGTACG	L GCTGG GATCG GAGGG C GGCGG G GGCCC A GGCCGG C A GGCCGG C R GGCGG C R GGCGG C R GGCGG C R GGCGG C R GGCGG C A GGCAG C C G C G C G C G C G C	L V AGAA AGGA AGGA E E ACAT D I CCGCG CGCGC A A CCGCG CGCG C CCG C C C C C C C C C C C	A CCCC P GATG CGAC A CCCC CCCC CCCC CCCC CCCC CC	GGAGG EGGAGG L EATCCC I TTGGC W CCACC H EGCGCCC A CCGCCC A A CCGCCC A A CCGCCC A A CCGCCC C A CCGCCC C A CCGCCC C C CCGCCC C C CCGCC C C C	H Q CACAA H K GCTAA R Y GGCGG GACGA GACGA C G C GACGA C C C C C C C C C C C C C C C C C C C	GGCA CGTGC CGTG CGTG CGTG CGTG CGTG CGTG	T R CGGA S AGCA S ACCA S CCGA C CGGA C CGGA C CGA C CGA C CGA C CGA C CGA C CGA C CGA C CGA C CGA C CGA C CGA C C CGA C C C C	A 1 ITCGG I 2 ITCG I 1 I 1 ITCGG V C ITCGG I ITCGG V C I I GGGC I I GGGC I I GGCC A C GGCC A C GGCC	M MC CGGCC A A IGGAC V E GCAA GGCAA GGCAA GGCAA GGCAA CGCC I I ACGCC V CGCT I I ACGCC V CCGG V CCGG V CCGG V CCGG V CCGG V I CCGG V I CCGG C CCGG C V I CCGC I I C CGCT I I C CGCT I I C CGCT I I C C C C I I I C C C C I I I C C C I I I C C C I I I C C C I I I C I C I I I C I I I C I I I I C I I I I I C I	I CGACC D GAACC D GAACC N CCCCG CCACC CCACC CCACC CCACC CCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCT CCCCCT CCCT CC	S ICC <sup>11</sup> S GCGC GCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C C C C C C C C C C C C	L CGCC S CCCG F GGCA CGCC C CGGGC C CGGGG C C C C C C C	G V TGCTG CCCGGG A R TCGTC I V TTCGAC L D TTCGGG L G TTCGGG V P CTTCGA F D CCGGCCC R F GGATC SGATC SGCCCCC	C G G T C G G G C G, A GA
1081 1141 1201 1261 1321 1381 1441 1501 1561 1621 1681 1741 1801 1801	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGGG D R CAGGACG Q D GCCGTCC A V CCGGCGG GCCCCCC	L GCTGG GACGG GCTGG GCTGG GCTGG GGCCC A GGCCGG G GCCGG C R GCCGGG R GCCGG R GCCGG R QCCCG C GTTGG GCCCC	L V AGAA AGGA AGGA AGGA CCGCGC A A CCGCGCA A CCGCG CCGCG CCGCG CCGCG CCGCG CGCGC CGCCG CGCCG CGCCG CGCCG CGCCG CGCCG CGCCG CGCCG CGCCG	A CCCC P GATG CGCC CGCC CGCC CCCCC CCCCC CCCCC CCCCCC	GGGGC EGGGCG LGATCC IGATCC WCCACC HGCTGC CGCCCC A BGCGCC CA CGCCCC A A CGCCCC CA CGCCCC CGCCCC CGGCC CGGCC CGGCC CGGCC CGGCC	H Q CACAA H K GCTAA R Y GGCGGG GACGGA GACGA V A GAGAT E I GAGGA C C C C C C C C C C C C C C C C C C C	T GGCA CCTCC F CCGTG CCGTG CCGTG CCGTG CCGTG CCGTG CCGTC	T R R SAGCG S SACCA S SACCA S S S C C S S S S S S S	A 1 ITCGG I 2 ITCG V TCCC V TCCC I 1 ITCGG FF C FF C FF C FF C FACAG E SACAG E CACAG C C C C	N MC CGGCC A A FIGGAC V E GCCAC GGCAC GGCAC GGCAC GGCAC CGCC CGC	I CGACC D GAAC CCAC CCAC CCAC CCAC CCAC C	S CCC S S S S S S S S S S S S S	L CGCC S GGCA GGCA CGGCA CGGCA CCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCC	G V TGCTG CCCCGG A R TCGTC I V TCGAC L D TCCGG C TCCGG C T A CCCGCC C CCGCCC C CCGCCC C CCGCCC C CCGCCC C CCGCCC C CCGCCCC C CCCCCC	C G G T C G G G C G, A GAA
1081 1141 1201 1261 1321 1381 1441 1501 1621 1681 1741 1801 1861	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G L Y GACAGGG D R CAGGACG Q D GCCGTCC A V CCGGCGGG GTGTACG GCCCCCG	L GCTGG GATCG GAGG GCTGG GCTGG GCCGGG GCCGGG GCCGGG CCGGG CCCGGG GCCCG GCCCGG	L V AGAA AGGAA AGGA AGGA CCGC A A CCGCG A R CCGCG A R CCGCG A V CCGT C C C C C C C C C C C C C C C C C	A CCCG GATG CGAC A CGAC A CGAC CCAC CCCAC CCCCC CCCCC CCCCC CCCCCCCC	G GGAGC E GCTGC L GATCC I GCTGC W CCACC H GCCCC A CCCCC A CCCCC A CCCCC A CCCCC A CCCCCC	H Q CACAA H K CGCTA R Y GGCGG GGCGG GGCGG GGCGG GGCGG GGCGG GGCGG GGCGG CGCCG GGCGG CCCCG CGCCG CGCCG CCCCG CGCCG CCCCG	GGA GGA CCTC CGTC CGTC CGTC CCTC CGTC CG	T R R CGGA S AGCCA S GCCA C G C G C G C G C G C G C G C G C	A 1 TTCGC TTCGC TTCGC TTCGC TTCGC FC CTCGC FC CTCGC FC CTCGC CTCGC CTCGC CCCC CCCC CCCCC CCCCC CCCCC CCCCCC	N MC CGGCC A A IGGAC V E GCCAA SGCAA SGCAA SGCAA SGCCA SGCCA SGCCA SGCCA CCGCT I CCGCT CCGCA CCGCA CCGCA CCGCA CCGCA CCGCA CCGCA CCGCA CCGCA CCGCA CCGCA	I CGACC D GAACC GAACC N GAACC P CCCAC CCCAC CCCAC CCCCC CCCGT CCCCC CCCCT CCCCC CCCCT CCCCC CCCCT CCCCC CCCCT CCCCC CCCCT CCCCCC	S ICC <sup>11</sup> S GCG( GCG( C GCC( C GCC( C C C C C C C C C C C C C	L CGCCG S CCCGG P GGCA CGGCA CCGGCA CCGGCA CCGGCA CCGCC CCGCC CCGCC CCGCC CCGCC CCGCC CCGCC CCGCC	G V TGCTG CCCCGG A R TCGTC I V TCGAC L D TTCGGG L G TTCGGG V P TTCCGG V P TTCCGG C Y CCGCCCG C C CCGCCC CCGCCCC CCCGCG CCCGGG	C G G T C G G G C G, A GAAC
1081 1141 1201 1261 1321 1381 1441 1501 1681 1681 1681 1741 1801 1861 1921	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V S C TGTCCG V S C TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGGG D R CAGGACG Q D CCGGCGG GCCGTCC A V CCGGCGGG GCCACCC	L GCTGG GATCG I CGAGG C GGCCC G GGCCC C GGCGG G C CCGGG G C CCGGG G C CCCCC GTCG GCTGC GCTGC C CCCCCC GCTGC	L V AGAA AGGAA E N AGGA E E ACAT D I CCGCGC A A CCGCGC A CCGCGC CCGCGC CCGCGC CGGGCC CGGGCC CGGCCCCCC	A CCCC P GATG CGCC CCCC CCCCC CCCCC CCCCC CCCCC CCCCCC	GAGG GGAGG E CTGC L CTGGC TTGGC TTGGC TTGGC TTGGC C CACC C C C C C C C C C C C C C C C	H Q CACAA H K CGCTAA R Y GGCGG G G GACGA D D GTCGCC V A GACGA C C C C C C C C C C C C C C C C C C C	T GGCA CTTC CGTG CGTG CGTG CGTG CGTG CGT	T CGGA R SAGCA SACCA SACCA SGTCI V CGGGI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C CGGTCI C C CGGTCI C C C C C C C C C C C C C C C C C C	A 1 ITCGG I A ITCGC I I I I ITCGG FF C CACGG FF C FF C FF C FF C FF C FF C FF C F	N MCGGCC   CGGCC GGCC   IGGAC V   EGCAAA GGCCAC   GGCAC GGCCAC   CGCTC I   GGCAC ORF2   GTGGC V   CCGCT I   CGCTC GTGCC   GTGCT CCGCT   GCGCT CCGCT   GCGCT CCGCT   GCGCT CCGCT   GCGCT CCGCT   GCGCT CCGCT   TTCC CCGCG   TTCCC I   TTCCC I   GGCCC A   GGCCC GGCCC	I CGACC D GAACC N CCCCG CCACC	SCI SCGC SCGC A GACC O CCAG CCAG CCAG CCAG CCAC CCAC	L CCCCG S CCCCG G CCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCC	G V TGCTG CCCGGG A R TCGTC I V TCGGCG L D TCCGGG L G TTCCGG V P TTCCG CACCGCC I Y CACCGCC I Y CACCGCC CGCACC GGATG CCCGGC CCCGGC GCAGT	C G G T C G G G C G A GAACA
1081 1141 1201 1261 1321 1381 1441 1501 1621 1681 1741 1801 1801 1803	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V A T TGTCCGG V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGGC D R CAGGACG Q D GCCGTCC A V CCGGCGG GGCCCCG GCCAGGC CCGACCA	L GCTGG GATCG I CGAGG C GCTGG G GCCG G GGCCG A GGCCG G GGCGG C CCGGG C C GGCGG C C CCGGG C C CCCCG GTGCG GCCCC GTGCG GCCCCCG GCCCCCCCC	L V AGAA AGGA AGGA AGGA CCGC CCGC CCGC CCG	A CCCG P GATG GATG CGATG CCCCC CCCCC CCCCCC CCCCCC CCCCCCCCCC	GGAGG EGAGG L GATCC L CTGGC W CCACC H GCCACC H GCCACC A CCGCC A A CCGCC A A CCGCC V A A CCGCC V A A CCGCC V C GCACC C CGCC C CGCACC C C C C C C C C	H Q CACAA H K GCTAA R Y GGCGG G G GACGA D L GTCGC V A GAGAT E J GAGGA C C C C C C C C C C C C C C C C C C C	T GGCA CCTTC F CCGTG CCGGTG CCGGC CCGCCG	T CGGA R SAGCA SAGCA SACCA SACCA SACCA CGGGI CGGGI CGGGI CGGI	A 1 ITCG( I 2 ITCC) V TTCC( V TTCC) V TTCC( V TTCG( V TTCG( V C V C V C V C V C V C V C V C V C V C	M MC CGGCC A A FIGGAC V E GCAAA R K GGCAA S H GAGTO CGCTO I CGCTO I CGCTO I CGCTO I CGCTO I CGCTO I CCGC V CGCTO I CCGC V CGCTO I CCGC CCGG CCGG CCGG CCGG CCGCC AGGCA ACGCC ACGCC ACGCC ACGCC CCCTO CCCGC CCCCC	I CGACC D GAACC N CCCCG CCACC CCCCG CCACC CCACC CCCCG CCACC CCCCG CCCCG CCCCG CCCCG GGCCC GGCCC GGCCC CCCCG GGCCC CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCG CCCCC CCCCG CCCCC CCCCG CCCCC CCCCG CCCCC CCCCC CCCCG CCCCC CCCCG CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCC	SCG SCG SCG GAC GAC CAC CAC CAC CCAC CCA	L CCCCG S CCCCG G G CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCC	G V TGCTG TGCTGG CCCGGG A R TCGTCG I V TTCGGG L D TTCGGG C C TTCGGG C P CCGGCC ACCGC CCGGGC CCGGGC CCGGGC CCGGGC CCGGGC	C G G T C G G G C G A GAACAT
1081 1141 1201 1261 1321 1381 1441 1501 1561 1621 1681 1741 1801 1801 1921 1981 2041	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V S G TCGAACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G I Y GACAGGG D R CAGGACG Q D GCCGTCC A V CCGGCGGG GGCCCCCC GGCCCCCC GCCAGCC CCGAAGGC	L GCTGG GATCG GAGGG GCGGG GCCCC A GGCGG GCCCC C GGCGG C R GGCGG C R GGCGG C R GGCGG C R GGCGG C C C GGTACT GATCG C C C C C C C C C C C C C C C C C C	L V AGAA AGGA AGGA CCGCG A ACAT D I CCGCGC A A CCGCG A A CCGCG C CCCCT C CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCC	A CCCC P GATG CGCC A CCCCC CCCCC CCCCC CCCCC CCCCC CCCCC CCCC	G GGAGC E GCTGC L GATCC W CCACC W CCACC W CCACC C GCACC A C GCGCC C GA C C GCACC C GA C C C GCACC C C GA C C C C	H Q CACAA H K GCTAA R Y GGCGG G G GACGA D E GTCGC V A GAGAT E I GAGGA C C C C C C C C C C C C C C C C C C C	T GGCA CCTCC F CCGTG CCGTG CCGTG CCGTG CCGTG CCGTC CCGTCC CCGTCC CCGTCC CCGTCC CCGTCC CCGTCC CCGTCC CCGTCC	T R R SAGCG S SACCA S SACCA S S S C C S S C S C S C	A 1 ITCGG I 2 ITCG V C ITCGG V C C C C C C C C C C C C C C C C C C C	N MC CGGCC A A FIGGAC V E GCAAC R K GGCAAC GGCAC CGCT I GGCAC CGCT I CGCT I CGCT I CGCT I CCGGG V CCGGG V CCGGG V CCGGG V CCGGG CCGGG CCGGAC CCGGAC CCGGAC CCGGAC CCCGAC CCGCAC CCGGAC CCGGAC CCCGGAC CCCGAC CCCGAC CCCCCCCC	I CGACC D GAAC D GAAC N GCAC P CCAC CCAC CCAC CCAC CCAC CCAC CC	S CGCI S GCGG GACC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C GCCC C C C C C C C C C C C C	L CGCC S GGCA GGCA CGGCC R CGGCC CGCCG	G V TGCTG TGCTGG CCCGGG A R TCGTC I V TCGAC L D TCGGG L G TTCGGG V P CACCGC T A CACCGC R F CGGACC CCGCC CGCACC CCGGC CCGGG CCGGG CCGGG CCGGG CCGGG CCGGG	C G G T C G G G C G A GAACAT
1081 1141 1201 1261 1321 1381 1441 1501 1621 1681 1681 1741 1861 1981 2041 2101	A V M TGGCGCT V A L CCCCGGC P P A TCGCCAC V S G TCGACG F E R AGAACCT Q N L GCCTGTC G L S GCATCTA G L Y GACAGGG D R CAGGACG Q D GCCGTCC A V CCGGCGGG GGCCCCCG GCCCCCCG GCCCCCCG GCCCCCCG GCCCCCCG GCCCCCCG GCCCCCCG GCCCCCCCC	L GCTGG GAGG GCTGG GCTGG GCTGG GGCCC A GGCCG GCCGG C CCGGG C CCCCG GGTGG GGTGG GGTGG GGTGG GGTGG GGTGC GGTGC GGTGC GGTGC GGTGC GGTGC	L V AGAA AGGAA AGGA AGGA CCGCGA A CCGCGCA A CCGCGCA A CCGCGCA CCCCC CCCCC CCCCCC CCCCCC CCCCCC CCCCCC	A CCCC GATC CGCC A CCCC CCCC CCCC CCCC C	GGGGG EGGGCG LGATCCC IGCTGGC W CCACCC H GCCCCC A GCCCCC A GCCCCC A GCCCCC A CCCCCC CGGA CCGCA CCGCA CCGCA CCGCA CCGCA CCGCA	H Q CACAA H K GCTAA R Y GGCGG G G GACGA D E GACGA V A GAGAT E I GAGAT C C C C C C C C C C C C C C C C C C C	T GGCA CCTC CCTC CCTC CCTC CCTC CCTC CCT	T R R SAGCG S SACCA S SACCA S S S S S S S S S S S S	A 1 TTCGG I 1 TTCGG V - TTCGG FACGG FACGG FACGG CFG CTGG CTGG CCTGG CCTG CCTGG CCTG CCT	N MCCGCC CCGCCC IGGAC V E GCAAC R K GGCAC S H GAGTC GGCC CCGCT I CCCCT I CCCGC V CCGC V CCGC V CCGC V CCGC V CCGC I TTCCC TCGCCA CCGC CCGC CCGC CCGC CCGC C	I CGACC D D D D D D D D D D D D D D D D D D	S ICC <sup>11</sup> S GCG( GCG( GCG( CCGCC CCGCC CCGCC CCGCC CCGCC CCGCC CCGCC CCGCC CCGCC CCGCC CCGCC	L CGCC S CCCG G CCCG CGCC CGCC CGCC CGCC	G V TGCTG CCCGGG A R TCGTC I V TCGAC L D TCGGG L G TTCGGG C Y P TTCCGG V P TTCCGG C Y F D CACCGCC C A CGCCCC GGATCG GGATG CCGCGC CCGCGC CCGCGC CCGCGC	C G G T C G G G C G A GAACAT

## 图 2 2.2 kb DNA 片段的全序列测定

Fig. 2 Nucleotide sequence of a 2.2kb DNA fragment

RBS:Ribosome binding site;ORF:Open reading frame. © 中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn





second and third base at 5 ' to 3 ' end respectively.

产色链霉菌 7100 的原生 质体,得到 112 个转化 子。随机挑选 12 个转化 子提取质粒并经酶切验 证均正确。将其中之一 接种到加有普霉素的 MM 培养基上 28℃培养 7d,制备孢子悬液。约 10<sup>6</sup>的孢子悬液涂布在含 有阿普霉素抗性的 MM 平皿上,39℃培养,每个 平皿上得到约 100 个破 坏子。

2.3.2 破坏子的 Southern blotting 杂交验证:随机挑选 3 个可能的破坏 子于 39℃培养后提取总 DNA,以圈卷产色链霉菌 野生型菌株 7100 总 DNA 为对照,分别用 Bg/II 酶切,经琼脂糖凝 胶电泳和 Southern 转移

基因功能研究 2.3 2.3.1sanH 基因的破坏: 由于 sanH 和 sanI 是共转录 的 故破坏位于上游的 sanH 后 sanI 也不能表达。将 sanH基因内部的 680bp 的 Smal片段亚克隆至载体 pKC1139 的 EcoRV 位点 \_\_\_\_」 2000(bp)上 构建为用于基因破坏的 重组质粒 pKC1139::∆sanH (图 4)。先转化大肠杆菌 ET12567,以消除重组质粒 在大肠杆菌 IM109 中可能 发生的修饰,然后转化圈卷



后,以地高辛标记的 680bp SmaI 片段为探针进行 Southern blotting 杂交。结果表明来自野生型菌株的总 DNA 只出现 1 条 2. 创本的阳性杂杂信品带亚西来自破扬子的总。DNA。分。

别出现 1.7kb 和 7.6kb 的 2 条阳性杂交信号 带(图5),其分子量与理论计算的大小是一 ы 致的,说明已获得了正确的破坏子。 **2.3.3** 破坏子的尼可霉素检测:将3个正确<sup>7.7</sup> 的破坏子分别划线接种在加有阿普霉素的 3.5-R2YE 平皿上 在 28℃培养 3 5 .7d 后进行了 1.9-尼可霉素产生的测定。结果表明所有破坏子 在测定时间内同野生型菌株一样明显地抑制 赤星灰霉的生长(图略)。说明 sanH 基因的 破坏不影响圈卷产色链霉菌产生的尼可霉素 的生物活性。

#### 1 2 3 4 5 1 2 3 Δ 5 9.3-2.7-00000 B

### 图 5 总 DNA Bgl II 酶切后的琼脂糖凝胶电泳(A) 及 Southern blotting 杂交(B)

#### 讨论 3

在链霉菌的分子生物学研究中,抗生素 生物合成基因的研究是该领域的一个热点。 在分子水平上研究抗生素生物合成基因的结

Fig. 5 Agarose gel electrophoresis( A ) and Southern blotting hybridization of total DNA digested with Bgl II( B) 1 ADNA/Eco T14I MW marker ; 2~4 : Total DNA from different disruptants ; 5 : Total DNA from wild type strain.

构及功能将有助于了解抗生素生物合成的途径,有利于阐明底物的进入及产物的释放之 间的确切关系 从而可揭示各组分之间的结构及作用机制。

抗生素的生物合成是多基因控制的,一般涉及紧密连锁的 20~30 个基因。尼可霉素 有 20 多种活性单组分 其生物合成途径相当复杂。要了解其复杂的生物合成途径 就必 需对与其生化合成途径相关的各基因的结构、功能及各基因之间的相互关系进行深入研 究。本实验室近期克隆到一个与尼可霉素生物合成相关的 7.0kb DNA 片段 部分序列测 定表明其中含有三个开放阅读框,分别命名为 sanF ,sanH 和 sanI。已证明 sanF 为圈卷 产色链霉菌尼可霉素生物合成中起重要作用的基因。因而推测位于 sanF 基因上游的 sanH和 sanI 也是尼可霉素生物合成基因簇中的一员。虽然 sanH 基因破坏的阻断突变 株同野生菌株一样能抑制抗生素检测指示菌赤星灰霉的生长 但这并不意味着它们产生 的尼可霉素活性组分是完全一样的。从结构上看 SanH 和 SanI 与参与链霉菌中呼吸链 电子传递的细胞色素 P450 和铁氧还蛋白有很高的同源性 因此它们在圈卷产色链霉菌 中可能也是起电子传递的作用。从尼可霉素的结构看,在其肽基组分中有两个羟基,它们 都可能是电子传递后的产物。而羟基缺失后的尼可霉素组分 Nikkomycin Lx 和 Lz 仍有 高的抗真菌活性<sup>10]</sup>。另外,据最新文献报道<sup>10]</sup>,在唐德链霉菌中也已克隆到了与细胞色 素 P450 和铁氧还蛋白编码基因高度同源的 nikF 和 nikG。 nikF 基因被破坏的阻断突变 株产生的尼可霉素组分为具高抗真菌活性的 Nikkomycin Lx 和 Lz。这进一步说明本文 对 sanH 和 sanI 基因功能的推测是正确的。但是,要更深入了解 sanH 和 sanI 在尼可霉 素生物合成途径中的作用机制,还需分离纯化并鉴定蛋白质,利用 HPLC 等手段检测破 坏子积累的中间产物等。虽然 sanH 基因的破坏不影响圈卷产色链霉菌产生的尼可霉素 的生物活性 但进行该基因的结构与功能的研究有利于阐明尼可霉素生物合成的途径。

#### 参考文献

- [1] Fiedler H P, Kurth R, Delzer J, et al. J Chem Tech Biotechnol, 1982, 32 271~280.
- [2] Graybill J R, Najvar L K, Bocanegra R, et al. Antimicrob Agents Chemother 1998 A2 2371~2374.
- [3] Bormann C, Mattern S, Schrempf H, et al. J Antibiot ,1989 A2 913~918.
- [4] Sambrook J, Fritsch E F, Maniatis T. Molecular Cloning : A Laboratory Manual, 2nd ed. New York : Cold Spring Harbor Laboratory, 1989.
- [5] 贾君永 李文利 陈 蔚 等.中国科学(C辑),2000 30:1~8.
- [6] Hopwood D A, Bibb M J, Chater K F, et al. Genetic Manipulation of Streptomyces. A laboratory manual. Norwich, England John Innes Foundation, 1988.
- [7] Tan H, Yang H, Tian Y, et al. Gene, 1998 212:137~146.
- [8] Bibb M J , Findlay P R , Johnson M W. Gene ,1984 30 :157~166.
- [9] Möhrle V, Roos U, Bormann C. Mol Microbiol ,1995, 15 561~571.
- [10] Bormann C, Lauer B, Kàlmànczhelyi A, et al. J Antibiot, 1999 52 582~585.

# STUDIES ON sanH AND sanI—GENES RELATED TO NIKKOMYCIN BIOSYNTHESIS OF STREPTOMYCES ANSOCHROMOGENES \*

Chen Wei Tian Yuqing Yang Haihua Tan Huarong

(Institute of Microbiology, The Chinese Academy of Sciences, Beijing 100080)

Abstract : A 7.0kb DNA fragment was cloned by reverse genetics, which was located in the gene cluster of nikkomycin biosynthesis of *Streptomyces ansochromogenes*. Besides sanF—a gene related to nikkomycin biosynthesis of *S. ansochromogenes*, the sequence of a 2.2kb *Bgl* II DNA fragment upstream sanF was determined. This 2.2kb DNA fragment cointains two complete open reading frames (ORF). ORF1 consists of 1233 base pairs, ORF2 consists of 195 base pairs, and they encode proteins with 410 amino acids and 64 amino acids respectively, which were designated as sanH and sanI. In search of databases, the deduced products of sanH has 46% amino acid identities and 62% amino acid similarities in comparison with the cytochrome P450 of *Streptomyces grise-olus*, and the deduced protein of sanI has 56% amino acid identities and 70% amino acid similarities with the ferredoxin of *S. griseolus*. The function of sanH didn 't influence nikkomycin biological activity in *S. ansochromogenes*.

Key words: Streptomyces ansochromogenes, Genes related to nikkomycin biosynthesis, Gene disruption

<sup>\*</sup> Project Supported by Natural Science Found tion of Ching 39830010) ②中国科学院微生物研究所期刊联合编辑部 http://journals.im.ac.cn