

浑球红细菌谷氨酸合酶大亚单位基因 (*gltB*)的序列分析*

鲁涛 吴永强** 宋鸿遇

(中国科学院上海植物生理研究所 上海 200032)

摘要 测定了浑球红细菌(*Rhodobacter sphaeroides*)谷氨酸合酶大亚单位基因(*gltB*)及其5'端和3'端的序列,全长为5510 bp。序列分析表明,*R. sphaeroides gltB*基因全长为4636 bp。从核苷酸序列推测其蛋白质分子量约为164 kD。*R. sphaeroides gltB*基因与*Azospirillum brasilense*和*Escherichia coli*的*gltB*基因DNA序列有很高的同源性。其蛋白质氨基酸序列与*A. brasilense gltB*基因产物GltB也具有很高的同源性。此外,还对*R. sphaeroides* GltB的各可能功能区进行了分析,发现它们具有很高的保守性。

关键词 浑球红细菌, 谷氨酸合酶基因, DNA序列

谷氨酸合酶(glutamate synthase, GOGAT; EC 2.6.1.15)是细菌氨同化的一个关键酶,当生物体处于氮限量的环境中时,它与谷氨酰胺合成酶(glutamine synthetase, GS; EC 6.3.1.2.)所催化的反应是合成谷氨酸的主要途径^[1]。在不同细菌中,编码GOGAT的*glt*基因的组成和调控特性各不相同^[2~7]。在光合细菌中,*glt*基因尚未克隆和测序。我们首先克隆了*R. sphaeroides*的*glt*基因并证明了*R. sphaeroides glt*基因与*E. coli*的*gltB*基因有同源性^[8]。通过进一步的Southern杂交、互补分析和启动子探测,把编码*R. sphaeroides* Glt的大亚单位的基因*gltB*定位于pLTRK272的一段5.5 kb的EcoRI-Sal I片段内(未发表)。本文测定了这一片段的DNA序列并对*Rhodobacter sphaeroides*的*gltB*基因的阅读框架、氨基酸编码和蛋白质功能区进行了分析,为研究*R. sphaeroides glt*基因自身的调控和其在氨同化及固氮调控中的作用奠定了基础。

1 材料和方法

1.1 菌株和质粒

实验所用菌株和质粒见表1。

1.2 培养条件

*E. coli*于37℃振荡培养。SOB、SOC和LB培养基参照文献[9]。根据实验需要添加不同抗生素,浓度如下: Ampicillin 50 µg/ml, Tetracycline 15 µg/ml。

1.3 DNA操作

* 国家高技术发展(863)计划和国家自然科学基金资助项目。

** 责任作者

本文于1996年4月18日收到。

表 1 菌株和质粒
Table 1 Strains and plasmids

	特 性 Relevant characteristics	来源 / 文献 Source / Reference
菌株 Strains		
<i>E. coli</i>		
JM109	rec A1 supE44 endA1 hsdR17 gyrA96 relA1 thiΔ (lac-proAB)	Yanish-Perron 1985
质粒 plasmids		
pLTRK272	10.5kb HindIII fragment of pLT27 cloned in pRK415, Tc ^r	[8]
pBluescriptKS	Ap ^r , sequencing vector	Short 1988
pBluescriptSK	Ap ^r , sequencing vector	Short 1988
噬菌体 phages		
M13mp18/19	Ap ^r , sequencing vector	Messing 1983

质粒 DNA 的制备和纯化、去磷酸化反应和连接反应、低熔点琼脂糖凝胶回收 DNA 方法、 α -互补现象的检测、大肠杆菌感受态细胞的制备及转化、M13 噬菌体感染方法、复制型及单链 DNA 抽提均按文献 [9] 进行；限制性内切酶反应按产家说明书进行。

1.4 M13 单向缺失克隆的构建

采用 Promega 公司的 Erase-a-Base 试剂盒，按产家说明书进行。

1.5 测序引物的合成

测序引物为 17 mer，在 Beckman Oligo 1000 DNA 合成仪上合成。

1.6 DNA 序列测定

采用 Sanger 的双脱氧终止法^[10]。测序反应按产家说明书进行。自动测序在 ABI Model 370 A DNA 顺序分析仪上进行。手动测序电泳采用 Bio-Rad 的 Sequi-GenII 核酸测序电泳系统。

1.7 主要药品来源

限制性内切酶、T4DNA 连接酶购自 Bio-lab 公司和华美生物工程公司，碱性磷酸脂酶购自 Boehringer Mannheim 公司，琼脂糖、抗生素购自 Sigma 公司，低熔点琼脂糖购自 BRL 公司，Erase-a-Base 试剂盒、T7 DNA 聚合酶测序试剂盒 TaqTrack 测序试剂盒 (Deaza) 购自 Promega 公司，Bst DNA 测序试剂盒购自 Bio-Rad 公司， $[\alpha\text{-}^{32}\text{P}]\text{dATP}$ 购自北京亚辉生物医学工程公司， $[\alpha\text{-}^{35}\text{S}]\text{dATP}$ 购自 Amersham 公司，其它所用药品均为分析纯。

2 结果和讨论

2.1 *R. sphaeroides* *gltB* 基因的序列测定

通过 Southern 杂交、互补分析和启动子探测，我们把编码 *R. sphaeroides* 的 GOGAT 大亚单位的基因 *gltB* 定位于 pLTRK272 的一段 5.5 kb 的 EcoRI-SalI 片段内(未发表)。按材料和方法中所述从正反两个方向测定了这一片段的 DNA 序列。由于 *R.*

sphaeroides 总 DNA 的高 G+C 百分含量^[11], 在测序中用 7-deaza-dGTP 代替 dGTP 以防止 GC compression。测序策略见图 1。经测定, 这一片段的全长为 5510bp, 全序列见图 2。

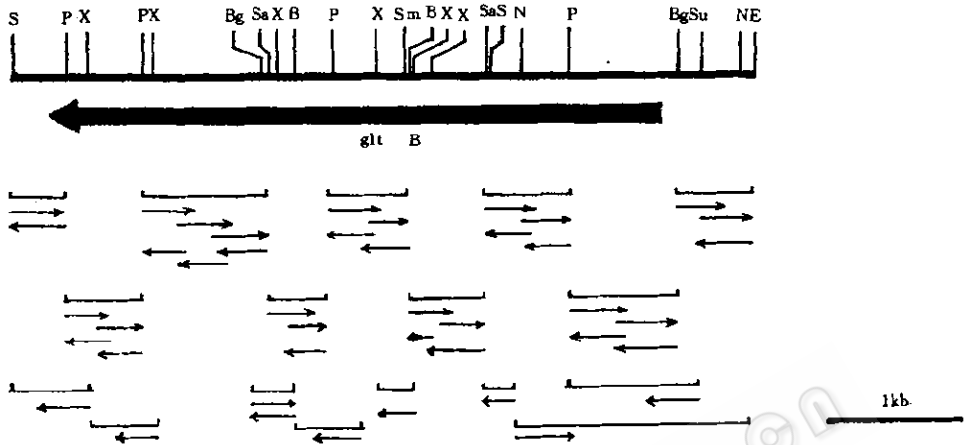


图 1 *R. sphaeroides gltB* 基因的测序策略

粗箭头表示 *gltB* 区域; 细箭头表示测序的方向和长短; 细线表示测序用的 M13 克隆。

Fig.1 The sequencing strategy of *R. sphaeroides gltB* region

The heavy arrow indicates the *gltB* regions. The thin arrows indicate the direction and extent of sequencing. The M13 clones used are represented by thin lines.

Abbreviations of restriction endonuclease are as follows: B, BamHI; Bg, BglII; E, EcoRI; H, HindIII; N, NaeI; P, PstI; Sa, SacI; S, SalI; Sm, SmaI; Su, StuI; X, XhoI.

序列分析表明这段 DNA 序列中存在一个主要的开放阅读框架(ORF), 从第 701 位的 ATG 起始, 终止于第 5237 位的 TGA。在此 ORF 起始密码子 ATG 前 6 个核苷酸存在一个可能的核糖体结合位点(RBS)。推测此 ORF 为 *R. sphaeroides* 的 *gltB* 基因。它从蛋氨酸(第 701 位核苷酸)起始翻译, 编码一个 1512 个氨基酸的蛋白质, 分子量为 164 kD, 见图 2。它与 *E. coli* GOGAT 大亚单位(166 kD)^[12] 和 *A. brasilense* GOGAT 大亚单位(162 kD)^[7] 大小相似。另外, 在 *R. sphaeroides gltB* 的上游未发现典型的依赖于 σ^* 型启动子的 DNA 保守序列: GG(-24)N₁₀CG(-12), 同时也未能发现类似于 *E. coli* 启动子上游 -10, -35 区的保守 DNA 序列。

2.2 G+C 百分含量及密码子使用频率分析

R. sphaeroides 的 *gltB* 基因 G+C 百分含量为 67.4%, 与 *R. sphaeroides* 总 DNA 的 G+C 百分含量 69%^[11] 相近。在密码子的利用上, 其三联密码子的第三位碱基倾向于利用 GC, 百分含量达到 92.1%。

2.3 *R. sphaeroides gltB* 基因与其它细菌 *gltB* 基因的同源性比较

我们把 *R. sphaeroides gltB* 基因 DNA 序列和已发表的 *A. brasilense gltB* 基因^[7] 及 *E. coli gltB* 基因的 DNA 序列^[12] 进行了比较。*R. sphaeroides gltB* 基因与 *A. brasilense gltB* 基因 DNA 序列同源性很高, 达到 73.3%, 与 *E. coli* 同源性为 58.6%。*R. sphaeroides gltB* 基因与 *A. brasilense gltB* 基因 DNA 序列具有很高的同源性, 与这两种菌中 DNA 的

GAATTCACCGAGACGAGCATCGGTGGCCACCGTCTCGCGCAGCGCCAGGGGAAGCAATATATTCATCCCGAATGGCGTCTCTTCTTCGCGC	90
GGCCCGACCCGATGGAGCGACGGCCACCGGGGCGACGCTGCCGCTGGACGGCAGCCCGTCGATTTTCGCTTCCCGCGGAGCCGGGCTTT	180
CTTGCCACCGAGCCGATCTTCGACTCGAGCGGATCCTGTCCAGGCGCTCGAGGACCGCCTCATCTGAAGGTCGCGCATGACCGCTGA	270
CCGGACCGGACALTCCTCTCGCGCGGGTGCAGATCGCOGTGGCGACCGCCTCGCAOAGATCGGCGCTGCCACTGCGAGATGTGCGCG	360
CCGCTGGACCGGCTCGCTTTCGTACAGATGATGTCGCCGAGGCGCCCATGACCGTGACCGGCGCGAGCATGTGAAGGCGCTATGCCTC	450
GTGCAATTTGGGCGAGCGCTCTCTCGCGCACCTCGGCTCGACGCTCTGGTATCGGCTGACCGGTCCGGGATGCGCGCGACTACTA	540
CATGGCGCGGGGCTTCTGGATGACCTGAGCGGCATCGGCTCGCCATGAGATCTACATCGAGCGCAAGCCGACCGCTCGAGCTTCGC	630
CGGTCCGACCCACAGATGACGGAAGCCGAGTTTCTCGCAOCTTCCAGCCAGCCGAGGATGAGCATGACGATCTATGATGAAGC	720
S.D. M T I Y D E A 7	
CTGGGTGAAAGCCGAAGAGGCGAAGGGAGCCTGGCTGGACGGAACGGACTCTACAAGCCGAGGACGAGCACGCTCTCGCGCGTGGG	810
W V K A E E A K G A W L D A N G L Y K A E D E H A S C G V G	37
CCTCGTCTGATCTCGGGCAGCCGCTCGCGAAGGTCGTGAGAGCGGCATCGCGCGCTGAAGGCGCTTGGCACCGGGGCGCGGT	900
L V V S I S G T P S R K V V E S G I A A L K A V W H R G A V	67
CGATGCCGACCGCAAGACCGGCGAGCGCGGGCATCCATGTCAGATCCCGTCAAGTCTTCTACGACCAGATCCGCGCACCGGCCA	990
D A D G K T G D G A G I H V Q I P V K F F Y D Q I R R T G H	97
CGAGCCGACATGCACCAAGCTCGTGGCGTGGGCGAGGTCTTCTCGCGCGCAOAGTCTCTCGGGCCAGGAACGCTGCCGACCATCGT	1080
E P T C T K L V A V G Q V F L P R T D L S G Q E R C R T I V	127
CGAGTCTGAAGTCTGCGCATGGGCCATTACATCTAAGGCTGGCGCATGTGCGGTGGACACCTCCGTCTCGCGGAGAAGGCCAATGC	1170
E S E V L R M G H Y I Y G W R H V P V D T S V L G E K A N A	157
CAOCCGCGGAGATGAGCAGATCCTGATCGGCTCGGAGAAGGACATGAGCAGCAGCAGTTCGAGCGCGAAGTCTACATCATCCGCG	1260
T R P E I E Q I L I R C E K D I D D E Q F E R E L Y I I R R	187
TCGGATCGAAGGCGGGCGAGCGGGCTCGATCCAGGGGCTCTACCTCTGCTCGTGTGTCGCGGTGCGATCATACAAGGGCATG	1350
R I E K A A Q A G S I Q G L Y L C S L S C R S I I Y K G M M	217
GCTGGCCGAGCAGGTGCCACCTTCTATCCGACCTGACGAGCAGCGGTCGAAAGCTCGTTTCGCGATCTACACCAGCGCTATTCCAC	1440
L A E Q V A T F Y P D L Q D E R F E S S F A I Y H C R Y S T	247
CAACACCTTCCGCAATGGTGGCTGGCGCAGCCCTTCGCGATGCTCGCCCAACCGCGAGATCAACACGCTGAAGGGCAACATCAACTG	1530
N T F P Q W W L A Q P F R M L A H N G E I N T L K G N I N W	277
GATGAAGAGCCACAGATCCGATCGCCCTCTCGGGCTTCGGGGATCGCGCGAGGACATCAAGCCGATGTCGCCGCGAGGCTCGTCCGA	1620
M K S S H E I R M A S S A F G D A A E D I K P I V P A G S S D	307
CTCGGGCGGCTGGATGCGGTGTTCGAGGTCATGGTGGCTCGGGCCGCTCGGGCGGATGGTCAAGACCATGATGGTCCGGGAAGCCTG	1710
S G A L D A V F E V M V R S G R S A P M V K T M M V P E A W	337
GTCGAAGACCACGACGACATGCCGAAGGCTGGCGGACATGTATGCTATTGCAACGCGGTGATGGAGCGGTGGGACGGGCGGGCGGC	1800
S K T T T D M P K A W A D M Y A Y C N A V M E P W D G P A A	367
GCTGGCCATGACGACGCGCTGGGTCTCGCGCGGGCTCGACCCAACCGGCTCGGCCGATGCGCTATGTGGTAACGGGTGACGGGAT	1890
L A M T D G R W V C G G L D R N G L R P M R Y V V T G D G M	397
GCTCATCGGGGCTCGGAAGCCGGCATGGTCCCGGTGGATGAGATGAACGTCGCGGAGAAGGGGCGCCTCGGGCCGGGTGAGCTCATCGC	1980
L I A G S E A G M V P V D E M N V R E K G A L G P G Q L I A	427
GGTCGACATGCCGAGGGCAAGCTCTACCAGACGCGGAGCTCAAGGACACGCTGGCGGCTCGCAGCCCTTCGGCGACTGGATCGAGAA	2070
V D M A E G K L Y H D A E L K D T L A A S Q P F G D W I E K	457
GGTCGTGGACCTGAACCGGATCCTCGCCGAGTGCCTGAGCAGCGCATGTTGCGCCGCGGAGCTGCGCAAGCGCCAGATCGCGGCGGG	2160
V V D L N A I L A D V P E Q R M F A P A E L R K R Q I A A G	487
CTTCTCGGTGAGGAGATCGAGCAGGTGCTGTCGCGATGGCCGAGGACGGCAAGGAGATGATCGCCTCGATGGCGATGACACGCGCCG	2250
E S V E E I E Q V L V P M A E D G K E M I A S M G D T P P	517
CGCGGTGCTGTCGTGGTCTACCGGCGCTCAGCCACTTCTCCGGCAGAATTGACCGAGGTCGCGCCCATCGATCGCTCGCTCG	2340
A V L S S V Y R P L S H F F R Q N F S Q V T N N P P I D S L R	547
CGAGAGCCGGTGTGAGCCCTCAAGACCCGGTTCGGCAACCTCAAGAKNVTGCTCGACGAACATTGAGCCAGACCGGATCCTGATCCT	2430
E S R V M S L K T R F G N L K N V L D E H S S Q T E I L I L	577
CGAGAGCCCTTCGTGCGAANTCCGAGTTCGAGACGATGCTGTGCGAGTTCGGCACCAACGTCGGCGAACCTGACTGCACTTCCCGGC	2520
E S P F V A N S E F E T M L S Q F G T N V A T I D C T F P A	607
CGACCATCACCAGATGCGCTGCGCCAGGGCTCGAGCGGATCCGGGCGAGGCGGAGGATGCGGTGCGCTCGGGTCCGGCCATATG	2610
D H H H D A L R H G L E R I R A E A E D A V R S G A A H I V	637
GCTGACCGACCAAGACCAGGGCCCGGGCGGTGCCGATGCGGATGATCCTCGGACTTCGGCGGTCCACTTCTGGCTGACCGCAAGGG	2700
L T D Q N Q G P G A V P M P M I L A T S A V H S W L T R K G	667

CACCGTGCTTTACGGGCGACGACGGGTTCTCTTCGCTGCCGGCCGCGGGGCGAGCGGTTCCGGGTGCGCAACTCGGGGCCAAGGT 4860
T V L Y G A T D G F L F A A G R A G E R F A V R N S G A K V 1387

GGTGGTCGAGGGCTGCGGCTCGAACGGCTGCGAATATATGACCGGGCGGCGTCCGGGTGATCCTCGGGCGGATCGGGCGCAACTCGGGCC 4950
V V E G C G S N G C E Y M T G G V A V I L G R I G A N F G A 1417

GGCATGACCGGGGGCATGGCTATCTCTACGACCCCTTCGGGGCTGGCAGAGGATTTTCATGAACCTCGAGACGCTCGTGACCTGTGCGAT 5040
G M T G G M A Y L Y D P S G V A E D F M N L E T L V T C A I 1447

GCTGCGGACCTTCACTCGATCAATGTGGCTCGGGCGAGTGCATCGACCCGCAATTATTCGGGTGCTGATCGGGCTCGGGCGCCACCAC 2790
L R T F T S I N V R S A E C I D P H Y F A V L I G C G A T T 697

GGTGAACGCTATCTGGGCGAGGACACCATCGCCGACCCATCGAGCGCGGCTCCTCGAGGGCAGCCTGAACGACGCAATGCBCCGCTA 2880
V N A Y L A Q D T I A D R I E R G L L E G S L N D A M R R Y 727

CCGCGACGCGATCAACCGGGCTTCTGAAGATCATGTGGAAGATGGGGATCTCGGTGATCTCCTCTATCGGGCGGTCTGAACCTCGA 2970
R D A I N A G L L K I M S K M G I S V I S S Y R G G L N F E 757

GGCGGTGGTCTTTCGGGGCCATGGTGGCCGAATATTTCCCGGCGATGCACAGCCGATCTCGGGGATCGGCACCTCGGGCATCCAGCA 3060
A V G L S R A M V A E Y F P G M H S R I S G I G T S G I Q H 787

CAAGCTGGAGCAGATCCACGCCAAGGGCTGGCTCGGGGGTCCGACGTCTGCCATCGGGGGCTTCTACAAGCGCGGCGCTCGGGCGA 3150
K L E Q I H A K G W L G G S D V L P I G G F Y K A R R S G E 817

GAACACGCTCGGAAGCCTCGACCATGCACATGCTCGACGAGCCTGCGACGGGCCAGCTACGACATCTGGAAGCAGTTCTCGGGCCAC 3240
K H A W E A S T M H M L Q Q A C D R A S S Y D I W K F C S A T 847

GCTCGGGGCCAACCCCGCATCCATCTCGCGATCTTCTGGACATCAAGACGCTCGGGCGTCCGCTCGAGGAGGTGGAATCGAT 3330
L R A N P P I H L R D L L D I K T L G R P V P I E E V E S I 877

CACTTCGATCCGCAAGCGGTTCGTGACGCGCGGCTGCGCTGGGCGGCTCTCGCCGAGGCGCACAGAAGCTGAACATCGCCATGAA 3420
T S I R K R F V T P G G M S L G A L S P E A H K T L N I A M N 907

CCGGATCGGGCCCAAGTCGAGCTCGGGCGAAGGCGGCGAGGATCGGCGCATTTCGTGCGGAGCCCAATGGCGACAAACCCCTCGGCCAA 3510
R I G A K S D S G E G G E D P A H F V P E P N G D N P S A K 937

GATCAAGCAGGTGGCTCGGGCGCTTCGGTGTGACGGCGGAATATCTGAACGCTCGGAAGAAGCTCGAGATCAAGGTGGCGCAGGGTGC 3600
I K Q V A S G R F G V T A E Y L N A C E B L E I K V A Q G A 967

CAAAACCGGGGAGGGCGGCCAGCTTCCGGGCATGAAGGTGACCGAGCTCATCGCGGGCTCCGGCATTGACGCGGGGCGTCCAGCTGAT 3690
K P G E G G Q L P G M K V T E L I A R L R H S T P G V T L I 997

CTCGCGCGCGCCACACGACATCTACTCGATCGAAGATCTCGCGCAGCTGATCTATGACCTGAAGCAGATCAACCGCGCGCCCAAGGT 3780
S P P P H H D I Y S I E D L A Q L I Y D L K Q I N P R A K V 1027

GACGGTCAAGCTGGTGGCGGCTCGGGCTGGCCAGGATTCGCGCGGGCGTGGCCAGGCCAAGGCCGACGCTGATCCTGATCTCGGGCCA 3870
T V K L V A A S G V G T I A A G V A K A K A D V I L I S G H 1057

CAACGGCGCACCGGGGCTCGCCCGGCACCTCGATCAAATATCGGGGCTGCGGTGGGAGATGGGGCTGACCGAGGCGCATCAGGTTCT 3960
N G G T G A S P G T S I K Y A G L P W E M G L T E A H Q V L 1087

GGCGATGAACAACTCGCGACCGGGTGCAGCTGCGCACGACGCGGGCCCTCAGGACGGGCGCGACATCGTCATGGCCGCGATGATGG 4050
A M N N L R D R V T L R T D G G L R T G R D I V M A A M M G 1117

GGCCGAGGAATATGGCATCGGCACCGCGCTGATCGCCATGGGCTGCATCGGTGCGCCAGTGCAGTGCAGCACCTGCCCGGTGG 4140
A E E Y G I G T A A L I A M G C I M V R Q C Q S N T C P V G 1147

CGTCTGCACGAGGACAAGAAGTTCGCGGAGAAGTTCACCGGCTCGCGGACAAAGGTGGTGAACCTCATCACCTTCTATGCCAGGAAGT 4230
V C T Q D K K L R E K F T G S A D K V V N L I T P Y A Q E V 1177

GCGCGAGATCCTCGCTCCATCGGGCCCGCTCGATGGACGAGATCATCGCGCGCGCGACCTGCTGACTCAGGTGAGCCGGGGGGCCGC 4320
R E I L A S I G A R S M D E I I G R A D L L T Q V S R G A A 1207

GCATCTCGACGACCTCGACCTCAACCGCTCCTCATCACCCTGGACGGGTGGAACCGCATCACTACGACCGGTGGAAGCCGCGCAACGC 4410
H L D D L D L N P L L I T V D G S N R I T Y D R S K P R N A 1237

GGTCCGGACACGCTCGATGCCGAGATCGTCAAGGATGGTGGCGCTTCTTCGAGGATGGCGAGAAGATGCAGCTCTCCTACCGGTGCG 4500
V P D T L D A E I V K D G A R F F E D G E K M Q L S Y A V R 1267

GAACACGCAACCGACCATCGGCACCGCGCTCGAGCCACATCGTGCAGAAATACGGGATCGGGAACAACCTGCAGCCCGACCATCTGAC 4590
N T H R T I G T R A S S H I V R K Y G M R N N L Q P D H L T 1297

GGTGAACCTGACCGGCTCCTCGGGCAGTCTCTGGGCGCTTCCGGCCAAAGGGGTGAAGATCGAGGTGCGGGCGATGCCAACGACTA 4680
V K L T G S C G Q S L G A F A A K G L K I E V A G D A N D Y 1327

TGTGGCCAAAGGCGTGGGGCGCACCATCGTGGTCCATCGCAGATGGAGAGCCCGCTTGTGGGGCGGAGAACAGATCATCGGCCAA 4770
V A K G L S G G T I V V H P Q M E S P L V A A E N T I I G N 1357

```

CGGCCATCGCACTGGGAGGCGCAGCTGAAGGGCCTGATCGAGCGCCATGTGCGCGAGACGGGCAGCCACGCTGCGCGCATCCTGAA 5130
G H P H W E A Q L K G L I E R H V R E T G S R H A A R I L N 1477

CGACTGGGAGACCGAGCGGGCGAATTCCTGCAGGTCTGCCCGAAGGAGATGCTGGTGCATCTGCCCTTCCGCTGTGCGGACGAGCCGCA 5220
D W E T E R A N F L Q V C P K E M L V H L P F P L S D E F Q 1507

GGCCGTCCCGCGGAGTGAAGAGACGGAGAGGGGGGCTGAGGCCCCCTTTCGATCCGCGTGCCTCCGCTCTCCCGAGCGGCGCGCGCCG 5310
A V P A E *

CGCGGGCGCGCAGAAATCGAGGAGGGCGCCCCGTCATGGCGCGGCCCTCGCAGCGCTGCTCCGGTGGCGGGAAGTTCBCTTGACCC 5400
GCCACGGCATCCGCTGCCACGGGGCCGCTCCGGCCGTCCGCGCAGACCCGACTCCGCTGGGGACGGTTCGTGTCCAAGGTCGGCTC 5490
ATCAGCGCGCAAGTCGAC 5510

```

图2 *R. sphaeroides* *gltB* 基因的核苷酸序列和氨基酸序列
线条表示可能的茎环结构。

Fig.2 Nucleotide and deduced amino acid sequences of *R. sphaeroides* *gltB* gene
The lines indicate the potential stem-loop structure.

高 G+C 百分含量有关。

我们还把 *R. sphaeroides* *gltB* 基因可能编码的蛋白质氨基酸序列与 *A. brasilense* GltB 及 *E. coli* GltB 氨基酸序列进行了比较。同样 *R. sphaeroides* GltB 与 *A. brasilense* GltB

```

A          **          **          *          *          *
R.s. 34  CGVGLVV----62HRGAVDADGKTGDG---240IYHQRYSTNT---381DRNG---
A.b. 1  CGVGFIA----30HRGAVDADGKTGDG---206IYHQRYSTNT---346DRNG---
E.c. 43  CGFGLIA----72HRGAILADGKTGDG---242LFHQRFSTNT---380DRNG---
Z.m. 98  CGVGFVA---127HRGGCGADSDSGDG---304IYHRRFSTNT---447DRNG---

B          CXXXXXXXXXXXXXCP
R.s. 1126 AALIAMGCIMVRQCSNTCPVGVCTQDKKLR 1156
A.b. 1095 ASLIAMGCIMVRQCHSNTCPVGVCVQDDKLR 1125
E.c. 1123 GPMVALGCKYLRICHLNNCATGVATQDDKLR 1153
Z.m. 1222 VAMIATGCVMARICHTNNCVPGVASQREELR 1252

C          *          *          *
R.s. 1030 K---1101DGGLRTGRD
A.b. 999 K---1070DGGGLKTGRD
E.c. 1027 K---1098DGGGLKTGVD
Z.m. 1126 K---1197DGGFRSGQD

D          - + + + + + + + +
R.s. 1378 FAVRNSGAKVVVEGCGSNGCEYMTGGVAVILGRIGANFGAGMTGGMAYLYDPSGVA 1433
A.b. 1346 FAVRNSGATVVVEGCGSNGCEYMTGGTAVILGRVGDNFAAGMTGGMAYVYDLDDSL 1401
E.c. 1376 FGVRNSGAIIVVEGIGDNGCEYMTGGIVCILGKTVGNFAGMTGGFAYVLDSEGDF 1431
Z.m. 1475 FAVRNSLCQAVVEGTGDHCCCEYMTGGCYVVLGKAGRNVAAAGMTGGLAYILDEDDTL 1530

E          - + - - + +
R.s. 1313 AKGLKIEVAGD 1323
A.b. 1281 VQGIKLEVMGD 1291
E.c. 1311 AGGVELYLTGD 1321
Z.m. 1410 TFGMNIIRLVGE 1420

```

图3 *R. sphaeroides* GltB 与 *A. brasilense*、*E. coli*、*Zea mays* GltB 保守区的比较

Fig.3 Comparison of GltB consensus domain among *R. sphaeroides*, *A. brasilense*, *E. coli* and *Zea mays*

A.谷氨酰胺氨基转移酶保守区 A glutamine amidotransferase domain;
B形成[3Fe-4S]中心的半胱氨酸簇 B the cysteine clusters for the formation of the [3Fe-4S] center;
C. FMN结合区 C the FMN binding domain; D. 腺苷酸结合位点 D adenylate binding fold;
E. FAD结合区 E FAD binding domain;

* 保守残基 * conserved residues; + 符合的残基 + residues that match; - 不符合的残基
- residues that do not match;

R. s.: *Rhodobacter sphaeroides*; A. b.: *Azospirillum brasilense*; E. c.: *Escherichia coli*; Z. m.: *Zea mays*

的同源性很高, 达到 66.3%, 而与 *E. coli* *GltB* 的同源性只有 38.8%。

2.4 *R. sphaeroides* *GltB* 的功能区分析

细菌的 *Glt* 和植物的 *Fd-Glt* 有一些共同的特点, 它们都有一个 FMN, 一个 FAD 和一个 $[3\text{Fe}-4\text{s}]^{+0}$ 中心。这些辅因子的结合区和谷氨酰胺氨基转移酶区一般都位于大亚单位。由于电子供体和辅因子的不同, 细菌的 *Glt* 还必须有一个 NADPH 结合区和一个形成二个 $[4\text{Fe}-4\text{S}]^{2+1+}$ 中心的半胱氨酸簇。

我们对 *R. sphaeroides* *GltB* 的功能区进行了分析, 并与 *A. brasilense*, *E. coli* 和玉米 (*Zea mays*) 的 *GltB* 进行了比较, 结果见图 3。

通过对 *R. sphaeroides*, *A. brasilense*, *E. coli* 和玉米 *GltB* 的比较可见, 在它们氨基末端都有一个保守的半胱氨酸残基, 这个保守的半胱氨酸残基参与了谷氨酰胺氨基基因的释放^[13, 14]。在 *R. sphaeroides* 中, 这个保守的半胱氨酸残基位于第 34 位(见图 3-A)。由此可知 *R. sphaeroides* *GltB* 具有一个类似于 PurF 的谷氨酰胺氨基转移酶保守区, 它和其它细菌 *GltB* 一样位于氨基末端的 350 个氨基酸残基以内。

由图 3-B 可见, 在 *R. sphaeroides* *GltB* 中, Cys-1133, Cys-1139 和 Cys-1144 形成了一个 CxxxxCxxxxCP 的半胱氨酸簇。这个半胱氨酸簇在 *A. brasilense*, *E. coli* 和玉米 *GltB* 中都是保守的, 它们参与了 $[3\text{Fe}-4\text{S}]^{+0}$ 中心的形成。

R. sphaeroides *GltB* 羧基末端 1000-1131 氨基酸残基与其它细菌 *GltB* 保守性很强, 特别是 Asp-1101 和 Arg-1105(见图 3C), 它们可能参与了与 FMN 核糖侧链的相互作用。还有第 1030 位的 Lys, 它参与了和 FMN 异咯嗪环的相互作用。

在一些黄素蛋白和依赖于吡啶核苷酸的酶中, FAD 和 NADPH 的腺苷酸部份通常作用于一个甘氨酸丰富区^[15, 16]。由图 3D 可见, *R. sphaeroides* *GltB* 的第 1386 到 1428 位的氨基酸残基中具有一个甘氨酸丰富区, 其中有 10 个氨基酸残基符合了所要求的 11 个腺苷酸结合位点^[15-17], 具有很强的保守性。

另外, *R. sphaeroides* *GltB* 的第 1313 到 1323 位的氨基酸残基可能具有一个结合 FAD 的保守序列, 它与^[18]所述的 FAD 结合位点保守性不是很强(见图 3E)。

参 考 文 献

- [1] Meers J L, Tempest D W, Brown C M. *J Gen Microbiol*, 1970, 64: 187 ~ 194.
- [2] Castano I, Bastarrachea F, Covarrubias A A. *J. Bacteriol*, 1988, 170: 821 ~ 827.
- [3] Castano I, Flores N, Valle F *et al. Mol Microbiol*, 1992, 6: 2733 ~ 2741.
- [4] Bohannon D E, Rosenkrantz M S, Sonenshein A L. *J Bacteriol*, 1985, 163: 957 ~ 964.
- [5] Bohannon D E, Sonenshein A L. *J Bacteriol*, 1989, 171: 4718 ~ 4727.
- [6] Donald R G K, Lapointe J, Ludwig R A. *J Bacteriol*, 1988, 170: 1197 ~ 1204.
- [7] Pelanda R, Vanoni M A, Percgo M *et al. J Biol Chem*, 1993, 268(5): 3099 ~ 3106.
- [8] 鲁涛, 吴永强, 宋鸿遇. 微生物学报, 1996, 36(1): 1 ~ 11.
- [9] Sambrook J, Fritsch, E F, Maniatis, T. *Molecular Cloning: A laboratory Manual*. 2nd ed. New York: Cold Spring Harbor Laboratory, 1989.
- [10] Sanger F, Nicklen S, Coulson A R. *Proc Natl Acad Sci USA*, 1977, 74: 5463 ~ 5467.
- [11] Pfennig N, Truper H G. *Ann Microbiol*, 1983, 134B: 9.
- [12] Oliver G, Gosset G, Sanchez-Pescador R *et al. Gene*, 1987, 60: 1 ~ 11.

- [13] Mei B, Zalkin H. *J Biol Chem*, 1989, **264**: 16613 ~ 16619.
- [14] Mei B, Zalkin H. *J Bacteriol*, 1990, **172**: 3512 ~ 3514.
- [15] Wierenga R K, De Maeyer M C H, Hol W G J. *Biochem*, 1985, **24**: 1346 ~ 1357.
- [16] McKie J H, Douglas K T. *FEBS Lett*, 1991, **279**: 5 ~ 8.
- [17] Karplus P A, Daniels M J, Herriott J R. *Science*, 1991, **251**: 60 ~ 66.
- [18] Eggink G, Engel H, Vriend G *et al*. *J Mol Biol*, 1990, **212**: 135 ~ 142.

NUCLEOTIDE SEQUENCE OF *gltB* GENE ENCODING THE LARGE SUBUNIT OF *RHODOBACTER SPHAEROIDES* GLUTAMATE SYNTHASE

Lu Tao Wu Yongqiang Song Hongyu

(*Institute of Plant Physiology, Chinese Academy of Sciences, Shanghai 200032*)

Abstract The complete nucleotide sequence of a 5.4-kb chromosomal EcoRI-SalI fragment was determined, which contains the structural gene (*gltB*) for the large subunit of *Rhodobacter sphaeroides* glutamate synthase, as well as the 5' - and 3' - flanking regions. A open reading frame of 4636 base pairs was identified as *R. sphaeroides gltB* gene. The MW of the large subunit, as deduced from the nucleotide sequence, was estimated as 164 kD. Comparison of the nucleotide sequences revealed a high similarity among *gltB* genes of *R. sphaeroides*, *Azospirillum brasilense* and *Escherichia coli*. The deduced amino acid sequence of *R. sphaeroides* GltB showed a high similarity with that of *A. brasilense* GltB.

Key words *Rhodobacter sphaeroides*, Glutamate synthase gene, DNA sequence