Volume 14 Number 11 1986 Nucleic Acids Research

The nucleotide sequence of the gene for γ-glutamylcysteine synthetase of *Escherichia coli*

Kunihiko Watanabe*, Yoshinori Yamano, Kousaku Murata and Akira Kimura

Research Institute for Food Science, Kyoto University, Uji, Kyoto 611, Japan

Received 1 May 1986; Accepted 15 May 1986

ABSTRACT

The nucleotide sequence of the gsh I gene for γ-glutamylcysteine synthetase(GSH I) of Escherichia coli B has been determined. The gsh I structural gene contains 1557 bases in length and predicted a ploypeptide of 518 amino acids with a calculated molecular weight of 58,251. The value is in good agreement with that obtained from gel filtration and SDS/PAGE of GSH I. The initiation codon 5 bp downstream of putative Shine-Dalgarno sequence was an unusual TTG, which encoded methionine. For transcription, two sets of consensus promoter signals(-10 and -35 regions) overlapping each other were identified. The terminator signal shows the favored stem-loop structure with an adequate free energy ΔG=-22.80 kcal/mol.

INTRODUCTION

We have previously reported the gene cloning of two glutathione biosynthetic enzymes in Escherichia coli B. They are gsh I for γ-glutamylcysteine sythetase(GSH I) and gsh II for glutathione synthetase(GSH II)(1-3). The nucleotide sequence of gsh II has already been determined(4). The cloned gsh I gene was derived from E. coli B having GSH I desensitized from feedback inhibition of reduced GSH(5). The GSH I encoded by the cloned gene is, therefore, insensitive to reduced GSH, although the wild enzyme is inhibited by reduced GSH(1). To clarify the structural change leading to the desensitization and to express the gene more efficiently, the nucleotide sequence of gsh I was determined and structural properties of the gene were discussed.

METHODS AND MATERIALS

Escherichia coli strains and plasmids

The strains C600 (F^-, thi, thr, leu, lacY, tonA, supE, hsd(r^-_B, m^-_B)), JM103 and JM109 and the vector plasmids mp10,

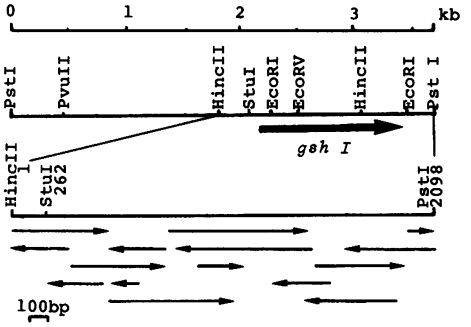

Figure 1. Restriction map of the gsh I gene and sequencing strategy

The gsh I gene was originally cloned in the 3.6 kb PstI-PstI fragment and the 2098 bp HincII-PstI fragment was subjected to sequencing. The length and direction of sequences obtained from M13 subclones are shown as thin arrows below the expanded HincII-PstI fragment, while the gsh I gene is denoted by a thick arrow.

mp11, mp18 and mp19 were all kindly provided by Dr. M. Takanami. The plasmid pGS100 containing gsh I gene has been described previously(1).

Nucleotide sequence determination

DNA fragment carrying the gsh I gene was derived from the plasmid pGS100. To obtain the gsh I sequence data between the HincII and PstI sites(2098bp), this region was digested with various restriction enzymes and subcloned into M13 mp10, mp11, mp18 and mp19 vectors. The single stranded phage DNA prepared from the resulting clones was subjected to the chain termination technique of DNA sequencing according to the suppliers' specifications (Takara Shuzo Co. Ltd.)

Other DNA manipulations

The procedures and conditions used for the isolation of DNAs, digestion with restriction enzymes, agarose gel electrophoresis and transformation of E. coli were the same as those described previously(6).

N-Terminal end analysis of γ-glutamylcysteine synthetase

The amino acid sequence of N-terminal region of GSH I has already been determined from the second to tenth residue by the use of Analytical Biosystem 470-A(7). To obtain the first N-terminal end amino acid, we performed the DABITC method (8), using the enzyme purified as shown before(7).

RESULTS

The nucleotide sequence of gsh I

Former work in this laboratory showed that the E. coli B gsh I gene is carried on the plasmid pGS100 in 3.6 kb length, encoding a 5 5 kD of GSH I(1). Since the gsh I gene was localized in one half of the 3.6 kb PstI fragment, our sequencing strategy was focussed on the 2098bp HincII-PstI fragment. A more detailed restriction map determined from the various clones prepared for the sequencing is shown in Fig. 1. The sequence data obtained from both strands of the 2098bp HincII-PstI fragment are shown in Fig. 2. A computer search for all possible initiation and termination codons in both strands reveals only one open reading frame. The possible open reading frame has an unusual TTG initiation codon at position 338 and a TGA termination codon at position 1891, and is large enough to encompass 5 5 kD polypeptide in molecular weight. There was a stop codon TAA upstreem at position 308 in phase with the open reading frame and no other initiation codons responsible for methionine were not found between the stop codon TAA at position 308 and the unique initiation codon TTG at position 338. The uniformly-spaced DNA ladder pattern neighbor the initiation codon TTG is shown in Fig. 3. The sequence obtained using another single-stranded DNA of the corresponding region also gave the same reading as the initiation codon. On the other hand, the amino terminal analysis with purified GSH I gave the first 10 amino acid residues as follows: Met-Ile-Pro-Asp-Val-Ser-Gln-Ala-Leu-Ala. These amino acid residues completely correspond with the first codon TTG to tenth. The polypeptide deduced from this open reading frame has a calculated molecular weight of 58,251 which is also consistent with that apparent one obtained from SDS-PAGE and molecular sieve behavior(7). The codon usage for the gsh I gene is presented in Table 1.

Translational and transcriptional signals

Promoters for E. coli RNA polymerase have been shown to contain two regions of conserved DNA sequence, located about 10 and 35 bp upstream of the transcriptional startsite(9). Thus, some homology with the E. coli consesus -35 and -10 promoter regions was searched, so that two sets of promoter systems were

37
GTTAACTAACGTTATCCTTGGGCGTTTCTTTCCATTG

87 137
CGTAAAACATCGCGCTGGCAAGAGAAAGCTACCGGCTGGCTGAAACGCTATGGTGCAGTCACGCTATTATTAAGCTGGATGCCCGTGGTTGGCGATTTAC

187 237
TGTGTCTGTTAGCGGGATGGATGCGCATCTCGTGGGGACCGGTAATCTTTTTTTGTGCCTTGGTAAAGCGTTACGCTATGTTGCAGTTGCAGCAGCGAC

-35(P2) 287 -10(P2) 337
CGTTCAGGGCATGATGTGGTGGCACTAATTGTAGGCCTGCACATATGGTCACCATTACAGTTATGCTAATTAAAACGATTTTGACAGGCGGGAGGTCAAT
-35(P1) -10(P1) SD SQ

352 382 412
TTG ATC CCG GAC GTA TCA CAG GCG CTG GCC TGG CTG GAA AAA CAT CCT CAG GCG TTA AAG GGG ATA CAG CGT GGG
Met Ile Pro Asp Val Ser Gln Ala Leu Ala Trp Leu Glu Lys His Pro Gln Ala Leu Lys Gly Ile Gln Arg Gly

427 457 487
CTG GAG CGC GAA ACT TTG CGT GTT AAT GCT GAT GGC ACA CTG GCA ACA ACA GGT CAT CCT GAA GCA TTA GGT TCC
Leu Glu Arg Glu Thr Leu Arg Val Asn Ala Asp Gly Thr Leu Ala Thr Thr Gly His Pro Glu Ala Leu Gly Ser

502 532 562
GCA CTG ACG CAC AAA TGG ATT ACT ACC GAT TTT GCG GAA GCA TTG CTG GAA TTC ATT ACA CCA GTG GAT GGT GAT
Ala Leu Thr His Lys Trp Ile Thr Thr Asp Phe Ala Glu Ala Leu Leu Glu Phe Ile Thr Pro Val Asp Gly Asp

577 607 637
ATT GAA CAT ATG CTG ACC TTT ATG CGC GAT CTG CAT CGT TAT ACG GCG CGC AAT ATG GGC GAT GAG CGG ATG TTG
Ile Glu His Met Leu Thr Phe Met Arg Asp Leu His Arg Tyr Thr Ala Arg Asn Met Gly Asp Glu Arg Met Trp

652 682 712
CCG TTA AGT ATG CCA TGC TAC ATC GCA GAA GGT CAG GAC ATC GAA CTG GCA CAG TAC GGC ACT TCT AAC ACC GGA
Pro Leu Ser Met Pro Cys Tyr Ile Ala Glu Gly Gln Asp Ile Glu Leu Ala Gln Tyr Gly Thr Ser Asn Thr Gly

727 757 787
CGC TTT AAA ACG CTG TAT CGT GAA GGG CTG AAA AAT CGC TAC GGC GCG CTG ATG CAA ACC ATT TCC GGC GTG CAC
Arg Phe Lys Thr Leu Tyr Arg Glu Gly Leu Lys Asn Arg Tyr Gly Ala Leu Met Gln Thr Ile Ser Gly Val His

802 832 862
TAC AAT TTC TCT TTG CCA ATG GCA TTC TGG CAA GCG AAG TGC GGT GAT ATC TCG GGC GCT GAT GCC AAA GAG AAA
Tyr Asn Phe Ser Leu Pro Met Ala Phe Trp Gln Ala Lys Cys Gly Asp Ile Ser Gly Ala Asp Ala Lys Glu Lys

877 907 937
ATT TCT GCG GGC TAT TTC CGC GTT ATC CGC AAT TAC TAT CGT TTC GGT TGG GTC ATT CCT TAT CTG TTT GGT GCA
Ile Ser Ala Gly Tyr Phe Arg Val Ile Arg Asn Tyr Tyr Arg Phe Gly Trp Val Ile Pro Tyr Leu Phe Gly Ala

952 982 1012
TCT CCG GCG ATT TGT TCT TCT TTC CTG CAA GGA AAA CCA ACG TCG CTG CCG TTT GAG AAA ACC GAG TGC GGT ATG
Ser Pro Ala Ile Cys Ser Ser Phe Leu Gln Gly Lys Pro Thr Ser Leu Pro Phe Glu Lys Thr Glu Cys Gly Met

1027 1057 1087
TAT TAC CTG CCG TAT GCG ACC TCT CTT CGT TTG AGC GAT CTC GGC TAT ACC AAT AAA TCG CAA AGC AAT CTT GGT
Tyr Tyr Leu Pro Tyr Ala Thr Ser Leu Arg Leu Ser Asp Leu Gly Tyr Thr Asn Lys Ser Gln Ser Asn Leu Gly

1102 1132 1162
ATT ACC TTC AAC GAT CTT TAC GAA TAC GTA GCG GGC CTT AAA CAG GCA ATC AAA ACG CCA TCG GAA GAG TAC GCG
Ile Thr Phe Asn Asp Leu Tyr Glu Tyr Val Ala Gly Leu Lys Gln Ala Ile Lys Thr Pro Ser Glu Glu Tyr Ala

1177 1207 1237
AAG ATT GGT ATT GAG AAA GAC GGT AAG AGG CTG CAA ATC AAC AGC AAC GTG TTG CAG ATT GAA AAC GAA CTG TAC
Lys Ile Gly Ile Glu Lys Asp Gly Lys Arg Leu Gln Ile Asn Ser Asn Val Leu Gln Ile Glu Asn Glu Leu Tyr

1252 1282 1312
GCG CCG ATT CGT CCA AAA CGC GTT ACC CGC AGC GGC GAG TCG CCT TCT GAT GCG CTG TTA CGT GGC GGC ATT GAA
Ala Pro Ile Arg Pro Lys Arg Val Thr Arg Ser Gly Glu Ser Pro Ser Asp Ala Leu Leu Arg Gly Gly Ile Glu

1327 1357 1387
TAT ATT GAA GTG CGT TCG CTG GAC ATC AAC CCG TTC TCG CCG ATT GGT GTA GAT GAA CAG CAG GTG CGA TTC CTC
Tyr Ile Glu Val Arg Ser Leu Asp Ile Asn Pro Phe Ser Pro Ile Gly Val Asp Glu Gln Gln Val Arg Phe Leu

1402 1432 1462
GAC CTG TTT ATG GTC TGG TGT GCG CTG GCT GAT GCA CCG GAA ATG AGC AGT AGC GAA CTT GCC TGT ACA CGC GTT
Asp Leu Phe Met Val Trp Cys Ala Leu Ala Asp Ala Pro Glu Met Ser Ser Ser Glu Leu Ala Cys Thr Arg Val

1477 1507 1537
AAC TGG AAC CGG GTG ATC CTC GAA GGT CGC AAA CCG GGT CTG ACG CTG GGT ATC GGC TGC GAA ACC GCA CAG TTC
Asn Trp Asn Arg Val Ile Leu Glu Gly Arg Lys Pro Gly Leu Thr Leu Gly Ile Gly Cys Glu Thr Ala Gln Phe

1552 1582 1612
CCG TTA CCG CAG GTG GGT AAA GAT CTG TTC CGC GAT CTG AAA CGC GTC GCG CAA ACG CTG GAT AGT ATT AAC GGC
Pro Leu Pro Gln Val Gly Lys Asp Leu Phe Arg Asp Leu Lys Arg Val Ala Gln Thr Leu Asp Ser Ile Asn Gly

1627 1657 1687
GGC GAA GCG TAT CAG AAA GTG TGT GAT GAA CTG GTT GCC TGC TTC GAT AAT CCC GAT CTG ACT TTC TCT GCC CGT
Gly Glu Ala Tyr Gln Lys Val Cys Asp Glu Leu Val Ala Cys Phe Asp Asn Pro Asp Leu Thr Phe Ser Ala Arg

1702 1732 1762
ATC TTA AGG TCT ATG ATT GAT ACT GGT ATT GGC GGA ACA GGC AAA GCA TTT GCA GAA GCC TAC CGT AAT CTG CTG
Ile Leu Arg Ser Met Ile Asp Thr Gly Ile Gly Gly Thr Gly Lys Ala Phe Ala Glu Ala Tyr Arg Asn Leu Leu

1777 1807 1837
CGT GAA GAG CCG CTG GAA ATT CTG CGC GAA GAG GAT TTT GTA GCC GAG CGC GAG GGC TTC GAA CGC CGT CAG CAG
Arg Glu Glu Pro Leu Glu Ile Leu Arg Glu Glu Asp Phe Val Ala Glu Arg Glu Gly Ser Glu Arg Arg Gln Gln

1852 1882 1917
GAA ATG GAA GCC GCT GAT ACC GAA CCG TTT GCG GTG TGG CTG GAA AAA CAC GCC TGA CAGAAAAGAAAAAGGCCACTCGT
Glu Met Glu Ala Ala Asp Thr Glu Pro Phe Ala Val Trp Leu Glu Lys His Ala ***

1967 2017
GAGTGGCCAAAATTTCATCTCTGAATTCAGGGATGATGATAACAAATGCGCGTCTTTCATATACTCAGACTCGCCTGGGAAGAAAGAGTTCAGAAAATTT

2067
TTAAAAAAATTACCGGAGGTGGCTAAATGCCGTTGTTAGATAGCTTCACAGTCGATCATACCCGGATGGAAGCGCCTGCAG

found; one(P1) is at position 256(GTGGCA) for -35 region and at position 281(TATGGT) for -10 region, and the other(P2) is at position 275(TGCACA) for -35 and at postion 299(TATGCT) for -10.

DISCUSSION

Some interesting features were found out from the determination of the nucleotide sequence of the gsh I gene. The most striking feature of the primary structure is the uncommon initiation codon TTG for methionine. Usually two kind of codons ATG and GTG have been accepted to initiate translation. However, since the first report of identification of the TTG codon as a translational initiation codon in vivo was published(10), several structural genes of prokaryotes(11-16) have been shown to have TTG as an initiation codon. As for the gsh I gene,the codon TTG encodes the first N-terminal residue Met obtained from purified GSH I. Reportedly, the isolated anticodon of an initiator tRNA ($tRNA_i$) binds to both ATG and GTG, while fragments of $tRNA_i$ containing the T as well as anticodon arms bind TTG(17) . Therefore, the T arm could have been utilized for pairing with TTG acting as an anticodon. As the result of searching for the sequence complementary to T loop(3'-UAAACΦU-5'), there was a heptanucleotides sequence 5'-ATTTTGA-3' between position 315 and 321, where five out of seven nucleotides were in agreement with complement of T arm. And also another report (18) suggested that the sequence complementary to the D loop(3'-ADGGUCCGA-5')(D, 5,6-dihydrouridine) of $tRNA_i$ exists in both upstream and downstream vicinities of an initiation codon. We found the interactive sequences to a D loop in both sides of TTG codon; 5'-TGACAGGCG-3' at position 319 and 5'-TCACAGGCG at position 353. (The underlined nucleotides indicate complement to D loop.) Furthermore, the polypurine strech known as Shine-Dalgarno sequence(SD sequence) is observed(GGAGG), preceeding 5 nucleotide to an initiator codon

Figure 2. Nucleotide sequence of the gsh I gene
The 2098 bp HincII-PstI fragment franking the open reading frame with its deduced 517 amino acids translation product are shown. The two sets of -10 and -35 promoter consensus sequences are indicated with P1 and P2, respectively. The ribosome-binding site of gsh I(Shine-Dalgarno sequence) is underlined with thick bar(SDsq). The inverted repeat sequence is shown with arrows.

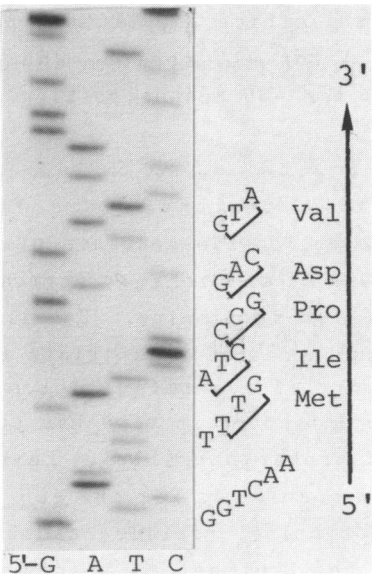

Figure 3. Autoradiogram of DNA sequencing gel showing the translational initiating region of the gsh I gene

The sequence ladder shows the existence of TTG codon for translational initiation. The single stranded DNA used for sequencing reaction was derived from the clone which harbored the 269 bp StuI-EcoRI fragment of gsh I in mp19. The sequence reading starts from position 331 bp in the direction of 5↳3'.

Table 1. Codon usage of the *gsh I* gene of *Escherichia coli* B

Phe	TTT	9	(1.7)*	Pro	CCT	4	(0.8)	Lys	AAA	19	(3.7)
	TTc	13	(2.5)		CCC	1	(0.2)		AAG	4	(0.8)
Leu	TTA	6	(1.2)		CCA	6	(1.2)	Asp	GAT	22	(4.2)
	TTG	5	(1.0)		CCG	14	(2.7)		GAC	5	(1.0)
	CTT	5	(1.0)	Thr	ACT	5	(1.0)	Glu	GAA	30	(5.8)
	CTC	4	(0.8)		ACC	11	(2.1)		GAG	13	(2.5)
	CTG	34	(6.6)		ACA	6	(1.2)				
Ile	ATT	19	(3.7)		ACG	7	(1.3)	Cys	TGT	4	(0.8)
	ATC	11	(2.1)	Ala	GCT	4	(0.8)		TGC	5	(3.0)
	ATA	1	(0.2)		GCC	9	(1.7)	Trp	TGG	8	(1.5)
Met	ATG	12	(2.3)		GCA	13	(2.5)	Arg	CGT	13	(2.5)
	TTG	1	(0.2)		CCG	17	(3.3)		CGC	16	(3.3)
Val	GTT	5	(1.0)	Tyr	TAT	10	(1.9)		CGA	1	(0.2)
	GTC	3	(0.6)		TAC	11	(2.1)		CGG	2	(0.4)
	GTA	4	(0.8)	His	CAT	4	(0.8)		AGG	2	(0.4)
	GTC	9	(1.7)		CAC	2	(0.4)	Gly	CGT	17	(3.3)
Ser	TCT	11	(2.1)	Gln	CAA	6	(1.2)		GCC	18	(3.5)
	TCC	2	(0.4)		CAG	14	(2.7)		GGA	3	(0.6)
	TCA	1	(0.2)	Asn	AAT	9	(1.7)		GGG	3	(0.6)
	TCG	7	(1.3)		AAC	9	(1.7)	Ter	TGA	1	(0.2)
	AGT	3	(0.6)								
	ACC	6	(1.2)								

*:(%)

TTG. The various evidence shown above strongly supports that the synthesis of this polypeptide is uniformly initiated at the assigned codon TTG.

The consensus sequences of promoters being used by E. coli RNA polymerase are recognized within 50 bp upstream ofSD sequence. Two sets of promoter(P1 and P2) of -10 and -35 consensus regions exist, overlapping each other. Each region holds the same homology ratio, i.e. four out of six nucleotides are matching with the consensus sequences for -10 and -35 regions. Besides this promoter system found in E. coli gsh I, the gene for adenylate cyclase is also reported to have consecutive promoter systems in the limited segment(16).

Two kinds of transcriptional terminator have been distinguished in E. coli with or without an additional factor rho to terminate. There exist two stable G-C rich dyad symmetry structures 10 bp downstream of the translational terminator TGA with ΔG=-22.80 kcal/mol (20). Moreover, there is rather A-T rich regions up and downstreams of two dyad symmetries. These characteristics meet well with the rho-independent termination signal. However, for lack of successive thymidine residues, the type of termination cannot be decided and such an argument needs more precise experiments.

The calculated molecular weight of GSH I deduced from the open reading frame was 58,251 D which was in good agreement with the molecular weight of 55 kD estimated from SDS-polyacrylamide gel electrophoresis and Sephadex G-150(7). The nonrandom codon usage in the gsh I gene was observed in accordance with the choice of synonymous codons in E. coli except for TTG for methionine (Table 1). Judging from the frequency of optimal codon usage, F_{op}=0.67, the gsh I gene is expressed modestly in E. coli cells(20). However, the initiation codon TTG for methionine is quite a biased choice in the gene expression. Therefore, it is of great interest to examine the expression of the gsh I gene by the replacement of TTG to ATG as shown in case of the gene for adenylate cyclase (21). And it is also of interest to compare the gene for the wild enzyme with the cloned gene for the desensitized enzyme, and then to discuss the mechanism of inhibition by reduced GSH.

ACKNOWLEDGEMENTS

We thank Dr. Mitsuru Takanami and Mr. Takashi Aoyama, Institute for Chemical Research of Kyoto University, for providing experimental installations and valuable discussion. And also we thank Dr. Morihiko Sakaguchi for his continuous encouragement.

*To whom correspondence should be addressed

REFERENCES

1. Gushima, H., Miya, T., Murata, K. and Kimura, A. (1983) Agric. Biol. Chem., 47, 1927-1928
2. Murata, K., Miya, T., Gushima, H. and Kimura, A. (1983) Agric. Biol. Chem., 47, 1381-1383
3. Gushima, H., Miya, T., Murata, K. and Kimura, A. (1983) J. Appl. Biochem., 5, 43-52
4. Gushima, H., Yasuda, S., Soeda, E., Yokota, M. and Kimura, A. (1984) Nucl. Acids Res., 12, 6797-6812
5. Murata, K. and Kimura, A. (1982) J. Gen. Microbiol., 128, 1047-1052
6. Manaiatis, T., Fritsh, E. F. and Sambrook, J., (1982) Molecular Cloning, Cold Spring Harbor
7. Watanabe, K., Murata, K. and Kimura, A. (1986) Agric. Biol. Chem., Submitted for publication
8. Chang, J. Y., Brauer, D., Wittman-Liebold, B. (1978) FEBS Lett., 93, 205
9. Rosenberg, M. and Court, D, (1979) Ann. Rev. Genet., 13, 319-353
10. Files, J. G., Weber, K., Coulondre, C. and Miller-Hill, B. (1975) J. Mol. Biol., 95, 327-330
11. Buchel, B. E., Gronenborn, B. and Muller-Hill, B. (1980) Nature(London), 283, 541-545
12. Mackie, G. A. (1981) J. Biol. Chem., 256, 8177-8182
13. Young I. G., Rodgers, B. L., Campbell, H. D., Jaworawsky, A. and Shaw, D. C. (1981) Eur. J. Biochem., 116, 165-170
14. Weyens, G., Rose, K., Falmagne, P., Glansdorff, N. and Pierard, A. (1985) Eur. J. Biochem., 150, 111-115
15. Valentin-Hansen, P., Hammer-Jespersen, K. Boetius, F. and Svendsen, I. (1984) EMBO J., 3, 179-183
16. Roy, A., Haziza, C. and Danchin, A. (1983) EMBO J., 2, 791-797
17. Dube, S. K., Rudland, P. S., Clark, B. F. C. and Marker, K. A. (1969) Cold Spring Habor Symp. Quant. Biol., 34, 161-166
18. Ganoza, M. C., Marbiere, E. C., Kofoid, E. C. and Louis, B. G. (1985) Proc. Natl. Acad. Sci., 82,4587-4591
19. Zucker, M. and Stiegler, P. (1981) Nucl. Acids Res., 9, 133-148
20. Ikemura, T. and Ozeki, H. (1983) Cold Spring Harbor Symp. Quant. Biol., 47, 1087-1097
21. Reddy, P., Peterkofski, A. and Mckenny, K. (1985) Proc. Natl. Acad. Sci., 82, 5656-5660