Biochemistry

 Volume 22, Number 23 November 8, 1983

Accelerated Publications

Complete Sequence of the *glt* A Gene Encoding Citrate Synthase in *Escherichia coli*†

Sarbjit S. Ner, Vipin Bhayana, Alexander W. Bell, Ian G. Giles, Harry W. Duckworth, and David P. Bloxham*

ABSTRACT: The sequence of a 3265 base pair *Hin*dIII–*Eco*RI restriction fragment spanning the *glt* A gene from *Escherichia coli* was determined. Translation in the direction *Hin*dIII → *Eco*RI revealed an open reading frame coding for a biosynthetic product of 427 amino acids, which corresponds to the chemically determined protein sequence for *E. coli* citrate synthase. The sequences for initiation and termination of transcription as well as ribosome binding are predicted. The frequency of codon usage within *glt* A is consistent with that expected for a gene which is strongly expressed. Comparison of the structure of *E. coli* with pig heart citrate synthase revealed strong sequence conservation of residues in the active site. An exception is that *E. coli* citrate synthase does not contain a residue equivalent to Arg-46 in the pig heart sequence. This residue is implicated in binding the 5′-diphosphate of coenzyme A, and the absence of this feature might contribute to diminished affinity for coenzyme A derivatives and altered regulatory properties of *E. coli* citrate synthase.

The citrate synthase [citrate oxaloacetate-lyase [coenzyme A (CoA)[1] acetylating], EC 4.1.3.7] gene locus is known as *glt* A and is located at 16.2 min on the *Escherichia coli* linkage map (Ashworth et al., 1965). The *glt* A gene has been located on an approximately 3240 base pair *Hin*dIII–*Eco*RI restriction fragment (Guest, 1981). This fragment, as well as the *Hin*dIII–*Sal*I fragment, has been subcloned into pBR322 and has proved especially useful in amplifying the expression of enzyme protein (Duckworth & Bell, 1982; Bloxham et al., 1983b). In order to complement the work on protein sequence, we have determined the corresponding gene sequence. Experiments proving the relationship of the *glt* A promoter to the translation start site have already been presented (Bloxham et al., 1983a).

Materials and Methods

Materials. Restriction endonucleases, T_4 DNA ligase, M13mp8, and M13mp9 were obtained from BRL Laboratories (Cambridge, England). The 15-mer sequence primer and dideoxynucleoside triphosphate were purchased from P-L Biochemicals (Northampton, England). DNA polymerase I (Klenow fragment) was from Boehringer Corp. Ltd. (Lewes, England).

Preparation of Double-Stranded DNA. Construction and isolation of a hybrid plasmid containing the citrate synthase structural gene and the mapping and isolation of the *Hin*dIII–*Eco*RI fragment containing the structural gene were performed as previously described (Bloxham et al., 1983a,b). From this fragment, *Bgl*II(2)–*Eco*RI (0.27-kb), *Bam*H1–*Eco*RI (2.53-kb), *Bam*H1–*Bgl*II (2-kb), *Hin*dIII–*Bgl*II(1) (2.72-kb), and *Hin*dIII–*Bam*H1 (0.73-kb) fragments were generated with double restriction enzyme digests as outlined previously (Bloxham et al., 1983a) and cloned in M13mp8 or M13mp9 (Messing & Vieira, 1982).

Restriction Endonuclease Digestion. DNA digestions were performed as follows: *Eco*RI in 100 mM Tris-HCl (pH 7.5), 50 mM NaCl, and 10 mM $MgCl_2$ at 37 °C; *Acc*I, *Bam*H1, *Bgl*II, *Hin*dIII, *Sau*3A, and *Taq*I in 6 mM Tris-HCl (pH 7.5), 60 mM NaCl, 6 mM $MgCl_2$, and 6 mM 2-mercaptoethanol

† From the Department of Biochemistry, University of Southampton, Southampton SO9 3TU, U.K. (S.S.N., I.G.G., and D.P.B.), and the Department of Chemistry, University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2 (V.B., A.W.B., and H.W.D.). *Received July 25, 1983.* This work was supported by a project grant (GR/B82639) from SERC (to D.P.B.) and by operating grants from the Natural Sciences and Engineering Research Council of Canada and the University of Manitoba Research Board (to H.W.D.).

[1] Abbreviations: CoA, coenzyme A; kb, kilobase; Tris, tris(hydroxymethyl)aminomethane; ATP, adenosine 5′-triphosphate; EDTA, ethylenediaminetetraacetic acid; NADH, reduced nicotinamide adenine dinucleotide; CoASH, sulfhydryl-containing coenzyme A.

0006-2960/83/0422-5243$01.50/0

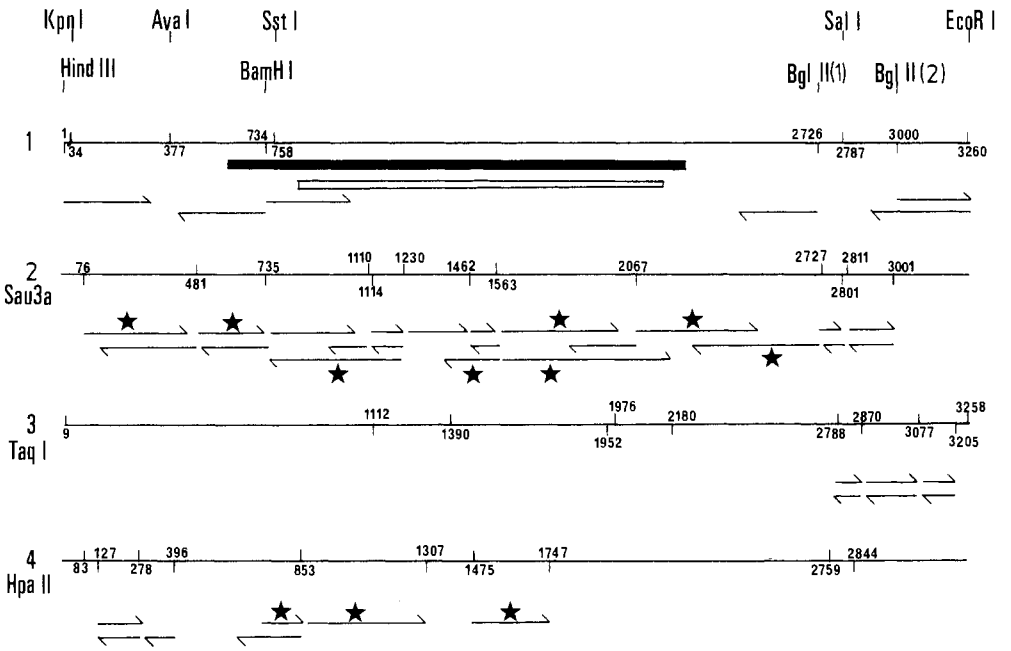


FIGURE 1: Sequence strategy for the *E. coli glt* A gene. The numbered lines (left) indicate the following: (1) restriction map of the *Hin*dIII–*Eco*RI fragment [the solid bar indicates the predicted length of the primary transcript of the *glt* A gene; residue 587 corresponds to the predicted site for initiation of transcription (Figure 2); the open bar indicates the translated region]; (2) *Sau*3A restriction map; (3) *Taq*I restriction map; (4) *Hpa*II restriction map. Numbers on restriction sites indicate the position of the first base. Sequences actually determined are indicated by the arrowed lines beneath the restriction map. The star indicates that sequences were located in partial digests of T_4 DNA ligase polymerized fragments.

at 37 °C except for *Taq*I (65 °C). Partial *Sau*3A digestion was performed in the appropriate buffer at 30 °C for 3 min. Partial *Hpa*II digestion was performed in the *Hpa*II buffer at 37 °C for 9 min.

Polymerization of DNA Fragments. Double-stranded *Hin*dIII–*Bgl*II(1) and *Bam*H1–*Bgl*II(1) fragments were individually polymerized to generate large molecular weight fragments. DNA (30–50 μg) in 20 mM Tris-HCl (pH 8.0) containing 10 mM $MgSO_4$, 1.5 mM dithiothreitol, and 2 mM ATP was incubated with 1 unit of T_4 DNA ligase at 4 °C for 48 h. Polymerized DNA was isolated by ethanol precipitation at −70 °C, and polymerization was confirmed by electrophoresis on 0.9% (w/v) agarose gels in 40 mM Tris–acetate buffer (pH 8.1) containing 2 mM EDTA on minigels (8 × 6 cm) run at 100 V for 2 h.

Isolation of Partially Digested Fragments. Polymerized *Hin*dIII–*Bgl*II(1) and *Bam*H1–*Bgl*II(1) fragments were partially digested, and DNA precipitated out from ethanol at −70 °C. Partially digested DNA was electrophoresed on 0.9% agarose gels in 40 mM Tris–acetate buffer (pH 8.1) containing 2 mM EDTA for 18 h at 1.5 V cm^{-1}. Bands of DNA greater than 300 base pairs were cut out and extracted by the freeze–thaw technique described by Bloxham et al. (1983a). The resultant DNA was dissolved in 50 μL of 10 mM Tris-HCl, pH 8.0, containing 0.1 mM EDTA, ready for ligation.

Construction of Sau3A, HpaII, and TaqI Restriction Fragment Libraries. Fragments from *Sau*3A digests were ligated into *Bam*H1-cleaved M13mp9. Fragments from *Hpa*II and *Taq*I digests were ligated into *Acc*I-cleaved M13mp9. Ligation reactions were carried out in 60 mM Tris-HCl buffer (pH 7.5) containing 10 mM $MgCl_2$, 10 mM dithiothreitol, and 0.05 mM Na_2ATP. T_4 DNA ligase (0.2 unit) was added and the mixture incubated at 4 °C for 24 h. Competent *E. coli* JM103 cells were prepared as outlined (Bloxham et al., 1983a) and were transformed by each ligation reaction mixture. M13 bacteriophage single-stranded DNA containing inserted fragments was isolated according to the procedure of Sanger et al. (1980). DNA was sequenced by a dideoxynucleoside triphosphate chain termination procedure (Sanger et al., 1977; Smith, 1980). Data were analyzed according to a program kindly provided by Dr. Roger Staden (Cambridge, England).

Results and Discussion

Sequence Strategy. DNA fragments sequenced in this work are shown in Figure 1 along with sites for restriction endonucleases. The relative positions for *Hin*dIII, *Ava*I, *Bam*H1, *Sst*I, *Bgl*II (sites 1 and 2), and *Eco*RI were originally placed by restriction mapping of plasmids from the *E. coli* gene bank (Guest, 1981; Bloxham et al., 1983b), and the current solution confirms their presence. The present sequence revises the position for *Kpn*I cleavage, which was incorrect in the original map (Guest, 1981). The presence of this single site was confirmed by restriction mapping with *Kpn*I.

Initially, the purified *Hin*dIII–*Eco*RI fragment, the *Hin*dIII–*Bam*H1 fragment, the *Hin*dIII–*Bgl*II(1) fragment, the *Bgl*II(2)–*Eco*RI fragment, and the *Bam*H1–*Eco*RI fragment were ligated into the appropriately cleaved M13mp8 and M13mp9 to construct a series of modified phage (M13 *glt* A1→10; Bloxham et al., 1983a). By studying the expression of citrate synthase in *E. coli* strain JM103 infected with these phage, we proved that the *glt* A promoter was on the 5′-side of the *Bam*H1 site whereas the initiation of translation was on the 3′-side of *Bam*H1 (Bloxham et al., 1983a). Single-stranded DNA from each of the new phage was sequenced. This gave approximately 45% of the complete sequence but revealed very little of the structural gene apart from the amino-terminal sequence (Bloxham et al., 1983a; Duckworth & Bell, 1982).

To generate additional sequences, we used a "shotgun" cloning procedure (Messing et al., 1981). Initially, we sequenced a library of *Sau*3A restriction fragments from the *Hin*dIII–*Eco*RI fragment. This resulted in a series of predominantly short sequences since these have the highest frequency of insertion into the cloning vector (M13mp9). To

Table I: Triplet Codon Usage in the *glt* A Gene

F	UUU	9	S	UCU	7	Y	UAU	6	C	UGU	4
F	UUC	14	S	UCC	7	Y	UAC	10	C	UGC	3
L	UUA	0	S	UCA	3	term	UAA	1	term	UGA	0
L	UUG	1	S	UCG	4	term	UAG	0	W	UGG	3
L	CUU	0	P	CCU	2	H	CAU	5	R	CGU	16
L	CUC	4	P	CCU	0	H	CAC	10	R	CGC	8
L	CUA	0	P	CCA	3	Q	CAA	1	R	CGA	0
L	CUG	30	P	CCG	14	Q	CAG	7	R	CGG	0
I	AUU	15	T	ACU	5	N	AAU	6	S	AGU	1
I	AUC	13	T	ACC	16	N	AAC	11	S	AGC	3
I	AUA	0	T	ACA	3	K	AAA	18	R	AGA	0
M	AUG	19	T	ACG	4	K	AAG	5	R	AGG	0
V	GUU	8	A	GCU	9	D	GAU	15	G	GGU	17
V	GUC	3	A	GCC	8	D	GAC	12	G	GGC	8
V	GUA	1	A	GCA	7	E	GAA	22	G	GGA	2
V	GUG	7	A	GCG	13	E	GAG	4	G	GGG	1

generate overlapping sequences, we modified the approach as follows. Essentially, the *Hin*dIII–*Bgl*II(1) fragment was end–end polymerized by incubation with T_4 DNA ligase and partially digested with *Sau*3A. Polymerization is essential prior to partial digestion since otherwise the fragments are derived predominantly from the 5′- and 3′-terminals. Following partial digestion, all fragments greater than 300 base pairs were isolated and sequenced after cloning. The sequence data from this library virtually completed the structure apart from two overlaps at 1230–1233 and 1563–1566. These overlaps were established by sequencing a library of *Hpa*II restriction fragments from the T_4 DNA ligase polymerized *Bam*H1–*Bgl*II(1) fragment. Additional sequence data were also obtained from a *Taq*I and *Hpa*II digest of the *Hin*dIII–*Eco*RI fragment. The final solution shows only a minimal overlap of six bases at the *Bgl*II(1) site (2726–2731).

Structure of the glt A Gene. The sequence of 3265 bases of the *Hin*dIII–*Eco*RI fragment containing the entire *glt* A gene is shown in Figure 2. The sequence is numbered from the putative site of transcription initiation. This is based on the location of a strong promoter sequence consisting of a T/A-rich region (−93 to −58), a −35 recognition site (−38 to −32), and a Pribnow box (−12 to −6) on the 5′-side of the *Bam*H1 restriction site (148–153; Bloxham et al., 1983a). Translation using the standard genetic code indicates that there is only one open reading frame sufficiently large to accommodate *E. coli* citrate synthase, and this is preceded by a repeated strong ribosome binding site (284–298; Bloxham et al., 1983a). The translation stop codon is followed by a sequence (1621–1652) that resembles other transcription termination structures in *E. coli* (Rosenberg & Court, 1979) in the possession of a stretch of T residues preceded by a G-C-rich region, which could form a hairpin-loop structure.

On the basis of DNA sequence we predict that *E. coli* citrate synthase is biosynthesized with 427 amino acids (mol wt 48 069). The protein molecular weight and predicted amino acid composition agree very closely with the most recent experimental estimates of these parameters (Duckworth & Bell, 1982). Furthermore, 377 residues of the protein sequence have been determined (underlined in Figure 2). The details of the chemical determination of protein sequence will be published at a later date. The residues that have been determined show agreement with the predicted sequence apart from the fact that the initiator codon (Met-1) is lost during biosynthesis. Asp-10 is present in the enzyme isolated by the Duckworth & Bell (1982) procedure rather than Asn-10 predicted by the gene sequence. By use of plasmid-encoded enzyme purified at Southampton (Bloxham et al., 1983b), the amino terminal was redetermined (K. A. Walsh, Seattle) and found to contain asparagine at residue 10, confirming the accuracy of the gene sequence prediction.

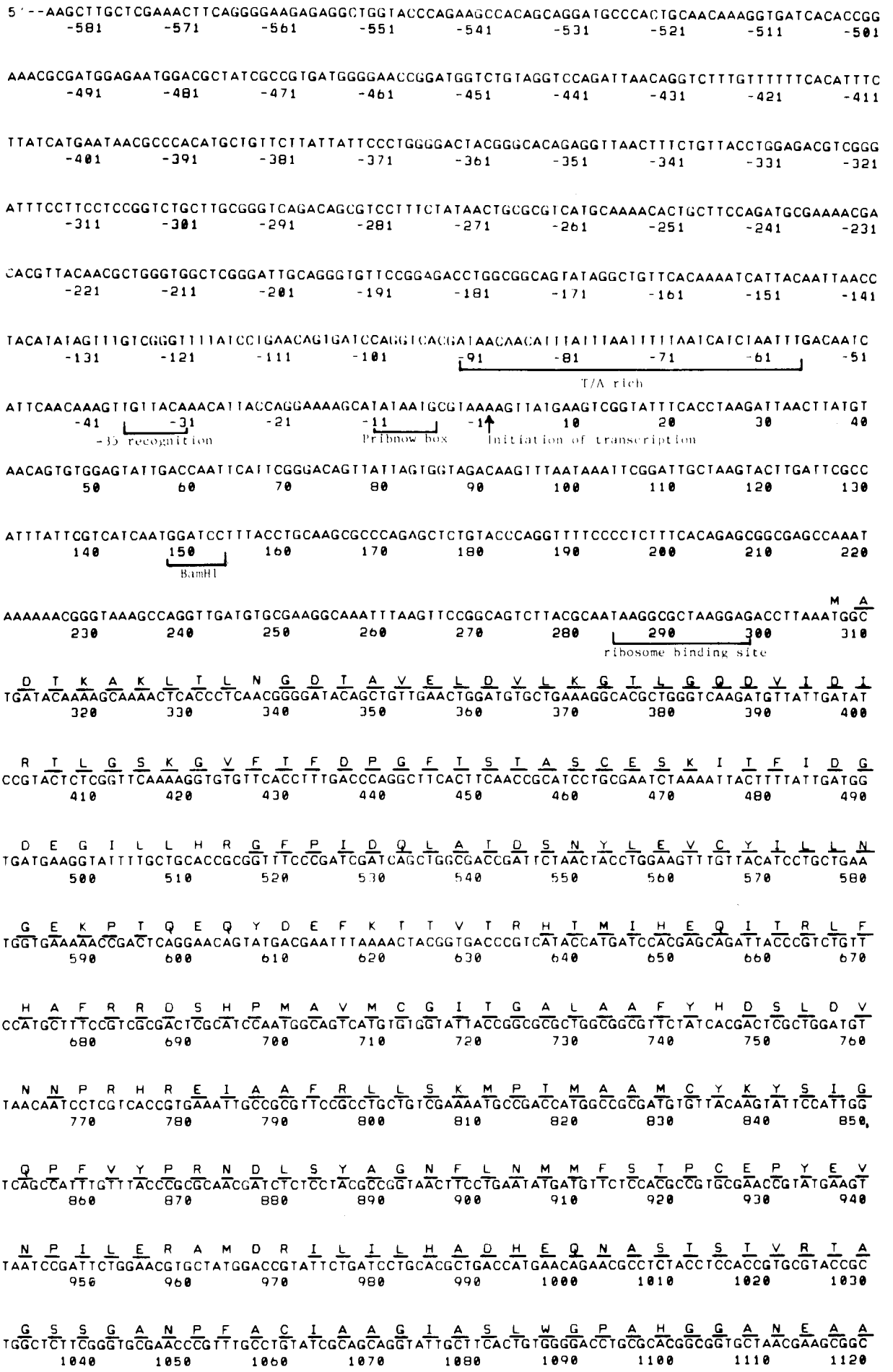
Table II: Prediction of Equivalent Residues in Pig Heart and *E. coli* Citrate Synthase[a]

ligand bound in crystal structure	residue in heart citrate synthase[b]	equivalent residue in *E. coli* citrate synthase
citrate	Arg-329	Arg-314
	Arg-401	Arg-387
	Arg-421′[c]	Arg-407
	His-235	His-226
	His-238	His-229
	His-274	His-264
	His-320	His-305
	Asp-375	Asp-362
CoASH 3′-phosphate	Arg-164′[c]	Arg-157
CoASH 5′-diphosphate	Arg-46	none
	Arg-324	Lys-309
	Me_3Lys-368	Lys-355 or -356

[a] Pig heart and *E. coli* citrate synthase were aligned to give the optimum sequence match for all residues. [b] Residues involved in ligand binding were identified from the X-ray crystal structure (Remington et al., 1982). [c] These side chains are donated by the second subunit in the dimer.

Codon Selection. Codon usage in the translated portion of the *glt* A gene (Table I) shows many of the features predicted for *E. coli* genes encoding strongly expressed proteins (Grantham et al., 1981; Grosjean & Fiers, 1982). In quartet codons (amino acids coded by four triplets) there is a preference for a pyrimidine at the third base position with the notable exception of CUG, CCG, and GCG. However, both CUG (leucine) and CCG (proline) are generally used preferentially in strongly expressed proteins. A further distinction in quartet codons is that U is preferred to C in the third base position. This is shown for GUU, CCU, GGU, CGU, and GCU but not for UCU and ACU. In contrast, for duet codons, C is frequently preferred to U in the third position, and this is shown for UUC, UAC, CAC, AAC, and AGC in the *glt* A gene. A final feature of high expression in *glt* A is the more frequent use of AAA, CAG, and GAA in duet codons where the choice is A or G at the third base position. Features of strong expression within the structural gene would be consistent with the strong promoter structure proposed for the *glt* A gene (see Figure 2; Bloxham et al., 1983b). Further, when *glt* A was inserted into M13 and used to infect *E. coli* strain JM103, citrate synthase accounted for nearly 30% of the total soluble protein in the bacteria (Bloxham et al., 1983a), showing that under appropriate conditions it is strongly expressed.

Sequence Homology. The availability of the amino acid sequence for both *E. coli* and pig heart (Bloxham et al., 1982) citrate synthase allows comparison of their sequence homology. Furthermore, the X-ray crystal structure of heart citrate synthase has defined a number of key residues that participate in citrate and CoASH binding (Remington et al., 1982). On the basis of sequence homology, equivalent residues in the *E. coli* enzyme can be predicted (Table II). The homologies commencing at pig heart citrate synthase residues 319 and 372 have already been identified by Bell et al. (1983), who drew attention to the conservation of active site residues. The conservation of sequence is well demonstrated in the four stretches of the sequence shown in Chart I, all of which contain catalytically essential residues in heart citrate synthase (marked with an asterisk).


5'--AAGCTTGCTCGAAACTTCAGGGGAAGAGAGGCTGGTACCCAGAAGCCACAGCAGGATGCCCACTGCAACAAAGGTGATCACACCGG
-581 -571 -561 -551 -541 -531 -521 -511 -501
AAACGCGATGGAGAATGGACGCTATCGCCGTGATGGGGAACCGGATGGTCTGTAGGTCCAGATTAACAGGTCTTTGTTTTTTCACATTTC
-491 -481 -471 -461 -451 -441 -431 -421 -411
TTATCATGAATAACGCCCACATGCTGTTCTTATTATTCCCTGGGGACTACGGGCACAGAGGTTAACTTTCTGTTACCTGGAGACGTCGGG
-401 -391 -381 -371 -361 -351 -341 -331 -321
ATTTCCTTCCTCCGGTCTGCTTGCGGGTCAGACAGCGTCCTTTCTATAACTGCGCGTCATGCAAAACACTGCTTCCAGATGCGAAAACGA
-311 -301 -291 -281 -271 -261 -251 -241 -231
CACGTTACAACGCTGGGTGGCTCGGGATTGCAGGGTGTTCCGGAGACCTGGCGGCAGTATAGGCTGTTCACAAAATCATTACAATTAACC
-221 -211 -201 -191 -181 -171 -161 -151 -141
TACATATAGTTTGTCGGGTTTTATCCTGAACAGTGATCCAGGTCACGATAACAACATTTATTTAATTTTTAATCATCTAATTTGACAATC
-131 -121 -111 -101 -91 -81 -71 -61 -51
T/A rich
ATTCAACAAAGTTGTTACAAACATTACCAGGAAAAGCATATAATGCGTAAAAGTTATGAAGTCGGTATTTCACCTAAGATTAACTTATGT
-41 -31 -21 -11 -1 10 20 30 40
-35 recognition
Pribnow box
Initiation of transcription
AACAGTGTGGAGTATTGACCAATTCATTCGGGACAGTTATTAGTGGTAGACAAGTTTAATAAATTCGGATTGCTAAGTACTTGATTCGCC
50 60 70 80 90 100 110 120 130
ATTTATTCGTCATCAATGGATCCTTTACCTGCAAGCGCCCAGAGCTCTGTACCCAGGTTTTCCCCTCTTTCACAGAGCGGCGAGCCAAAT
140 150 160 170 180 190 200 210 220
BamHI
M A
AAAAAACGGGTAAAGCCAGGTTGATGTGCGAAGGCAAATTTAAGTTCCGGCAGTCTTACGCAATAAGGCGCTAAGGAGACCTTAAATGGC
230 240 250 260 270 280 290 300 310
ribosome binding site
D T K A K L I L N G D I A V E L D V L K G I L G Q D V I D I
TGATACAAAAGCAAAACTCACCCTCAACGGGGATACAGCTGTTGAACTGGATGTGCTGAAAGGCACGCTGGGTCAAGATGTTATTGATAT
320 330 340 350 360 370 380 390 400
R I L G S K G V F I F D P G F I S I A S C E S K I I F I D G
CCGTACTCTCGGTTCAAAAGGTGTGTTCACCTTTGACCCAGGCTTCACTTCAACCGCATCCTGCGAATCTAAAATTACTTTTATTGATGG
410 420 430 440 450 460 470 480 490
D E G I L L H R G F P I D Q L A I D S N Y L E V C Y I L L N
TGATGAAGGTATTTTGCTGCACCGCGGTTTCCCGATCGATCAGCTGGCGACCGATTCTAACTACCTGGAAGTTTGTTACATCCTGCTGAA
500 510 520 530 540 550 560 570 580
G E K P T Q E Q Y D E F K T T V T R H I M I H E Q I T R L F
TGGTGAAAAACCGACTCAGGAACAGTATGACGAATTTAAAACTACGGTGACCCGTCATACCATGATCCACGAGCAGATTACCCGTCTGTT
590 600 610 620 630 640 650 660 670
H A F R R D S H P M A V M C G I T G A L A A F Y H D S L D V
CCATGCTTTCCGTCGCGACTCGCATCCAATGGCAGTCATGTGTGGTATTACCGGCGCGCTGGCGGCGTTCTATCACGACTCGCTGGATGT
680 690 700 710 720 730 740 750 760
N N P R H R E I A A F R L L S K M P T M A A M C Y K Y S I G
TAACAATCCTCGTCACCGTGAAATTGCCGCGTTCCGCCTGCTGTCGAAAATGCCGACCATGGCCGCGATGTGTTACAAGTATTCCATTGG
770 780 790 800 810 820 830 840 850
Q P F V Y P R N D L S Y A G N F L N M M F S T P C E P Y E V
TCAGCCATTTGTTTACCCGCGCAACGATCTCTCCTACGCCGGTAACTTCCTGAATATGATGTTCTCCACGCCGTGCGAACCGTATGAAGT
860 870 880 890 900 910 920 930 940
N P I L E R A M D R I L I L H A D H E Q N A S I S I V R I A
TAATCCGATTCTGGAACGTGCTATGGACCGTATTCTGATCCTGCACGCTGACCATGAACAGAACGCCTCTACCTCCACCGTGCGTACCGC
950 960 970 980 990 1000 1010 1020 1030
G S S G A N P F A C I A A G I A S L W G P A H G G A N E A A
TGGCTCTTCGGGTGCGAACCCGTTTGCCTGTATCGCAGCAGGTATTGCTTCACTGTGGGGACCTGCGCACGGCGGTGCTAACGAAGCGGC
1040 1050 1060 1070 1080 1090 1100 1110 1120

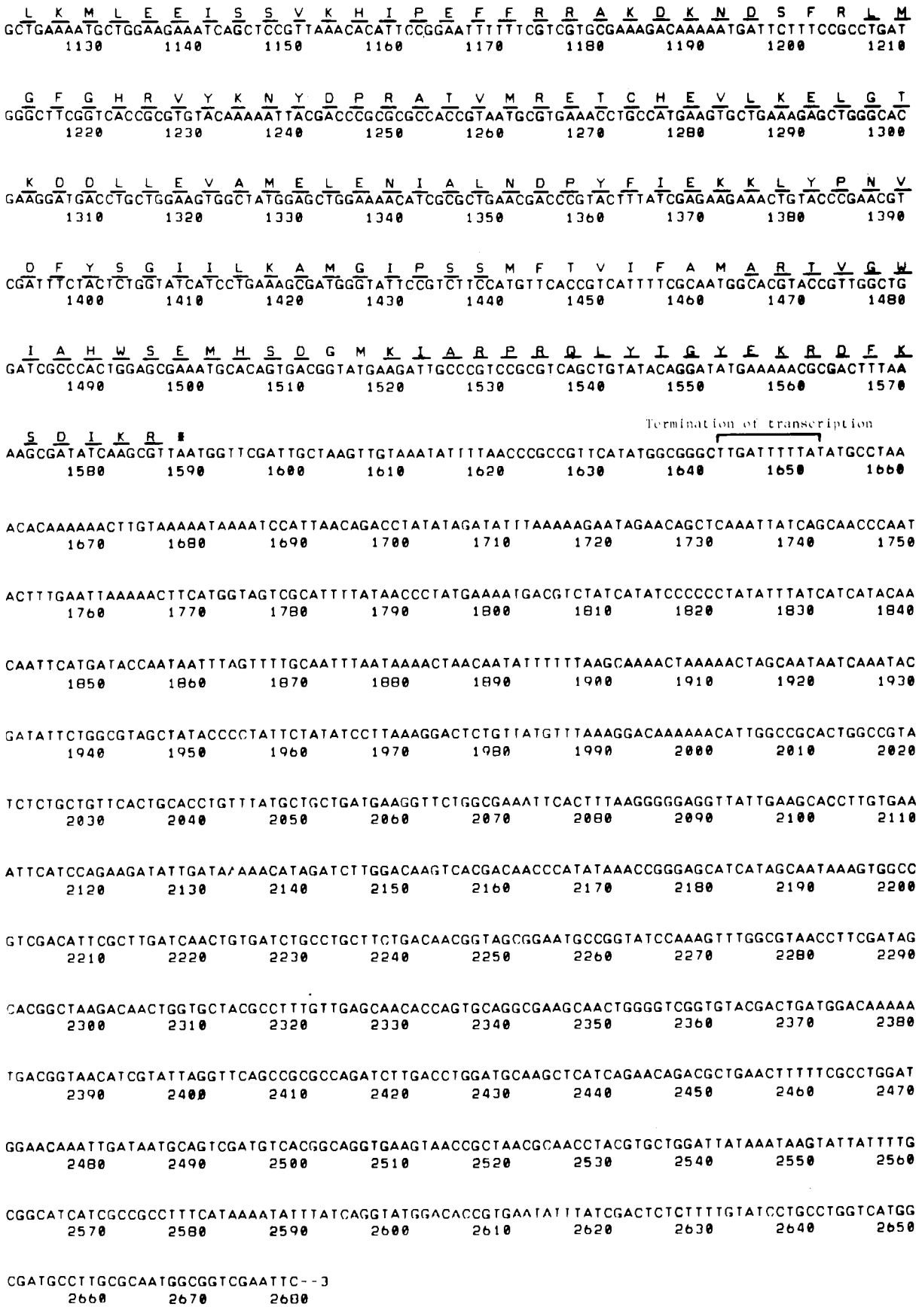


FIGURE 2: Sequence of the anti sense strand of the *glt* A gene and the predicted protein sequence of *E. coli* citrate synthase. The gene sequence is numbered from the predicted site for initiation of transcription (Bloxham et al., 1983a) and indicates potential regulatory sites based on strong sequence homology. The predicted protein sequence was obtained by using the standard genetic code. The protein is numbered from the first residue (Ala-1) after the initiator methionine. Residues that are underlined have been identified by chemical sequencing of the protein product. The single letter amino acid code is as follows: A, alanine; C, cysteine; D, aspartate; E, glutamate; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline, Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

Chart I

```
          232       *       *
heart      L  T  I  H  S  D  H  E
          222
E. coli    L  I  L  H  A  D  H  E

          271       *
heart      G  P  L  H  G  L  A  N  Q
          261
E. coli    G  P  A  H  G  G  A  N  E

          315             *           *           *
heart      V  P  G  Y  G  H  A  V  L  R  K  T  D  P  R  Y  T
          300
E. coli    L  M  G  F  G  H  R  V  Y  K  N  Y  D  P  R  A  T

          371          *
heart      W  P  N  V  D  A  H  S  G  V  L  L
          358
E. coli    Y  P  N  V  D  F  Y  S  G  I  I  L
```

The crystal structure of heart citrate synthase (Remington et al., 1982) predicts that Asp-375, His-274, and His-320 are the most likely residues to participate in bond-forming and -cleavage events during catalysis. Clearly, sequence conservation (i.e., Asp-362, His-264, and His-305 in *E. coli* citrate synthase) in these regions is very striking and emphasizes the importance of these residues.

The crystal structure of citrate synthase showed that the adenine of CoA was bound by a series of hydrogen bonds from the peptidyl backbone of the main chain (residues 314–320) that wrapped around the adenine ring in an edge-on manner. Such a structure, which does not rely on side-chain interactions, might constitute a primitive binding site (Remington et al., 1982). Interestingly, there is reasonable homology in this region between the two proteins (315–320 in pig heart; 300–305 in *E. coli*), indicating that the side chains may well be important in establishing this loop structure although participation of the side chains in bonding is not required.

Although the preceding discussion emphasizes areas of similarity between eukaryotic and prokaryotic citrate synthase, there are critical differences between the two proteins. Thus, the eukaryotic enzyme exists as a dimer (Singh et al., 1970; Srere, 1975) whereas the *E. coli* enzyme exists in multiple oligomeric forms. In the presence of 0.1 M KCl, the hexamer appears to be the most stable and catalytically active form (Tong & Duckworth, 1975). Furthermore, whereas eukaryotic enzymes are inhibited by ATP (competitive with acetyl-CoA) (Hathaway & Atkinson, 1965), the enzyme from *E. coli* is only weakly inhibited by this nucleotide (Weitzman, 1966; Jangaard et al., 1968). Rather, the *E. coli* enzyme (Weitzman & Danson, 1976) is inhibited by NADH (competitive with acetyl-CoA) and α-ketoglutarate (competitive with oxaloacetate).

Comparing sequence homology between *E. coli* and pig heart citrate synthase using DIAGON plots (Staden, 1982) reveals that the least statistically significant structural correlation can be identified in the amino-terminal region (residues 1–100). Homology within this region appears to be fortuitous. This situation can be contrasted with sequence studies on chicken and pig heart citrate synthase, which show excellent homology in the amino-terminal region (Beeckmans & Kanarek, 1983). It may be suggested that structural variations in the amino-terminal region may be responsible for differences between the prokaryotic and eukaryotic enzymes. As an example, Arg-46 is one of the groups capable of forming an ionic interaction with the 5′-diphosphate of CoA in heart citrate synthase (Remington et al., 1982), and there is no equivalent homologous residue in the linear sequence of the *E. coli* protein. This suggests that, if the loss of this ionic interaction is significant, then the *E. coli* enzyme should bind CoA derivatives less strongly than the heart enzyme. This is consistent with K_m measurements of acetyl-CoA which are 4–6 μM for heart (Johansson & Pettersson, 1974; Beeckmans & Kanarek, 1983) and 50–400 μM for *E. coli* (Jangaard et al., 1968; Faloona & Srere, 1969) citrate synthase, respectively. The variation in K_m of acetyl-CoA for the *E. coli* enzyme reflects the fact that this value is very sensitive to ionic strength.

Conclusion

The current work provides a complete description of the *glt* A gene from *E. coli*, which allows identification of a number of sequences that might control expression. Furthermore, the complete amino acid sequence of the protein can be predicted, and this corroborates experiments on the chemical determination of the protein structure. The strong homology in predicted active site aspartate, histidine, and arginine residues between *E. coli* and pig heart sequences emphasizes the fundamental nature of the catalytic mechanism predicted by the crystallographic study of heart citrate synthase. Comparison of the two sequences reveals critical differences in residues that bind the 5′-diphosphate of CoA, which could be the basis for the difference in sensitivity of the two enzymes to inhibition by ATP.

Acknowledgments

We thank Christopher J. Herbert for technical assistance.

Registry No. Citrate synthase, 9027-96-7; DNA (*E. coli* gene *glt* A), 87226-15-1; citrate synthase (*E. coli* reduced), 87226-16-2.

References

Ashworth, J. M., Kornberg, H. L., & Nothmann, D. L. (1965) *J. Mol. Biol. 11*, 654–657.

Beeckmans, S., & Kanarek, L. (1983) *Int. J. Biochem. 15*, 469–478.

Bell, A. W., Bhayana, V., & Duckworth, H. W. (1983) *Biochemistry 22*, 3400–3405.

Bloxham, D. P., Parmelee, D. C., Kumar, S., Walsh, K. A., & Titani, K. (1982) *Biochemistry 21*, 2028–2036.

Bloxham, D. P., Herbert, C. J., Giles, I. G., & Ner, S. S. (1983a) *Mol. Gen. Genet. 192*, 499–506.

Bloxham, D. P., Herbert, C. J., Ner, S. S., & Drabble, W. T. (1983b) *J. Gen. Microbiol. 129*, 1889–1897.

Duckworth, H. W., & Bell, A. W. (1982) *Can. J. Biochem. 60*, 1143–1147.

Faloona, G. R., & Srere, P. A. (1969) *Biochemistry 8*, 4497–4503.

Grantham, R., Gautier, C., Gouy, M., Jacobzone, M., & Mercien, R. (1981) *Nucleic Acids Res. 9*, 443–473.

Grosjean, H., & Fiers, W. (1982) *Gene 18*, 199–209.

Guest, J. R. (1981) *J. Gen. Microbiol. 124*, 17–23.

Hathaway, J. A., & Atkinson, D. E. (1965) *Biochem. Biophys. Res. Commun. 20*, 661–665.

Jangaard, N. O., Unkeless, J., & Atkinson, D. E. (1968) *Biochim. Biophys. Acta 151*, 225–235.

Johansson, C.-J., & Pettersson, G. (1974) *Eur. J. Biochem. 42*, 383–388.

Messing, J., & Vieira, J. (1982) *Gene 19*, 269–276.

Messing, J., Crea, R., & Seeburg, P. H. (1981) *Nucleic Acids Res. 9*, 309–321.

Remington, S., Wiegand, G., & Huber, R. (1982) *J. Mol. Biol. 158*, 111–152.

Rosenberg, M., & Court, D. (1979) *Annu. Rev. Genet. 13*, 319–353.

Sanger, F., Nicklen, S., & Coulsen, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.

Sanger, F., Coulson, A. R., Barrell, B. G., & Smith, A. J. H. (1980) *J. Mol. Biol. 143*, 161–178.
Singh, M., Brooks, G. C., & Srere, P. A. (1970) *J. Biol. Chem. 245*, 4636–4640.
Smith, A. J. H. (1980) *Methods Enzymol. 65*, 560–580.
Srere, P. A. (1975) *Adv. Enzymol. Relat. Areas Mol. Biol. 43*, 57–101.
Staden, R. (1982) *Nucleic Acids Res. 10*, 2951–2961.
Tong, E. K., & Duckworth, H. W. (1975) *Biochemistry 14*, 235–241.
Weitzman, P. D. J. (1966) *Biochim. Biophys. Acta 128*, 213–215.
Weitzman, P. D. J., & Danson, M. J. (1976) *Curr. Top. Cell Regul. 10*, 161–204.

Articles

Effects of Deletions near *Escherichia coli rrn*B Promoter P_2 on Inhibition of in Vitro Transcription by Guanosine Tetraphosphate†

Robert E. Kingston*

ABSTRACT: The regulatory nucleotide guanosine tetraphosphate (ppGpp) inhibits in vitro transcription from the *rrn*B P_2 promoter. In order to locate the sequences responsible for this inhibition, plasmids have been constructed that contain deletions that extend to various points near the *rrn*B P_2 initiation site. Guanosine tetraphosphate has little or no effect on transcription of certain of the altered promoters. The loss of ability of a P_2 promoter to be inhibited by ppGpp correlates with changes around the transcription initiation site, suggesting that these sequences are involved in the interactions that determine sensitivity of a promoter to ppGpp in vitro. RNA polymerase that initiates at *rrn*B promoter P_1 has been shown to pause in the region of P_2 in vitro. The effects of the deletions on these pause sites have been determined. The results strongly support the hypothesis that RNA polymerase at P_2 can cause RNA polymerase that has initiated at promoter P_1 to pause for a substantial length of time.

When growing *Escherichia coli* are deprived of an amino acid or otherwise blocked in the aminoacylation of tRNA, they undergo what has been termed a stringent response [Stent & Brenner, 1961; reviewed by Gallant & Lazzarini (1976)]. The level of expression of numerous operons changes substantially [review by Nierlich (1978) and Gallant (1979)]; in particular, the rate of ribosomal RNA synthesis decreases 10–20-fold. Genetic and physiological studies, as well as in vitro transcription studies, have implicated the nucleotide guanosine tetraphosphate (ppGpp)[1] in these changes (Nierlich, 1978; Gallant, 1979; Travers, 1976a,b; van Ooyen et al., 1976; Travers & Baralle, 1976; Debenham & Travers, 1977; Oostra et al., 1977; Travers et al., 1978; Hamming et al., 1979, 1980). ppGpp specifically inhibits transcription from each of the tandem promoters (P_1, P_2) found in rRNA operons in vitro, though it has no effect on transcription from several phage promoters (Hamming et al., 1979, 1980; Kingston et al., 1981a,b). These results suggest that there are specific areas in promoters involved in the stringent response that render them sensitive to ppGpp. After comparison of the known sequences of promoters involved in the stringent response, Travers proposed that a specific sequence near the transcription initiation site determines sensitivity to ppGpp (Travers, 1980a). A $tRNA^{Tyr}$ promoter altered in this region has been reported to show a changed response to ppGpp (Travers, 1980b). Here the sequences in the *rrn*B P_2 promoter that determine its sensitivity to ppGpp in vitro are defined.

The region around the initiation site of promoter P_2 is of interest not only because of its possible involvement in determining sensitivity to ppGpp but also because this region apparently determines three sites where RNA polymerase pauses during elongation from promoter P_1 in vitro (Kingston & Chamberlin, 1981). These pause sites could potentially play a role in regulating rRNA expression by limiting the maximal rate of transcription of *rrn* operons ("turnstile attenuation"; Kingston & Chamberlin, 1981; Kingston et al., 1981b). Plasmids have been constructed containing deletions extending into the initiation site of promoter P_2 (J. Brosius, unpublished results). The effects of these deletions both on the ability of ppGpp to inhibit transcription from P_2 and on the pausing of RNA polymerase in this region are reported here.

Experimental Procedures

Materials. Plasmids were constructed by and were the gracious gift of Dr. Jürgen Brosius (Columbia University). DNA templates containing P_2 deletions were prepared for transcription by digesting crude plasmid DNA with the appropriate restriction enzyme (pKK34-121, *Pst*I; pKK35-120 and derivatives, *Bam*HI), separating the linearized plasmid DNA from chromosomal DNA and RNA on an agarose gel, and isolating the plasmid DNA by the method of Vogelstein

† From the Department of Biochemistry, University of California, Berkeley, California 94720. *Received January 13, 1983; revised manuscript received June 21, 1983.* This work was supported by Grant GM 12010 from the National Institute of General Medical Sciences (to Dr. Michael Chamberlin) and by a predoctoral National Institutes of Health training grant.

* Address correspondence to this author at the Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139.

[1] Abbreviations: ppGpp, guanosine tetraphosphate; EDTA, ethylenediaminetetraacetic acid.

0006-2960/83/0422-5249$01.50/0