Primary Structure of the Thermostable Formyltetrahydrofolate Synthetase from *Clostridium thermoaceticum*[†,‡]

Charles R. Lovell,*[,§] Alan Przybyla,[‖] and Lars G. Ljungdahl[‖]

Department of Biological Sciences, University of South Carolina, Columbia, South Carolina 29208, and Department of Biochemistry and Center for Biological Resource Recovery, University of Georgia, Athens, Georgia 30602

Received December 28, 1989; Revised Manuscript Received March 29, 1990

ABSTRACT: The complete nucleotide sequence of the *Clostridium thermoaceticum* formyltetrahydrofolate synthetase (FTHFS) was determined and the primary structure of the protein predicted. The gene was 1680 nucleotides long, encoding a protein of 559 amino acid residues with a calculated subunit molecular weight of 59 983. The initiation codon was UUG, with a probable ribosome binding site 11 bases upstream. A putative ATP binding domain was identified. Two Cys residues likely to be involved in subunit aggregation were tentatively identified. No characterization of the tetrahydrofolate (THF) binding domain was possible on the basis of the sequence. A high level of amino acid sequence conservation between the *C. thermoaceticum* FTHFS and the published sequences of *C. acidiurici* FTHFS and the FTHFS domains of the *Saccharomyces cerevisiae* C_1-THF synthases was found. Of the 556 residues shared between the two clostridial sequences, 66.4% are identical. If conservative substitutions are allowed, this percentage rises to 75%. Over 47% of the residues shared between the *C. thermoaceticum* FTHFS and the yeast C_1-THF synthases are identical, 57.4% if conservative substitutions are allowed. Hydrophobicity profiles of the *C. acidiurici* and *C. thermoaceticum* enzymes were very similar and did not support the idea that large hydrophobic domains play an important role in thermostabilizing the *C. thermoaceticum* FTHFS.

The activation of formate by 10-formyltetrahydrofolate synthetase (EC 6.3.4.3) is a key reaction in acetogenic and purinolytic bacteria. Acetogens synthesize acetyl-CoA via the Wood pathway of autotrophic CO_2 fixation, which uses this enzyme to activate formate in the presence of tetrahydrofolate and ATP (Wood et al., 1986; Fuchs, 1986; Ljungdahl, 1986). Purine fermenters synthesize acetate via the glycine synthase/glycine reductase pathway, which also utilizes FTHFS[1] in a CO_2 fixation process (Waber & Wood, 1979; Schiefer-Ullrich et al., 1984; Ljundahl, 1984). FTHFS is present in both acetogenic and purinolytic bacteria at high levels.

Many of the physical, chemical, and catalytic properties of FTHFSs purified from acetogenic and purinolytic clostridia are very similar. FTHFS has been purified to homogeneity from two acetogens, the thermophile *Clostridium thermoaceticum* (Ljungdahl et al., 1970) and the mesophile *C. formicoaceticum* (O'Brien et al., 1976), and from two mesophilic purine fermenters, *C. acidiurici* and *C. cylindrosporum* (Rabinowitz & Pricer, 1962). Each of these enzymes consists of four subunits of M_r = 60 000 that are identical within the enzyme from each source (MacKenzie & Rabinowitz, 1971). The substrate turnover rates, activation energies, and half-saturation constants of these enzymes are very similar, as are their amino acid compositions (O'Brien et al., 1976). The subunits of synthetases from *C. thermoaceticum* and *C. cylindrosporum* have been observed to form dimers, which in turn yield the tetrameric enzyme (Mayer et al., 1982). The binding regions holding the monomers together are known to differ from those linking the dimers. Chemical modification studies on the FTHFS active site have been performed (Elliot & Ljungdahl, 1982) but were limited by the reactivity of tetrahydrofolate. Further elucidation of active site structure and substrate binding domains and identification of key catalytic amino acid residues have been hindered by lack of knowledge of the amino acid sequences of acetogen synthetases.

The *C. thermoaceticum* enzyme differs from the mesophile enzymes in having a higher optimum temperature for catalysis and a higher thermostability (O'Brien et al., 1976). While this is not unusual in enzymes purified from thermophiles (Lungdahl, 1979), no mechanism explaining the higher optimum temperature and thermostability of the *C. thermoaceticum* FTHFS has been developed. It has been previously speculated that a high content of hydrophobic amino acids would account for high thermostability (Lungdahl et al., 1970). In order to test this idea for the case of the *C. thermoaceticum* FTHFS, to begin characterization of amino acids important in subunit aggregation, and to continue the characterization of the FTHFS active site structure and substrate binding domains, we cloned the structural gene encoding this enzyme into *Escherichia coli* (Lovell et al., 1988). The clone expressed very high levels of catalytically active, immunologically cross-reactive, thermostable FTHFS. We now report the complete nucleotide sequence of this gene and the predicted primary structure of the protein. We also compare this sequence to the published sequences for the purine-fermenting, mesophilic *C. acidiurici* FTHFS (Whitehead & Rabinowitz, 1988) and for FTHFS domains of the trifunctional *Saccharomyces cerevisiae* cytoplasmic and mitochondrial C_1-THF synthases (Staben & Rabinowitz, 1986; Shannon & Rabi-

[†] This research was supported by NIH Biomedical Research Support Grant Award S07 RR07160, the Carolina Venture Fund, and the National Institute of Diabetes and Digestive and Kidney Diseases, Grant DK 27323. A preliminary account of this work was part of a presentation given at the annual meeting of the American Society for Microbiology, New Orleans, LA, 1989 (Lovell & Hui, 1989).

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02911.

[§] University of South Carolina.

[‖] University of Georgia.

[1] Abbreviations: FTHFS, 10-formyltetrahydrofolate synthetase; kb, kilobase; THF, tetrahydrofolate; LB, Luria-Bertani.

0006-2960/90/0429-5687$02.50/0

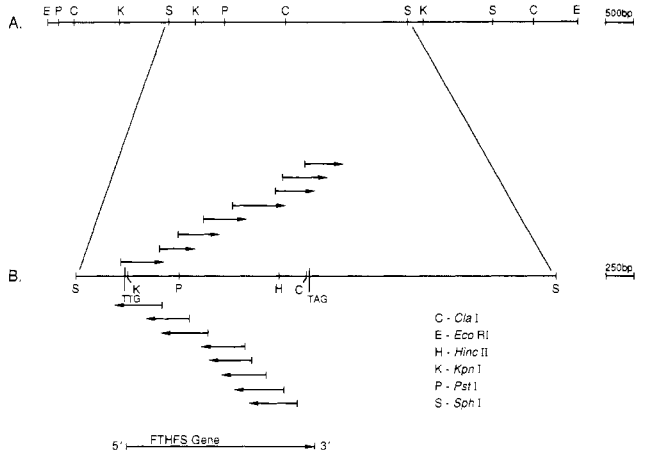

FIGURE 1: Derivation from (A) the 9.5-kb *Clostridium thermoaceticum* DNA fragment of (B) the 4.3-kb fragment in the subclone pCRL 471. Restriction map and sequencing strategy for the *C. thermoaceticum* FTHFS gene.

nowitz, 1988) and report a putative ATP binding sequence and a conserved cysteine residue that may be involved in subunit aggregation in *C. thermoaceticum* FTHFS.

Experimental Procedures

Materials. Modified bacteriophage T7 DNA polymerase (Sequenase), sequencing grade deoxy- and dideoxynucleotide triphosphate mixes and 7-deaza-2′-deoxyguanosine 5′-triphosphate were from United States Biochemical Corp. (Cleveland, OH). Other biochemicals were from Boehringer Mannheim Biochemicals or Sigma (St. Louis, MO). Radionuclides were from New England Nuclear (Wilmington, DE).

Bacterial Strains and Media. Escherichia coli strain JM105, used as the cloning host and maintenance strain, was maintained on M9 minimal medium (42 mM Na_2HPO_4, 22 mM KH_2PO_4, 19 mM NH_4Cl, 8.6 mM NaCl, 11.1 mM D-glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 1 mM thiamine hydrochloride) (Amersham M13 Cloning and Sequencing Handbook, Amersham Corp., Arlington Heights, IL). Cultures used in transformation experiments were grown at 37 °C in LB broth (Maniantis et al., 1982). LB agar and broth supplemented with 50 μg of ampicillin/mL were used for selecting and cultivating transformants. The construction of *Escherichia coli* strain CRL 47, which contains a hybrid plasmid derived from pBR 322 and a 9.5 kb *Eco*RI fragment of *Clostridium thermoaceticum* genomic DNA containing the FTHFS structural gene, has been described previously (Lovell et al., 1988). A 4.3-kb fragment containing the *C. thermoaceticum* FTHFS structural gene (see Figure 1) was subcloned in both orientations into pIBI 31 (International Biotechnologies, Inc., New Haven, CT), producing strains CRL 471 and CRL 472. Both subclones expressed high levels of catalytically active FTHFS. These subclones were the starting strains for producing deletion subclones for nucleotide sequence determination.

Construction and Selection of Deletion Subclones. The exonuclease III method of Henikoff (1984) was used to produce deletion subclones from CRL 471 and CRL 472. Transformants selected on LB with ampicillin were chosen and incubated at 37 °C with vigorous shaking (300 rpm) overnight in 1.5 mL of LB ampicillin (LB-Amp) for plasmid minipreparation (Holmes & Quigley, 1981). Plasmids from deletion subclones were selected on the basis of apparent molecular weight after *Eco*RI digestion and fractionation on agarose gels. A set of overlapping subclones giving virtually complete coverage of both orientations of the FTHFS structural gene was assembled (see Figure 1).

Purification of Plasmids for Sequencing. Overnight cultures of selected clones were grown in LB-Amp at 37 °C with 250 rpm shaking. Flasks containing 30 mL of LB-Amp were inoculated with 100 μL of the overnight cultures and incubated at 37 °C with shaking at 250 rpm for 5–6 h (early stationary phase). Cells were harvested and plasmids purified by using a modification of the alkaline extraction method of Birnboim and Doly (1979), followed by precipitation of the plasmid DNA with 6.5% (final concentration) poly(ethylene glycol).

Double-Stranded Plasmid Sequencing. The dideoxy chain termination sequencing method of Sanger et al. (1977), as modified for use with double-stranded vectors (Chen & Seeburg, 1985), 7-deaza-2′-deoxyguanisine 5′-triphosphate (Mizusawa et al., 1986), and deoxyadenosine 5′-([α-^{35}S]-thiotriphosphate) label (Biggin et al., 1983) was used. The specific activity of the [^{35}S]dATP was 18.5 TBq/mmol. Modified bacteriophage T7 DNA polymerase (Sequenase) was used to synthesize labeled termination products (Tabor & Richardson, 1987). These labeled chains were size fractionated on buffer-gradient polyacrylamide gels (Biggin et al., 1983).

Sequence Analysis. Routine sequence analysis was performed by using the Beckman (Palo Alto, CA) MicroGenie sequence analysis software (Queen & Korn, 1984). Hydrophobicity plots and homology searches of sequence databases were produced by using features of the University of Wisconsin Genetics Computer Group software package (Devereux et al., 1984).

Results

The deletion subclones used in determining the sequence of the *C. thermoaceticum* FTHFS structural gene provided coverage of all but 152 bases of the gene in both orientations (see Figure 1). Where the gene was sequenced in only one orientation (the final 152 bases of the 3′-end of the sense strand), nine templates and two different primers were used. Each arrow in Figure 1 represents from two to six separate subclones; a total of 46 subclones were sequenced.

The *C. thermoaceticum* derived DNA sequence contains 1680 nucleotides, encoding a protein of 559 amino acids with a calculated molecular weight of 59 983 (Figure 2). This was in good agreement with subunit molecular weight estimates of 60 000 derived from SDS–polyacrylamide gel electrophoresis and sedimentation velocity (Brewer et al., 1970; O'Brien et al., 1976). The initiator codon was UUG, with a probable ribosome binding site (GAGG) 11 bases upstream (see Figure 2). The terminator codon was UAG, followed by a probable termination diad symmetry region. No promoter based on the *E. coli* consensus sequences was identified upstream of the gene. Although only 55 bases upstream from the initiation codon were sequenced in both orientations, another 200 bases upstream of the gene were sequenced in one orientation only from M13mp18 subclones. The base composition of the *C. thermoaceticum* FTHFS gene is 58.3 mol % G + C, which is in good agreement with the value of 54 mol % G + C for the *C. thermoaceticum* genome (Ljungdahl, 1986), and is reflected in third position nucleotides of preferred codons. The G + C content of the *C. thermoaceticum* FTHFS gene differs significantly, however, from the 36 mol % G + C found for the *C. acidiurici* FTHFS gene (Whitehead & Rabinowitz, 1988) and 28 mol % G + C for the genome of that organism. Codon usage in the *C. thermoaceticum* FTHFS gene is given in Table I.

The amino acid compositions of the *C. thermoaceticum* and *C. acidiurici* FTHFS subunits determined through protein

Table I: Codon Usage in the mRNA Encoding the Formyltetrahydrofolate Synthetase of *C. thermoaceticum*

amino acid	codon	usage	amino acid	codon	usage
Phe	UUU	7	Tyr	UAU	10
	UUC	9		UAC	7
Leu	UUA	0	Ter	UAA	0
	UUG	3		UAG	1
	CUU	6		UGA	0
	CUC	13			
	CUA	2	His	CAU	1
	CUG	33		CAC	8
Ile	AUU	10	Gln	CAA	3
	AUC	25		CAG	7
	AUA	2			
Met	AUG	14	Asn	AAU	7
				AAC	16
Val	GUU	5			
	GUC	22	Lys	AAA	21
	GUA	7		AAG	15
	GUG	9			
			Asp	GAU	14
Ser	UCU	0		GAC	19
	UCC	9			
	UCA	1	Glu	GAA	15
	UCG	3		GAG	17
	AGU	3			
	AGC	5	Cys	UGU	0
				UGC	7
Pro	CCU	2			
	CCC	13	Trp	UGG	2
	CCA	1			
	CCG	9	Arg	CGU	5
				CGC	8
Thr	ACU	4		CGA	0
	ACC	24		CGG	9
	ACA	2		AGA	0
	ACG	4		AGG	4
Ala	GCU	7	Gly	GGU	17
	GCC	50		GGC	26
	GCA	3		GGA	3
	GCG	3		GGG	8

Table II: Amino Acid Compositions of the Formyltetrahydrofolate Synthetase Subunits of *Clostridium thermoaceticum* and *C. acidiurici*

amino acid	no. of residues per subunit in FTHFS from[a]	
	C. thermoaceticum	*C. acidiurici*[b]
Ala	63 (60)	64 (63)
Arg	26 (24)	17 (16)
Asn	23 (Asx = 52)	25 (Asx = 69)
Asp	33	44
Cys	7 (6)	6 (6)
Gln	10 (Glx = 45)	12 (Glx = 45)
Glu	32	33
Gly	54 (49)	51 (51)
His	9 (8)	12 (11)
Ile	37 (37)	38 (31)
Leu	57 (52)	56 (52)
Lys	47 (33)	36 (46)
Met	14 (13)	14 (11)
Phe	16 (16)	19 (19)
Pro	25 (24)	20 (19)
Ser	21 (17)	16 (16)
Thr	34 (30)	28 (28)
Trp	2 (4)	1 (2)
Tyr	17 (14)	9 (9)
Val	43 (42)	44 (38)
end	1	1
total	560	557
acidic (Asp + Glu)	65	77
basic (Arg + Lys)	73	53
Ar (Phe + Trp + Tyr)	35	29
hydrophobic (Ar + Ile + Leu + Met + Val)	186	181
mol wt (calcd)	59983	59597

[a] Values in parentheses were determined by amino acid analysis of hydrolyzed protein samples (O'Brien et al., 1976). The other values are predicted from the DNA sequences. [b] *C. acidiurici* values were determined from the published data of Whitehead and Rabinowitz (1988).

hydrolysis followed by amino acid analysis (O'Brien et al., 1976) and predicted from the amino acid sequences (this study and Whitehead & Rabinowitz, 1988) are given in Table II. The results of the amino acid analysis agree well with the data derived from the nucleotide sequences. The amino acid compositions reflect the high degree of protein homology. The clostridial synthetases also showed a high degree of homology to the FTHFS domains of cytoplasmic and mitochondrial C_1-THF synthases from *Saccharomyces cerevisiae* (Staben & Rabinowitz, 1986; Shannon & Rabinowitz, 1988; see Figure 3).

Examination of the protein sequences given in Figure 3 revealed the presence of putative ATP binding sequences from residues 58–85 and 113–124 (numbers based on the *C. thermoaceticum* sequence). The homologous FTHFS and C_1-THF synthase sequences in this region are compared to the sequences of several known ATP binding proteins (see Figure 4 caption for notation definition) in Figure 4. No obvious folate binding domain was observed.

Kyte–Doolittle hydropathy plots (Kyte & Doolittle, 1982) of *C. thermoaceticum* (Figure 5A) and *C. acidiurici* (Figure 5B) FTHFS subunit sequences were examined for changes in hydrophobic and hydrophilic domains between these two proteins. Both enzymes showed several pronounced hydrophobic and hydrophilic domains and were very similar overall. Several of the hydrophilic domains found in both sequences were broader in the *C. thermoaceticum* sequence. The most prominent hydrophobic domain in each sequence, encompassing about 25 amino acids in the *C. thermoaceticum* enzyme and about 18 amino acids in the *C. acidiurici* enzyme, coincided with the location of a conserved Cys residue (Cys 210 in *C. thermoaceticum* and Cys 207 in *C. acidiurici*).

Discussion

In our previous study we cloned the *C. thermoaceticum* structural gene encoding the highly thermostable FTHFS into *E. coli* (Lovell et al., 1988). The clone expressed a high level (32% of the soluble protein in the *E. coli* cells) of catalytically active, immunologically cross-reactive, thermostable FTHFS. A much lower, but still significant, level of FTHFS expression from the cloned *C. acidiurici* gene was also observed by Whitehead and Rabinowitz (1986). Although Whitehead and Rabinowitz (1988) reported recognizable *E. coli* style promoter sequences upstream of the FTHFS gene in *C. acidiurici*, no such sequences were found in our examination of the sequence up to 200 bases upstream of the *C. thermoaceticum* FTHFS gene. The initiation codon used in the predicted *C. thermoaceticum* mRNA is UUG. This identification is based on results of amino terminal sequencing of the *C. thermoaceticum* FTHFS and comparison with the *C. acidiurici* FTHFS sequence, which support no alternative start codon. UUG is used as the initiation codon in 17% of messages in another Gram-positive spore former, *Bacillus* (Hager & Rabinowitz, 1985), but in only 1% of *E. coli* messages (Stormo, 1986). Use of UUG as initiation codon does not affect translation rates in *E. coli*, and in fact, many highly expressed proteins use this initiation codon (Stormo, 1986). There is a recognizable ribosome binding sequence (Shine & Dalgarno, 1974) 11 bases 5′ of the initiation codon.

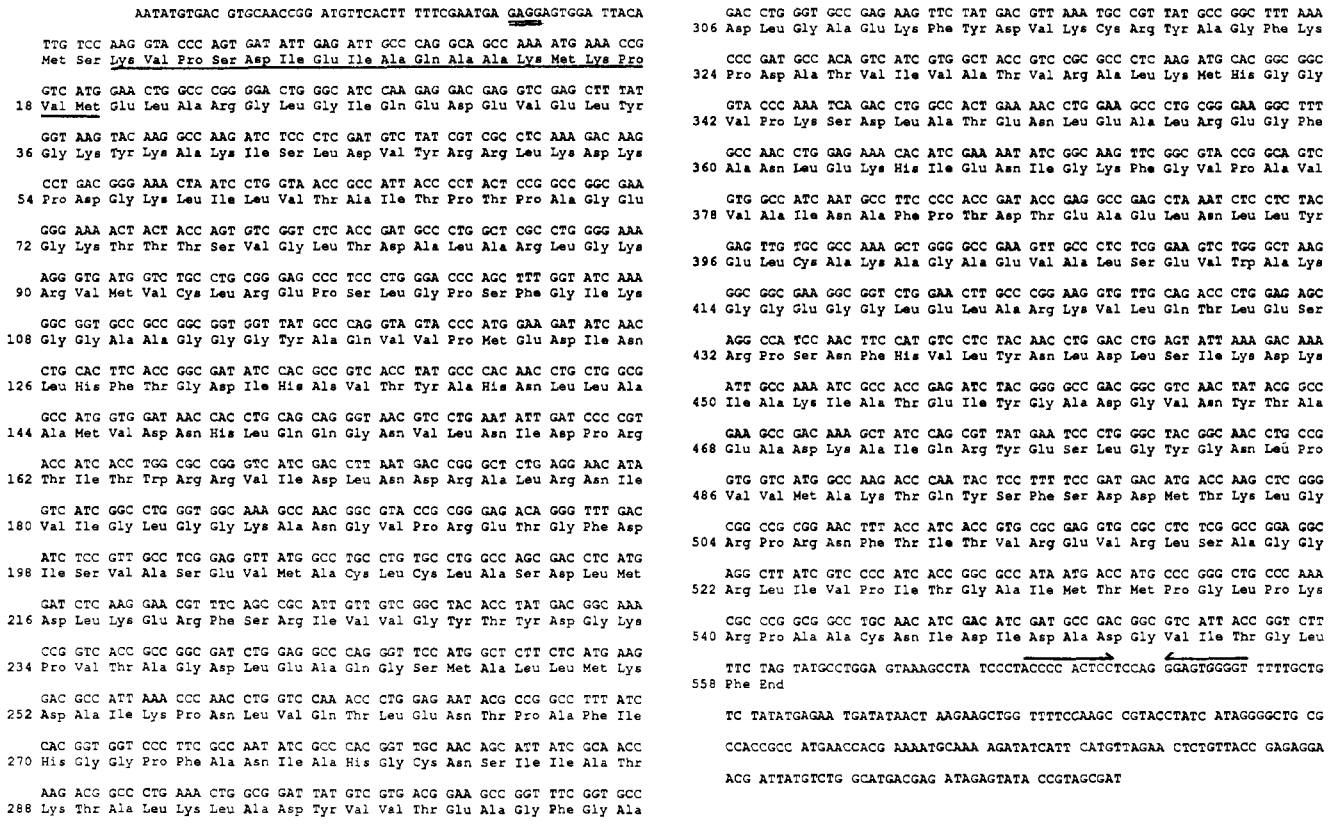

FIGURE 2: Nucleotide and predicted amino acid sequences of the *Clostridium thermoaceticum* formyltetrahydrofolate synthetase structural gene. Residues identified by amino terminal sequencing are underlined. The proposed ribosome binding site is indicated by double underlining and the termination diad symmetry region is indicated by facing arrows.

Perhaps the most striking results of this study were the extremely high levels of amino acid sequence homology observed between the clostridial synthetases, these proteins, and the FTHFS domains of the yeast C_1-THF synthases. Considering the difference in G + C contents of the clostridial genomes (*C. thermoaceticum* 54, *C. acidiurici* 28) and the evolutionary distance between either *Clostridium* and the yeast *Saccharomyces cerevisiae*, such high levels of sequence conservation are noteworthy.

The predicted amino acid sequences of the clostridial enzymes shared 556 residues. A total of 369 of these residues are conserved, 66.4% of the total (Figure 3). When conservative substitutions (Ile-Leu-Val, Ser-Thr, Phe-Tyr, Arg-Lys, and Asp-Glu) are accepted, 417 positions, or 75% of the total amino acid sequence, are conserved. The longest unbroken stretch of amino acid sequence homology between these two proteins is 25 residues, extending from Val 260 to Ser 284.

The *C. thermoaceticum* FTHFS amino acid sequence also has 556 residues in common with the *S. cerevisiae* cytoplasmic and mitochondrial C_1-THF synthase domains mentioned above. Of these, 264 residues (47.5% of the total) are conserved intact among the three sequences. When the conservative substitutions listed above are accepted, 319 residues, or 57.4% of the total amino acid sequence, are conserved. The bacterial enzymes display quite high levels of homology to each other and, perhaps unexpectedly, to the FTHFS domains of both *S. cerevisiae* C_1-THF synthases. Such a high degree of sequence conservation must be essential for the continued efficient functioning of these enzymes and examination of the conserved residues in these sequences provides some information about residues important to substrate binding domains and subunit aggregation (see below).

Interestingly, in regard to the high level of amino acid sequence conservation observed between the clostridial enzymes and the FTHFS domains of the yeast C_1-THF synthases, little to no sequence homology with other folate binding proteins was found. Many enzymes, such as thymidylate synthase and dihydrofolate reductase, bind folylpolyglutamate instead of tetrahydrofolate, perhaps explaining the observed lack of similarity (Maley et al., 1982; Smith et al., 1982). Unfortunately, no homology between the FTHFS sequences and another tetrahydrofolate binding enzyme, *E. coli* folylpolyglutamate synthetase–dihydrofolate synthetase (Bognar et al., 1987), was detected either. The only folate-binding enzymes we examined that showed any homology to the FTHFS sequences were the methylene-THF reductases (*metF* products) of *E. coli* and *Salmonella typhimurium* (Saint-Girons et al., 1983; Stauffer & Stauffer, 1988). Residues 145–150 of both methylene-THF reductases corresponded exactly with residues 197–202 of *C. thermoaceticum* FTHFS, a region conserved among all of the FTHFS sequences at both the amino acid and nucleotide sequence level. It is not possible at this time to draw any hard conclusions about folate binding at the FTHFS active site on the basis of our sequence information, but the hexapeptide from 197 to 202 is certainly intriguing.

Sequences exhibiting significant levels of homology to known ATP binding sequences are approximately residues 58–85 and residues 113–124 (Walker et al., 1982; Robson, 1984; Duncan et al., 1986; Fry et al., 1986). The FTHFS sequences correspond closely to those defined by Fry et al. (1986), although the first sequence, like that of the *E. coli recA* protein, is not as glycine rich as usual for this type of sequence. As is the case for nitrogenase Fe protein and some protein kinases, there is considerable divergence from the spacing of ATP binding domains in ATPases (Fry et al., 1986), but the overall similarity is clear. It is also interesting that 13 of the 28 residues of the first sequence and 9 of the 12 residues of the second are conserved among the clostridial and yeast FTHFS se-

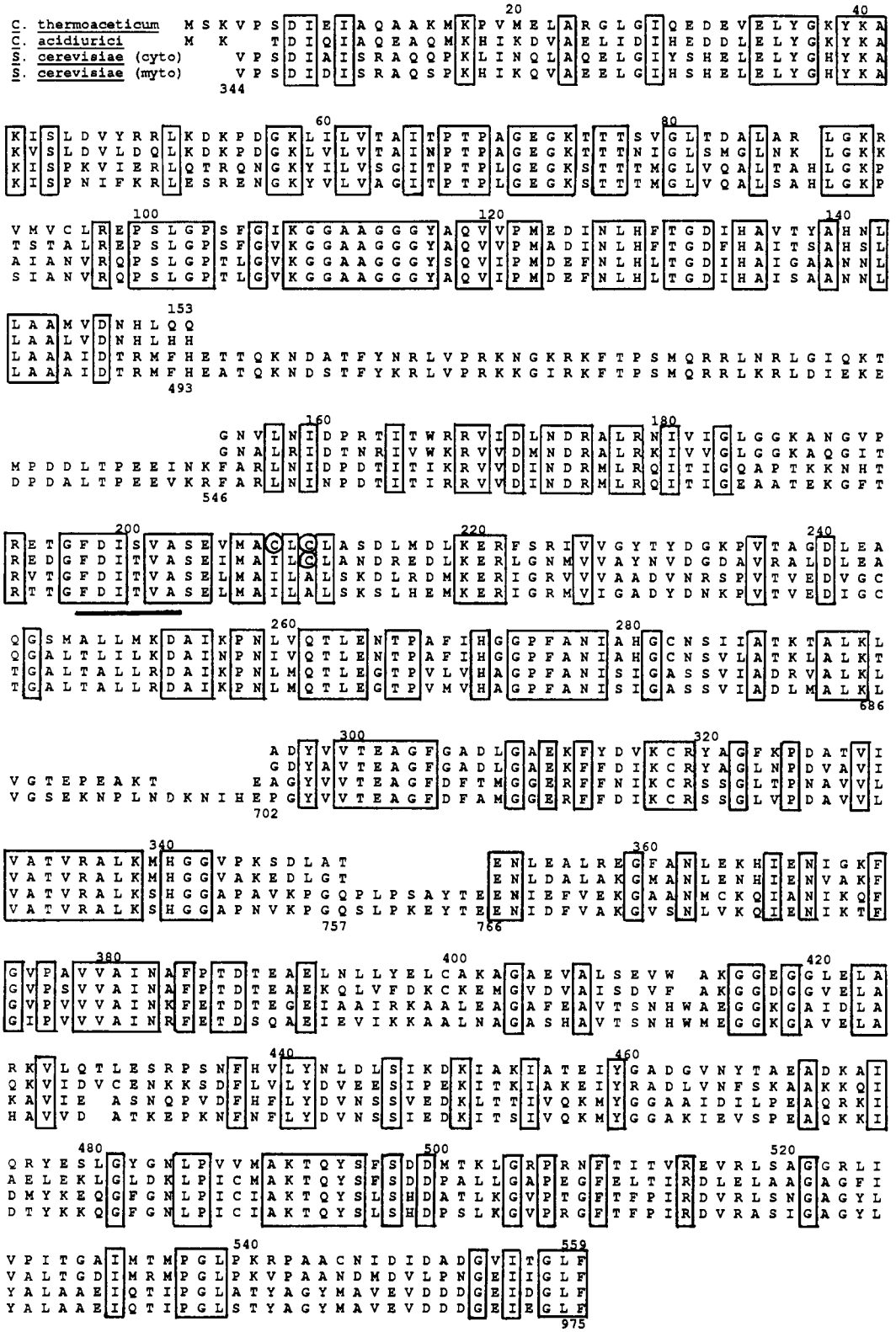

FIGURE 3: Amino acid homology between the formyltetrahydrofolate synthetase proteins encoded by *Clostridium thermoaceticum* and *Clostridium acidiurici* and the FTHFS domains of the cytoplasmic and mitochondrial C_1-THF synthases from *Saccharomyces cerevisiae*. Cys 208 and 210 of the *C. thermoaceticum* and Cys 207 of the *C. acidiurici* sequences are circled. Residues homologous to residues in the *Escherichia coli* and *Salmonella typhimurium metF* products are underlined. The *C. thermoaceticum* and *S. cerevisiae* mitochondrial sequences are numbered for reference.

quences. These highly conserved residues include Lys 58 and 74, Ile 65, Gly 71 and 73, and Thr 76 and 77 of the first domain and Gly 113, 114, and 115 and Asp 124 of the second domain. If these sequences are indeed ATP binding sequences, the level of conservation of these specific rsidues may indicate their importance in the function of the ATP binding domain of FTHFS. Unfortunately, no conclusions can be drawn about interactions between the ATP binding domain of this sequence and the folate or formate binding sequences or residues at this time.

Chemical modification studies by Elliott and Ljungdahl (1982) showed just one Cys residue in the tetrameric native *C. thermoaceticum* FTHFS to be chemically reactive (exposed to the external solution). It was further found that two Cys

```
      59                                                   85      113                       124
Ct    K L I L V T A I T P T P A G E G K T T T S V G L T D A L       G G G Y   A Q V V    P M E D
Ca    K L V L V T A I N P T P A G E G K T T T N I G L S M G L       G G G Y   A Q V V    P M A D
Scc   K Y I L V S G I T P T P L G E G K S T T T M G L V Q A L       G G G Y   S Q V I    P M   D E
Scm   K Y V L V A G I T P T P L G E G K S T T T M G L V Q A L       G G G Y   A Q V I    P M   D E

An    R Q I       A F   Y G K G G I G K S T T S Q N T L A A M
Av    R Q C       A I   Y G K G G I G K S T T T Q N L V A A L
Cp    R Q V       A I   Y G K G G I G K S T T T Q N L T S G L
Kp    R Q C       A I   Y G K G G I G K S T T T Q N L V A A L
Rm    R Q I       A F   Y G K G G I G K S T T S Q N T L A A L

Eca   Q R E L I I G D       R G G T G K T A L A I D     A I I           G E D A L I I    Y     D D
Ecb   K   V     G L F     G G A G V G K T   V N M M     E L I           G R D   V L L    F   V D
Era   R   I V E I Y G   P     E S S G K T T L T L Q V I A A A
Nm      Q S M L I     T   G E S G A G K T       E N T K K V I
Pak   K       I I   F V V G G P G S G K G T     Q C E K   I V           G   Q P T L L L    Y V D A G
```

FIGURE 4: FTHFS amino acid sequence comparisons with some known ATP binding proteins. Some important conserved residues are in bold. Abbreviations: Ct, *C. thermoaceticum* FTHFS, residues 58–85, 113–124 (this paper); Ca, *C. acidiurici* FTHFS, residues 55–82, 110–121 (Whitehead & Rabinowitz, 1988); Scc, *Saccharomyces cerevisiae* cytoplasmic C_1-THFS, residues 374–401, 430–441 (Staben & Rabinowitz, 1986); Scm, *S. cerevisiae* mitochondrial C1-THFS, residues 398–425, 454–465 (Shannon & Rabinowitz, 1988); An, *Anabaena* nitrogenase Fe protein, residues 7–30 (Mevarech et al., 1980); *Av*, *Azotobacter vinlandii* nitrogenase Fe protein, residues 3–26 (Hausinger & Howard, 1982); Cp, *Clostridium pasteurianum* nitrogenase Fe protein, residues 2–25 (Tanaka et al., 1977); Kp, *Klebsiella pneumoniae* nitrogenase Fe protein, residues 4–28 (Sundaresan & Ausubel, 1981); Rm, *Rhizobium meliloti* nitrogenase Fe protein, residues 5–29 (Torok & Kondorosi, 1981); Eca, *Escherichia coli* ATPase, α-subunit, residues 163–184, 253–262 (Gay & Walker, 1981); Ecb, *E. coli* ATPase β-subunit, residues 145–164, 235–243 (Saraste et al., 1981); Era, *Escherichia coli* Rec A protein, residues 61–84 (Horii et al., 1980); Nm, nematode myosin, residues 164–184 (see Walker et al., 1982); Pak, porcine adenylate kinase, residues 9–29, 110–121 [see Walker et al. (1982)]. No second domain occurs in nitrogenase sequences.

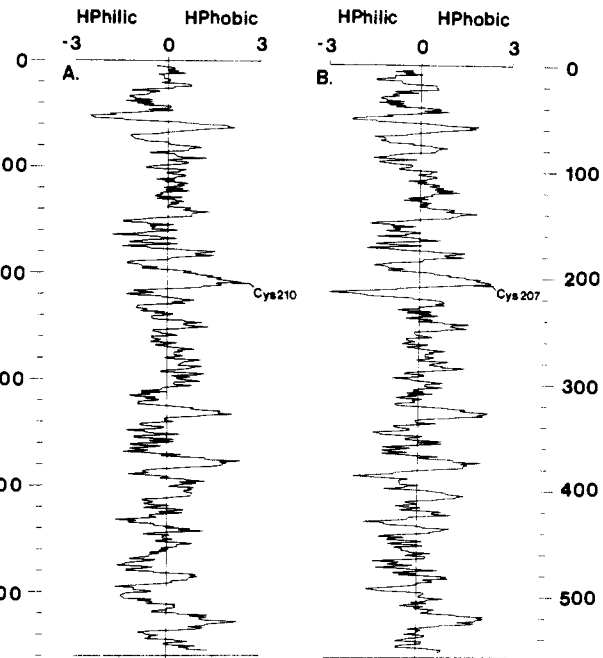

FIGURE 5: Kyte–Doolittle hydrophilicity plots of the predicted primary protein sequences for (A) *Clostridium thermoaceticum* FTHFS and (B) *Clostridium acidiurici* FTHFS. Cys 210 in the *Clostridium thermoaceticum* sequence and Cys 207 in the *Clostridium acidiurici* sequence are indicated.

residues appeared to be located at one of the regions of subunit association. It can be assumed that these Cys residues would be found in a hydrophobic domain of the protein, since they are not exposed to the external aqueous solution. We propose Cys 208 and Cys 210 to be the residues associated with subunit aggregation in the *C. thermoaceticum* enzyme, since no corresponding residues are found in the yeast enzymes (see Figure 3) and they are located in the largest hydrophobic domain of the protein (see Figure 5). Cys 210 corresponds to Cys 207 in the *C. acidiurici* sequence, while Cys 205 has no counterpart in the *C. acidiurici* sequence. Other than Cys 210, only Cys 282 and Cys 399 are conserved in both clostridial sequences but in neither yeast sequence. Neither Cys 282 nor Cys 399 are located in regions of the sequence having substantial hydrophobicity (see Figure 5). Absence of a homologue to Cys 208 or Cys 210 in either yeast sequence may be due to substantial differences in protein structure between the yeast trifunctional enzyme and the bacterial monofunctional proteins.

Elliott and Ljungdahl (1982) also found two chemically reactive Tyr residues in the *C. thermoaceticum* FTHFS. These residues could be protected from modification by folic acid, a competitive inhibitor of tetrahydrofolate binding, which implies they are located in the active site of the enzyme. Unfortunately, these residues cannot be deduced from our sequence information. There are seven Tyr residues conserved among the clostridial and yeast sequences and no compelling evidence implicating any of them in the active site of the enzyme.

The *C. thermoaceticum* FTHFS is enriched in Arg, Lys, and Tyr relative to the *C. acidiurici* FTHFS (Table II), with 20 more basic residues overall. The p*K* values for the R groups of Arg, Lys, and Tyr are 12.5, 10.8, and 10.1, respectively, so these amino acids have the potential to contribute strongly basic groups to ionic interactions and salt bridges within the thermophile protein. Increased content of basic amino acid residues in thermostable proteins has also been observed in ferredoxins (Bruschi & Guerlesquin, 1988) and stabilization of ferredoxins by salt bridges has been well documented (Perutz & Raidt, 1975; Hase et al., 1983; Bruschi & Guerlesquin, 1988). In several cases the nonconserved Arg, Lys, and Tyr residues in the *C. thermoaceticum* FTHFS are associated with hydrophilic regions of the sequence that are broader than their counterparts in the *C. acidiurici* FTHFS. For example, nonconserved Lys 502 and Arg 505 are found in a hydrophilic region that extends for about 20 residues in the *C. thermoaceticum* enzyme but only for about six in the *C. acidiurici* enzyme (see Figure 5). In each case where Tyr substitutes for a nonaromatic residue or Lys or Arg substitutes for a nonbasic residue, the residue is located in a region of the *C. thermoaceticum* sequence that has either reduced hydrophobicity or enhanced hydrophilicity relative to the *C. acidiurici* sequence. When codons having the same bases in the first two positions but a different third base (i.e., wobble base changes) encode amino acids with different properties in the two sequences, the net effect of these amino acid changes is to increase hydrophilicity at these positions of the *C. thermoaceticum* FTHFS. Overall, the trend of changes in amino acids in nonconserved positions is to increase hydrophilicity and the content of basic amino acids in the *C. thermoaceticum* enzyme relative to the *C. acidiurici* enzyme. We speculate that some of these changes may result in ionic interactions or

salt bridges that enhance thermostability. Factors contributing to thermostability in proteins can be quite subtle, as in the cases of the α-subunit of tryptophan synthase of *E. coli* (Yutani et al., 1977) and the kanamycin-inactivating enzyme (Matsumura et al., 1984), where single amino acid substitutions have been reported to enhance thermostability. Elucidation of these factors can be difficult [see the discussion by Argos et al. (1979)]. However, we are able to conclude that the predicted increase in the number or extent of hydrophobic domains in the thermophile FTHFS sequence relative to that in the mesophile sequence was not observed, unless the region of 390–410 in the *C. thermoaceticum* sequence is such a domain. Indeed, the overall content of hydrophobic residues is very similar in the two enzymes. The major differences in the amino acid compositions of these two proteins are in the content of acidic (lower in *C. thermoaceticum*) and basic (higher in *C. thermoaceticum*) residues. We now think these residues to be important in the thermostability of the *C. thermoaceticum* FTHFS, although the mechanism(s) of their participation in FTHFS thermostabilization is as yet unknown.

Acknowledgments

We acknowledge the very helpful advice of Drs. Bert Ely, Vicki Vance, and Farouk Kambaty on sequencing techniques essential to the success of this study. We also thank Kristin French for assistance in preparing figures for publication and Cathy Abbott for final preparation of the manuscript.

References

Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., & Tratschin, J. D. (1979) *Biochemistry 18*, 5698–5703.

Biggin, M. D., Gibson, T. J., & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 3963–3965.

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

Bognar, A. L., Osborne, C., & Shane, B. (1987) *J. Biol. Chem. 262*, 12337–12343.

Brewer, J. M., Ljungdahl, L., Spencer, T. E., & Neece, S. H. (1970) *J. Biol. Chem. 245*, 4798–4503.

Bruschi, M., & Guerlesquin, F. (1988) *FEMS Microbiol. Rev. 54*, 155–176.

Chen, E. Y., & Seeburg, P. H. (1985) *DNA 4*, 165–170.

Devereux, J., Haeberli, P., & Smithies, O. (1984) *Nucleic Acids Res. 12*, 387–395.

Duncan, T. M., Parsonage, D., & Senior, A. E. (1986) *FEBS Lett. 208*, 1–6.

Elliott, J. I., & Ljungdahl, L. G. (1982) *Arch. Biochem. Biophys. 215*, 245–252.

Fry, D. C., Kuby, S. A., & Mildvan, A. S. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 907–911.

Fuchs, G. (1986) *FEMS Microbiol. Rev. 39*, 181–213.

Gay, N. J., & Walker, J. E. (1981) *Nucleic Acids Res. 9*, 2187–2194.

Hager, P. W., & Rabinowitz, J. C. (1985) in *The Molecular Biology of the Bacilli* (Dubnau, D., Ed.) Vol. 2, pp 1–32, Academic Press, New York.

Hase, T., Masubara, H., Koike, H., & Katoh, S. (1983) *Biochim. Biophys. Acta 744*, 46–52.

Hausinger, R. P., & Howard, J. B. (1982) *J. Biol. Chem. 257*, 2483–2490.

Henikoff, S. (1984) *Gene 28*, 351–359.

Holmes, D. S., & Quigley, M. (1981) *Anal. Biochem. 114*, 193–197.

Horii, T., Ogawa, T., & Ogawa, H. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 313–317.

Kyte, J., & Doolittle, R. F. (1982) *J. Mol. Biol. 157*, 105–132.

Ljungdahl, L. G. (1979) *Adv. Microb. Physiol. 19*, 149–243.

Ljungdahl, L. G. (1984) in *Folates and Pterins* (Blakley, R. L., & Benkovic, S. J., Eds.) Vol. 1, pp 555–579, Wiley-Interscience, New York.

Ljungdahl, L. G. (1986) *Annu. Rev. Microbiol. 40*, 415–450.

Ljungdahl, L. G., Brewer, J. M., Neece, S. H., & Fairwell, T. (1970) *J. Biol. Chem. 245*, 4791–4797.

Lovell, C. R., & Hui, Y. (1989) *Abstr. Annu. Meet. Am. Soc. Microbiol. 1101*, 234.

Lovell, C. R., Przybyla, A., & Ljungdahl, L. G. (1988) *Arch. Microbiol. 149*, 280–285.

MacKenzie, R. E., & Rabinowitz, J. C. (1971) *J. Biol. Chem. 246*, 3731–3736.

Maley, G. F., Maley, F., & Baugh, C. M. (1982) *Arch. Biochem. Biophys. 216*, 551–558.

Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, pp 1–545, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Matsumura, M., Katakura, Y., Imanaka, T., & Aiba, S. (1984) *J. Bacteriol. 160*, 413–420.

Mayer, F., Elliott, J. I., Sherod, D., & Ljungdahl, L. G. (1982) *Eur. J. Biochem. 124*, 397–404.

Mevarech, M., Rice, D., & Haselkorn, R. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 6476–6480.

Mizusawa, S., Nishimura, A., & Seela, F. (1986) *Nucleic Acids Res. 14*, 1319–1324.

O'Brien, W. E., Brewer, J. M., & Ljungdahl, L. G. (1976) in *Enzymes and Proteins from Thermophilic Microorganisms, Structure and Functions* (Zuber, H., Ed.) pp 249–262, Birkhauser, Boston, Basel, and Stuttgart.

Perutz, M. F., & Raidt, H. (1975) *Nature 255*, 256–259.

Queen, C., & Korn, L. J. (1984) *Nucleic Acids Res. 12*, 581–599.

Rabinowitz, J. C., & Picer, W. E., Jr. (1962) *J. Biol. Chem. 237*, 2898–2902.

Robson, R. L. (1984) *FEBS Lett. 173*, 394–398.

Saint-Girons, I., Duchange, N., Zakin, M. M., Park, I., Margarita, D., Ferrara, P., & Cohen, G. N. (1983) *Nucleic Acids Res. 11*, 6723–6732.

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

Saraste, M., Gay, N. J., Eberle, A., Runswick, M. J., & Walker, J. E. (1981) *Nucleic Acids Res. 9*, 5287–5296.

Schiefer-Ullrich, H., Wagner, R., Durre, P., & Andreesen, J. R. (1984) *Arch. Microbiol. 138*, 345–353.

Shannon, K. W., & Rabinowitz, J. C. (1988) *J. Biol. Chem. 263*, 7717–7725.

Shine, J., & Dalgarno, L. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 1342–1346.

Smith, D. R., Rood, J. I., Bird, P. I., Sneddon, M. K., Calvo, J. M., & Morrison, J. F. (1982) *J. Biol. Chem. 257*, 9043–9048.

Staben, C., & Rabinowitz, J. C. (1986) *J. Biol. Chem. 261*, 4629–4637.

Stauffer, G. V., & Stauffer, L. T. (1988) *Mol. Gen. Genet. 212*, 246–251.

Stormo, G. D. (1986) in *Maximizing Gene Expression* (Reznikoff, W., & Gold, L., Eds.) pp 195–224, Butterworths, Boston.

Sundaresan, V., & Ausubel, F. M. (1981) *J. Biol. Chem. 256*, 2808–2812.

Tabor, S., & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4767–4771.

Tanaka, M., Haniu, M., Yasonobu, K. T., Mortenson, L. E. (1977) *J. Biol. Chem. 252*, 7093–7100.

Torok, I., & Kondorosi, A. (1981) *Nucleic Acids. Res. 9*, 5711-5723.
Waber, L. J., & Wood, H. G. (1979) *J. Bacteriol. 140*, 468-478.
Walker, J. E., Saraste, M., Runswick, M. J., & Gay, N. J. (1982) *EMBO J. 1*, 945-951.
Whitehead, T. R., & Rabinowitz, J. C. (1986) *J. Bacteriol. 167*, 205-209.
Whitehead, T. R., & Rabinowitz, J. C. (1988) *J. Bacteriol. 170*, 3255-3261.
Whitehead, T. R., Park, M., & Rabinowitz, J. C. (1988) *J. Bacteriol. 170*, 995-997.
Wood, H. G., Ragsdale, S. W., & Pezacka, E. (1986) *FEMS Microbiol. Rev. 39*, 345-362.
Yutani, K., Ogasahara, K., Sugino, Y., & Matsushiro, A. (1977) *Nature 267*, 274-275.

Excited-State Properties of *Escherichia coli* DNA Photolyase in the Picosecond to Millisecond Time Scale†

Paul F. Heelis,*,‡ Tadashi Okamura,§ and Aziz Sancar‖

Faculty of Science and Innovation, North East Wales Institute, Deeside, Clwyd CH5 4BR, United Kingdom, Laboratory of Physical Chemistry, Tokyo Zokei University, Hachioji, Tokyo 193, Japan, and Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina 27599

Received January 8, 1990; Revised Manuscript Received March 1, 1990

ABSTRACT: *Escherichia coli* DNA photolyase contains a stable flavin radical that is readily photoreduced in the presence of added electron donors. Picosecond, nanosecond, and conventional flash photolysis technique have been employed to investigate the events leading to photoreduction from 40 ps to tens of milliseconds following flash excitation. Direct light absorption by the flavin radical produces the first excited doublet state which undergoes rapid (within 100 ps) intersystem crossing to yield the lowest excited quartet ($n\pi^*$) state. In contrast, light absorption by the folate chromophore produces a new intermediate state via interaction of the folate excited singlet state with the ground-state flavin radical, leading to an enhanced yield of the excited radical doublet state and hence quartet state. Subsequent reaction of the excited quartet state involves hydrogen atom abstraction from a tryptophan residue. Secondary electron transfer from added electron donors occurs to the oxidized tryptophan radical with rate constants ranging from 10^4 (dithiothreitol) to 4×10^6 M^{-1} s^{-1} (*n*-propyl gallate). The low value of the latter rate compared to reduction of the tryptophan radical in lysozyme suggests that the reactive tryptophan is highly buried in photolyase. A redox potential diagram has been constructed for the ground and excited states involved. It is concluded that the one-electron reduction potential of the excited quartet state of the flavin radical must be at least 1.23 V more positive than the ground state, in agreement with the value of $\Delta E > 1.77$ V calculated from spectroscopic data.

DNA photolyase (EC 4.1.99.3) catalyzes the photochemical conversion of pyrimidine dimers in DNA into pyrimidines, thus reversing the effect of far-UV (200-300 nm) radiation. The enzyme from *Escherichia coli* contains the FAD blue neutral radical (Sancar & Sancar, 1984; Jorns et al., 1984) and a second chromophore that was recently identified as 5,10-methenyltetrahydrofolate (Johnson et al., 1988). Recent studies (Payne et al., 1987; Sancar et al., 1987b) have shown that in vitro photoreactivation proceeds via photoreduction of the flavin radical followed by dimer repair by the photoexcited reduced form. Hence, the mechanism of radical photoreduction is clearly important in understanding the in vitro photoreactivation mechanism.

We suggested previously (Heelis & Sancar, 1986; Heelis et al., 1987; Okamura et al., 1989) that excitation of the flavin radical (**1** → **2**, Figure 1) results in an intramolecular electron (or H atom) transfer from an unknown amino acid residue (DH) to the excited quartet state of the radical (**3** → **4**, Figure 1), to transiently form a reduced flavin-amino acid radical intermediate (**4**, Figure 1). A dark reversal of this electron-transfer process occurs (**4** → **1**, Figure 1) over 20 ms (Heelis et al., 1987) so that no net change in the enzyme is detected upon long-term irradiation. However, in the presence of external electron donors (RH, e.g., thiols, glycyltyrosine, or NADH) reduction of the oxidized amino acid radical occurs (**4** → **6**, Figure 1), resulting in permanent reduction (in the absence of O_2) of the flavin radical. The present study involves the use of both picosecond and nanosecond laser flash photolysis techniques together with conventional photolysis experiments to investigate in more detail the nature of the intermediates and the kinetics of the processes involved in radical photoreduction.

Experimental Procedures

DNA photolyase was prepared as described previously (Sancar et al., 1984, 1987a). The enzyme was in the "blue form", essentially free of oxidized FAD as shown by absorbance and fluorescence measurements. The stock solutions of the enzyme were stored at −20 or −80 °C in storage buffer containing 50 mM Tris·HCl, pH 7.5, 50 mM NaCl, 10 mM DTT, 1 mM EDTA, and 50% glycerol. Experiments were carried out by diluting the stock solution from 2- to 60-fold into a buffer containing 50 mM Tris·HCl, pH 7.4, 100 mM NaCl, and other components as indicated. Enzyme concentrations were in the 10^{-5}-10^{-4} range with respect to the flavin

† This work was supported by Grant GM31082 from the National Institutes of Health and partly by grants from NATO, the SERC, and the Wellcome Trust.
* To whom correspondence should be addressed.
‡ North East Wales Institute.
§ Tokyo Zokei University.
‖ University of North Carolina.

0006-2960/90/0429-5694$02.50/0