Structure of Dihydrofolate Reductase: Primary Sequence of the Bovine Liver Enzyme†

Por-Hsiung Lai,‡ Yu-Ching E. Pan,§ John M. Gleisner,‖ Darrell L. Peterson,⊥ Katie R. Williams, and Raymond L. Blakley*

ABSTRACT: The primary structure of dihydrofolate reductase from bovine liver has been established by Edman degradation of the intact carboxymethylated protein and of peptides obtained from the protein by the action of cyanogen bromide, trypsin, and the protease from *Staphylococcus aureus*, respectively. Since separation of some of the peptide mixtures by classical methods proved impossible, new systems were developed for the use of high-performance liquid chromatography to separate such mixtures. Some of the cleavage procedures used to obtain peptides gave atypical results at certain peptide bonds. The results are discussed in terms of the residues involved in these unexpectedly resistant or sensitive bonds. The sequence of the bovine liver enzyme is compared with those published for the enzyme from other sources, and known or probable functions of invariant residues are described. Sequences of vertebrate and bacterial reductases are compared and contrasted, and a possible role is considered for the residues which are invariant in bacterial reductases, but different in vertebrate reductases, in determining the selective inhibitory action of trimethoprim on bacterial reductases.

Extensive reviews have appeared on the metabolic importance of the reduction of dihydrofolate to tetrahydrofolate catalyzed by dihydrofolate reductase (EC 1.5.1.3), the clinical significance of the inhibition of the mammalian enzyme by antitumor drugs like methotrexate, and the mode of the specific interaction between bacterial reductase and certain 2,4-diaminopyrimidines like trimethoprim that are useful in treating bacterial infections (Blakley, 1969, 1981; Hitchings & Roth, 1980; Hitchings & Smith, 1980).

Although the primary structures have been reported for the reductase from a murine lymphoma (Stone et al., 1979), pig liver (Smith et al., 1979), and chicken liver (Kumar et al., 1980), and there is considerable homology between the reductases from these vertebrate sources, considerable importance attaches to the primary structure of the bovine liver reductase. In the first place, the more sequences of the vertebrate enzyme that can be compared, the more certain are the conclusions that can be drawn from sequence comparisons about the role of specific residues in the maintenance of protein structure, in the determination of substrate specificity, in the binding of substrates, and in enzyme catalysis. Second, the structure of the bovine liver enzyme is of particular interest because it has been used extensively in detailed comparisons of inhibitory properties of numerous diaminopyrimidines (Blaney et al., 1979; Dietrich et al., 1980; Li et al., 1981) and triazines (Hansch et al., 1981; Dietrich et al., 1980).

The establishment of the primary structure of the bovine liver enzyme is presented here, and the implications of sequence in comparison with sequences of the enzyme from other sources are discussed.

Experimental Procedures

Materials

Dihydrofolate reductase was prepared from beef liver as previously described (Peterson et al., 1975). The purified reductase had a specific activity of 100 units/mg and was homogeneous as judged by analytical ultracentrifugation, polyacrylamide gel electrophoresis, and titration with methotrexate. Iodo[1-^{14}C]acetic acid (14.73 mCi/mmol) was obtained from New England Nuclear. Iodoacetic acid was recrystallized 3 times from petroleum ether (boiling range 30–60 °C) before use. Citraconic anhydride and polybrene were purchased from Aldrich Chemical Co. Citraconic anhydride was redistilled under reduced pressure before use. Acetyl chloride and cyanogen bromide were from Eastman. Carboxypeptidases A and B, pyroglutamate aminopeptidase, and TPCK[1]-treated trypsin were obtained from Sigma, and *Staphylococcus aureus* V8 protease was from Miles Laboratories. All sequencer solvents and reagents were either Pierce sequenal grade or Beckman sequence chemicals. HPLC solvents were Millipore-filtered reagent grade or HPLC grade. Reagent-grade pyridine was redistilled from sodium hydroxide pellets before use. Fluorescamine was obtained from Roche. Phosphocellulose was purchased from Sigma, and Bio-Gel P-2, P-4, P-6, and P-10 columns were from Bio-Rad Laboratories.

Methods

Protein Carboxymethylation. Carboxymethylation of the single cysteine residue of the reductase by iodoacetate was accomplished by the method of Crestfield et al. (1963). The S-carboxymethylated protein was then dialyzed against 5% acetic acid in water to remove excess reagents and urea. When

† From the Department of Biochemistry, College of Medicine, University of Iowa, Iowa City, Iowa 52242. *Received November 10, 1981.* This work was supported in part by U.S. Public Health Service Research Grant CA 13840 from the National Cancer Institute. Sequencing was carried out with instruments funded by U.S. Public Health Service Instrumentation Grant GM 27441 from the National Institute for General Medical Sciences. A preliminary report of this work was presented at the Sixth International Symposium on the Chemistry and Biology of Pteridines, La Jolla, CA, 1978 (Lai et al., 1979).

* Correspondence should be addressed to this author at the Division of Biochemical and Clinical Pharmacology, St. Jude Children's Research Hospital, Memphis, TN 38101.

‡ Present address: Sequemat Inc., Watertown, MA 02172.

§ Present address: National Institutes of Environmental Health Sciences, Research Triangle Park, NC 27709.

‖ Present address: Virginia Mason Research Center, Seattle, WA 98101.

⊥ Present address: Department of Biochemistry, Medical College of Virginia, Virginia Commonwealth University, Richmond, VA 23298.

[1] Abbreviations: TPCK, L-1-[(*p*-toluenesulfonyl)amido]-2-phenylethyl chloromethyl ketone; HPLC, high-performance liquid chromatography; PTH, phenylthiohydantoin; TLC, thin-layer chromatography; ODS, octadecylsilane; DMAA, *N,N*-dimethyl-*N*-allylamine; $NaDodSO_4$, sodium dodecyl sulfate; DHFR, dihydrofolate reductase.

0006-2960/82/0421-3284$01.25/0

iodo[1-^{14}C]acetic acid was employed for carboxymethylation of protein, a small amount of 2-mercaptoethanol was included in the 5% acetic acid used for dialysis.

Cleavage by Cyanogen Bromide. The carboxymethylated dihydrofolate reductase (12 mg/mL) was treated with cyanogen bromide (100-fold molar excess over methionine) in 70% formic acid at room temperature as previously described (Gleisner et al., 1975). After 18 h, the reaction mixture was diluted 10-fold with cold water and freeze-dried twice from the same volume of cold water.

Citraconylation of Lysine Residues. Reversible blocking of lysine amino groups by citraconic anhydride was performed as described previously (Gleisner et al., 1975) except that the sample of carboxymethylated dihydrofolate reductase (500 mg) was prepared for treatment by dissolving in 55 mL of 2 M urea and then adjusting the pH to 8.5 with triethylamine. After the reaction, the mixture was dialyzed at 4 °C for 4 h each against three changes of 4 L of 0.05 M ammonia, pH 10.5.

Enzymatic Hydrolysis. Digestion of citraconylated carboxymethylated dihydrofolate reductase by TPCK-treated trypsin was carried out as before (Gleisner et al., 1975) with an enzyme:substrate ratio of 1:100 (w/w). However, the freeze-dried protein sample (500 mg) was prepared for digestion by dissolving in 25 mL of 0.05 N NaOH and then adjusting the pH to 8.0 with formic acid. Digestion of peptide with *S. aureus* V8 protease was carried out at 37 °C for 16 h in 0.1 M NH_4HCO_3 with an enzyme:substrate ratio of 1:50 (w/w). Removal of the amino-terminal pyroglutamate residue from peptide CB6 by bovine liver pyroglutamate aminopeptidase was performed according to Podell & Abraham (1978) except that 0.05 M NH_4HCO_3 buffer, pH 8.0, was used and glycerol was omitted from the reaction mixture.

Peptide Fractionation. A variety of different procedures was employed for the purification of peptides generated by chemical or enzymic procedures. Bio-Gel P-2, P-4, P-6, and P-10 columns equilibrated with 30% or 50% formic acid or with 0.05 M ammonia, pH 10, were used for peptide separation by gel filtration. Chromatography of certain peptides on phosphocellulose was performed as previously described (Gleisner et al., 1975). Peptides were assayed in the effluent fractions by the absorbance at 215 or 280 nm. In some instances, peptides were purified by preparative paper electrophoresis in pyridine–acetic acid–water (100:4:900) and subsequently eluted from the paper with 20% acetic acid.

Reversed-phase HPLC on C_{18} columns has also been used to isolate some peptides. For the isolation of large peptides produced by the action of cyanogen bromide or by limited tryptic digestion, a gradient of 2-propanol in 0.1% (v/v) phosphoric acid (pH 2.5) was used. For separation of peptides obtained with *S. aureus* protease and for desalting of peptide samples, a gradient of 2-propanol in 0.1 M pyridinium formate (pH 3.5) was used. It was, of course, necessary to evaporate under reduced pressure any 2-propanol from fractions to be desalted. Fractions were assayed for peptide content according to Böhlen et al. (1973) except that a Hitachi fluorometer was used and peptide samples (volume 10–20 μL) were mixed with 2 mL of 0.3 M lithium borate buffer, pH 9.1, and 0.5 mL of fluorescamine solution (15 mg/50 mL of acetone).

Amino Acid Analysis. Peptide samples were hydrolyzed in sealed, N_2-flushed, evacuated tubes in constant-boiling HCl containing 0.5% (v/v) phenol and 0.5% (v/v) 2-mercaptoethanol at 110 °C for 24 h. Amino acid analyses were performed either with the two-column system of a JLC-6AH automatic amino acid analyzer or with a Beckman 121MB analyzer. For homoserine-containing peptides, the dried hydrolysates were redissolved in pyridine–acetic acid–water (10:0.4:90), pH 6.5, and kept at 105 °C for 1 h to convert homoserine lactone to homoserine. Analysis of homoserine with the JLC-6AH analyzer was performed as described previously (Gleisner et al., 1974).

Edman Degradations. In earlier work, automated Edman degradation was performed according to the method of Edman & Begg (1967) on a JEOL JAS-47K automatic sequence analyzer as previously described (Gleisner et al., 1975). In later work, a Beckman 890C sequencer equipped with a cold trap and a Sequemat P-6 autoconverter was employed. The repetitive yields obtained with the JAS-47K were 90 ± 3%, and those obtained with the 890C were 94 ± 2%. Automatic sequencing of large cyanogen bromide peptides and limited tryptic peptides in the Beckman 890C sequencer was performed in the presence of 3 mg of polybrene. Before sample addition, the polybrene was purified in the cup by running three complete cycles. A modification of the Beckman protein program (no. 11978) was used for automatic sequencing. The modification includes replacing 1.0 M Quadrol buffer with 0.33 M Quadrol buffer and the use of prolonged N_2 drying (2 min) after the second wash, cleavage, and extraction steps. After sample addition, prewashing and double coupling were performed before the first cycle was started, as in the program used by Hunkapillar & Hood (1978).

Manual Edman degradation was performed in a 0.6-mL Reactivial according to the method of Tarr (1977) with the following modifications: (1) coupling time was increased from 3 to 4.5–5 min; (2) in all cases, washing after coupling was carried out with two washes of heptane–ethyl acetate (10:1), followed by two washes with heptane–ethyl acetate (2:1); (3) cleavage time was increased to 3 min from 1.5 min.

In some cases, quantitative determination of anilinothiazolinone or PTH derivatives was carried out by amino acid analysis following hydrolysis in 30% HI at 150 °C for 5 h in sealed evacuated tubes. In certain cases, hydrolysis with 0.1 M $Na_2S_2O_4$ in 0.2 N NaOH was also carried out to distinguish between alanine and serine (Smithies et al., 1971). Thin-layer chromatography of the PTH derivatives on silica gel, with $CHCl_3$–MeOH (80:20) as developing solvent (Pataki, 1966), was carried out for distinguishing amides from corresponding acids. In later experiments, HPLC identification of PTH-amino acids was used to analyze residues obtained from the 890C sequencer–P-6 autoconverter system.

When conversion of anilinothiazolinone acids to PTH-amino acids was performed with the P-6 autoconverter, about 0.5 mL of the conversion acid (MeOH–acetyl chloride) was added to the conversion cell after the butyl chloride extract had been transferred from the sequencing cup to the conversion cell before drying of extract was started. After the extract was dried, the standard MeOH–acetyl chloride reagent addition for conversion was made as in the standard procedure. This procedure improved the yield of PTH-serine and PTH-threonine. Direct detection of these two derivatives at 254 nm was then possible (M. J. Horn, Sequemat Inc., unpublished results).

Samples of PTH-amino acids were recovered from the sequencer–P-6 autoconverter in dichloroethane–methanol and dried under a stream of ultra-high-purity nitrogen at 40–50 °C. The dried PTH-amino acids were kept in a freezer at −100 °C until analysis. After solution in an appropriate volume (15–100 μL) of methanol, 5 μL of the solution was then drawn into a microsyringe so as to sandwich it between 5- and 10-μL volumes of 20% methanol in 0.01 M sodium acetate buffer,

pH 6.0. This sample was injected into an Altex or Spectra Physics-Altex HPLC system. The Altex system consisted of two Model 110A pumps, Model 420 microprocessor, Model 153 UV–visible detector with analytical flow cell, Model C-RIA integrator–recorder, and Model 210 injector. The Spectra Physics system consisted of a Model 8700 solvent delivery system, with an Altex Model 153 UV–visible detector and analytical flow cell, Altex Model C-RIA integrator–recorder, and Altex Model 210 injector.

The PTH-amino acids were identified by HPLC on an Altex Ultrasphere–ODS 5-μm column (4.6 × 250 mm) by a gradient formed from the following components: 20% MeOH in 0.01 M sodium acetate buffer, pH 6.0 (solvent A), and 88% MeOH in 0.01 M sodium acetate buffer, pH 6.0 (solvent B). The gradient was developed as follows: 0% solvent B for 2 min; 0–35% solvent B over 8 min; 35% solvent B for 2 min; 35–50% solvent B over 5 min; 50% solvent B for 8 min; 50–30% solvent B over 3 min; 30% solvent B for 12 min; 30–0% solvent B over 5 min. The flow rate was 1.0 mL/min at room temperature. In later experiments, a short Altex Ultrasphere–ODS 5-μm column (4.6 × 150 mm) was used, and PTH-amino acids were separated by the following gradient of solvent B in solvent A: 0% solvent B for 2 min; 0–35% solvent B over 8 min; 35% solvent B for 2 min; 35–55% solvent B over 10 min; 55% solvent B for 3 min; 55–0% solvent B over 10 min. The flow rate was 1.0 mL/min at room temperature. PTH-amino acids were detected by their absorbance at 254 nm.

The short-column method for HPLC analysis of PTH-amino acids gave excellent results [Figure 6 of supplementary material; see paragraph at end of paper regarding supplementary material] and was especially convenient because the total analysis time was considerably less than the time for an Edman degradation cycle on the Beckman 890C sequencer.

Results

The complete amino acid sequence for the beef liver dihydrofolate reductase is shown in Figure 1, where the various fragments used to establish the sequence of the enzyme are also indicated. The procedures involved and the sequence data obtained are outlined below and described in greater detail in the supplementary material.

Automated Sequence Analysis of Intact Carboxymethylated Protein. Two separate determinations were carried out, one through 40 cycles and the other through 43 cycles. The results of the latter are shown in Table I. The position of the single cysteine residue in the N-terminal region of the beef liver enzyme sequence was confirmed by automated sequence analysis of the ^{14}C-labeled S-carboxymethylated enzyme through 21 cycles. A single radioactive peak which accounted for 90% of the total radioactivity incorporated into the protein appeared at cycle 6, a result establishing 6 as the position of the single cysteine residue rather than position 11 as reported previously (Lai et al., 1979) from attempted identification by TLC of PTH-amino acids from unlabeled carboxymethylated protein.

Cyanogen Bromide Cleavage. Amino acid analysis of the enzyme indicated the apparent presence of 4.2 methionines (Table II). Similarly, amino acid analysis of the products of cyanogen bromide treatment of the carboxymethylated protein indicated the presence of four homoserine residues and 0.4 methionine residue. Ultimately, however, complete solution of the sequence as shown in Figure 1 revealed that, in fact, five methionine residues are present.

Methods used for the isolation of the pure cyanogen bromide peptides are summarized in Figure 7, which together with details of the purification procedures not described below can be found in the supplementary material. Peptides were obtained in sufficient purity for sequence determination with the exception of fraction 5-1. The three small peptides this fraction contained were not purified further because their sequences were already known (from fractions 4-1, 4-2, and 3-2).

A major difficulty in obtaining the sequences of all cyanogen bromide peptides was presented by the incomplete resolution of these peptides by ion exchange on phosphocellulose or by gel filtration in 50% formic acid on Bio-Gel P-4. This difficulty was resolved by employing reversed-phase HPLC on a C_{18} column as shown in Figure 2. The solvent system used (a gradient of 2-propanol in 0.1% v/v phosphoric acid) worked efficiently despite the poor solubility of the peptide mixture under many other conditions. Detection of peptides by a manual procedure performed on each fraction (see Experimental Procedures) gave much more satisfactory results than

Table I: Automated Sequence Analysis of Carboxymethylated Dihydrofolate Reductase[a]

cycle	product of hydrolysis: by HI	product of hydrolysis: by NaOH	yield (nmol)	TLC result	residue assigned
1	Val		393		Val
2	Arg		259		Arg
3	Pro		490		Pro
4	Leu		642		Leu
5	Asp		346	Asn	Asn
6	Ala	no Ala	264 (0)[b]		(Cys)[c]
7	Ile + alle		818		Ile
8	Val		352		Val
9	Ala	Ala	262 (204)[b]		Ala
10	Val		442		Val
11	Ala	Ala	156 (62)[b]		(Ser)[d]
12	Glu		274	Gln	Gln
13	Asp		244	Asn	Asn
14		Met	96		Met
15	Gly		404		Gly
16	Ile + alle		349		Ile
17	Gly		338		Gly
18	Lys		150		Lys
19	Asp		148	Asn	Asn
20	Gly		196		Gly
21	Asp		152	Asp	Asp
22	Leu		276		Leu
23	Pro		138		Pro
24	Gly + Ala		142		Trp
25	Pro		112		Pro
26	Pro		74		Pro
27	Leu		192		Leu
28	Arg		ND[e]		Arg
29	Asp		86	Asn	Asn
30	Glu		100	Glu	Glu
31	Phe		80		Phe
32	Glu		96	Gln	Gln
33	Tyr		26		Tyr
34	Phe		34		Phe
35	Glu		52		Glx[f]
36	Arg		4		Arg
37	NI[i]				(Met)[g]
38	Abu		ND[e]		Thr
39	Abu		ND[e]		Thr
40	Val		38		Val
41	Ala		29		(Ser)[h]
42	Ala		32		(Ser)[h]
43	Val		36		Val

[a] Analysis was carried out with 1600 nmol of carboxymethyl-DHFR and the use of a JAS-47K JEOL sequence analyzer. Abbreviations: alle, alloisoleucine; Abu, α-aminobutyric acid. [b] Yield from hydrolysis with NaOH. [c] Known to be Cys from results of sequencing ^{14}C-labeled *S*-carboxymethylated DHFR. [d] Known to be Ser from CB1. [e] ND, not determined but appeared as major peaks in amino acid analysis. [f] Known to be Gln from LT2. [g] Known to be Met from LT3. [h] Known to be Ser from LT3. [i] NI, not identified.

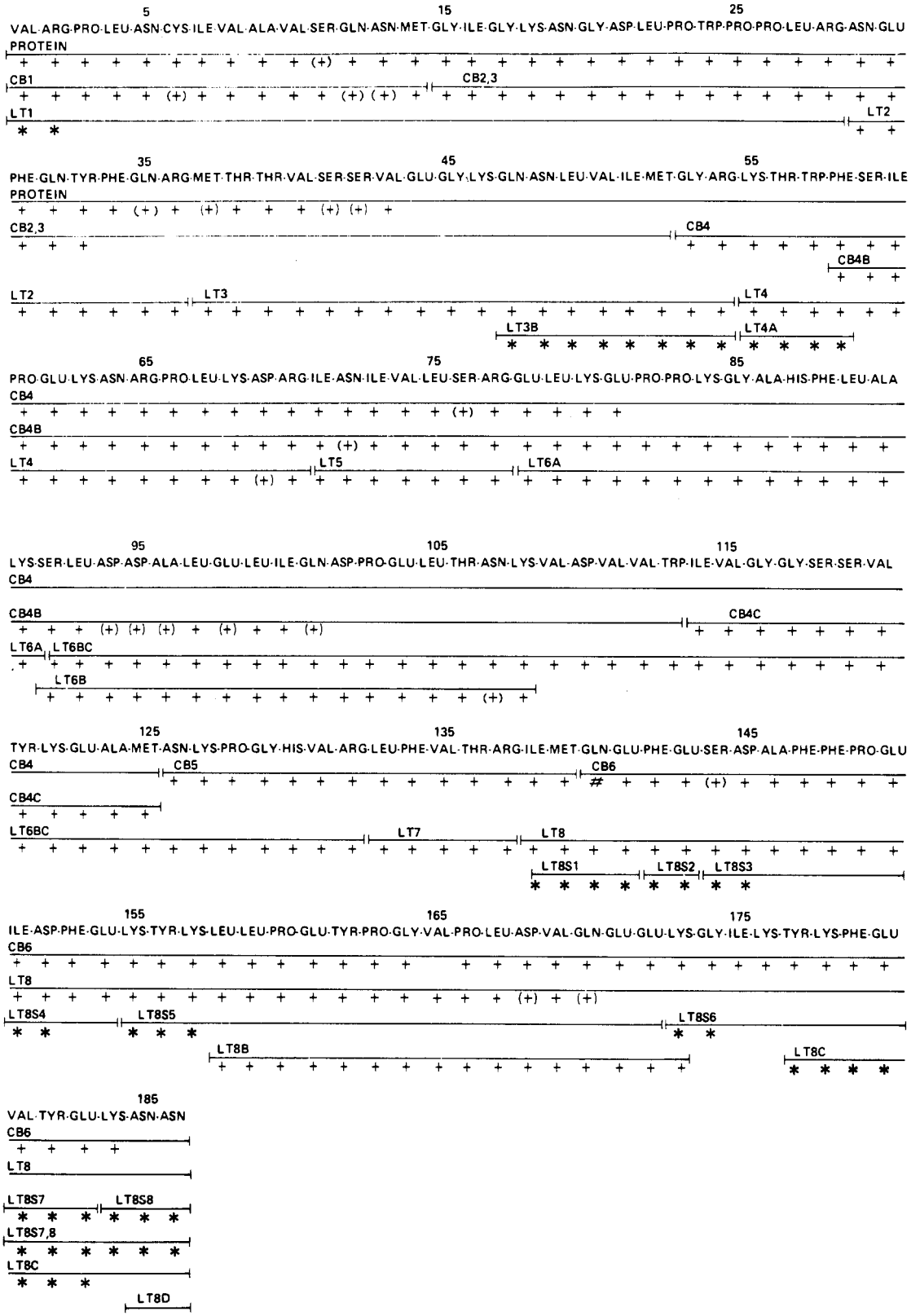

FIGURE 1: Summary of the peptide fragments used to establish the complete amino acid sequence of dihydrofolate reductase from bovine liver. Peptides generated with cyanogen bromide and by limited tryptic digestion are designated CB and LT, respectively, and are numbered sequentially from the N terminus according to their position in the protein. Peptides obtained by *S. aureus* protease action on LT8 are designated LT8S followed by a number indicating the position of the peptide within LT8. Solid lines indicate the extent of the isolated peptide. Symbols below the solid lines indicate sequencing methods used: (+) automated Edman degradation; (*) manual Edman degradation; (#) pyroglutamate removed by pyroglutamate aminopeptidase. Parentheses indicate unresolved ambiguity between Ala–Ser, Asp–Asn, or Glu–Gln.

detection by ultraviolet absorbance. No loss of the resolution achieved by HPLC was occasioned by this procedure over that resulting from the necessity to collect in fractions during preparative chromatography.

The identity of the peptides obtained in the HPLC fractions (Figure 2) was indicated in the first place by N-terminal analysis with the manual Edman procedure and confirmed by amino acid analysis. N-Terminal analysis of fraction 5-3 gave

Table II: Amino Acid Composition of Dihydrofolate Reductase and of Peptides Obtained by Cleavage with Cyanogen Bromide[a]

amino acid	CB1	CB2,3	CB4	CB4B[b]	CB4C[b]	CB5	CB6	DHFR[c]
Asp	1.6 (2)	3.9 (4)	8.7 (8)	6.9 (8)		1.1 (1)	5.6 (5)	20.2 (20)
Thr	0.5 (0)	0.4 (2)	2.2 (2)	1.0 (1)		0.9 (1)		6.0 (5)
Ser	1.0 (1)	0.8 (2)	5.4 (5)	2.7 (3)	1.7 (2)		0.3 (1)	10.1 (9)
Glu	1.2 (1)	4.0 (5)	7.5 (7)	8.9 (6)	1.0 (1)		11 (11)[d]	24.0 (24)
Pro	1.0 (1)	3.0 (3)	5.7 (5)	5.4 (5)		0.7 (1)	4.4 (4)	14.3 (14)
Gly	0.7 (0)	3.3 (4)	4.6 (4)	3.0 (1)	1.8 (2)	1.1 (1)	3.5 (2)	11.9 (11)
Ala	0.6 (1)		4.2 (4)	2.9 (3)	0.9 (1)		0.5 (1)	7.7 (6)
Val	2.0 (3)	2.8 (3)	7.2 (6)	4.7 (4)	1.5 (2)	1.6 (2)	3.8 (3)	17.0 (17)
Ile	0.6 (1)	1.8 (2)	4.5 (5)	2.7 (4)	0.5 (1)	0.7 (1)	2.0 (2)	9.0 (11)
Leu	0.9 (1)	2.6 (3)	8.3 (8)	5.0 (8)		1.0 (1)	3.6 (3)	14.4 (16)
Tyr		1.0 (1)	1.9 (1)	1.7 (0)	0.8 (1)		3.8 (4)	6.8 (6)
Phe		2.7 (2)	3.1 (2)	3.5 (2)		0.8 (1)	4.6 (5)	10.0 (10)
His			0.9 (1)	0.8 (1)		0.6 (1)		2.0 (2)
Lys	0.5 (0)	1.7 (2)	7.9 (8)	6.0 (6)[d]	1.0 (1)	1.0 (1)	6.3 (6)	17.1 (17)
Arg	0.6 (1)	2.0 (2)[d]	4.0 (4)[d]	2.2 (3)		1.4 (2)		8.5 (9)
Met		trace (1)						4.2 (5)
Hse	1.0 (1)[d]	1.4 (1)	1.3 (1)	trace (0)	1.0 (1)[d]	1.0 (1)[d]		
Trp		trace (1)	trace (2)	trace (1)				2.1 (3)
Cys (Cm)	0.6 (1)							
Cys (O_3H)								1.6 (1)[f]
yield (%)	25	43	75[e]	30	35	29	83[e]	
residues	14	38	73	56	12	14	47	187 (186)
position	1-14	15-52	53-125	58-113	114-125	126-139	140-186	

[a] In this and succeeding tables, values in parentheses were obtained from sequencing. Other values, except those on the intact protein, were obtained by 24-h hydrolysis. [b] This peptide resulted from cleavage at Trp. [c] As reported previously by Peterson et al. (1975). Corrected values were obtained from hydrolysis of the protein for three time periods. [d] The analytical value was set equal to the appropriate integer for this residue and the number of residues of other amino acids calculated accordingly. [e] Yield obtained from HPLC separation. [f] Analysis of the carboxymethylated protein gave a value of 0.9 carboxymethylcysteine in agreement with the presence of a single cysteine in the protein.

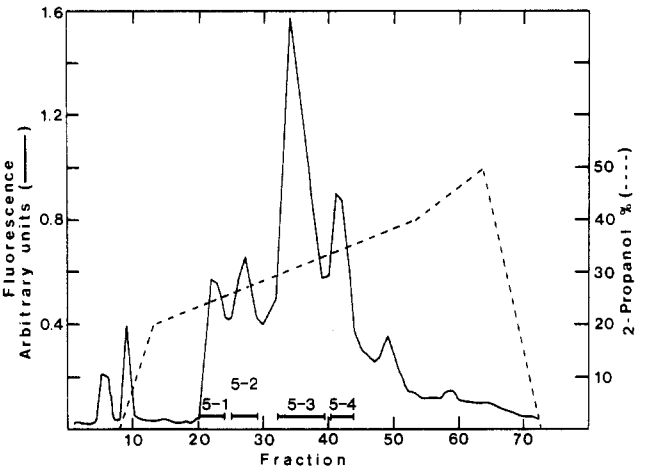

FIGURE 2: Separation by HPLC of cyanogen bromide peptides. The peptide mixture (from 20 mg of carboxymethylated protein) was dissolved in 2 mL of 0.1% (v/v) phosphoric acid and loaded onto the μBondapack C_{18} column (3.9 × 300 mm). Flow rate was 0.25 mL/min; fraction volume was 0.75 mL. For other details, see Methods.

low yields of PTH-glutamine and PTH-glutamic acid methyl ester (a byproduct of the conversion of glutamine anilinothiazolinone to PTH-glutamine). The low yields suggested that the N terminus had been largely blocked as a pyroglutamate residue. This was confirmed by treatment of desalted fraction 5-3 with pyroglutamate aminopeptidase. End-group analysis of the freeze-dried digest gave an excellent yield of PTH-glutamic acid and its methyl ester (Glu is the second residue in the sequence of CB6). The method used in desalting fractions from the HPLC described in Figure 2 is described in the supplementary material.

The amino acid composition of dihydrofolate reductase and derived cyanogen peptides is shown in Table II. The analyses indicate that some of the peptides were accompanied by contaminants, but since sequence analysis gave a single major residue at each cycle further purification was not considered necessary. The details of the sequence determinations of cyanogen bromide fragments, as summarized in Figure 1, are presented in the supplementary material.

Limited Tryptic Digestion. A summary of the various procedures used for separating and purifying peptides produced by tryptic digestion of citraconylated, carboxymethylated protein, followed by deblocking at pH 3, is given in the supplementary material. The first separation was achieved on the basis of solubility. As deblocking proceeded at pH 3, certain peptides precipitated and were separated by centrifugation. Further fractionation was carried out as described in Figure 8, which together with other details is presented in the supplementary material.

It may be seen from Figure 1 that some of the peptides obtained in these fractionation procedures (e.g., LT3B) resulted from cleavage at Lys residues, due to incomplete acylation, and others (e.g., LT4A) resulted from cleavage at an aromatic residue, due to chymotryptic-like activity.

Peptide LT8 was sequenced automatically through 33 cycles of Edman degradation. However, it was important to obtain the complete sequence in order to determine the C-terminal region of the protein sequence. For achievement of this, a sample of LT8 was desalted on a C_{18} HPLC column and subjected to digestion with *S. aureus* protease. Peptides in the protease digest were then separated by reversed-phase HPLC with use of a gradient of 2-propanol in 0.1 M pyridinium formate, pH 3.5. The results, shown in Figure 3, again illustrate the power of HPLC in separating mixtures of peptides, this time of shorter chain length. As indicated in Figure 1, three of these peptides (LT8S7, LT8S8, and LT8S7,8) determined the C-terminal region of the protein.

Further details of the isolation procedures and sequence determinations for the limited tryptic peptides are provided in the supplementary material. The amino acid composition of the limited tryptic peptides is shown in Table III and that of the peptides derived from LT8 by *S. aureus* protease in Table IV. The analysis in Table III suggests that LT6BC

Table III: Amino Acid Composition of Limited Tryptic Peptides[a]

amino acid	LT1	LT2	LT3	LT4	LT5	LT6A[b]	LT6B[b]	LT6BC	LT7	LT8	LT8B	LT8C[c]	LT8D[b]
Asp	4.4 (4)	1.0 (1)	0.9 (1)	2.1 (2)	1.1 (1)		4.2 (4)	6.0 (6)[c]	0.3 (0)	5.5 (5)	1.5 (1)	1.3 (2)	2.0 (2)[c]
Thr			2.0 (2)	1.0 (1)			1.0 (1)	0.8 (1)	0.8 (1)	0.8 (0)			
Ser	1.2 (1)		2.1 (2)	1.0 (1)	0.8 (1)		0.9 (1)	2.0 (3)	0.2 (0)	2.1 (1)			
Glu	1.4 (1)	2.9 (3)	2.7 (2)	1.1 (1)	0.3 (0)	2.1 (2)	3.4 (3)	9.0 (4)	0.4 (0)	11.4 (11)	4.2 (4)	1.9 (2)	
Pro	3.7 (4)			2.0 (2)		1.8 (2)	1.0 (1)	3.8 (2)		5.0 (4)	2.5 (3)		
Gly	3.2 (3)		2.1 (2)			1.4 (1)		3.3 (3)		3.3 (2)	1.3 (1)		
Ala	1.1 (1)		0.3 (0)			2.1 (2)	1.1 (1)	1.9 (2)		1.7 (1)			
Val	3.3 (3)		3.0 (3)		0.9 (1)			5.2 (6)	1.0 (1)	3.8 (3)	2.5 (2)	1.1 (1)	
Ile	2.0 (2)		0.9 (1)	1.0 (1)	1.7 (2)		1.1 (1)	2.3 (2)		2.7 (3)			
Leu	3.4 (3)		1.1 (1)	1.0 (1)	1.0 (1)	2.0 (2)	4.6 (2)	4.1 (4)	1.0 (1)	4.1 (3)	2.9 (3)		
Tyr		0.8 (1)						2.4 (1)		4.1 (4)	0.5 (1)	1.0 (2)	
Phe		2.0 (2)		1.0 (1)[c]	0.2 (0)	1.0 (1)[c]		2.5 (0)	1.0 (1)	5.0 (5)[c]		1.0 (1)[c]	
His						0.9 (1)		0.7 (1)					
Lys	1.0 (1)[c]		1.1 (1)	3.0 (3)		3.1 (3)	1.0 (1)[c]	2.7 (3)	0.2 (0)	6.1 (6)	1.0 (1)[c]	2.5 (2)	
Arg	1.6 (2)	1.0 (1)[c]	1.0 (1)[c]	2.4 (2)	1.0 (1)[c]			0.7 (1)	1.0 (1)[c]	0.5 (0)			
Met	1.1 (1)		2.2 (2)					0.9 (1)		0.8 (1)			
Trp	trace (1)			trace (1)				trace (1)					
Cys(Cm)	1.0 (1)												
yield (%)	10	10	9	11	67	12	11	3.5	68	5.5	20	5.3	1.1
residues	28	8	18	16	7	14	17	41	5	49	16	10	2
position	1-28	29-36	37-54	55-70	71-77	78-91	92-108	92-132	133-137	138-186	158-173	177-186	185-186

[a] Values were obtained by 24-h hydrolysis. [b] This peptide resulted from cleavage at lysine. [c] Analytical values adjusted to give the appropriate integer for this amino acid and the number of other amino acids adjusted accordingly.

Table IV: Amino Acid Composition and Sequences of Peptides Derived from LT8 by Treatment with *S. aureus* Protease

	fraction						
amino acid	12-1	12-2	12-3	12-4	12-5	12-6	12-7
Lys	1.0 (1)	1.0 (1)			3.0 (30)		2.0 (2)
Asp	1.9 (2)	2.0 (2)			1.0 (1)	1.0 (1)	1.0 (1)
Ser						0.9 (1)	
Glu		1.4 (1)	1.0 (1)	3.0 (3)	2.2 (2)	1.1 (1)	4.0 (4)
Pro						1.0 (1)	2.4 (3)
Gly					1.0 (1)		0.9 (1)
Ala						0.9 (1)	
Val		1.4 (1)		1.0 (1)			2.1 (2)
Met				0.8 (1)			
Ile				1.1 (1)	1.9 (2)		
Leu							3.0 (3)
Tyr		1.0 (1)		1.0 (1)	1.0 (1)		1.6 (2)
Phe			0.9 (1)		1.8 (2)	1.8 (2)	
yield (%)	62	17	95	50	100	100	100
peptide	LT8S8	LT8S7,8	LT8S2	LT8S1, LT8S7	LT8S4, LT8S6	LT8S3	LT8S5
residues	3	6	2	7	12	7	18
position	184-186	181-186	142-143	138-141, 181-183	151-154, 173-180	144-150	155-172

was contaminated by LT8 which is rich in Glu, Phe, and Pro. This would account for the high values for these amino acids in the apparent composition of LT6BC. However, sequence analysis on LT6BC gave clear-cut results, so that further purification appeared unnecessary.

Alignment of Peptides. It was possible to align the two cyanogen bromide fragments, CB1 and CB2,3, and the three limited tryptic fragments, LT1, LT2, and LT3, on the basis of the automatic sequence data obtained with the intact carboxymethylated enzyme. The methionine-containing limited tryptic fragments LT3, LT3B, LT6BC, and LT8 provided sequence information for aligning CB4, CB4B, CB4C, CB5, and CB6. The complete sequence of the C-terminal cyanogen bromide fragment, CB6, was confirmed by LT8 and protease peptides derived from LT8 as indicated in Figure 1.

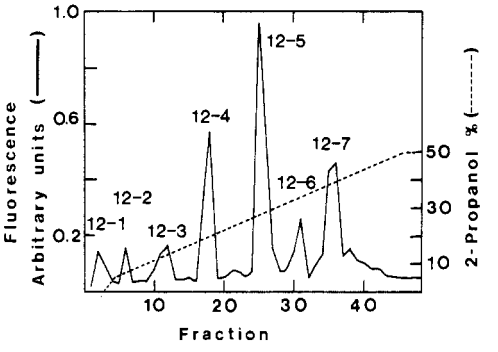

FIGURE 3: Separation by HPLC of peptides produced by digestion of peptide LT8 with *S. aureus* protease. The digest from 318 nmol of LT8 was dried under vacuum and the residue dissolved in 2 mL of 0.1 M pyridinium formate (pH 3.5) and injected onto the μBondapack C_{18} column (3.9 × 300 mm). Flow rate was 0.5 mL/min; fraction volume was 750 μL.

Discussion

Correction of Assignments. The sequence of bovine liver reductase reported here differs from that reported previously (Lai et al., 1979) in the assignments of residues at positions 6, 11, 13, 69, 94, 126, and 170. Assignment of a cysteine residue at position 6 was indicated by data obtained when the

intact protein was subjected to Edman degradation after carboxymethylation with iodo[^{14}C]acetate. The results confirmed the report of Freisheim et al. (1979), who used the same method. Residues at other positions were reinvestigated for acid–amide assignments by HPLC methods. The acid–amide assignments at positions 21, 44, 110, and 123 were also confirmed by HPLC methods. Other tentatively assigned residues in our previous report (Lai et al., 1979) were all reinvestigated and unambiguously identified. These residue identifications include Val-111, Trp-113, Ser-144, Glu-172, Lys-173, and Asn-186.

Sequencing Methodology. Unsuccessful attempts to digest S-carboxymethylated intact enzyme with carboxypeptidases A, B, or Y under unspecified conditions were reported for mouse L1210 reductase (Stone et al., 1979) and for chicken liver reductase (Kumar et al., 1980). Attempted carboxypeptidase digestion of intact carboxymethylated bovine reductase without $NaDodSO_4$ and heat treatment also released no significant amount of amino acid. Presumably, the C-terminal region of vertebrate reductase is not accessible to carboxypeptidase digestion without extensive unfolding of the backbone. However, evidence for the C terminus of the beef liver enzyme was obtained by carboxypeptidase digestion of the carboxymethylated intact protein in the presence of 0.25% $NaDodSO_4$, according to the method of Guidotti (1960). The protein solution was heated at 100 °C in the presence of the $NaDodSO_4$ for 5 min before digestion to further ensure chain unfolding. A nonstoichiometric yield from the carboxypeptidase digestion precluded determination of the number of asparagines released, but the C-terminal sequence of the protein was later determined by manual Edman degradation of peptides formed from LT8 by *S. aureus* protease. The terminal sequence Asn-Asn is also consistent with the isolation of a small peptide (LT8D) obtained by limited tryptic digestion, which gave only aspartate on acid hydrolysis, but eluted much earlier than aspartate in amino acid analysis.

It is clear (Figure 1) that in one experiment in which the carboxymethylated protein was treated with cyanogen bromide cleavages occurred at Trp-57 and at Trp-113, producing fragments CB4B and CB4C, respectively. Cleavage of the polypeptide chain by cyanogen bromide at tryptophanyl bonds has previously been reported (Ozols & Gerard, 1977), but neither the mechanism of such cleavage nor the product of the reaction at tryptophanyl residues is known. In another experiment in which the bovine liver reductase was treated with CNBr, cleavage at these tryptophanyl bonds was not observed, and it was possible to isolate the large fragment CB4 by HPLC. The variation in conditions responsible for the different result in these two experiments was not apparent.

One methionine residue in the beef liver reductase sequence precedes threonine (Met-37, Thr-38). Such a Met–Thr peptide bond has previously been reported to be resistant to CNBr (Schroeder et al., 1969). Similarly, in the beef liver reductase, this bond was cleaved by CNBr to the extent of only about 10%. This was determined from the amount of α-aminobutyric acid recovered after hydrolysis with HI of the anilinothiazolinones obtained when one cycle of Edman degradation was performed on the products of CNBr cleavage. The chemical basis of the resistance of the Met–Thr bond to CNBr is uncertain.

The C-terminal peptide released by CNBr cleavage, CB6, has glutamine at the amino terminus. During preparation and isolation, this glutamine residue cyclized, resulting in a blocked, pyroglutamyl N terminus. The same observation was reported by Stone et al. (1979) for the fragment CB6 derived from the mouse L1210 enzyme. Our simplified procedure for deblocking CB6 with pyroglutamate aminopeptidase produced a deblocked peptide (desGln-CB6) of high purity (above 95%) that could be sequenced almost completely by Edman degradation (44 residues out of 46). In view of the rather frequent formation of blocking pyroglutamate N termini during protein sequence studies, our procedure may have general application.

The results of limited tryptic digestion were complicated by the occurrence of some cleavage at lysine residues, due either to incomplete citraconylation or to hydrolysis of some of the blocking groups during tryptic digestion. There was also some chymotryptic-like cleavage of the enzyme during prolonged digestion with trypsin (see Methods). Such autolytic development of chymotryptic activity during tryptic digestion has been reported previously (Keil-Dlouhá et al., 1971). Both these problems probably resulted from the use of too extended a period of tryptic digestion. This long incubation period was employed because of the incomplete solubility of the substrate. However, some of the unexpected peptides produced were valuable. Thus, isolation and subsequent sequencing of one of the peptides with C-terminal lysine (LT6B) provided essential information for the construction of the complete sequence.

During the course of this study, we have developed methods for reversed-phase HPLC separation of peptides generated from either chemical or enzymic digestion of a protein. The reversed-phase HPLC procedure requires only short operation times (2–4 h) and gives good recovery (45–100%) and good resolution of peptides from mixtures that had proved intractable to separation by classical methods. The method readily yields peptides in quantities and in a state of purity suitable for sequencing or other analytical studies. Since the method is particularly useful for efficient and rapid preparative separation of large peptides, it facilitates protein sequence studies. The isolation of cyanogen bromide peptide CB6, and limited tryptic peptides LT6B and LT8, illustrates the value of such reversed-phase HPLC peptide separation.

The method of identification of PTH-amino acids by HPLC on a short, 5-μm, C_{18} column (see Results) developed during the study is sensitive (less than 100 pmol is necessary) and rapid (35 min between consecutive samples). This method is capable of keeping pace with the output of a Beckman 890C sequencer. While this work was in progress, Bhown et al. (1981) reported similar use of a short, 5-μm, C_{18} column for the identification of PTH-amino acids. In general, the elution profiles for PTH-amino acids, and the resolution and speed, are similar to those in our method, but Bhown et al. did not indicate whether the PTH derivatives of the methyl esters of aspartate and glutamate were also resolved in a single analysis.

Separation of Peptides by HPLC. Although many reports have appeared on the purification of proteins and peptides by HPLC, few have concerned the separation of mixtures of peptides in the course of sequence studies. Reports of the latter kind that have appeared have dealt with the resolution of relatively small, enzymatically generated peptides (Wilson et al., 1979; Gerard & Hugli, 1980; Setlow & Ozols, 1980).

Among several solvents used for HPLC of peptides, aqueous 0.1% phosphoric acid (pH 2.1) was found to be particularly useful, at least for the mixture of peptides produced from the reductase by cyanogen bromide. Although formic acid in concentrations up to 80% has been used (Gerber et al., 1979; Takagaki et al., 1980), this solvent causes serious column instability because of the low pH.

In preliminary studies, many relatively polar solvents, such as acetonitrile, ethanol, and methanol, were tried as eluting

agents. These were found unsatisfactory for elution of hydrophobic peptides, and at higher concentrations (>20%), they caused precipitation of many peptides from solution. On the contrary, the more hydrophobic 2-propanol was efficient at moderate concentrations for peptide elution. Larger peptides required a higher concentration of 2-propanol for their elution than small peptides, so that with the type of gradient used peptides eluted approximately in order of size. However, all of the peptides used in the present work were eluted with 40% propanol. The restriction of the organic solvent to this concentration avoids problems of inorganic salt precipitation which may occur with some of the less efficient organic eluting agents mentioned above. Rubinstein et al. (1979) have reported the use of 1-propanol for peptide elution, but in our experience, better resolution of peptides was obtained with a 2-propanol gradient.

The utility of the pyridine–formate–2-propanol system for separating mixtures of small peptides was illustrated by the separation of the peptides produced by *S. aureus* protease from LT9. Of nine peptides formed, with chain lengths of 2–18 residues, five were obtained pure and the other four in two peaks each containing two peptides (Figures 1, 3, and 9). It is clear that members of the peptide pairs that failed to resolve either had very similar compositions (e.g., LT9S1 and LT9S7) or had compositions which would permit strong interaction (LT9S4 and LT9S6). It is likely that selection of other conditions, for example, a different pH of the aqueous mobile phase or the use of ion pairing reagents, would permit resolution of such mixtures.

Quantitative recovery of peptides, as in the protease digest of LT8 provides a sensitive analytical method for ascertaining the products of chemical or enzymatic cleavage of polypeptides, and for elucidating the specificity of the cleavage procedure. Thus, in the hydrolysis of LT8 by *S. aureus* protease, HPLC clearly demonstrated that there was no cleavage of the bond between Glu-161 and Tyr-162 or of the bond between Glu-171 and Glu-172. Similarly, the isolation of LT8S7,8 in 17% yield (Table IV) indicated that only partial hydrolysis of the bond between Glu-183 and Lys-184 had occurred. Yet complete hydrolysis of two other Glu–Lys bonds (156–155 and 172–173) was evident from the quantitative recovery of the resulting peptides. Similar conclusions about *S. aureus* protease have been reported by Stone et al. (1979), who studied the action of the protease on peptide CB6 from L1210 dihydrofolate reductase, and by Kumar et al. (1980), who observed its action on peptide CB5 from chicken liver dihydrofolate reductase. In these cases, the Glu-161 to Tyr-162 bond was also resistant to the protease digestion. The susceptibility of the glutamyl bond in the sequence Glu-X-Pro to hydrolysis by the protease remains to be determined.

Primary Structure of the Bovine Liver Reductase. The amino acid sequence of beef liver dihydrofolate reductase as shown in Figure 1 contains 186 residues and has a calculated molecular weight of 21 447. This value is in good agreement with the molecular weight of 22 078 ± 303 determined by analytical ultracentrifugation (Peterson et al., 1975). It contains equal numbers of acidic and basic amino acid residues (26 of each) consistent with the reported isoelectric point of 6.8 (Kaufman & Kemerer, 1976). Charged residues constitute 28% of the total and are irregularly distributed, except that no charged residue occurs in the regions 3–17 and 111–121. The first half of the molecule is relatively basic (15 basic residues and 7 acidic), while the second half is relatively acidic (11 basic residues and 19 acidic).

Table V: Functions of Invariant Residues in Dihydrofolate Reductase

residue[a]	reason for conservation	type of interaction
Ala-9	ligand interaction	hydrophobic, with pteridine ring
Ile-16	ligand interaction	hydrophobic, with NMN-ribose, nicotinamide
Gly-17	ligand interaction	hydrophobic, with NMN-ribose
Pro-23	structure maintenance	backbone conformation
Trp-24	ligand interaction	hydrophobic, with nicotinamide
Leu-27	structure maintenance	hydrophobic packing
Phe-34	ligand interaction	hydrophobic, with pteridine ring
Thr-38	structure maintenance	H bond, with Arg-70
Lys-46	structure maintenance	charge interaction, with Asp-110
Gly-53	ligand interaction	hydrophobic, with AMP-ribose
Thr-56	ligand interaction	H bond, with pyrophosphate
Leu-67	ligand interaction	hydrophobic, with *p*-aminobenzoyl
Arg-70	ligand interaction	charge interaction, glutamate
Leu-75	structure maintenance	hydrophobic packing
Gly-116	structure maintenance	steric constraints
Gly-117	ligand interaction	hydrophobic, with pyrophosphate
Thr-136	ligand interaction	H bond, with pteridine 2-amino group
Glu-143	structure maintenance	charge interaction, with Lys-18
Asp-145	structure maintenance	H bond, to Thr-146[b]

[a] Numbering is for vertebrate enzyme as in Figure 4. [b] Except for beef liver, which has Ala-146.

Sequence Comparisons. The beef liver dihydrofolate reductase sequence is highly homologous with that of porcine liver reductase, with that of L1210 murine lymphoma reductase, and, to a lesser extent, with that of chicken liver reductase. There is much less homology with the bacterial dihydrofolate reductases. A comparison of the known sequences of dihydrofolate reductase from various sources is shown in Figure 4. The sequence homology among the vertebrate reductases, among the bacterial reductases, and between these two groups is presented in Figure 5. Among the vertebrate enzymes, 77% of the residues (142 out of 186) are identical in all sequences. Among bacterial sequences, only 13% of the residues (21 out of 160) are identical in all sequences.

Only 10% of the residues (19 out of 186) are identical in all sequences. Of these 19 invariant residues, 11 are involved in interactions with NADPH or methotrexate in the case of enzymes for *Escherichia coli* and *Lactebacillus casei* as indicated by X-ray crystallography (Matthews et al., 1977–1979). Probable functions of these invariant residues which may explain their strict conservation were suggested from consideration of the X-ray crystallographic model of the complex of *L. casei* reductase with methotrexate and NADPH, and are summarized in Table V. Some of the residues not involved in ligand interactions are inserted in closely packed hydrophobic regions (e.g., Leu-27 and Leu-75), and it must be assumed that variants at these positions would have side chains that could not pack in the way needed to maintain the conformation of the backbone, including secondary structure. Other invariant residues (Thr-38, Lys-46, Glu-143, and Asp-145), some at the surface, make charge or H-bond interactions with other side chains that seem to be important for structure maintenance. This is particularly clear for Thr-38 which maintains Arg-70 in the proper position for charge interaction with the α-carboxyl of bound methotrexate. A corresponding interaction presumably occurs also with bound folate derivatives. Variants at Pro-23 would distort the backbone conformation at a critical region on one side of the active site. Substitutions at Gly-116 and Gly-117 would cause

```
                                                                                    •
         5  * * *        15* •    * * *  •25       • • •  *35           45        • *55 •   • •
BL    VRPLNCIVAVSQNMGIGKNGDLPWPPLRNEFQYFQRMTTVSSVEGKQNLVIMGRKTWFSI
PL    VRPLNCIVAVSQNMGIGKNGDLPWPPLRNEYKYFQRMTTTSSVEGKQNLVIMGRKTWFSI
ML    VRPLNCIVAVSQNMGIGKNGDLPWPPLRNEFKYFQRMTTTSSVEGKQNLVIMGRKTWFSI
CL    VRSLNSIVAVCQNMGIGKDGNLPWPPLRNEYKYFQRMTSTSHVEGKQNAVIMGKKTWFSI
ECI     MISLIAALAVDRVIGMENAMPWN LPADLAWFKRNTL     N K  PVIMGRHTWESI
LC       TAFLWAQNRDGLIGKDGHLPWH LPDDLHYFRAQTV      GK  IMVVGRRTYESF
SFII    MFISMWAQDKNGLIGKDGLLPWR LPNDMRFFREHTM     D K  ILVMGRKTYEGM
SFI     MLAAIWAQDENGLIGKEDQLPWR LPNDLKFFKQMTEA    N    TLVMGRKTFEGM
R67     MERSSNEVSNPVAGNFVFPSNATPGMGDRVRKKSGAAWQGQIVGWYCTNLTPEGYAVE
                   *                                                 •
         65 •  •     75 * •       85        * *  95         105          115 • • • •
BL    PEKNRPLKDRINIVLSRELKEPPKGAHFLAKSLDDALELIEDPELTNKVDVVWIVGGSSV
PL    PEKNRPLKDRINIVLSRELKEPPQGAHFLAKSLDDALKLTEQPELKDKVDMVWIVGGSSV
ML    PEKNRPLKDRINIVLSRELKEPPRGAHFLAKSLDDALRLIEQPELASKVDMVWIVGGSSV
CL    PEKNRPLKDRINIVLSRELKEAPKGAHYLSKSLDDALALLDSPELKSKVDMVWIVGGTAV
ECI   G   RPLPGRKNIILSSQPGTD DRV TWVKSVDEAIAACGDVP       EIMVIGGG RV
LC    P K RPLPERTNVVLTHQEDYQAQ GAVVVHDVAAVFAYAKQHLDQ    E LVIAGGAQI
SFII  G K LSLPYRHIIVLTTQKDFKVEKNAEVLHSIDELLAYAKDIPE      D IYVSGGSRI
SFI   G (K) (R)P L(P)
R67   SEAHPGSVQIYPVAALERIN

        125          135*          145*              155            165
BL    YKEAMNKPGHVRLFVTRIMQEFESDAFF PEIDFE   K   YKLLPEYPGVPLDVQEEKG
PL    YKEAMNKPGHIRLFVTRIMKEFESDTFF PEIDLE   K   YKLLSECSGVPSDVQEEKG
ML    YEQAMNEPGHLRLFVTRIMQEFESDTFF PEIDLG   K   YKLLPEYPGVLSEVQEEDG
CL    YKAAMEKPINHRLFVTRILHEFESDTFF PEIDYK   D   FKLLTEYPGVPADIQEEDG
ECI   YEQFLPK  AQKLYLTHIDAQVEGDTHF PDYEPDDWESVFSEFHDA         DAQNS
LC    FTAFKDD V DTLLVTRLAGSFEGDTKMIP LNWDDFTKVSSRTVED          TNPA
SFII  FQALL  PETKIIWRTLIDAEFEGDT FIGEIDFTSFELVEEHEGI      VNQE NQYPH
SFI                                                       (M)DEKNPYAH

     175       185
BL    IKYKFEVYEKNN
PL    IKYKFEVYEKNN
ML    IKYKFEVYEKKD
CL    IQYKFEVYQKSVLAQ
ECI   HSYCFEILERR
LC    LTHTYEVWQKKA
SFII  RFQKWQKMSKVV
SFI      QFETYQRKRK
```

FIGURE 4: Comparison of the sequences of dihydrofolate reductase from a variety of sources: BL, bovine liver; PL, porcine liver (Smith et al., 1979); ML, murine L1210 lymphoma (Stone et al., 1979); CL, chicken liver (Kumar et al., 1980); EC_I, *E. coli* RT500 (Baccanari et al., 1981) and *E. coli* K12 (Smith & Calvo, 1980); LC, *L. casei* (Bitar et al., 1977); SF_{II}, *S. faecium* isoenzyme II (Gleisner et al., 1974); SF_I, partial sequence of *S. faecium* isoenzyme I (S. Pongsamart, P.-H. Lai, K. R. Williams, and R. L. Blakley, unpublished results); R67, plasmid R67 (Stone & Smith, 1979). Residues underlined in the EC_I sequence are reported differently for the *E. coli* MB1428 sequence (Bennett et al., 1978) as follows: Asn-37 → Asp; Asp-87 → Asn; Gln-118 → Glu; Asp-142 → Asn; Glu-154 → Lys. In addition, the *E. coli* K12 sequence also differs from *E. coli* RT500 by the change Gln-118 → Glu. Asterisks indicate residues of the *L. casei* sequence interacting with ligands in the methotrexate–NADPH complex (Matthews et al., 1978, 1979).

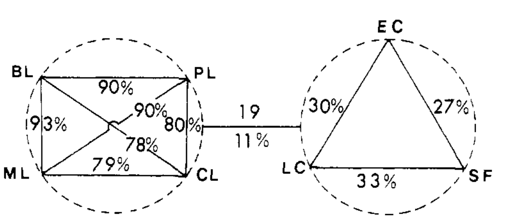

FIGURE 5: Sequence homology among dihydrofolate reductases from various sources. Sources are indicated as follows: EC, *E. coli* RT500; LC, *L. casei*; SF, *S. faecium* isoenzyme II; BL, bovine liver; PL, porcine liver; CL, chicken liver; ML, mouse L1210 lymphoma. Numbers indicate the percent of residues that are identical in the sequences of the reductase from sources connected by solid lines. In enzyme from vertebrate sources, 142 residues (77%) are identical in all four sequences. There are 21 residues (13%) identical in all three bacterial sequences. As shown, 19 residues (11%) are identical in all sequences.

steric hindrance for the binding of folates and NADPH, respectively.

Trimethoprim Selective Toxicity. The examination of the sequence homology of reductase from diverse sources not only provides information about the relative position of species in the evolutionary scheme and about structure–function relationships but also has the potential to elucidate the molecular basis for the specificity of certain ligands for enzymes from particular sources.

Since the crystal structure was solved by X-ray diffraction for the methotrexate complex of *E. coli* dihydrofolate reductase and for the methotrexate–NADPH complex of *L. casei* reductase by Matthews and co-workers (1977–1979), the structure of the reductase binding site for methotrexate and NADPH has been extensively discussed by Hitchings & Smith (1980), Kumar et al. (1980), Baccanari et al. (1981), and Blakley (1981). The preliminary X-ray crystallographic study by Baker et al. (1981) of the *E. coli* reductase complex with trimethoprim has recently provided information about the binding of this inhibitor in the active site also. However, the molecular basis for the selective inhibition of the bacterial reductase by trimethoprim, which permits its clinical use as an antibacterial agent, is still obscure.

When sequences of animal and bacterial dihydrofolate reductases are compared (Figure 4), the residues at positions 28, 58, 68, 78, and 144 are of interest (bovine reductase

Table VI: Amino Acid Residues Invariant within Animal and Bacterial Dihydrofolate Reductases but Differing in the Two Groups of Reductases[a]

sequence position[b]	amino acid residue: all animal reductases	amino acid residue: all bacterial reductases
28	Arg-28	Pro (−25 EC)[c] (−24 LC)
58	Phe-58	Glu (−48 EC) (−47 LC)
68	Lys-68	Pro (−55 EC) (−55 LC)
78	Glu-78	Gln (−65 EC) (−65 LC)
144	Ser-144	Gly (−121 EC) (−124 LC)

[a] The plasmid R67 sequence is not included in this comparison. [b] Numbering as in Figure 4, based on the bovine liver sequence. [c] EC, *E. coli* sequence, with the appropriate number in that sequence. LC, *L. casei* sequence, with the appropriate number in that sequence.

numbering). The differences between the bacterial and animal reductases at these positions are summarized in Table VI. All animal reductases have invariant Arg, Phe, Lys, Glu, and Ser in these positions, respectively, whereas the bacterial reductases (excluding the plasmid reductase) have invariant Pro, Glu, Pro, Gln, and Gly at the corresponding positions, respectively. In the bacterial reductases, Pro-28 and Pro-68 are situated on either side of the active site pocket and at its outer edges. This is indicated by the X-ray crystallographic results of Matthews et al. (1977–1979). The corresponding residues in the vertebrate reductases, Arg-28 and Lys-68, presumably have their positive charges in the vicinity of the edges of the active site cleft also, since in general there is considerable similarity in the folding of the backbones of the bacterial and vertebrate reductases (Dr. D. A. Matthews, unpublished results). Whether this introduction of positive charges is responsible for, or at least contributes to, the weaker binding of trimethoprim to vertebrate reductase cannot be determined without additional crystallographic evidence. It may be noted in this connection, however, that replacement of Leu-31 by Arg in the *E. coli* sequence (numbering as in Figure 4) greatly decreases the affinity of the active site for both trimethoprim and methotrexate (Baccanari et al., 1981). This is attributed to charge interaction between Asp-30 and Arg-31. Such a juxtaposition and interaction of Asp-30 and Arg-31 might also account for the insensitivity of R-plasmid reductase if the alignment of this sequence in Figure 4 is correct. Although there is very little homology between the plasmid reductase sequence and that of other reductases, the alignment shown places the only aspartate in the position of the catalytic residue.

Acknowledgments

We acknowledge with pleasure the technical assistance of Glen Wilson in performing amino acid analyses and Edman degradations. We are grateful to Dr. David A. Matthews for making available the atomic coordinates and ψ–φ angles for the methotrexate–NADPH complex of the *L. casei* reductase.

Supplementary Material Available

Figures 6–16 and Tables VII–XXIII providing the following information: (a) identification of PTH-amino acids by HPLC; (b) details of further separations of peptides by ion-exchange chromatography, gel filtration, and HPLC; and (c) details of sequence determinations by Edman degradation on various peptides, including yield in nanomoles and methods of amino acid identification (31 pages). Ordering information is given on any current masthead page.

References

Baccanari, D. P., Stone, D., & Kuyper, L. (1981) *J. Biol. Chem. 256*, 1738–1747.

Baker, D. J., Beddell, C. R., Champness, J. N., Goodford, P. J., Norrington, F. E. A., Smith, D. R., & Stammers, D. K. (1981) *FEBS Lett. 126*, 49–52.

Bennett, C. D., Rodkey, J. A., Sondey, J. M., & Hirschmann, R. (1978) *Biochemistry 17*, 1328–1337.

Bhown, A. S., Mole, J. E., & Bennett, J. C. (1981) *Anal. Biochem. 110*, 355–359.

Bitar, K. G., Blankenship, D. T., Walsh, K. A., Dunlap, R. B., Reddy, A. V., & Freisheim, J. H. (1977) *FEBS Lett. 80*, 119–122.

Blakley, R. L. (1969) *The Biochemistry of Folic Acid and Related Pteridiness*, pp 139–187, North-Holland Publishing Co., Amsterdam.

Blakley, R. L. (1981) in *Molecular Actions and Targets for Cancer Chemotherapeutic Agents* (Sartorelli, A. C., Lazo, J. S., & Bertino, J. R., Eds.) pp 303–332, Academic Press, New York.

Blaney, J. M., Dietrich, S. W., Reynolds, M. A., & Hansch, C. (1979) *J. Med. Chem. 22*, 614–617.

Böhlen, P., Stein, S., Dairman, W., & Udendriend, S. (1973) *Arch. Biochem. Biophys. 155*, 213–220.

Crestfield, A. M., Moore, S., & Stein, W. H. (1963) *J. Biol. Chem. 238*, 622–627.

Dietrich, S. W., Blaney, J. M., Reynolds, M. A., Jow, P. Y. C., & Hansch, C. (1980) *J. Med. Chem. 23*, 1205–1212.

Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1*, 80–91.

Freisheim, J. H., Kumar, A. A., & Blankenship, D. T. (1979) in *Chemistry and Biology of Pteridines* (Kisliuk, R. L., & Brown, G. M., Eds.) pp 419–429, Elsevier/North-Holland, New York.

Gerard, C., & Hugli, T. E. (1980) *J. Biol. Chem. 255*, 4710–4715.

Gerber, G. E., Anderegg, R. J., Herlihy, W. C., Gray, C. P., Biemann, K., & Khorana, H. G. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 227–231.

Gleisner, J. M., Peterson, D. L., & Blakley, R. L. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71*, 3001–3005.

Gleisner, J. M., Peterson, D. L., & Blakley, R. L. (1975) *J. Biol. Chem. 250*, 4937–4944.

Guidotti, G. (1960) *Biochim. Biophys. Acta 42*, 177–179.

Hansch, C., Dietrich, S. W., & Fukunaga, J. Y. (1981) *J. Med. Chem. 24*, 544–549.

Hitchings, G. H., & Roth, B. (1980) in *Enzyme Inhibitors as Drugs* (Sandler, M., Ed.) Macmillan, London.

Hitchings, G. H., & Smith, S. L. (1980) *Adv. Enzyme Regul. 18*, 349–371.

Hunkapillar, M. W., & Hood, L. E. (1978) *Biochemistry 17*, 2124–2133.

Kaufman, B. T., & Kemerer, V. F. (1976) *Arch. Biochem. Biophys. 172*, 289–300.

Keil-Dlouhâ, V., Sylber, N., Imhoff, J.-M., Tong, N.-T., & Keil, B. (1971) *FEBS Lett. 16*, 291–295.

Kumar, A. A., Blankenship, D. T., Kaufman, B. T., & Freisheim, J. H. (1980) *Biochemistry 19*, 667–678.

Lai, P.-H., Pan, Y.-C., Gleisner, J. M., Peterson, D. L., & Blakley, R. L. (1979) in *Chemistry and Biology of Pteridines* (Kisliuk, R. L., & Brown, G. M., Eds.) pp 437–440, Elsevier/North-Holland Publishing Co., Amsterdam.

Li, R.-L., Dietrich, S. W., & Hansch, C. (1981) *J. Med. Chem. 24*, 538–544.
Matthews, D. A., Alden, R. A., Bolin, J. T., Freer, S. T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M., & Hoogsteen, K. (1977) *Science (Washington, D.C.) 197*, 452–455.
Matthews, D. A., Alden, R. A., Bolin, J. T., Filman, D. J., Freer, S. T., Hamlin, R., Hol, W. G. J., Kisliuk, R. L., Pastore, E. J., Plante, L. T., Xuong, N., & Kraut, J. (1978) *J. Biol. Chem. 253*, 6946–6954.
Matthews, D. A., Alden, R. A., Freer, S. T., Xuong, N., & Kraut, J. (1979) *J. Biol. Chem. 254*, 4144–4151.
Ozols, J., & Gerard, C. (1977) *J. Biol. Chem. 252*, 5986–5989.
Pataki, G. (1966) *J. Chromatogr. 21*, 133–135.
Peterson, D. L., Gleisner, J. M., & Blakley, R. L. (1975) *Biochemistry 14*, 5261–5267.
Podell, D. N., & Abraham, G. N. (1978) *Biochem. Biophys. Res. Commun. 81*, 176–185.
Rubinstein, M., Rubinstein, S., Familletti, P. C., Miller, R. S., Waldman, A. A., & Pestka, C. (1979) *Proc. Natl. Acad. Sci. 76*, 640–644.
Schroeder, W. A., Shelton, J. B., & Shelton, J. R. (1969) *Arch. Biochem. Biophys. 130*, 551–556.
Setlow, P., & Ozols, J. (1980) *J. Biol. Chem. 255*, 10445–10450.
Smith, D. R., & Calvo, J. M. (1980) *Nucleic Acids Res. 8*, 2255–2274.
Smith, S. L., Patrick, P., Stone, D., Phillips, A. W., & Burchall, J. J. (1979) *J. Biol. Chem. 254*, 11475–11484.
Smithies, O., Gibson, D., Fanning, E. M., Goodfliesh, R. M., Gilman, J. G., & Ballantyne, D. L. (1971) *Biochemistry 10*, 4912–4921.
Stone, D., & Smith, S. L. (1979) *J. Biol. Chem. 254*, 10857–10861.
Stone, D., Paterson, S. J., Raper, J. H., & Phillips, A. W. (1979) *J. Biol. Chem. 254*, 480–488.
Takagaki, Y., Gerber, G. E., Nihei, K., & Khorana, H. G. (1980) *J. Biol. Chem. 255*, 1536–1541.
Tarr, G. E. (1977) *Methods Enzymol. 47*, 335–357.
Wilson, J. B., Lam, H., Pravatmuang, P., & Huisman, T. H. J. (1979) *J. Chromatogr. 179*, 271–290.

Complete Amino Acid Sequence of the 4Fe–4S, Thermostable Ferredoxin from *Clostridium thermoaceticum*†

James I. Elliott, Shiow-Shong Yang,‡ Lars G. Ljungdahl,* James Travis, and Christopher F. Reilly

ABSTRACT: The complete amino acid sequence of the 4Fe–4S ferredoxin from the thermophilic bacterium *Clostridium thermoaceticum* has been determined. The protein is extremely thermostable and is the only known clostridial ferredoxin to contain a single [4Fe–4S] cluster. The sequence totals 63 residues and includes the first tryptophan (Trp-26) reported for a clostridial ferredoxin, and other amino acids not commonly found in clostridial or clostridial-like ferredoxins: methionine (Met-1), histidine (His-33), arginine (Arg-49), and leucine (Leu-9, -19, and -31). Sequence homology to clostridial and other 8Fe–8S ferredoxins is limited to eight to nine residues at the amino-terminal sulfhydryl grouping (Cys-10, -13, -16, and -20) and two to five residues in the carboxy-terminal region. This ferredoxin is, thus, sequentially distinct from all known clostridial ferredoxins and from other bacterial ferredoxins in both the 8Fe–8S and 4Fe–4S classes.

Yang et al. (1977) isolated an unusual ferredoxin from the homoacetate-fermenting, thermophilic, anaerobic bacterium *Clostridium thermoaceticum*. In addition to its high thermostability (60 min at 80 °C), iron and sulfur analyses and extrusion studies established only a single [4Fe–4S] cluster present per molecule of protein. This was in direct opposition to the previously coined term "clostridial ferredoxin" which was used to denote a ferredoxin with two [4Fe–4S] clusters. Moreover, the *C. thermoaceticum* ferredoxin contained 63 residues, rather than the 54–56 residues consistently reported for clostridial and related ferredoxins, and included a unique tryptophan residue, an amino acid not previously observed in a clostridial ferredoxin.

Several nonclostridial ferredoxins are now known to contain a single, low-potential [4Fe–4S] cluster: *Bacillus polymyxa* (Yoch & Valentine, 1972), *Bacillus stearothermophilus* (Mullinger et al., 1975), *Desulfovibrio gigas* (LeGall & Dragoni, 1966), *Desulfovibrio desulfuricans* (Zubieta et al., 1973), and *Spirochaeta stenostrepta* (Johnson & Canale-Parola, 1973). Thus, the sequence of the thermostable ferredoxin from *C. thermoaceticum* is of interest for comparison to the ferredoxins from thermophilic and mesophilic clostridia, as well as to the 4Fe–4S ferredoxins from other bacterial systems.

Materials and Methods

C. thermoaceticum was grown at 58 °C according to Yang et al. (1977). Ferredoxin was isolated as described previously (Yang et al., 1977) in yields of 7–9.5 mg of protein/310 g of cell paste.

Iodo[2-^{14}C]acetic acid (40–60 mCi/mmol) was obtained from Amersham, Arlington Heights, IL. Porcine trypsin and yeast carboxypeptidase Y came from Sigma Chemical Co., St. Louis, MO. Hydrazine was purchased from Fisher Sci-

† From the Department of Biochemistry, The University of Georgia, Athens, Georgia 30602. *Received January 4, 1982.* This investigation was supported by U.S. Public Health Service Grant AM-27323 from the National Institute of Arthritis and Metabolic Diseases and U.S. Public Health Service Grant HL-26887 from the National Heart, Lung, and Blood Institute.

‡ Present address: Corn Products, Moffett Technical Center, Summitt Argo, IL 60501.

0006-2960/82/0421-3294$01.25/0