Dihydrofolate Reductase: The Amino Acid Sequence of the Enzyme from a Methotrexate-Resistant Mutant of *Escherichia coli*†

Carl D. Bennett,* John A. Rodkey, John M. Sondey, and Ralph Hirschmann

ABSTRACT: The determination of the amino acid sequence of the enzyme dihydrofolate reductase (5,6,7,8-tetrahydrofolate:NADP$^+$ oxidoreductase, EC 1.5.1.3) from a mutant of *Escherichia coli* B is described. The 159 residues were positioned by automatic Edman degradation of the whole protein, of the reduced and alkylated cyanogen bromide fragments, and of selected tryptic, chymotryptic, and thermolytic digestion products. An *N*-bromosuccinimide produced fragment of the largest cyanogen bromide peptide was also used in the sequence determination.

The enzyme dihydrofolate reductase (5,6,7,8-tetrahydrofolate:NADP$^+$ oxidoreductase, EC 1.5.1.3) which catalyzes the NADPH-dependent reduction of dihydrofolate occupies a key role in DNA synthesis, and thus in cell proliferation. Information about the structure and biophysical parameters of this enzyme from diverse sources will aid in the understanding of its activity and in the design of species specific inhibitors. It may even give some insight into characteristics of rapidly proliferating cells such as those found in cancer and during development. Dihydrofolate reductase is the target enzyme for chemotherapeutic agents such as aminopterin and methotrexate used in the treatment of neoplastic disease (Thomas & Storb, 1971; Condit, 1971) as well as the antibacterial agent trimethoprin (Hitchings & Burchall, 1965).

The dihydrofolate reductase under study in this paper was purified from a methotrexate-resistant mutant of *E. coli* B (strain MB 1428) (Poe et al., 1972) and has been the object of intense and extensive biophysical study (Williams, 1975, and references therein). As isolated, the enzyme binds 1 mol of dihydrofolate and 2 mol of NADPH per molecular weight of 17 958.

It is hoped that further study will correlate the results of the biophysical studies noted in those papers with the primary structure of the enzyme and with the crystallographic studies by Dr. J. Kraut and his associates (Matthews et al., 1978). A preliminary report of this sequence has been presented (Bennett, 1974a) in which homology with the larger dehydrogenases was suggested (see, however, Bennett, 1974b). Subsequent studies using a Beckman sequencer and HPLC[1] quantitation of the polar Pth-amino acids revealed eight corrections in the sequence. Details of the earlier study using manual Edman degradations are presented as confirmation of the sequence determination and of the side chain amide assignments.

Materials and Methods

The enzyme for this study was supplied by Drs. M. Poe and M. Williams, Merck Sharp & Dohme Research Laboratories, Rahway, N.J., and had been purified as described by Poe et al. (1972) as modified by Williams et al. (1973b). Most of the enzyme used for the sequence determination had been recovered after biophysical studies were completed and had been dialyzed against distilled water, heat denatured (90 °C, 2 min), and lyophilized. The two free cysteines of the active enzyme (M. Williams, personal communication) were found to be oxidized to form a disulfide bridge during these operations. The enzyme was shown to be free of carbohydrate by the periodic acid–Schiff reaction as detailed in Product Bulletin PS from Millipore and by the anthrone method of Spiro (1966).

Amino Acid Analysis. Quantitative amino acid analysis was performed on an automated amino acid analyzer (Beckman amino acid analyzer Model 121). For acid hydrolysis, 50–100 nmol of salt-free peptide or protein was dissolved in 1 mL of constant boiling HCl containing 1% phenol. The solution was placed in a small tube, purged of air by evacuation followed by the introduction of nitrogen. This was done three times, and then the tube was sealed under vacuum. Hydrolysis occurred in a constant-temperature forced-air oven (110 ± 0.2 °C) for 20, 40, or 72 h, or in a constant-boiling toluene bath (110.6 °C). After hydrolysis the solution was taken to dryness over NaOH pellets under vacuum and dissolved in pH 2.2 sodium citrate buffer. The sample was analyzed by the 4-h protein hydrolysate program provided by Beckman Instruments; or more recently, by a 2-h single column methodology (Durrum PICO buffer system II) using the Beckman 121 with Durrum DC-6A resin. Serine and threonine values were extrapolated to zero time (usually 10 and 5% destruction, respectively, in 20 h). Cysteine was determined as cysteic acid after performic acid oxidation (Moore, 1963), or as CM-cysteine after derivatization (see below). Tryptophan was determined after hydrolysis in *p*-toluenesulfonic acid using 0.2 M tryptamine to prevent oxidation of the tryptophan (Liu & Chang, 1966) or by hydrolysis in 4 M $Ba(OH)_2$ (Blackburn, 1968). The presence of tryptophan in peptides was determined by spraying a spot containing at least 5 nmol of peptide with the 4-(*N,N*-dimethylamino)cinnaminaldehyde reagent of Harley-Mason and Archer (1958).

Reduction and Alkylation. A 2% solution of the protein in 6 M guanidine hydrochloride, 0.2 M Tris-HCl (pH 8.1) was flushed with nitrogen and 100 μmol of dithiothreitol added per μmol of protein. After incubation for 2 h at 60 °C, the solution was cooled to room temperature and 250 μmol of sodium iodoacetate (iodoacetic acid freshly recrystallized from chloroform and neutralized with NaOH) was added. In one case, 57 μCi of iodo[2-^{14}C]acetic acid (Amersham/Searle) was

† From the Department of Medicinal Chemistry, Merck Sharp & Dohme Research Laboratories, West Point, Pennsylvania 19486. *Received December 29, 1975; revised manuscript received November 29, 1977.*

[1] Abbreviations used: HPLC, high-pressure liquid chromatography; NBS, *N*-bromosuccinimide; CNBr, cyanogen bromide; DHFR, dihydrofolate reductase; Pth, phenylthiohydantoin; Tos-Phe-CH_2Cl, tosyl-L-phenylalanine chloromethyl ketone; CM, carboxymethyl; TAPS, tris(hydroxymethyl)methylaminopropanesulfonic acid; $NaDodSO_4$, sodium dodecyl sulfate; TLC, thin-layer chromatography. The single letter amino acid abbreviations used are those previously suggested [(1968) *Biochemistry 7*, 2703].

0006-2960/78/0417-1328$01.00/0

used. The mixture was kept in the dark for 20 min and then dialyzed against three changes of deionized water. The lyophilized product was the CM-protein.

Cyanogen Bromide Cleavage. Sixteen milligrams of protein (0.64 μmol) was dissolved in 300 μL of 70% formic acid (5% protein solution). To this was added about 30 mg of CNBr. Cellulose acetate electrophoresis of the products (see below) showed that the reaction was essentially complete after 14 h at room temperature. After 14–18 h at room temperature, the reaction mixture was diluted to 3 mL with deionized water and lyophilized.

Electrophoresis. Paper electrophoresis was performed at 4 °C on a noncommercial flat bed unit using 48-cm Whatman 3M paper using 40 to 50 V/cm. Buffers used included pH 6.4 pyridine–acetate (0.83 M), pH 7.3 2,6-lutidine–acetate (0.13 M), pH 4.7 pyridine–acetate (0.33 M), and pH 1.9 formic acid (8%). Peptides were detected using the neutral ninhydrin spray reagent of Easley (1962) followed by the chlorination reagent of Rydon & Smith (1952) using sodium hypochlorite. The mobility of each peptide at pH 6.4 (or 4.7 when histidine was present) relative to free aspartic acid was measured, and the net charge on the peptide was calculated from this mobility and from the known molecular weight (Offord, 1966).

Free flow electrophoresis separations were carried out with the Brinkman free flow apparatus, Model FF-I (Brinkman Instruments, Inc.), using 2,6-lutidine–acetate buffer (pH 7.0), 0.39 M in the electrode rinsing chamber and 0.13 M in the separation chamber (manuscript in preparation). For tryptic peptides, samples of from 1 to 4 μmol were applied in 3 mL of 0.39 M lutidine–acetate with a dosing rate of 1.8 mL per h and a flow rate of the separating buffer of 96 mL per h at 3 °C. The effluent was collected over a range of 48 tubes with acidic peptides migrating toward tube 1. The voltage used and point of application in the electrophoretic field were determined by analysis of the sample on paper electrophoresis in the 0.13 M buffer (separations were usually much better on the free flow instrument than on paper electrophoresis). The separation of tryptic peptides was evaluated using thin-layer silica gel G with BuOH:pyridine:HOAc:H_2O (15:10:3:12) as the solvent and neutral ninhydrin followed by chlorination (*tert*-butyl hypochlorite) sprays for detection.

Cellulose acetate electrophoresis was performed on a Millipore PhoroSlide System (Millipore Corp., Bedford, Mass.), using 1 × 3 in. slides. For CNBr peptides, 0.25 μL was spotted from a solution containing approximately 5 μg of peptide in 5–10 μL of the running buffer (8% formic acid) and electrophoresis was performed for 10 min at 100 V. Slides were immediately immersed in a 0.1% solution of Coomassie Brilliant Blue-R250 in methanol:acetic acid:water (5:1:5) for 2 min and rinsed exhaustively in a water solution containing both 7.5% acetic acid and 5% methanol.

Enzyme Digestion. Tryptic and chymotryptic digestions were carried out in a solution or suspension of 100 μg of peptide per mL of 0.01 M NaTAPS (tris(hydroxymethyl)methylaminopropanesulfonic acid) buffer (pH 9.1) for 4 h at 37 °C. Tos-Phe-CH_2Cl-treated trypsin, Worthington, was dissolved in 0.001 M HCl and stored at 4 °C at least 4 h before use. Chymotrypsin, Worthington, was dissolved in the TAPS buffer immediately before use. The equivalent of 1 μg of trypsin or chymotrypsin was added per 100 μg of peptide to be digested. Digestion was stopped by adding 5 drops of glacial HOAc per mL of solution or by boiling.

Thermolysin (Calbiochem) digestion was carried out in 0.01 M NH_4OAc buffer (pH 8.3) at 37 °C for 18 h with an enzyme to substrate ratio of 1/50. *Staphylococcus aureus,* strain V8, protease (Miles) was used in both ammonium bicarbonate, 0.05 M, pH 7.8, and in phosphate buffer, 0.05 M, pH 7.8, as reported by Houmard & Drapeau (1972).

Carboxypeptidase A and B, Worthington, were used as described by Bennett et al. (1970) in sodium barbital buffer (pH 8.5). For digestion of the whole protein and large peptides, sodium dodecyl sulfate was added to a final concentration of 0.08%.

Dansylation. Equal volumes (usually 100 μL each) of 0.2 M $NaHCO_3$ (pH 9.8) and 0.25% dansyl chloride in acetone were added to 50–100 pmol of peptide. Ten picomoles of an amino acid is easily visible after dansylation and chromatography if the cautions mentioned by Hartley (1970) are observed. However, the lack of quantitative transfer from the hydrolysis tube to the polyamide sheet to avoid excessive transfer of dansyl acid makes 50 pmol a practical lower limit for definitive identification. After incubation at 37 °C for 30 min, the solvents were lyophilized and 200 μL of constant-boiling HCl was added. The tube was sealed under vacuum and heated to 110 °C for 4 h (Gros & Labouesse, 1969). The tubes were opened and the HCl removed under vacuum. When the peptide or protein was insoluble in 0.2 M $NaHCO_3$, 0.1% sodium dodecyl sulfate was included in the 0.2 M $NaHCO_3$ buffer, and, after the 30-min incubation, the dansylated protein was precipitated with excess acetone and washed with acetone three times. Hydrolysis was then performed as above. Chromatography was performed according to the method of Hartley (1970) as detailed by Weiner et al. (1972) on 5 × 5 cm polyamide sheets. However, the unknown dansyl amino acids were spotted from acetone instead of pyridine, and, for solvent IV, 1% NH_4OH in water was used instead of 0.05 M Na_3PO_4 aqueous ethanol. The NH_4OH solvent was easier to prepare and gave a better separation of dansylarginine from α-dansylhistidine and ε-dansyllysine in our hands.

Cleavage at Tryptophan. The *N*-bromosuccinimide method of Ramachandean & Witkop (1967) was used.

Sequence Determination. The manual Edman degradation method of Niall et al. (1969) was used except that the coupling reaction was carried out at 55 °C for 25 min and the cleavage for 5 min at 55 °C. Automated Edman degradations were performed using a Beckman Model 890C sequencer (method of Edman & Begg, 1967). All reagents used were purchased from Beckman Instruments (Spinco Division, Palo Alto, Calif.) and the Beckman Fast Protein Program was used without modification.

Pth-amino acid derivatives generated by either the manual or automatic procedures were identified by a combination of gas and thin-layer chromatography. A Beckman GC-65 gas chromatograph equipped with a flame ionization detector and a 3% SP-400 column (1.3 m) was used to identify all of the Pth-amino acids, except arginine, histidine, aspartic acid, and glutamic acid, and to quantitate the Pths of the nonpolar amino acids. In addition, a 1.3-m column of 1.5% AN-600 (Analabs GP91V) prepared on Chromasorb WHP 100–120 mesh was used to differentiate Pth-leucine from Pth-isoleucine.

Thin-layer chromatography of the nonpolar Pth-amino acid derivatives on nonfluorescent silica gel G (Analtec), developed with heptane:propionic acid:ethylene chloride 58:17:25 (Jeppsson & Sjöquist, 1967, solvent V), was used in conjunction with gas chromatography. The Pths of polar amino acids were separated but not quantitated using solvent H of Edman (1970) (ethylene chloride:acetic acid, 30:7). Plates were placed in a development tank containing solid iodine to locate the Pth-amino acids and the polar residues were further identified by the colors developed by the ninhydrin test of Roseau & Pantel (1969). Pth-histidine and Pth-arginine, remaining in the water layer, were identified either by thin-layer chroma-

TABLE I: Amino Acid Content of Dihydrofolate Reductase and the CNBr Peptides.[a]

	Whole protein		CNBr I	CNBr II	CNBr III	CNBr IV	CNBr V	CNBr VI
Lys	7.1	(7)				2.0	2.0	3.4
His	4.8	(5)					0.8	3.9
Arg	8.7	(9)		1.0		1.1	4.0	3.3
Asx	18.1	(19)		1.1	1.0	4.0	5.3	7.2
Thr	6.2	(6)				1.0	3.0	2.0
Ser	9.2	(9)		1.0			4.0	3.3
Glx	16.0	(15)			1.0	0.2	4.3	10.2
Pro	9.6	(10)				2.6	3.7	2.9
Gly	10.5	(10)		1.0		0.3	5.0	4.2
Ala	13.2	(13)		3.2	1.0	2.3	3.1	4.0
Val	10.7	(11)		1.8		1.1	3.9	3.7
Ile	10.8	(12)		2.7		1.0	3.8	2.8
Leu	10.6	(11)		2.1		3.0	2.0	3.9
Tyr	4.0	(4)						3.9
Phe	6.1	(6)				1.0		5.0
Trp	4.9[b]	(5)				+(2)	+(2)	+(1)
Cys	2.0	(2)					0.7	0.6
Met	4.3	(5)	1.0	0.9	0.9	1.4	0.7	
Total		159	1	15	4	22	50	67
Yield (%)			70	65	90	92	75	62
N terminal	Met			Ile	Glx	Pro	Gly	Val

[a] Values for Val, Ile, and Leu are reported for 70-h hydrolysis. Met values are for homoserine and its lactone for the CNBr peptides, and Cys values are for cysteic acid after performic acid oxidation of the whole protein and for carboxymethyl-Cys for peptides CNBr V and VI. Values in parentheses are those found by sequence. Yield is not corrected for manipulatory losses. [b] After alkaline hydrolysis.

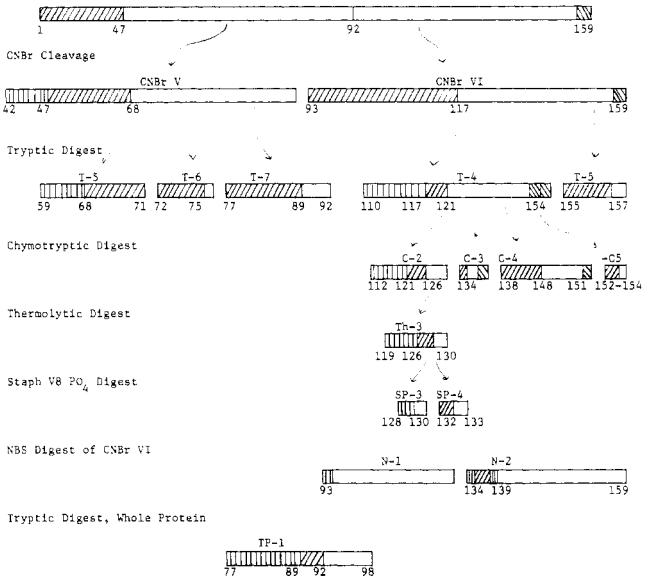

FIGURE 1: Peptides used to determine the sequence of DHFR. (▨) Represents sequences determined by Edman degradation; (▧) areas determined by CPA + B digestion; (▥) sequences known from a previous degradation, i.e., overlaps. Blank areas are characterized by amino acid composition only. See also Figures 2, 6, 10, and 12 for more details.

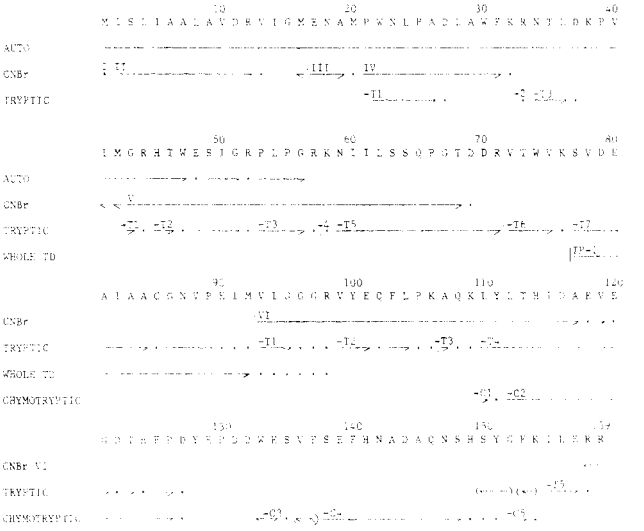

FIGURE 2: A summary of the peptides used in determining the sequence of the first 100 residues of the protein. → are those residues identified by Edman degradation. — are those identified by CPA + B digestion. • were not identified in that particular sequence study. Residues identified in the automatic sequence determination of the whole protein are also shown.

tography in chloroform/methanol/heptafluorobutyric acid (70:30:0.5) or by the spray reagents of Sanger & Tuppy (1951) for Pth-histidine and the phenanthrenequinone reagent of Yamada & Itano (1966) for Pth-arginine. Subsequently, the Pths of the polar amino acids were separated and quantitated by HPLC as described by Rodkey & Bennett (1976).

Nomenclature. The peptides produced by CNBr cleavage are labeled CNBr-I through CNBr-VI. Tryptic peptides are prefixed with T, cymotryptic peptides with C, thermolytic peptides with Th, and Staph V8 protease peptides produced in bicarbonate buffer with S, and in phosphate buffer with SP. Tryptic peptides from the whole protein are prefixed with TP, and NBS oxidation products by N.

Results

The initial Edman degradation established the position of 54 of the first 56 residues and the complete sequence of the first four CNBr peptides. CNBr V and VI were then isolated and their sequences established from tryptic, chymotryptic, thermolytic, and Staph V8 proteolytic peptides. In one case, the overlap peptide between CNBr V and VI was isolated from a tryptic digest of the whole protein. NBS oxidation was used to establish the overlap of CNBr VI-T4-C3 with -C4. These sequence determinations are depicted in Figures 1 and 2.

TABLE III: Sequence Determination of 300 Nanomoles of the CM-protein.[a]

Residue	Amino acid	GC	HPLC
1	Met	290	
2	Ile	252	
3	Ser	264	
4	Leu	248	
5	Ile	217	
6	Ala	234	
7	Ala	226	
8	Leu	239	
9	Ala	217	
10	Val	200	
11	Asp		196
12	Arg		140
13	Val	180	
14	Ile	163	
15	Gly	168	
16	Met	180	
17	Glu		156
18	Asn		111
19	Ala	132	
20	Met	130	
21	Pro	116	
22	Trp	125	
23	Asn		115
24	Leu	107	
25	Pro	113	
26	Ala	94	
27	Asp		68
28	Leu	89	
29	Ala	77	
30	Trp	80	
31	Phe	47	
32	Lys		20
33	Arg		29
34	Asn		20
35	Thr	67	
36	Leu	61	
37	Asp		9
38	Lys		23
39	Pro	46	
40	Val	39	
41	Ile	43	
42	Met	20	
43	Gly	14	
44	Arg		6
45	His		4
46	Thr	15	
47	Trp	14	
48	Glu	*b*	
49	Ser	7.8	
50	Ile	9.9	
51	Gly	7.0	
52	Arg	*b*	
53	Pro	6.3	
54	Leu	7.6	
55	Pro	7.8	
56	Gly	3.1	

[a] Ile/Leu identification was made by AN-600; Pro/Thr, by HPLC. Yields are nanomoles of Pth-amino acid above background. [b] Assigned from the sequence determination of CNBr V.

Whole Protein. The amino acid analysis of the whole protein agreed well with the sequence subsequently determined (Table I), except for the values for Asp, Glu, and Ile. Even after hydrolysis for 70 h, the Ile value was 10% below the expected. The Val-Ile sequence (residues 40–41) and the Ile-Ile-Leu sequence (residues 60–62) were found to be particularly resistant to hydrolysis, even in the isolated tryptic peptides. Digestion of

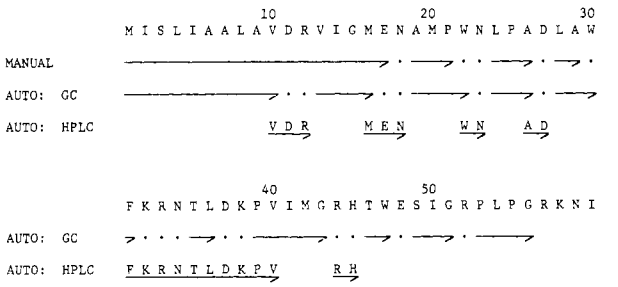

FIGURE 3: The amino-terminal sequence of *E. coli* DHFR, determined by the manual and automatic Edman procedure using GC, TLC, and HPLC for Pth identification.

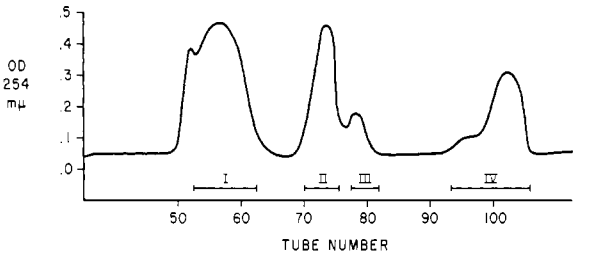

FIGURE 4: Gel filtration of the peptides produced by CNBr digestion of DHFR. The column (0.9 × 100 cm) was eluted at a rate of 10 mL/h with 50% HOAc.

the CM-protein with CPA + B for 24 h released only 1.8 mol of arginine per mol of protein. Dansylation of the CM-protein in $NaDodSO_4$ produced dansylmethionine and dansyl-Met-Ile after 4 h of acid hydrolysis. The protein was shown to be carbohydrate free by the periodic acid–Schiff stain after cellulose acetate electrophoresis.

Edman degradations of the CM-protein (Figure 3) were performed both manually (Table II; supplementary material; see paragraph at end of paper) and automatically (Table III). The repetitive yield in the sequencer was 96% through step 25, and then dropped off, probably due to extraction of the shortened protein at succeeding steps. Identification of the polar Pth-amino acids by HPLC helped to extend the sequence to the 56th residue. This method identified the sequence Asp-Lys at residues 37–38 which had been reversed previously (Bennett, 1974a). The initial sequence determination established the sequence of the first four CNBr peptides and 12 of the first 14 residues of CNBr V. Additional confirmatory sequence determinations of CNBr I through IV are presented at the end of this section and as supplementary material (Figures 8 and 9 and Tables II, VI, VIII, XI, XII, XIV, and XV).

Since this initial degradation provided the position of five of the six peptides expected from CNBr cleavage, the order of the CNBr peptides in the protein was established.

CNBr Cleavage. Cleavage of the CM-protein with CNBr produced four peptides identifiable by cellulose acetate electrophoresis [CNBr I (homoserine) and III (a tetrapeptide) were too small to be retained on the plate during staining]. Dansylation of the CNBr digest identified a new end group, valine, in addition to the end groups expected from the known sequence (Ile_2, Glx_{17}, Pro_{21}, and Gly_{43}). The valine end group, therefore, must have been derived from CNBr VI (Table I).

When the CNBr digestion products were subjected to gel filtration on Sephadex G-50 in 50% HOAc (Figure 4), the first peak contained peptides CNBr V and VI plus small amounts of uncleaved protein. The other three peaks contained CNBr IV, II, and III plus I, respectively.

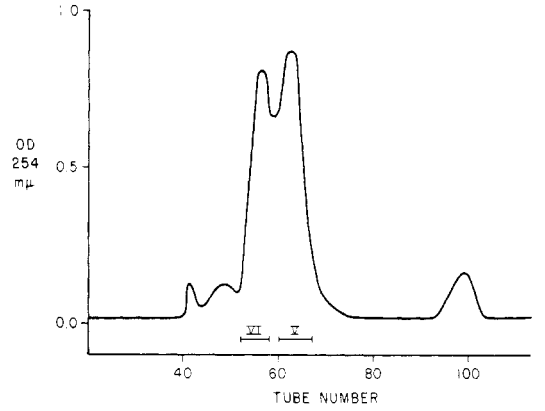

FIGURE 5: Gel filtration of peak I (Figure 4) through Sephadex G-50 (1.2 × 60 cm) in 0.02 M NH_4OAc (pH 8.3). Fractions containing CNBr VI and V are indicated.

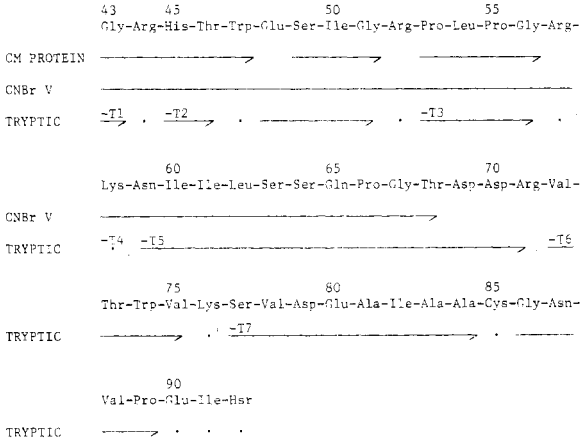

FIGURE 6: The sequence determination of CNBr V. Tryptic peptides were isolated by free flow electrophoresis (see text). The sequence of CNBr V-T7 was verified and extended by the sequence determination of peptide TP-1 (residues 77–98, Table VII).

TABLE IV: Sequence Determination of 150 Nanomoles of CNBr V (Residues 43–92).[a]

Residue	Amino acid	GC	HPLC
43	Gly	120	
44	Arg		90
45	His		80
46	Thr	20	
47	Trp	60	
48	Glu		65
49	Ser	23	
50	Ile	51	
51	Gly	49	
52	Arg		30
53	Pro	45	
54	Leu	39	
55	Pro	51	
56	Gly	44	
57	Arg		15
58	Lys		30
59	Asn		20
60	Ile	23	
61	Ile	26	
62	Leu	25	
63	Ser	4	
64	Ser	4	
65	Gln		15
66	Pro	30	
67	Gly	20	
68	Thr	10	

[a] See Table III for additional information.

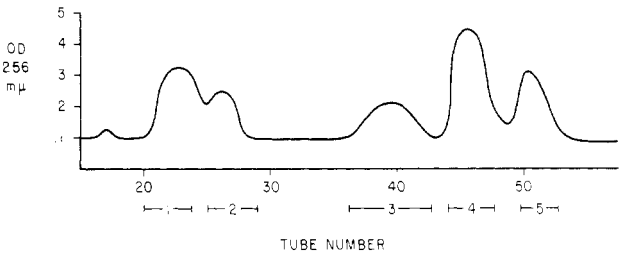

FIGURE 7: Gel filtration of the peptides from a tryptic digest of CNBr V through a Sephadex G-25 column (0.9 × 100 cm) with 0.02 M NH_4HCO_3 (pH 9.1). Fraction 1 contained -T7, 2 contained -T5, and 3 contained -T1, -T3, and -T4. Fractions 4 and 5 were retarded beyond the salt peak and contained -T2 and -T6, respectively.

After lyophilization, peak I, Figure 4, was dissolved to make a 0.25% solution in 0.5 M NH_4OAc and the pH adjusted to 9.0 with 1.0 M ammonium hydroxide solution. A small amount of flocculent precipitate was removed by centrifugation and the supernatent was gel filtered on Sephadex G-50 in 0.02 M NH_4OAc, pH 8.3 (Figure 5). The small initial peaks contained uncleaved protein and aggregated material. The second peak contained CNBr VI and the third peak, CNBr V.

Amino acid analysis and dansyl end groups of each of the isolated CNBr peptides (Table I) agreed with the subsequent sequence determinations, except for Ile values in CNBr V, which were 25% low even after 70-h acid hydrolysis. The separation of the CNBr peptides was monitored by cellulose acetate electrophoresis at each step.

CNBr V (Figure 6). The sequence determination of the whole protein had established 12 of the first 14 residues of this peptide. CPA + B digestion released no amino acids after 24 h. Automatic Edman degradation of this peptide (Table IV and Figure 6) established the sequence of residues 43 through 68, and determined the order of all of the tryptic peptides from this CNBr fragment, since T7 would contain the C-terminal homoserine and T6 could be placed by difference.

CNBr V-T5, -T6, and -T7 were isolated from a tryptic digest of CNBr V (Figure 7) and by free flow electrophoresis of a tryptic digest of CNBr V + VI (Figures 8 and 9; supplementary material). Amino acid analysis (Table V) showed -T7 to contain homoserine and thus was the C-terminal tryptic peptide of CNBr V. The manual Edman degradation of -T5 showed it to contain the N-terminal sequence of -T5, found in the degradation of CNBr V (residues 59–68). The degradation of this peptide was remarkable in that it proceeded with very little extraction of the residual peptide at each step (Table VI; supplementary material). A 10% repetitive yield was found during the manual Edman degradation of CNBr V-T6, showing a sequence of Val-Thr-Trp-(Val,Lys), so that the dansyl-Edman method of Gray & Smith (1970) was used on 25 nmol of this peptide to establish the complete sequence.

From the tryptic digestion of the whole protein (see below), the peptide containing residues 77–98 was isolated and its sequence determined through residue 92 (Table VII). This placed all of the residues in CNBr V and provided additional evidence for the overlap between CNBr V and CNBr VI.

CNBr VI (Figure 10). Automatic Edman degradation of CNBr VI proceeded through residue 117 and identified residue 119 as Val. Tryptic digestion of CNBr VI for 4 h produced five

TABLE V: Amino Acid Analyses (20 h) of the Tryptic Peptides Isolated from CNBr V and the Whole Protein.

	CNBr V				
	-T5	-T6	-T7	TP-1	TP-3
Lys		1.0			0.9
Arg	1.1				1.0
CM-Cys			0.9	0.9	
Asx	3.0		1.6	1.4	2.1
Thr	1.0	1.0			1.0
Ser	2.1		0.6	0.8	
Glx	1.0		2.0	2.0	
Pro	1.0		1.0	1.0	0.9
Gly	1.1		1.1	4.2	1.1
Ala			2.9	3.0	
Val		1.9	2.0	2.9	0.6
Met			0.9[a]	0.8	0.9
Ile	1.6		1.0	1.9 (3)	0.6
Leu	0.6				1.0
Trp		+			
N terminal	Asx	Val	Ser	Ser	Asx
Residues	59-71	72-76	77-92	77-98	34-44
Yield (%)	82	67	85	73	61

[a] Homoserine.

TABLE VII: Edman Degradation of 50 Nanomoles of Peptide TP-1 (Residues 77-98).

Residue	Amino acid	GC	HPLC
77	Ser		30
78	Val	36	
79	Asp		22
80	Glu		30
81	Ala	46	
82	Ile	29	
83	Ala	37	
84	Ala	35	
85	CM-Cys		25
86	Gly	18	
87	Asp		15
88	Val	11	
89	Pro	4.5	
90	Glu		3.3
91	Ile	2.1	
92	Met	2.0	

peptides. CNBr VI-T4 precipitated during the digestion and was isolated by centrifugation. The other four peptides were separated by gel filtration (Figure 11). The sequences of CNBr VI-T1 (peak II), -T2 (peak I), and -T3 (peak IV) were confirmed and the sequence of -T5 (peak III) was established by manual Edman degradation, amino acid analysis, and tryptic specificity (Figure 10). CNBr VI-T5 was placed at the carboxy terminal of the CNBr peptide by the fact that it contained the arginine known to be the carboxy-terminal residue of the protein (the other arginine of CNBr VI was known to be residue 98) and by the fact that CNBr VI-T1 through -T4 were positioned by the initial sequence determination of CNBr VI.

CNBr VI-T4 (Figure 12). This 45 residue peptide was isolated from tryptic digests of CNBr VI, CNBr V + VI, and the CM-protein. Edman degradation identified residues 110–121 and 123, 125, and 126 (Figure 12) (Table VIII; supplementary material). Digestion of -T4 with CPA + B for 4 and 24 h suggested a carboxy-terminal sequence of (Ser,Tyr,Cys)-(Phe,Lys) (Table IX). A chymotryptic digestion of 1.1 μmol of this peptide produced five peptides which were isolated by

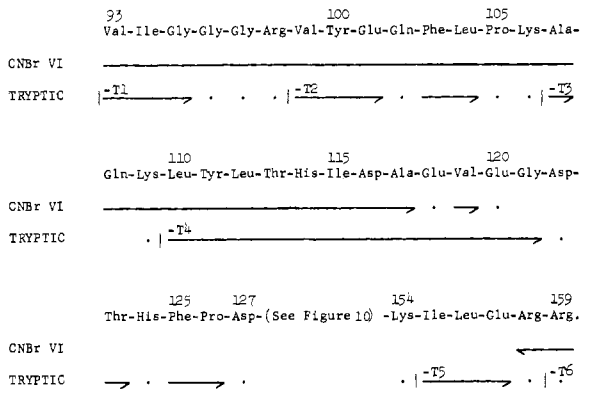

FIGURE 10: The partial sequence determination of CNBr VI. See Figure 12 for the sequence determination of residues 110–154.

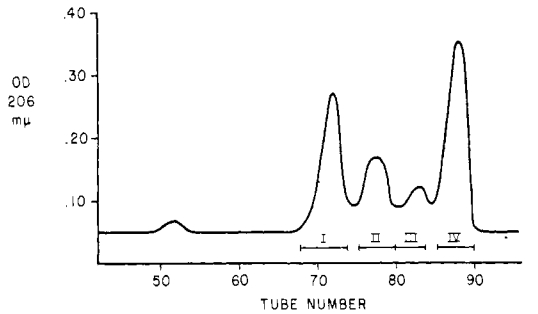

FIGURE 11: Gel filtration of the soluble peptides from a tryptic digest of CNBr VI through a Bio-Gel P-4 column (0.9 × 100 cm) in 0.02 M TAPS buffer (pH 9.1).

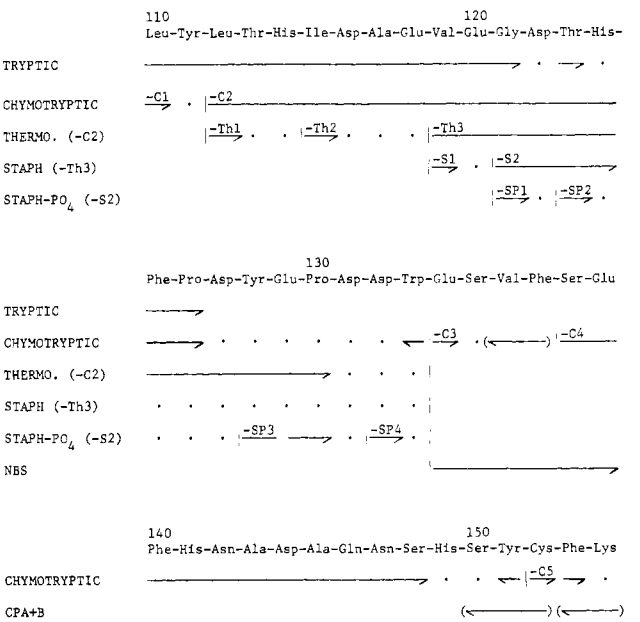

FIGURE 12: The sequence determination of CNBr VI-T4. The chymotryptic peptide -C2 was digested by thermolysin to produce -Th3, which was digested with staphylococcal protease in bicarbonate buffer to produce -S2 which was digested with the same protease in phosphate buffer to produce -SP1 through -SP4.

gel filtration through Sephadex G-25 (Figure 13 and Table X). CNBr VI-T4-C1 and -C5 (Figure 12) were further purified by gel filtration through Sephadex G-10. The position of -C1 and -C2 was known from the initial sequence determination of -T4. The tripeptide, -C5, was carboxy terminal in the tryptic peptide because it contained the lysine known to be the carboxy

TABLE IX: Moles of Amino Acid Released per Mole of CNBr VI-T4 Digested by CPA+B.

	4 h	24 h
Lys	0.92	1.05
Phe	0.91	0.98
CM-Cys	0	0.58
Tyr	0	0.60
Ser	0	0.61

terminal of the peptide. The lack of tyrosine in -C3 and its presence in -C4 suggested the sequence -C3 followed by -C4, since serine and tyrosine had been released by CPA + B digestion of CNBr VI-T4. This order of peptides was confirmed by NBS oxidation of CNBr VI (see below). Manual Edman degradation confirmed the sequence of -C5, established the sequence of residues 138 through 148 of -C4 and residues 112 through 126 of -C2 (Figure 12). The tetrapeptide, -C3, could not be sequenced using the manual method (the residual peptide was extracted completely, along with Pth-Glu, at the first step); but CPA digestion for 4 h of this peptide released Val and Phe. Its sequence was later established from the sequence determination of the NBS oxidation product derived from this area (see below). CPA digestion of -C2 produced only Trp which was thus the C-terminal amino acid of -C2.

CNBr VI-T4-C2 (Figure 10). A thermolysin digest of 600 nmol of this 21 residue peptide produced three peptides which were separated on Bio-Gel P-4 (Figure 14). The sequences of -Th1 and -Th2 were known, and manual Edman degradation of -Th3 positioned residues 119 through 130. Staphylococcal protease digestion of 200 nmol of -Th3 in 0.05 M NH_4HCO_3 buffer (pH 7.8) produced only two peptides, -S1 (Val-Glu, residues 119–120) and -S2 (residues 121–133) which were separated on Bio-Gel P-2. When 120 nmol of CNBr VI-T4-C2-Th3-S2 was digested with the same enzyme in phosphate buffer (pH 7.8), four peptides could be separated by paper chromatography. Amino acid analysis and dansylation identified them as -SP1 (residues 121–122), -SP2 (residues 123–127), -SP3 (residues 128–131), and -SP4 (residues 132–133). Manual Edman degradation of -SP3 confirmed the sequence assigned to -Th3 (Figure 12). Since residue 133 was known to be Trp, 132 must be Asp, based on dansylation, amino acid analysis of -SP4, and electrophoretic mobility.

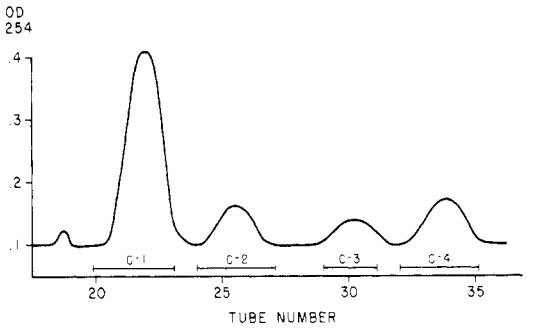

FIGURE 13: Gel filtration of the chymotryptic digest of CNBr VI-T4 through a Sephadex G-25 column (0.9 × 100 cm) with 0.02 M NH_4HCO_3 (pH 9.1) at a rate of 6 mL/h.

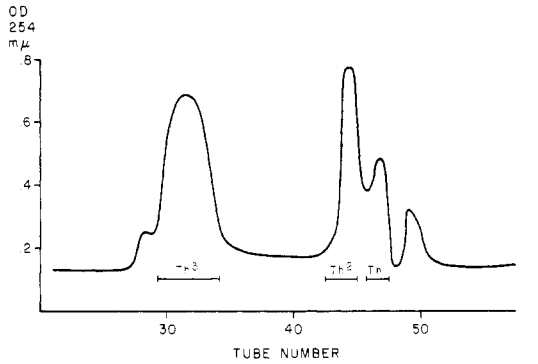

FIGURE 14: Gel filtration of the thermolytic digest of CNBr VI-T4-C2 through a Bio-Gel P-4 column (0.9 × 85 cm) with 0.01 M TAPS buffer (pH 9.1).

NBS Oxidation of CNBr VI. Ninety nanomoles of [^{14}C]-CM-CNBr VI was treated for 0.5 h with 180 nmol of NBS in 0.2 mL of 50% HOAc. The digest, after lyophilization, was gel

TABLE X: Amino Acid Analyses (20 h) of the Peptide Isolated from Digests of CNBr VI.

	-T4	-T4 chymotryptic digest (Figure 13)				-T4-C2-			-T5	N-2
		Peak I	Peak II	Peak III	Peak IV	Th3	SP3	SP4		
Lys	0.9				1.1					1.2
His	3.4	2.1	1.7			0.9				1.9
Arg									1.0	2.1
CM-Cys	0.9				0.7					0.9
Asx	7.2	4.8	3.2			4.2	1.0	1.1		2.9
Thr	1.9	1.9				0.9				
Ser	3.8		3.0	1.2						3.8
Glx	5.7	3.1	2.0	1.1	0.4	2.1	1.0		1.0	4.1
Pro	2.2	1.7				2.0	1.0			
Gly	1.2	1.2				1.1				
Ala	3.2	1.2	2.1							2.0
Val	2.1	1.1		0.9		0.9				1.1
Ile	1.1	1.0							0.9	1.0
Leu	2.2	0.9			1.0				0.9	1.0
Tyr	2.7	0.9	0.8		1.3	0.9	0.9			0.8
Phe	3.8	1.1	0.9	0.8	0.8	1.0				2.6
Trp	+	+				0.8		0.9		
N terminal	Leu	Leu	Ser	Glx	Leu + CM-Cys	Val	Tyr	Asp	Ile	Glx
Residues	110–154	112–133	138–151	134–137	110–111 + 152–154	119–133	128–131	132–133	155–158	134–159
Yield (%)	98	91	70	30	96 + 72				75	22

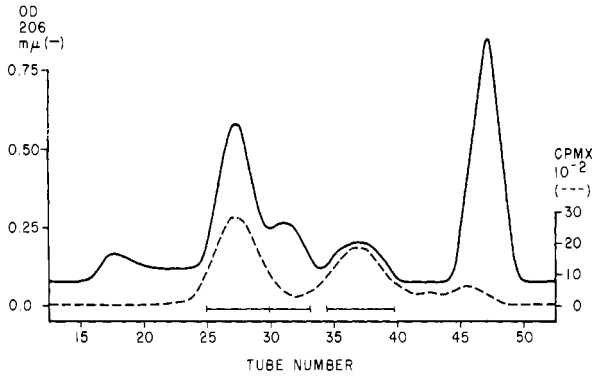

FIGURE 15: Gel filtration of the NBS oxidation products of CNBr VI through a column (1.5 × 50 cm) of Sephadex G-50 in 0.02 M NH_4OAc (pH 8.3). The cysteine (residue 152) was labeled with [^{14}C]iodoacetic acid.

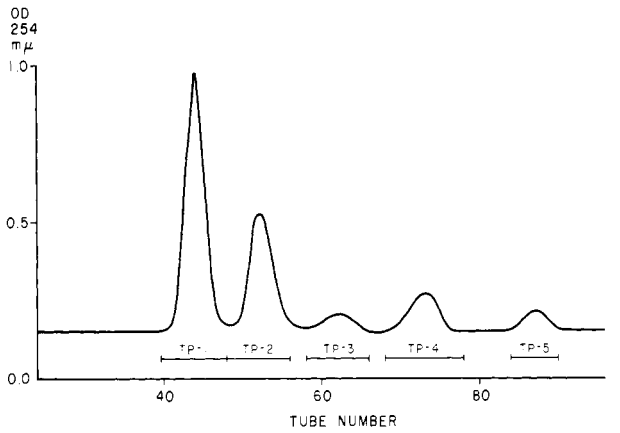

FIGURE 16: Gel filtration of the soluble tryptic peptides from the CM-protein through a Bio-Gel P-4 column (0.9 × 100 cm) eluted with 0.01 M TAPS buffer (pH 9.1).

filtered through Sephadex G-50. The first peak (Figure 15) was uncleaved starting material, while the second peak had the amino acid composition of residues 93–133, and an amino terminal valine. The third peak had the amino acid composition of residues 134–159 (Table X), including the [^{14}C]CM-cysteine (residue 152), and a Glu or Gln amino terminal. Six steps of manual Edman degradation established the overlap sequence, residues 134 through 139, between CNBr VI-T4-C3 and -C4.

Tryptic Digestion of the CM-Protein. The CM-protein (500 nmol) was digested for 4 h with trypsin. During this time a precipitate formed and was removed by centrifugation at the end of the digestion. Based on amino acid analysis and N-terminal identification, it contained residues 1–12, 13–32, and 110–154. The supernatant was gel filtered through a Bio-Gel P-4 column. The initial peak (Figure 16) contained a pure peptide, TP-1, residues 77–98 (Table V). Manual Edman degradation of 200 nmol of this peptide established the position of residues 77–92 (Table VII). Based on this sequence determination and the amino acid composition of the peptide, the overlap of CNBr V and VI was also confirmed. The third peak (Figure 16) contained pure residues 34–44 (Table V), confirming the overlap between CNBr IV and V.

CNBr I through IV. Confirmatory sequence determinations were performed on these peptides (Table VI (supplementary material), Table I, and Figure 2). CNBr I and II were isolated from peak IV, Figure 4, by gel filtration on G-10. Residues 17–19 were confirmed by manual Edman degradation. Paper electrophoresis (pH 6.4) at the first two steps confirmed the amide assignment (Glu-Asn) for residues 17–18.

The first 11 residues of CNBr II were confirmed by Edman degradation, but at step 11 the residual peptide was completely extracted into the EtAc layer, apparently due to the loss of the charged arginine (residue 12) from the peptide (Table XI; supplementary material).

CNBr II was extremely insoluble at neutral pH, so that the small amount which was present in CNBr IV was removed by centrifugation in 0.01 M NH_4HCO_3. After 2 h of tryptic digestion of CNBr IV, two peptides and free arginine were detected on thin-layer cellulose chromatographed with the upper phase of BuOH:HOAc:H_2O (4:1:5). No cleavage occurred at the Lys-Pro bond (residues 38–39). The tryptic digest was then distributed between the two phases of BuOH:HOAc:H_2O (4:1:5). As suggested by its R_f of 0.9 on thin-layer cellulose, CNBr IV-T1 was extracted quantitatively into the upper phase (Table XII; supplementary material). The lower phase contained free arginine (residue 33) and -T3 (residues 34–42), which were subsequently separated on Sephadex G-10. Six and three steps of manual Edman degradation identified -T1 and -T3, respectively (Figure 2).

Discussion

The structure of dihydrofolate reductase is of interest not only because of its key role in DNA synthesis, but also because it is a small protein (159 residues) which binds a relatively large substrate (dihydrofolate) and 2 mol of cofactor (NADP). The three-dimensional structure, which is being determined by x-ray crystallography (Matthews et al., 1978) should help explain these multiple binding sites in this small protein.

The extended Edman degradation which was made possible by quantitation of all of the Pth-amino acids using GC and HPLC placed 94 of the 159 residues of the protein. This was done using the whole protein and the largest two CNBr peptides. The Edman degradation of the whole protein also aligned all of the CNBr peptides and determined the complete sequence of four of them. Tryptic digestion provided peptides whose sequence could be determined entirely by the manual Edman degradation except for CNBr VI-T4 (residues 110–154). This 45-residue peptide was digested with chymotrypsin, thermolysin, and staphylococcal protease in order to completely establish its sequence. In addition, cleavage of CNBr VI by NBS oxidation of its only tryptophan produced evidence for the ordering of the chymotryptic peptides of CNBr VI-T4.

The enzymes used generally showed their expected specificities, including the cleavage by trypsin of an Arg-Pro sequence (residues 52–53) but not a Lys-Pro sequence (38–39). No unexpected cleavages occurred; but, disappointingly, several bonds failed to cleave in the large acidic peptide CNBr VI-T4. Under our conditions: (1) chymotrypsin failed to cleave Asp-Tyr-Glu (127–129) and Glu-Phe-His (139–141); (2) thermolysin failed to cleave Asp-Ala-Glu (116–118), although it did cleave Glu-Val-Glu (118–120), Staphylococcal protease in bicarbonate buffer cleaved Glu-Gly-Asp (120–122) but not Glu-Pro-Asp (129–131). This last sequence was not even cleaved by staphylococcal protease in phosphate buffer. CPA failed to cleave the C-terminal homoserine and Ile from CNBr V even after 24 h.

Asp_{131} and His_{149} were positioned by the amino acid composition of SP-3 and -C4, respectively.

The method of Offord for determining the charge of the peptides worked well at pH 6.4 and pH 4.7. In one case, however, there was serious tailing (residues 93–98). Fortunately, the only charged residue this peptide contained was arginine,

TABLE XIII.

Bennett (1974)		Present study	
Residue	Reported	Residue	Found
37–38	Lys-Asp	37–38	Asp-Lys
57	Ser	57	Arg
72	Arg	72	Val
88	Gln	87	Asn
90a,b	Not found	90–91	Glu-Ile
119	Asp	120	Glu
126–130	Asn-Glu-Tyr-Pro-Glu	127–132	Asp-Tyr-Glu-Pro-Asp-Asp
147a	Not found	150	Ser

so that its mobility was immaterial to its structure determination. More serious, however, was the low mobility of the two large, acidic peptides (77–92 and 112–133). The spots were symmetrical and showed no signs of tailing. However, the identification of the aspartic and glutamic acid residues during the sequence determination of these peptides was positive (Figures 6 and 12).

Initially (Bennett, 1974a), the sequence of this protein was established by the manual Edman degradation of CNBr and tryptic peptides using gas chromatography and TLC to identify the Pths released. The tryptic peptides were produced from a mixture of CNBr V and VI, since no satisfactory separation could be found. Subsequent studies, undertaken to obtain additional confirmatory experimental data, were carried out using the Beckman sequencer and HPLC quantitation of polar Pth-amino acids. These studies focused also on those parts of the molecule in which our structural assignments differed from the sequence of the enzyme from a trimethopterin resistant strain of *E. coli* which was elucidated more recently by Stone et al. (1977). Results of x-ray crystallographic studies (Matthews et al., 1978) led us to reexamine and subsequently revise the original assignment at positions 57 and 72.

These further studies have shown that eight corrections are necessary to the sequence previously published (Table XIII). The justification for these corrections is detailed in the Results section. Most of these changes can be attributed to amino acid analyses using very small amounts of material, difficulty in TLC identification of the Pth-amino acids of Ser, Arg, Asp, and Glu and by positioning amino acid residues by difference or inference.

This dihydrofolate reductase from a mutant of *E. coli* is homologous to the enzyme isolated from a mutant of *Streptococcus faecium* (as pointed out by Gleisner et al., 1974), to the first 51 residues of the enzyme isolated from a methotrexate-resistant strain of *Lactobacillus casei* (Batley & Morris, 1977) and to the enzyme isolated from a methotrexate-resistant line of the mouse lymphoma L1210 (Stone & Phillips, 1977).

Rossman et al. (1974) have pointed out homologous NAD^+ binding areas in glyceraldehyde-3-phosphate, lactate, and alcohol dehydrogenases based on their three-dimensional structures. These areas contain 130 to 150 amino acid residues. *E. coli* DHFR binds 2 mol of NADPH per mol of enzyme (Williams et al., 1973a) with fewer than 159 residues, since some amino acids must be involved in the binding of the dihydrofolate substrate. The amino acid sequence is apparently not homologous to the dehydrogenases (Bennett, 1974b) and the X-ray crystal structure will be needed to determine how the tertiary structure of these two cofactor binding sites is related to the well-studied binding site of the dehydrogenases.

The sequence of the DHFR of a trimethoprim resistant strain of *E. coli* reported earlier this year (Stone et al., 1977) differs from the present sequence at three positions: 118 Gln; 142 Asp; and 154 Glu. These changes can all be related by single base changes in the DNA coding for this protein. It can be suggested that the change from 154 Lys to 154 Glu could explain the difficulty in alkylating Cys-152 in the trimethoprim resistant strain.

Acknowledgment

We wish to thank Drs. M. Poe and M. Williams for the pure dihydrofolate reductase, Messrs. Milton Davis and Carl Homnick for amino acid analyses, and Drs. Hugh Niall and Karst Hoogsteen for helpful discussions of this work. We are indebted to Dr. J. Kraut for communicating to us the results of his x-ray crystallographic studies which led us to reexamine the assignments in positions 57 and 72.

Supplementary Material Available

Figures 8 and 9 and Tables II, VI, VIII, XI, XII, XIV, and XV (5 pages). Ordering information is given on any current masthead pages.

References

Batley, K. E., & Morris, H. R. (1977) *Biochem. Biophys. Res. Commun. 75,* 1010.
Bennett, C. D. (1974a) *Nature (London) 248,* 67.
Bennett, C. D. (1974b) *Nature (London) 252,* 511.
Bennett, C. D., Konigsberg, W. H., & Edelman, G. M. (1970) *Biochemistry 9,* 3181.
Blackburn, S. (1968) *Amino Acid Determination,* p 22, Marcel Dekker, New York, N.Y.
Condit, P. T. (1971) *Ann. N.Y. Acad. Sci. 186,* 475.
Easley, C. W. (1965) *Biochim. Biophys. Acta 107,* 386.
Edman, P. (1970) in *Protein Sequence Determination* (Needleman, S., Ed.) p 211, Springer-Verlag, New York, N.Y.
Edman, P., & Begg, G. (1967) *Eur. J. Biochem. 1,* 80.
Gleisner, J. M., Peterson, D. L., & Blakley, R. L. (1974) *Proc. Natl. Acad. Sci. U.S.A. 71,* 3001.
Gray, W. R., & Smith, J. F. (1970) *Anal. Biochem. 33,* 36.
Gros, C., & Labouesse, B. (1969) *Eur. J. Biochem. 7,* 463.
Harley-Mason, J., & Archer, A. A. P. G. (1958) *Biochem. J. 69,* 60.
Hartley, B. S. (1970) *Biochem. J. 119,* 805.
Hitchings, G. H., & Burchall, J. J. (1965) *Adv. Enzymol. 27,* 417.
Houmard, J., & Drapeau, G. R. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69,* 3506.
Jeppsson, J.-O., & Sjöquist (1967) *Anal. Biochem. 18,* 264.
Liu, T.-Y., & Chang, Y. H. (1971) *J. Biol. Chem. 246,* 2842.
Matthews, D. A., Alden, R. A., Bolin, J. T., Freer, S. T., Hamlin, R., Xuong, N., Kraut, J., Poe, M., Williams, M. N., & Hoosteen, K. (1978) *Science* (in press).
Moore, S. (1963) *J. Biol. Chem. 238,* 235.
Morris, H. R., Batley, K. E., Harding, N. G. L., Bjur, R. A., Dann, J. G., & King, R. W. (1974) *Biochem. J. 137,* 409.
Niall, H. D., Keutmann, H. T., Copp, D. H., & Potts, J. T., Jr. (1969) *Proc. Natl. Acad. Sci. U.S.A. 64,* 771.
Offord, R. E. (1966) *Nature (London) 211,* 591.
Peterson, D. L., Gleisner, J. M., & Blakley, R. L. (1975) *J. Biol. Chem. 250,* 4945.
Poe, M., Greenfield, N. J., Hirschfield, J. M., Williams, M. N., & Hoogsteen, K. (1972) *Biochemistry 11,* 1023.
Ramachandean, L. K., & Witkop, B. (1967) *Methods Enzy-*

mol. 11, 283.
Rodkey, J. A., & Bennett, C. D. (1976) *Biochem. Biophys. Res. Commun. 72,* 1407.
Roseau, G., & Pantel, P. (1969) *J. Chromatogr. 44,* 392.
Rossman, M. G., Moras, D., & Olsen, K. W. (1974) *Nature (London) 250,* 194.
Rydon, H. N., & Smith, N. G. (1952) *Nature (London) 169,* 922.
Sanger, F., & Tuppy, H. (1951) *Biochem. J. 49,* 463.
Spiro, R. G. (1966) *Methods Enzymol. 8,* 3.
Stone, D., & Phillips, A. W. (1977) *FEBS Lett. 74,* 85.
Stone, D., Phillips, A. W., & Burchall, J. J. (1977) *Eur. J. Biochem. 72,* 613.
Thomas, E. D., & Storb, R. (1971) *Ann N.Y. Acad. Sci. 186,* 467.
Weiner, A. M., Platt, T., & Weber, K. (1972) *J. Biol. Chem. 247,* 3242.
Williams, M. N. (1975) *J. Biol. Chem. 250,* 322.
Williams, M. N., & Bennett, C. D. (1978) *J. Biol. Chem.* (in press).
Williams, M. N., Greenfield, N. J., & Hoogsteen, K. (1973a) *J. Biol. Chem. 248,* 6380.
Williams, M. N., Poe, M., Greenfield, N. J., Hirschfield, J. M., & Hoogsteen, K. (1973b) *J. Biol. Chem. 248,* 6375.
Yamada, S., & Itano, H. A. (1966) *Biochim. Biophys. Acta 130,* 538.

Homogeneous Rabbit Immunoglobulin Lacking Group a Allotypes: Amino Acid Sequence Analysis of the Heavy Chain†

Alan P. Johnstone,‡ Allen L. Thunberg,§ and Thomas J. Kindt*

ABSTRACT: The partial amino acid sequence of rabbit a-negative heavy chain has been determined for residues 1–43 as: <EEQLEESGGGLVQPGGSLKLSCKGSGFDFSVY-GVTWVRQAPGK; and for residues 64–120 as: MNG-RFTISSDNAQNRLYLQLNSLTAADTATYFCARSMV-VVAGVHSYFDVWGPGTLVTV. Comparison of this sequence with the human heavy chain subgroup III shows homology of 78% suggesting that a common ancestral variable region gene existed in mammals prior to speciation. The constant region of the a-negative chain is structurally identical with that of a-positive chains, whereas the variable region differs substantially between a-positive and a-negative molecules. These findings support the concept that two genes encode one immunoglobulin polypeptide chain and demonstrate the existence in the rabbit of variable region subgroups similar to those reported for humans and other species. A novel approach to the initial fragmentation of the heavy chain was developed in this study. This method, which involved digestion of the H chain with the protease V8, produced a free N terminus and should have wide application in future studies on heavy chains with blocked amino terminals.

Rabbit immunoglobulin heavy (H[1]) chain V regions carry several genetically determined antigenic markers (Oudin, 1960a,b), analyses of which have provided insight into basic immunogenetic questions. Although the majority of rabbit V_H regions carry group a allotypes, molecules lacking these specificities were revealed by quantitative determination of the group a markers in immunoglobulin pools (Dray et al., 1963), by homozygous allotype suppression experiments (David & Todd, 1969), and by the production of homogeneous antibodies that do not react with antisera directed against the group a specificities (Kindt et al., 1970). Kim & Dray (1972) have shown that there are allotypic markers other than those of group a present on this minority population.

Amino acid sequence studies on the H chains from immunoglobulin pools (Wilkinson, 1969; Mole et al., 1971; Johnstone & Mole, 1977) and from homogeneous antibodies (Fleischman, 1971; Jaton, 1975; Margolies et al., 1977) have demonstrated differences between the three group a allotypes, a1, a2, and a3. Structural differences between molecules possessing the group a allotypes and those lacking them were indicated by compositional differences and by sequence analysis of N-terminal peptides (Prahl et al., 1973; Tack et al., 1973). More recently, the sequence of a-negative molecules in an immunoglobulin pool was found to be 80% homologous to the human V_HIII subgroup in the N-terminal 20 residues (Johnstone & Mole, 1977).

A homogeneous immunoglobulin lacking group a allotypes was produced by rabbit 3547 upon hyperimmunization with group A streptococci. The L chain of this immunoglobulin was of the b4 allotype and its V region structure has been determined (Thunberg & Kindt, 1976). The present study, concerning sequence analysis on the a-negative H chain of this immunoglobulin, was initiated to investigate the structural basis of V region serological markers and to determine whether structurally identical C regions are associated with both a-positive and a-negative V regions (Dreyer & Bennett, 1965). Preliminary accounts of these findings have been presented

† From The Rockefeller University, New York, New York 10021. *Received August 25, 1977.* This work was supported by U.S. Public Health Service Grant AI08429 from the National Institute of Allergy and Infectious Diseases.

‡ Recipient of a NATO Postdoctoral Fellowship. Current address: National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7-1AA, England.

§ Current address: Eastman Kodak Company, Rochester, New York 14650.

* Current address: Laboratory of Immunogenetics, Building 5, NIAID-NIH, Bethesda, Maryland 20014.

[1] Abbreviations used: H chain, heavy chain; L chain, light chain; V region, variable region; C region, constant region; V_HI, V_HII, V_HIII, human heavy chain variable region subgroups I, II, and III, respectively; CmCys, *S*-carboxymethylcysteine; Dnp-, 2,4-dinitrophenyl-; CN, V8, T, TL, peptides from CNBr, V8-protease, tryptic and thermolytic cleavage, respectively; SPITC, 4-sulfophenyl isothiocyanate.

0006-2960/78/0417-1337$01.00/0