

## THE AMINO ACID SEQUENCE OF DIHYDROFOLATE REDUCTASE FROM L1210 CELLS

D. STONE and A. W. PHILLIPS

*Department of Protein Chemistry, Microbiology Research Division, The Wellcome Research Laboratories, Langley Court, Beckenham, Kent BR3 3BS, England*

Received 5 January 1977

### 1. Introduction

Dihydrofolate reductase (5678 tetrahydrofolate: NADP<sup>+</sup>-oxidoreductase, E C 1.5.1.3) has been intensively studied because of its role as a target enzyme for antibacterial drugs. The antibacterial activity of these drugs, particularly trimethoprim, has been shown to depend on the much higher affinity with which they bind to the bacterial enzymes compared to the mammalian enzymes [1]. This difference in susceptibility to inhibitors has stimulated interest in the structures of dihydrofolate reductases from both bacterial and mammalian sources, although work has generally focussed on the bacterial enzymes because of their greater availability. Thus the complete amino acid sequences of the dihydrofolate reductases from *E. coli* MB1428 [2], *E. coli* RT500 [3] and *S. faecium* [4] have been established while, in contrast, the only data for mammalian enzymes are two reports of the amino terminal 34 and 19 residues of the bovine liver enzymes [5,6] and the amino terminal 25 residues of the mouse sarcoma 180 enzyme [7].

This present paper reports the amino acid sequence of the dihydrofolate reductase from a Methotrexate-resistant line of the mouse lymphoma L1210 and compares this sequence with those of other dihydrofolate reductases.

### 2. Experimental

L1210 cells were grown in DBA/2 mice and the dihydrofolate reductase was isolated essentially as described by Whiteley *et al.* [8].

Dihydrofolate reductase in which the cysteine resi-

dues had been alkylated with iodo [<sup>14</sup>C] acetic acid was cleaved with cyanogen bromide and the resulting fragments separated into three fractions by gel-filtration on Sephadex G-75. The largest fragment was eluted as a single component. Further purification of the intermediate and low molecular weight fractions was carried out by chromatography on DEAE-cellulose and high-voltage electrophoresis respectively. A total of 6 fragments were isolated by these procedures.

The amino acid sequence of the three fragments comprising the low molecular weight fraction were determined by the dansyl-Edman procedure. The major parts of the sequences of the three larger fragments were determined by automatic sequencer analysis in the Beckman 890C automatic sequencer using the peptide programme number 102974. The information required to complete these sequences was obtained from manual dansyl-Edman sequence determinations of peptides isolated from either tryptic or staphylococcal protease digests of the fragments.

The alignment of the three amino terminal CNBr fragments was determined by automatic sequencer analysis of the intact protein in the Beckman 890C using the protein programme number 122974. In order to obtain overlap data for the remaining fragments, tryptic peptides were isolated from dihydrofolate reductase in which the methionines had been alkylated with iodo [2-<sup>14</sup>C] acetic acid. These radioactive peptides were isolated by gel-filtration on Sephadex G-25 followed by high-voltage electrophoresis and their sequences determined either by automatic sequencer analysis or by the dansyl-Edman procedure. Full details of these procedures will be published elsewhere.

The methods for CNBr cleavage, enzymic digestion, amino acid analysis, high-voltage electrophoresis, the

dansyl-Edman procedure and automatic sequencer analysis are as described previously [3].

### 3. Results and discussion

The sequence of L1210 dihydrofolate reductase is shown in fig.1. No particular problems were encountered in the sequence determination although the Met<sub>37</sub>Thr<sub>38</sub> bond failed to cleave with CNBr. The failure of Met–Thr bonds to cleave with CNBr has been reported for other proteins [9,10] although in these cases partial cleavage occurred. In the present case however, no detectable cleavage occurred and the separation of the CNBr fragments was not complicated by the presence of partial cleavage products. Consequently only six CNBr fragments were isolated instead of the seven anticipated.

The sequence determined contains 186 residues and corresponds to mol 21 458. This value is in fairly good agreement with the value of 22 500 determined by SDS–polyacrylamide gel electrophoresis (J. Raper, unpublished results) and the value of 20 000 reported by Neef and Huennekens [11] for the enzyme from L1210/R6 cells. The L1210 enzyme is therefore longer than the *S.faecium* and *E.coli* enzymes by 18 and 26 residues respectively. As shown in fig.1, all these enzymes show a considerable number of areas of identical sequence, the degree of identity between the L1210 enzyme and both bacterial enzymes being 29%. The most extended regions of homology between all three enzymes are those between residues 49 and 60 and 138 and 149, although the following 4 regions also contain a high proportion of identical residues: 16–30, 65–78, 91–98, 111–122. On the basis of a comparison of the sequences of the *S.faecium* and

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	
L1210	VAL	ARG	PRO	LEU	ASN	CYS	ILE	VAL	ALA	VAL	SER	GLN	ASN	MET	GLY	ILE	GLY	LYS	ASN	GLY	ASP	LEU	PRO	TRP	PRO	PRO	LEU	ARG	ASP	GLN	
<u>S.faecium</u>			MET	PHE	ILE	SER	MET	TRP	ALA	GLN	ASP	LYS	ASN	GLY	LEU	ILE	GLY	LYS	ASP	GLY	LEU	LEU	PRO	TRP	ARG	LEU	PRO	ASN	ASP	MET	
<u>E.coli</u>			MET	ILE	SER	LEU	ILE	ALA	ALA	LEU	ALA	VAL	ASP	ARG	VAL	ILE	GLY	MET	GLU	ASN	ALA	MET	PRO	TRP	ASN	LEU	PRO	ALA	ASP	LEU	
	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	
	PHE	LYS	TYR	PHE	GLN	ARG	MET	THR	THR	THR	SER	SER	VAL	GLN	GLY	LYS	GLN	ASN	LEU	VAL	ILE	MET	GLY	ARG	LYS	THR	TRP	PHE	SER	ILE	
	ARG			PHE	PHE	ARG	GLU	HIS	THR	MET	ASP					LYS	ILE		LEU	VAL		MET	GLY	ARG	LYS	THR	TYR	GLU	GLY	MET	
	ALA	TRP		PHE	LYS	ARG	ASN		THR	LEU	ASP					LYS	PRO				VAL	ILE	MET	GLY	ARG	HIS	THR	TRP	GLU	SER	ILE
	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	
	PRO	GLU	LYS	ASN	ARG	PRO	LEU	LYS	ASP	ARG	ILE	ASN	ILE	VAL	LEU	SER	ARG	GLU	LEU	LYS	GLU	PRO	PRO	ARG	GLY	ALA	HIS	PHE	LEU	ALA	
	GLY	LYS	LEU	SER	LEU	PRO	TYR			ARG	HIS	ILE	ILE	VAL	LEU	THR	THR	GLN	LYS	ASP	PHE	LYS	VAL	GLU	LYS	ASN	ALA	GLU	VAL	LEU	
	GLY			ARG	PRO	LEU	PRO	GLY	ARG	LYS	ASN	ILE	ILE	LEU	SER	SER	GLN	PRO	GLY	THR	ASP	ASP	ARG	VAL	THR	TRP	VAL				
	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	
	LYS	SER	LEU	ASP	ASP	ALA	LEU	ARG	LEU	ILE	GLN	GLU	PRO	GLU	LEU	ALA	SER	LYS	VAL	ASP	MET	VAL	TRP	ILE	VAL	GLY	GLY	SER	SER	VAL	
	HIS	SER	ILE	ASP	GLU	LEU	LEU	ALA	TYR	ALA	LYS	ASP	ILE	PRO	GLU	ASP	ILE	TYR	VAL	SER					GLY	GLY	SER	ARG	ILE		
	LYS	SER	VAL	ASP	GLU	ALA	ILE	ALA	ALA	CYS	GLY	ASN	VAL	PRO	GLU	ILE				MET	VAL			ILE	GLY	GLY	GLY	ARG	VAL		
	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	
	TYR	GLU	GLN	ALA	MET	ASN	GLU	PRO	GLY	HIS	LEU	ARG	LEU	PHE	VAL	THR	ARG	ILE	MET	GLN	GLU	PHE	GLU	SER	ASP	THR	PHE	PHE	PRO	GLU	
	PHE		GLN	ALA	LEU	LEU		PRO	GLU	THR	LYS	ILE	ILE	TRP	ARG	THR	LEU	ILE	ASP	ALA	GLU	PHE	GLU	GLY	ASP	THR	PHE	ILE	GLY	GLU	
	TYR	GLU	GLN	PHE	LEU			PRO	LYS	ALA	GLN	LYS	LEU	TYR	LEU	THR	HIS	ILE	ASP	ALA	GLN	VAL	GLU	GLY	ASP	THR	HIS	PHE	PRO	ASP	
	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	
	ILE	ASP	LEU	GLY	LYS	TYR	LYS	LEU	LEU	PRO	GLU	TYR	PRO	GLY	VAL	LEU	SER	GLN	VAL	GLN	GLU	GLU	ASP	GLY	ILE	LYS	TYR	LYS	PHE	GLU	
	ILE	ASP	PHE	THR	SER	PHE	GLU	LEU	VAL	GLU	GLU	HIS	GLU	GLY	ILE	VAL	ASN	GLU	GLU	ASN	GLN	TYR	PRO	HIS	ARG	PHE	GLN	LYS	TRP	GLN	
	TTR	GLU	PRO	ASP	ASP	TRP	GLU	SER	VAL	PHE	SER	GLU	PHE	HIS	ASP	ALA	ASP	ALA	GLN	ASN	SER	HIS	SER	TTR	CYS	PHE	GLU	ILE	LEU	GLU	
	181	182	183	184	185	186																									
	VAL	TYR	GLU	LYS	LYS	ASP																									
	LYS	MET	SER	LYS	VAL	VAL																									
	ARG	ARG																													

Fig.1. Comparison of the amino acid sequences of dihydrofolate reductases from L1210, *E.coli*, *S.faecium*. The numbering is that for the L1210 enzyme. The sequences have been aligned to maximise the homology between each of the bacterial enzymes and the L1210 enzyme without introducing gaps into the latter sequence.

*E.coli* enzymes, Gleisner *et al.* [4] drew attention to two domains of homology which exist between the two molecules and it is interesting to note that the regions of those enzymes which show the most extensive homology with the L1210 reductase also fall within these domains.

A comparison of the N-terminal part of the L1210 enzyme with the N-terminal portions of other mammalian dihydrofolate reductases shows very close similarity. The first 30 residues are identical with the pig liver enzyme (our unpublished results) with the possible exception that in the latter case residue 6 is either Cys or Ser. The sequence of the first 25 residues is identical to that of the mouse sarcoma 180 [7]. However the sequence of the first 19 residues of a bovine liver enzyme reported by Bauman and Wilson [5] differs in having Ser at position 6 while the sequence of the first 35 residues of a bovine liver enzyme reported by Peterson *et al.* [6] differs from L1210 at positions 6 (Ala for Cys), 21 (Tyr for Asp) and 32 (Lys for Glu).

#### Acknowledgements

The authors are indebted to Dr. R. Jackson for a gift of Methotrexate-resistant L1210 cells, to Dr

A. Whitaker for maintaining the cells in culture and to Mr. J. Raper for isolating the L1210 dihydrofolate reductase.

We also thank Mrs. S. Paterson, Miss S. Couper and Mr R. J. Foreman for skilled technical assistance.

#### References

- [1] Burchall, J. J. and Hitchins, G. H. (1965) *Mol. Pharmacol.* 1, 126–136.
- [2] Bennett, C. D. (1974) *Nature (London)* 248, 67–68.
- [3] Stone, D., Phillips, A. W. and Burchall, J. J. (1977) *Eur. J. Biochem.* (in press)
- [4] Gleisner, J. M., Peterson, D. L. and Blakley, R. L. (1974) *Proc. Natl. Acad. Sci. USA* 71, 3001–3005.
- [5] Bauman, H. and Wilson, K. J. (1975) *Eur. J. Biochem.* 60, 9–15.
- [6] Peterson, D. L., Gleisner, J. M. and Blakley, R. L. (1975) *Biochemistry* 14, 5261–5267.
- [7] Rodkey, J. A. and Bennett, C. D. (1976) *Biochem. Biophys. Res. Commun.* 72, 1407–1413.
- [8] Whiteley, J. M., Jackson, R. C., Mell, G. P., Drais, J. H. and Huennekens, F. M. (1972) *Arch. Biochem. Biophys.* 150, 15–22.
- [9] Schroeder, W. A., Shelton, J. B. and Shelton, J. R. (1969) *Arch. Biochem. Biophys.* 130, 551–556.
- [10] DeLange, R. J. (1970) *J. Biol. Chem.* 245, 907–916.
- [11] Neef, V. G. and Huennekens, F. M. (1975) *Arch. Biochem. Biophys.* 171, 435–443.