The Amino Acid Sequence of *Neurospora* NADP-Specific Glutamate Dehydrogenase

PEPTIC AND CHYMOTRYPTIC PEPTIDES AND THE COMPLETE SEQUENCE

By ANTHONY A. HOLDER, JOHN C. WOOTTON, ANDREW J. BARON,
GEOFFREY K. CHAMBERS and JOHN R. S. FINCHAM
Department of Genetics, University of Leeds, Leeds LS2 9JT, U.K.

(Received 23 December 1974)

Peptic and chymotryptic peptides were isolated from the NADP-specific glutamate dehydrogenase of *Neurospora crassa* and substantially sequenced. Out of 452 residues in the polypeptide chain, 265 were recovered in the peptic and 427 in the chymotryptic peptides. Together with the tryptic peptides [Wootton, J. C., Taylor, J. G., Jackson, A. A., Chambers, G. K. & Fincham, J. R. S. (1975), *Biochem. J.* **149**, 739-748] and staphylococcal proteinase peptides [Wootton, J. C., Baron, A. J. & Fincham, J. R. S. (1975) *Biochem. J.* **149**, 749-755], these establish the complete sequence of the chain, including the acid and amide assignments, except for seven places where overlaps are inadequate. These remaining alignments are deduced from information on the CNBr fragments obtained in another laboratory [Blumenthal, K. M., Moon, K. & Smith, E. L. (1975), *J. Biol. Chem.* **250**, 3644-3654]. Further information has been deposited as Supplementary Publication SUP 50054 (17 pages) with the British Library (Lending Division), Boston Spa, Wetherby, W. Yorkshire LS23 7BQ, U.K., from whom copies may be obtained under the terms given in *Biochem. J.* (1975) **145**, 5.

In the two preceding papers (Wootton *et al.*, 1975*a,b*) the amino acid sequences of tryptic and staphylococcal proteinase peptides obtained from *Neurospora crassa* NADP-specific glutamate dehydrogenase (EC 1.4.1.4) were described. These sequences allow a substantial part of the entire sequence of the glutamate dehydrogenase to be assembled, but there are still numerous gaps and a considerable number of acid/amide assignments remain to be made. In this paper we report sequences of peptic and chymotryptic peptides which supply most of the missing information, including all amide assignments. Additional information on the CNBr fragments, obtained in the laboratory of E. L. Smith (Blumenthal *et al.*, 1975), allow a complete sequence to be deduced. The previously published provisional complete sequence (Wootton *et al.*, 1974) contained two undetermined acid/amide assignments (261 and 218) and typographical errors at two residues (2 and 421) which are rectified here (Fig. 1).

Materials and Methods

Source of the protein

Glutamate dehydrogenase from wild-type *Neurospora* was used in the pepsin digestion, but most of the information on chymotryptic peptides was obtained from a digest of glutamate dehydrogenase isolated from the mutant *am*¹⁹ (Fincham & Stadler, 1965; Coddington *et al.*, 1966). The purification procedure for the glutamate dehydrogenase from mutant

*am*¹⁹ was essentially the same as that for the wild-type and has been described elsewhere (Ashby *et al.*, 1974). The assay of the glutamate dehydrogenase activity from mutant *am*¹⁹ during the purification required the special assay system described in Coddington *et al.* (1966) and referred to as System S. As is shown by the sequences presented in Table 2 of the present paper, the glutamate dehydrogenase from mutant *am*¹⁹ has a Lys → Met replacement at residue 141; otherwise the sequences from mutant *am*¹⁹ and wild-type appear to be identical. Some chymotryptic peptides were derived from one or other of two prototrophic revertants derived from strain *am*¹⁹, called R1 and R4 (Stadler, 1966). These, like another revertant R26 (Wootton *et al.*, 1975*b*), retain the methionine-141 characteristic of the progenitor mutant. They presumably each have a second amino acid replacement somewhere else in the chain, and this has probably been identified in the case of strain R1 (A. A. Holder, unpublished work), but the peptides from strain R1 and R4 cited in this paper contain only normal sequences as shown by comparisons with wild-type tryptic and peptic peptides.

From each fungal strain about 100 mg of purified glutamate dehydrogenase was obtained per kg of damp mycelium. The protein was carboxymethylated before digestion (Hirs, 1967).

Digestion procedures

For chymotrypsin digestion 150-200 mg (3-4 μ mol) of carboxymethylated glutamate dehydrogenase

Table 1. *Amino acid sequences of peptic peptides*

—, Round of Edman degradation with end group identified; . . . , round of Edman degradation without clear end-group identification. Net charges were determined at pH 6.5, sometimes after some rounds of Edman degradation (1E, 2E etc.); some amides were assigned after limited digestion with aminopeptidase-M (AP-M) or carboxypeptidase A+B [CP(A+B)]. Residues in parentheses, consistent with the analysis, are given in the order known from other kinds of peptides.

Peptide	Sequence	Evidence for amide assignments
P2	10 Glx-Glx-Ala-Tyr-Lys-Glx-Leu-Ala	
P2A	13 Tyr-Lys-Glu-Leu-Ala	Charge 0
P3	18 Tyr-Thr-Leu (Glu)	Charge -1
P5	33 Arg-Thr-Ala-Leu	
P6	37 Thr-Val-Ala-Ser-Ile-Pro-Glu-Arg-Val-Ile-Gln	Charge 0 (glutamate-43 from T4)
P7	49 Arg-Val-Val-Trp-Glx-Asx-Asx(Gly)	
P8	58 Asn-Val-Gln-Val-Asn-Arg-Gly	Charge +1 (after 1E and 3E also)
P10	89 Ser-Ile(Leu)	
P11	92 Lys-Phe-Leu (Gly)	
P12	97 Glu-Gln-Ile-Phe-Lys-Asn-Ala-Leu-Thr-Gly-Leu	Charge 0 (glutamate-97, glutamine-98 from T11)
P13	107 Leu-Ser-Met-Gly-Gly-Gly-Lys-Gly-Ala-Asp-Phe-Asp(Pro)Lys	Charge -1

P15	136 Met-Ala-Glx-Leu-His-Lys-His-Ile-Gly-Ala (Asx, Thr, Asx, Val, Pro, Ala, Gly, Asx, Ile, Gly, 160 Val, Gly, Gly, Arg, Glx)	
P15A	His-Lys-His-Ile-Gly-Ala (Asx, Thr etc.)	
P16	161 163 Ile-Gly(Tyr, Met)	
P19	191 201 Ile-Arg-Pro-Glu-Ala-Thr-Gly-Tyr-Gly-Leu-Val	Charge 0
P19A	Ile-Arg-Pro-Glu (Ala, Thr, Gly, Tyr, Gly, Leu)	Charge 0
P21	210 231 Tyr-Ser-Gly-Ala-Gly-Ser-Tyr-Ala-Gly-Lys-Arg-Val-Ala-Leu-Ser-Gly-Ser-Gly-Asx-Val-Ala-Glx	
P22	232 237 Tyr-Ala-Ala-Leu-Lys-Leu	
P22A	233 237 Ala-Ala-Leu-Lys-Leu	
P23	238 244 Ile-Glu-Leu-Gly (Ala, Thr, Val)	Charge -1
P24	245 254 Val-Ser-Leu-Ser (Asp, Ser, Lys, Gly, Ala, Leu)	Charge 0
P24A	Leu-Ser-Asp-Ser-Lys-Gly-Ala-Leu	Charge 0
P24B	Ser-Asp-Ser-Lys-Gly-Ala-Leu	Charge 0
P23/24	241 254 Gly-Ala-Thr-Val (Val, Ser, Leu, Ser, Asp, Ser, Lys, Gly, Ala, Leu)	
P23/24A	241 247 Gly-Ala-Thr-Val-Val-Ser(Leu)	

Table 1 (*continued*)

Peptide	Sequence	Evidence for amide assignments
P25	255 Val-Ala-Thr-Gly-Glu-Ser-Gly-Ile-Thr-Val-Glu 265	CF(A + B) and AP-M (complete digestion)
P27	272 Ala-Ile-Lys-Glx-Ala-Arg-Glx-Ser-Leu 280	
P29	306 Ile-Ala (Leu) 308	
P31	316 Val-Ser-Lys-Glu-Glu-Ala-Glu-Gly-Leu-Leu 325	Charge -2
P31A	318 Lys-Glu-Glu-Ala-Glu-Gly-Leu 324	Charge -2
P32	326 Ala-Ala-Gly-Cys-Lys-Phe 331	
P33	332 Val-Ala-Glu-Gly-Ser-Asn-Met 338	Charge -1, Asn by AP-M
P34	339 Gly-Cys-Thr-Leu-Glu-Ala 344	Charge -2 (Glu, CnCys)
P33/34	332 Val-Ala-Glx-Gly-Ser-Asx-Met-Gly-Cys-Thr-Leu-Glx-Ala 344	
P36	379 Glu-Met-Ala-Gln-Asn-Ser-Gln-Arg-Leu 387	Charge 0 (cf. P36A)
P36A	Met-Ala-Gln-Asn-Ser-Gln (Arg, Leu)	Charge +1 (cf. P36)
P36B	Ala-Glx-Asx-Ser-Glx-Arg-Leu	
P38	393 Glu-Val-Asp-Glu-Lys-Leu-Lys-Asp 400	Charge -2

Charge 0 and +1 (partial
deamidation ?)

Charge -2

Charge 0

P38A	Val-Asx-Glx (Lys, Leu, Lys, Asx)	
	→ → → → →	
	407	
P40	Phe-Asn-Gly-Leu-Asn-Thr-Ala-Lys-Thr	415
	→ → → → →	→ → → → →
P40A	Phe-Asx-Gly (Leu, Asx, Thr)	
	→ → → → →	
P40B	Ala-Lys (Thr)	
	→ → → → →	
	419	
P42	Ala-Ala-Glu-Gly-Glu-Leu	424
	→ → → → →	→ → → → →
	428	
P44	Val-Ala-Gly-Ser-Asn-Ile-Ala-Gly	435
	→ → → → →	→ → → → →
	437	
P45	Val-Lys-Val-Ala-Glx-Ala-Met-His-Asx-Glx-Gly-Asx (Trp)	449
	→ → → → →	→ → → → →

was suspended and dissolved so far as possible in 0.1M-NH₄HCO₃ (20-30ml). For each 100mg of glutamate dehydrogenase, 2mg of chymotrypsin (Worthington Biochemical Corp., Freehold, N.J., U.S.A.; three times crystallized), dissolved in a small volume (about 3 ml) of the same buffer, was added. Digestion was allowed to proceed at 37°C for 3h, the reaction was stopped by adjusting the pH to 4 with acetic acid, and the digest was freeze-dried.

For the pepsin digestion, 180mg of wild-type carboxymethylated glutamate dehydrogenase was dissolved in 49ml of 0.01M-HCl; 3.5mg of pepsin [Sigma (London) Chemical Co., London S.W.6, U.K.; twice crystallized and freeze-dried] was dissolved in 0.35ml of 3% (v/v) acetic acid+3% (v/v) pyridine, mixed to give pH4.0, and was added to the glutamate dehydrogenase. Digestion was allowed to proceed at room temperature for 1h, after which the reaction was stopped by adjusting the pH to 5 with pyridine, and the digest was freeze-dried.

Purification of peptides

The procedures, utilizing columns of Dowex AG50 (X4), Dowex AG1 (X2) and Bio-Gel P4, were as described by Wootton *et al.* (1975a). Details for each peptide are given in Tables 1 and 2 of Supplementary Publication SUP 50054.

Amino acid analysis and sequence determination

Hydrolysis of peptides was as previously described (Wootton *et al.*, 1975a). Amino acid analyses were performed in a Bio-Cal BC-200 amino acid analyser. The recovery of methionine was usually very low, and only a qualitative indication of presence or absence was obtained in most analyses, the exceptions being when 4% (v/v) thioglycollic acid was included in the hydrolysis mixture. In a few analyses (e.g. that of peptide C45/46) the value for lysine was obviously too high, probably because of interference by a peak running close to NH₃ and apparently being derived from some batches of deionized water. The analyses are given in Tables 3 and 4 of Supplementary Publication SUP 50054.

All sequence determination was by the dansyl-Edman procedure with the modifications described by Wootton *et al.* (1975a).

Amide assignments

Most acid/amide assignments were made on the basis of net charge as determined by high-voltage paper electrophoresis at pH6.5 (Offord, 1966), either on the intact peptide or on the peptide both intact and after some rounds of Edman degradation. In some otherwise difficult cases the residue in question was identified as the phenylthiohydantoin derivative after the appropriate Edman step. The butyl acetate extract obtained after the cyclization reaction was dried, and the amino acid 2-anilinothiazolinone was

Fig. 1. Complete sequence proposed for the monomer of *Neurospora crassa* NADP-specific glutamate dehydrogenase

Tryptic, staphylococcal proteinase, peptic and chymotryptic peptides, and CNBr fragments, are indicated by T, SP, P, C and CN respectively.

C1	P2	C2	P3	C3
10			20	30
AcSer-Asn-Leu-Pro-Ser-Glu-Pro-Glu-Phe-Glu-Gln-Ala-Tyr-Lys-Glu-Leu-Ala-Tyr-Thr-Leu-Glu-Asn-Ser-Ser-Leu-Phe-Gln-Lys-His-Pro-				
CNI				
SP1	SP2	SP3	SP4	SP5
T1			T2	
P5	P6		P7	
C5	C6	C7	C8	60
40		50		
Glu-Tyr-Arg-Thr-Ala-Leu-Thr-Val-Ala-Ser-Ile-Pro-Glu-Arg-Val-Ile-Gln-Phe-Arg-Val-Trp-Glu-Asp-Gly-Asn-Val-Gln-				
CNI				
T3	T4	T5	T6	
SP6				
P8	C9	C10	C11	C12
				C13
				C14
				90
Val-Asn-Arg-Gly-Tyr-Arg-Val-Gln-Phe-Asn-Ser-Ala-Leu-Gly-Pro-Tyr-Lys-Gly-Gly-Leu-Arg-Leu-His-Pro-Ser-Val-Asn-Leu-Ser-Ile-				
CNI				
T7	T8	T9	T10	
P11	P12		P13	
C15	C16	C17	C18	
	100			120
Leu-Lys-Phe-Leu-Gly-Phe-Glu-Gln-Ile-Phe-Lys-Asn-Ala-Leu-Thr-Gly-Leu-Ser-Met-Gly-Gly-Lys-Gly-Ala-Asp-Phe-Asp-Pro-				
CNI				
T11	T12	T13		
			P15	
		C20	C21	150
	130			
Lys-Gly-Lys-Ser-Asp-Ala-Glu-Ile-Arg-Arg-Phe-Cys-Cys-Ala-Phe-Met-Ala-Glu-Leu-His-Lys-His-Ile-Gly-Ala-Asp-Thr-Asp-Val-Pro-				
CNI				
T14	T15	T16		
SP9B		SP11		

	P16	C22	C23	C24	C25	
	160			170		180
Ala-Gly-Asp-Ile-Gly-Val-Gly-Gly-Arg-Glu-Ile-Gly-Tyr-Met-Phe-Gly-Ala-Tyr-Arg-Lys-Ala-Ala-Asn-Arg-Phe-Glu-Gly-Val-Leu-Thr-						
CN3						CN4
T16		T17				T19
		SP12A				
			P19			
C26	C27/28/29			C30		C31
	190			200		210
Gly-Lys-Gly-Leu-Ser-Trip-Gly-Gly-Ser-Leu-Ile-Arg-Pro-Glu-Ala-Thr-Gly-Tyr-Gly-Leu-Val-Tyr-Tyr-Val-Gly-His-Met-Leu-Glu-Tyr-						
CN4						
			T20			
	P21				P22	
C32	C33		C34		C35	C36
	220			230		240
Ser-Gly-Ala-Gly-Ser-Tyr-Ala-Gly-Lys-Arg-Val-Ala-Leu-Ser-Gly-Ser-Gly-Asn-Val-Ala-Gln-Tyr-Ala-Ala-Leu-Lys-Leu-Ile-Glu-Leu-						
CN5						
			T21			
P23	P24			P25		
C37		C38				
	250			260		270
Gly-Ala-Thr-Val-Val-Ser-Leu-Ser-Asp-Ser-Lys-Gly-Ala-Leu-Val-Ala-Thr-Gly-Glu-Ser-Gly-Ile-Thr-Val-Glu-Asp-Ile-Asn-Ala-Val-						
CN5						
T22				T23		
	SP16			SP17		SP18
P27						
C40		C41	C42	C43	C44	
	280			290		300
Met-Ala-Ile-Lys-Glu-Ala-Arg-Gln-Ser-Leu-Thr-Ser-Phe-Gln-His-Ala-Gly-His-Leu-Lys-Trp-Ile-Glu-Gly-Ala-Arg-Pro-Trp-Leu-His-						
CN6						
T24			T25			
			SP19			

Fig. 1 (continued)

P29	C45	P31	P32
			C46
Val-Gly-Lys-Val-Asp-Ile-Ala-Leu-Pro-Cys-Ala-Thr-Glu-Asp-Glu-Val-Ser-Lys-Glu-Glu-Ala-Gly-Leu-Leu-Ala-Ala-Gly-Cys-Lys-	310	320	330
	CN6		
T26B	T27		
SP20	SP21/22		SP23
P33	P34		
C47	C48	C49	
Phe-Val-Ala-Glu-Gly-Ser-Asn-Met-Gly-Cys-Thr-Leu-Glu-Ala-Ile-Glu-Val-Phe-Glu-Asn-Asn-Arg-Lys-Gly-Lys-Gly-Glu-Ala-Val-	340	350	360
	CN7		
	T28	T29	
SP24	SP25	SP26	SP27
			SP28
		P36	
C50		C51	C52
Trp-Tyr-Ala-Pro-Gly-Lys-Ala-Ala-Asn-Cys-Gly-Gly-Val-Ala-Val-Ser-Gly-Leu-Glu-Met-Ala-Gln-Asn-Ser-Gln-Arg-Leu-Asn-Trp-Thr-	370	380	390
	CN7		CN8
T30	T31		
SP29			SP30
P38		P40	
C53	C54	C56	
Gln-Ala-Glu-Val-Asp-Glu-Lys-Leu-Lys-Asp-Ile-Met-Lys-Asn-Ala-Phe-Phe-Asn-Gly-Leu-Asn-Thr-Ala-Lys-Thr-Tyr-Val-Glu-Ala-Ala-	400	410	420
	CN8	CN9	
T32	T33	T34	T35
SP31	SP32		SP33
P42	P44	P45	
C57		C58	C59
			C60
Glu-Gly-Glu-Leu-Pro-Ser-Leu-Val-Ala-Gly-Ser-Asn-Ile-Ala-Gly-Phe-Val-Lys-Val-Ala-Gln-Ala-Met-His-Asp-Gln-Gly-Asp-Trp-Ser-Lys-Asn			
	CN9	CN10	CO2H
T36	T37	T38	
	SP34/35		

isomerized to the phenylthiohydantoin essentially as described by Laursen (1972). Identification of the phenylthiohydantoin derivative was achieved by one-dimensional t.l.c. on silica gel G (Niederwieser, 1972). The solvent used was chloroform-methanol (9:1, v/v), and spots were located by staining with iodine vapour or by quenching of fluorescence on illumination at 254 nm, and were identified by comparison with standards.

A few assignments were made by amino acid analysis after limited digestion of the peptide with carboxypeptidase A (or A+B) or aminopeptidase M (Wootton *et al.*, 1975a).

Results and Discussion

The sequences of the isolated peptic peptides are presented in Table 1, which indicates, by the numbering of the terminal residues of each peptide, the position in the complete proposed sequence (Fig. 1) where the peptide is found to fit. The peptides cover 265 of the sequence of 452 residues. They show a considerable degree of degeneracy and many were recovered in very low yields. In most cases the sequence obtained agreed well with the amino acid composition, the exceptions being peptides P5, P6, P13, P15, P21, P22, P23/24, P24A, P31A and P42, which gave contaminated analyses but clear sequences. We have used the peptic peptides for many acid/amide assignments (Table 1; see also Table 1 of Supplementary Publication SUP 50054) for which purpose their generally small size is an advantage. They also provide the definitive overlaps between the chymotryptic peptides C29 and C30 (P19) and between the following tryptic peptides (Wootton *et al.*, 1975a): T1 and T2 (P2), T11 and T12 (P12), T12 and T13 (P13), T15 and T16 (P15), T20 and T21 (P21), T21 and T22 (P22, together with C36), T36 and T37 (P45). They also confirm the detailed sequences obtained on peptides from other sources.

Although the catalogue of peptic peptides was incomplete, nearly all of the glutamate dehydrogenase chain (427 of the 452 residues) was recovered in the form of chymotryptic peptides. Table 2 shows the amino acid sequences of these peptides, numbered according to their deduced sequence from the *N*- to the *C*-terminus of the chain. In Table 2, also, the acid/amide assignments from all sources are collected, with the source indicated in each case. Three small sections of the sequence, known from other sources, are missing in the chymotryptic peptides. These are residues 132–135, 255–271 and 408–411. It is not clear why the two shorter fragments were missed (though the CmCys*-CmCys-Ala-Phe sequence between residues 132 and 135 may have been dispersed through partial oxidation of the carboxymethyl cysteine), but the 254–271 fragment includes the greater part of the

highly insoluble tryptic peptide T23 (Wootton *et al.*, 1975a) and would be expected to be itself insoluble.

Most chymotryptic peptides gave sequences in good agreement with the amino acid analyses, the exceptions being peptides C6, C7/8, C18, C22, C36/37, C47, C49, C53, C54 and C59 which gave contaminated analyses but clear sequences, and three cases where two peptides were analysed and sequenced as a mixture (see footnote to Table 4 of Supplementary Publication SUP 50054). Many chymotryptic peptides were used for amide assignments (Table 2; for primary data see Table 5 of Supplementary Publication SUP 50054). Peptides C24, C28, C27/28/29, C30, C31, C43 and C44 were of considerable value in giving detailed sequences not obtained from the tryptic, staphylococcal proteinase or peptic peptides. Chymotryptic peptides provide definitive overlaps between tryptic peptides T4 and T5 (C6), T16 and T17 (C21, using both the sequence and analysis), T19 and T20 (C26), T23, T24 and T25 (C40), T28, T29 and T30 (C49), T32 and T33 (C53) and between the staphylococcal proteinase peptides SP19 and SP20 (C44). They also confirm detailed sequences obtained on peptides from other sources.

The relatively full analysis of the chymotryptic peptides provides a good demonstration of the specificity of chymotrypsin. The bonds on the carboxyl side of aromatic residues were all cleaved with the sole exception of Phe-Glu at residues 9–10. There are no aromatic amino acid-proline sequences in the protein. Leucyl bonds were also cleaved with high probability but here there were many more exceptions; 13 out of 36, including all three Leu-Pro sequence were not seen to split, although, of course, some products may have been formed but missed in the fractionation. Of the eight methionyl bonds, two were shown to be susceptible to chymotrypsin, as were six of the 18 asparaginyl bonds. Glutaminyl bonds were not cleaved except for Gln-Val at residues 60–61, and nor were histidyl bonds except, interestingly enough, at both of the His-Met sequences in glutamate dehydrogenase from mutant *am*¹⁹. Only two bonds of other types were found to be cleaved: one was Val-Ser in the Val-Val-Ser sequence at residues 244–246 and the other was Lys-Lys at 355–356 (the only Lys-Lys in the entire chain).

Although the sequence set out in Fig. 1 and implied in the layout of the Tables in this and the two preceding papers (Wootton *et al.*, 1975a,b) is largely defined by our data, it stands in need of additional support in some places. This support is provided by the parallel study by Blumenthal *et al.* (1975) on the CNBr fragments. The particular parts of the sequence for which the CNBr fragments are important are as follows. In the *N*-terminal part of the sequence the connexions between the tryptic peptides spanning residues 49–50, 77–78, 81–82 and 92–93 are made, on the basis of peptic and chymotryptic peptides,

* Abbreviation: CmCys, carboxymethylcysteine.

Table 2. *Amino acid sequences of chymotryptic peptides*

—, Round of Edman degradation with end group identified; ···, round of Edman degradation without clear end group identification. Charges were determined by paper electrophoresis at pH 6.5. CP = carboxypeptidase, AP-M = aminopeptidase M, as described in the text.

Peptide	Sequence	Amide assignments
C1	¹ Ser-Asn-Leu-Pro-Ser-Glu-Pro-Glu-Phe-Glu-Gln-Ala-Tyr ¹³ ¹⁴ Lys-Glu-Leu-Ala-Tyr ¹⁸	From SP1, SP2, SP3, T2
C2	Lys-Glu-Leu-Ala-Tyr ¹⁸	From SP2, T2, P3
C2A	Lys-Glu-Leu	Charge 0; see also SP3, T2, P2A
C3	¹⁹ (Thr, Leu, Glu, Asn, Ser, Ser, Leu, Phe) ²⁶	From T2
C4	²⁷ Gln-Lys-His-Pro-Glu-Tyr ³² ³³ Arg-Thr-Ala-Leu ³⁶	From T2, T3
C5	Arg-Thr-Ala-Leu ³⁶	
C6	³⁷ Thr-Val-Ala-Ser-Ile-Pro-Glu-Arg-Val-Ile-Gln-Phe ⁴⁸	From SP6, T4, T5, P6
C7	⁴⁹ Arg-Val-Val-Trp ⁵²	
C7/8	⁴⁹ Arg-Val-Val-Trp-Glx-Asx-Asx-Gly(Asn)Val-Gln-Val(Asn) ⁶²	From P8, T6, T6Cl
C8(R1)	⁵³ Glu-Asp-Asp-Asp(Gly, Asn, Val, Gln, Val, Asn)	Charge -4, -3, -2, -1, 0 after 0, 1, 2, 3, 4 Edman steps
C8(R4)	Glu-Asp-Asp-Asp-Gly-Asn-Val-Gln-Val-Asn	From P8, C8(R1)
C8/9	⁶¹ Val-Asn-Arg-Gly-Tyr ⁶⁵	

C9							63 Arg-Gly-Tyr † <u>→ → →</u>	
C10		66	69				Arg-Val-Gln-Phe <u>→ → →</u>	Charge +1; see also T8
C11		70		76			Asn-Ser-Ala-Leu-Gly-Pro-Tyr <u>→ → →</u>	From T8
C12		77	80				Lys-Gly-Gly-Leu <u>→ → →</u>	
C13		81		88			Arg-Leu-His-Pro-Ser-Val-Asn-Leu <u>→ → →</u>	Charge +2; see also T10
C14		89	91				Ser-Ile-Leu <u>→ → →</u>	
C15		92		100			Lys-Phe-Leu-Gly-Phe-Glu-Gln-Ile-Phet <u>→ → → . . *</u>	From T11, P12
C15A							Leu-Gly-Phe <u>→ → →</u>	
C15B							Glu-Gln-Ile-Phe <u>→ → →</u>	From T11, P12
C16		101	104				Lys-Asn-Ala-Leu <u>→ → →</u>	
C16A							Lys-Asn <u>→ → →</u>	
C17		105	107				Thr-Gly-Leu <u>→ → →</u>	
C16/17		101		107			Lys-Asn-Ala-Leu-Thr-Gly-Leu <u>→ → →</u>	
C18		108					Ser-Met-Gly-Gly-Lys-Gly-Gly-Ala-Asp-Phe-Asp-Pro-Lys-Gly-Lys-Ser (Asp, Ala, Glu, Ile, Arg, Arg, Phe) <u>→ → → . . * → → →</u> 131	From SP9B, p13, T14A

Table 2 (continued)

Peptide	Sequence	Amide assignments
C20	136 Met-Ala-Glu-Leu	From T15B
C20/21	140 Met-Ala-Glu-Leu-His	
C21	140 His-Met-His-Ile-Gly-Ala-Asp-Thr-Asp-Val-Pro-Ala-Gly-Asp-Ile-Gly-Val-Gly-Gly-Arg (Glu, Ile, Gly, Tyr) 163	From T16, T17, T17A
C21A	Met-His-Ile-Gly (Ala, Asp etc.)	
C22	164 165 Met-Phe	
C23	166 168 Gly-Ala-Tyr	
C24	169 173 Arg-Lys-Ala-Ala-Asn	Charge +2; chymotrypsin specificity
C25	174 179 Arg-Phe-Glu-Gly-Val-Leu	
C25A	Glu-Gly-Val-Leu	From T19
C26	180 184 Thr-Gly-Lys-Gly-Leu	
C27/28/29	185 Ser-Trp-Gly-Gly-Ser-Leu-Ile-Arg-Pro-Glu-Ala-Thr-Gly-Tyr† 198	From P19
C28	187 190 Gly-Gly-Ser-Leu	

C29		191	Ile-Arg-Pro-Glu-Ala-Thr-Gly-Tyr	198
			→ → → → → → → → → →	
C27	185 186		Ser-Trp	
			→ →	
C30	199	203	Gly-Leu-Val-Tyr-Tyr	
			→ → → → →	
C30A			Gly-Leu-Val-Tyr	
			→ → → →	
C31		204	Val-Gly-His-Met-Leu-Glu-Tyr	210
			→ → → → → → →	
C30/31			Tyr-Val-Gly-His-Met-Leu-Glu-Tyr†	
			→ → → → → → →	
C30/31A			Tyr-Val-Gly-His	
			→ → → →	
C32	211	216	Ser-Gly-Ala-Gly-Ser-Tyr	
			→ → → → → →	
C33	217	223	(Ala)Gly-Lys-Arg-Val-Ala-Leu	
			→ → → → → → →	
C34	224	232	Ser-Gly-Ser-Gly-Asn-Val-Ala-Gln-Tyr	
			→ → → → → → →	
C35	233	235	(Ala,Ala,Leu)	
			→ → →	
C36	236	240	Lys-Leu-Ile-Glu-Leu	
			→ → → → →	
C37	241	247	Gly-Ala-Thr-Val-Val-Ser-Leu	
			→ → → → → →	

CP-A

Charge 0, also from T21

From T22, P23, P24

Table 2 (continued)

Peptide	Sequence	Amide assignments
C37A	Gly-Ala-Thr-Val-Val → → → → →	
C37B	Ser-Leu → →	
C36/37	236 Lys-Leu-Ile-Glu-Leu-Gly-Ala (Thr, Val, Val, Ser, Leu) → → → → →	247
C38	248 Ser-Asp(Ser) Lys-Gly-Ala-Leu → → → → →	254 From T22
C40	272 Ala-Ile-Lys-Glu-Ala-Arg-Gln-Ser-Leu → → → → →	280 From T24, T25
C41	281 283 Thr-Ser-Phe → → →	
C42	284* 289 Gln-His-Ala-Gly-His-Leu → → → → →	From T25 and blocked N-terminus
C43	290 291 Lys-Trp → →	
C44	292 Ile-Glu-Gly-Ala-Arg-Pro-Trp → → → → →	298 From SP19
C45	299 Leu-His-Val-Gly-Lys-Val-Asp-Ile-Ala-Leu(Pro, Cys) Ala-Thr-Glu-Asp-Glu-Val-Ser(Lys) → → → → →	Asp(305), Glu(313), Asp(314), Glu(315) by t.l.c.; others from P31
C46	326 331 Ala-Ala-Gly-Cys-Lys-Phe → → → → →	325 Glu-Glu-Ala-Glu(Gly, Leu, Leu) → → → → →
C45/46	325 331 Leu-Ala-Ala-Gly-Cys-Lys-Phe → → → → →	

C47	332 Val-Ala-Glu-Gly-Ser-Asn-Met-Gly-Cys-Thr-Leu	342	From P33
C48	343 Glu-Ala-Ile-Glu-Val-Phe	348	From SP24, SP25
C49	349 Glu-Asn-Asn-Arg-Lys-Glu-Lys-Lys-Gly-Glu-Ala-Val	361 (Trp)	From SP26, SP27, SP28
C49A	Lys-Gly-Glu-Ala-Val	(Trp)	
C50	362 Tyr-Ala-Pro-Gly-Lys-Ala-Ala-Asn-Cys-Gly-Gly-Val-Ala-Val	378 (Ser)Gly-Ileu	Charge +1; Asn(369) by t.l.c, also from T31
C50A	Ala-Pro-Gly-Lys-Ala-Ala-Asn	(Cys)Gly-Val-Ala-Val-Ser-Gly-Leu	
C51	379 Glu-Met-Ala-Gln-Asn	383	From P36, P36A
C52	384 Ser-Gln-Arg-Leu-Asn	389 (Trp)†	From T32C1, P36A
C52A	Ser-Gln-Arg-Leu		
C53	390 Thr-Gln-Ala-Glu-Val-Asp-Glu-Lys-Lys-Leu-Lys-Asp-Ile-Met	402	Gln(391) by t.l.c.; others from P38, SP30, SP31, T34
C54	403 Lys-Asn-Ala-Phe-Phe	407	From T35
C54A	Ala-Phe-Phe		
C56	412 Thr-Ala-Lys-Thr-Tyr	416	Charge -1
C57	417 Val-Glu-Ala-Ala-Glu-Gly-Glu-Leu-Pro	436 (Ser, Leu, Val, Ala, Gly, Ser, Asn, Ile, Ala, Gly, Phe)	Glu(418) by t.l.c.; others from P42, P44, SP33, SP33/34

Table 2 (continued)

Peptide	Sequence	Amide assignments
C57A		433 Ile-Ala-Gly-Phe
C58	437 Val-Lys-Val-Ala-Gln, Ala, Met)	By CP-A, AP-M
C58A	Val-Lys-Val-Ala-Gln	From T37, C59 Charge -1
C59	444 His-Asp-Gln-Gly-Asp-Trp	CN fragment (K. Blumenthal, personal communication)
C58/59	Val-Lys-Val-Ala-Gln-Ala (Met, His, Asp, Gln, Gly, Asp, Trp)	
C60	450 452 Ser-Lys-Asn	Charge +1; CP on intact chain

* Blocked N-terminus, unblocked by methanolic HCl treatment.

† Sequenced as a mixture with other known component; peptide C9 with C52, C15 with C30/31, C27/28/29 with C2 (which was also sequenced as a pure peptide).

only by ambiguous one-residue overlaps, and the sequence given is one of two equally plausible ones, even assuming that no tryptic peptides have been missed. The data of Blumenthal *et al.* (1975) on the CNBr fragment CN1 confirm the 77-78 overlap and this, together with the close agreement found between the amino acid composition of CN1 and those of its supposed component tryptic peptides, leaves only one consistent interpretation of this part of the chain. The overlap between tryptic peptides at residues 129-130 is only tenuously established by our data since it depends on the residual amino acid composition of the chymotryptic peptide C18 which was not sequenced right to the end. This overlap is firmly established by Blumenthal *et al.* (1975) through the sequencing of a staphylococcal proteinase peptide derived from fragment CN2. The sequence between T17 and T19, which we have recovered only in chymotryptic peptides C24 and C25, is overlapped definitively by the N-terminal sequence (residues 165-179) of fragment CN4 determined by Blumenthal *et al.* (1975). A further overlap which is not fully established by our data is that linking chymotryptic peptides C26 and C27 (residues 184-185). The insoluble tryptic peptide T20, which, on the basis of its amino acid composition, appears to cover this region, was found to have N-terminal Gly-Leu, but the rest of its sequence was not obtained. Fortunately, the overlap here is confirmed unequivocally by the sequence obtained by Blumenthal *et al.* (1975) in a tryptic derivative of fragment CN4.

The complete sequence derived from our data and those of Blumenthal *et al.* (1975) is presented in Fig. 1. The partial homology with bovine glutamate dehydrogenase has been discussed by Wootton *et al.* (1974) and Blumenthal *et al.* (1975). The probable location of two coenzyme-binding domains in the sequences of *Neurospora* and vertebrate glutamate dehydrogenases has been discussed by Wootton (1974). These putative domains were identified and located by a combination of secondary-structure predictions and sequence comparisons with other dehydrogenases of known structure.

We thank the Science and Medical Research Councils for financial support, and Margaret Burnley for her great help in performing the amino acid analyses. It is a pleasure to acknowledge the free exchange of information with Emil L. Smith and his colleagues at the University of California, Los Angeles, during the course of this work.

References

- Ashby, B., Wootton, J. C. & Fincham, J. R. S. (1974) *Biochem. J.* **143**, 317-329
- Blumenthal, K. M., Moon, K. & Smith, E. L. (1975) *J. Biol. Chem.* **250**, 3644-3654

- Coddington, A., Fincham, J. R. S. & Sundaram, T. K. (1966) *J. Mol. Biol.* **17**, 503–512
- Fincham, J. R. S. & Stadler, D. R. (1965) *Genet. Res.* **6**, 121–129
- Hirs, C. W. H. (1967) *Methods Enzymol.* **11**, 325–329
- Laursen, R. A. (1972) *Methods Enzymol.* **25**, 344–359
- Niederwieser, A. (1972) *Methods Enzymol.* **25**, 60–99
- Offord, R. E. (1966) *Nature (London)* **211**, 591–593
- Stadler, D. R. (1966) *Genet. Res.* **7**, 18–31
- Wootton, J. C. (1974) *Nature (London)* **252**, 542–546
- Wootton, J. C., Chambers, G. K., Holder, A. A., Baron, A. J., Taylor, J. G., Fincham, J. R. S., Blumenthal, K. M., Moon, K. & Smith, E. L. (1974) *Proc. Natl. Acad. Sci. U.S.A.* **71**, 4361–4365
- Wootton, J. C., Taylor, J. G., Jackson, A. A., Chambers, G. K. & Fincham, J. R. S. (1975a) *Biochem. J.* **149**, 739–748
- Wootton, J. C., Baron, A. J. & Fincham, J. R. S. (1975b) *Biochem. J.* **149**, 749–755