

Limited proteolysis and sequence analysis of the 2-oxo acid dehydrogenase complexes from *Escherichia coli*

Cleavage sites and domains in the dihydrolipoamide acyltransferase components

Leonard C. PACKMAN and Richard N. PERHAM*

Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.

The structures of the dihydrolipoamide acyltransferase (E2) components of the 2-oxo acid dehydrogenase complexes from *Escherichia coli* were investigated by limited proteolysis. Trypsin and *Staphylococcus aureus* V8 proteinase were used to excise the three lipoyl domains from the E2p component of the pyruvate dehydrogenase complex and the single lipoyl domain from the E2o component of the 2-oxoglutarate dehydrogenase complex. The principal sites of action of these enzymes on each E2 chain were determined by sequence analysis of the isolated lipoyl fragments and of the truncated E2p and E2o chains. Each of the numerous cleavage sites (12 in E2p, six in E2o) fell within similar segments of the E2 chains, namely stretches of polypeptide rich in alanine, proline and/or charged amino acids. These regions are clearly accessible to proteinases of M_r 24000–28000 and, on the basis of n.m.r. spectroscopy, some of them have previously been implicated in facilitating domain movements by virtue of their conformational flexibility. The limited proteolysis data suggest that E2p and E2o possess closer architectural similarities than would be predicted from inspection of their amino acid sequences. As a result of this work, an error was detected in the sequence of E2o inferred from the previously published sequence of the encoding gene, *sucB*. The relevant peptides from E2o were purified and sequenced by direct means; an amended sequence is presented.

INTRODUCTION

The 2-oxo acid dehydrogenase complexes from *Escherichia coli* share a common architectural design. The pyruvate dehydrogenase complex comprises a structural core (24-mer) of dihydrolipoamide acetyltransferase (E2p) subunits (EC 2.3.1.12), arranged with octahedral symmetry, which binds multiple copies of pyruvate decarboxylase [pyruvate dehydrogenase (lipoamide), E1p, EC 1.2.4.1] and dihydrolipoamide dehydrogenase (E3, EC 1.8.1.4, formerly EC 1.6.4.3). The corresponding enzymes in the 2-oxoglutarate dehydrogenase complex are 2-oxoglutarate decarboxylase [2-oxoglutarate dehydrogenase (lipoamide), E1o, EC 1.2.4.2], dihydrolipoamide succinyltransferase (E2o, EC 2.3.1.61) and the same E3 (Reed, 1974; Perham, 1983). The lipoyl-lysine residues, which act as swinging arms to transfer substrate between successive active sites in the mechanism, are located in segments of the E2p and E2o chains that protrude between the E1 and E3 subunits. These segments, in the form of lipoyl domains, can be selectively released from the complexes under suitably judicious conditions of limited proteolysis (Bleile *et al.*, 1979; Hale & Perham, 1979a; Perham & Roberts, 1981; Packman *et al.*, 1984). However, there is a significant difference between the structures of the E2p and E2o chains. The E2p chain appears able to carry up to three lipoyl-lysine residues (Hale & Perham, 1979b), in repeating sequences (Hale & Perham, 1980), whereas the E2o chain contains only one (White *et al.*, 1980). The amino acid sequences of the E2p and E2o chains have been inferred from the DNA sequences of their respect-

ive genes, *aceF* and *sucB* (Stephens *et al.*, 1983; Spencer *et al.*, 1984). The *N*-terminal half of the E2p chain is found to comprise three highly homologous lipoyl segments in tandem array, and these may be released as three individual functional domains by limited proteolysis of the complex with *Staphylococcus aureus* V8 proteinase (Packman *et al.*, 1984). The E2o chain bears only one such lipoyl domain, which may be removed by proteolysis with trypsin (Perham & Roberts, 1981).

The combined approaches of limited proteolysis and ^1H -n.m.r. spectroscopy have been successfully applied to demonstrate the existence of conformational mobility in both 2-oxo acid dehydrogenase complexes. The ^1H -n.m.r. spectrum of intact pyruvate dehydrogenase complex features sharp signals characteristic of segments of mobile polypeptide chain rich in alanine or threonine residues (Perham *et al.*, 1981). Cleavage of E2p by trypsin at Lys-316 (Packman *et al.*, 1984), leaving a residual complex comprising the inner E2p core to which the E1 and E3 subunits remain bound, removes many of the sharp signals (Perham *et al.*, 1981). For this and other reasons, a major source of these signals is thought to be the three long sequences in E2p, rich in alanine, proline and some charged amino acids, which link the lipoyl domains to each other and to the inner core (subunit-binding or catalytic) domain (Fig. 3) (Stephens *et al.*, 1983; Packman *et al.*, 1984; Graham *et al.*, 1986). In the 2-oxoglutarate dehydrogenase complex, similar polypeptide chain mobility is suggested by the presence of equally sharp resonances in the ^1H -n.m.r. spectrum, but removal of the lipoyl domain by treatment with trypsin has little effect on these sharp signals (Perham & Roberts, 1981). This

Abbreviations used: DABITC, 4-*NN*-dimethylaminoazobenzene 4'-isothiocyanate; DABTH, 4-*NN*-dimethylaminoazobenzene-4'-thiohydantoin.

* To whom correspondence should be addressed.

implies a cleavage site for trypsin, as yet unidentified, between the single lipoyl domain and the mobile segment in the E2o chain.

A comparison of the E2p and E2o chains would clearly be very fruitful, since the numbers and organization of the lipoyl domains in these chains are different, yet the basic mechanism of the two complexes remains the same. A knowledge of the proteolytic cleavage sites is essential if full benefit is to be derived from the $^1\text{H-n.m.r.}$ spectroscopy of both complexes, leading to more precise identification of the conformationally mobile regions of the respective E2 chains and of their possible contribution to the mechanism. We have therefore made a detailed study of the proteolysis of both complexes with *S. aureus* V8 proteinase and with trypsin, uncovering significant differences as well as important similarities.

In the course of this study, it became clear that the published amino acid sequence of E2o inferred from the sequence of the *sucB* gene (Spencer *et al.*, 1984) was inconsistent with our protein chemical data. We have determined the sequence of this region of the polypeptide chain by direct methods and an amendment to the inferred sequence is presented.

MATERIALS AND METHODS

Enzymes and reagents

The pyruvate dehydrogenase complex was isolated from a constitutive strain of *E. coli* K12 (Danson *et al.*, 1979). The 2-oxoglutarate dehydrogenase complex was isolated by a similar method from a derivative of *E. coli* K12 carrying the *aceEF-lpd* operon on a plasmid (pGS110 or pGS156); the chromosomal *sucAB* genes are expressed well in this strain (Guest *et al.*, 1985; Graham *et al.*, 1986). Pyroglutamate aminopeptidase and endoproteinase Lys-C were purchased from Boehringer Corp. and serum thymic factor was from Sigma Chemical Co. The sources of proteinases and sequencing reagents are given in Packman *et al.* (1984).

Enzyme assays

Whole-complex activity was measured by monitoring the reduction of NAD^+ in the presence of pyruvate or 2-oxoglutarate as substrate (Danson & Perham, 1976).

Polyacrylamide-gel electrophoresis

Analytical polyacrylamide-gel electrophoresis, in the presence and in the absence of SDS, was carried out with the gel and buffer systems described by Packman *et al.* (1984). Gel compositions are given by the parameters T and C, where T is the final concentration of acrylamide monomer and C is the percentage cross-linking. For consistency, all apparent molecular-mass values quoted are those derived from electrophoresis in gels buffered with Tris/glycine and may not match exactly the previously reported values derived from phosphate-buffered gels.

Preparative SDS/polyacrylamide-gel electrophoresis was carried out in slabs (7.5% T, 2.6% C; 14 cm \times 14 cm \times 1.5 mm) with a continuous buffer system of 0.1 M-sodium phosphate, pH 6.7, containing 0.1% SDS. Electrode buffers were supplemented with 2 mM-2-mercaptoacetic acid as a free-radical scavenger (Packman *et al.*, 1984). Each sample was loaded across the full width of the gel. After electrophoresis, protein

bands were detected by immersing the gel in 1 M-KCl (Walker *et al.*, 1980), and the relevant areas were excised and soaked overnight at room temperature in 50 mM- NaHCO_3 containing 5% SDS. Electro-elution of the proteins into dialysis bags was carried out at 200 V for 24 h in 50 mM- NaHCO_3 containing 2 mM-2-mercaptoacetic acid and 0.1% SDS. The bags were dialysed against several changes of 10% (v/v) ethanol, and each protein was then freeze-dried, washed with acetone/water (4:1, v/v) at 0 °C and dried. It was dissolved in anhydrous trifluoroacetic acid and clarified by centrifugation, and a portion (about 2 nmol) was transferred to an acid-washed Dreyer tube for sequence analysis.

Amino acid and N-terminal sequence analysis

Amino acid analysis was performed as described by Perham (1978) with an LKB 4400 analyser. Manual sequencing of proteins, by using the DABITC method (Chang, 1983), was carried out as described previously (Packman *et al.*, 1984) with one modification for proteins larger than 20 kDa: before each coupling round with DABITC, the protein was dissolved in 20 μl of trifluoroacetic acid, which was immediately removed under vacuum. This extra step greatly enhanced the solubility of the protein in the coupling buffer. DABTH-Ile and DABTH-Leu were not distinguished.

Preparation of lipoyl domains

Lipoyl domains were excised and purified from the 2-oxo acid dehydrogenase complexes as follows, based on the method of Packman *et al.* (1984). The appropriate 2-oxo acid dehydrogenase complex (5–10 mg, 10 mg/ml in 20 mM-sodium phosphate buffer, pH 7, containing 2.7 mM-EDTA and 0.02% NaN_3) was digested at 0 °C with 0.5% (w/w) *S. aureus* V8 proteinase or 0.2% (w/w) trypsin until the whole-complex activity had declined to less than 5% of its initial value. Digestion with *S. aureus* V8 proteinase was terminated by addition of 5 mg of α_2 -macroglobulin (Boehringer Corp.). Trypsin was inactivated by adding a 5-fold molar excess of soya-bean trypsin inhibitor.

The lipoyl domains were recovered from the digest by gel filtration on a TSK3000SW column (60 cm \times 0.75 cm) at room temperature in 50 mM-sodium phosphate buffer, pH 7. Fractions containing lipoyl domains were identified by SDS/polyacrylamide-gel electrophoresis, with *N*-ethyl [2,3- ^{14}C]maleimide-labelled lipoyl domain (Packman *et al.*, 1984) as a marker, and pooled. The solution of lipoyl domain was diluted 4-fold with 20 mM- NH_4HCO_3 (buffer A), and individual domains were separated by fast protein liquid chromatography on Mono Q resin with successive linear gradients: 0–15% (v/v) buffer B over 5 min, then 15–45% (v/v) buffer B over 50 min, where buffer B is 1 M- NH_4HCO_3 . Detection was at 230 nm. Pure samples of each domain were freeze-dried ready for amino acid and sequence analysis.

Preparation of E2p and E2o inner cores (residual E2p and E2o chains)

Separate samples of pyruvate dehydrogenase complex (2–5 mg, 10 mg/ml in 20 mM-sodium phosphate/EDTA/azide buffer, pH 7) were digested with *S. aureus* V8 proteinase (0.5%, w/w) for 2 h at 0 °C or at 37 °C for 1 h. The incubation conditions were chosen to maximize the yield of products of apparent mass 37.2 kDa and 34.5 kDa (generated at 0 °C) and 29 kDa (generated

more rapidly at 37 °C). The digestion was terminated by addition of an equal volume of double-strength SDS/polyacrylamide-gel-electrophoresis sample buffer at 100 °C. After 5 min at 100 °C, the proteins were separated by preparative SDS/polyacrylamide-gel-electrophoresis and prepared for amino acid sequence analysis as described above. The 2-oxoglutarate dehydrogenase complex was treated in a similar way with 0.2% (w/w) trypsin at 0 °C for 35 min to generate the 37.3 kDa fragment of E2o or with 0.5% (w/w) *S. aureus* V8 proteinase for 2 h at 0 °C to generate the 31.6 kDa fragment.

An alternative starting point for the preparation of some of these proteins was the excluded material from the gel-filtration step in the lipoyl domain preparation (see the preceding subsection).

All proteins gave unambiguous *N*-terminal sequences, except for the 37.3 kDa tryptic fragment from E2o, whose *N*-terminus appeared to be blocked (see the Results section). A different purification strategy was devised to circumvent this problem, as follows. After the initial treatment with trypsin, the 2-oxoglutarate dehydrogenase complex was gel-filtered on the TSK3000SW column to remove excised fragments, and the residual complex was subjected to secondary digestion with *S. aureus* V8 proteinase. This material was gel-filtered again. The peptide from E2o released by the *S. aureus*-V8-proteinase digestion, whose *N*-terminus was formed by the initial tryptic digestion, was recovered from the 230 nm-absorbing material present towards the included volume (V_i) of the TSK3000SW column and purified for sequence analysis by reverse-phase (C_{18}) h.p.l.c. in

0.1% trifluoroacetic acid with acetonitrile as organic modifier.

RESULTS

Lipoyl domains of E2p

S. aureus V8 proteinase cuts E2p in the native pyruvate dehydrogenase complex at Glu-92, Glu-192 and Glu-292 to release the lipoyl domains LpV1, LpV2 and LpV3 (referred to as L1, L2 and L3 respectively in Packman *et al.*, 1984). The *C*-terminus of each of these domains has subsequently been established by isolation of *C*-terminal peptides from complete *S. aureus*-V8-proteinase digests of the purified domains (L. C. Packman, unpublished work). LpV1 therefore comprises residues 1–92, LpV2 comprises residues 93–192 and LpV3 comprises residues 193–292.

Cleavage of native E2p with trypsin (see the Materials and methods section) generated five lipoyl fragments separable by fast protein liquid chromatography (LpTa–LpTe; Fig. 1), all of whose *N*-terminal sequences were determined (Table 1). Four of the five fragments (LpTb–LpTe) evidently arise from partial cleavage at Lys-90 and Lys-91, generating four differently charged species (residues 1–90, 1–91, 91–190 and 92–190) from two lipoyl domains. Fragments LpTc and LpTd were poorly resolved by fast protein liquid chromatography. The sequence analysis was therefore carried out on the mixture of these two components, but the results were entirely consistent with the presence of two peptide fragments beginning with residues 1 and 92. The

Table 1. Fragments of E2p generated by limited proteolysis with trypsin and *S. aureus* V8 proteinase

Lipoyl fragments released by trypsin					
Fragment	<i>N</i> -Terminal sequence	Position in sequence*			
LpTa	QEAAP-	191–301			
LpTb	KEAAP-	91–190			
LpTc	EAAPA-	92–190			
LpTd	$\begin{matrix} I & I \\ A_L & X_L \end{matrix}$ -	1–91			
LpTe	$\begin{matrix} I & I \\ A_L & E_L \end{matrix}$ -	1–90			
Inner-core fragments					
Proteinase† used to generate fragment	Apparent molecular mass of fragment (kDa)	<i>N</i> -Terminal sequence	Position in sequence*	Computed molecular mass (Da)	
V8	37.2	AAAPAPA-	293–629	36457	
V8	34.5	FAENDXY-	319–629	34158	
V8	29.0	AAPAATG-	373–629	28050	
T	36.5‡	—	302–629	35708	
T	35.5§	—	313–629	34759	
T	35.0	—	317–629	34374	
T	29.0	RAEAAPA-	370–629	28406	

* The *C*-terminus of each fragment is inferred (see the text).

† Abbreviations: T, trypsin; V8, *S. aureus* V8 proteinase.

‡ 39 kDa on phosphate-buffered gels (Bleile *et al.*, 1979; Hale & Perham, 1979a).

§ 36 kDa on phosphate-buffered gels (Bleile *et al.*, 1979; Hale & Perham, 1979a).

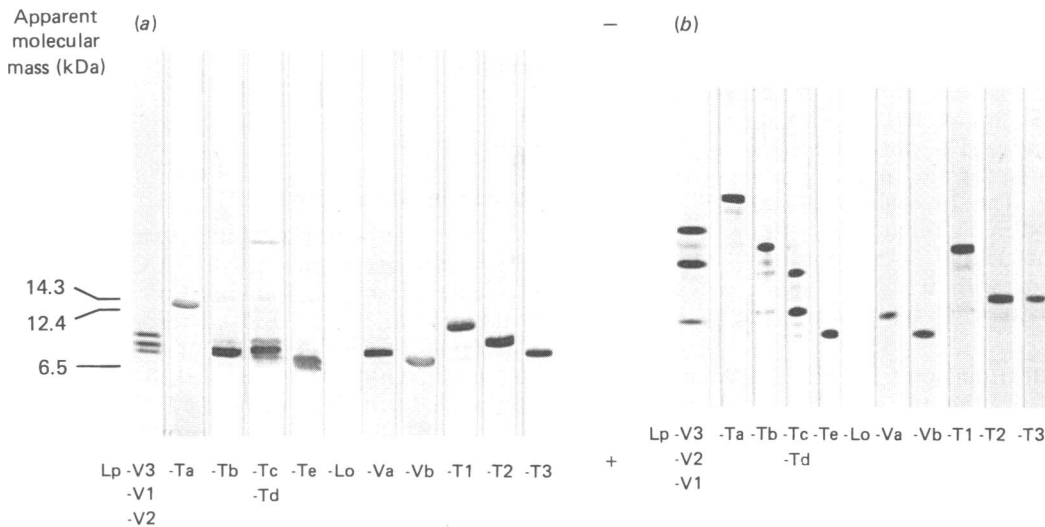


Fig. 1. Polyacrylamide-gel electrophoresis of the lipoyl fragments of E2p and E2o

The lipoylated fragments of E2p and E2o were prepared from trypsin and *S. aureus* V8 proteinase digests of the relevant 2-oxo acid dehydrogenase complex and separated by fast protein liquid chromatography. Samples of each preparation were examined by polyacrylamide-gel electrophoresis (27%T, 0.3%C) in the presence (a) and in the absence (b) of SDS.

C-terminal residue of each fragment was not determined directly. However, it could be inferred in part from the amino acid composition data but also from the relative mobilities of the fragments during SDS/polyacrylamide-gel electrophoresis, coupled with their respective charge characteristics inferred from non-denaturing polyacrylamide-gel electrophoresis and chromatography on Mono Q resin.

The fifth lipoyl domain, LpTa, was recovered in 4 times the yield of the other fragments, and seemed to be much larger as judged by SDS/polyacrylamide-gel electrophoresis (Fig. 1). Fragment LpTa was resistant to further tryptic digestion, but was readily cleaved with *S. aureus* V8 proteinase to release a peptide whose amino acid composition corresponded to residues 293–301. This peptide must derive from the C-terminus of fragment

LpTa, which therefore comprises residues 191–301 of E2p. Prolonged digestion of fragment LpTa with *S. aureus* V8 proteinase additionally removed the N-terminal dipeptide, Gln-Glu, to generate, as expected, a fragment (residues 193–292) with electrophoretic and chromatographic properties identical with those of the lipoyl domain LpV3 (Packman *et al.*, 1984). The peptides excised from fragment LpTa by *S. aureus* V8 proteinase (total molecular mass of peptides removed 1 kDa) contribute some 3.6 kDa to the apparent molecular mass of the lipoyl domain, as judged by SDS/polyacrylamide-gel electrophoresis. This curious ability of an (alanine + proline)-rich region of the E2p chain to cause anomalously slow migration during SDS/polyacrylamide-gel electrophoresis has already received comment (Guest *et al.*, 1985; Graham *et al.*, 1986).

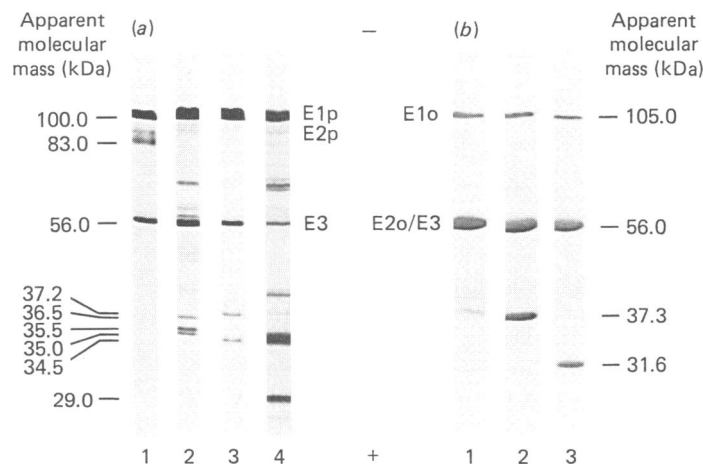


Fig. 2. Limited proteolysis of the 2-oxo acid dehydrogenase complexes

Samples of 2-oxo acid dehydrogenase complex were proteolysed at 0 °C as described in the text, and the products were separated by SDS/polyacrylamide-gel electrophoresis (20%T, 0.3%C). Staining was with Coomassie Brilliant Blue R. (a) Pyruvate dehydrogenase complex. Lane 1, untreated complex; lane 2, complex treated with trypsin; lane 3, complex treated with *S. aureus* V8 proteinase; lane 4, complex treated with *S. aureus* V8 proteinase at 37 °C. Lane 4 was additionally stained with silver to enhance detection of the 34.5 kDa and 29 kDa bands. (b) 2-Oxoglutarate dehydrogenase complex. Lane 1, untreated complex; lane 2, complex treated with trypsin; lane 3, complex treated with *S. aureus* V8 proteinase.

Inner-core domains of E2p

Two inner-core fragments of E2p (apparent molecular masses 37.2 and 34.5 kDa) could be generated by digestion of the pyruvate dehydrogenase complex with *S. aureus* V8 proteinase at 0 °C, and a smaller inner-core fragment (apparent molecular mass 29.0 kDa) could be obtained by prolonged digestion at 0 °C or more readily if the temperature of the digestion was raised to 37 °C (Fig. 2). These three fragments were purified and their *N*-terminal sequences were determined (Table 1). The cleavage sites were identified as Glu-292, Glu-318 and Glu-372.

Similarly, three inner-core fragments of E2p (apparent molecular masses 36.5, 35.5 and 35.0 kDa) could be generated by digesting the complex with trypsin (Fig. 2). The three tryptic fragments of E2p are clearly intermediate in size between the 37.2 kDa and 34.5 kDa fragments of E2p generated by digestion with *S. aureus* V8 proteinase through cuts at Glu-292 and Glu-318. Given the specificity of trypsin, the cleavage sites for this enzyme must therefore be Lys-301, Lys-312 and, as already demonstrated by Packman *et al.* (1984), Lys-316. Prolonged tryptic digestion of intact pyruvate dehydrogenase complex, or isolated E2p, produces a 29 kDa fragment (Packman *et al.*, 1984). This fragment was purified from a digest, performed in 20 mM buffer, of isolated E2p (Packman *et al.*, 1984) or E2p-E3 subcomplex (S. Radford, unpublished work). At this low ionic strength the desired product formed a precipitate, and was recovered by centrifugation. The *N*-terminal sequence of this peptide was also determined (Table 1), establishing the cleavage site as Lys-369, just three residues to the *N*-terminal side of the comparable *S. aureus*-V8-proteinase-cleavage site, Glu-372. A second sequence (AEAAPAA-) was detected in the sequence analysis of the tryptic peptide, indicative of a secondary tryptic cleavage at Arg-370. This may be due to trypsin beginning to release the *N*-terminal arginine residue after the primary cleavage has occurred at Lys-369.

The *C*-terminus of each inner-core fragment of E2p was taken as Met-629. This was inferred from the observation (Hale & Perham, 1979a) that E2p is not cleaved by trypsin when the lysine side chains are blocked by amidation; the five arginine residues in the *C*-terminal 40 residues of the protein are probably unavailable to proteinase. As there are no lysine or glutamic acid residues in this area of E2p, cleavage at the *C*-terminus of E2p by trypsin or *S. aureus* V8 proteinase is unlikely.

Lipoyl domains of E2o

Treatment of the native 2-oxoglutarate dehydrogenase complex with *S. aureus* V8 proteinase released two lipoylated fragments, LoVa and LoVb, whereas the action of trypsin released three such fragments, LoT1, LoT2 and LoT3 (Fig. 1). In each case, prolonged exposure to the proteinase degraded the larger fragments to a single product, LoVb and LoT3 respectively. All the fragments were found to possess an *N*-terminal sequence (SSVD-) identical with that of intact E2o. Their amino acid compositions were also compared with the amino acid sequence of E2o inferred from the DNA sequence of the *sucB* gene (Spencer *et al.*, 1984). On the basis of the known specificities of trypsin and *S. aureus* V8 proteinase, the best fit of the composition of each

fragment to the potential amino acid sequence was used to determine the position of each *C*-terminus. For fragments LoVa, LoVb, LoT2 and LoT3, the compositions precisely fitted those expected for the polypeptide chains released by cleavage at Glu-92, Glu-85, Lys-93 and Lys-89 respectively, as shown in Table 2.

However, even allowing for the sometimes unusual specificity of trypsin, no satisfactory fit could be achieved for fragment LoT1, the best fit being to residues 1-117 of E2o with a correlation coefficient of 0.969. Repurification of the fragment on Mono Q resin brought about no improvement. We therefore suspected an error in the inferred sequence of E2o. This possibility was investigated by digesting fragment LoT1 with *S. aureus* V8 proteinase, which generated a mixture of fragments with the electrophoretic properties of fragments LoVa and LoVb. The smaller peptides released from fragment LoT1 by this process were purified by reverse-phase (C₁₈) h.p.l.c. The major peptide recovered had the amino acid composition TSEPA₂KR, and its sequence, established by complete degradation, was found to be KASTPAQR; the purified minor peptides were identified as degradation products of LoVa. This sequence corresponds with that proposed for residues 93-100 of E2o by Spencer *et al.* (1984), but with residue-100 (alanine) replaced by arginine. With this amendment to the sequence, the amino acid composition of fragment LoT1 precisely fitted residues 1-100 of E2o (Table 2).

Inner-core domains of E2o

An inner-core fragment of E2o (apparent molecular mass 31.6 kDa) was generated by digestion of the 2-oxoglutarate dehydrogenase complex with *S. aureus* V8 proteinase, whereas a comparable but slightly larger fragment (apparent molecular mass 37.3 kDa) was generated by the action of trypsin (Fig. 2). The *N*-terminal sequence of the 31.6 kDa fragment was successfully obtained, identifying its *N*-terminus as Ser-157 (Table 2). However, the 37.3 kDa fragment presented a significant problem. Successive preparations of the fragment failed to yield a satisfactory *N*-terminal sequence, each giving a low yield of DABTH-Gln for the *N*-terminus. Further cycles provided no clearly identifiable DABTH-amino acids. An *N*-terminal pyroglutamic acid residue was suspected, but precautionary measures such as the exclusion of acid steps in the preparation of the fragment had no beneficial effect. The blocked *N*-terminus was similarly unaffected by treatment with pyroglutamate aminopeptidase under conditions that allowed efficient deblocking of serum thymic factor.

An alternative approach to the problem was therefore sought, as follows. The residual 2-oxoglutarate dehydrogenase complex left after treatment with trypsin was submitted to a secondary digest with *S. aureus* V8 proteinase, and the peptide released from the already shortened E2o chain was purified by h.p.l.c., all as described in the Materials and methods section. The *N*-terminus of this peptide must have been formed by the initial tryptic digestion, but this time the *N*-terminal sequence was readily found to be QQAS_L^IEEQN-.

Although the yield of DABTH-Gln was again low compared with that found during the sequencing of an equivalent amount (1 nmol) of the peptide KASTPAQR (see above), the continued clear signal and absence of background allowed unambiguous determination of the

Table 2. Fragments of E2o generated by limited proteolysis with trypsin and *S. aureus* V8 proteinase

Lipoyl fragments					
Proteinase* used to generate fragment	Fragment	N-Terminal sequence	Position in sequence	Fit of fragment composition with expected composition from inferred sequence of E2o (correlation coefficient)	
V8	LoVa	SSVD ^I _L -	1-92	0.992	
V8	LoVb	SSVD ^I _L -	1-85	0.999	
T	LoT1	SSVD-	1-100†	0.997	
T	LoT2	SSVD-	1-93	0.998	
T	LoT3	SSVD-	1-89	0.996	

Inner-core fragments					
Proteinase* used to generate fragment	Apparent molecular mass of fragment (kDa)	N-Terminal sequence	Position in sequence†	Computed molecular mass (Da)	
T	37.3	Q-	101-404‡	33329	
V8	31.6	SAPAA(A)AP-	157-404‡	27329	

* Abbreviations: T, trypsin; V8, *S. aureus* V8 proteinase.
† Amended sequence as presented in Fig. 3.
‡ The C-terminus of each fragment is inferred (see the text).

sequence, which could be matched to that proposed for E2o (Spencer *et al.*, 1984), with some changes:

E2o sequence, residues 93-108 (Spencer *et al.*, 1984) -KASTPAQAAASLEEQN-
Peptides identified in the present study -KASTPAQR



Given the specificity of trypsin, these data suggested a juxtaposition of the sequence QQASLEEQN- with the C-terminus (-KASTPAQR) of fragment LoT1. The sequence QQASLEEQN- was therefore thought likely to be that at the N-terminus of the 37.3 kDa tryptic fragment of E2o. In support of this view, the amino acid composition of the 37.3 kDa fragment was found to fit closely with that calculated from the revised sequence of the E2o chain (correlation coefficient 0.997). This close fit also provided evidence that no significant cleavage of E2o by trypsin had occurred at the C-terminus of the protein. Equally, no peptides were found in the trypsin/*S. aureus* V8 proteinase double-digest experiment that corresponded to the C-terminal sequence of E2o. It therefore seems likely that this region of the protein is unavailable to the proteinase.

Correction to the amino acid sequence of E2o

In order to obtain definitive evidence for the proposed juxtaposition of fragment LoT1 and the 37.3 kDa inner-core fragment in intact E2o, the sequential digest experiment was repeated, but with endoproteinase Lys-C (which cleaves at lysine but not at arginine residues) in place of trypsin. This enzyme generated an initial E2o inner-core fragment of apparent molecular mass 38.0 kDa, as judged by SDS/polyacrylamide-gel electrophoresis. The peptide released by the subsequent *S. aureus* V8 proteinase digest was purified as before and found to possess the N-terminal sequence ASTPAQRQQAS-, confirming the suspected continuity

of sequence across this region. Endoproteinase Lys-C had evidently cleaved E2o at Lys-93, as predicted from the cleavage sites found for trypsin (Fig. 3).

The revision to the E2o sequence therefore involves a change of two amino acid residues and the addition of a third. This requires a renumbering of the residues in E2o beyond Gln-99, and the amended sequence is shown in Fig. 3, together with an indication of all the proteinase-cleavage sites in both the E2p and E2o chains. The revised molecular mass for E2o is computed as 43877 Da, excluding lipoic acid.

In order to check that the proposed change of sequence was not the product of mutations induced at this site during genetic manipulation of the host organism, the 2-oxoglutarate dehydrogenase complex was purified from *E. coli* Crookes strain (Reed & Mukherjee, 1969) and subjected to limited proteolysis with trypsin and *S. aureus* V8 proteinase. The pattern of proteolytic fragments after polyacrylamide-gel electrophoresis in the presence and in the absence of SDS was the same as that produced by similar treatment of the 2-oxoglutarate dehydrogenase complex from *E. coli* K12 carrying plasmids pGS110 or pGS156 (Guest *et al.*, 1985; Graham *et al.*, 1986), confirming the tryptic site (i.e. Arg-100) to be present in both proteins.

DISCUSSION

When the pyruvate dehydrogenase and 2-oxoglutarate dehydrogenase multienzyme complexes are exposed to

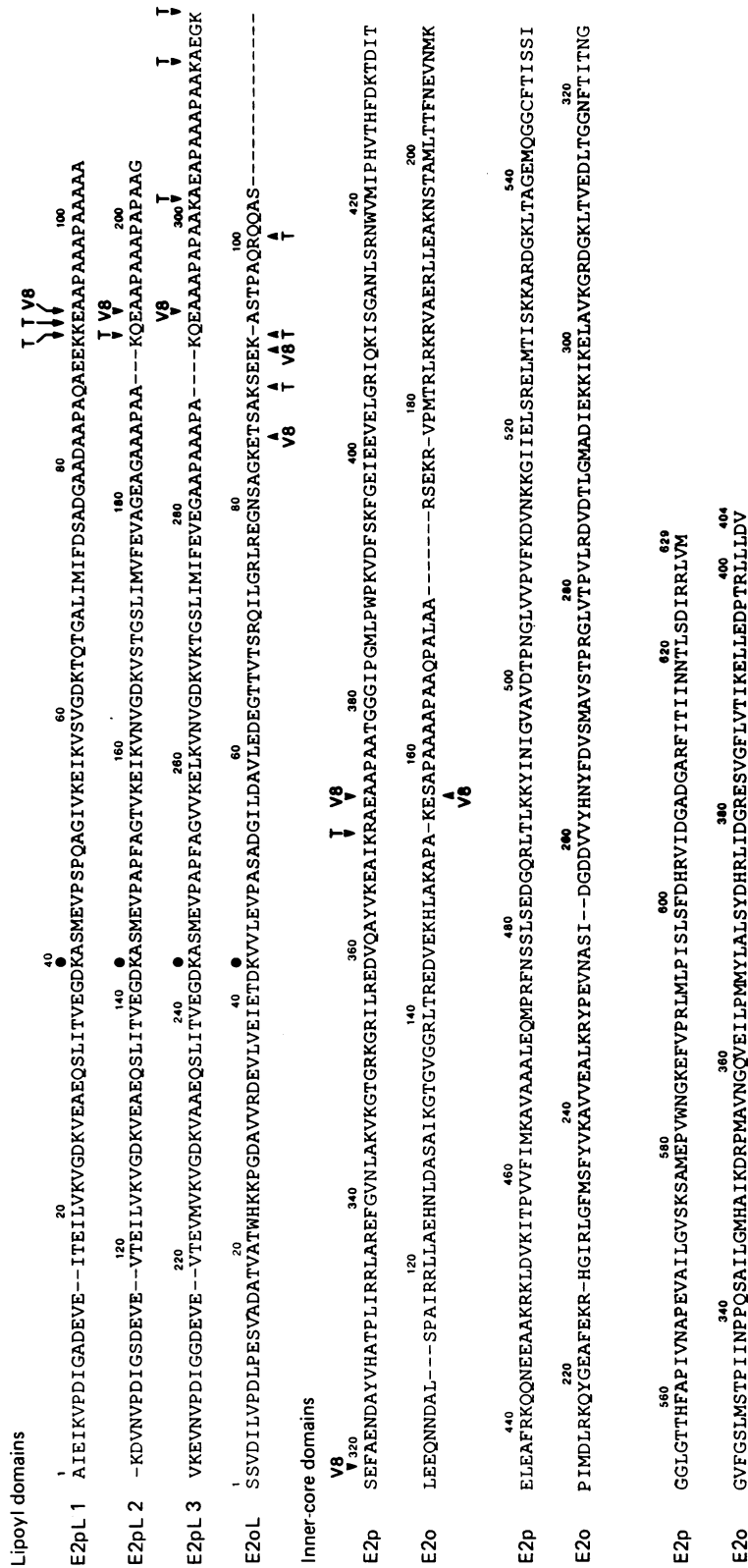


Fig. 3. Cleavage sites for proteinases in the E2p and E2o polypeptide chains

The amino acid sequences of E2p and E2o are aligned for maximum sequence homology (Spencer *et al.*, 1984). The E2o sequence contains the amendments reported in the present paper. The cleavage sites of trypsin (T) and *S. aureus* V8 proteinase (V8) are shown. ●, Site of lipoylation of lysine residues.

trypsin or *S. aureus* V8 proteinase, a series of cleavages takes place in the E2p and E2o chains, a full catalogue of which is now available from the results of the present study (Fig. 3). The cleavage sites are found in common regions of amino acid sequence in E2p and E2o that must be accessible to proteinases of molecular mass 24–28 kDa. This undoubtedly reflects a similar structural organization of the two proteins. In E2p the cleavage sites for trypsin and *S. aureus* V8 proteinase are generally within a few residues of each other, and, significantly, are mostly contained within the three long (alanine+proline)-rich regions of polypeptide chain implicated in conformational mobility (Packman *et al.*, 1984; Graham *et al.*, 1986). One puzzling feature is the inability of trypsin to cleave at Lys-290 either in intact E2p or in the excised LpTa fragment (residues 191–301). This residue is contained in the sequence -AKQEAA-, exactly duplicating the sequence surrounding Lys-190, which is a target for trypsin. Both Glu-192 and Glu-292 are cleavage sites for *S. aureus* V8 proteinase. Despite the sequence homology between these two regions, their different susceptibilities to trypsin may point to a significant difference in structure between these two segments of polypeptide chain, a difference that is retained even in the excised LpTa fragment.

At least one additional section of E2p is accessible to proteinases, albeit poorly, as evidenced by the slow cleavages at Lys-369 and Glu-372 (Fig. 3). In E2o, however, *S. aureus* V8 proteinase cleaves readily at the homologous site (Glu-156), although trypsin is without effect during reasonable incubation times. The significance of this is unclear, but it is conceivable that this region of the E2o core in the 2-oxoglutarate dehydrogenase complex may present a more open structure to the proteinase than does the corresponding region of E2p in the pyruvate dehydrogenase complex.

These results may be significant in terms of conformational mobility in the E2 chains. Removal of the lipoyl domains from the 2-oxoglutarate dehydrogenase complex by treatment with trypsin causes little diminution in the abundant sharp ¹H-n.m.r. signals observed in the 270 MHz spectrum of the complex, many of which are attributable to alanine (or threonine) residues (Perham & Roberts, 1981). Given the amendment to the sequence of E2o and the identification of the tryptic-cleavage site at Arg-100 described here, it is tempting to attribute some at least of the sharp ¹H-n.m.r. signals to the distinct (alanine+proline)-rich segment of polypeptide chain at approximately residues 150–172 in E2o (Fig. 3). This sequence encompasses the cleavage site (Glu-156) identified for *S. aureus* V8 proteinase, which would be consistent with free exposure to solvent and the possibility of conformational mobility implied by the sharp resonances in the ¹H-n.m.r. spectrum.

Similarly, there is evidence from ¹H-n.m.r. spectroscopy for some polypeptide chain flexibility remaining in the residual pyruvate dehydrogenase complex that is formed by release of the E2p lipoyl domains by tryptic cleavage at Lys-316 (Perham *et al.*, 1981; Packman *et al.*, 1984). In view of the existence in E2p of cleavage sites for trypsin and *S. aureus* V8 proteinase at Lys-369 and Glu-372 respectively, it is conceivable that the E2p chain in this region can also enjoy some conformational mobility. There is a small (alanine+proline)-rich sequence on the C-terminal side of Lys-369 and Glu-372,

together with some nearby charged side chains, but this type of sequence is not as pronounced as it is in the corresponding region of E2o (Fig. 3). Conversely, alanine and proline residues are not conspicuous in the region of E2o (approximately residues 82–104) encompassing the major set of trypsin- and *S. aureus*-V8-proteinase-cleavage sites, unlike the three corresponding regions of E2p (Fig. 3).

Thus it may well be that the differences in structure between the E2o and E2p chains are less marked than their respective amino acid sequences would superficially suggest, particularly in those regions connecting the lipoyl domain(s) to the inner core. The distribution of charged side chains, mostly those of glutamic acid and lysine residues in these particular sequences (Fig. 3), is probably important for the accessibility and potential flexibility of the polypeptide chain. These ideas are of course very tentative, but they can now be tested by more detailed ¹H-n.m.r. experiments and by the direct methods of protein engineering (Graham *et al.*, 1986).

We are grateful to the Science and Engineering Research Council for financial support. We thank Professor J. R. Guest and his colleagues for discussion of their predicted amino acid sequence of E2o and Mr. John Lester for skilled assistance with the amino acid analyser.

REFERENCES

- Bleile, D. M., Munk, P., Oliver, R. M. & Reed, L. J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4385–4389
- Chang, J.-Y. (1983) *Methods Enzymol.* **91**, 455–466
- Danson, M. J. & Perham, R. N. (1976) *Biochem. J.* **159**, 677–682
- Danson, M. J., Hale, G., Johnston, P., Perham, R. N., Smith, J. & Spragg, P. (1979) *J. Mol. Biol.* **129**, 603–617
- Graham, L. D., Guest, J. R., Lewis, H. M., Miles, J. S., Packman, L. C., Perham, R. N. & Radford, S. E. (1986) *Philos. Trans. R. Soc. London Ser. A* **317**, 391–404
- Guest, J. R., Lewis, H. M., Graham, L. D., Packman, L. C. & Perham, R. N. (1985) *J. Mol. Biol.* **185**, 743–754
- Hale, G. & Perham, R. N. (1979a) *Eur. J. Biochem.* **94**, 119–126
- Hale, G. & Perham, R. N. (1979b) *Biochem. J.* **177**, 129–137
- Hale, G. & Perham, R. N. (1980) *Biochem. J.* **187**, 905–908
- Packman, L. C., Hale, G. & Perham, R. N. (1984) *EMBO J.* **3**, 1315–1319
- Perham, R. N. (1978) in *Techniques in Protein and Enzyme Biochemistry* (Kornberg, H. L., Metcalfe, J. C., Northcote, D. H., Pogson, C. I. & Tipton, K. F., eds.), vol. B110, pp. 1–39, Elsevier/North-Holland, Amsterdam
- Perham, R. N. (1983) in *Frontiers in Biochemical and Biophysical Studies of Proteins and Membranes* (Liu, T.-Y., Sakakibara, S., Schechter, A. N., Yagi, K., Yajima, H. & Yasunobu, K., eds.), pp. 249–258, Elsevier, New York
- Perham, R. N. & Roberts, G. C. K. (1981) *Biochem. J.* **199**, 733–740
- Perham, R. N., Duckworth, H. W. & Roberts, G. C. K. (1981) *Nature (London)* **292**, 474–477
- Reed, L. J. (1974) *Acc. Chem. Res.* **7**, 40–46
- Reed, L. J. & Mukherjee, B. B. (1969) *Methods Enzymol.* **13**, 55–61
- Spencer, M. E., Darlison, M. G., Stephens, P. E., Duckenfield, I. K. & Guest, J. R. (1984) *Eur. J. Biochem.* **141**, 361–374
- Stephens, P. E., Darlison, M. G., Lewis, H. M. & Guest, J. R. (1983) *Eur. J. Biochem.* **133**, 481–489
- Walker, J. E., Auffret, A. D., Carne, A. F., Naughton, M. A. & Runswick, M. J. (1980) in *Methods in Peptide and Protein Sequence Analysis* (Birr, C., ed.), pp. 257–266, Elsevier/North-Holland, Amsterdam
- White, R. H., Bleile, D. M. & Reed, L. J. (1980) *Biochem. Biophys. Res. Commun.* **94**, 78–84