Amino acid sequence analysis of the lipoyl and peripheral subunit-binding domains in the lipoate acetyltransferase component of the pyruvate dehydrogenase complex from *Bacillus stearothermophilus*

Leonard C. PACKMAN, Adolfo BORGES and Richard N. PERHAM

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

The pyruvate dehydrogenase multienzyme complex from *Bacillus stearothermophilus* comprises a structural core, composed of 60 dihydrolipoamide acetyltransferase (E2p) subunits, which binds multiple copies of pyruvate decarboxylase (E1p) and dihydrolipoamide dehydrogenase (E3) subunits. After limited proteolysis with chymotrypsin, the *N*-terminal lipoyl domain of E2p was excised, purified and sequenced. The residual complex, which remained assembled, was then digested with trypsin under mild conditions. This treatment promoted complete disassembly of the complex and the various components were separated by gel filtration and h.p.l.c. A folded fragment of E2p containing about 50 amino acid residues was identified as being responsible for binding the E3 subunits, although, unlike the corresponding region of the E2p or E2o chains of the pyruvate dehydrogenase or 2-oxoglutarate dehydrogenase complexes from *Escherichia coli*, the fragment also bound E1p molecules. Further peptide purification and sequence analysis allowed the determination of the first 211 amino acid residues of the *B. stearothermophilus* E2p chain, thus providing the complete primary structure of the lipoyl domain, the E1p/E3-binding domain and the regions of polypeptide chain, probably highly flexible in nature, that link the domains to each other and to the inner-core (E2p-binding) domain. Several of the proteolytically sensitive sites were also identified. The sequence of the *B. stearothermophilus* E2p chain shows close homology with the sequences of the E2p and E2o chains from *E. coli*, although significant differences in structure are apparent. Detailed evidence for the sequence of the peptides obtained by limited proteolysis and further chemical and enzymic cleavages have been deposited as Supplementary Publication SUP 50142 (11 pages) at the British Library Lending Division, Boston Spa, Wetherby, West Yorkshire LS23 6BQ, U.K., from whom copies may be obtained as indicated in Biochem. J. (1988) 249, 5.

INTRODUCTION

The 2-oxo acid dehydrogenase multienzyme complexes from bacterial and mammalian sources catalyse the oxidative decarboxylation of 2-oxo acids (pyruvate, 2-oxoglutarate and branched-chain 2-oxo acids) to provide the corresponding acyl-CoA and NADH. All such complexes studied to date share a common architectural design, comprising a multisubunit core enzyme, dihydrolipoamide acyltransferase (E2), which binds the peripheral subunits of 2-oxo acid decarboxylase (E1) and dihydrolipoamide dehydrogenase (E3) [for reviews see Reed (1974), Yeaman (1986) and Perham *et al.* (1987)]. In the pyruvate dehydrogenase complex from *Escherichia coli* the 24-subunit (octahedral) dihydrolipoamide acetyltransferase component (E2p, EC 2.3.1.12) is exquisitely sensitive to cleavage by a variety of proteinases under non-denaturing conditions, as is the corresponding octahedral dihydrolipoamide succinyltransferase (E2o, EC 2.3.1.61) component of the 2-oxoglutarate dehydrogenase complex from the same organism. Various experiments on each of these enzymes have shown that discrete segments of the E2 molecule can be excised under mild conditions, suggesting the existence of

structural domains separated by proteolytically sensitive sites. A knowledge of the amino acid sequence of each E2 component (Stephens *et al.*, 1983; Spencer *et al.*, 1984; Packman & Perham, 1987), combined with the determination of the cleavage sites of trypsin and *Staphylococcus aureus* V8 proteinase (Packman *et al.*, 1984a; Packman & Perham, 1986, 1987), has allowed the approximate limits of each domain to be mapped. From this work, a model has been derived of the domain structure of these enzymes (Packman & Perham, 1986).

The E2p polypeptide chain (66 kDa) of *E. coli* embodies, from *N*- to *C*-terminus, three highly homologous lipoyl domains, an E3-binding domain and an E2p subunit-binding domain (which also houses the acetyltransferase active site). Each domain is separated from its neighbour by stretches of polypeptide chain rich in alanine, proline and charged amino acids. These regions are the source of unexpectedly sharp resonances observed in the ¹H-n.m.r. spectrum of the intact complex (Perham *et al.*, 1981; Radford *et al.*, 1986, 1987; Texter *et al.*, 1988), indicating that they are highly mobile; they probably facilitate the movement of the lipoyl domains between the active sites of E1p, E2p and E3 as part of the catalytic mechanism of the complex. In the similar E2o

chain (44 kDa) of the 2-oxoglutarate dehydrogenase complex, however, there is only a single lipoyl domain (White *et al.*, 1980; Perham & Roberts, 1981; Spencer *et al.*, 1984) and the flexible link between this and the E3-binding domain lacks the high proportion of alanine and proline residues, being richer in lysine, glutamine and glutamic acid (Spencer *et al.*, 1984; Packman & Perham, 1986).

Limited-proteolysis experiments on the pyruvate dehydrogenase complex from *Bacillus stearothermophilus*, which is assembled round a 60-subunit (icosahedral) E2 core, have shown that its E2p chain bears an *N*-terminal lipoyl domain that can also be isolated as a folded functional unit, and the ^1H -n.m.r. spectrum of the complex again suggests the existence of mobile regions of polypeptide chain linking the lipoyl and inner-core domains (Duckworth *et al.*, 1982; Packman *et al.*, 1984b). In the absence of detailed amino acid sequence information for this complex, however, interpretation of the ^1H -n.m.r. data is necessarily limited. We have therefore determined the amino acid sequence of the *N*-terminal half of the *B. stearothermophilus* E2p molecule and, in addition, some of the primary cleavage sites with trypsin, chymotrypsin and *S. aureus* V8 proteinase. The sequence determined covers the lipoyl domain, a domain that binds E1p and E3, and extends into the inner-core or E2-binding domain. There turn out to be some striking differences as well as similarities in the structural organization of this enzyme compared with that of the E2p and E2o components of the octahedral complexes from *E. coli*. Knowledge of this amino acid sequence now opens the way for a more detailed study of the *B. stearothermophilus* complex by means of ^1H -n.m.r. spectroscopy and protein engineering, complementary to that being undertaken on the pyruvate dehydrogenase complex of *E. coli* (Guest *et al.*, 1985; Graham *et al.*, 1986; Miles *et al.*, 1987; Radford *et al.*, 1987; Texter *et al.*, 1988).

MATERIALS AND METHODS

Enzymes and reagents

The pyruvate dehydrogenase complex was purified from *B. stearothermophilus* by the method of Henderson & Perham (1980). Chymotrypsin, pepsin and *N*-tosyl-L-phenylalanylchloromethane-treated trypsin were from Worthington Biochemical Corp., Freehold, NJ, U.S.A. *S. aureus* V8 proteinase was from Miles Scientific, now ICN Biochemicals, High Wycombe, Bucks., U.K. Reagents used in amino acid sequence analysis were from Applied Biosystems, Warrington, Lancs., U.K. BNPS-skatole was from Pierce and Warriner, Chester, Cheshire, U.K. Aquacide II was from Calbiochem, La Jolla, CA, U.S.A. All other reagents used were of analytical grade.

Limited proteolysis of pyruvate dehydrogenase complex

Digestion of native pyruvate dehydrogenase complex (10.7 mg, 14 mg/ml in 20 mM-sodium phosphate buffer, pH 7.0, containing 2.7 mM-EDTA and 0.02% NaN_3) with chymotrypsin was performed at 0 °C with 23 μg (0.2%, w/w) of the proteinase (Duckworth *et al.*, 1982). A second addition of the proteinase (10 μg) was made at 65 min, and at 80 min the incubation temperature was raised to 23 °C to accelerate the final stages of digestion. When the catalytic activity of the complex had fallen to

below 5% of its starting value, phenylmethanesulphonyl fluoride was added to a final concentration of 1 mM. The sample was centrifuged (11 500 *g* for 5 min) to remove particulate matter and applied to a Superose 12 column (1 cm $\times$ 30 cm) equilibrated and developed with 50 mM-sodium phosphate buffer, pH 7.0, containing 2.7 mM-EDTA and 0.02% NaN_3 at 20 °C. A residual complex, assembled round an E2p core comprising truncated E2 chains (apparent molecular mass 45 kDa), emerged at the void volume, V_0 , of the column; later peaks comprised a small quantity of unbound E1p α , E1p β and E3 subunits (which arise from a small amount of degraded E2p in the original preparation of the complex) and the excised lipoyl domains.

The lipoyl domain was purified by ion-exchange chromatography on a Mono Q resin (Pharmacia) column (0.5 cm $\times$ 5 cm) with a linear gradient of 10–500 mM- NH_4HCO_3 . Where necessary, any remaining contaminants were removed by reverse-phase chromatography on a Brownlee Aquapore (C_8) RP300 column (2 mm $\times$ 3 cm) equilibrated with 0.3% (v/v) trifluoroacetic acid, with acetonitrile (containing 0.3% trifluoroacetic acid) as organic modifier.

The residual chymotrypsin-treated complex was concentrated to 6.4 mg/ml by immersing the sample in a dialysis bag in Aquacide II, and a portion (0.55 ml, 3.5 mg) was subjected to further proteolysis with 11 μg (0.3%, w/w) of trypsin at 0 °C in 50 mM-sodium phosphate buffer, pH 7.0, for 1 h. Digestion was arrested by adding 32 μg of soya-bean trypsin inhibitor. The components of the now disassembled complex were separated by gel filtration at 20 °C on a Superose 12 column (1 cm $\times$ 30 cm) equilibrated and developed with 1% (w/v) NH_4HCO_3 .

Cleavage of peptides, peptide separation and purification

Some products of limited proteolysis and all the peptides generated by further chemical or enzymic cleavages were purified by reverse-phase chromatography as described above but with a Spherisorb 5 μm -particle-size (C_{18}) ODS2 column (3 mm $\times$ 10 cm) as support.

Chemical cleavage of the lipoyl domain (5 nmol) with BNPS-skatole (Fontana *et al.*, 1980), recrystallized from acetone, was carried out in 70 μl of 61% (v/v) acetic acid containing 30 μg of L-tyrosine for 24 h at 23 °C. The bulk of the reagent was precipitated by dilution with 230 μl of 0.3% (v/v) trifluoroacetic acid, and the reaction products in the supernatant were separated by reverse-phase chromatography. CNBr digestion was attempted with a 30-fold (w/w) excess of reagent in 70% (v/v) formic acid for 18 h at 23 °C in the dark, under argon.

The lipoyl domain had to be denatured before carrying out enzymic digestion with trypsin or *S. aureus* V8 proteinase, to overcome the proteinase-resistance of the folded protein substrate at pH 7–8 (Packman *et al.*, 1984b). This was achieved by incubating the salt-free domain in 70% (v/v) acetonitrile containing 1% (v/v) trifluoroacetic acid for 5 min at room temperature. The protein was recovered under vacuum, resuspended in 0.5% NH_4HCO_3 , and immediately treated with proteinase.

Total enzymic digestion of lipoyl domain and other peptides was carried out for 18 h with 1% (w/w) trypsin or 1–2% (w/w) *S. aureus* V8 proteinase or for 4 h with 1% (w/w) chymotrypsin, each at 37 °C in 0.5% NH_4HCO_3 . Digestion with 1% (w/w) pepsin was at

37 °C for 1 h in 5 % (v/v) formic acid. The peptides from the peptic digest of the lipoyl domain were injected directly on to the C_{18} reverse-phase column, equilibrated with 0.3 % (v/v) trifluoroacetic acid, but the peptides from the other digests were first acidified with an equal volume of 1 % (v/v) trifluoroacetic acid.

Amino acid analysis and sequence determination

Peptides were hydrolysed under vacuum or in an atmosphere of argon by the vapour from 6 M-HCl containing 1 % (v/v) 2-mercaptoacetic acid, for 18–24 h at 110 °C. The amino acid composition was determined with an LKB 4400 amino acid analyser. Amino acid sequence analysis was performed on an Applied Biosystems 470A gas-phase sequencer coupled to a 120A amino acid phenylthiohydantoin analyser. All samples were applied to a glass-fibre disc coated with 3 mg of Biobrene Plus and precycled before use in accordance with the manufacturer's instructions. All sequence runs were performed with the use of the 03RPTH programme, modified to contain 03CPRO cycles as necessary.

Other techniques

SDS/polyacrylamide-gel electrophoresis (Tris/glycine buffer) and enzyme assays were all performed as previously described (Packman & Perham, 1987).

Accuracy of the determined amino acid sequence

Several regions of sequence have been represented by a single peptide only (Fig. 1). However, the sequence determined appeared unambiguous in these areas and correlated well with the amino acid composition of the relevant peptide (see Supplementary Publication SUP 50142). There is therefore no reason to suspect the presence of wrong assignments. Where ambiguous or unexpected low-yield residues occurred in a sequence determination, a separate peptide covering the doubtful region was isolated from a different enzymic digest and sequenced. All ambiguities were resolved with the second peptide.

RESULTS

Sequence analysis of the lipoyl domain

Treatment of the native pyruvate dehydrogenase complex from *B. stearothermophilus* with chymotrypsin releases the lipoyl domain from the E2p chain (Duckworth *et al.*, 1982; Packman *et al.*, 1984b). The lipoyl domain, purified as described above, was submitted to *N*-terminal sequence analysis, but only 25 residues could be identified unambiguously owing to a low repetitive yield (87 %) caused by sample wash-out. Sequence analysis of the first six residues of a sample of intact E2p (Henderson & Perham, 1980) confirmed the lipoyl domain to be at the *N*-terminus of the molecule. Cleavage of the lipoyl domain (5 nmol) by BNPS-skatole at tryptophan-21 produced a large *C*-terminal fragment (BNPS-2), which was purified by reverse-phase h.p.l.c. A sample of this fragment (1 nmol) was subjected to *N*-terminal sequence analysis, but sample wash-out was again significant. However, this enabled the sequence to be extended up to position 50, although no clear residue was recorded at positions 36 and 42.

The remainder of the sequence of the lipoyl domain was established by sequencing selected peptides from total digests of the whole domain. A peptic digest of the

domain (5 nmol) yielded a number of peptides, which were separated by means of h.p.l.c. and subjected to amino acid analysis. Any peptide recognized as not being from the *N*-terminal half of the molecule was sequenced and its likely position in the primary structure assessed by comparison with the known sequences of the lipoyl domains from the E2p and E2o chains of *E. coli*. This information influenced the choice of proteinase used in further enzymic digests of the lipoyl domain, which were designed to generate the peptides required to overlap the existing sequences and to confirm some more tentative assignments towards the end of the sequence of the large BNPS-skatole-cleavage fragment.

Two further digests of the lipoyl domain (4–5 nmol per digest) were prepared. Peptides from a tryptic digest were separated by means of h.p.l.c. and the desired ones were selected for sequence analysis on the basis of their amino acid compositions.

For an *S. aureus* V8 proteinase digest, pyridethylated domain was first prepared. Lipoyl domain (10 nmol) was reduced for 30 min with 2 mM-dithiothreitol in 150 mM-Tris/HCl buffer, pH 8.5, containing 6 M-guanidinium chloride, and then incubated with 11 mM-vinylpyridine for 30 min, all under argon at room temperature. The modified domain was purified by h.p.l.c. and subjected to amino acid analysis, which revealed the presence of a single pyridethyl-cysteine residue per molecule. The position of this residue was located by sequence analysis of a peptide isolated from the *S. aureus* V8 proteinase digest of the modified domain (peptide V6; Fig. 1).

The same digest produced another peptide (V10; Fig. 1), which was found to contain a lysine residue by amino acid analysis. This residue failed to appear in the sequence analysis, but a blank cycle was observed at position 5. Examination of the peptide by fast-atom-bombardment m.s. revealed its mass to be 1403 Da, consistent with the missing residue being S^6S^8 -(bispyridethyl)-lipoyl-lysine. Lysine-42 is therefore taken to be the site of lipoylation in this protein.

The *C*-terminal peptide of the lipoyl domain, Asn-Met-Thr-Phe, was identified from the *S. aureus* V8 proteinase digest as lacking a *C*-terminal glutamic acid residue but having a chymotrypsin-sensitive residue, phenylalanine, at its *C*-terminus.

Treatment of the lipoyl domain with CNBr repeatedly failed to cleave the chain significantly at methionine-83, although conversion into homoserine was complete. This was established by amino acid and sequence analysis of the *C*-terminal peptide (P11; Fig. 1) of the CNBr-treated domain, isolated from a peptic digest.

The sequences of the peptides used to establish the primary structure of the lipoyl domain are summarized in Fig. 1.

Sequence analysis of the E1p/E3 binding domain

When the chymotrypsin-treated pyruvate dehydrogenase complex, which remained assembled and lacked only the lipoyl domains, was treated with trypsin under mild conditions, complete disassembly of the complex ensued (Duckworth *et al.*, 1982). Gel filtration of the digest on Superose 12 gave a number of peaks (Fig. 2), which were submitted to SDS/polyacrylamide-gel electrophoresis. The material emerging at the V_0 of the column (peak I) consisted of the E2p inner core, composed of 28 kDa fragments of the E2p chains. Peak II, containing the E1p and E3 subunits, contained in

Fig. 1. Amino acid sequence of the first 211 residues of E2p

The strategy used for isolating the peptides is described in the text. Peptides derived from limited proteinase digestion with trypsin, chymotrypsin and *S. aureus* V8 proteinase are prefixed with LT, LC and LV respectively. Peptides from total enzymic digests with these proteinases and pepsin are prefixed with T, C, V and P respectively. V-LV denotes a peptide from a total enzymic digest with *S. aureus* V8 proteinase of a fragment first isolated by limited digestion with the same proteinase. ←, Complete peptide sequence determined; —, partial sequence of a larger peptide; ●, position of lipoyl group. The cleavage sites for limited digestion with trypsin (T), chymotrypsin (C) and *S. aureus* V8 proteinase (V8) are shown by vertical arrows; other cleavage sites are likely between residues 85 and 127 but have not been characterized. Peptide V9a co-purifies in part with peptide V9.

addition some material of lower molecular mass, which arises in part from degradation of the E1p subunits, principally E1p α . However, this peak also contained a fragment of apparent mass 6 kDa, identical in size with a fragment that emerged later (peak III) in the effluent from the column. The fractions containing this latter fragment were submitted to reverse-phase h.p.l.c., a process that revealed the additional presence of a number of smaller-sized fragments (see the next section). Each fragment present was subjected to *N*-terminal sequence analysis. The largest fragment (LT21.15) was 53 residues long, of which only the last five residues were not clearly identified in the sequence analysis. These were determined from the sequence of a peptide (C2; Fig. 1) isolated from a complete digest of fragment LT21.15 (2 nmol) with chymotrypsin. The same digest curiously produced a peptide, Val-Arg-Lys-Tyr, which arose from chain cleavage by chymotrypsin after serine-133. The sequence of fragment LT21.15 (Fig. 1) was recognized as being highly homologous with the sequence of the E3-binding domain from the E2p and E2o components of the 2-oxo acid dehydrogenase complexes from *E. coli* (Packman & Perham, 1986, 1987) (see below).

Acidification of the E1p- and E3-containing fractions (peak II) from the Superose 12 column with 1% (v/v) trifluoroacetic acid preferentially precipitated the larger proteins. The smaller components remained soluble and were again purified by h.p.l.c. The main component, apparent mass 6 kDa, proved to have the same

N-terminal sequence as fragment LT21.15 described above. The common occurrence of this fragment of E2p in peaks II and III from the Superose 12 column strongly suggests that it forms a domain involved in the binding of E1p and E3 in the *B. stearothermophilus* pyruvate dehydrogenase complex.

Treatment of the *B. stearothermophilus* complex with *S. aureus* V8 proteinase causes complete disassembly; from such a digest, the putative E1p/E3-binding domain was isolated by using the same protocol as that just described for fragment LT21.15. Direct sequence analysis of this domain (LV22; Fig. 1) showed it to have a six-to-eight-residue *N*-terminal extension compared with the domain obtained by successive chymotryptic-tryptic digestion of E2p, as described above. The *N*-terminus was ragged, arising from cleavage at glutamic acid residues (positions 119 and 121) that were two residues apart. The *C*-terminus was not identified directly, but is likely to be one residue shorter than the corresponding domain released by trypsin.

The polypeptide chain between the lipoyl and E1p/E3-binding domains

Sequence analysis of the two smaller fragments (peptides LT21.8 and LT21.9) purified by h.p.l.c., which were co-eluted with the E1p/E3-binding domain in peak III on the Superose column, showed them to have arisen from the region of polypeptide chain that links the lipoyl domain to the E1p/E3-binding domain. The shorter of

Fig. 2. Gel filtration of the limited tryptic digest of chymotrypsin-treated pyruvate dehydrogenase complex

Chymotrypsin-treated pyruvate dehydrogenase complex (3.5 mg) was treated with trypsin under mild conditions as described in the Materials and methods section. The cleavage products were separated by gel filtration on Superose 12 equilibrated with 1% NH_4HCO_3 . The compositions of peaks I, II and III are discussed in the text. The remaining peaks probably comprise degradation products of the E1p component. The large peak at fraction 25 comprises buffer salts and EDTA.

the two peptides was derived from the larger by cleavage with trypsin (Fig. 1). The complete amino acid sequence of this region was established directly, as follows.

E2p was purified from the pyruvate dehydrogenase complex by gel filtration in 2 M-KI (Henderson & Perham, 1980). Limited proteolysis of the E2p core (500 μg) with 1% (w/w) chymotrypsin at 0 °C removed the lipoyl domain to yield a smaller core composed of truncated chains of apparent molecular mass 45 kDa. This was purified by gel filtration on Superose 12 in 50 mM- NH_4HCO_3 and subjected to *N*-terminal sequence analysis. After 13 residues, sequence already obtained from peptide LT21.9 described above was reached (Fig. 1).

The overlap between the lipoyl domain and the 45 kDa fragment of E2p generated by chymotrypsin was established as follows. A sample of the pyruvate dehydrogenase complex was treated with *S. aureus* V8

proteinase to release the lipoyl domain. This was purified by gel filtration and h.p.l.c. as described above. The domain was denatured and subjected to complete digestion by the same proteinase. The peptides were separated by means of h.p.l.c. and submitted to amino acid analysis. A likely *C*-terminal peptide (V-LV4) was identified in this way and sequenced to confirm that the *C*-terminus of the lipoyl domain generated by chymotryptic cleavage at phenylalanine-85 and the *N*-terminus of the 45 kDa fragment, generated by the same means, are contiguous in the complete sequence (Fig. 1).

The inner core of E2p

Treatment of the complex with trypsin or *S. aureus* V8 proteinase produced an inner E2p core composed of fragments of the E2p chains with apparent molecular mass 28 kDa. *N*-Terminal sequence analysis of each of these fragments (LT28 and LV28 respectively; Fig. 1) showed them to differ by a single additional lysine residue at the *N*-terminus of the *S. aureus* V8 fragment. The *C*-terminal sequence of the E1p/E3-binding domain was overlapped with the inner core as follows. The isolated E2p core (500 μg) was treated with 0.5% (w/w) chymotrypsin as described above but the incubation temperature was raised to 37 °C. This induced a cleavage at phenylalanine-165 to generate a larger version (29 kDa, fragment LC29) of the E2p inner core. *N*-Terminal sequence analysis of this fragment was carried out until the amino acid sequence of the 28 kDa fragment was recognized (Fig. 1).

DISCUSSION

The first 211 residues of the amino acid sequence of the E2p component of the pyruvate dehydrogenase complex from *B. stearothermophilus* are shown in Fig. 1. The limits of the domains are necessarily approximate; they are based upon the accessibility of the structure to proteinases rather than any precise physical measurement. The *N*-terminal lipoyl domain bears a lipoyl group bound to lysine-42, and, unlike any lipoyl domain from other 2-oxo acid dehydrogenase complexes studied so far, it also contains a cysteine residue, located at position 36. Whether this residue is capable of partaking in any of the reactions of the domain is not yet known.

Limited proteolysis of the *B. stearothermophilus* pyruvate dehydrogenase complex with chymotrypsin releases the lipoyl domain from E2p (cleavage site now identified as phenylalanine-85; Fig. 1) without causing disassembly of the complex (Duckworth *et al.*, 1982). Limited digestion of this residual complex with trypsin (Fig. 2) further truncates the E2p chains and releases the E1p and E3 components, exposing an inner E2p core composed of polypeptide fragments of apparent molecular mass 28 kDa (Duckworth *et al.*, 1982; Packman *et al.*, 1984b). The tryptic cleavage site that generates the 28 kDa fragment of E2p has now been identified as lysine-180 (Fig. 1). Similarly, limited proteolysis of the complex with *S. aureus* V8 proteinase causes release of the E1p and E3 components and generates an inner E2p core composed of polypeptide fragments of apparent molecular mass 28 kDa (Packman *et al.*, 1984b). The cleavage site in E2p responsible for this effect has now been identified as glutamic acid-179 (Fig. 1). This strongly suggests that the region of E2p polypeptide chain between residues 85 and 179 is involved

in binding the E1p and E3 components to the E2p core.

In support of this idea a fragment (LT21.15; apparent molecular mass 6 kDa) of the E2p chain was found free and also associated with the peak consisting of the E1p and E3 subunits during gel filtration of the tryptic digest of the chymotryptic-digest core complex on Superose 12 (Figs. 1 and 2). The amino acid sequence of this fragment was determined (Fig. 1) and found to represent residues 128 (cleavage at arginine-127) to 180 (cleavage at lysine-180). There are four arginine residues and five lysine residues (one of which is followed by a proline residue) between arginine-127 and lysine-180, making eight potential cleavage sites for trypsin in this stretch of the E2p chain. Thus the failure of trypsin to digest this fragment under the mild conditions used for the proteolysis suggests that it forms a folded protein domain. Further support comes from the results of limited proteolysis of the native complex with *S. aureus* V8 proteinase. A similar fragment of the E2p chain was isolated from this digest by the same means and found to extend from residues 120 or 122 (the *N*-terminus is ragged owing to cleavage of E2p at glutamic acid-119 or glutamic acid-121) to probably residue 179 (cleavage at glutamic acid-179), as shown in Fig. 1. There are two more glutamic acid residues in this stretch of the E2p chain, but, again, cleavage with the proteinase was not observed.

A comparable domain has already been described for the E2p (Packman & Perham, 1987; Radford *et al.*, 1987) and E2o (Packman & Perham, 1986) chains of the 2-oxo acid dehydrogenase complexes from *E. coli*. It too is some 50–60 amino acid residues long and is located between the lipoyl domain(s) and the inner-core (acyl transfer) domain. However, there is a striking difference in that this domain in the octahedral complexes from *E. coli* appears to be responsible only for binding E3. The reason for this difference is not known, but it appears to be a general distinction between the octahedral and icosahedral complexes, for the mammalian pyruvate dehydrogenase complexes (icosahedral, like the *B. stearothermophilus* enzyme) are likewise disassembled, losing their E1p and E3 components, when the E2p chains are similarly treated with various proteinases (Kresze *et al.*, 1980; Bleile *et al.*, 1981). Given that these various peripheral subunit-binding domains can now be purified, it should be possible to study their interactions with their respective E1p and E3 components and so to learn more about the part they play in the assembly, and perhaps also the catalytic mechanism, of the complexes.

The similarities of the E2p chain of the *B. stearothermophilus* pyruvate dehydrogenase complex and the E2p and E2o chains of the *E. coli* pyruvate and 2-oxoglutarate dehydrogenase complexes are very evident when the primary structures of their *N*-terminal regions are compared (Fig. 3). There is extensive homology between the five lipoyl domains (three from the E2p chain of *E. coli*) and also between the three E3- or E1p/E3-binding domains, making it certain that the related domains of the three complexes are folded similarly. However, the homology is less striking for the start of the inner-core (acyltransferase) domains, but the regions compared are still somewhat short and more sequence information will be needed here.

Of particular interest perhaps are the inter-domain regions of the E2p and E2o polypeptide chains. In the

Fig. 3. Sequence comparison between the *N*-terminal half of E2p from *B. stearothermophilus* and the corresponding regions of E2p and E2o from *E. coli*

★, Identity between E2p from *B. stearothermophilus* and E2p or E2o from *E. coli*; ●, position of lipoyl group.

B. stearothermophilus E2p sequence, the region of polypeptide chain linking the lipoyl domain with the E1p/E3-binding domain (Figs. 1 and 3) can be divided into two distinct sections, namely residues 86–109, in which charged side chains are abundant, and residues 110–128, which are relatively rich in alanine and proline with some charged residues present. The first section is very similar to the corresponding sequence in the E2o chain of the *E. coli* complex, which is rich in lysine, glutamate and glutamine residues but which lacks the preponderance of alanine and proline found in the same region of the E2p chain of *E. coli* (Figs. 1 and 3). The second section of this region of the *B. stearothermophilus* sequence is richer in alanine and proline residues and therefore more closely resembles E2p. In contrast, the region of polypeptide chain (approx. residues 165–190) linking the E1p/E3-binding domain to the inner-core (acyltransferase) domain in the *B. stearothermophilus* E2p chain is very rich in alanine and proline (residues 167–189), and this is followed by a short section at the start of the inner-core domain (residues 190–201) that is dominated by hydrophilic and charged amino acids, notably glutamic acid residues (Fig. 1). In this region, therefore, there is more obvious similarity with the E2o than with the E2p chain of the *E. coli* enzymes (Fig. 3).

The inter-domain regions of the E2p and E2o chains of the *E. coli* enzymes harbour the principal cleavage sites for the limited proteolysis of the native enzyme complexes (Packman *et al.*, 1984a; Packman & Perham, 1986, 1987). Several such sites have now been identified in the E2p chain from *B. stearothermophilus*, although it is likely that others exist, particularly in the lysine- and glutamic acid-rich segment between residues 85 and 119. In *E. coli* the inter-domain regions of the E2p chains have also been identified (Radford *et al.*, 1986, 1987; Texter *et al.*, 1988) as the major sources of the sharp resonances observed in the ^1H -n.m.r. spectra of the native pyruvate dehydrogenase complex (Perham *et al.*, 1981). Similar considerations apply to results of the limited proteolysis and ^1H -n.m.r. spectroscopy of the 2-oxoglutarate dehydrogenase complex of *E. coli* (Perham & Roberts, 1981; Spencer *et al.*, 1984; Packman & Perham, 1986). These regions are therefore thought to be conformationally flexible and to facilitate movement of the domains as part of the mechanism of substrate transfer in the complexes. The present results on the limited proteolysis of the *B. stearothermophilus* pyruvate dehydrogenase complex also conform to this pattern (Fig. 1) and permit a fuller interpretation of the data currently available from ^1H -n.m.r. spectroscopy of the complex (Duckworth *et al.*, 1982; Packman *et al.*, 1984b).

Thus treatment of the *B. stearothermophilus* complex with chymotrypsin releases a lipoyl domain whose ^1H -n.m.r. spectrum is that of a small folded protein (Packman *et al.*, 1984b), but the sharp resonances in the ^1H -n.m.r. spectrum of the complex are unaffected (Duckworth *et al.*, 1982). The cleavage site for this proteinase has now been identified as phenylalanine-85 (Fig. 1), which allows removal of the lipoyl domain but leaves the inter-domain sequences intact as part of the residual complex. On the other hand, treatment of the complex with trypsin or *S. aureus* V8 proteinase promotes disassembly of the complex by cleaving the E1p/E3-binding domain and its associated E1p and E3 subunits

from the icosahedral inner core. Such cleavage (at lysine-180 or glutamic acid-179 respectively) produces an E2p chain containing only six alanine residues in its *N*-terminal region, and it is probably these residues that give rise to the small sharp resonance observed in the ^1H -n.m.r. spectrum of the isolated inner core (Packman *et al.*, 1984b). There is therefore every reason to suppose that the inter-domain sequences in the *B. stearothermophilus* E2p chain resemble their counterparts in the E2p and E2o chains of *E. coli* in conformational flexibility and, presumably, in function.

The 2-oxo acid dehydrogenase complexes of *E. coli* and *B. stearothermophilus*, though evidently similar in basic structure and overall mechanism, are interestingly different. One notable difference is that the E2p chain of the octahedral *E. coli* pyruvate dehydrogenase complex contains three lipoyl domains in tandem array, whereas that of the octahedral 2-oxoglutarate dehydrogenase and icosahedral *B. stearothermophilus* pyruvate dehydrogenase complexes contains just one. However, protein engineering experiments on the *E. coli* E2p chain have revealed that two of the three domains can be deleted (Guest *et al.*, 1985; Graham *et al.*, 1986), and the (alanine + proline)-rich region that links this domain to the E3-binding domain can be shortened by up to one-third (Miles *et al.*, 1987), without significant effect on the enzyme. Another notable difference, established in the present work, is that the E3-binding domain in the E2p chain of the icosahedral *B. stearothermophilus* pyruvate dehydrogenase complex appears to play a major part also in binding the E1p subunits to the E2 core, unlike the corresponding domain in the E2p and E2o chains of the *E. coli* enzymes. The reason for this may lie in the symmetry differences or in the nature of the split E1p (E1p α , E1p β) chain in the icosahedral complexes. Further study of the icosahedral enzymes, especially by the methods of protein engineering, should prove rewarding.

We thank the Science and Engineering Research Council and the Wellcome Trust for financial support and Girton College, Cambridge, for the award of a Maria Luisa de Sanchez Studentship to A. B. We are grateful to Mr. J. Lester for skilled technical assistance, to Dr. P. Booth and Dr. D. H. Williams for the results of fast-atom-bombardment m.s. of a lipoylated peptide, and to Dr. J. Walker for initial help with automated sequence analysis of proteins.

REFERENCES

- Bleile, D. M., Hackert, M. L., Pettit, F. H. & Reed, L. J. (1981) *J. Biol. Chem.* **256**, 514–519
- Duckworth, H. W., Jaenicke, R., Perham, R. N., Wilkie, A. O. M., Finch, J. T. & Roberts, G. C. K. (1982) *Eur. J. Biochem.* **124**, 63–69
- Fontana, A., Savige, W. E. & Zamboni, M. (1980) in *Methods in Peptide and Protein Sequence Analysis* (Birch, C., ed.), pp. 309–322, Elsevier/North-Holland, Amsterdam
- 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 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
- Henderson, C. E. & Perham, R. N. (1980) *Biochem. J.* **189**, 161–172
- Kresze, G.-B., Ronft, H. & Dietl, B. (1980) *Eur. J. Biochem.* **105**, 371–379

- Miles, J. S., Guest, J. R., Radford, S. E. & Perham, R. N. (1987) *Biochim. Biophys. Acta* **913**, 117–121
- Packman, L. C. & Perham, R. N. (1986) *FEBS Lett.* **206**, 193–198
- Packman, L. C. & Perham, R. N. (1987) *Biochem. J.* **242**, 531–538
- Packman, L. C., Hale, G. & Perham, R. N. (1984a) *EMBO J.* **3**, 1315–1319
- Packman, L. C., Perham, R. N. & Roberts, G. C. K. (1984b) *Biochem. J.* **217**, 219–227
- 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
- Perham, R. N., Packman, L. C. & Radford, S. E. (1987) *Biochem. Soc. Symp.* **54**, 67–81
- Radford, S. E., Laue, E. D. & Perham, R. N. (1986) *Biochem. Soc. Trans.* **14**, 1231–1232
- Radford, S. E., Laue, E. D., Perham, R. N., Miles, J. S. & Guest, J. R. (1987) *Biochem. J.* **247**, 641–649
- Reed, L. J. (1974) *Acc. Chem. Res.* **7**, 40–46
- 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
- Texter, F. L., Radford, S. E., Laue, E. D., Perham, R. N., Miles, J. S. & Guest, J. R. (1988) *Biochemistry* **27**, 289–296
- White, R. H., Bleile, D. M. & Reed, L. J. (1980) *Biochem. Biophys. Res. Commun.* **94**, 78–84
- Yeaman, S. J. (1986) *Trends Biochem. Sci.* **11**, 293–296

Received 3 November 1987; accepted 14 January 1988