Segmental structure and protein domains in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*

Genetic reconstruction *in vitro* and ^1H -n.m.r. spectroscopy

Sheena E. RADFORD,* Ernest D. LAUE,* Richard N. PERHAM,*† John S. MILES† and John R. GUEST†

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

†Department of Microbiology, University of Sheffield, Sheffield S10 2TN, U.K.

A deletion *in vitro* can be made in the *aceEF-lpd* operon encoding the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*, which causes deletion of two of the three homologous lipoyl domains that comprise the *N*-terminal half of each dihydrolipoamide acetyltransferase (E2p) polypeptide chain. An active complex is still formed and ^1H -n.m.r. spectroscopy of this modified complex revealed that many of the unusually sharp resonances previously attributed to conformationally mobile segments in the wild-type E2p polypeptide chains had correspondingly disappeared. A further deletion was engineered in the long (alanine + proline)-rich segment of polypeptide chain that linked the one remaining lipoyl domain to the *C*-terminal half of the E2p chain. ^1H -n.m.r. spectroscopy of the resulting enzyme complex, which was also active, revealed a further corresponding loss in the unusually sharp resonances observed in the spectrum. These experiments strongly support the view that the sharp resonances derive, principally at least, from the three long (alanine + proline)-rich sequences which separate the three lipoyl domains and link them to the *C*-terminal half of the E2p chain. Closer examination of the 400 MHz ^1H -n.m.r. spectra of the wild-type and restructured complexes, and of the products of limited proteolysis, revealed another sharp but smaller resonance. This was tentatively attributed to another, but smaller, (alanine + proline)-rich sequence that separates the dihydrolipoamide dehydrogenase-binding domain from the inner core domain in the *C*-terminal half of the E2p chain. If this sequence is also conformationally flexible, it may explain previous fluorescence data which suggest that dihydrolipoamide dehydrogenase bound to the enzyme complex is quite mobile. The acetyltransferase active site in the E2p chain was shown to reside in the inner core domain, between residues 370 and 629.

INTRODUCTION

The pyruvate dehydrogenase (PDH) multienzyme complex of *Escherichia coli* catalyses the following overall reaction:

It contains a structural core of 24 dihydrolipoamide acetyltransferase (E2p, EC 2.3.1.12) polypeptide chains, arranged with octahedral symmetry, around which are assembled multiple copies of pyruvate decarboxylase (E1p, EC 1.2.4.1) and dihydrolipoamide dehydrogenase (E3, EC 1.8.1.4, formerly EC 1.6.4.3) subunits [for reviews, see Reed (1974) and Perham (1983)]. The primary structure of the E2p chain, inferred from the DNA sequence of the *aceF* gene (Stephens *et al.*, 1983), is notable for having, as its *N*-terminal half, three highly homologous sequences, each of about 80 amino acid residues in tandem array. These sequences are separated from each other and linked to the *C*-terminal half of the E2p chain by three long segments (20–30 residues) of polypeptide chain unusually rich in alanine, proline and charged amino acids (Fig. 1).

The substrate is transferred between the physically separate active sites in the complex by means of lipoyl-

lysine swinging arms (Reed, 1974; Ambrose & Perham, 1976; Grande *et al.*, 1976). There are three such lipoyl-lysine residues per E2p chain (Packman *et al.*, 1984a), one in each of the three repeated sequences that form the *N*-terminal half of the molecule (Fig. 1). The lipoylated regions protrude from an inner region of the E2p core between the E1p and E3 subunits and can be released from the PDH complex, without causing disassembly of the particle, by limited proteolysis with trypsin at Lys-316 (Bleile *et al.*, 1979; Hale & Perham, 1979; Packman *et al.*, 1984a). They are folded to form three distinct and functional lipoyl domains that can be isolated after limited proteolysis of the complex with *Staphylococcus aureus* V8 proteinase, which readily cleaves the E2p chain in each of the long (alanine + proline)-rich segments (Packman *et al.*, 1984a). Harsher treatment of the complex with the same proteinase brings about an additional cleavage of E2p at Glu-372 (Packman & Perham, 1987), and this is accompanied by dissociation of the E3 component but not of E1p (Fig. 1). Thus it is likely that the region of E2p between Lys-316 and Glu-372, which may represent another folded domain, is of particular importance in the binding of E3 (Packman & Perham, 1986).

On the basis of fluorescence energy transfer experiments, it has been reported that the distance between the

Abbreviations used: PDH, pyruvate dehydrogenase; TSP, 3-(trimethylsilyl)-2,2,3,3-tetradeuteriopropionic acid.

† To whom correspondence should be addressed.

Fig. 1. Primary structure of the E2p chain of the pyruvate dehydrogenase complex of *E. coli*

The covalently-bound lipoyl groups are indicated (●) and the (alanine + proline)-rich regions are underlined. The sequences of the lipoyl domains are aligned for maximum homology. Abbreviations: T, tryptic cleavage site; V, *S. aureus* V8 proteinase cleavage site.

E1p and E3 active sites is not less than 5 nm, which is more than the span (2.8 nm) of a single lipoyl-lysine swinging arm (Angelides & Hammes, 1979). The lipoyl-lysine arm must also be able to reach the acetyltransferase active site, which is known to reside in the inner core domain of the E2p polypeptide chain (Bleile *et al.*, 1979; Angier *et al.*, 1987). There are other unusual features of the enzyme mechanism. For example, acetyl groups can be transferred between lipoic acid residues of different E2p subunits in the same enzyme core (Bates *et al.*, 1977; Collins & Reed, 1977; Packman *et al.*, 1983) and a given lipoyl group can visit the active site of several different E1p subunits in the same complex (Stepp *et al.*, 1981; Berman *et al.*, 1981; Hackert *et al.*, 1983).

The explanation for these novel systems of active site coupling undoubtedly lies in the highly segmented structure of the E2p polypeptide chain. The appearance of the E2p core in the electron microscope and its susceptibility to limited proteolysis have led to suggestions that the lipoylated regions are conformationally flexible (Bleile *et al.*, 1979; Hale & Perham, 1979). Strong evidence in favour of the existence of particular regions of the E2p chain with substantial conformational mobility has come from ¹H-n.m.r. spectroscopy combined with limited proteolysis of the complex (Perham *et al.*, 1981; Roberts *et al.*, 1983). Since the n.m.r. spectra point to a high proportion of alanine or threonine residues in the flexible regions (Perham *et al.*, 1981), it has been proposed that the three long segments oddly rich in alanine residues which separate the lipoyl domains are those primarily concerned (Packman *et al.*, 1984a; Spencer *et al.*, 1984). The fact that the sharp resonances in the ¹H-n.m.r. spectrum of the intact PDH complex closely resemble the ¹H-n.m.r. spectrum of a 32-residue synthetic peptide with the same amino acid sequence as the innermost of the three long (alanine + proline)-rich segments of the E2p polypeptide chain (Fig. 1) strongly supports this view (Radford *et al.*, 1986).

Specific deletions *in vitro* in the *aceF* gene have recently allowed the creation of *E. coli* PDH complexes assembled round E2p cores comprising truncated E2p chains from which one or more of the lipoyl domains have been excised (Guest *et al.*, 1985). Lowering the number of lipoyl domains from three to two or even to one per E2p chain (pGS110-encoded complex, Fig. 2) was found to cause no detectable impairment in catalytic activity or in

the intramolecular coupling of active sites. The outer two lipoyl domains of the E2p chain can therefore be regarded as redundant in this context. Similarly, a 12-residue deletion was engineered in the (alanine + proline)-rich segment that links the lipoyl domain to the E3-binding domain in a one-lipoyl domain E2p chain (pGS156-encoded complex, Fig. 2), again without impairment in catalytic activity or intramolecular coupling of active sites in the resultant PDH complex (Graham *et al.*, 1986).

The construction of these genetically engineered complexes (Guest *et al.*, 1985; Graham *et al.*, 1986) has allowed us to renew the study of the *E. coli* pyruvate dehydrogenase multienzyme complex by means of ¹H-n.m.r. spectroscopy. In the present paper, we describe a further correlation of the sharp resonances in the 400 MHz ¹H-n.m.r. spectra of the complexes with the length of the (alanine + proline)-rich segments in the E2p chain and identify a new sharp resonance that may derive from another conformationally mobile segment in the same chain. We have also been able to define more precisely the region of the E2p chain that houses the acetyltransferase active site.

Fig. 2. Schematic representation of the structure of the wild-type, pGS110- and pGS156-encoded E2p chains

The E2p chain of the pGS110-encoded complex was formed by deleting residues 33 to 236 inclusive and that of the pGS156-encoded complex was formed by a further deletion of residues 284 to 295 inclusive. The lipoyl domains are differentially shaded; ~, sequences rich in alanine, proline and charged residues; ●, lipoyl group.

MATERIALS AND METHODS

Materials

Phosphotransacetylase was purchased from Sigma and tosylphenylalanylchloromethane ('TPCK')-treated trypsin was from Worthington. Deuterium oxide (99.8%) was bought from Norsk Hydro. Soyabean trypsin inhibitor and 3-(trimethylsilyl)-2,2,3,3-tetradeuteropropionic acid (TSP) were from BDH.

Purification of PDH complexes

The wild-type PDH complex was purified from a constitutive mutant of *E. coli* K12 as described by Danson *et al.* (1979). Complexes from mutated strains of *E. coli* were purified in the same manner, as modified by Guest *et al.* (1985).

Enzyme assays

The PDH complex activity was measured in the direction of NAD⁺-reduction (Danson & Perham, 1976). The acetyltransferase activity of the E2p component was assayed by the method of Reed & Willms (1966), except that the formation of acetyldihydrolipoamide was monitored spectrophotometrically at 240 nm instead of colorimetrically after reaction with hydroxylamine (Packman *et al.*, 1984b).

Resolution of the PDH complex

The E1pE2p subcomplex was prepared by removal of the E3 component from the PDH complex on hydroxyapatite in 6 M-urea (Perham *et al.*, 1981). The E2pE3 subcomplex was prepared by treating the intact complex at pH 10.0 for 3 h, followed by gel filtration at pH 10.0 on Sepharose CL-6B (Coggins *et al.*, 1976). Both subcomplexes were concentrated by ultracentrifugation (122000 g for 5 h), dissolved and dialysed exhaustively against ²H₂O buffer containing 20 mM-potassium phosphate, 2.7 mM-EDTA and 0.02 % NaN₃, p²H 7.0. The E1p and E3 components removed from the complex by these procedures were concentrated by means of ultrafiltration and dialysed against the same ²H₂O buffer.

Preparation of the tryptic-core complexes for ¹H-n.m.r. spectroscopy

The wild-type and pGS156-encoded complexes (25 mg, 5 mg/ml in 20 mM-sodium phosphate buffer, pH 7.0, containing 2.7 mM-EDTA and 0.02 % NaN₃) were treated with 0.3 % (w/w) trypsin at 0 °C (Perham *et al.*, 1981). When complex activity had fallen to less than 5 % of its original value (approx. 80 min), proteolysis was arrested by the addition of a 10-fold excess (w/w over trypsin) of soyabean trypsin inhibitor. The tryptic cores of the complexes were then purified by gel filtration either on Sepharose 6B at 4 °C, or on a TSK 3000SW column (7.5 mm × 60 cm) at room temperature, in 50 mM-sodium phosphate buffer, pH 7.0, containing 2.7 mM-EDTA and 0.02 % NaN₃. The protein that emerged at the void volume was collected, diluted 2-fold with 2.7 mM-EDTA and concentrated by ultracentrifugation (135000 g for 3 h). The resulting pellets were resuspended in the 20 mM-potassium phosphate/²H₂O buffer described above, and dialysed exhaustively against the same buffer.

Generation of the 29 kDa fragment of E2p

The pGS156-encoded complex (25 mg) was resolved into E1p and the E2pE3 subcomplex, as described above. The subcomplex (0.8 mg/ml in 20 mM-sodium phosphate buffer, pH 7.0, containing 2.7 mM-EDTA and 0.02 % NaN₃) was treated with trypsin (1.5 % w/w) at 37 °C for 2 h. Proteolysis was then arrested by the addition of a 5-fold excess (w/w over trypsin) of soyabean trypsin inhibitor. The residual E2p core, composed of 29 kDa fragments of E2p, was only partly soluble in this low-ionic-strength buffer. The insoluble fraction was collected by centrifugation at 120000 g for 15 min and the soluble fraction collected by centrifugation at 135000 g for 6 h. The resulting pellets were dissolved in ²H₂O buffer containing 200 mM-potassium phosphate, 2.7 mM-EDTA and 0.02 % NaN₃, p²H 7.0. The solutions were then combined and dialysed against the same buffer.

¹H-n.m.r. spectroscopy

¹H-n.m.r. spectra were obtained at 400 MHz with a Bruker AM400 spectrometer, using a 10 kHz spectral width and a 90° pulse. The free-induction decay was recorded in 32 K data points, and 500–1000 measurements were taken and averaged. The sample temperature was 25 (± 1) °C. Before Fourier transformation, the free-induction decay was multiplied by a negative exponential corresponding to a line broadening of 2 Hz. Chemical shifts are expressed relative to TSP. Spin-echo spectra were obtained using the pulse sequence (90°-*t*-180°-*t*-acquire-*T*)*n* where the variable delay (*t*) was 0.5 ms and *T* is a fixed delay of 10 s (equivalent to 5*T*₁ of the methyl protons of TSP). The intensity of the peaks was calculated by cutting them out from chart paper and weighing them. Comparison was made with the weight of the TSP peak, for which the concentration was known.

Other methods

Proteins were analysed by means of SDS/polyacrylamide gel electrophoresis, as described by Packman *et al.* (1984a). Gels were stained with Coomassie Brilliant Blue. The polypeptide chain ratios of the PDH complexes and of their subcomplexes were determined by using the radioamidination method at pH 10 (Hale *et al.*, 1979). Protein concentrations were estimated spectrophotometrically.

RESULTS

¹H-n.m.r. spectroscopy of the wild-type and mutant PDH complexes

The 400 MHz ¹H-n.m.r. spectra of the wild-type, pGS110- and pGS156-encoded PDH complexes are shown in Fig. 3. The three spectra are extremely similar, consisting of a broad envelope of overlapping resonances upon which are superimposed many sharp peaks, characteristic of a section of polypeptide with substantial conformational flexibility (Perham *et al.*, 1981). A close inspection of the spectra, however, reveals that the intensity of the sharp signals differs between the three complexes. The progressive decrease in intensity of the sharp signals between the wild-type, pGS110- and pGS156-encoded complexes (Fig. 3), as exemplified by the resonance at 1.39 p.p.m. that may arise from the methyl protons of alanine or threonine side chains (Perham *et al.*, 1981), matches the progressive loss of

Fig. 3. 400 MHz ^1H -n.m.r. spectra of wild-type and mutant pyruvate dehydrogenase complexes

(a) Wild-type complex, 24 mg/ml; (b) pGS110-encoded complex, 23 mg/ml; (c) pGS156-encoded complex, 26 mg/ml. Each complex was dialysed exhaustively against $^2\text{H}_2\text{O}$ buffer containing 20 mM-potassium phosphate, 2.7 mM-EDTA and 0.02% NaN_3 , p^2H 7.0. The very sharp resonances marked E and $\times$ arise respectively from EDTA and non-protein contaminants (deriving from the dialysis tubing). The source of the resonance at 1.52 p.p.m. labelled * is discussed in the text.

(alanine + proline)-rich regions in the corresponding E2p chains (Fig. 2). This suggests that the flexible segments of polypeptide chain lie principally in the long (alanine + proline)-rich regions of E2p (the origin of the peaks marked with an asterisk is discussed below) and that the resonance at 1.39 p.p.m. arises principally from alanine residues.

It is impossible to measure accurately the intensities of the sharp peaks in these spectra owing to the difficulties in drawing an appropriate baseline. However, by the use of spin-echo spectra (which, with a total delay time, t , of 1 ms, greatly reduces the intensity of the broad envelope of signals arising from the bulk of the protein but leaves the intensity of the sharp signals virtually undiminished), an attempt was made to compare the intensity of the alanine $\beta\text{-CH}_3$ resonance (at 1.39 p.p.m.) in the three complexes (Table 1). For simplicity, the lengths of the (alanine + proline)-rich sequences in the E2p chains were assumed to be those illustrated in Fig. 1, but it must be emphasized that their precise beginnings and ends are still uncertain. Given the unavoidable errors in these calculations, the correlation between the observed and predicted numbers of mobile alanine residues per E2p chain is remarkably good. Furthermore, if each amino acid residue in the mobile segments is not equally mobile, with those residues situated closer to the lipoyl and E3-binding domains being relatively less flexible, the expected number of freely mobile alanine residues per

Table 1. Comparison of the number of mobile alanine residues, resonating at 1.39 p.p.m., in the flexible regions of the wild-type, pGS110- and pGS156-encoded complexes

The number of mobile alanine residues in the E2p chains of each complex was assessed by cutting and weighing the peak at 1.39 p.p.m. (arising from $\beta\text{-CH}_3$ protons of alanine residues) and comparing this with the weight of the peak arising from a known concentration of TSP. The concentration of E2p in each complex was calculated from the protein concentration and polypeptide chain ratios (Guest *et al.*, 1985; Graham *et al.*, 1986). The lengths of the three (alanine + proline)-rich regions in the *N*-terminal half of E2p were assumed to be those illustrated in Fig. 1, from which the number of potentially mobile alanine residues was calculated.

Complex	Number of mobile alanine residues/E2p chain	
	Observed	Potential
Wild-type	36	50
pGS110-encoded	16	20
pGS156-encoded	7	13

E2p chain will be an overestimate. This might explain the poorer correlation for the pGS156-encoded complex, since the error in estimating the number of potentially mobile residues would be correspondingly larger in the shorter length of polypeptide chain.

Characterization of the pGS156-encoded complex

The ^1H -n.m.r. spectrum of the pGS156-encoded complex differed from the spectra of the wild-type and pGS110-encoded complexes in that it clearly displayed a sharp resonance at 1.52 p.p.m., which was not initially apparent in the other two spectra (Fig. 3). The source of this novel resonance, presumably another flexible region of polypeptide chain, was therefore sought. From the chemical shift, it was again thought likely that $\beta\text{-CH}_3$ protons of alanine residues could be responsible. The E1pE2p and E2pE3 subcomplexes were prepared and their ^1H -n.m.r. spectra were examined (Fig. 4). Since E2p is the only polypeptide chain common to both subcomplexes and the sharp peak was retained at least in part in the spectra of both structures, the extra mobile piece(s) of polypeptide chain must be located within E2p. This conclusion was supported by the fact that neither the E1p nor the E3 subunits, released during the preparation of the two subcomplexes, displayed in their ^1H -n.m.r. spectra any suitably sharp resonances at the appropriate chemical shift (results not shown). These spectra of E1p and E3 closely resembled the 270 MHz ^1H -n.m.r. spectra of the same components from the wild-type complex described previously (Perham *et al.*, 1981).

The best estimates that can be made are that only about seven of the 13 potentially mobile alanine residues in the long (alanine + proline)-rich segment of the pGS156-encoded E2p chain are contributing to the sharp resonance at 1.39 p.p.m. (Table 1), and that approx. five alanine residues per E2p chain account for the adjacent peak (at 1.52 p.p.m.). It was conceivable, therefore, that the alanine residues putatively acting as the source of the

Fig. 4. 400 MHz ¹H-n.m.r. spectra of the pGS156-encoded complex and its subcomplexes

Panel I. 400 MHz ¹H-n.m.r. spectra of (a) intact complex, 26 mg/ml; (b) E2pE3 subcomplex, 9 mg/ml; (c) E1pE2p subcomplex, 7 mg/ml. The very sharp resonances labelled E and x arise from EDTA and non-protein contaminants, respectively. The peak at 1.52 p.p.m. is marked with an asterisk. The vertical gain of each spectrum has been adjusted according to the protein concentration used so that the spectra are more directly comparable. Panel II. The complexes were also examined by SDS/polyacrylamide-gel electrophoresis (20 % T, 0.3 % C). (a) Intact complex, (b) E2pE3 subcomplex, (c) E1pE2p subcomplex. The polypeptide chain ratios (E1p:E2p:E3) of the intact complex, the E2pE3 and the E1pE2p subcomplexes were respectively 0.9:1.0:1.1, 0.1:1.0:0.7 and 1.0:1.0:0.1.

resonance at 1.52 p.p.m. could reside in the same mobile segments, but that some alteration in protein structure had caused a change in their chemical shift. Alternatively, these residues could lie elsewhere in the E2p chain.

To investigate this further, the pGS156-encoded complex was submitted to limited proteolysis with trypsin at pH 7.0 and 0 °C (Perham *et al.*, 1981). This treatment releases the lipoyl domains from the E2p chains by cleavage at Lys-316 (wild-type E2p numbering); the residual complex retains its E1p and E3 chains and remains assembled around an octahedral E2p core composed of truncated E2p chains each with an M_r of approx. 36000 (Packman *et al.*, 1984a). In the 400 MHz ¹H-n.m.r. spectrum of this tryptic core complex, the large sharp resonance at 1.39 p.p.m. deriving from the long (alanine + proline)-rich segment, was no longer observed (Fig. 5), as expected from identical experiments previously carried out on the wild-type complex (Perham *et al.*, 1981). On the other hand, the smaller sharp resonance at 1.52 p.p.m. was retained and, moreover, was undiminished in its intensity (Fig. 5).

The linewidth of the peak at 1.52 p.p.m. was found to be 13 ± 5 Hz. The spin-spin relaxation time (T_2) of these protons, estimated by the spin-echo method, was 33 ms, equivalent to a theoretical linewidth of approx. 10 Hz. Thus, chemical shift heterogeneity does not contribute greatly to the width. Evidently, in the E2p chain of the pGS156-encoded complex, there must be a region of polypeptide chain, possibly including several alanine residues, that is conformationally flexible and which lies to the C-terminal side of Lys-316. This region is distinct from the long (alanine + proline)-rich segment that links the lipoyl domain to the E3-binding domain in the E2p core (Fig. 1).

The sharp resonance at 1.52 p.p.m. is a feature of all E2p chains

Given that the sharp resonance at 1.52 p.p.m. in the 400 MHz ¹H-n.m.r. spectrum of the pGS156-encoded complex derives from residues that lie to the C-terminal side of Lys-316 (see above), it should also be present in the spectra of wild-type and pGS110-encoded complexes.

Fig. 5. 400 MHz ¹H-n.m.r. spectra of the pGS156-encoded complex and its tryptic core complexes

Panel I. 400 MHz ¹H-n.m.r. spectra of (a) intact complex, 26 mg/ml; (b) tryptic core complex (36 kDa E2p), 17 mg/ml; (c) 29 kDa fragment of E2p, 2 mg/ml. The very sharp resonances labelled E and x arise from EDTA and non-protein contaminants, respectively. The peak at 1.52 p.p.m. is marked with an asterisk. The vertical gain of each spectrum has been adjusted according to the protein concentration used. Panel II. The complex and its tryptic fragments were also examined by SDS/polyacrylamide-gel electrophoresis (20%T, 0.3%C). (a) Intact complex; (b) tryptic core complex (36 kDa E2p); (c), 29 kDa fragment of E2p.

A more searching inspection of the spectra of the wild-type and the pGS110-encoded complexes (Fig. 3) revealed the presence of just such a small peak at 1.52 p.p.m. It had not been immediately obvious, being largely hidden by the larger resonance at 1.39 p.p.m. arising from the long (alanine+proline)-rich segment(s) of the E2p chain.

In the spectrum of the wild-type complex (Fig. 3), this peak is barely resolved. To confirm its presence, the 400 MHz ¹H-n.m.r. spectrum of the core complex produced by limited proteolysis of the wild-type complex with trypsin at pH 7.0 and 0 °C was examined (results not shown). The resonance could now be clearly distinguished. It formed a small distinct peak just downfield of the small residual hump (at 1.39 p.p.m.) that remains when the (alanine+proline)-rich regions which lie to the N-terminal side of Lys-316 (Packman *et al.*, 1984a) are removed.

The difficulty of fully resolving this resonance even in 400 MHz spectra of the genetically engineered complexes described above exacerbated the problem of assessing the number of amino acid residues contributing to this other conformationally mobile segment in the E2p chain. The

Table 2. Acetyltransferase activity of the wild-type, pGS156- and pGS110-encoded PDH complexes, and their tryptic fragments

The acetyltransferase activity of each complex was assayed as described in the Materials and methods section. The formation of acetyldihydroipoamide was followed at 240 nm and the rates are expressed in arbitrary units.

Enzyme	E2p activity (ΔA_{240} /min per nmol of E2p)
Wild-type complex	0.77
Tryptic core from wild-type complex (36 kDa E2p)	0.72
pGS110-encoded complex	0.64
pGS156-encoded complex	1.01
Tryptic core from pGS156-encoded complex (36 kDa E2p)	1.03
pGS156-encoded 29 kDa fragment of E2p	0.98

best estimate, from the pGS156-encoded complex is that it might contain roughly five alanine residues, together with any other amino acids involved.

¹H-n.m.r. spectroscopy of the 29 kDa fragment of the pGS156-encoded E2p chain

Under more vigorous conditions of limited proteolysis (pH 7.0, 37 °C) treatment of the PDH complex with trypsin causes cleavage of the E2p polypeptide chain at Lys-369, leaving the C-terminal half of the E2p chain as a fragment of M_r 29000 (Packman & Perham, 1987). This offered an opportunity to define more closely the origin of the sharp resonance at 1.52 p.p.m. Preliminary experiments suggested that the E2pE3 subcomplex is more susceptible to this attack by trypsin. The pGS156-encoded E2pE3 subcomplex was therefore digested with trypsin (1.5% w/w) at pH 7.0 and 37 °C for 2 h and the products were then separated by ultracentrifugation, as described in the Materials and methods section. When the pellet was analysed by means of SDS/polyacrylamide-gel electrophoresis, it was found to consist only of the expected 29 kDa fragment of the E2p chain. All the E3 chains had become dissociated from this residual E2p core (Fig. 5).

The 29 kDa fragment of E2p was shown to have remained as an aggregate, presumably as the original 24-mer, by gel filtration on a Superose 12 column in 0.5 M-sodium phosphate buffer, pH 7.0, in which it is soluble (Packman & Perham, 1987). The protein was eluted at the void volume of the column. Since cleavage of the E2p chain with trypsin at Lys-316 is not accompanied by dissociation of the E3 component (Perham *et al.*, 1981; Packman *et al.*, 1984a), it is evident that the region of E2p between residues 316 and 369 is of special importance for the binding of the E3 polypeptide chains (Packman & Perham, 1986).

When the inner core fragment of E2p, M_r 29000, in its native aggregated form (see above), was examined by means of 400 MHz ¹H-n.m.r. spectroscopy, the peak at 1.52 p.p.m. was found to be retained in the spectrum although it was reduced in intensity, corresponding to a residual two flexible residues per E2p chain (Fig. 5). Thus, at least part of the flexible piece of polypeptide chain responsible for this resonance must lie somewhere between Lys-369 and the C-terminus (Met-629) of the E2p chain. In addition, a peak at 1.39 p.p.m. reappeared in the spectrum, its intensity corresponding to about four mobile residues per E2p chain. The origin of this peak is unknown.

Location of the acetyltransferase active site

To determine the location of the acetyltransferase active site in E2p, the acetyltransferase activities in the wild-type, pGS110- and pGS156-encoded complexes were assayed and compared with the activities remaining in various tryptic fragments of E2p (Table 2). The acetyltransferase activities of the wild-type, pGS110- and pGS156-encoded complexes were similar, since the relevant active site is not part of the lipoyl domains but resides within the 36 kDa inner-core region of E2p (Bleile *et al.*, 1979; Guest *et al.*, 1985; Angier *et al.*, 1987). As expected, the full acetyltransferase activity was found in the tryptic core complex assembled around truncated E2p chains of M_r approx. 36000. However, the catalytic activity was also retained in the naked E2p inner core composed of E2p fragments of M_r 29000. The

acetyltransferase active site of E2p must therefore be housed between residues 370 and 629 of the polypeptide chain.

DISCUSSION

The unusual and surprisingly sharp resonances in the 270 MHz ¹H-n.m.r. spectra of all 2-oxo acid dehydrogenase multienzyme complexes [M_r (5–10) × 10⁶] have been attributed to the presence of conformationally flexible segments in the E2 polypeptide chains (Perham *et al.*, 1981; Roberts *et al.*, 1983; Packman *et al.*, 1984b), which might contribute to the catalytic properties of the complexes by enabling movements of the lipoyl domains. The E2p chain of the *E. coli* PDH complex has been perhaps the best studied. Since the n.m.r. spectra of this protein indicate that its flexible regions may contain about 100 amino acid residues, among which alanine or threonine figure prominently (Packman *et al.*, 1982; Roberts *et al.*, 1983), it has been thought that the three long (alanine+proline)-rich segments (20–30 residues each) which separate the lipoyl domains (Figs. 1 and 2) are most likely to be those concerned (Packman *et al.*, 1984a; Spencer *et al.*, 1984). The 400 MHz ¹H-n.m.r. spectra of the genetically reconstructed PDH complexes described above amply confirm this supposition.

There is a progressive decrease in the intensity of the sharp resonances in the spectra of the wild-type, pGS110- and pGS156-encoded complexes (Fig. 3) that accompanies the deletions in the E2p polypeptide chains (Fig. 2). These deletions all embody losses of (alanine+proline)-rich regions and, notwithstanding the unavoidable errors in estimating the intensity of the ¹H-n.m.r. signals, the correlation between the observed and potential numbers of mobile alanine residues per E2p chain (Table 1) is good. Moreover, the ¹H-n.m.r. spectra of all these complexes closely resemble the spectrum of a 32-residue synthetic peptide with the same amino acid sequence as the innermost (alanine+proline)-rich region of E2p (Radford *et al.*, 1986). Thus, there can be little doubt that the sharp resonances seen in the spectrum of the wild-type complex arise from the (alanine+proline)-rich regions of E2p, as surmised by Packman *et al.* (1984a) and Spencer *et al.* (1984), or that these regions form flexible links between the lipoyl domains.

The 400 MHz ¹H-n.m.r. spectrum of the pGS156-encoded complex was of particular interest in that it revealed the presence of an additional sharp resonance, at 1.52 p.p.m., which was not readily detected in the spectrum of the wild-type or pGS110-encoded complexes owing to the presence of the much larger signal at 1.39 p.p.m. (Fig. 3). This additional resonance clearly arose from the E2p chains (Fig. 4). Further n.m.r. spectroscopy of the complex remaining after tryptic cleavage of the E2p chains at Lys-316 (Fig. 5) and of the naked residual E2p core formed by tryptic cleavage of the E2p chains at Lys-369, demonstrated that the amino acid sequence(s) responsible must lie to the C-terminal side of Lys-369 in the E2p polypeptide chain. Our best estimate, albeit very rough, is that the equivalent of about five alanine residues per E2p chain are responsible for this peak. The same resonance at 1.52 p.p.m. could be detected in the 400 MHz ¹H-n.m.r. spectrum of the wild-type complex, proving that it must arise from a normal feature of the E2p chain. The signal had not been visible in the earlier 270 MHz ¹H-n.m.r. spectrum (Perham *et al.*, 1981),

probably because of its relatively small size and the poorer resolving power of the n.m.r. spectrometer then in use.

It has recently been shown (Packman & Perham, 1986, 1987) that the three long (alanine + proline)-rich regions of the E2p chains are not the only ones susceptible to limited proteolysis in the wild-type PDH complex, as evidenced by the slow cleavages at Lys-369 (with trypsin) and at Glu-372 (with *S. aureus* V8 proteinase). Cleavage of the E2p chains at Glu-372 (Packman & Perham, 1986) or at Lys-369 (see Fig. 5) is accompanied by dissociation of the E3 component. This strongly suggests that the region of polypeptide chain between Lys-316 and Lys-369, probably in the form of another separately folded domain (Packman & Perham, 1986), is of particular importance in binding E3 subunits. It should also be noted that full acetyltransferase activity was retained by the 29 kDa fragment of the E2p chain formed by tryptic cleavage at Lys-369 (Table 2). This active site must therefore be housed between Arg-370 and the C-terminus, Met-629.

Susceptibility to proteolysis is often an indication of exposure to solvent and conformational mobility in a polypeptide chain. Just to the C-terminal side of Lys-369 in E2p there exists a short region of polypeptide chain containing five alanine residues, one proline residue and several charged amino acids (residues 370–377, Fig. 1). The resemblance to the longer (alanine + proline)-rich regions is quite marked and, given the ^1H -n.m.r. results described above, it is tempting to suggest that this region of polypeptide chain enjoys some form of conformational flexibility and gives rise, at least in part, to the newly recognized sharp resonance at 1.52 p.p.m. (Fig. 5).

We have assigned the peak at 1.39 p.p.m. in the intact complex to alanine residues in a solvent-exposed and conformationally mobile environment. We suspect that the peak at 1.52 p.p.m. is also due to alanine residues, also conformationally mobile, but in a somewhat different environment. The chemical shift could again be consistent with the $\beta\text{-CH}_3$ protons of alanine residues (although the $\gamma\text{-CH}_3$ protons of threonine, the $\gamma\text{-CH}_2$ protons of lysine or isoleucine and the $\beta\text{-CH}_2$ and $\gamma\text{-CH}$ protons of leucine residues are also possible candidates on chemical shift grounds) (Jardetzky & Roberts, 1981). The chemical shift and the intensity of the peak, coupled with the knowledge of the proteolytic cleavage sites, point to alanine residues in the sequence 370–377 being the source of the peak at 1.52 p.p.m. The reappearance of a peak at 1.39 p.p.m. in the complex after tryptic cleavage at Lys-369 might be due to a few (approx. four) of these alanine residues entering a similar environment to those of the long (alanine + proline)-rich region, approx. two remaining in their original environment and resonating still at 1.52 p.p.m. However, it has to be emphasized that this is a speculative interpretation, one that can only be tested by further experiments.

The suggestion of conformational mobility in the region of residues 370–377 of E2p is strongly supported by comparable experiments with the 2-oxoglutarate dehydrogenase complex of *E. coli*. Each E2o chain in this complex contains only one lipoyl domain (Spencer *et al.*, 1984; Perham & Roberts, 1981) which can be removed by limited proteolysis with trypsin at Arg-100 (Perham & Roberts, 1981; Packman & Perham, 1987). Such proteolysis 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) side-chains (Perham & Roberts, 1981). Thus, it is likely that at least some of these sharp resonances derive from the distinct (alanine + proline)-rich segment of the E2o polypeptide chain at approx. residues 150 to 172, which strongly resembles the three long (alanine + proline)-rich segments in the E2p chain (Spencer *et al.*, 1984), and which also encompasses sites of cleavage for limited proteolysis (Packman & Perham, 1987). Both these lines of evidence point to solvent exposure and conformational flexibility for this stretch of polypeptide chain. The position of this sequence in the E2o chain corresponds with that of the much shorter (alanine + proline)-rich sequence (residues 370–377) of the E2p chain referred to above, although the sequences of the E2p and E2o chains are not otherwise markedly homologous in this region (Spencer *et al.*, 1984). If this region of the E2p and E2o polypeptide chains, which links the E3-binding domain to the inner core-forming catalytic domain, does enjoy some degree of conformational freedom in the intact complexes, it might provide a structural basis for interpreting the hitherto puzzling fluorescence data which suggest that E3 bound to the PDH complex is quite mobile (Grande *et al.*, 1980). Further genetic reconstructions of the E2p and E2o polypeptide chains should enable unequivocal answers to these questions to be obtained.

Flexible segments of polypeptide chain have been identified in a variety of other protein systems, such as rabbit myosin A₁ light chains (Henry *et al.*, 1982), ox β -crystallin B₁-chains (Berbers *et al.*, 1983), immunoglobulins (Endo & Arata, 1985) and histones (Cary *et al.*, 1981). The amino acid sequences of flexible regions are frequently rich in alanine and proline residues, interspersed with charged amino acids, which may indicate an emerging pattern for the design of such regions. ^1H -n.m.r. spectroscopy of a 37-residue synthetic peptide with the same amino acid sequence as the conformationally mobile segment from the N-terminus of the myosin light chain has suggested that the (alanine + proline)-rich region of this peptide exists in an extended conformation (Bhandari *et al.*, 1986). This structure is imposed by the stereochemistry of the *trans* Ala-Pro dipeptide unit, which is repeated seven times in the middle of the peptide. Similarly, ^1H -n.m.r. spectroscopy of a 32-residue synthetic peptide with the same amino acid sequence as the innermost of the three long (alanine + proline)-rich segments of the E2p polypeptide chain has revealed that its six Ala-Pro peptide bonds are also all-*trans* (Radford *et al.*, 1986). These peptides therefore do not enjoy unrestricted conformational flexibility in free solution. Further n.m.r. spectroscopy of synthetic peptides and genetically reconstructed 2-oxo acid dehydrogenase complexes can be expected to throw more light on this problem and on the part such peptides play in biological activity.

We are grateful to the Science and Engineering Research Council for financial support. We thank the S.E.R.C. and the Medical Research Council for the award of Research Studentships to S. E. Radford and J. S. Miles, respectively, and Professor G. C. K. Roberts for valuable discussion. The Cambridge authors are members of the Cambridge Biochemical NMR Group, whose facilities were funded jointly by the S.E.R.C. and the M.R.C.

REFERENCES

- Ambrose, M. C. & Perham, R. N. (1976) *Biochem. J.* **155**, 429–432
- Angelides, K. J. & Hammes, G. G. (1979) *Biochemistry* **18**, 1223–1229
- Angier, S. J., Miles, J. S., Srere, P. A., Engel, P. C. & Guest, J. R. (1987) *Biochem. Soc. Trans.*, in the press
- Bates, D. L., Danson, M. J., Hale, G., Hooper, E. A. & Perham, R. N. (1977) *Nature (London)* **268**, 313–316
- Berbers, G. A. M., Hoekman, W. A., Bloemendal, H., de Jong, W. C., Kleinschmidt, T. & Braunitzer, G. (1983) *FEBS Lett.* **161**, 225–229
- Berman, J. N., Chen, G.-X., Hale, G. & Perham, R. N. (1981) *Biochem. J.* **199**, 513–520
- Bhandari, D. G., Levine, B. A., Trayer, I. P. & Yeadon, M. E. (1986) *Eur. J. Biochem.* **160**, 349–356
- Bleile, D. M., Munk, P., Oliver, R. M. & Reed, L. J. (1979) *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4385–4389
- Cary, P. D., Crane-Robinson, C., Bradbury, E. M. & Dixon, G. H. (1981) *Eur. J. Biochem.* **119**, 545–551
- Coggins, J. R., Hooper, E. A. & Perham, R. N. (1976) *Biochemistry* **15**, 2527–2533
- Collins, J. H. & Reed, L. J. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 4223–4227
- Danson, M. J. & Perham, R. N. (1976) *Biochem. J.* **159**, 677–682
- Danson, M. J., Hale, G., Johnson, P., Perham, R. N., Smith, J. & Spragg, P. (1979) *J. Mol. Biol.* **129**, 603–617
- Endo, S. & Arata, Y. (1985) *Biochemistry* **24**, 1561–1568
- 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
- Grande, H. J., Van Telgen, H. J. & Veeger, C. (1976) *Eur. J. Biochem.* **71**, 509–518
- Grande, H. J., Visser, A. J. W. G. & Veeger, C. (1980) *Eur. J. Biochem.* **106**, 361–369
- Guest, J. R., Lewis, H. M., Graham, L. D., Packman, L. C. & Perham, R. N. (1985) *J. Mol. Biol.* **185**, 743–754
- Hackert, M. L., Oliver, R. M. & Reed, L. J. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 2907–2911
- Hale, G. & Perham, R. N. (1979) *FEBS Lett.* **105**, 263–266
- Hale, G. H., Hooper, E. A. & Perham, R. N. (1979) *Biochem. J.* **177**, 136–137
- Henry, G. D., Dalgarno, D. C., Marcus, G., Scott, M., Levine, B. A. & Trayer, I. P. (1982) *FEBS Lett.* **144**, 11–15
- Jardetzky, O. & Roberts, G. C. K. (1981) *NMR in Molecular Biology*, Academic Press, New York
- Packman, L. C. & Perham, R. N. (1986) *FEBS Lett.* **139**, 155–158
- Packman, L. C. & Perham, R. N. (1987) *Biochem. J.* **242**, 531–538
- Packman, L. C., Perham, R. N. & Roberts, G. C. K. (1982) *Biochem. J.* **205**, 389–396
- Packman, L. C., Stanley, C. J. & Perham, R. N. (1983) *Biochem. J.* **213**, 331–338
- 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. (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)* **282**, 474–477
- Radford, S. E., Laue, E. D. & Perham, R. N. (1986) *Biochem. Soc. Trans.* **14**, 1231–1232
- Reed, L. J. (1974) *Acc. Chem. Res.* **7**, 40–46
- Reed, L. J. & Willms, C. R. (1966) *Methods Enzymol.* **9**, 247–265
- Roberts, G. C. K., Duckworth, H. W., Packman, L. C. & Perham, R. N. (1983) *Ciba Found. Symp.* **93**, 47–62
- 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
- Stepp, L. R., Bleile, D. M., McRorie, D. K., Pettit, F. H. & Reed, L. J. (1981) *Biochemistry* **20**, 4555–4560

Received 5 February 1987/14 May 1987; accepted 27 July 1987