The EMBO Journal vol.10 no.13 pp.4267-4277, 1991

A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA synthetase

Claire Cerini, Pierre Kerjan[1], Martine Astier, Danielle Gratecos, Marc Mirande[1] and Michel Sémériva

Laboratoire de Génétique et Biologie Cellulaires, CNRS, Luminy, Case 907, 13288 Marseille, Cedex 9, France, and [1]Laboratoire d'Enzymologie, CNRS, 91198 Gif-sur-Yvette, Cedex, France.

Communicated by R.Buckingham

In higher eukaryotes, nine aminoacyl-tRNA synthetases are associated within a multienzyme complex which is composed of 11 polypeptides with molecular masses ranging from 18 to 150 kDa. We have cloned and sequenced a cDNA from *Drosophila* encoding the largest polypeptide of this complex. We demonstrate here that the corresponding protein is a multifunctional aminoacyl-tRNA synthetase. It is composed of three major domains, two of them specifying distinct synthetase activities. The amino and carboxy-terminal domains were expressed separately in *Escherichia coli*, and were found to catalyse the aminoacylation of glutamic acid and proline tRNA species, respectively. The central domain is made of six 46 amino acid repeats. In prokaryotes, these two aminoacyl-tRNA synthetases are encoded by distinct genes. The emergence of a multifunctional synthetase by a gene fusion event seems to be a specific, but general attribute of all higher eukaryotic cells. This type of structural organization, in relation to the occurrence of multisynthetase complexes, could be a mechanism to integrate several catalytic domains within the same particle. The involvement of the internal repeats in mediating complex assembly is discussed.

Introduction

Aminoacyl-tRNA synthetases are ubiquitous enzymes that catalyse the same overall reaction, the esterification of a specific tRNA by the cognate amino acid. This family of 20 enzymes, endowed with a common catalytic function, displays strikingly different structural parameters. In prokaryotes, these enzymes are monomers (α), dimers (α_2) or tetramers (α_4 and $\alpha_2\beta_2$), with subunit molecular weights ranging from 37 000 for one protomer of the dimeric tryptophanyl-tRNA synthetase to 108 000 for the monomeric valyl-tRNA synthetase (Schimmel, 1987).

During the past 10 years, extensive efforts have been made to clone and sequence the genes encoding aminoacyl-tRNA synthetases. From alanyl-tRNA synthetase in 1981 (Putney *et al.*, 1981) to cysteinyl-tRNA synthetase in 1991 (Eriani *et al.*, 1991; Hou *et al.*, 1991), the genes coding for all prokaryotic synthetases have now been characterized. The knowledge of the primary structures of the 20 enzymes has revealed the existence of two classes of aminoacyl-tRNA synthetases, depending on conserved sequence motifs (Eriani *et al.*, 1990a). Enzymes from class I (ArgRS, CysRS, GlnRS, GluRS, IleRS, LeuRS, MetRS, TrpRS, TyrRS, ValRS) display the two consensus sequences HIGH and KMSKS, while those of class II share three conserved motifs (AsnRS, AspRS, HisRS, LysRS, PheRS, ProRS, SerRS, ThrRS) or at least one of them (AlaRS, GlyRS). The aminoacyl-tRNA synthetases from eukaryotes for which the primary sequences are known have the same consensus sequences as their prokaryotic counterparts. The three-dimensional crystal structures of three class I [TyrRS (Brick *et al.*, 1989), MetRS (Brunie *et al.*, 1990), GlnRS (Rould *et al.*, 1989)] and two class II [SerRS (Cusack *et al.*, 1990), AspRS (Ruff *et al.*, 1991)] enzymes have been described. Whereas class I synthetases possess a common nucleotide binding domain arranged according to a Rossmann fold, class II enzymes do not. From an evolutionary point of view, the partition of aminoacyl-tRNA synthetases into two structural classes suggests that this family of 20 enzymes has evolved from at least two different ancestral proteins.

Molecular and biochemical studies on eukaryotic aminoacyl-tRNA synthetases have led to the identification of several features that distinguish these enzymes from their prokaryotic homologues (Mirande, 1991). As a general rule, eukaryotic aminoacyl-tRNA synthetases have subunit masses significantly higher than those of the homologous enzymes from prokaryotes. In addition, 10 aminoacyl-tRNA synthetases from higher eukaryotes display a specific structural property: the formation of high molecular weight complexes. Two distinct complexes have been described. One of them is composed of a single aminoacyl-tRNA synthetase, valyl-tRNA synthetase, that associates with the heavy form of the eukaryotic elongation factor eEF-1 (Motorin *et al.*, 1988; Bec *et al.*, 1989). The other high molecular weight structure, referred to as the multisynthetase complex, contains the nine aminoacyl-tRNA synthetases specific for proline, glutamic acid, isoleucine, leucine, methionine, glutamine, lysine, arginine and aspartic acid (Mirande *et al.*, 1985).

The multisynthetase complex seems to be a common denominator of all higher eukaryotic cells or tissues studied so far, from the nematode *Ascaris suum* to humans. The multisynthetase complexes purified from several sources are composed of 11 polypeptides with masses ranging from 18 to 150 kDa (Mirande, 1991). Seven of the nine aminoacyl-tRNA synthetases were identified as individual polypeptide components of this complex. Their associated subunit molecular weights are essentially similar to those of the corresponding enzymes from lower eukaryotes. One exception is the mammalian glutamyl-tRNA synthetase that has been assigned to the M_r 150 000 component, the largest polypeptide chain that composes the multisynthetase complex. The homologous enzyme of prokaryotic origin is a monomer with a much lower molecular weight: 54 000 Daltons. Recently, the ninth synthetase component of the complex, prolyl-tRNA synthetase, was also shown to

correspond to a polypeptide displaying an electrophoretic mobility identical to that of glutamyl-tRNA synthetase (P.Kerjan, M.Triconnet and J.P.Waller, to be published).

In order to test the possibility that, in higher eukaryotes, a single polypeptide could share two distinct aminoacyl-tRNA synthetase activities, efforts were made to clone, sequence and characterize a cDNA encoding for glutamyl- and/or prolyl-tRNA synthetase. This paper describes the isolation, from *Drosophila*, of a large cDNA that encodes both glutamyl- and prolyl-tRNA synthetase activities. This multifunctional enzyme is composed of an amino-terminal domain specifying glutamyl-tRNA synthetase and of a carboxy-terminal domain specifying prolyl-tRNA synthetase. In addition, the two synthetase moieties of this large polypeptide are separated by a central domain made of several repeated units composed of ~75 amino acid residues. The involvement of the latter domain in mediating the assembly of the other aminoacyl-tRNA synthetase components of the multisynthetase complex will be discussed. This is the first example of a naturally occurring chimeric aminoacyl-tRNA synthetase, that could be referred to as the 'gluprolyl-tRNA synthetase'.

Results

Isolation of cDNAs encoding the 190 kDa component of the multisynthetase complex

As will be described elsewhere (D.Gratecos, C.Cerini, M.Astier and M.Sémériva, in preparation), a λgt11 expression library constructed from *Drosophila* embryonic poly(A)$^+$ mRNA (Nolan *et al.*, 1986) has been screened with a rabbit polyclonal antiserum directed against an antigen specifically expressed in the cardioblasts of *Drosophila* embryos. Amongst several clones thus obtained, a 3.1 kbp cDNA clone (λGD51) was isolated which is unrelated to the initial antigen and has the potentiality of coding for an aminoacyl-tRNA synthetase. The sequence of the deduced protein (see later) showed extensive homologies with the C-terminal region of the putative human glutaminyl-tRNA synthetase deduced from the sequence of the incomplete Pz cDNA (Thömmes *et al.*, 1988). Since then, the complete sequence of the human synthetase has been reported (Fett and Knippers, 1991) and has largely confirmed the suspected identity of the *Drosophila* protein.

Search for a full-length cDNA was achieved by using the two *Eco*RI fragments of the λGD51 clone in a screen of an embryonic cDNA λZAP library. Eighteen independent cDNAs were recovered and the two longest ones (λZ6 and λZ7) were analysed in detail for further sequencing. The cDNA λZ6 contained a small unspliced intron at position 1242 that introduced two stop codons in the open reading frame. By contrast, this intron was not present in λZ7. A simplified restriction map of the λZ7 cDNA is shown in Figure 1.

The λZ7 cDNA is 5.3 kbp long and accounts for nearly 90 % of the full-length mRNA. Figure 2 summarizes a transcript analysis carried out with different fragments of the λZ7 cDNA. The 1.3 kbp *Eco*RI fragment from the 3′ end hybridized with two transcripts (6.1 kbp and 3.4 kbp respectively). The *Eco*RI−*Hin*dIII fragment (1.85 kbp) located in the 5′ part of the cDNA only recognized the longest transcript, whereas the central *Hin*dIII−*Bst*XI fragment (1.2 kbp) as well as the central *Hin*dIII−*Eco*RI fragment (1.8 kbp) hybridized with both transcripts. Finally, the 0.40 kbp *Eco*RI fragment situated in the 5′ end of the λZ7 cDNA only recognized the 6.1 kbp transcript. It appears therefore that all probes derived from the end 5′ of the central *Hin*dIII site could only hybridize with the longest 6.1 kbp transcript whereas all probes derived from the central region and the 3′ moiety could recognize both transcripts. In addition, whereas the transcript corresponding to the full-length mRNA is present at all stages of development, the shorter transcript, encompassing the 3′ end of the cDNA, is especially abundant in 5−10 h embryos until the first larval stage.

The transcript analysis shows that the λZ7 cDNA lacks ~800 nucleotides to be full-length. To extend the cDNA

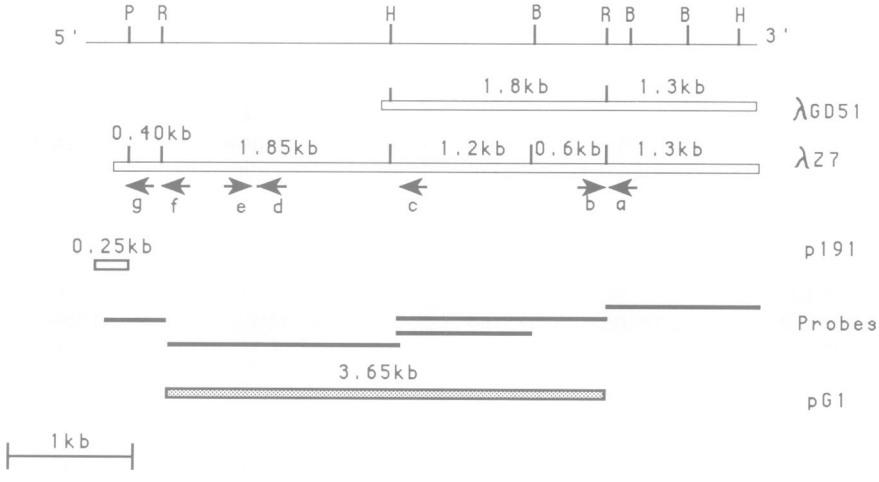

Fig. 1. cDNAs and sequence strategies. The 1.8 and 1.3 kbp *Eco*RI fragments of λGD51, the 1.85 kbp *Eco*RI−*Hin*dIII fragment of λZ7 and the 0.40 kbp *Eco*RI fragment of λZ7 were independently sequenced in both orientations. The arrows show the orientations and the locations of some of the oligonucleotides used in the sequencing. Oligonucleotides a and b were deduced from the sequence of the 1.8 and 1.3 kbp fragments of λGD51. They have been used as primers to sequence throughout the *Eco*RI site in λZ7 cDNA. Oligonucleotide c was devised to check the overlap across the *Hin*dIII site between λGD51 and λZ7. The same was done with oligonucleotide f close to the *Eco*RI site in the 5′ end of the λZ7 cDNA. Oligonucleotides d and e were used to ascertain the absence of the small intron in the cDNA. The λZ7 cDNA was extended towards its 5′ end by priming with oligonucleotide f or g (191). pG1 represents the insert used to produce fusion proteins in the vector pGEX-2T. The black lines represent the different fragments used as probes for the Northern analysis. B, *Bst*X1; H, *Hin*dIII; R, *Eco*RI; P, *Pst*I.

towards the 5′ end we used the Frohman's methodology (Frohman, 1990) which consists of a primer extension on the total poly(A)$^+$ mRNA followed by a polymerase chain reaction. As a primer we designed an oligonucleotide (191) downstream and close to the first *Pst*I site (see Figure 1) and complementary to the mRNA sequence. The longest cloned amplified (p191) sequence was 250 bp long and, as expected, partially overlapped with the λZ7 cDNA (Figure 1).

An antiserum raised against the fusion protein produced by λGD51 recognized in *Drosophila* embryos one major antigen with a molecular weight of ~180 kDa and a minor one of ~100 kDa (not shown). Experiments are in progress to determine whether these two different proteins are encoded by the two transcripts.

The cytological location of the gene was assigned by *in situ* hybridization of the 3.1 kbp cDNA on polytene chromosomes of third instar *Drosophila* larvae. Hybridization was detected at a single site in the distal region of the third chromosome in 95 CD (data not shown). To date no mutations that map to this site have been reported.

Nucleotide and deduced amino acid sequences

The combined overlapping cDNA sequence extends for 5455 nucleotides and contains a long open reading frame for 1714 amino acids (Figure 3) with a calculated molecular mass of 189 kDa. This molecular weight is consistent with that determined on Western blots by using an antiserum directed against the λGD51 fusion protein (see above). The presumptive initiation codon was ascribed to the first in-frame ATG at position 112.

The 5′ non-coding region is composed of at least 111 nucleotides with a TAG stop codon located at position 30, in-frame with the downstream initiation codon. The 3′ untranslated sequence, from the TAA termination codon at position 5254 to the first adenine residue of the poly(A) tract at position 5430, is relatively short. A putative polyadenylation signal ATTAAA (Breathnach and Chambon, 1981) is present at position 5407, 17 nucleotides before the poly(A) tract. The short intron, made up of 57 nucleotides, remaining in clone λZ6 described above is located at position 1242, within the open reading frame. The codon usage within the whole open reading frame is in good agreement with that of the consensus determined for proteins from *Drosophila* (Dessen *et al.*, 1990). The overall amino acid composition is rather hydrophilic, with a relative basicity of 1.21 (K+R+H/D+E). The primary structure of the encoded polypeptide was found to be homologous to the partial sequence of the corresponding human cDNA (Thömmes *et al.*, 1988) which has been tentatively assigned to human glutaminyl-tRNA synthetase (see Discussion), and to its recently published full-length sequence (Fett and Knippers, 1991).

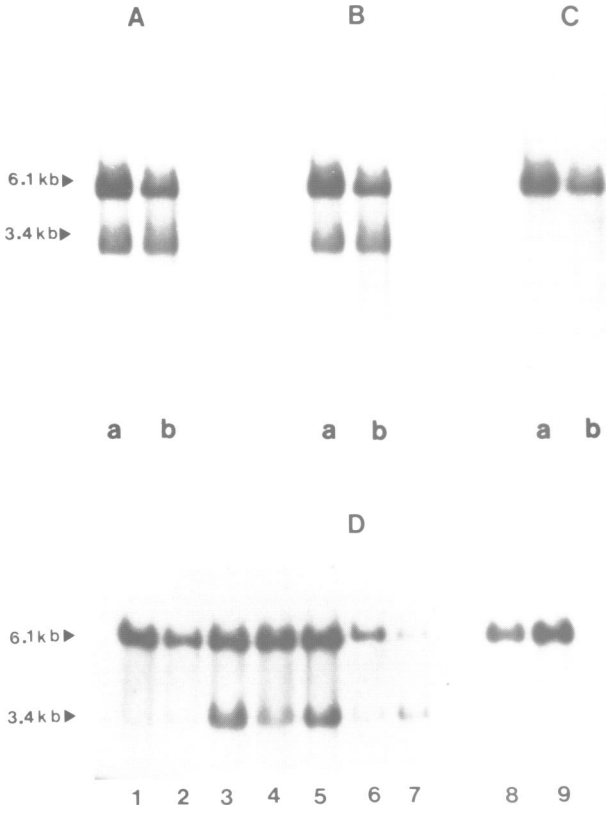

Fig. 2. Northern blot analysis. Each lane in A, B and C was loaded with 5 μg of poly(A)$^+$ mRNAs extracted from 0–20 h embryos (lane a) and third instar larvae (lane b). The blots were respectively probed with (**A**) the 1.3 kbp *Eco*RI fragment, (**B**) the 1.2 kbp *Hin*dIII–*Bst*XI fragment, (**C**) the 1.85 kbp *Eco*RI–*Hin*dIII fragment or the 0.4 kbp fragment. (**D**) Developmental Northern blot probed with the 1.8 kbp *Hin*dIII–*Eco*RI fragment. Lanes 1–9 were loaded with 5 μg of poly(A)$^+$ mRNA extracted from 0–2 h embryos, 2–5 h embryos, 5–10 h embryos, 12–24 h embryos, first instar larvae, third instar larvae, pupae, male adults and female adults respectively.

Homologies with other known aminoacyl-tRNA synthetases

As assessed by sequence analysis, the primary sequence of the *Drosophila* protein can be divided into three regions. Two of them, domain I (residues 1–751) and domain III (residues 1174–1714) have the structure requirements to account for functional aminoacyl-tRNA synthetases, according to sequence homologies with already known synthetases.

Domain I displays convincing homologies with *Escherichia coli* glutamyl-tRNA synthetase (22% of identical amino acids) and even stronger homologies with glutaminyl-tRNA synthetases from *E.coli* (37%) and *Saccharomyces cerevisiae* (35%) (Figure 4). It contains the two short highly conserved sequences identified in some other aminoacyl-tRNA synthetases: the HIGH signature sequence (Schimmel, 1987) at position 217, and a derivative of the KMSKS consensus sequence (Hountondji *et al.*, 1986), VLSKR at position 438. Therefore, this protein domain belongs to class I aminoacyl-tRNA synthetases (Eriani *et al.*, 1990a).

The extreme amino-terminal region, from residues 1 to 200 displays no significant homology with other known synthetases, including the corresponding region of the human enzyme (Fett and Knippers, 1991), which is composed of only 123 amino acids. No particular secondary structure can be predicted. Glutaminyl-tRNA synthetase from *S.cerevisiae* has a large amino-terminal extension of 249 amino acids (Ludmerer and Schimmel, 1987a; Figure 4). No evidence of a specific role for this polypeptide moiety could be obtained (Ludmerer and Schimmel, 1987b).

The region from the human cDNA (Fett and Knippers, 1991) homologous to domain I of the *Drosophila* cDNA has >66% of identical residues (Figure 4), without taking into account many conservative changes, and must therefore specify the same synthetase activity. The finding that glutaminyl-tRNA synthetases from other species display the

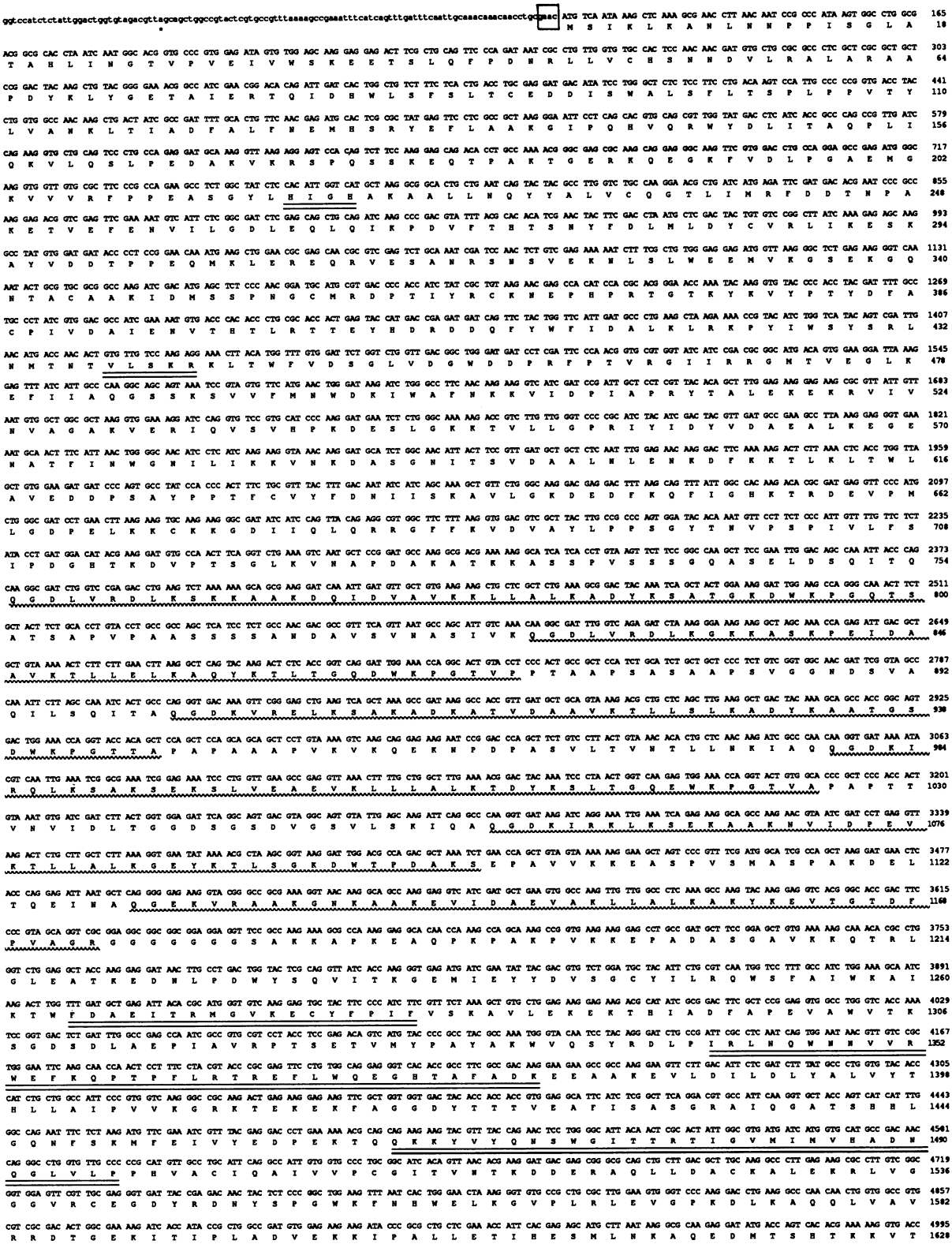

Fig. 3. Nucleotide and deduced amino acid sequences of the overlapping cDNAs. Translation initiation and polyadenylation signal consensus sequences are boxed. The upstream in-frame stop codon and the termination codon are marked with an asterisk. The consensus motifs of class I and class II synthetases as well as the six repeated motifs are underlined.

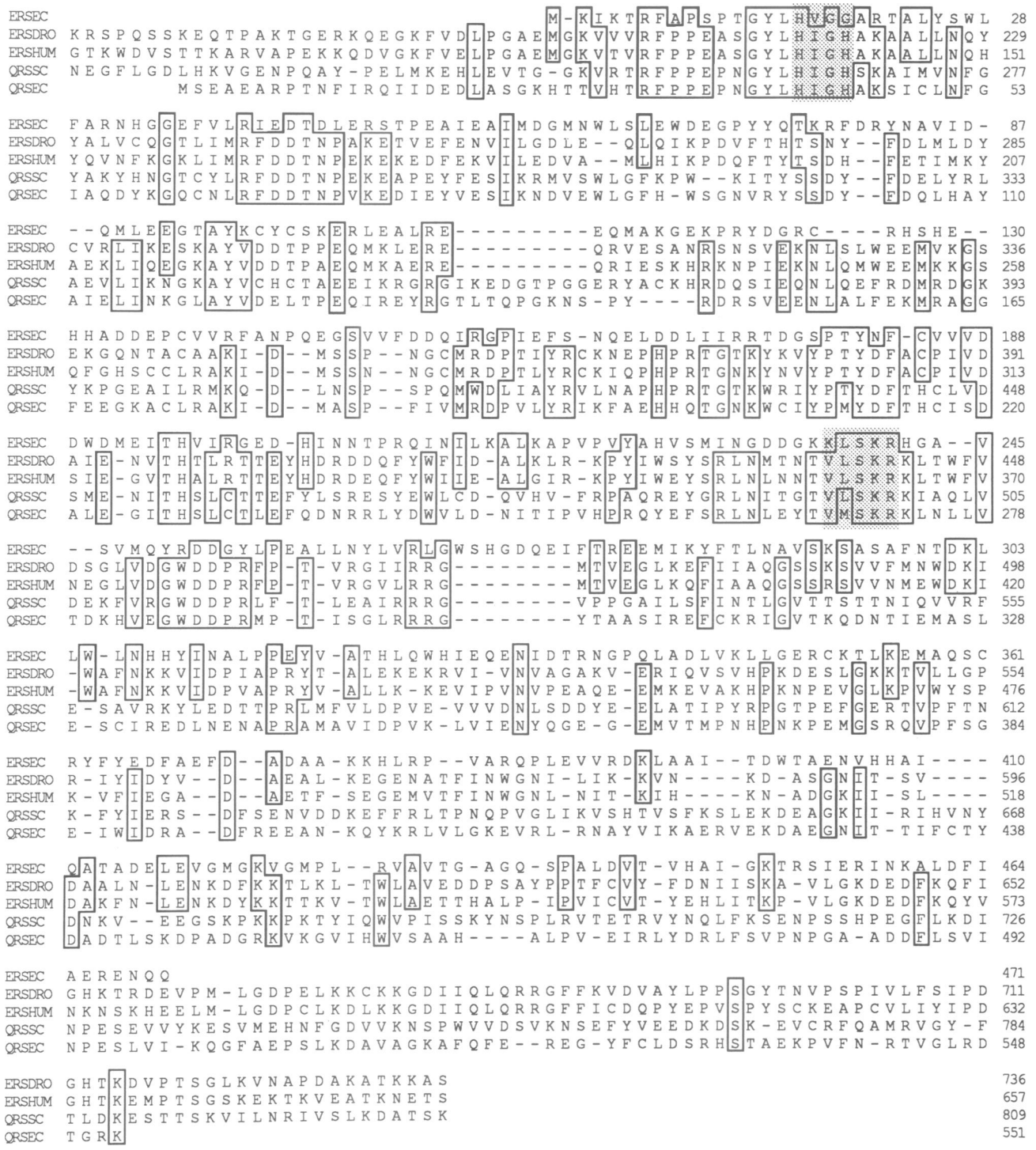

Fig. 4. Comparison of glutamyl- and glutaminyl-tRNA synthetase sequences. The sequences of glutamyl-tRNA synthetases from *E.coli* [ERSEC, (Breton *et al.*, 1986)], *Drosophila* (ERSDRO) and human [ERSHUM, (Fett and Knippers, 1991)] are aligned with those of glutaminyl-tRNA synthetases from *S.cerevisiae* [QRSSC, (Ludmerer and Schimmel, 1987a)] and *E.coli* [QRSEC, (Hoben *et al.*, 1982)]. Positions where amino acid sequences from the two amino-terminal domains for cDNAs from *Drosophila* and human are identical to the two glutaminyl-tRNA synthetase species, to *E.coli* glutamyl-tRNA synthetase, or both, are boxed. The two consensus sequences characteristic of aminoacyl-tRNA synthetases from class I are shaded.

more convincing homologies prompted Thömmes *et al.* (1988) to assign this polypeptide domain to a glutamine-specific synthetase. However, according to our previous assignments (Mirande *et al.*, 1982b), this polypeptide should correspond to glutamyl-tRNA synthetase. This question will be further addressed by measuring synthetase activity expressed in *E.coli* by the corresponding cDNA region (see below).

The carboxy-terminal domain of the *Drosophila* protein (domain III) displays convincing homologies with prolyl-tRNA synthetase from *E.coli* (Figure 5). This observation raises the possibility that a single polypeptide may support

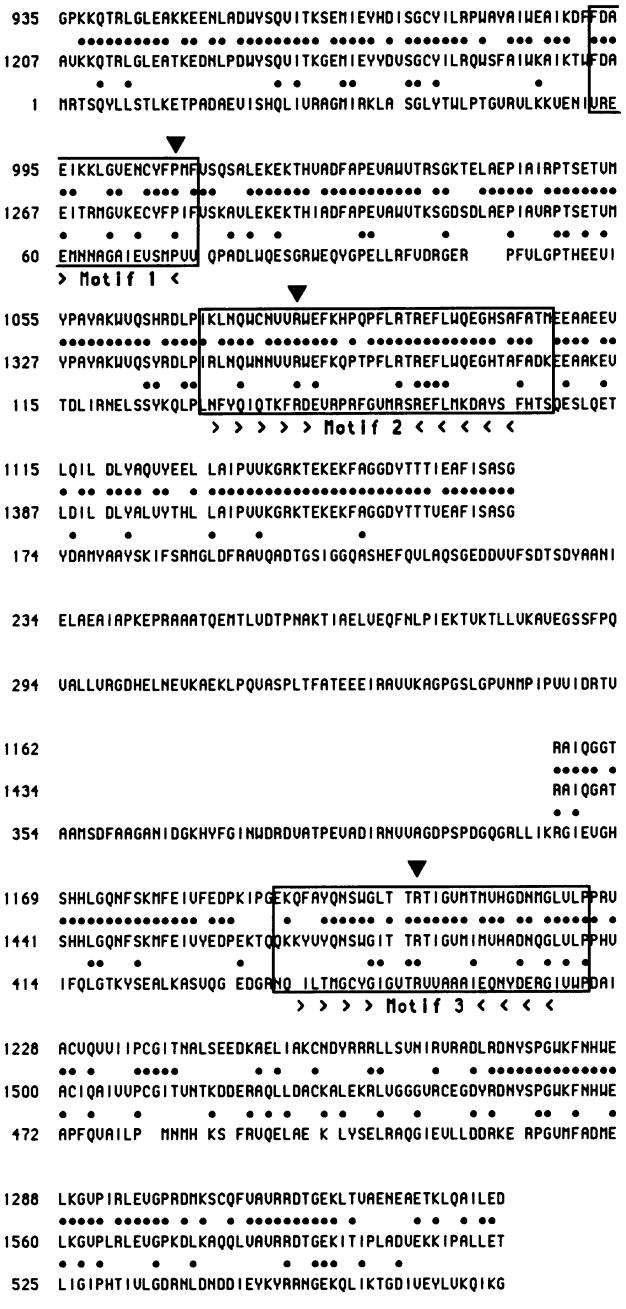

Fig. 5. Comparison of prolyl-tRNA synthetase sequences. The amino acid sequence of the carboxy-terminal domain for the cDNA from *Drosophila* (residues 1207–1607, central sequence) is aligned with the corresponding region for the cDNA from human [(Fett and Knippers, 1991), residues 935–1335, upper sequence] and with the primary structure of *E.coli* prolyl-tRNA synthetase [(Eriani *et al.*, 1990a), lower sequence]. Positions with identical residues are indicated by dots. The three sequence motifs characteristic of class II aminoacyl-tRNA synthetases (Eriani *et al.*, 1990a) are boxed. The arrowheads point at the invariant residues.

glutamyl- as well as prolyl-tRNA synthetase activities. The three conserved motifs found in class II synthetases (Eriani *et al.*, 1990a) are also present in the fly protein. The carboxy-terminal region of the so-called human glutaminyl-tRNA synthetase (Fett and Knippers, 1991) also shows extensive homologies with domain III of the *Drosophila* protein (>73% of identical amino acids). This comparison suggests that the human polypeptide may also possess both aminoacyl-tRNA synthetase specificities.

The central region of the *Drosophila* protein (domain II) is composed of six repeated motifs that are highly conserved. They are composed of 46 amino acids, separated by 26–33 variable amino acids (Figure 6). According to secondary structure prediction analysis, a high probability of α-helical structures should prevail for the central region of these repeated units, whereas the flanking sequences would adopt a disordered structure. Only three of these repeats could be detected in the human polypeptide (Fett and Knippers, 1991), although with the same organization and predicted secondary structures. A computer assisted search of protein data banks with the consensus sequence deduced from the homologies between the human and *Drosophila* motifs revealed the presence of related motifs within the extreme amino-terminal part of histidyl-tRNA synthetase from human or CHO cells, as already reported by Fett and Knippers (1991), of tryptophanyl-tRNA synthetase from human or bovine, and of peptide chain release factor from rabbit (Figure 6). In those cases, only one copy of the conserved motif is present.

The amino- and carboxy-terminal domains encoded by the large cDNA specify glutamyl- and prolyl-tRNA synthetase activities, respectively

The analysis of sequence homologies strongly suggests that the two putative synthetase domains may support aminoacylation of tRNA with glutamate or proline. In order to test this hypothesis, we expressed in *E.coli* isolated domain I and domain III, and measured the ability of fusion proteins to aminoacylate tRNA with glutamate and proline, respectively.

Domain I was produced from the large *Eco*RI fragment of 3.65 kbp (Figure 1) encoding residues 139–1355 of the large polypeptide. This fragment was subcloned into the expression vector pGEX-2T. The resulting plasmids pG1 and pG2 correspond to a proper or opposite orientation with respect to the IPTG inducible *lac* promoter, respectively. The fusion protein produced from pG1 had a molecular mass of ~150 kDa (result not shown), which is consistent with the size of the insert-encoded protein plus that of the bacterial glutathione S-transferase (25 kDa), whereas the protein produced from pG2 corresponded to glutathione S-transferase alone.

In a first series of experiments, the recombinant bacteria were grown at 37°C and IPTG induction of the expression of fusion proteins was performed at 37°C. As shown in Table I, when assayed with beef liver tRNA, glutamyl-, and to a lesser extent glutaminyl-tRNA synthetase activities could be detected in crude extracts from pG1 transformants. Most of the expressed fusion protein was found to be insoluble, forming inclusion bodies. We also expressed in pGEX-2T the 1.85 kbp *Eco*RI–*Hin*dIII fragment (Figure 1). The resulting fusion protein was found to be even more insoluble. The ability of recombinant proteins to form inclusion bodies could be reduced following expression at lower temperature (Schein and Noteborn, 1988). The insolubility is believed to be the reflection of temperature-sensitive denaturation of the polypeptide chain. An abnormal conformation of the pG1-encoded protein could *a priori* explain the relaxed specificity we observed. In accordance with this proposal, only glutamyl-tRNA synthetase activity was detected in crude extracts from cells grown and induced at 30°C, by using either beef liver or *Drosophila* tRNA (Table I). Homologous tRNA from *Drosophila* was found to be 3-fold more

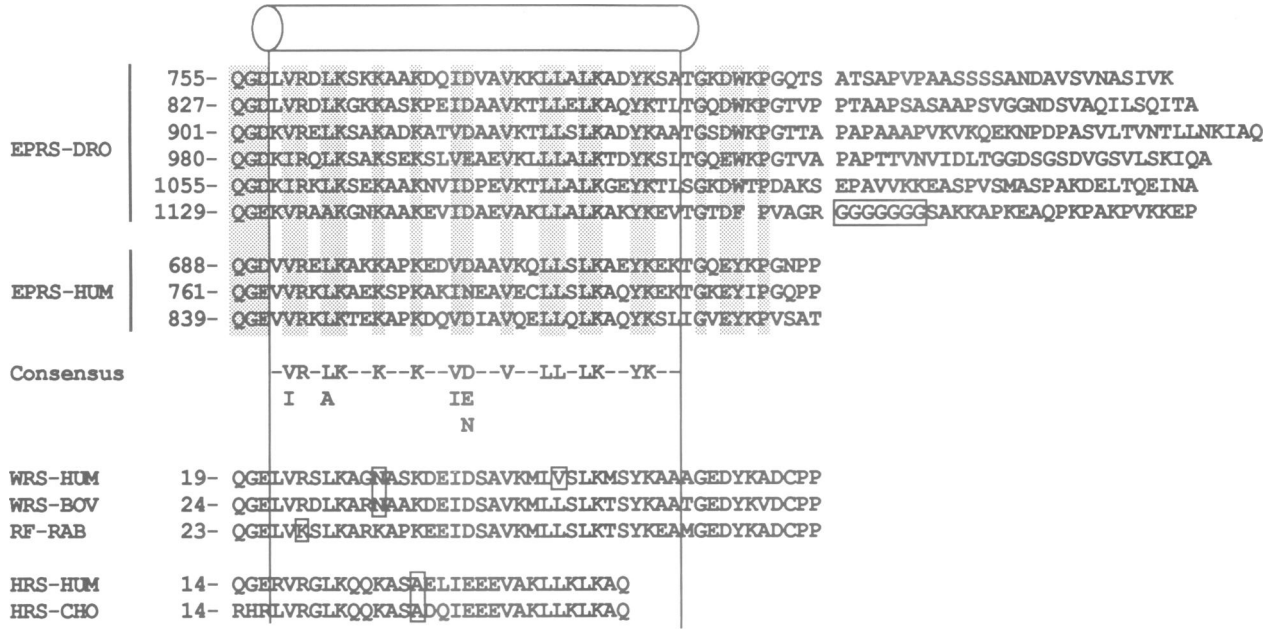

Fig. 6. Comparison of the six repeated motifs of domain II with the same motifs present in other proteins. The segment from residues 755 to 1201 from the *Drosophila* polypeptide, corresponding to domain II of gluprolyl-tRNA synthetase (EPRS-DRO), is represented by aligning the six repeated units. The glycine spacer that follows the last repeat is boxed. The three repeated units from the human gluprolyl-tRNA synthetase (Fett and Knippers, 1991) are indicated (EPRS-HUM). Positions where the same amino acid or a conservative substitution is present are shaded. The consensus sequence deduced from the comparison of the nine motifs is shown for the region that is presumed to be α-helical. The homologous sequences from other proteins are indicated below the consensus sequence: tryptophanyl-tRNA synthetase from human (WRS-HUM; EMBL Library, code HSWRSAA) or bovine (WRS-BOV; EMBL Library, code BTTRS2); peptide chain release factor from rabbit [RF-RAB, (Lee *et al.*, 1990)]; histidyl-tRNA synthetase from human or Chinese hamster ovary cells [HRS-HUM and HRS-CHO, respectively, (Tsui and Siminovitch, 1987)]. Residues that diverge from the consensus sequence are boxed.

Table I. Aminoacyl-tRNA synthetase activities in crude extracts from *E.coli*

			Aminoacyl-tRNA synthetase assayed[a]						
			A[b]		B[c]		C[d]		
Strain	Plasmid or phage	IPTG	GluRS	GlnRS	GluRS	GlnRS	GluRS	GlnRS	ProRS
JM101	pG1	−	16.8	1.3	7.5	1.1	5.8	2.3	
JM101	pG1	+	44.0	1.3	15.3	1.1	27.5	7.4	
JM101	pG2	−			1.0	1.3	1.0	1.1	
JM101	pG2	+	1.0	1.0	1.0	1.0	1.0	1.0	
Y1089	λGD51	−							1.0
Y1089	λGD51	+							62.5
Y1089	λgt11	−							1.0
Y1089	λgt11	+							1.0

[a] For each aminoacyl-tRNA synthetase assayed, results are expressed in relative specific activities.
[b] A: cells grown at 30°C, and synthetase activities measured with crude tRNA from *Drosophila*.
[c] B: cells grown at 30°C, and synthetase activities measured with crude tRNA from beef liver.
[d] C: cells grown at 37°C, and synthetase activities measured with crude tRNA from beef liver.

efficiently aminoacylated. Eukaryotic tRNAs were poorly aminoacylated by the bacterial, endogenous enzymes. The 44-fold increase in specific activity unambiguously ascertains the assignment of domain I to a eukaryotic glutamyl-tRNA synthetase enzyme domain.

Domain III was produced from the λGD51 clone which encodes the carboxy-terminus of the cDNA, starting from residue 733. The size of the fusion protein (210 kDa) produced after induction by IPTG was in agreement with that expected from the fusion with bacterial β-galactosidase. Prolyl-tRNA synthetase activity was assayed using beef liver tRNA, a poor amino acid acceptor tRNA for the *E.coli* enzyme. As shown in Table I, IPTG induction of the fusion protein resulted in a 62-fold increase in prolyl-tRNA synthetase activity.

Thus, glutamyl- and prolyl-tRNA synthetases are two distinct enzyme domains of the same polypeptide chain which deserves the name of gluprolyl-tRNA synthetase.

Discussion

The glutamyl- and prolyl-tRNA synthetase domains

In mammalian cells, and almost certainly in all metazoan species, nine aminoacyl-tRNA synthetases are encountered as components of a stable, high molecular weight structure (Mirande, 1991). The ubiquitous multisynthetase complex

contains the synthetases specific for proline, glutamic acid, isoleucine, leucine, methionine, glutamine, lysine, arginine and aspartic acid. It includes 11 polypeptides with masses ranging from 18 to 150 kDa. In order to elucidate molecular features involved in complex assembly, we have undertaken the cloning and analysis of cDNAs coding for synthetase components of the complex. In the present study, we describe the cloning and sequencing of a cDNA from *Drosophila* encoding the largest polypeptide of this complex. We show that the encoded protein is a multifunctional enzyme bearing both glutamyl- and prolyl-tRNA synthetase activities.

This conclusion is drawn from two types of evidence: (i) analysis of the primary sequence and its comparison with those of prokaryotic enzymes; (ii) direct activity measurements from fusion proteins produced in *E.coli*. These latter experiments allow unambiguous assignment of the 5′ and 3′ coding regions of the cloned cDNA to glutamyl- and prolyl-tRNA synthetase, respectively.

Recently Fett and Knippers (1991) have reported the deduced amino acid sequence of what is likely to be the human homologue of the *Drosophila* gluprolyl-tRNA synthetase. Approximately 70% of identities are recovered within the catalytic domains I and III of the enzymes from the two species. Considering this strong similarity, the human and *Drosophila* enzymes are likely to display the same aminoacylation specificities. This conclusion is at variance with the assignment of the human cDNA to glutaminyl-tRNA synthetase (Fett and Knippers, 1991).

However, this latter assignment was only based on sequence comparisons with no direct evidence that the cloned human cDNA actually expresses glutaminyl-tRNA synthetase activity. The observation that antibodies raised against the cDNA-encoded protein allowed immunoprecipitation of glutaminyl-tRNA synthetase activity (Thömmes *et al.*, 1988), which is also one of the synthetase components of the complex containing glutamyl-tRNA synthetase, does not prove the identity of the human protein. Indeed, any antibody able to recognize one of the entities of the multisynthetase complex co-immunoprecipitates all of the other components (Mirande *et al.*, 1982c).

Noteworthy is the finding that glutamyl-tRNA synthetase from *Drosophila* or human is more homologous to *E.coli* glutaminyl-tRNA synthetase (37%) than to *E.coli* glutamyl-tRNA synthetase (22%). As a general rule, aminoacyl-tRNA synthetases from higher eukaryotes are more closely related to their lower eukaryotic counterparts than to the prokaryotic ones. Among other examples, there are 26% and 21% identities between the aspartyl-tRNA synthetase from *E.coli* and the homologous enzymes from yeast and mammals respectively (Eriani *et al.*, 1990b ; Sellami *et al.*, 1986; Mirande and Waller, 1989), whereas 58% identities are recovered when the yeast and the higher eukaryotic enzymes are compared. The sequence of the yeast glutamyl enzyme is not yet known. However, the finding that identical homology scores are obtained when the *Drosophila* or human sequences are compared with the bacterial or yeast glutaminyl-tRNA synthetase reinforces their assignment to glutamyl-tRNA synthetase. This also agrees with previously reported results establishing that rabbit glutamyl-tRNA synthetase corresponds to the largest polypeptide of the complex with a mass of 150 kDa, whereas glutaminyl-tRNA synthetase was attributed to the M_r 96 000 component (Mirande *et al.*, 1982b).

Glutamyl- and glutaminyl-tRNA synthetases are two closely evolutionarily related enzymes, displaying a high level of sequence similarity (Breton *et al.*, 1990). Moreover, all Gram-positive bacteria as well as plant and animal organelles lack glutaminyl-tRNA synthetase activity, glutamyl-tRNA synthetase ensuring aminoacylation of $tRNA^{Glu}$ and $tRNA^{Gln}$ by glutamate, the formation of $Gln\text{-}tRNA^{Gln}$ requiring a transamidation step (Schön *et al.*, 1988, and references therein). It is generally believed that modern bacteria and eukaryotic cells arose from a common ancestral organism. One can speculate that this unknown ancestor was devoid of glutaminyl-tRNA synthetase activity, as already discussed (Schön *et al.*, 1988). If glutaminyl-tRNA synthetase appeared after divergence between modern prokaryotes and eukaryotes, independent evolution of the Glx enzyme family in these organisms may explain the above mentioned paradox; i.e. the closer relationship observed between glutamyl-tRNA synthetase from higher eukaryotes and glutaminyl-tRNA synthetase from *E.coli* than between the two homologous glutamyl-enzymes.

The carboxy-terminal moiety of the 190 kDa polypeptide (residues 1207−1607) is closely homologous to the corresponding region of the human glutaminyl-tRNA synthetase (Fett and Knippers, 1991), henceforth referred to as human glutamyl-tRNA synthetase (see above). The alignment shown in Figure 5 reveals the presence of 283 conserved amino acids out of 506, and several other conservative changes. This implies that these two domains are likely to fulfil the same function. According to immunological studies, it has been claimed that the C-terminal half of human glutamyl-tRNA synthetase is covered by other components of the multisynthetase complex (Schray *et al.*,1990) and may therefore be involved in the formation of the multisynthetase complex.

On the other hand, we have found that the carboxy-terminal domains of the 190 kDa *Drosophila* and human polypeptides share the three sequence motifs characteristic of aminoacyl-tRNA synthetases from class II. Comparison with primary structures of all known aminoacyl-tRNA synthetases, including the 10 prokaryotic enzymes from class II, allowed detection of significant homologies with prolyl-tRNA synthetase from *E.coli*, whose sequence has recently been published (Eriani *et al.*, 1990a). The extent of homology is rather low (22% of identical amino acids), but the composition and location of motifs 1, 2 and 3 are well conserved. Expression in *E.coli* of prolyl-tRNA synthetase activity by the 3′ region of the cDNA from *Drosophila* unambiguously demonstrates that it encodes an active enzyme.

The alignment shown in Figure 5 made clearly apparent a deletion of 184 amino acids in the eukaryotic enzymes, between motifs 2 and 3 of the bacterial synthetase. It should be noticed that the alignment of the motifs region for class II synthetases (Eriani *et al.*, 1990a) revealed a conserved spacing between the two first motifs (40−80 residues) and a more variable gap length between motifs 2 and 3 (70−300 residues). This suggests that the latter intermotif region is not important for the catalytic activity.

The internal repeated units: a template for complex assembly?

A most striking structural feature of this multifunctional aminoacyl-tRNA synthetase is the presence of six repeated units between the synthetase domains. This spacer region,

composed of ~75 amino acids repeated six times, constitutes ~25% of the molecule. According to several algorithms of secondary structure prediction (Chou and Fasman, 1978; Garnier *et al.*, 1978), the central region of these repeats is presumed to be α-helical. The inter-helical segments are rich in proline and glycine residues, and may thus contain many turns.

The helices are composed of 32 amino acids, comprising 13–17 hydrophobic residues and 12 or 13 charged residues per helix motif. The six helical motifs are highly homologous (60–70% identity), but the conserved residues are not evenly distributed. This is clearly apparent from the helical net diagram shown in Figure 7, where the consensus sequence alone is reported. The invariant residues are twisted counter-clockwise around the helix. Furthermore, the corresponding helical sections from the three tandemly repeated units of the human polypeptide display exactly the same consensus sequence, which is organized similarly. Since the conserved residues are not randomly distributed we believe that they may serve as an important biological interface.

One particular feature that distinguishes mammalian aminoacyl-tRNA synthetases from the corresponding enzymes from *E.coli* and yeast, is the occurrence of a multienzyme complex comprising nine synthetases, including glutamyl- and prolyl-tRNA synthetases. Earlier studies carried out on purified aminoacyl-tRNA synthetase complexes have shown that hydrophobic interactions are involved in the assembly of their components (Johnson *et al.*, 1980; Sihag and Deutscher, 1983; Lazard *et al.*, 1985; Cirakoglu and Waller, 1985). Jacobo-Molina *et al.* (1989) observed that the amino-terminal extremity of human aspartyl-tRNA synthetase may form an amphiphilic helix, and therefore may account for the complex association domain of mammalian aspartyl-tRNA synthetase. The structural features displayed by the central region of the 190 kDa polypeptide, made of helical domains rich in hydrophobic amino acids, should be optimally suited to serve as a template for association of the other components of the multisynthetase complex. In addition, since lysyl- and aspartyl-tRNA synthetases, two individual components of this complex, are dimers, it must be inferred that the complex displays two symmetrical association sites. Determination of the relative stoichiometries of the polypeptide components of the complex from rabbit liver indicated that two copies of the corresponding multifunctional protein are found per complex particle (Mirande *et al.*, 1982a). Therefore, helical repeated units, present at two copies per complex molecule, are a good candidate for mediating complex assembly.

The organization of the gluprolyl-tRNA synthetase is quite different from that described for other known multifunctional enzymes (Coggins and Hardie, 1986, and references therein). In most cases, the discrete functional domains are linked by connective polypeptides essentially devoid of secondary structure. In recent years, numerous proteins containing internal repeats have been characterized, and most have been classified in a relatively small number of families. In cases where the repeated sequences have a known function, they are usually involved in physical interactions with other proteins (EGF-like repeats, extracellular matrix proteins, cytoskeleton proteins, muscular proteins ...). It is therefore tempting to ascribe to the central domain of gluprolyl-tRNA synthetase a role in protein–protein interactions.

In connection with their possible involvement in complex assembly, two observations are noteworthy. First, the

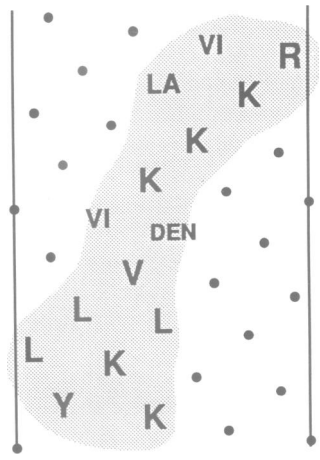

Fig. 7. Helical net-diagram of the consensus sequence of the repeated units. The α-helical structure predicted for the central region of the six imperfectly repeated motifs is represented by a net-diagram (Dunnill, 1968). Only the invariant residues from the consensus sequence shown in Figure 6 are indicated. Dots represent the positions of residues that are not conserved in all repeats.

corresponding polypeptide from the interdomain region of the human counterpart is composed of only three repeats (Fett and Knippers, 1991). However, whereas the multisynthetase complexes from mammalian sources have been extensively studied, only preliminary studies have been carried out in *Drosophila* (Shafer *et al.*, 1976). The actual composition of the latter complex remains to be established. Second, as indicated in Figure 6, a single copy of the repeated motif is found at the amino-terminal extremity of two synthetases that have never been reported as components of multienzyme complexes, tryptophanyl-tRNA synthetase from human or bovine and histidyl-tRNA synthetase from human or CHO cells, and of peptide chain release factor from rabbit. However, tryptophanyl-tRNA synthetases display one or two substitutions as compared with the consensus motif (Lys→Asn, Leu→Val), and histidyl-tRNA synthetases have an incomplete motif with one substitution (Lys Ala), which would suggest that these motifs do not fulfill the same function.

This motif is not encountered in other aminoacyl-tRNA synthetases from higher eukaryotes whose primary sequences are presently known and in synthetases from prokaryotes or lower eukaryotes, enzymes that do not associate into high molecular weight structures. Whereas all these observations strongly argue in favour of a triggering role of these tandemly repeated structures in the association of the other synthetase components of the complex, the possibility that they may serve another function cannot be completely dismissed.

Why does evolution select for a multifunctional aminoacyl-tRNA synthetase?

Our results demonstrate that prolyl- and glutamyl-tRNA synthetases from *Drosophila* are encoded by the same cDNA. These two synthetase activities are expressed from independent and autonomous domains of a large polypeptide chain. The finding that the isolated synthetase domains are able to aminoacylate their cognate tRNA suggests that this naturally occurrring chimeric protein arose by a gene fusion mechanism between individual, pre-existing genes. The fusion of monofunctional genes that are not clustered in the genome implies a genetic transposition step. In *E.coli*, the

proS and *gltX* genes, coding for prolyl- and glutamyl-tRNA synthetases, are located at 5 and 52 min, respectively, on the genetic map (Bachmann, 1990).

The 190 kDa polypeptide described in the present study is one of the few examples of a multifunctional enzyme that catalyses two non-consecutive steps in a biosynthetic pathway (Coggins and Hardie, 1986, and references therein). This is quite unusual, although a similar situation has been described for aspartokinase-homoserine dehydrogenase I from *E.coli* (Katinka *et al.*, 1980). However, in this latter case, the two activities are regulated by threonine at a single allosteric site.

It is generally believed that the formation of multifunctional enzymes confers to them selective advantages,i.e. protection of intermediates, substrate channelling, regulation of catalytic activity or co-ordinate regulation of the expression of individual components. Although the latter advantage undoubtedly applies to the multisynthetase protein, we suggest that the major advantage conferred by this type of structural organization could be related to spatial organization of the translational machinery in higher eukaryotic cells. It is conceivable that the formation of multifunctional proteins is the reflection of an ultimate step towards cellular compartmentalization of the protein synthesis machinery. Covalent attachment of synthetases could provide a more efficient alternative to promote stable association of individual proteins.

Materials and methods

Isolation of cDNAs

Isolation of the initial cDNA clone (λGD51) was performed by screening an embryonic λgt11 library (Nolan *et al.*, 1986) essentially as described in Huynh *et al.* (1985), with the use of a rabbit antiserum directed against a heart-specific antigen in *Drosophila* (D.Gratecos, C.Cerini, M.Astier and M.Sémériva, in preparation). The secondary antibody was coupled to alkaline phosphatase (Promega).

To obtain more complete cDNAs, the two *Eco*RI fragments (1.3 and 1.8 kbp) of the λGD51 clone were labelled with ^{32}P by random priming (Multiprime oligolabelling, Amersham) and used to probe a cDNA λZAP library constructed from *Drosophila* embryonic poly(A)$^+$ mRNA (Stratagene). The λZAP clones were excised (Short *et al.*, 1988) to produce the cDNAs in the phagemids BlueScript SK$^-$ which were further analysed. The different cDNAs (18) were aligned by restriction analysis. The two longest ones λZ6 (5.45 kbp) and λZ7 (5.30 kbp) were used for the subsequent studies.

Subcloning and sequencing

The relevant regions from phage, plasmid and PCR-generated cDNAs were subcloned into BlueScript (Stratagene) in both orientations by standard methods (Sambrook *et al.*, 1989). Single-stranded templates were sequenced directly or subjected to deletions (Henikoff, 1987). DNA sequencing was performed by the dideoxy method (Sanger *et al.*, 1977) using ^{35}S-labelled d(thio)ATP and Sequenase (US Biochemicals). Synthetic oligonucleotides were used as primers to fill any gap encountered in the nested deletions. The sequence was entirely determined on both DNA strands.

The 5′ end amplification of cDNA was carried out essentially as described by Frohman (1990). The specific 20 base primer (191) for reverse transcriptase (Mu MLV, BRL) was complementary to the mRNA sequence with its 3′ end at position 238. After tailing with ATP by the terminal-deoxynucleotidyl transferase (BRL), the cDNA was amplified using the specific primer 191 and a 35 base oligonucleotide with the sequence 5′-CCCGGGGAATTCCTGCAG $(T)_{17}$-3′, containing a *Pst*I restriction site. After digestion with *Pst*I, the amplified sequences were subcloned into BlueScript. Five different subclones containing inserts of slightly different sizes (~250 bp) and in the two different orientations were independently sequenced.

Northern blots and in situ hybridization

Total RNA was prepared according to Chirgwin *et al.* (1979) from embryos, larvae and adults. Poly(A)$^+$ mRNA was isolated using oligo(dT)-cellulose (BRL). RNAs (5 μg per lane) were separated in agarose gel under denaturing conditions (Sambrook *et al.*, 1989), blotted onto a nitrocellulose filter (Hybond C extra, Amersham) and probed with cDNA radiolabelled with the random priming procedure (Multiprime system, Amersham).

In situ hybridization to polytene chromosomes was performed according to Pardue (1985).

Preparation of fusion proteins and antisera

The λGD51 phage was used to induce lysogens according to Huynh *et al.* (1985). Two major fusion proteins, with molecular weights of 210 and 150 kDa, were observed following induction by IPTG. The two proteins were used as immunogens to elicit the production of specific antisera in mice as described in Harlow and Lane (1988).

For the production of the pG1 and pG2 fusion proteins, the 3.65 kbp *Eco*RI fragment was subcloned in the pGEX-2T vector (Pharmacia). Fusion proteins were produced according to the manufacturer.

Determination of enzymatic activities

Aminoacyl-tRNA synthetase activities produced in recombinant bacteria were assayed by the aminoacylation of tRNA, as previously described (Mirande *et al.*, 1985). Unfractionated brewer's yeast tRNA (Boehringer) was a poor acceptor tRNA for prolyl- and glutamyl-tRNA synthetases from *Drosophila*. Prolyl-tRNA synthetase was assayed with beef liver tRNA and glutamyl- and glutaminyl-tRNA synthetases with beef liver and *Drosophila* embryo tRNA. Crude tRNA (10 mg) from *Drosophila* embryos (60 g) was prepared essentially as described by White and Tener (1973). [^{14}C]amino acids were from the Commissariat à l'Energie Atomique and Amersham. Protein concentration was determined according to Gornall *et al.* (1949).

Acknowledgements

We thank Dr J.P.Waller for his encouragements and support in this work, and Professor E.Kubli for fruitful advice in the preparation of tRNA from *Drosophila*. This work was supported by grants from the Association pour la Recherche sur le Cancer (Nos 6450 and 6486) and Centre National de la Recherche Scientifique.

References

Bachmann,B.J. (1990) *Microbiol. Rev.*, **54**, 130–197.
Bec,G., Kerjan,P., Zha,X.D. and Waller,J.P. (1989) *J. Biol. Chem.*, **264**, 21131–21137.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
Breton,R., Sanfaçon,H., Papayannopoulos,I., Biemann,K. and Lapointe,J. (1986) *J. Biol. Chem.*, **261**, 10610–10617.
Breton,R., Watson,D., Yaguchi,M. and Lapointe,J. (1990) *J. Biol. Chem.*, **265**, 18248–18255.
Brick,P., Bhat,T.N. and Blow,D.M. (1989) *J. Mol. Biol.*, **208**, 83–98.
Brunie,S., Zelwer,C. and Risler, J.L. (1990) *J. Mol. Biol.*, **216**, 411–424.
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.
Chou,P.Y. and Fasman,G.D. (1978) *Adv. Enzymol.*, **47**, 45–148.
Cirakoglu,B. and Waller,J.P. (1985) *Eur. J. Biochem.*, **151**, 101–110.
Coggins,J.R. and Hardie,D.G. (1986) In Hardie,D.G. and Coggins,J.R. (eds), *Multidomain Proteins: Structure and Evolution*. Elsevier, Amsterdam, pp.1–12.
Cusack,S., Berthet-Colominas,C., Härtlein,M., Nassar,N. and Leberman,R. (1990) *Nature*, **347**, 249–255.
Dessen,P., Fondrat,C., Valencien,C. and Mugnier,C. (1990) *CABIOS*, **6**, 355–356.
Dunnill,P. (1968) *Biophys. J.*, **8**, 865–875.
Eriani,G., Delarue,M., Poch,O., Gangloff,J. and Moras,D. (1990a) *Nature*, **347**, 203–206.
Eriani,G., Dirheimer,G. and Gangloff,J. (1990b) *Nucleic Acids Res.*, **18**, 7109–7118.
Eriani,G., Dirheimer,G. and Gangloff,J. (1991) *Nucleic Acids Res.*, **19**, 265–269.
Fett,R. and Knippers,R. (1991) *J. Biol. Chem.*, **266**, 1448–1455.
Frohman,M.A. (1990) In Innis,M.A., Gelfand,D.H., Sninsky,J.J. and White,T.J. (eds), *PCR Protocols: A Guide to Methods and Applications*. Academic Press, San Diego, pp. 28–38.
Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97–120.
Gornall,A.G., Bardawill,C.J. and David,M.M. (1949) *J. Biol. Chem.*, **177**, 751–766.
Harlow,E. and Lane,D. (1988) *Antibodies: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Henikoff,S. (1987) *Methods Enzymol.*, **155**, 156–165.
Hoben,P., Royal,N., Cheung,A., Yamao,F., Biemann,K. and Söll,D. (1982) *J. Biol. Chem.*, **257**, 11644–11650.
Hou,Y.M., Shiba,K., Mottes,C. and Schimmel,P. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 976–980.
Hountondji,C., Dessen,P. and Blanquet,S. (1986) *Biochimie*, **68**, 1071–1078.
Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), *DNA Cloning: A Practical Approach*. IRL Press, Oxford, Vol. I, pp. 49–78.
Jacobo-Molina,A., Peterson,R. and Yang,D.C.H. (1989) *J. Biol. Chem.*, **264**, 16608–16612.
Johnson,D.L., Dang,C.V. and Yang,D.C.H. (1980) *J. Biol. Chem.*, **255**, 4362–4366.
Katinka,M., Cossart,P., Sibilli,L., Saint-Girons,I., Chalvignac,M.A., LeBras,G., Cohen,G.N. and Yaniv,M. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5730–5733.
Lazard,M., Mirande,M. and Waller,J.P. (1985) *Biochemistry*, **24**, 5099–5106.
Lee,C.C., Craigen,W.J., Muzny,D.M., Harlow,E. and Caskey,C.T. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 3508–3512.
Ludmerer,S.W. and Schimmel,P. (1987a) *J. Biol. Chem.*, **262**, 10801–10806.
Ludmerer,S.W. and Schimmel,P. (1987b) *J. Biol. Chem.*, **262**, 10807–10813.
Mirande,M. (1991) *Prog. Nucleic Acid Res. Mol. Biol.*, **40**, 95–142.
Mirande,M. and Waller,J.P. (1989) *J. Biol. Chem.*, **264**, 842–847.
Mirande,M., Kellermann,O. and Waller,J.P. (1982a) *J. Biol. Chem.*, **257**, 11049–11055.
Mirande,M., Cirakoglu,B. and Waller,J.P. (1982b) *J. Biol. Chem.*, **257**, 11056–11063.
Mirande,M., Gache,Y., LeCorre,D. and Waller,J.P. (1982c) *EMBO J.*, **1**, 733–736.
Mirande,M., LeCorre,D. and Waller,J.P. (1985) *Eur. J. Biochem.*, **147**, 281–289.
Motorin,Y.A., Wolfson,A.D., Orlovsky,A.F. and Gladilin,K.L. (1988) *FEBS Lett.*, **238**, 262–264.
Nolan,J.M., Lee,M.P., Wyckoff,E. and Hsieh,T.S. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3664–3668.
Pardue,M.L. (1985) In Hames,B.D. and Higgins,S.L. (eds), *Nucleic Acid Hybridization: A Practical Approach*. IRL Press, Oxford, pp. 179–202.
Putney,S.D., Royal,N.J., Neuman de Vegvar,H., Herlihy,W.C., Biemann,K., and Schimmel,P. (1981) *Science*, **213**, 1497–1501.
Rould,M.A., Perona,J.J., Söll,D. and Steitz,T. (1989) *Science*, **246**, 1135–1142.
Ruff,M., Krishnaswamy,S., Boeglin,M., Poterszman,A., Mitschler,A., Podjarny,A., Rees,B., Thierry,J.C. and Moras,D. (1991) *Science*, **252**, 1682–1689.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schein,C.H. and Noteborn,M.H.M. (1988) *Bio/Technology*, **6**, 291–294.
Schimmel,P. (1987) *Annu. Rev. Biochem.*, **56**, 125–158.
Schön,A., Kannangara,C.G., Gough,S. and Söll,D. (1988) *Nature*, **331**, 187–190.
Schray,B., Thömmes,P. and Knippers,R. (1990) *Biochim. Biophys. Acta*, **1087**, 226–234.
Sellami,M., Fasiolo,F., Dirheimer,G., Ebel,J.P. and Gangloff,J. (1986) *Nucleic Acids Res.*, **14**, 1657–1666.
Shafer,S.J., Olexa,S. and Hillman,R. (1976) *Insect Biochem.*, **6**, 405–411.
Short,J.M., Fernandez,J.M., Sorge,J.A. and Huse,W.D. (1988) *Nucleic Acids Res.*, **16**, 7583–7600.
Sihag,R.K. and Deutscher,M.P. (1983) *J. Biol. Chem.*, **258**, 11846–11850.
Thömmes,P., Fett,R., Schray,B., Kunze,N. and Knippers,R. (1988) *Nucleic Acid Res.*, **16**, 5391–5406.
Tsui,F.W.L. and Siminovitch,L. (1987) *Nucleic Acid Res.*, **15**, 3349–3367.
White,B.N. and Tener,G.M. (1973) *Can. J. Biochem.*, **51**, 896–902.

Received on August 16, 1991

Note added in proof

The sequence data are available from EMBL sequence data bank under the accession number M74104