PHENYLALANYL-tRNA SYNTHETASE FROM BAKER'S YEAST

Repeated sequences in the two subunits

Sylviane ROBBE-SAUL, Franco FASIOLO and Yves BOULANGER
Laboratoire de Biochimie, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue Descartes, 67000 Strasbourg, France

Received 7 October 1977

1. Introduction

Aminoacyl-tRNA synthetases can be divided into five classes on the basis of protomer size and binding capacity:

(i) The 'short monomers' of about 75 000 daltons containing one binding site for amino acid and tRNA, e.g., arginyl-tRNA synthetase from yeast [1].
(ii) The 'long monomers' of about 110 000 daltons containing one (maybe two) binding site for amino acid and tRNA, e.g., valyl-tRNA synthetase from *Bacillus stearothermophilus* [2,3].
(iii) Those containing two identical protomers, each of about 40 000 daltons (with one binding site per protomer) e.g., tyrosyl-tRNA synthetase from *B. stearothermophilus* [4,5].
(iv) Those containing two identical protomers, each of about 85 000 daltons, e.g., methionyl-tRNA synthetase from *E. coli* [6].
(v) Finally those containing dissimilar subunits, e.g., phenylalanyl-tRNA synthetase from yeast which has an $\alpha_2\beta_2$ oligomeric structure (α, 73 000 daltons and β, 62 000 daltons) [7].

It has recently been shown that those enzymes with polypeptide chains of 85 000 daltons or greater than 85 000 daltons (classes 2 and 4) contain a significant proportion of repeated sequence [2,8–11], which probably arises from gene duplication and fusion rather than post-transcriptional or post-translational fusion [10].

On the other hand aminoacyl-tRNA synthetases with subunits smaller than 50 000 daltons (class 3) do not contain such long repeats [2].

One could therefore wonder whether or not aminoacyl-tRNA synthetases with protomeric size ranging between 50 000 and 80 000 daltons contained long repeated units in their polypeptide chain (class 1 or 5).

In this paper we report some of our latest results concerning yeast phenylalanyl-tRNA synthetase (PheRS EC 6.1.1.20): they clearly demonstrate the existence of repeating sequences in both α and β subunits.

2. Materials and methods

The yeast phenylalanyl-tRNA synthetase was purified as described [12].

Urea was from Riedel de Haën and was recrystallised from ethanol. Urea solutions were deionised just before use by passage through a column of Dowex 2 X 8 (OH^- form). Maleic acid was from Fluka and was recrystallised from acetic acid. All other chemicals were from Merck, analytical grade, and so was trypsin (treated with L-1-tosylamido-2-phenylethyl-chloromethyl-ketone, 2 U/mg, ref. 24581). Sephadex G-50 was obtained from Pharmacia (Sweden) and DEAE-cellulose DE52 from Whatman. Thin-layer cellulose plates were from Macherey-Nagel (Polygram C400, 20 X 20 cm) and polyamide sheets from Schleicher and Schüll (DC-Fertigfolien F1700, Micropolyamid 15 X 15 cm).

2.1. *Preparation of radioactive maleic acid*

Radioactive maleic acid was prepared as follows: 250 μCi [2,3–^{14}C]maleic anhydride (Amersham, spec. act. 50 mCi/mmol, ref. CFA 490) was incubated overnight at room temperature with 50 μl of 1 N NaOH and then diluted with 250 μl 1 M cold maleic acid adjusted to pH 7.0 with Tris base.

2.2. *S-succinylation of the enzyme and separation of the α and β subunits*

Enzyme samples (40–50 mg) were alkylated with the above solution of radioactive maleic acid and subjected to chromatography on DEAE-cellulose as described [13] with the following modification: to achieve complete labelling a 50-fold molar excess of the reagent over the thiol-groups was used and the incubation was carried out for 15 h at 37°C.

2.3. *Amino acid analysis*

Samples were thoroughly dialysed against dilute acetic acid (0.5% v/v), freeze-dried and hydrolysed under nitrogen in 6 N HCl (containing 0.5% thiodiglycol, v/v, and a crystal of phenol) at 110°C for 24 h. Analyses were performed with a Beckman Unichrom Analyser. Cysteine and methionine were determined as cysteic acid and methioninesulphone respectively after hydrolysis of the performic-oxidised enzyme according to [14].

Tryptophan was estimated according to the method [15].

2.4. *End-group analysis*

The N-terminal residues were determined by means of the technique [16] modified as follows:

The protein solution containing 5–10 nmol was dialysed against dilute acetic acid, freeze-dried and redissolved in 1 ml 0.5 M $NaHCO_3$ in 8 M urea. To this was added 10 mg DNS-Cl in 0.5 ml acetone and the mixture was incubated for 3 h at 37°C. Another 10 mg DNS-Cl was added and the incubation was continued for another 3 h. The labelled protein was precipitated with 2 ml 10% trichloroacetic acid, spun down, washed twice with 4 ml of the same solution, twice with 4 ml acetone and dried. Acid hydrolysis of the dansylated protein was carried out as described for amino acid analyses. The dry hydrolysate was extracted twice with 100 μl ethylacetate saturated with water. Both the extract and the residue were dissolved in 50% pyridine (v/v) and chromatographed on polyamide thin-layer according to [17].

2.5. *Tryptic digestion*

Protein solutions (10–20 mg in 1–2 ml) were dialysed against 1% NH_4HCO_3 (w/v). Trypsin was added (1:100, w/v) and the mixture was incubated at 37°C for 2 h. Another sample of trypsin was added and the incubation was continued for another 2 h. The slight precipitate which formed was removed by centrifugation and the clear supernatant absorbed onto a column of Sephadex G-50.

2.6. *Gel-filtration of the tryptic peptides*

The column of Sephadex G-50 fine (1.6 × 104 cm) was equilibrated and eluted with 1% NH_4HCO_3 at a flow rate of 18 ml/h. Fractions of 4 ml were collected. Peptides were monitored by absorbance measurements at 235 nm and 280 nm, radioactive peptides by scintillation counting (in Bray's solution).

2.7. *Fingerprinting techniques*

Fractions were subjected to thin-layer electrophoresis and chromatography on cellulose plates. Electrophoresis was run in the first dimension, in a Camag apparatus, at pH 2 (8% acetic acid, 2% formic acid, v/v) and 6.5 (10% pyridine, 0.5% acetic acid, v/v) for 1.5 h at 400 V. The plates were dried in a fume cupboard and run in the second dimension in BAWP (butan-1-ol/acetic acid/water/pyridine, 15 : 3 : 2 v/v/v/v). With these combined techniques about 100 μg peptides in 10 μl could be fingerprinted and analysed. High voltage paper electrophoresis (Gilson Model D apparatus) and paper chromatography were also used (with the same buffers and solvent, Whatman 3 MM paper).

Peptides were stained with 0.3% ninhydrin in ethanol (containing 3% collidine and 10% acetic acid). Peptides containing tryptophan were detected with the Ehrlich reagent (*p*-dimethylaminobenzaldehyde 1 g, 12 N HCl 10 ml, acetone 90 ml). Radioactive peptides were detected by autoradiography of the map.

3. Results and discussion

3.1. *Separation of the subunits*

We have shown [4] that the α and β subunits

could be separated by ion-exchange chromatography of the dissociated pCMB (*para*-chloromercuribenzoate) labelled enzyme. This method turned out to be not suitable for sequence studies because of the lability of the Hg–S bond. Indeed a chemically stable label was highly desirable and among the various reagents tested, maleic acid proved to be the most satisfactory in terms of specificity and efficiency of labelling with respect to the thiol groups: side reactions with other groups of the enzyme occurred at a very low if not negligible level.

3.2. *Amino acid compositions of the two subunits*

The amino acid compositions of the α and β subunits are listed in table 1 and compared to that of native PheRS. Table 1 also contains the composition of the β′ subunit of the trypsin-modified $\alpha_2\beta'_2$ enzyme [18].

The composition of native PheRS calculated from those of α and β yields figures very close to the experimental ones (the minor differences observed are within the limits of experimental error).

The most striking difference between α and β is the very low cysteine content of the latter (2 residues) as compared to that of the former (10 residues). These figures further explain and confirm the results of the labelling experiments with pCMB, DTNB (5,5′-dithiobis-(2 nitrobenzoate)) [14] or maleic acid (above results) and consequently the separation of labelled α and β on the DEAE-cellulose column: indeed labelling of PheRS with any of these reagents introduces 10 new negative charges on each α-subunit versus only 2 on each β-subunit.

It must be noticed that β′ contains the two cysteines of β.

3.3. *End group analysis*

The N-terminal residues, determined as described in section 2.4, were proline for the native enzyme and the α-subunit and isoleucine for β′ whereas β gave no detectable DNS-amino acid, thus suggesting a blocked N-terminus for this subunit. These results also

Table 1
Amino acid compositions of native phenylalanyl-tRNA synthetase, α-, β- and β′-subunits

Amino acid	α	β	2 (α + β)	Native enzyme	β′	β–β′
Cys	10	2	24	25	2	0
Asp	73	62	270	284	40	22
Thr	30	29	118	128	22	7
Ser	39	29	136	136	23	6
Glu	77	70	294	292	40	30
Pro	30	22	104	100	17	5
Gly	34	30	128	134	23	7
Ala	41	28	138	142	18	10
Val	34	25	118	120	19	6
Met	10	12	44	45	9	3
Ile	36	27	126	126	14	13
Leu	60	62	244	236	33	29
Tyr	20	16	72	68	13	3
Phe	30	28	116	114	20	8
Trp	5	7	24	24	5–6	1–2
His	14	13	54	56	10	3
Lys	47	40	174	176	25	15
Arg	27	20	94	90	15	5
Total	617	522	2278	2296	348	174 (175)
Mol. wt	70 000	60 000	260 000	260 000	40 000	20 000

suggest that the proteolytic fragmentation of β which gives rise to β′ takes place mainly in the N-terminal part of the polypeptide chain.

3.4. *Fractionation of the tryptic digests of α and β subunits*

Under the conditions used β gave very little insoluble material on tryptic digestion and α virtually none. The clear soluble part of each digest was first fractionated by gel-filtration chromatography on a column of Sephadex G-50. For both α and β four fractions, S1, S2, S3, S4, were separated on the basis of optical density and liquid scintillation counting (fig. 1a, b).

Each pooled fraction was fingerprinted by a combination of high-voltage paper electrophoresis (at pH 6.5 and pH 2.0), paper chromatography, thin-layer electrophoresis and chromatography.

These combined fingerprinting techniques gave us good estimates of the number of total tryptic peptides and of some particular ones, such as those containing cysteine and tryptophan. They also enabled us to purify some of these peptides for sequence studies.

Table 2 summarises the results: clearly enough α and β subunits yield far fewer peptides than expected on the basis of their (Lys + Arg) contents. These figures gave the first indication that these two subunits might contain repeating sequences but more convincing evidence came from the estimates of the molar yields of the cysteine peptides. It was assumed that yields greater than 1 mol peptide/mol protomer would prove the existence of at least two copies of the corresponding sequence in the intact protomer.

These yields were calculated from amino acid analyses and radioactivity measurements which enabled corrections for losses to be made at each step in purification.

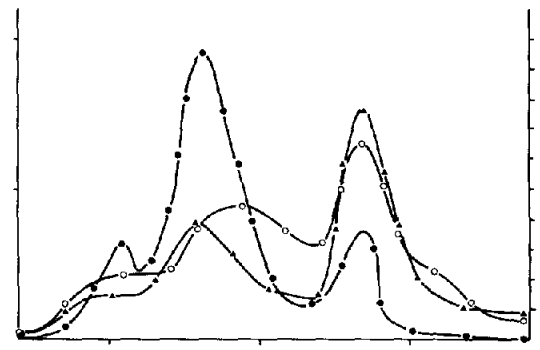

Fig.1a

Fig.1b

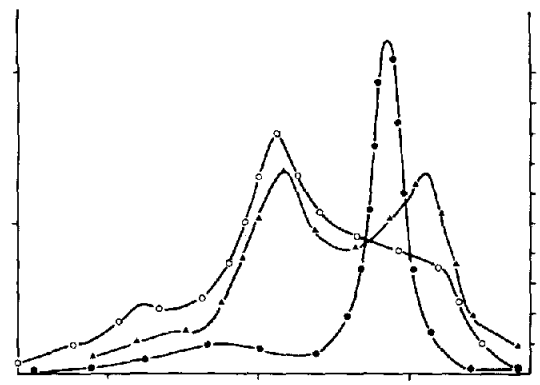

Fig.1. Fractionation of the soluble tryptic peptides of *S*-succinylated α (2a) and β (2b) on a column of Sephadex G-50. (○) Absorbance at 280 nm. (▲) Absorbance at 235 nm. (●) Radioactivity measurements. Fractions were pooled according to the following scheme: (2a): α-S1 (fractions 28–32), α-S2 (fractions 33–43), α-S3 (fractions 44–48) and α-S4 (fractions 49–54). (2b): β-S1 (fractions 27–34), β-S2 (fractions 35–44), β-S3 (fractions 45–52) and β-S4 (fractions 53–57).

Table 2
Estimation of the number of tryptic peptides

Subunit	Total peptides		Tryptophane peptides		Cysteine peptides	
	Expected	Observed	Expected	Observed	Expected	Observed
β	61	43–47	7	4–5	2	1
α	75	37–42	5	3	10	6

Table 3
Amino acid composition and molar yields of the cysteine containing peptides of α and β subunits

Peptide	Composition	Molar yield (mol/mol protomer)
β-T-S3-M	Asp Glu_2 Pro Cys^a Tyr Trp^b Lys_2	1.4
α-T-S3-M	Asp_2 Cys^a Ala_2 Leu_2 Arg	1.8
α-T-S2-M1	Asp_2 Ser_3 Glu Pro_2 Cys^a Gly Val Ile Leu_2 His Lys Arg	2.0
α-T-S2-M2a	Asp_2 Ser_2 Glu_2 Cys^a Gly_2 Ala Ile Leu_2 Lys	0.8
α-T-S2-M2b	Asp_2 Thr Ser_2 Glu_4 Pro Cys^a Gly_2 Ala Val_2 Ile Tyr Phe His Arg	0.8
α-T-S2-M3	Asp_2 Thr_2 Ser_3 Glu_2 Pro Cys^a Gly_2 Ala_2 Val_2 Ile Leu_3 Phe Trp^b His Lys	1.5
α-T-S1-H-M	Asp_2 Ser Glu_2 Cys^a Gly_2 Ala Leu Phe	2.0

[a] Cys was determined as *S*-succinyl-cysteine, eluted just before aspartic acid in the analytical system used

[b] Presence of tryptophan deduced from a positive staining of the peptide with the Ehrlich reagent

Abbreviations: T, soluble tryptic peptide; S, Sephadex G-50; H, thermolytic peptide; M, labelled peptide (S-succinylated with [^{14}C]maleic acid)

The results listed in table 3 show that most of the cysteine peptides are repeated in both α and β subunits: this is the case for the unique cysteine peptide of β whilst only two of the six cysteine peptides of α appear to come from a non-repeated portion of the polypeptide chain. Although their complete amino acid sequences are not available yet, the discriminating fractionation procedures used make it most unlikely that all these peptides are mixtures. In each case the molar proportions of the amino acids present were good. The probability that all these peptides could be duplicated by random chance is of course negligible.

These tryptic peptides listed in table 3 account for the two cysteines of β and eight of the ten cysteines of α. Another repeated cysteine sequence was identified in fraction α-T-S1 which corresponds to fairly large tryptic peptides, most probably partial digestion products. As these peptides do not migrate well on cellulose plates and paper an aliquot of this α-T-S1 fraction was digested with thermolysin under conditions identical with those described for trypsin and fingerprinted on cellulose plates: one radioactive peptide was thus isolated named α-T-S1-H-M; its amino acid composition and molar yield are given in table 3.

Peptides α-T-S3-M and β-T-S3-M have been sequenced with the help of Wittmann-Liebold et al. using a solid-phase sequencing technique derived from [19–21]. The two following sequences were obtained:

α-T-S3-M Leu–Ala–Ala–Asn–Leu–Asn–Cys–Arg

β-T-S3-M Tyr–Asn–Trp–Lys–Pro–Glu–Glu–Cys–Lys

It must be emphasized that this unique cysteine peptide present twice in β was not found in α: indeed none of the cysteine peptides of α contains both tyrosine and tryptophan. This result rules out completely the hypothesis [22] according to which the $\alpha_2\beta_2$ enzyme could be a slightly proteolysed form of a native α_4 tetramer. As already pointed out if this assumption were proven right 80% of the thiols would be located in the 10 000 dalton piece supposedly released from α to yield the β chain. Furthermore the unique cysteine sequence left in β should then be found in α which is obviously not the case.

The repeating units found so far in aminoacyl-tRNA synthetases are fairly large, ranging from about 200 residues for methionyl-tRNA synthetase (protomeric size of 85 000 daltons) to 300–400 residues for the long monomeric enzymes of 100 000–120 000 daltons [2,8–11]. Therefore it would be interesting to determine the size of the repeats in both α and β-subunits of phenylalanyl-tRNA synthetase, whose molecular weights are significantly lower than the above figures.

Obviously more data must be accumulated to give a conclusive answer which will eventually come from the complete sequences of both α and β-subunits.

In conclusion the results presented in this paper make more consistent the possibility that all 20 of the aminoacyl-tRNA synthetases have evolved from a common ancestral precursor through gene duplication and fusion: the small subunits of 40 000 daltons (synthetases from class 2) would have elongated in this manner to yield the larger protomers of the enzymes from the other classes.

Acknowledgements

This work was partly supported by a grant from the Délégation Générale à la Recherche Scientifique et Technique. We thank the Management of the Levurerie Fala, Strasbourg, for their generous gift of Baker's yeast. The excellent technical assistance of Mrs C. Rether and G. Nussbaum is gratefully acknowledged. We are indebted to Dr Wittmann-Liebold and co-workers for their help in the use of the solid-phase sequencer. Finally we thank Professor J. P. Ebel for his helpful advice and stimulating discussions.

References

[1] Gangloff, J., Schutz, A. and Dirheimer, G. (1976) Eur. J. Biochem. 65, 177–182.
[2] Koch, G. L. E., Boulanger, Y. and Hartley, B. S. (1974) Nature 249, 316–320.
[3] Fersht, A. R. (1975) Biochemistry 14, 5–12.
[4] Koch, G. L. E. (1974) Biochemistry 13, 2307–2312.
[5] Bosshard, H. R., Koch, G. L. E. and Hartley, B. S. (1975) Eur. J. Biochem. 53, 493–498.
[6] Koch, G. L. E. and Bruton, C. J. (1974) FEBS Lett 40, 180–182.
[7] Fasiolo, F., Befort, N., Boulanger, Y. and Ebel, J. P. (1970) Biochim. Biophys. Acta 217, 305–318.
[8] Kula, M. R. (1973) FEBS Lett. 35, 299–302.
[9] Waterson, R. M. and Konigsberg, W. H. (1974) Proc. Natl. Acad. Sci. USA 71, 376–380.
[10] Bruton, C. J., Jakes, R. and Koch, G. L. E. (1974) FEBS Lett. 45, 26–28.
[11] Bruton, C. J. (1975) Biochem. J. 147, 191–192.
[12] Fasiolo, F. and Ebel, J. P. (1974) Eur. J. Biochem. 49, 257–263.
[13] Fasiolo, F. and Boulanger, Y. (1975) FEBS Lett. 60, 156–160.
[14] Moore, S. (1963) J. Biol. Chem. 238, 235–237.
[15] Liu, T. Y. and Chang, Y. H. (1971) J. Biol. Chem. 246, 2842–2848.
[16] Gray, W. R. and Hartley, B. S. (1963) Biochem. J. 89, 379–380.
[17] Woods, K. P. and Wang, K. T. (1967) Biochim. Biophys. Acta 133, 369–370.
[18] Fasiolo, F., Boulanger, Y. and Ebel, J. P. (1974) Eur. J. Biochem. 53, 487–492.
[19] Laursen, R. A. (1971) Eur. J. Biochem. 20, 89–102.
[20] Previero, A., Derancourt, J., Coletti-Previero, M. A. and Laursen, R. A. (1975) FEBS Lett. 33, 135–138.
[21] Wittmann-Liebold, B. and Lehmann, A. (1975) in: Solid-Phase Methods in Protein Sequence Analysis (Laursen, R. A. ed.) pp. 81–90, Pierce Chemical Company, Rockford, USA.
[22] Hossain, A (1974) Fed. Proc. Fed. Am. Soc. Exp. Biol. 33, 1421.