other than cholinergic nerve terminals.

Since mitochondria are involved in pyruvate-derived acetylcholine formation and since there is little detectable citrate cleavage enzyme in the cholinergic synaptoplasm, it follows that, in the absence of other known precursors of cerebral acetylcholine (*18–22*), acetyl CoA itself must be able, in presence of Ca^{2+}, to leave the mitochondria. The amount involved (*4, 5*) may be only a small percentage of that formed in the mitochondria during respiration. Oxaloacetate, a potent inhibitor of pyruvate-derived acetylcholine formation, both in the absence or presence of Triton X-100 (Table 1), must combine with acetyl CoA, since citrate synthase (E.C. 4.1.3.7) is present both in the mitochondria of the cholinergic synaptosome and in the Triton-treated preparation. Succinate (2 m*M*) suppresses the formation of pyruvate-derived acetylcholine (Table 1), part of it being converted to oxaloacetate in the cholinergic mitochondria. Succinate has little or no effect in the Triton-treated preparation (Table 1) when the mitochondrial membranes undergo rupture and succinate dehydrogenase activity is suppressed.

It was shown many years ago (*23*) that phospholipase A from venoms stimulates glucose-derived acetylcholine formation. We find that mammalian phospholipase A_2 from the pancreas stimulates the formation of labeled acetylcholine from labeled pyruvate in a preparation from rat brain synaptosomes in the presence of CoA and NAD (Table 3). This stimulation is enhanced by Ca^{2+} (Table 3), which stimulates the uptake of NAD into rat liver mitochondria (*24*). Since phospholipase A_2 is present in brain (*25*) and other mitochondria (*26*) and is stimulated by Ca^{2+} (*25–27*), it seems likely that Ca^{2+} alters the permeability of the mitochondrial membrane to CoA and acetyl CoA and that this effect may be mediated by the activity of endogenous mitochondrial phospholipase A_2. These results may be correlated with findings that indicate a role for intramitochondrial phospholipase A_2 in the processes that establish the permeability properties of the inner mitochondrial membrane (*28*) and with our finding (*29*) that phospholipase A_2 causes an efflux of CoA and acetyl CoA from the synaptosomal fraction into the incubation medium.

A. M. BENJAMIN
J. H. QUASTEL

Division of Neurological Sciences, Department of Psychiatry, University of British Columbia, Vancouver, Canada V6T 1W5

References and Notes

1. J. M. Lowenstein, in *The Metabolic Roles of Citrate*, T. W. Goodwin, Ed. (Academic Press, New York, 1968), p. 61.
2. J. Sollenberg and B. Sorbo, *J. Neurochem.* **17**, 201 (1970); G. H. Sterling and J. J. O'Neill, *ibid.* **31**, 525 (1978).
3. P. Lefresne, J. C. Beaujouan, J. Glowinski, *Nature (London)* **274**, 497 (1978).
4. S. Tucek, *Biochem. J.* **104**, 749 (1967).
5. ———, *Acetylcholine Synthesis in Neurons* (Halsted, New York, 1978).
6. R. S. Jope, *Brain Res. Rev.* **1**, 313 (1979).
7. A. M. Benjamin, Z. H. Verjee, J. H. Quastel, *J. Neurochem.* **35**, 67 (1980).
8. D. S. Grewaal and J. H. Quastel, *Biochem. J.* **132**, 1 (1973).
9. I. Pull, H. McIlwain, R. L. Ramsay, *ibid.* **116**, 181 (1970).
10. P. Lefresne, P. Guyenet, J. C. Beaujouan, J. Glowinski, *J. Neurochem.* **25**, 415 (1975).
11. That the synaptosomal mitochondria are involved in the formation of acetylcholine is also evident from the fact that labeled β-hydroxybutyrate forms labeled acetylcholine in a rat brain synaptosomal preparation; a concentration as low as 0.1 m*M* will form 7 nmole of acetlycholine per 100 mg of protein per hour. This reaction involves mitochondrial β-hydroxybutyrate dehydrogenase in the cholinergic synaptosomes. Any formation of acetoacetate by the mitochondria external to these synaptosomes would be so highly diluted (100-fold) by the incubation medium as to render it ineffective as a precursor of acetylcholine.
12. G. E. Gibson and J. P. Blass, *J. Neurochem.* **27**, 37 (1976); *J. Biol. Chem.* **251**, 4127 (1976); R. S. Jope and D. J. Jenden, *J. Neurochem.* **35**, 318 (1980).
13. A. Szutowicz and W. Lysiak, *J. Neurochem.* **35**, 775 (1980).
14. G. E. Gibson and M. Shimada, *Biochem. Pharmacol.* **29**, 167 (1980).
15. Synaptosomes were ruptured by hypo-osmotic shock by exposure to 60 m*M* sucrose for 30 minutes at 10°C. Free mitochondria seem not to be affected under these conditions [L. Salganicoff and E. DeRobertis, *J. Neurochem.* **12**, 287 (1965)].
16. A. Szutowicz, M. Stephen, W. Lysiak, S. Angielski, *ibid.* **25**, 11 (1975).
17. S. C. Cheng and E. A. Brunner, *ibid.* **20**, 1421 (1978).
18. The suggestion that the acetyl group is transferred as acetylcarnitine is made unlikely by the demonstration of the inactivity of carnitine injections in vivo on brain acetylcholine levels (*19*) and by the lack of effect of carnitine, or of acetylcarnitine, on the rate of [^{14}C]acetylcholine formation from [2-^{14}C]pyruvate (*20*) or from [U-^{14}C]glucose (*14*). Moreover, carnitine acetyltransferase seems to be confined to the mitochondria (*13, 21*), a conclusion supported by the fact that an electrical lesion of a cholinergic tract that almost (> 90 percent) destroys choline acetyltransferase (E.C. 2.3.1.6) has a relatively small inhibitory effect (35 percent) on the activity of carnitine acetyltransferase resembling that on the activities of other mitochondrial enzymes, such as fumarase (36 percent) and pyruvate dehydrogenase (36 percent) (*22*).
19. S. Tucek and S. C. Cheng, *J. Neurochem.* **22**, 893 (1974).
20. P. L. Polak, P. C. Molenaar, P. Braggar-Schapp, in *Cholinergic Mechanisms and Psychopharmacology*, D. F. Jenden, Ed. (Plenum, New York, 1978), p. 511; P. Lefresne *et al.*, *Biochimie* **59**, 197 (1977).
21. R. E. McCaman, M. W. R. McCaman, M. L. Stafford, *J. Biol. Chem.* **241**, 930 (1966).
22. S. H. Sterri and F. Fonnum, *J. Neurochem.* **35**, 249 (1980).
23. B. M. Braganca and J. H. Quastel, *Nature (London)* **169**, 695 (1952).
24. M. D. Greenspan and J. L. Purvis, *Biochim. Biophys. Acta* **99**, 167 (1965).
25. M. F. Cooper and G. R. Webster, *J. Neurochem.* **17**, 1543 (1970); H. Woelk and G. Porcellati, *Hoppe-Seylers Z. Physiol. Chem.* **354**, 90 (1973); ———, A. Gaiti, *Neurochem. Res.* **4**, 535 (1979); T. L. Eyrich, D. Barrett, P. A. Rock, *J. Neurochem.* **26**, 1079 (1976).
26. M. Waite, *Biochemistry* **8**, 2536 (1969); ——— and P. Sisson, *ibid.* **10**, 2377 (1971).
27. W. Y. Cheung, *Science* **207**, 19 (1980).
28. M. C. Beatrice, J. W. Palmer, D. R. Pfeiffer, *J. Biol. Chem.* **255**, 8663 (1980).
29. A. M. Benjamin and J. H. Quastel, *Soc. Neurosci. Abstr.* **6**, 514 (1980).
30. We thank the British Columbia Health Care Research Foundation for financial assistance.

25 August 1980; revised 19 February 1981

Primary Structure of a Large Aminoacyl-tRNA Synthetase

Abstract. *The complete primary structure of* Escherichia coli *alanyl-tRNA synthetase, a 95,000-dalton polypeptide, was established by sequencing the gene encoding the enzyme and by sequencing oligopeptides in hydrolyzates of the protein by gas chromatography–mass spectrometry. Contrary to expectation, this long polypeptide contains no lengthy duplications. One 13-residue peptide is homologous with* Escherichia coli *tyrosyl-tRNA synthetase, and clusters of charged amino acids occur in several sections of the structure.*

Aminoacyl-tRNA sythetases (tRNA, transfer RNA) play a crucial role in protein biosynthesis by attaching amino acids to their corresponding tRNA's (*1*). We have sequenced the gene that encodes *Escherichia coli* alanyl-tRNA synthetase (*2*) and, in addition, have identified short randomly located oligopeptides of the protein by gas chromatography–mass spectrometry (GC-MS) (*3, 4*). Reports of repeating sequences in tRNA synthetases of large molecular weight (M_r > 60,000) have led to the hypothesis that these enzymes evolved by gene duplication (*1, 5–9*). Alanyl-tRNA synthetase provides the first complete test of this hypothesis because it is the only large (875 residues) synthetase to be sequenced in its entirety. The sequence of the alanine enzyme also allows for a more complete examination of sequence homologies among this class of enzymes, because primary structure data are available only on two small (M_r < 50,000) aminoacyl-tRNA synthetases (*10–12*). Finally, alanyl-tRNA synthetase binds to a specific DNA sequence as part of a transcriptional control mechanism (*13*), and extensive data are available on the general location of catalytic sites (*2*). Therefore, knowledge of the primary structure is essential for understanding the relation between structure and function of this enzyme.

Figure 1 shows the restriction map of the gene for alanyl-tRNA synthetase. The arrows above and below these strands indicate the length and direction

0036-8075/81/0925-1497$01.00/0

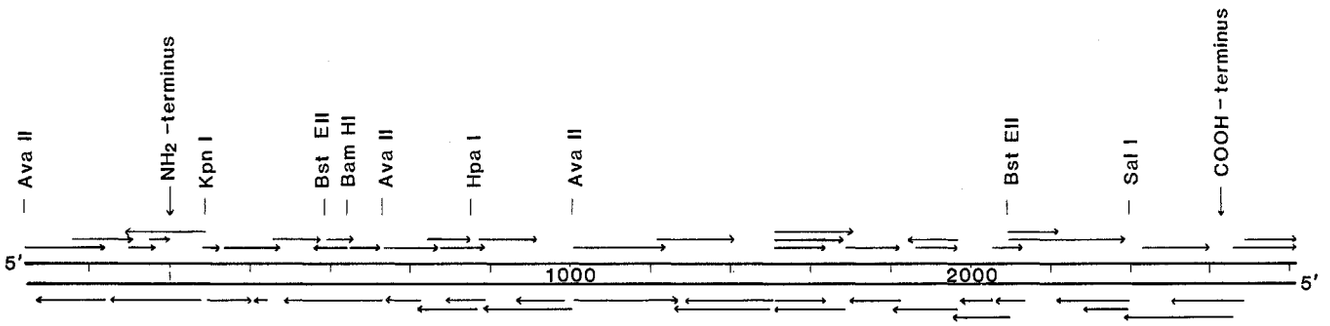

Fig. 1. Restriction map of alanyl-tRNA synthetase showing sequencing strategy. Both strands of DNA are shown, and the top strand is the coding strand. The structural gene begins at position 1 and terminates at position 2625. Vertical tick marks are placed every 200 nucleotides. Each arrow, the beginning of which indicates the position of the radioactive label, shows the strand and direction of sequencing. Sequencing was done by the chemical method (*14*).

of sequencing with the method of Maxam and Gilbert (*14*). The top strand is the coding strand, and position 1 indicates the beginning of the structural gene. The location of the amino terminal coding region was established by matching the DNA sequence with the amino terminal sequence of the intact subunit (*2*). The region encoding the structural gene consists of 2625 nucleotides, and the protein is 875 amino acids long (Fig. 2). Also shown are 142 nucleotides that follow the structural gene. The sequence of the region preceding the structural gene has been reported (*15*).

The lines below the amino acid sequence (Fig. 2) indicate the location of the oligopeptides sequenced by GC-MS. Application of the GC-MS approach to the first 165 amino acids of alanyl-tRNA synthetase has been described (*4*). In applying this method to the sequencing of alanyl-tRNA synthetase, we took advantage of a 40,000-dalton amino terminal fragment of the enzyme (*4*). This fragment, denoted T-1, is generated by limited digestion with trypsin and was separated from the resulting mixture of fragments (termed T-2) by gel filtration chromatography. The GC-MS experiments were carried out separately on T-1 and T-2. Overall, approximately 25 percent of the amino acid sequence has been confirmed by GC-MS sequencing.

All of the peptides found by mass spectrometric analysis are present in the designated coding region of the DNA sequence of Fig. 2. The first in-phase stop codon (TAA) occurs after amino acid 875 (Gln) (*16*). On carboxypeptidase A digestion, the most abundant amino acids released are Glx (Glu or Gln) and Leu. In addition, a string of 13 consecutive amino acid residues shortly preceding the stop codon was confirmed by mass spectrometric analysis. The placement of the carboxyl terminus at this point gives a polypeptide length in close agreement with that measured independently (*2*).

Alanyl-tRNA synthetase is the largest synthetase to be sequenced thus far. The smaller (~ 330 residues) tryptophanyl-tRNA synthetases from *Bacillus stearothermophilus* (*10*) and from *E. coli* (*12*) have been sequenced, in addition to a large portion of the approximately 400-residue subunit of tyrosyl-tRNA synthetase from both of these organisms (*11*). Because the tRNA synthetases each bind adenosine 5′-triphosphate (ATP) and tRNA and catalyze similar reactions, there is reason to expect sequence homologies between synthetases specific for different amino acids from the same organism (*1*). However, as judged by the limited polypeptide sequence data that have been obtained, such homologies are not immediately obvious (*1*).

Because of the lack of sequence information, a complete comparison of the alanine enzyme with most other tRNA synthetases cannot be made. However, a computer search (*17*) for sequence homologies was conducted between the alanyl-tRNA synthetases and the tyrosyl- and tryptophanyl-tRNA synthetases from *E. coli*, as well as the tryptophanyl-tRNA synthetases from *B. stearothermophilus*. (The search identifies identical amino acids occurring in two sequences, regardless of the location of the homologous regions in the two enzymes.) No significant homologies were found with two unrelated *E. coli* proteins (as expected). However, the search revealed a significant homology between the alanyl-tRNA synthetases and the partially sequenced *E. coli* tyrosyl-tRNA synthetase.

The homology (Fig. 3A) is between residues 370 to 382 of alanyl-tRNA synthetase and residues 9 to 25 of tyrosyl-tRNA synthetase. Within a stretch of 13 consecutive residues in alanyl-tRNA synthetase, 11 match those in a 17-residue segment of the tyrosine enzyme. The possible significance of this homology is apparent when the functions of the various regions of the alanyl-tRNA synthetase subunit are considered. A 48,000-dalton fragment, approximately 75 amino acids longer than the T-1 fragment, which extends from the amino terminus to approximately residue 440, has full ATP–inorganic pyrophosphate exchange activity, in that it forms the aminoacyl adenylate just as efficiently, per mole, as the native enzyme subunit (*2*). In addition, unlike the native subunit (which aggregates to form a tetramer), this fragment does not associate. More recent experiments have shown that when the larger fragment is treated with trypsin, the T-1 fragment is generated. Measurement of ATP–inorganic pyrophosphate activity revealed that T-1 is only 20 percent as active as the larger fragment (*18*). This result suggests that the 75 amino acids removed to produce the stable T-1 piece are important for full adenylate formation activity (Fig. 3B).

The section that is homologous to tyrosyl-tRNA synthetase is contained within (or is immediately adjacent to) the 75 amino acids that are cleaved from the larger amino terminal fragment to produce T-1 (Fig. 3B). Because this region

Fig. 2. The nucleotide sequence of the structural gene for alanyl-tRNA synthetase and the translated amino acid sequence. Peptides from protein hydrolyzates sequenced by gas chromatography–mass spectrometry (GC-MS) are underlined, and only GC-MS peptides that occur once in the sequence are included. The undecapeptide at the amino terminus was sequenced by Edman degradation (*2*). The last peptide sequenced from the T-1 fragment (Glu-Gln-Phe) ends at position 366 and, therefore, the carboxyl terminus of this fragment must be just beyond this point (see Fig. 3B). There are significantly more peptides underlined in the T-1 than in the T-2 region because more experiments were done with T-1 in developing the methodology.

```
AGC AAG AGC ACC GCT GAG ATC CGT CAG GCG TTT CTC GAC TTT TTC CAT AGT AAG GGA CAT CAG GTA GTT GCC AGC AGC TCC AGG TAC CCC ATA ACC GAC CCA ACT
Ser Lys Ser Thr Ala Glu Ile Arg Gln Ala Phe Leu Asp Phe Phe His Ser Lys Gly His Gln Val Val Ala Ser Ser Ser Arg Tyr Pro Ile Thr Asp Pro Thr  35

TTG TTG TTT ACC AAC GCC GGG ATG AAC CAG TTC AAG GAT GTG TTC CTT GGG CTC GAC AAG CGT AAT TAT TCC CGC GCT ACC ACT TCC CAA CGC TGC GTG CGT GCG
Leu Leu Phe Thr Asn Ala Gly Met Asn Gln Phe Lys Asp Val Phe Leu Gly Leu Asp Lys Arg Asn Tyr Ser Arg Ala Thr Thr Ser Gln Arg Cys Val Arg Ala  70

GGT GGT AAA CAC AAC GAC CTG GAA AAC GTC GGT TAC ACC GCG CGT CAC CAT ACC TTC TTC GAA ATG CTG GGC AAC TTC AGC TTC GGC GAC TAT TTC AAA CAC GAT
Gly Gly Lys His Asn Asp Leu Glu Asn Val Gly Tyr Thr Ala Arg His His Thr Phe Phe Glu Met Leu Gly Asn Phe Ser Phe Gly Asp Tyr Phe Lys His Asp 105

GCC ATT CAG TTT GCA TGG GAA CTG CTG ACC AGC GAA AAA TGG TTT GCC CTG CCG AAA GAG CGT CTG TGG GTT ACC GTC TAT GAA AGC GAC GAC GAA GCC TAC GAA
Ala Ile Gln Phe Ala Trp Glu Leu Leu Thr Ser Glu Lys Trp Phe Ala Leu Pro Lys Glu Arg Leu Trp Val Thr Val Tyr Glu Ser Asp Asp Glu Ala Tyr Glu 140

ATC TGG GAA AAA GAA GTA GGG ATC CCG CGC GAA CGT ATT ATT CGC ATC AAC GAT AAC AAG GGT GCG CCA TAC GCA TCT GGC AAC TTC TGG CGG ATG GGT GGC ACT
Ile Trp Glu Lys Glu Val Gly Ile Pro Arg Glu Arg Ile Ile Arg Ile Asn Asp Asn Lys Gly Ala Pro Tyr Ala Ser Gly Asn Phe Trp Arg Met Gly Gly Thr 175

GGT CCG TGC GAC CCG TGC ACC GAA ATC TTC TAC GAT CAC GGC GAC CAC ATT TGG GGG GGC CCT CCG GGA AGC CCG GAA GAA GAC GGC GAC CGC TAC ATT GAG ATC
Gly Pro Cys Asp Pro Cys Thr Glu Ile Phe Tyr Asp His Gly Asp His Ile Trp Gly Gly Pro Pro Gly Ser Pro Glu Glu Asp Gly Asp Arg Tyr Ile Glu Ile 210

TGG AAC ATC GTC TTC ATG CAG TTC AAC CGC CAG GCC GAT GGC ACG ATG GAA CCG CTG CCG AAG CCG TCT GTA GAT ACC GGT ATG GGT CTG GAG CGT ATT GCT GCG
Trp Asn Ile Val Phe Met Gln Phe Asn Arg Gln Ala Asp Gly Thr Met Glu Pro Leu Pro Lys Pro Ser Val Asp Thr Gly Met Gly Leu Glu Arg Ile Ala Ala 245

GTG CTG CAA CAC GTT AAC TCT AAC TAT GAC ATC GAC CTG TTC CGC ACG CTG ATC CAG GCG GTA GCG AAA GTC ACT GGC GCA ACC GAT CTG AGC AAT AAA TCG CTG
Val Leu Gln His Val Asn Ser Asn Tyr Asp Ile Asp Leu Phe Arg Thr Leu Ile Gln Ala Val Ala Lys Val Thr Gly Ala Thr Asp Leu Ser Asn Lys Ser Leu 280

CGC GTA ATC GCT GAC CAC ATT CGT TCT TGT GCG TTC CTG ATC GCG GAT GGC GTA ATG CCG TCC AAT GAA AAC CGT GGT TAT GTA CTG CGT CGT ATC ATT CGT CGC
Arg Val Ile Ala Asp His Ile Arg Ser Cys Ala Phe Leu Ile Ala Asp Gly Val Met Pro Ser Asn Glu Asn Arg Gly Tyr Val Leu Arg Arg Ile Ile Arg Arg 315

GCA GTG CGT CAC GGT AAT ATG CTC GGC GCG AAA GAA ACC TTC TTC TAC AAA CTG GTT GGT CCG CTG ATC GAC GTT ATG GGC TCT GCG GGT GAA GAC CTG AAA CGC
Ala Val Arg His Gly Asn Met Leu Gly Ala Lys Glu Thr Phe Phe Tyr Lys Leu Val Gly Pro Leu Ile Asp Val Met Gly Ser Ala Gly Glu Asp Leu Lys Arg 350

CAG CAG GCG CAG GTT GAG CAG GTG CTG AAG ACT GAA GAA GAG CAG TTT GCT CGT ACT CTG GAG CGC GGT CTG GCG TTG CTG GAT GAA GAG CTG GCA AAA CTT TCT
Gln Gln Ala Gln Val Glu Gln Val Leu Lys Thr Glu Glu Glu Gln Phe Ala Arg Thr Leu Glu Arg Gly Leu Ala Leu Leu Asp Glu Glu Leu Ala Lys Leu Ser 385

GGT GAT ACG CTG GAT GGT GAA ACT GCT TTC CGT CTG TAC GAC ACC TAT GGC TTC CCG GTT GAC CTG ACG GCT GAT GTT TGT CGT GAG CGC AAC ATC AAA GTT GAC
Gly Asp Thr Leu Asp Gly Glu Thr Ala Phe Arg Leu Tyr Asp Thr Tyr Gly Phe Pro Val Asp Leu Thr Ala Asp Val Cys Arg Glu Arg Asn Ile Lys Val Asp 420

GAA GCT GGT TTT GAA GCT GCA ATG GAA GAG CAG CGT CGT CGC GCG CGC GAA GCC AGC GGC TTT GGT GCC GAT TAC AAC GCA ATG ATC CGT GTT GAC AGT GCA TCT
Glu Ala Gly Phe Glu Ala Ala Met Glu Glu Gln Arg Arg Arg Ala Arg Glu Ala Ser Gly Phe Gly Ala Asp Tyr Asn Ala Met Ile Arg Val Asp Ser Ala Ser 455

GAA TTT AAA GGC TAT GAC CAT CTG GAA CTG AAC GGC AAA GTG ACT GCG CTG TTT GTT GAT GGT AAA GCG GTT GAT GCC ATC AAT GCA GGC CAG GAA GCT GTG GTC
Glu Phe Lys Gly Tyr Asp His Leu Glu Leu Asn Gly Lys Val Thr Ala Leu Phe Val Asp Gly Lys Ala Val Asp Ala Ile Asn Ala Gly Gln Glu Ala Val Val 490

GTG CTG GAT CAA ACG CCA TTC TAT GCG GAA TCC GGC GGT CAG GTT GGC GAT AAA GGC GAA CTG AAA GGC GCT AAC TTC TCC TTT GCG GTG GAA GAT ACG CAG AAA
Val Leu Asp Gln Thr Pro Phe Tyr Ala Glu Ser Gly Gly Gln Val Gly Asp Lys Gly Glu Leu Lys Gly Ala Asn Phe Ser Phe Ala Val Glu Asp Thr Gln Lys 525

TAC GGC CAG GCG ATT GGT CAC ATC GGT AAA CTT GCT GCG GGT TCT CTG AAA GTG GGC GAC GCG GTG CAG GCT GAT GTT GAT GAG GCT CGT CGC GCC CGT ATT CGT
Tyr Gly Gln Ala Ile Gly His Ile Gly Lys Leu Ala Ala Gly Ser Leu Lys Val Gly Asp Ala Val Gln Ala Asp Val Asp Glu Ala Arg Arg Ala Arg Ile Arg 560

CTG AAT CAC TCC GCA ACG CAC CTG ATG CAC GCT GCG CTG CGC CAG GTT CTG GGT ACT CAT GTA TCG CAT AAA GGT TCA CTG GTT AAC GAC AAG GTG CTG CGC TTC
Leu Asn His Ser Ala Thr His Leu Met His Ala Ala Leu Arg Gln Val Leu Gly Thr His Val Ser His Lys Gly Ser Leu Val Asn Asp Lys Val Leu Arg Phe 595

GAC TTC TCA CAC AAC GAA GCG ATG AAA CCA GAA GAG ATT CGT GCG GTC GAA GAC CTG GTG AAC ACA CTG ATT CGT CGC AAT TTG CCG ATC GAA ACC AAC ATC ATG
Asp Phe Ser His Asn Glu Ala Met Lys Pro Glu Glu Ile Arg Ala Val Glu Asp Leu Val Asn Thr Leu Ile Arg Arg Asn Leu Pro Ile Glu Thr Asn Ile Met 630

GAT CTC GAA GCG GCG AAA GCG AAA GGT GCG ATG GCG CTG TTC GGC GAG AAG TAT GAT GAG CGC GTA CGC GTG CTG AGC ATG GGC GAT TTC TCT ACC GAG TTG TGT
Asp Leu Glu Ala Ala Lys Ala Lys Gly Ala Met Ala Leu Phe Gly Glu Lys Tyr Asp Glu Arg Val Arg Val Leu Ser Met Gly Asp Phe Ser Thr Glu Leu Cys 665

GGC GGT ACT CAC GCC AGC CGC ACT GGT GAT ATT GGT CTG TTC CGC ATC ATC TCT GAA TCG GGT ACT GCT GCA GGC GTT CGT CGT ATC GAA GCG GTA ACC GGA GAA
Gly Gly Thr His Ala Ser Arg Thr Gly Asp Ile Gly Leu Phe Arg Ile Ile Ser Glu Ser Gly Thr Ala Ala Gly Val Arg Arg Ile Glu Ala Val Thr Gly Glu 700

GGT GCT ATC GCC ACC GTT CAT GCA GAC AGC GAT CGC TTA AGC GAA GTC GCG CAT CTG CTG AAA GGC GAT AGC AAT AAT CTG GCT GAT AAA GTG CGC TCA GTA CTG
Gly Ala Ile Ala Thr Val His Ala Asp Ser Asp Arg Leu Ser Glu Val Ala His Leu Leu Lys Gly Asp Ser Asn Asn Leu Ala Asp Lys Val Arg Ser Val Leu 735

GAA CGT ACG CGT CAG CTG GAA AAA GAG TTA CAA CAG CTT AAA GAA CAA GCT GCC GCA CAG GAG AGC GCA AAT CTT TCC AGT AAG GCA ATT GAT GTT AAT GGT GTT
Glu Arg Thr Arg Gln Leu Glu Lys Glu Leu Gln Gln Leu Lys Glu Gln Ala Ala Ala Gln Glu Ser Ala Asn Leu Ser Ser Lys Ala Ile Asp Val Asn Gly Val 770

AAG CTG TTG GTT AGC GAG CTT AGC GGT GTT GAG CCG AAA ATG TTG CGT ACC ATG GTT GAC GAT TTA AAA AAT CAG CTG GGG TCG ACA ATT ATC GTG CTG GCA ACG
Lys Leu Leu Val Ser Glu Leu Ser Gly Val Glu Pro Lys Met Leu Arg Thr Met Val Asp Asp Leu Lys Asn Gln Leu Gly Ser Thr Ile Ile Val Leu Ala Thr 805

GTA GTC GAA GGT AAG GTT TCT CTG ATT GCA GGC GTA TCT AAG GAC GTC ACA GAT CGT GTG AAA GCA GGG GAA CTG ATT GGT ATG GTC GCT CAG CAG GTG GGC GGC
Val Val Glu Gly Lys Val Ser Leu Ile Ala Gly Val Ser Lys Asp Val Thr Asp Arg Val Lys Ala Gly Glu Leu Ile Gly Met Val Ala Gln Gln Val Gly Gly 840

AAG GGT GGT GGA CGT CCT GAC ATG GCG CAA GCC GGT GGT ACG GAT GCT GCG GCC TTA CCT GCA GCG TTA GCC AGT GTG AAA GGC TGG GTC AGC GCG AAA TTG CAA
Lys Gly Gly Gly Arg Pro Asp Met Ala Gln Ala Gly Gly Thr Asp Ala Ala Ala Leu Pro Ala Ala Leu Ala Ser Val Lys Gly Trp Val Ser Ala Lys Leu Gln 875

TAATATAAGCGTCAGGCAACGCCGTGGACTCGCTTCACGGCATTCGCATTAACGCTATCGACAACGATAAAGTCAGGTTGAAGTTGTGTATATCGCGTAAACTTAGGTTTAACAGAATGTAATGCCATGACTGCTTAGATGT
```

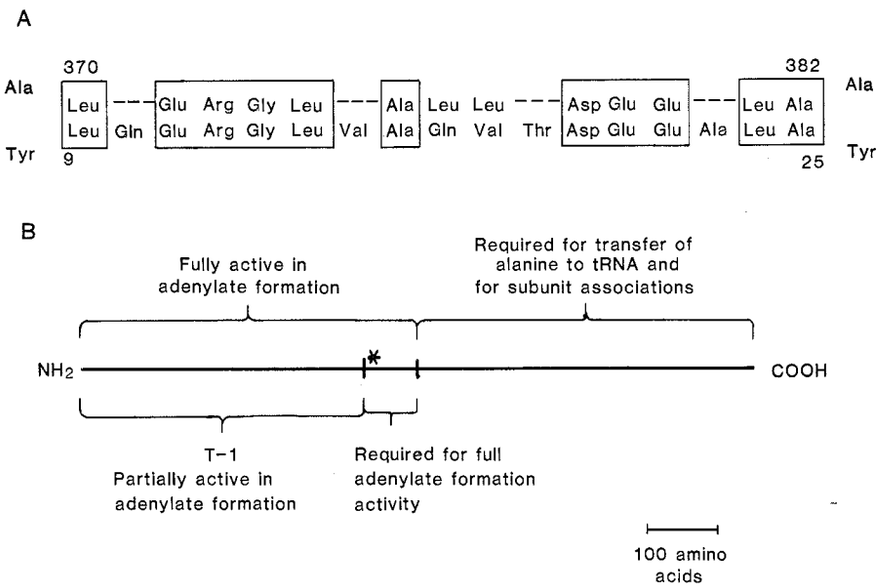

Fig. 3. (A) Homology between *E. coli* alanyl (Ala)- and tyrosyl (Tyr)-tRNA synthetases. The sequence on top is that of Ala-tRNA synthetase, and that on the bottom is from the Tyr-tRNA synthetase. The homologous region begins in Ala-tRNA synthetase at residue 370, ends at residue 382, and covers residues 9 to 25 in the tyrosine enzyme. The residues at either side of the dashes are actually adjacent in the linear sequence, but were separated here in order to achieve maximum homology. Amino acids that are identical in both sequences are boxed. (B) Structure and function map of alanyl-tRNA synthetase polypeptide. The horizontal line represents the subunit, and the vertical bar in the very center of this line indicates the carboxyl terminus of the 48,000-dalton amino terminal proteolytic fragment. The location of the T-1 fragment is indicated, and the asterisk is placed at the region of the subunit having homology with a portion of *E. coli* tyrosyl-tRNA synthetase.

of the subunit is important for full adenylate formation activity, and because adenylate formation is a reaction common to all tRNA synthetases, it is perhaps significant that homology is found in this region. Further evidence of the importance of this homology is the extensive homology (11 of 17 amino acids are identical) in this region between the tyrosine enzymes from *E. coli* and *B. stearothermophilus* (*11*). Because this homology occurs in an important region, it may be found between other tRNA synthetases.

As for comparisons with the *E. coli* tryptophanyl-tRNA synthetase, 15 nucleotides in a block of 18 in the carboxyl terminal half of the tryptophan enzyme are identical to the sequence of nucleotides 2221 to 2238 in the carboxyl terminal half of alanyl-tRNA synthetase. The resulting amino acid sequences for these codons are

Leu-Glu-Lys-Glu-Leu-Gln
Leu-Glu-Lys-Glu-Phe-Glu

where the top sequence is the alanine enzyme. Therefore this match in the nucleotide sequences occurs in the same reading frame, but we do not know enough about the enzymes to comment further on this match.

Because lengthy internal repeats are thought to occur in tRNA synthetases with subunits larger than molecular weight 60,000 (*1, 5–9*), we expected the alanine enzyme to contain similar repeats. However, repeated sequences longer than tetrapeptides are not present, and only five tetrapeptides occur twice in the sequence. The lack of lengthy sequence repeats may mean that gene duplication and fusion were not part of the history of this tRNA synthetase. It also raises questions about earlier assertions of sequence repeats, because such assertions were not based upon complete sequence information.

Another notable feature of the sequence is the clustering of charged amino acids. For example, there are clusters of arginine residues (four of the five at positions 432 to 436, and four of the six at positions 555 to 560). Such charged areas could interact with the tRNA or ATP and may be present in other aminoacyl-tRNA synthetases; they may also participate in binding to specific DNA segments (*13*). In addition, there are groups of mixed charged residues (for instance, positions 133 to 140, 201 to 206, 346 to 350, and 645 to 650). Yeast aspartyl-tRNA synthetase has such clusters of charged residues (*19*) as does the S7 ribosomal protein which interacts with 16*S* RNA (*20*).

Because of the speed and reliability of recently developed methods for sequencing DNA (*14, 21*), it is attractive to use these techniques to determine primary sequences of encoded polypeptides. This practice, however, can lead to erroneous protein sequences. For example, there is at least one instance where different laboratories determined the DNA sequence of the same gene and from it deduced the amino acid sequence without direct data from the protein (*22–24*). Although the DNA sequences agreed at more than 98 percent of the nucleotides, only about 71 percent of the translated amino acids were the same. Almost all of the disagreement stemmed from the accidental insertion or deletion of a single nucleotide, which altered the reading frame for long stretches of amino acids.

In the sequence determination of alanyl-tRNA synthetase we found that, even though more than 80 percent of the DNA was sequenced on both strands, the protein is so large that frameshift errors did occasionally occur, and premature termination codons did not always reveal the mistake. Having independently derived data from the protein greatly increased our confidence in the polypeptide sequence.

SCOTT D. PUTNEY, NANCY J. ROYAL
HENRY NEUMAN DE VEGVAR
WALTER C. HERLIHY
K. BIEMANN, PAUL SCHIMMEL
Departments of Biology and
Chemistry, Massachusetts Institute of Technology, Cambridge 02139

References and Notes

1. P. R. Schimmel and D. Soll, *Annu. Rev. Biochem.* **45**, 601 (1979).
2. S. D. Putney, R. T. Sauer, P. R. Schimmel, *J. Biol. Chem.* **256**, 198 (1981).
3. K. Biemann, *Pure Appl. Chem.* **50**, 149 (1978).
4. W. C. Herlihy, N. J. Royal, K. Biemann, S. D. Putney, P. R. Schimmel, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 6531 (1980).
5. M. R. Kula, *FEBS Lett.* **35**, 299 (1973).
6. G. L. E. Koch, Y. Boulanger, B. S. Harley, *Nature (London)* **249**, 316 (1974).
7. R. M. Waterson and W. H. Konigsberg, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 376 (1974).
8. C. J. Bruton, R. Jakes, G. L. E. Koch, *FEBS Lett.* **45**, 26 (1974).
9. S. Robbe-Saul, F. Fasiolo, Y. Boulanger, *ibid.* **84**, 57 (1977).
10. G. P. Winter and B. S. Hartley, *ibid.* **80**, 340 (1977).
11. D. Barker, personal communication.
12. K. Muench and C. Yanofsky, personal communication.
13. S. D. Putney and P. R. Schimmel, *Nature (London)* **291**, 632 (1981).
14. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
15. S. D. Putney, D. L. Melendez, P. R. Schimmel, *J. Biol. Chem.* **256**, 205 (1981).
16. Abbreviations: Gln, glutamine; Glu, glutamic acid; Leu, leucine; Lys, lysine; and Phe, phenylalanine.
17. J. M. Hood, A. V. Fowler, I. Zabin, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 113 (1978).
18. R. Co and P. Schimmel, unpublished observations.
19. Y. Boulanger, S. Potier, P. Walter, N. Hounwanou, J. Reinbolt, in *Abstracts, tRNA Workshop* (European Molecular Biology Organization/Federation of European Biochemical Societies, Strasbourg, France, 1980).
20. B. Ehresmann, J. Reinbolt, C. Backendorf, D. Tritsch, J. P. Ebel, *FEBS Lett.* **67**, 316 (1976).

21. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
22. J. Chou, M. J. Casadaban, P. G. Lemaux, S. N. Cohen, *ibid.* **76**, 4020 (1979).
23. F. Heffron, B. J. McCarthy, H. Ohtsubo, E. Ohtsubo, *Cell* **17**, 1153 (1979).
24. J. Chou, P. G. Lemaux, M. J. Casadaban, S. N. Cohen, *Nature (London)* **282**, 801 (1979).

25. Supported by NIH grants GM05472 (K.B.) and GM23562 (P.S.). S.P. was a National Institutes of Health predoctoral trainee (1977–80). We thank B. Meeusen for technical assistance and J. M. Hood for supplying computer programs used to locate sequence homologies.

8 June 1981; revised 29 July 1981

Familial Retinoblastoma and Chromosome 13 Deletion Transmitted via an Insertional Translocation

Abstract. *Surviving persons from a kindred in which retinoblastoma occurred over four generations, transmitted by eight unaffected individuals, underwent chromosomal analysis. The results revealed that the development of retinoblastoma was associated with a constitutional chromosome deletion del(13)(q13.1q14.5) and that the unaffected transmitting state was associated with a balanced insertional translocation. These findings indicate that predisposition to retinoblastoma may be attributed to the loss of specific genetic material and that a chromosomal mechanism may explain apparent lack of gene penetrance in certain families. The development of unilateral, and not bilateral, retinoblastoma suggests either that the chromosome deletion is different from the mutation of heritable retinoblastoma in general, or that the chromosome deletion lessens the probability of subsequent somatic carcinogenic events.*

Familial retinoblastoma has generally been attributed to an autosomal dominant mutation, with high penetrance and expressivity (unilateral as opposed to bilateral involvement) noted in prospective studies. However, low penetrance and expressivity have been described in some families and attributed to host resistance genes, gonadal mosaicism, or delayed mutation (*1–4*). We have used chromosomal and biochemical markers to define more precisely the genetics of retinoblastoma in a family with unilateral retinoblastoma and many unaffected transmitting relatives.

This family, which has been described (*5*), was brought to our attention through an unaffected individual who expressed concern over the frequency of retinoblastoma in his family. He reported that retinoblastoma had occurred in nine relatives over three generations, transmitted by seven unaffected individuals. Common ancestors of the affected individuals were identified and medical records, vital statistics, and church records were sought to verify all reported births, deaths, and illnesses in those ancestors and their descendants. Medical histories, physical examinations, and peripheral blood samples for chromosomal and biochemical genetic markers were sought for all living kindred members and spouses. Chromosomal analysis included high-resolution banding techniques, with amethopterin cell synchronization being used for prometaphase and prophase preparations (*6*). Esterase D was examined by electrophoresis and its quantitative activity was determined as described (*7*). Twenty-five other biochemical markers were analyzed by standard electrophoretic and immunologic methods (*8*).

The family (Fig. 1) was traced to individuals I-1 and I-2 who immigrated to the United States from Ireland in the mid-1800's. Among their descendants nine cases of retinoblastoma were confirmed, and an additional case was detected during the study. The ages at diagnosis of retinoblastoma ranged from 14 to 48 months, with a median of 31 months. Two affected patients had prior negative examinations under anesthesia between 1 and 9 months of age. Retinoblastoma was reported as unilateral in all cases. The presence of an unaffected retina was confirmed by ophthalmologic examination in the four living patients, and by postmortem examination in two patients (IV-21 and IV-22). Clinical characteristics of the retinoblastoma patients included failure to thrive in infancy, developmental delay, and mild to severe mental retardation. No consistent malformations were present. Six of the patients died within the year of the diagnosis of retinoblastoma, of postoperative complications (II-9, IV-21, and IV-22), progressive retinoblastoma (III-5, III-17), or whooping cough (III-3).

There was no evidence of retinoblastoma or retinal anomalies in any other individuals, including the parents of the affected patients. Adult onset cancers occurred in three individuals, including cancer of the liver (I-1), cancer of the cervix (III-2), and cancer of the prostate (III-16).

Chromosomal analysis of the three living patients with retinoblastoma and an infant hospitalized for failure to thrive at 3 months of age revealed an interstitial deletion of the long arm of chromosome 13, del(13)(q13.1q14.5), in all cells analyzed. The infant (V-12) was examined under anesthesia soon after the chromo-

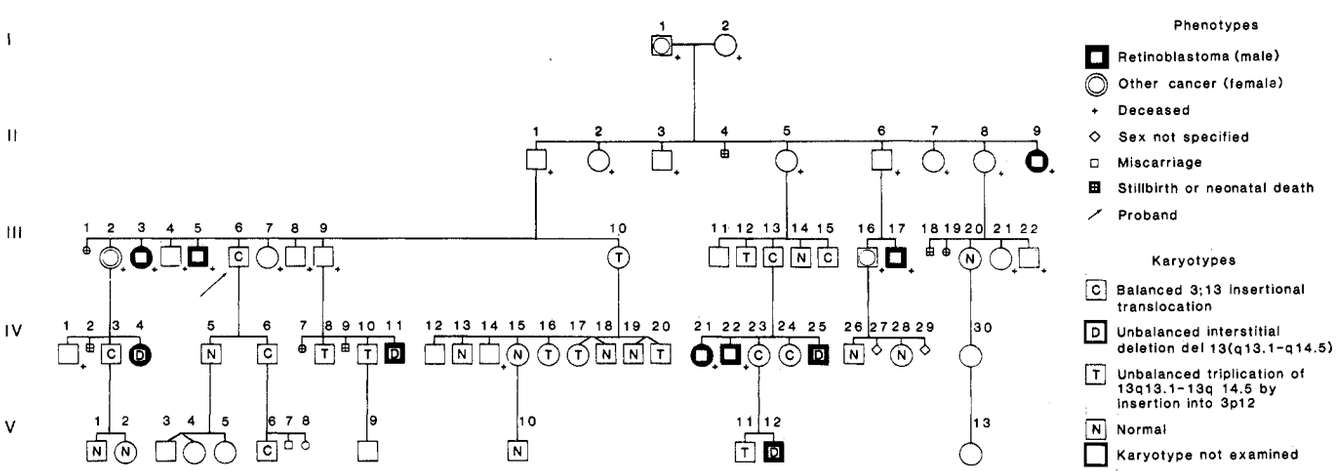

Fig. 1. Pedigree of the kindred by clinical phenotype and karyotype.

 0036-8075/81/0925-1501$01.00/0