The Yeast Lysyl-tRNA Synthetase Gene

EVIDENCE FOR GENERAL AMINO ACID CONTROL OF ITS EXPRESSION AND DOMAIN STRUCTURE OF THE ENCODED PROTEIN*

Marc Mirande‡ and Jean-Pierre Waller
From the Laboratoire d’Enzymologie, Centre National de la Recherche Scientifique, 91190 Gif-sur-Yvette, France

(Received for publication, June 8, 1988)

The nucleotide sequence of a 3.6-kilobase pair DNA fragment containing the structural gene for yeast cytoplasmic lysyl-tRNA synthetase (KRS1) and its flanking regions was determined. The encoded protein of 67,881 kDa displays a cluster of 11 lysines within a 29-amino acid residue segment at its amino-terminal extremity. Evidence is presented that this segment is responsible for the affinity displayed by the native enzyme toward polyanionic carriers. The transcription initiation sites of the KRS1 gene were determined. Upstream from the TATA box, putative control elements corresponding to the consensus sequences for the RPG box and the general amino acid control system were identified. Evidence for transcriptional induction of the KRS1 gene via the general amino acid control system is presented.

We previously reported (Cirakoglu and Waller, 1985a) that yeast aminoacyl-tRNA synthetases, like the corresponding enzymes from higher eukaryotes (Alzhanova et al., 1980, 1982), but unlike those of prokaryotic origin, display the property of binding to polyanionic carriers through ionic interactions. The behavior of purified yeast lysyl-tRNA synthetase was examined in detail (Cirakoglu and Waller, 1985a). It was shown that the native dimeric enzyme (subunit M, 73,000) interacts strongly with immobilized heparin or tRNA as well as with negatively charged liposomes under conditions where the corresponding native enzyme from Escherichia coli (subunit M, 65,000) does not. Moreover, truncated yet fully active homodimeric forms of the yeast enzyme, with subunit molecular weights of 67,000 and 65,500, respectively, were generated by controlled proteolysis with elastase and papain. The affinity of these modified forms for polyanionic carriers was found to be strongly attenuated in the case of the elastase-modified enzyme and abolished in the case of the papain-modified form. Based on these results, a structural model was proposed according to which each subunit of the native lysyl-tRNA synthetase from yeast is composed of a functional domain which is similar in size to the subunit of the corresponding prokaryotic enzyme (M, 65,000), contiguous to a structural domain of M, ~8,000 carrying a cationic net charge responsible for association with polyanionic carriers.

In order to test this model at the molecular level, the cloning and sequencing of the corresponding gene were undertaken. Cloning of the yeast lysyl-tRNA synthetase gene was recently achieved by probing a λgt11 DNA expression library with antibodies directed against the purified enzyme (Mirande et al., 1986). The cloned gene was positively identified as coding for this enzyme by showing that E. coli lysogens harboring the recombinant DNA phage produced an active yeast lysyl-tRNA synthetase displaying the proper size.

In this paper, we report the nucleotide sequence of the corresponding gene. The interpreted protein sequence reveals the presence of a cluster of cationic amino acids (11 lysines and 2 arginines versus 1 glutamic acid) located between residues 33 and 61 from the amino-terminal extremity. The amino-terminal polypeptide sequence of the elastase-modified enzyme was also determined over a length of 20 amino acid residues. The latter sequence precisely matches the stretch from residues 58 to 77 deduced from the DNA sequence, thereby confirming the identity of the cloned gene and locating the elastase cleavage point between Ala and Ser.

These results are discussed in light of the previously established behavior of the native and proteolytically modified forms of yeast lysyl-tRNA synthetase toward polyanionic carriers (Cirakoglu and Waller, 1985a). They confirm our earlier prediction concerning the existence, in yeast lysyl-tRNA synthetase, of an autonomous structural domain rich in cationic residues responsible for binding to polyanions and unrelated to the expression of catalytic activity. Comparison of the known primary sequences of several yeast aminoacyl-tRNA synthetases suggests that a similar structural organization may be a common feature of most of these lower eukaryotic enzymes.

Concerning the regulation of the expression of yeast lysyl-tRNA synthetase, the nucleotide sequence determined 5’ upstream from the first ATG codon of its structural gene harbors several putative regulatory elements. A RPG box, responsible for coordinate expression of ribosomal protein genes in yeast (Leer et al., 1985), and nucleotide sequences relevant to the general amino acid control system in yeast (Jones and Fink, 1982) were identified. It is shown that transcriptional induction of KRS1 mRNA levels via the general amino acid control system is accompanied by translational repression of the expression of the corresponding gene product.

EXPERIMENTAL PROCEDURES

Materials—Restriction endonucleases and DNA-modifying enzymes were purchased from Boehringer Mannheim, Appligene, and New England Biolabs and used as recommended by the suppliers. Radionucleotides and [14C]lysine were from Du Pont-New England Nuclear and Commissariat a l’Energie Atomique (Saclay, France), respectively.

*This work was supported in part by Grant 90.1821 from the Centre National de la Recherche Scientifique (Action Thématique Programmée “Organisation et Expression du Génome”) and by the Fondation pour la Recherche Medicale. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04186.

† To whom correspondence should be addressed.

‡ To whom correspondence should be addressed.
Homogeneous lysyl-tRNA synthetase from Saccharomyces cerevisiae and the corresponding antibodies were as previously described (Cirakoglu and Waller, 1985a; Mirande et al., 1986). Wild-type yeast genomic DNA and RNA were gifts from D. Thomas and D. Henry, (Laboratoire d’Enzymologie, Centre National de la Recherche Scientifique). Bacterial and yeast strains are listed in Table I. Yeast strains were kindly supplied by Y. S. S. K. J. N. E. C. (Laboratoire d’Enzymologie, CNRS), J.M. O. T. by Y. M. C. (Laboratoire de Biochimie, Ecole Polytechnique), and B. N. I. N. 3 by B. G. (Centre de Génétique Moléculaire, CNRS).

Analytical Procedures—Sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis was performed according to Laemmli (1970) on 10% acrylamide gels. The protein blotting procedure was conducted as previously described (Mirande et al., 1982) using affinity-purified goat anti-rabbit IgG conjugated to peroxidase (Biosys, Compiegne, France).

For Northern blot analysis, RNA was subjected to electrophoresis in formaldehyde-agarose gels as described (Maniatis et al., 1982) and blotted onto GeneScreen Plus filters (Du Pont-New England Nucleon, Boston). Hybridization was conducted as described in Eaton press blocks (In uitro transcription with bacteriophage SP6 polymerase, were used as probes for Northern and Southern experiments.

DNA and protein sequences were analyzed by using the BISANCE programs from the Centre Interuniversitaire de Traitement de l’Information (Paris).

Table I

<table>
<thead>
<tr>
<th>Bacterial and yeast strains used in this study</th>
<th>Strain</th>
<th>Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacteria</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JM1017Tr</td>
<td>Δ(lac-pro, supE, thi, recA56, strPhi10F (F’traD96, proAB, lac, lacZAM15))</td>
<td></td>
</tr>
<tr>
<td><strong>BYN103</strong></td>
<td>ΔlacU169, proA*, Δlon, araD139, strA, thi, hfl</td>
<td>A150 [chr:Trn10]</td>
</tr>
<tr>
<td><strong>Y1089</strong></td>
<td>BNN103 (pMC9)</td>
<td></td>
</tr>
<tr>
<td><strong>Yeast</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FL100</td>
<td>Wild-type MAT-α</td>
<td></td>
</tr>
<tr>
<td>Δ1278b</td>
<td>Wild-type MAT-α</td>
<td></td>
</tr>
<tr>
<td>MG409</td>
<td>arg11</td>
<td></td>
</tr>
</tbody>
</table>

DNA probes were isolated by high performance size exclusion chromatography (Schmitter et al., 1986). Protein Sequencing—Homogeneous yeast lysyl-tRNA synthetase (subunit M, 73,000) and the esterase-modified, fully active form (subunit M, 67,000) (Cirakoglu and Waller, 1985a) were subjected to N-terminal amino acid sequence analysis.

Automated Edman degradations were performed by J.-P. Le Caer, (Service Commun de Microsequençage, Laboratoire de Physiologie Nerveuse, Centre National de la Recherche Scientifique) using an Applied Biosystems 470A Sequencer. Phenylthiodyantoin derivatives were identified by reverse-phase high performance liquid chromatography.

DNA Sequencing—DNA from A. lycopersicum, a derivative of Agt11 carrying a 3.5-kilobase pair yeast genomic insert encoding lysyl-tRNA synthetase in the EcoRI site (Mirande et al., 1986), was purified from induced lyogen Y1089/Aly8 by the procedure of Maniatis et al. (1982). Large fragments of the inserted DNA were subcloned in M13mp18 or M13mp19 DNA (Yanisch-Perron et al., 1985). The complete sequence was determined by the dideoxyxenucleotide chain termination method (Sanger et al., 1977). Sets of overlapping fragments were obtained by limited Bal31 digestion, followed by end-filling with DNA polymerase I (Klenow fragment). The truncated inserts were excised, treated with calf alkaline phosphatase, separated from the M13 DNA by size exclusion chromatography, and subcloned in suitable non-denatured M13mp18 or M13mp19 DNA which was made free of its excised polynucleotide linker. The ligation mixture was used to transform JM1017Tr.

1 The abbreviations used are: SDS, sodium dodecyl sulfate; bp, base pair(s); PIPES, 1,4-piperazinediethanesulfonic acid; ApA, 5',5'-diaminohexane tetrathosphate; ORF, open reading frame.

RESULTS

Nucleotide Sequence of Yeast Genomic DNA Insert from A. lycopersicum Lysyl-tRNA Synthetase—A Agt11 recombinant DNA clone, Aly8, allowing the expression of a native lysyl-tRNA synthetase from yeast origin was previously isolated by probing a DNA expression library with antibodies directed against the purified enzyme (Mirande et al., 1986). The location of the KR51 gene, corresponding to yeast lysyl-tRNA synthetase, was determined by the characterization of two other recombinant clones, one of them allowing the expression of a β-galactosidase/lyssy1-tRNA synthetase fusion protein (Mirande et al., 1986). In this work, the nucleotide sequence of the 3.6-kilobase pair DNA insert from Aly8 was established. The complete DNA insert was sequenced on both strands by the strategy shown in Fig. 1. Generation of overlapping fragments by Bal31 deletions represents a very convenient and rapid procedure for sequencing large fragments.
A long open reading frame of 1773 bp, corresponding to the KRS1 gene, was found (positions +1 to +1773 in Fig. 2). In addition, a second open reading frame was observed 726 bp upstream from the initiation codon of the KRS1 gene. These two transcription units present opposite polarities. By using a 600-nucleotide-long anti-tRNA probe corresponding to the internal EcoRI-EcoRI fragment of the KRS1 gene, a unique polyadenylated mRNA of 1900-2000 nucleotides was detected by Northern blot analysis (Fig. 3). That the putative ORF2 gene actually encodes a protein is supported by the observation that a polyadenylated mRNA of 1700-1800 nucleotides can be detected by using an 1150-nucleotide-long anti-tRNA probe corresponding to the EcoRI-EcoRI fragment encompassing the 3'-ends of the DNA insert shown in Fig. 1.

The KRS1 translation unit has a codon bias index of 0.54, characteristic of moderately expressed proteins (Bennetzen and Hall, 1982). The corresponding value for the fragment of the unidentified ORF2 gene is 0.17, suggestive of a regulatory, low abundance protein.

**Characteristics Features of KRS1 Gene Product**—The predicted protein encoded by the KRS1 gene is 591 amino acid residues in length, with a calculated molecular weight of 67,881, in reasonable agreement with that estimated by SDS-polyacrylamide gel analysis of the purified yeast cytoplasmic lysyl-tRNA synthetase (subunit M, 73,000).

It was previously shown that the native lysyl-tRNA synthetase can be converted to a fully active modified dimer of subunit M, 67,000 upon elastase treatment (Cirakoglu and Waller, 1985a). These two homogeneous forms were subjected to automated Edman degradations. Analysis of the native enzyme reveals the presence of a blocked N-terminal residue. Taking in account the observed patterns of the effect of the penultimate amino acid on post-translational modifications in yeast (Huang et al., 1987) showing that methionine is removed and that serine is blocked when the penultimate residue is a serine, this result is in accordance with the N-terminal amino acid sequence determined for the KRS1 gene product: Met-Ser... Treatment of the native enzyme with elastase leads to exposure of a single free N-terminal residue. The following amino acid sequence was obtained: N-Ser-Lys-Lys-Lys-Thr-Asp-Leu-Phe-Ala-Asp-Leu-Asp-Pro-Ser-Gln-Tyr-Phe-Glu-Thr-Arg... corresponding to residues 58-77 of the predicted protein encoded by the KRS1 gene. This result confirms the unambiguous assignment of the DNA insert to the KRS1 gene, previously based on the observation that the λlys8 DNA insert allows the expression of an active yeast lysyl-tRNA synthetase in E. coli lysogens (Mirande et al., 1986).

It was previously shown that the aptitude of the native lysyl-tRNA synthetase to interact with polyanionic carriers in vitro was strongly decreased on elastase conversion to a fully active modified dimer of subunit M, 67,000 and was completely abolished on papain conversion to a fully active enzyme of subunit M, 65,500 (Cirakoglu and Waller, 1985a). Knowledge of the protein sequence of lysyl-tRNA synthetase provides a rational explanation for this behavior. Cleavage by elastase leads to deletion of an N-terminal fragment of 57 amino acid residues comprising a cluster of 8 lysines located between residues 33 and 55, with consequent attenuation of the affinity of the truncated enzyme for polyanionic carriers. Upon further elimination of lysine residues 59-61 by papain treatment, which is expected to remove approximately 10 additional residues (M, 67,000 → 65,500), this residual affinity is abolished. From the amino acid sequence of the KRS1 gene shown in Fig. 2, an isoelectric point of 6.36 can be calculated for the native full-length monomer of M, 67,881. Similarly, a pI of 5.79 for the elastase-modified enzyme (residues 58-591) and of 5.64 for the papain-modified enzyme (taking into account residues 67-591) can be determined. Correspondingly, the polypeptides from residues 1 to 57 and 1 to 66 display pI values of 10.09 and 10.16, respectively.

Taken together, these results strongly support the earlier proposal according to which each subunit of yeast lysyl-tRNA synthetase is composed of a functional domain which is similar in size to the subunit of the prokaryotic enzyme, contiguous to a binding domain responsible for association to negatively charged carriers (Cirakoglu and Waller, 1985a), the latter being localized in the N-terminal part of the molecule. The significance of this finding will be further addressed under "Discussion."

**Characteristics Features of KRS1 Transcription Unit**—In order to determine whether or not lysyl-tRNA synthetase from yeast is encoded by a single nuclear gene, Southern analysis of total genomic yeast DNA was performed. As shown in Fig. 4, by probing with the RNA transcript from pSP65-YKRS15, unique DNA fragments of 600, 2100, 3700, and 5000 bp were detected following digestion with EcoRI, BglII, BamHI, or HindIII, respectively. Whereas the EcoRI
FIG. 2. Nucleotide sequence of 3.6-kilobase pair yeast DNA insert containing KRSi gene and deduced amino acid sequences. The nucleotide sequence of the strand corresponding to the KRSi gene open reading frame is indicated by upper-case letters; and lysine residues are indicated by asterisks. The putative TATAA transcriptional initiation signal (position -75) is underlined, and the RNA initiation sites (positions -23, -49, and -69) are overlined. The sequence 5'-TCCTTCC-3' (positions -40 to -34), complementary to the 3'-end of 18 S ribosomal RNA (3'-AGGAAGG-5'), is underlined. The RPG box (positions -227 to -238) is indicated by an arrow, and the consensus sequences for general amino acid control (located between positions -188 and -129) are underlined. The tripartite consensus sequence for transcription termination (TAG...TATGA...TTT) and two putative AATAAA polyadenylation sites located in the 3'-flanking region are underlined. The deduced amino acid sequence of the nonsense strand corresponding to the unidentified open reading frame (ORF2) from positions -727 to -1522 is shown above the DNA sequence. Its putative TATAA box at position -693 is underlined.

fragment is internal to the KRSi gene, digestion with BglII or BamHI gives rise to large fragments encompassing the 3'-flanking regions of the KRSi gene, and digestion with HindIII corresponds to its 5'upstream sequence. These results indicate that yeast cytoplasmic lysyl-tRNA synthetase is encoded by a unique nuclear gene.

The mRNA initiation site of the KRSi gene was determined by S1 nuclease protection experiments. As shown in Fig. 5, five major initiation sites were found at positions -69, -49, -43, -32, and -23 upstream from the translation initiation codon. A Goldberg-Hogness TATAA box, corresponding to the consensus sequence for the transcription initiation signal (Chen and Struhl, 1985) is located at positions -75 to -71 (Fig. 2). At positions -40 to -34, a tripartite consensus sequence for transcription termination (TAG...TATGA...TTT) and two putative AATAAA polyadenylation sites located in the 3'-flanking region are underlined. The deduced amino acid sequence of the nonsense strand corresponding to the unidentified open reading frame (ORF2) from positions -727 to -1522 is shown above the DNA sequence. Its putative TATAA box at position -693 is underlined.
Structure of the Yeast Lysyl-tRNA Synthetase Gene

Fig. 3. Identification of \( KRS_1 \) and ORF2 transcripts. Poly(A)+ mRNA (lane 1, 2 \( \mu \)g; lane 2, 1 \( \mu \)g) or total cellular RNA (lane 3, 10 \( \mu \)g) was run in a 1% agarose gel containing formaldehyde. Northern blot analysis was conducted using the labeled SP6 transcripts corresponding to anti-\( KRS_1 \) mRNA (A) or anti-ORF2 mRNA (B). Ribosomal RNAs from Chinese hamster ovary cells (4713 and 1879 nucleotides) and yeast (3392 and 1789 nucleotides) were used as internal markers.

Fig. 4. Southern hybridization analysis of yeast genomic DNA. Total genomic yeast DNA was digested with EcoRI (lane A), BamHI (lane B), BglII (lane C), or HindIII (lane D) and subjected to fractionation in a 1% agarose gel. Southern blot analysis was carried out using an RNA probe corresponding to the 600-bp-long internal EcoRI-EcoRI fragment (positions +386 to +1009 in Fig. 2) of the \( KRS_1 \) gene. The DNA fragments of 3700 and 2100 bp visualized following digestion with BamHI (position +204) or BglII (position +229) encompass the 3′-flanking region of the \( KRS_1 \) gene, whereas the 5000-bp fragment maps the 5′-flanking region (HindIII, position +1348).

Concerning the nucleotide sequence determined 5′ upstream from the TATA box, several stretches of putative regulatory elements could be observed. Located between positions -238 and -227, a 12-nucleotide sequence (5′-TAGTGTATGGGT-3′), complementary to 5′-ACCCATACCTA-3′, is closely homologous to the sequence 5′-ACCCTACATTTA-3′, corresponding to the RPG box which is supposed to be involved in the coordinate control of ribosomal protein gene expression in yeast (Leer et al., 1985). Such conserved sequences are generally found 200-450 bp upstream from the ATG codon, either on the coding or the noncoding strand of the gene.

In addition, nucleotide sequences which were found to be relevant to the general amino acid control system in yeast (Jones and Fink, 1982) are found downstream from the TAA stop codon. Taking into account the initiation site at position -49, the polyadenylation signal at position +1899, and a poly(A) tract of about 50 adenine residues, an mRNA species of ~2000 nucleotides is expected, in accordance with the size of the transcript detected in Fig. 3.

et al., 1978). In the 3′-flanking region of the \( KRS_1 \) gene, a sequence (TAG.. .TATGA.. .TTT) homologous to the tripartite consensus sequence required for transcription termination in yeast (Zaret and Sherman, 1982) as well as twice repeated potential eukaryotic polyadenylation signal (AA-TAAA; Proudfoot and Brownlee, 1976) are found downstream from the TAA stop codon. Taking into account the initiation site at position -49, the polyadenylation signal at position +1899, and a poly(A) tract of about 50 adenine residues, an mRNA species of ~2000 nucleotides is expected, in accordance with the size of the transcript detected in Fig. 3.

Fig. 5. Determination of 5′-ends of \( KRS_1 \) transcript. Analysis of the 5′-ends of the \( KRS_1 \) transcript by S1 mapping was carried out as described under “Experimental Procedures.” The 5′-end-labeled DNA fragment (positions -272 to +68) was subjected to base-specific chemical cleavages (G + A, T + C) or hybridized to poly(A)+ mRNA. Lane 1, S1 nuclease-resistant fragments; lane 2, BamHI-HpaII fragment used as DNA probe. At left, the numbers refer to the positions in the DNA sequence shown in Fig. 2. At right, the arrows indicate the major S1 nuclease-protected ends. The S1 nuclease-resistant faint band in lane 1 migrating as a 220-bp fragment is a contamination of the probe with the downstream HpaII-EcoRI fragment of \( KRS_1 \) (positions +167 to +386).
genes involved in the biosynthetic pathways of several amino acids in yeast (Hinnebusch and Fink, 1983; Donahue et al., 1983; Hill et al., 1986; Crabell et al., 1985). As in the case of the HIS3 gene (Hill et al., 1986), a T-rich sequence precedes the TGACTC box.

General Amino Acid Regulation of KRS1 mRNA Levels—
Taking into account the putative regulatory element TGACTC, we have investigated the possibility that KRS1 gene expression may be under control of the general amino acid system in yeast. A wild-type strain (Σ1278b) and an arginine leaky auxotrophic strain (MG409) were grown in minimal medium supplemented with arginine for repression or starved for arginine for derepression of amino acid biosynthetic enzymes. Northern blot analysis of total RNA was performed to determine the steady-state level of KRS1 mRNA. An actin anti-mRNA probe was used as an internal marker. As can be seen in Fig. 6, whereas similar amounts of KRS1 mRNA were detected from Σ1278b grown in minimal medium and MG409 grown in the same medium supplemented with arginine, an approximately 5-fold increase in the KRS1 mRNA level was observed from MG409 grown under arginine restriction, with the amount of actin mRNA remaining unchanged. This result strongly argues in favor of general amino acid regulation of KRS1 mRNA expression. Lysyl-tRNA synthetase activity was determined in the corresponding crude extracts. As shown in Table II, no elevation of lysyl-tRNA synthetase activity was observed following induction of its corresponding mRNA in the presence of limiting amounts of arginine in the culture medium. As a control, L-ornithine carbamoyltransferase, an enzyme that is subject to general amino acid control in yeast (Crabeel et al., 1985), was assayed in the same extracts. A 3-fold increase of L-ornithine carbamoyltransferase activity was observed following induction by arginine starvation, as compared to the activity determined in the wild-type strain. We have verified by the protein blotting procedure that the polypeptide of M 73,000, corresponding to lysyl-tRNA synthetase, was not overexpressed in MG409 cells grown in the presence of limiting amounts of arginine (Fig. 7). Thus, the transcriptional induction via the general amino acid control system is counteracted by a post-transcriptional mechanism, leading to translational repression of the expression of lysyl-tRNA synthetase.

**DISCUSSION**

**Yeast Aminoacyl-tRNA Synthetases Are Composed of Two Independent Domains—**Lysyl-tRNA synthetases from *E. coli* and yeast display the same oligomeric structure of the α2-β2 type, yet differ significantly in their apparent subunit molecular weights: 65,000 for the *E. coli* enzyme (Hirshfield et al., 1976) and 73,000 for the yeast enzyme (Cirakoglu and Waller, 1985a). The chain extension that characterizes yeast lysyl-tRNA synthetase was shown to be responsible for the polyanionic binding property displayed by this enzyme. Whereas native lysyl-tRNA synthetase from yeast (subunit M73,000) binds strongly to polyanionic carriers, a papain-modified fully active form (subunit M65,500), like the corresponding native enzyme from *E. coli* (M65,000), does not (Cirakoglu and Waller, 1985a). In light of the results reported in this study, the binding property displayed by yeast lysyl-tRNA synthetase can be assigned to the N-terminal moiety of its polypeptide chain. Indeed, among the 66 amino-terminal residues corresponding to the portion of the molecule removed following papain treatment, the interpreted protein sequence reveals the presence of 13 lysines and 2 arginines versus 3 aspartic acids and 5 glutamic acids, accounting for a calculated pI of 10.16 versus 6.36 for the native enzyme. Moreover, a cluster of basic residues (11 lysines + 2 arginines) is located between residues 33 and 61. By using the algorithm of Garnier et al. (1978), an α-helix secondary structure can be predicted for the N-terminal part of yeast lysyl-tRNA synthetase. Assuming the validity of this prediction, a helical wheel representation for residues 33–60 leads to the striking observation that the lysine residues are distributed on one-half of the helix section (Fig. 8).

The difference in behavior toward polyanionic carriers which characterizes lysyl-tRNA synthetases from *E. coli* and
yeast appears to be a general property of the aminoacyl-tRNA synthetases from these sources. Whereas prokaryotic aminoacyl-tRNA synthetases do not interact with immobilized polyanionic carriers (Alzhanova et al., 1980), those from lower (Cirakoglu and Waller, 1985a) or higher (Alzhanova et al., 1980; Cirakoglu and Waller, 1985b) eukaryotes do. Comparison of the primary structures of those seven aminoacyl-tRNA synthetases, the sequences of which are presently known from E. coli as well as from S. cerevisiae (Fig. 9), leads to the following observations. (i) The subunit sizes of the yeast enzymes are invariably larger than those of the corresponding enzymes from E. coli. (ii) Following alignment of their primary structures, the polypeptide chain extensions that characterize the yeast enzymes are found to be located at one extremity of the molecules, rather than as insertions within the conserved regions. (iii) These chain extensions display remarkably basic isoelectric points and are generally located in the N-terminal moiety of the molecules, accounting for at least 10% of the primary structures (28% in the case of yeast glutaminyl-tRNA synthetase). One exception is yeast seryl-tRNA synthetase, which displays a C-terminal extension of 13 residues comprising a cluster of 6 lysines. It is noteworthy that similar feature prevails in the case of all eukaryotic elongation factors 1α, which harbor C-terminal extensions of 18–20 amino acid residues comprising 7 lysines relative to their prokaryotic or mitochondrial elongation factor Tu counterparts (Brands et al., 1986). It was previously shown that elongation factor 1α binds to polyanionic carriers under conditions where the corresponding elongation factor Tu from E. coli does not (Demogatsky et al., 1978). However, the relevance of this C-terminal cationic extension to polyanion binding was not tested experimentally. Isoleucyl-tRNA synthetase from yeast also carries a C-terminal extension relative to its prokaryotic counterpart. This stretch is composed of primarly acidic residues (pl 5.00 for residues 830–1073). However, a cluster of 7 lysines, absent from the E. coli enzyme, is present between residues 906 and 916 in the region just preceding the chain extension that characterizes the yeast enzyme. In the case of methionyl-tRNA synthetase, in addition to the N-terminal cationic chain extension found in the yeast monomeric enzyme, a C-terminal extension is found for the E. coli enzyme corresponding to a C-terminal domain implicated in the dimerization of the E. coli enzyme (Cassio and Waller, 1971).

In addition to the seven enzyme couples listed in Fig. 9, the primary sequence of aspartyl-tRNA synthetase from yeast is also available (Sellami et al., 1986). The N-terminal moiety of this molecule likewise displays markedly cationic properties: pl 8.87 for residues 1–100, compared to 6.84 for the native enzyme. Furthermore, helical wheel representations of the lysine-rich N-terminal domains of aspartyl-, threonyl-, and valyl-tRNA synthetases from yeast reveal an anisotropic distribution of the lysine residues (Lorber et al., 1988), as shown here for yeast lysyl-tRNA synthetase (Fig. 8).

Thus, comparison of the known primary structures of E. coli and yeast aminoacyl-tRNA synthetases suggests that a

**Fig. 8.** Helical wheel representation of lysine-rich N-terminal domain of yeast lysyl-tRNA synthetase. Amino acid residues 31–60 from yeast lysyl-tRNA synthetase, for which an α-helix secondary structure can be predicted, are arranged in a helical wheel representation as described for aspartyl-, threonyl-, and valyl-tRNA synthetases from yeast (Lorber et al., 1988).

**Fig. 9.** Comparison of homologous aminoacyl-tRNA synthetases from E. coli and yeast. Alignment of homologous aminoacyl-tRNA synthetases is based on the Kenehia alignment program (Kanehisa et al., 1984) by weighting with the mutation data matrix (Dayhoff et al., 1983). Aminoacyl-tRNA synthetases are designated as follows: XRSEC and XRSSCC, aminoacyl-tRNA synthetase specific for the amino acid X (one-letter symbol) from E. coli (EC) or S. cerevisiae cytoplasm (SCC). The region of homology is indicated by solid bars, and N- or C-terminal extensions are shown by discontinuous bars. Number 1 corresponds to the N-terminal amino acid residue. At right, calculated isoelectric points (in boldface) are shown for amino acid residues x to y. Protein sequences are deduced from the nucleotide sequences of the corresponding cloned genes: HRSEC (Freedman et al., 1985), HRSSEC (Natsoulis et al., 1986), IRSEC (Webster et al., 1984), IRRSSEC (Englsich et al., 1987), MRSEC (Barker et al., 1982; Dardel et al., 1984) MRSSCC (Walter et al., 1983), QRSEC (Hoben et al., 1982), QRRSSEC (Ludmerer and Schimmel, 1987a), SRSEC (Hartlein et al., 1987b), SRSSEC (Weygand-Durasevic et al., 1987): TRSEC (Mayaux et al., 1983), TRSSCC (Pape and Tzagoloff, 1985), VRSEC (Hartlein et al., 1987a), and VRSSCC (Jordana et al., 1987).
similar structural organization may be a common feature of most, if not all, of the lower eukaryotic enzymes. These observations support the view that lower eukaryotic aminoacyl-tRNA synthetases have evolved from an ancestral enzyme similar in size to that from prokaryotes by acquiring a structural domain conferring to them the ability to bind to polyanionic carriers. As regards the functional significance of this evolutionary acquisition, it was proposed earlier that this binding domain may promote the compartmentalization of these enzymes within the cytoplasm, at or near the site of protein synthesis, through electrostatic interactions with as yet unidentified cellular components carrying negative charges (Cirakoglu and Waller, 1985a). The anisotropic distribution of the cationic charges within the binding domain may be optimally suited to ensure oriented anchorage of these enzymes to polyanionic surfaces.

Tentative experiments were recently carried out to test the significance of the large amino-terminal extension borne by glutaminyl-tRNA synthetase from *S. cerevisiae* (Ludmer and Schimmel, 1987b). No evidence for a specific role was found. However, the internal deletions examined did not concern the entire N-terminal extension of 290 residues. In particular, the 9 lysines located between positions 188 and 204 were present in all of the constructions analyzed.

**Cytoplasmic Yeast Lysyl-tRNA Synthetase Is Encoded by Unique Nuclear and Inducible Genes**

Lysyl-tRNA synthetase from *E. coli* is encoded by two genes: a constitutive gene (*lysS*) and a heat-inducible gene (*lysU*) which were mapped at 62.1 and 92 min on the *E. coli* chromosome, respectively (Emmerich and Hirshfield, 1987; Van Bogelen et al., 1983). By contrast, Southern analysis of total genomic yeast DNA reveals the presence of a single nuclear gene encoding lysyl-tRNA synthetase.

It was shown that lysyl-tRNA synthetase from *E. coli*, yeast, or sheep liver catalyzes the *in vitro* synthesis of 5',5'-diadenosine tetraphosphate (Blancuet et al., 1983). Several lines of evidence have suggested that this unusual dinucleotide may act as a pleiotropic alarmone (Varshavsky, 1983). Assuming that in vivo the *lysU* gene product is implicated in APnA synthesis following the heat-shock response, we have to consider that in yeast its synthesis may be ensured in another way. However, the possibility that *KRS1* gene expression may be heat-inducible or that another form of yeast lysyl-tRNA synthetase may be encoded by a gene harboring too low homology with *KRS1* to be detected by hybridization cannot be completely dismissed.

### Regulation of Expression of Yeast Lysyl-tRNA Synthetase

Whereas structural elements involved in the regulation of the expression of the three bacterial aminoacyl-tRNA synthetases specific for the amino acids alanine (Putney and Schimmel, 1981), threonine (Springer et al., 1985), and phenylalanine (Fayat et al., 1983) have been identified and characterized, examination of the 5'-upstream regions of the yeast aminoacyl-tRNA synthetase genes sequenced to date has provided little insight on the regulatory mechanisms implicated in their expression. Concerning the expression of yeast lysyl-tRNA synthetase, several putative regulatory elements were identified 5' upstream from the ATG codon of the *KRS1* gene.

The genes encoding the majority of yeast ribosomal proteins and elongation factor 1α were shown to harbor conserved sequences (HOMOL1 and RPG box) in their 5'-flanking regions (Leer et al., 1986). It was proposed that a DNA-binding protein component may be a general factor for transcriptional activation of these genes (Huet et al., 1985). By extension, it may be inferred that the RPG box found in the 5'-noncoding region of the *KRS1* gene allows coordinated regulation of its expression, in relation to a large family of genes coding for the translational machinery in yeast.

Located downstream from the RPG box, sequence elements relevant to the general amino acid control system were found. It was shown that these regulatory elements alone are able to confer general amino acid control on the *E. coli* β-galactosidase gene (Silverman et al., 1982) or on the yeast cytochrome c gene (Hinnebusch et al., 1985). Moreover, a single mutation in the TGACTC sequence affects the binding of the GCN4 activator protein (Hill et al., 1986). The results presented in this report unambiguously demonstrate that transcription of the *KRS1* gene is controlled by the general amino acid system. A similar observation was previously reported for the *ILS1* gene encoding yeast isoleucyl-tRNA synthetase (Meusdoerff and Fink, 1983). The nucleotide sequence of that gene was recently determined (Englisch et al., 1987). In the 5'-upstream region of the *ILS1* gene, a TGACTC sequence can be observed (positions -102 to -97) preceded by a T-rich sequence (positions -151 to -134). However, no data concerning the expression of the *ILS1* gene product are available. In the case of yeast lysyl-tRNA synthetase, we have observed a post-transcriptional repression of its expression following arginine starvation of a leaky auxotrophic strain. The molecular features responsible for this secondary effect remain to be determined.

Concerning the physiological significance of these findings, it must be pointed out that arginine restriction should result in an increased level of lysine biosynthesis since the enzymes from the corresponding biosynthetic pathway are subject to the general amino acid control system (Jones and Fink, 1982).

In the case of the expression of bacterial aminoacyl-tRNA synthetases, it was shown that the level of tRNA aminoacylation is involved in the regulation of the cognate synthetases specific for threonine (Springer et al., 1985) and phenylalanine (Fayat et al., 1983). Assuming that tRNA is involved in the regulation of yeast lysyl-tRNA synthetase, it may be predicted that conditions which promote a decrease in the *in vivo* aminoacylation level of tRNA would allow overexpression of lysyl-tRNA synthetase following activation of *KRS1* transcription. This possibility will be tested by submitting a lysine leaky auxotrophic strain to lysine starvation. In considering the mechanisms implicated in translational repression, we have to envisage the possibility that lysyl-tRNA synthetase may repress its own synthesis by interacting with *KRS1* mRNA through its cationic N-terminal moiety. However, an autogenous repression mechanism cannot be a priori, the finding that the level of enzyme is not significantly altered following a 5-fold derepression of its messenger RNA.

Two additional features of the 5'-end of *KRS1* mRNA are noteworthy. First, this portion of the molecule can be folded into stable secondary structures encompassing nucleotides -50 to +100. In particular, a four-stem structure, accounting for a total stability of -32.7 kcal, can be generated within the N-terminal part of *KRS1* mRNA. Second, the amino acid composition of the N-terminal protein sequence encoded by nucleotides +1 to +183 is highly biased, comprising 12 alanine residues in addition to the 13 lysines mentioned earlier (37 alanines and 48 lysines among the 591 residues of the full-length protein). The possibility that the 5'-end of *KRS1* mRNA is involved in the repression of lysyl-tRNA synthetase expression via the formation of stable secondary structures and/or its abnormal codon composition is being investigated.

**Acknowledgments**—We are indebted to J. Rossier and J.-P. Le Caer for the protein sequence analyses, to R. Giege for communicating results prior to publication, to Y. Surdin-Kerjan and D. Thomas for fruitful advice, and to M.-T. Latreille for excellent technical assistance during part of this work.