Vol. 165, No. 1, 1989 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1989 Pages 464-473

PRIMARY STRUCTURE OF THE S. CEREVISIAE GENE ENCODING URIDINE MONOPHOSPHOKINASE

Patricia Liljelund[1*], Ambaliou Sanni[2], James D. Friesen[3], and François Lacroute[4]

[1] Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario M5G 1L6, Canada

[2] Departement de Biochimie et de Biologie Cellulaire, Université Nationale du Benin, République Populaire du Benin

[3] Research Institute, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada

[4] Centre de Génétique Moléculaire, C.N.R.S., 91191 Gif-sur-Yvette, France

Received October 17, 1989

In the yeast Saccharomyces cerevisiae, the biosynthesis of both pyrimidine nucleoside triphosphates UTP and CTP is dependent on the activity of the uridine monophosphokinase step. We have determined the sequence of the uridine monophosphokinase gene. The coding region is 615 base pairs long and encodes 205 amino acids (22,500 daltons). The 5' terminus is comprised of a 17 amino acid-long hydrophobic leader sequence which is not present in genes encoding adenylate kinases. The coding region shows a strong degree of homology with the cytosolic adenylate kinases of vertebrates, and a lesser degree of homology with yeast and E. coli adenylate kinases.

In the budding yeast S. cerevisiae, pyrimidines are synthesized endogenously via a pathway comprised of six enzymatic steps encoded by five structural genes (1). Both endogenously synthesized and exogenously or catabolically-derived pyrimidines enter the pathway at uridine monophosphate (2), which is sequentially converted to UDP, UTP, and CTP. The enzyme uridine monophosphokinase catalyzes the transfer of a phosphate from ATP to UMP to produce a molecule of UDP and one of ADP. It hence represents the pyrimidine biosynthetic enzyme analogous to the extensively studied enzyme adenylate kinase (3-14) which transfers a phosphate from ATP to AMP to produce two molecules of ADP in the biosynthesis of purines. The essential role of uridine monophosphokinase in furnishing the cell with pyrimidines requires that corresponding mutants be selected as conditional lethals. A S. cerevisiae mutant for this gene displaying thermosensitivity and 5-fluorouracil resistance as well

*To whom correspondence should be addressed.

0006-291X/89 $1.50

as thermolabile uridine monophosphokinase activity was obtained; the observation that UTP and CTP pools are reduced to 10% of their normal size already thirty minutes after shifting the mutant to non-permissive temperature demonstrates the importance of this enzymatic step in regulation of the cellular RNA precursor pools (15). It was therefore possible to clone the wild-type gene by complementation of the mutant phenotype (15). To date, this yeast gene (URA6) is the only eukaryotic or prokaryotic gene encoding a pyrimidine nucleoside kinase to have been isolated, although cold-sensitive mutants for the uridine monophosphokinase (16) and nucleoside diphosphokinase (17) loci have been selected in S. typhimurium.

In this report we show the nucleotide sequence of the gene encoding yeast uridine monophosphokinase and examine various structural implications.

METHODS

Strains Used

The S. cerevisiae strain FL100a was the wild-type strain for all yeast-related manipulations. The bacterial strain JM103 was the host for M13 transformation.

Preparation and Analysis of RNA

Total yeast RNA was extracted using a phenol-cresol method (18). The procedure of Thomas (19) was followed for northern blotting.

Preparation and Analysis of DNA

Plasmid DNA purification, restriction enzyme digestion and ligation, filling of recessed termini, and nick translation were carried out according to Maniatis et al. (20). Phage DNA was purified as specified by the suppliers (Biolabs and P. L. Biochemicals).

Nucleotide Sequencing

Fragments cloned into the appropriate replicative form of M13 mp18 or mp19 phages (21) were sequenced by the dideoxynucleotide termination chain method following the protocol of Sanger et al. (22). Both specific cloning and shotgun approaches were used; in several instances oligonucleotide primers were necessary, in particular, to sequence over the two NcoI sites.

RESULTS AND DISCUSSION

Two genes are present

The gene URA6 coding for the enzyme uridine monophosphokinase in S. cerevisiae was originally cloned by complementation of a conditional mutant displaying a thermosensitive and 5-fluorouracil-resistant phenotype (15). A 2 kilo-base HpaI-ClaI fragment was shown to complement the mutant phenotype, and a poly A^+ transcript of 2.2 kb. to hybridize to probes derived from the central region (15).

However, the nucleotide sequence of this fragment reveals the presence of two major open reading frames on the strand established by DNA-RNA filter hybridization to be the coding strand (data not shown), one relatively short one (615 nucleotides) and one much longer one which extends at least 0.5 kb. in the 3′ direction beyond the original 2 kb. complementing fragment. (See Fig. 1.) In confirmation of this reassessed situation, a Northern blot probed with the entire 2 kb. HpaI-ClaI fragment indicates the presence of two polyadenylated transcripts of 0.85 kb. and 2 kb. (Figure 2). It had previously been observed (15) that if the original complementing clone, the 2 kb. HpaI-ClaI fragment, was deleted between the PvuII and HpaI sites, thus selectively removing the promoter region and the ATG codon for the smaller open reading frame, both the ability to complement the thermosensitive phenotype and to restore sensitivity to 5-fluorouracil (the wild-type phenotype) were lost. We thus conclude that the 615 nucleotide open reading frame situated towards the 5′ end of the 2 kb. HpaI-ClaI fragment and representing a coding capacity of 22.500 daltons corresponds to the uridine monophosphokinase gene. The comparable small size of 20,000-26,000 daltons found for adenylate kinases (23) reinforces this conclusion.

DNA sequence of the uridine monophosphokinase gene

The codon usage in the 615 nucleotide open reading frame corresponds to a codon bias index (24) of 0.213, a value situated between those for URA3 (0.204)

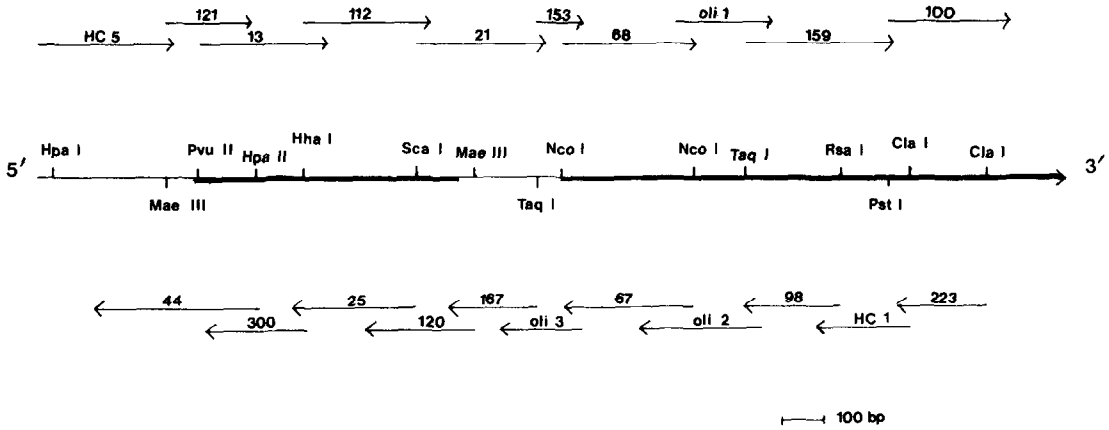

Figure 1. Restriction map and sequencing strategy for the URA6 gene. The two open reading frames are represented as thickened lines.

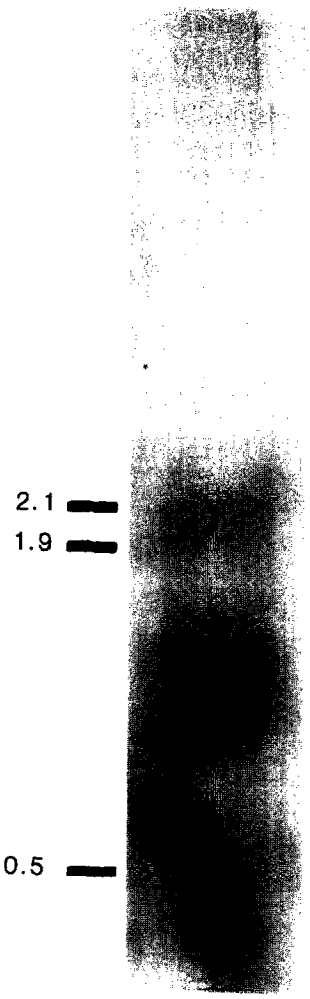

Figure 2. Northern blot showing presence of two transcripts. 10 µg. of poly(A)$^+$ RNA were probed with the nick-translated HpaI-ClaI fragment. Labelled HindIII-digested λ DNA was used as a size marker.

and HIS1 (0.232) (25), and thus in the range of a structural gene expressed at an average level. This is supported by the extent of hybridization observed when total yeast RNA is probed with a URA6 fragment (data not shown). An average level of expression for uridine monophosphokinase contrasts with that of adenylate kinase, which at least in E. coli has been shown to represent one of the more abundant cellular proteins (9). Such a discrepancy would in fact correlate well with the relative sizes of the ATP and UTP pools, in yeast for example, 3.15 mM and 0.9 mM respectively (15), as well as with the continuous and intense cellular needs for energy, requiring high levels of ATP synthesis as

opposed to a lower level of nucleoside triphosphate synthesis required to assure the pyrimidine pool levels. Similar amounts of basic (28) and acidic (27) amino acids are present, and no regions of particular charge or polarity internal to the gene are evident. Nevertheless, an interesting feature unique to the uridine monophosphokinase sequence is the presence of a 17 residue hydrophobic "leader" at the 5′ end. Whereas the adenylate kinases are characterized by a short and highly charged 5′ region including both acidic and basic residues (2/9 and 4/9 respectively), only one charged residue (1/17) is found in the long uridine monophosphokinase leader segment. (See Fig. 3.)

Such a striking difference at the 5′ terminal when both the size of the proteins and the overall amino acid sequences (see farther ahead) are very

```
      GAATCCACGCTACAGGTTTATACTGAGATTTTATTCTTCGAAAAAACCAACTTTCCACAATACAGCACCTTCGAAGACCA
      ACGTAAACGTACCACGCGCTAACAAAAGTCAAAGCAAGGGTAAACACAAAGGCAAATTGTTGGTACTCGTAGGAACTTTA
      GCACTCGTGACATCAGTTATTTCGGTAAACTATCAAAAGAACGAACCAGTTGAATTTCTTGAATAAGACAA ATG ACA
                                                                                Met Thr

    7 GCT GCC ACT ACA TCA CAG CCA GCT TTC TCG CCT GAC CAA GTT TCC GTG ATC TTC GTT CTA
      Ala Ala Thr Thr Ser Gln Pro Ala Phe Ser Pro Asp Gln Val Ser Val Ile Phe Val Leu

   67 GGA GGA CCC GGT GCA GGC AAG GGT ACT CAG TGT GAA AAA CTA GTT AAG GAC TAT TCA TTT
      Gly Gly Pro Gly Ala Gly Lys Gly Thr Gln Cys Glu Lys Leu Val Lys Asp Tyr Ser Phe

  127 GTC CAT TTG TCA GCC GGA GAC CTT CTA CGT GCT GAG CAG GGC AGA GCA GGT TCC CAA TAT
      Val His Leu Ser Ala Gly Asp Leu Leu Arg Ala Glu Gln Gly Arg Ala Gly Ser Gln Tyr

  187 GGG GAA TTG ATC AAG AAC TGC ATC AAA GAG GGC CAG ATT GTC CCT CAA GAG ATT ACT TTG
      Gly Glu Leu Ile Lys Asn Cys Ile Lys Glu Gly Gln Ile Val Pro Gln Glu Ile Thr Leu

  247 GCG CTT TTA CGC AAC GCT ATT TCC GAT AAC GTC AAG GCG AAC AAG CAT AAG TTC TTA ATT
      Ala Leu Leu Arg Asn Ala Ile Ser Asp Asn Val Lys Ala Asn Lys His Lys Phe Leu Ile

  307 GAC GGA TTT CCT AGG AAG ATG GAT CAA GCC ATT TCC TTT GAA AGA GAC ATC GTT GAA AGC
      Asp Gly Phe Pro Arg Lys Met Asp Gln Ala Ile Ser Phe Glu Arg Asp Ile Val Glu Ser

  367 AAA TTC ATC CTG TTC TTT GAC TGC CCT GAA GAT ATC ATG TTA GAG AGA CTA TTG GAG CGT
      Lys Phe Ile Leu Phe Phe Asp Cys Pro Glu Asp Ile Met Leu Glu Arg Leu Leu Glu Arg

  427 GGC AAG ACC AGT GGT AGA AGC GAT GAC AAC ATT GAG TCC ATT AAG AAG AGA TTT AAC ACT
      Gly Lys Thr Ser Gly Arg Ser Asp Asp Asn Ile Glu Ser Ile Lys Lys Arg Phe Asn Thr

  487 TTC AAG GAG ACT AGT ATG CCC GTC ATC GAG TAC TTT GAA ACC AAA TCG AAA GTC GTC CGT
      Phe Lys Glu Thr Ser Met Pro Val Ile Glu Tyr Phe Glu Thr Lys Ser Lys Val Val Arg

  547 GTT CGT TGC GAC AGA TCC GTC GAA GAT GTG TAC AAA GAC GTC CAA GAC GCT ATC CGT GAT
      Val Arg Cys Asp Arg Ser Val Glu Asp Val Tyr Lys Asp Val Gln Asp Ala Ile Arg Asp

  607 AGC TTA TAG ACTAAGACGTATTTTATATAGTTTACGCTATAAGTTACGGCGATATTATATATATATGTAAGAATAT
      Ser Leu End

      ATACATATATATTAAAATGATTATATGCCCTCCCCTACATATCAGACCTTTACAGGGTAACCTTAAAATAAACTGCTAAA
      TTATTTG
```

Figure 3. DNA and predicted protein sequence of the yeast uridine monophosphokinase gene. Shown is the sense strand of DNA sequence from 5′ to 3′.

similar is strongly suggestive of a specific biological function for this structure. This hydrophobic amino terminal could represent a signal or "topogenic" sequence (26), specifically directing the enzyme molecule across an intracellular membrane and into another cellular compartment. It is in fact possible that in yeast, pyrimidine biosynthesis occurs in the nucleus since evidence exists (27) that the first enzyme of the pathway, the aspartate transcarbamylase-carbamyl phosphate synthetase complex, migrates into the nucleus after being synthesized in the cytoplasm.

Homology of uridine monophosphokinase with adenylate kinase

Comparison of the amino acid sequence predicted for the uridine monophosphokinase protein with that of adenylate kinase in various organisms shows strikingly close homology which extends over the entire gene. The adenylate kinases form a large family of isoenzymes which can be divided into two classes (23): the mammalian cytosolic enzymes or "small variants," and the mitochondrial and microbial enzymes or "large variants," so named because they possess an insertion of 31 amino acids in a region shown to be part of the active site (13). Unexpectedly, yeast uridine monophosphokinase is much more homologous with the "small variants" or mammalian cytosolic enzymes than it is with either the yeast cytosolic adenylate kinase or with the E. coli enzyme, and in fact entirely lacks the insertion typical of the "large variant" adenylate kinase (situated between residues 145 and 146 using umpk numbering). For example, porcine (3), rabbit (8), bovine (8), human (5), and chick (12) adenylate kinase respectively show 43.4%, 42.3%, 42.9%, 42.3%, and 40.1% identity with the uridine monophosphokinase gene at the amino acid level, homologies dispersed across the entire coding region. On the other hand, yeast adenylate kinase (11) shows only 25.6% identity and E. coli adenylate kinase (9), 24.7% identity. (See Fig. 4.)

Why is yeast uridine monophosphokinase evolutionarily closer to the mammalian cytosolic adenylate kinase than to either the yeast cytosolic adk or to the E. coli adk? We suggest an evolutionary scheme by which the uridine monophosphokinase gene evolved by classical mechanisms of duplication followed

Porcine adk (4)						5'	MET	GLU	GLU	LYS	LEU	LYS	LYS	SER	LYS	ILE	ILE	PHE		
Rabbit adk (9)							MET	GLU	GLU	LYS	LEU	LYS	LYS	ALA	LYS	ILE	ILE	PHE		
Bovine adk (9)							MET	GLU	GLU	LYS	LEU	LYS	LYS	ALA	LYS	ILE	ILE	PHE		
Human adk (6)							MET	GLU	GLU	LYS	LEU	LYS	LYS	THR	LYS	ILE	ILE	PHE		
Chick adk (13)						MET	SER	THR	GLU	LYS	LEU	LYS	HIS	HIS	LYS	ILE	ILE	PHE		
Yeast adk (12)										SER	SER	GLU	SER	ILE	ARG	MET	VAL	---		
E. coli adk (10)															MET	ARG	ILE	ILE	LEU	
Yeast umpk	MET	THR	ALA	ALA	THR	THR	SER	GLN	PRO	ALA	PHE	SER	PRO	ASP	GLN	VAL	SER	VAL	ILE	PHE

Figure 4. Comparison of amino acid sequences of the 5′ leader regions. Adenylate kinase and uridine monophosphokinase are abbreviated as adk and umpk respectively. Charged residues have been boxed.

by divergence to specific functions such as has been proposed in other cases, for example, the genes encoding the largest subunits of the three RNA polymerases in yeast (28) - and that the real "maverick" here is the large variant adenylate kinase, which is a special form developed to meet the heavy energy demands of rapidly growing microorganisms.

Structural predictions for uridine monophosphokinase

Uridine monophosphokinase and adenylate kinase share the property of being small proteins, an advantage which has allowed crystallization and structural analysis of adenylate kinase, in contrast with the much larger protein kinases. In view of the similar size, overall amino acid sequence, and function, it seems reasonable to extend the structural analysis of adenylate kinase to uridine monophosphokinase. The immediately noticeable homology between the two enzymes is the complete conservation of the glycine-rich loop (residues 23-30 in umpk), attesting to its universal importance in binding of the ATP molecule. This structure has been shown to form an "anion hole" implicated in phosphate binding (29), and is highly conserved among mononucleotide binding proteins (23).

There is some conflict in the literature between X-ray and NMR determination of residues involved in binding ATP and AMP, but the generally accepted model is the presence of two hydrophobic "pockets" in the adenylate kinase molecule, which would bind the adenosine moiety of Mg ATP and AMP respectively (7, 10). The active site is seen as a deep cleft in the middle of the molecule (4), which has been shown to exist in two crystallographic forms corresponding to bound and unbound to substrates (6). If one considers the residues implicated in formation of the nucleotide binding sites in adenylate

kinase, the hydrophobic pocket established by X-ray diffraction as the binding site for Mg ATP in porcine cytosolic adk (7) is lined by Val 67, Leu 69, Val 72, Leu 73, Leu 76 on one side, and Ile 92, Tyr 95, Arg 97, and Gln 101 on the other side, all of which are present in uridine monophosphokinase except Leu 69 and Val 72, and Tyr 95, a conservative replacement. Furthermore, the residue His 36, which has been shown to be situated deep in the active site cleft of adenylate kinase and can provoke interconversion of the two structural forms of the enzyme by being protonated (30), is also found in uridine monophosphokinase, suggesting a similar conformation and mechanism of the active site in the two enzymes.

The nature of the AMP binding site in adenylate kinase remains somewhat ambiguous. The original AMP-binding site proposed for the small variant porcine cytosolic adk using X-ray diffraction (7) - probably most pertinent to uridine

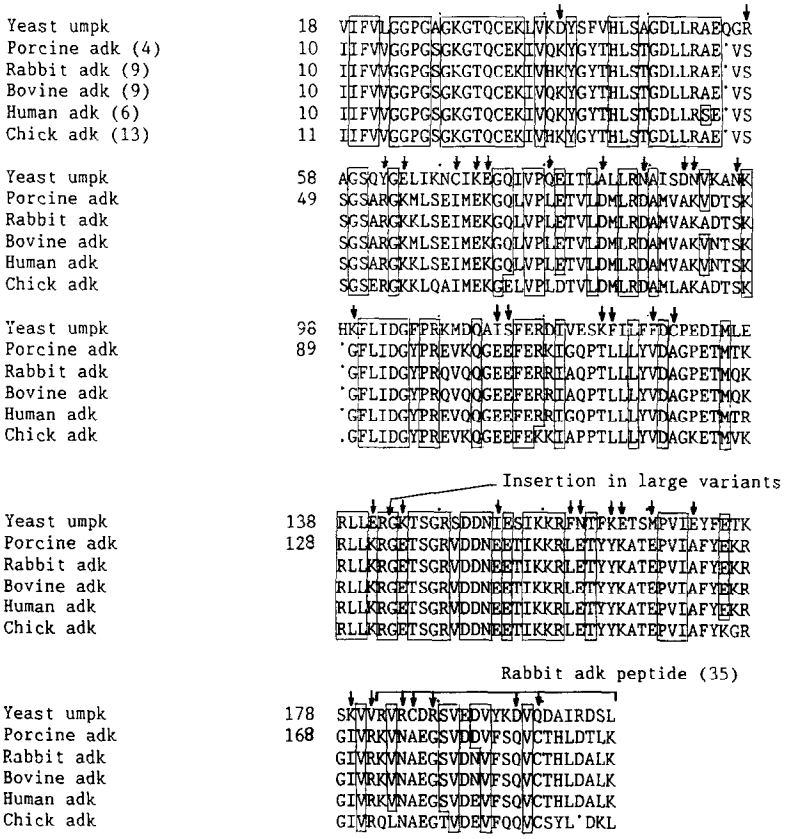

Figure 5. Alignments between protein sequences of uridine monophosphokinase and adenylate kinases (small variants). Identical residues have been boxed. Residues completely conserved among the adenylate kinases but which correspond to non-conservative changes between adk and umpk are indicated by an arrow.

monophosphokinase on account of its closest homology to the small variants - is bordered by Ile 11, Val 13, Leu 91, Leu 116, Ser 19, Gly 20, Gln 24, Ile 28, and Val 186, all found in uridine monophosphokinase except Ser 19 and Ile 28, which are conservatively replaced. Using NMR techniques, Hamada et al. (31) observed that a peptide of residues 171-193 in rabbit adenylate kinase binds AMP. The corresponding region in uridine monophosphokinase (residues 182-201) shows overall conservative homology to the rabbit sequence with nevertheless the introduction of several charged amino acids (using rabbit adk numbers, Asn 173 → Arg, Gly 176 → Arg, Ser 183 → Lys) and polar residues, (Ala 174 → Cys, Leu 189 → Gln). These positions are interesting because the corresponding adk residues are highly conserved between species, implying that these changes in umpk may indeed play a role in binding specificity, as highly conserved residues would be expected to have a particular functional necessity. A number of additional residues in uridine monophosphokinase show strikingly different polarity or charge compared with the amino acid at the homologous position in adenylate kinases (indicated in Figure 5). It could be surmised that among these residues are those that confer the binding specificity for UMP on the uridine monophosphokinase enzyme, a mechanism which might involve, for instance, some sort of charge relay system or H-bonding interaction with the two carbonyl groups of uracil.

It will be interesting to try to alter the nucleotide binding specificity of the uridine monophosphokinase molecule by using in vitro mutagenesis to change certain of these umpk-specific residues.

ACKNOWLEDGMENTS

We are grateful to G. Schulz for discussion and to K. Schappert and P. Walther for assistance with computers. P. L. was supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. Lacroute, F. (1968) J. Bacteriol. 95:824-842.
2. Jund, R. and Lacroute, F. (1970) J. Bacteriol. 102:607-615.
3. Heil, A., Muller, G., Noda, L.H., Pinder, T. Schirmer, R.H., Schirmer, I. and Von Zabern, I. (1974) Eur. J. Biochem. 43:131-144.

4. Schulz, G.E., Elzinga, M., Marx, F. and Schirmer, R.H. (1974) Nature **250**:120-124.
5. Von Zabern I., Wittman-Liebold, B.,Untucht-Grau, R., Schirmer, R.H., and Pai, E.F. (1976) Eur. J. Biochem. **68**:281-290.
6. Sachsenheimer, W. and Schulz, G.E. (1977) J. Mol. Biol. **114**:23-36.
7. Pai, E.F., Sachsenheimer, W., Schirmer, R.H. and Schulz, G.E. (1977) J. Mol. Biol. **114**:37-45.
8. Kuby, S.A., Palmieri, R.H., Frischat, A.H., Wu, L.H., Maland, L. and Manship, M. (1984) Biochemistry **23**:2392-2399.
9. Brune, M., Schumann, R. and Wittinghofer, A. (1985) Nucl. Acids Res. **13**:7139-7151.
10. Fry, D.C., Kuby, S.A. and Mildvan, A.S. (1985) Biochemistry **24**:4680-4694.
11. Tomasselli, A.G., Mast, E., Janes, W. and Schiltz, E. (1986) Eur. J. Biochem. **155**:111-119.
12. Kishi, F., Maruyama, M., Tanizawa, Y. and Nakazawa, A. (1986) J. Biol. Chem. **261**:2942-2945.
13. Egner, U., Tomasselli, A.G. and Schulz, G.E. (1987) J. Mol. Biol. **195**:649-658.
14. Saint-Girons, I., Gilles, A.-M., Margarita, D., Michelson, S., Monnot, M., Fermandjian, S., Danchin, A. and Barzu, O. (1987) J. Biol. Chem. **262**:622-629.
15. Liljelund, P. and Lacroute, F. (1986) Mol. Gen. Genet. **205**:74-81.
16. Ingraham, J.L. and Neuhard, J. (1972) J. Biol. Chem. **247**:6259-6265.
17. Ginther, C.L. and Ingraham, J.L. (1974) J. Bacteriol. **118**:1020-1026.
18. Waldron, C. and Lacroute, F. (1975) J. Bacteriol. **122**:855-865.
19. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. USA **77**:5201-5205.
20. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
21. Messing, J. (1983) Methods Enzymol. **101**:20-78.
22. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**:5463-5467.
23. Schulz, G.E. (1987) Cold Spring Harbor Symp. Quant. Biol. Vol. **LII**:429-439.
24. Bennetzen, J.F. and Hall, B.D. (1982) J. Biol. Chem. **257**:3026-3031.
25. Kammerer, B. Guyonvarch, A. and Hubert, J.C. (1984) J. Mol. Biol. **180**:239-250.
26. Blobel, G. (1980) Proc. Natl. Acad. Sci. USA **77**:1496-1500.
27. Nagy, M. Laporte, J. Penverne, B. and Herve, G. (1982) J. Cell Biol. **92**:790-794.
28. Sentenac, A. (1985) CRC Critical Reviews in Biochemistry, **18**:31-90.
29. Dreusicke, D. and Schulz, G.E. (1986) FEBS Lett. **208**:301-304.
30. Dreusicke, D. and Schulz, G.E. (1988) J. Mol. Biol. **203**:1021-1028.
31. Hamada, M., Palmieri, R.H., Russell, G.A. and Kuby, S.A. (1979) Arch. Biochem. Biophys. **195**:155-177.