The Primary Structure of the Aspartate Transcarbamylase Region of the URA2 Gene Product in *Saccharomyces cerevisiae*

FEATURES INVOLVED IN ACTIVITY AND NUCLEAR LOCALIZATION*

(Received for publication, November 28, 1988)

Marie Nagy[‡][§], Martine Le Gouar[‡], Serge Potier[¶], Jean-Luc Souciet[¶], and Guy Hervé[‡]

From the ‡Laboratoire d'Enzymologie, Centre National de la Recherche Scientifique, 91198 Gif-sur-Yvette Cedex, France and the ¶Laboratoire de Génétique Physiologique, Institut de Biologie Moléculaire et Cellulaire du Centre National de la Recherche Scientifique, 15 rue René Descartes, 67084 Strasbourg, France

The yeast URA2 locus encodes a multifunctional protein which possesses the carbamylphosphate synthetase and aspartate transcarbamylase activities and which catalyzes the first two reactions of the pyrimidine pathway. We report here the nucleotide sequence of the central and the 3' region of this locus. The latter encodes that part of the multifunctional protein which has the aspartate transcarbamylase activity. The deduced amino acid sequence shows a high degree of homology with the known aspartate transcarbamylases of various organisms from Escherichia coli to mammals. The amino acid residues that have been shown to be involved in the catalytic site of the E. coli enzyme are all conserved suggesting that, in the more complex structure of the yeast protein, the catalytic sites are also located at subunit interfaces. There is also an important conservation of the amino acid pairs that, in E. coli, are implicated in intra- and interchain interactions. As well as the oligomeric structure suggested by these two features, the three-dimensional structure of the yeast enzyme must also be organized to account for the channeling of carbamylphosphate, from the carbamylphosphate synthetase catalytic site to that of aspartate transcarbamylase, and for the concomitant feedback inhibition of the two activities by the end product UTP. The URA2 gene product was shown to be localized in the nucleus. With the aim of identifying the regions that may be involved in this transport, we have determined by electron microscopy the subcellular distribution of aspartate transcarbamylase in three strains expressing different fragments of the URA2 locus. In the first strain the protein lacks 190 residues at the N terminus, but accumulates normally in the nucleus. In the second strain the protein lacks 382 residues in the central part and seems impaired in the nuclear transport process. In the third strain the 476-residue protein encoded by the 3' region of URA2 locus and catalyzing the aspartate transcarbamylase reaction is able by itself to migrate to and accumulate in the nucleus. This suggests that two regions are involved in the nuclear accumulation. On the basis of their conservation in analogous proteins of other eukaryotes and their similarity to sequences already identified as nuclear location signals, a sequence in the central region of the protein and two short sequences in the C-terminal region are good candidates for the nuclear location signal involved in the targeting of the URA2 product.

In Saccharomyces cerevisiae, the first two reactions of the pyrimidine pathway are catalyzed by a multifunctional protein which possesses the carbamylphosphate synthetase and aspartate transcarbamylase (ATCase)¹ activities. The carbamylphosphate synthetase (including the structural feature responsible for the glutamine amidotransferase activity) and the ATCase activities are carried respectively by the Nterminal and the C-terminal domains of this single polypeptide chain which has a M_r of 244,000 (Souciet et al., 1989). Sucrose gradient centrifugation experiments suggest that the native form of this multienzyme protein in oligomeric (Denis-Duphil et al., 1981). As in other organisms, S. cerevisiae uracilcarbamylphosphate synthetase exhibits a complex saturation curve for ATP (Aitken et al., 1975; Belkaïd et al., 1988), but no detectable cooperativity in the utilization of the two other substrates. In contrast to the Escherichia coli enzyme, the yeast ATCase does not show any cooperative phenomena. Both activities of the multienzymatic protein are feedbackinhibited by the end product UTP (Kaplan et al., 1967; Aitken et al., 1975). In addition, channeling of carbamylphosphate from the catalytic site of carbamylphosphate synthetase toward the catalytic site of ATCase, where it is used as a substrate has been demonstrated (Aitken et al., 1975; Belkaïd et al., 1988). In view of these properties, it was of interest to determine the nucleotide sequence and to compare the structure-function relationships of the deduced primary structure with the corresponding nonassociated enzymes present in other organisms, especially E. coli in which ATCase structure and function are known in great detail.

It has been previously reported that, in yeast, the carbamylphosphate synthetase-ATCase multienzymatic protein is located in the nucleus (Nagy *et al.*, 1982), as is the case with other eukaryotic organisms, such as *Neurospora crassa* (Bernhardt and Davis, 1972), *Drosophila melanogaster*,² and mammals (Chaparian and Evans, 1987). Many proteins of high molecular weight are known to be localized in the nucleus at a rate of translocation much faster than that predicted for free diffusion through the nuclear pores. Consequently, some

^{*} Sequence data treatments were analyzed using the computer facilities at CITI1, Paris, with the help of the French Ministère de la Recherche et de l'Enseignement Supérieur. 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 GenBankTM/EMBL Data Bank with accession number(s) J04711.

[§] To whom correspondence should be addressed.

¹ The abbreviations used are: ATCase, aspartate transcarbamylase; NLS, nuclear location signals.

² B. P. Jarry, personal communication.

specific feature of mature nuclear proteins, designated as "nuclear location signals" (NLS) are believed to account for their rapid translocation and accumulation in the nucleus. Several short karyophilic sequences responsible for the nuclear location of proteins and able to target non-nuclear proteins to the nucleus were identified (for review, see Smith et al., 1985; Dingwall and Laskey, 1986). Most of them concern viral proteins, the best known being that of the simian virus large T antigen (Lanford and Butel, 1984; Kalderon et al., 1984a, 1984b) which consists of a stretch of basic amino acids, PKKKRKV. Similar sequences have been found in a variety of nuclear proteins. However, other karyophilic sequences, unrelated to the SV40 large T antigen have been reported, especially those involved in nuclear targeting of yeast proteins. This is the case of $MAT\alpha^2$ protein (Hall et al., 1984), the GAL4 product (Silver et al., 1984), and the ribosomal L3 protein (Moreland et al., 1985). All three are able to direct cytoplasmic β -galactosidase into the nucleus. In order to determine whether the URA2 product possesses any karyophilic domains, we examined the nuclear accumulation of URA2 proteins deleted in different regions of this locus. The 476-amino acid part of the URA2 product which bears the ATCase activity is able to accumulate in the nucleus, an observation which provided an additional interest in the determination of the nucleotide sequence corresponding to this region.

MATERIALS AND METHODS

Strains

JLS-133—JLS-133 is the uracil auxotroph FL100 ura2-60 transformed for prototrophy by pJLS1 (Amp'ura2 ATCase) (Souciet *et al.*, 1982). This plasmid is derived from the pFL1 plasmid (Chevallier *et al.*, 1980) by insertion of the yeast URA2 fragment BamHI(b)-BamHI(c) (Fig. 1) at the BamHI restriction site of pBR322.

SP903-31 (or: URA2 ΔfD)—SP903-31 (or: URA2 ΔfD) is the α GRF-18 strain in which the nuclear URA2 gene has been replaced by a BglII(e)-BglII(f) deleted allele (Potier et al., 1987) (Fig. 1).

MN1—MN1, the uracil auxotroph FL100 trp1 ura2 (9-15-30) (Exinger and Lacroute, 1979) was transformed by the pMC2010 plasmid bearing the 2.2-kilobase pair ClaI(b)-BamHI(c) fragment of the URA2 locus (Fig. 1). The yeast DNA was inserted in the BamHI site of the vector polylinker and the ClaI restriction site of the bacterial *lacZ*. The transformants were selected by uracil complementation.

Medium and Cultures

The yeast were grown at 28 °C on minimal YNB medium containing 0.67% yeast nitrogen base (Difco) and 2% glucose. Cells were collected during the logarithmic growth phase at $A_{590 \text{ nm}} = 1.0-1.5$.

Subcellular Localization of ATCase by Electron Microscopy

All the steps necessary for the cytochemical staining of the proteins by the formation of lead phosphate deposits by ATCase activity (Gomori reaction) were as described by Nagy *et al.* (1982). The cells were examined with a Philips model EM410 electron microscope.

Determination of Aspartate Transcarbamylase Specific Activity

The activity was measured both in permeabilized cells (as described by Penverne and Hervé, 1983) and in protoplasts lysed in the assay mixture. ATCase was assayed as described by Perbal and Hervé (1972) and protein was estimated as described by Bradford (1976).

DNA Sequencing

The ClaI-BamHI(c) URA2 subfragment was subcloned in the M13mp19 vector. For sequencing the complementary strand, the 1.5-kilobase pair BgIII(f)-BgIII(g) URA2 subfragment was subcloned in M13mp19 and the orientation complementary to the ClaI-BamHI(c) fragment was selected. Deletions of varying sizes were generated from one end of the insert by the 3' to 5' exonuclease activity of T4 DNA polymerase as described by Dale *et al.* (1985), except that the prelim-



FIG. 1. Restriction map of the URA2 locus (Souciet *et al.*, **1982**) and the encoded wild type (W.T.) protein. Deletion mutations in the JLS133, SP903, and MN1 strains are indicated by *hatched bars*.

inary EcoRI digestion step of the recombinant phage was performed at 42 °C overnight.

The BglII(e)-BglII(f) subfragment cloned in M13mp19 was also sequenced in both orientations using synthetic oligonucleotides.

DNA sequences were determined by the method of Sanger et al. (1977).

Immunoblots

Cells were harvested during the exponential phase of growth, and 5×10^8 were suspended in 300 μ l of sodium dodecyl sulfate sample buffer (Laemmli, 1970) containing 2 mM phenylmethylsulfonyl fluoride and 1 µg each of pepstatin A, antipain, leupeptin, and chymostatin. Cells were broken by vortexing the suspension for 3 min in the presence of 300 mg of glassbeads. The homogenate was heated for 5 min at 100 °C and centrifuged for 10 min at $11,000 \times g$. The supernatant was subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis, 7% polyacrylamide Laemmli-type gel, 50 µl/lane (about 300-500 µg of protein) at 20 mA overnight. The transfer of the proteins to a nitrocellulose sheet was performed at 3 °C for 24 h at 40 V (Burnett, 1981). The nitrocellulose blot was washed for 10 min in 200 ml of Tris-buffered saline (10 mM Tris-HCl, pH 7.5, 0.5 M NaCl), incubated for 1 h at room temperature in Tris-saline containing 3% bovine serum albumin, and then for 2 h in a sealed bag with 20 ml of Tris-saline containing 1% bovine serum albumin and a polyclonal antiserum raised against a purified, native form of URA2 product.3 The nitrocellulose was washed briefly with water and then for 20 min in two changes of 200 ml of Tris-saline. The sheet was then incubated for 2 h in sealed bag with 20 ml of Tris-saline containing 1% bovine serum albumin and goat anti-rabbit IgG (Institut Pasteur, Paris) conjugated with horseradish peroxidase. The nitrocellulose sheet was then washed as described above and developed in 100 ml of Tris-saline to which was added 20 ml of methanol containing 60 mg of 4-chloro-1-naphtol and 60 µl of 30% hydrogen peroxide.

RESULTS

Comparison of Yeast ATCase Structure with That of ATCases from Other Organisms

In order to determine the primary sequence of yeast ATCase, to compare it with other known ATCases, and to look for putative NLS, the nucleotide sequence of the URA2 locus was determined from the BglII(e) site to the *stop* of the reading frame (see the restriction map in Fig. 1). The nucleotide and the deduced amino acid sequences are presented in Fig. 2.

Primary Structure—In Fig. 3 the primary structure of yeast ATCase is compared to those of the corresponding enzymes from Bacillus subtilis, E. coli (catalytic subunit), Dictyostelium

³ Unpublished results from G. Hervé Laboratory.

Yeast Aspartate Transcarbamylase

1268	Arg S AGA T	er H CT 1	^{he P} TC C (e)	ro Ph CA TT	e 1 CA	le TT	Ser TCA	Lys AAG	Val GTT	Val GTT	G1y GGC	Val GTC	Asn AAT	Leu CTG	Ile ATT	Glu GAA	Leu TIC	Ala GCA	Thr ACA	Lys AAC	Ala GCC	Ile ATA	Met ATC	Gly GCT	Leu TTG	Pro CCT	Leu TTG	Thr ACG	Pro CCT	Tyr TAT	Pro CCT	Val GTT	Glu GAA	Lys AAA	Leu TTA	Pro CCA	Asp GAT	Asp CAT	Tyr TAT
1307	Val A	la V	al L	ys Va	1 P	ro	G1n	Phe	Ser	Phe	Pro	Arg	l.eu	Ala	Gly	Ala	Asp	Pro	Val	Leu	G1y	Val	Glu	Met	Ala	Ser	Thr	Gly	Glu	Val	Ala	Thr	Phe	G1y	His	Ser	Lys	Tyr	Glu
	GTC G	CC (TT A	AA GT	A C	CA	CAA	TTC	TCT	TTC	CCA	CGT	TTA	GCA	GGA	GCT	GAT	CCA	GTC	TTA	GGT	GTT	GAA	ATG	GCC	TCT	ACT	GGT	GAA	GTC	GCT	ACT	TTT	GGC	CAC	TCG	AAG	TAT	GAA
1346	Ala T	yr I	eu L	ys Se	r L	eu :	Leu	Ala	Thr	Cly	Phe	Lys	Leu	Pro	Lys	Lys	Asn	Ile	Leu	Leu	Ser	Ile	G1y	Ser	Tyr	Lys	Glu	Lys	Gln	Glu	Leu	Leu	Ser	Ser	Val	Gln	Lys	Leu	Tyr
	GCA T	AC J	TA A	AG TC	T T	TG :	TTG	GCA	ACC	GGC	TTC	AAA	CTT	CCA	AAG	AAC	AAT	ATT	TTA	TTG	TCT	ATT	GGT	TCT	TAC	AAG	GAA	AAA	CAA	GAA	TTG	CTT	TCT	TCC	GTA	CAA	AAA	CTA	TAC
1385	Asn M AAC A	et (IG (GA T	yr Ly AT AA	s L A T	eu TA 1	Phe TTT	Ala GCA	Thr ACA	Ser TCA	Gly GGT	Thr ACT	Ala GCT	Asp GAC	Phe TTT	Leu TTA	Ser TCT	Glu GAA	His CAT	G1y GGC	lle ATT	Ala GCC	Val GTC	Gln CAA	Tyr TAT	Leu CTG	Ser AGT	Leu TTA	Asn AAC	Lys AAG	Asp GAT	Asp CAT	Asp GAT	Asp GAT	G1n CAA	Lys AAA	Ser TCA	Glu GAA	Tyr TAC
1424	Ser L	eu 1	hr G	ln Hi	s L	eu .	Ala	Asn	Asn	Glu	Ile	Asp	Leu	Tyr	Jle	Asn	Leu	Pro	Ser	Ala	Asr	Arg	Phe	Arg	Arg	Pro	Ala	Ser	Tyr	Val	Ser	Lys	Gly	Tyr	Lys	Thr	Arg	Arg	Leu
	TCA C	FT #	CT C	AA CA	T T	TG (GCT	AAT	AAT	GAA	ATT	GAC	CTT	TAC	ATC	AAC	TTG	CCT	TCT	GCC	AAC	AGG	TTC	CGT	CGT	CCT	CCA	TCC	TAT	GTT	TCA	AAG	GGG	TAT	AAA	ACA	CGT	CGT	TTG
1463	Ala Va GCT G	al / FC (AT T	yr Se AT TC	r V. G G	al I TT (Pro CCG	Leu TTG	Val GTT	Thr ACT	Asn AAC	Val GTT	Lys A AA	Cys TGT	Ala GCA	Lys AAA	Leu TTG	Leu TTG	Ile ATT	Glu GAA	Ala GCC	lle ATT	Ser TCA	Arg AGA	Asn AAT	11e ATC	Thr ACT	Leu T TA	Asp GAT	Val GTT	Ser TCT	Glu GAA	Arg CGT	Asp GAT	Ala GCA	G1n CAA	Thr ACT	Ser TCC	His CAC
1502	Arg T	hr 1	le T	hr Le	u P	ro (G1y	Leu	Ile	Asn	Ile	Ala	Thr	Tyr	Val	Pro	Asn	Ala	Ser	His	Val	Ile	Lys	G1y	Pro	Ala	Glu	Leu	Lys	G1u	Thr	Thr	Arg	Leu	Phe	Leu	Glu	Ser	G1y
	AGA A	CT /	TT A	CC TT	A C	CT (GGT	TTA	ATC	AAT	ATC	GCA	ACT	TAT	GTT	CCG	AAT	GCA	TCC	CAT	GTT	ATC	AAA	GGC	CCA	GCT	GAA	CTG	AAG	GAG	ACC	ACA	CCT	CTA	TTT	TTC	GAA	TCC	GGT
1541	Phe T	hr 1	yr C	ys G1	n L	eu l	Met	Pro	Arg	Ser	Ile	Ser	Gly	Pro	Val	Ile	Thr	Asp	Val	Ala	Ser	Leu	Lys	Ala	Ala	Asn	Ser	Val	Ser	Gln	Asp	Ser	Ser	Tyr	Thr	Asp	Phe	Ser	Phe
	TTT A	CA 1	AC T	GT CA	A T	TG l	ATG	CCC	ACA	TCC	ATC	AGC	G GA	CCT	GTT	ATT	ACT	GAC	GTT	GCA	TCC	TTG	A A G	GCC	GCA	AAC	TCT	GTT	TCC	CAA	GAT	TCA	TCT	TAC	ACT	GAC	TTT	TCT	TTC
1580	Thr M ACT A	et A TG (la G	ly Th GT AC	r A A G	la I CA (His CAC	Lys AAG	Ala GCT	His CAT	Ser AGT	G1y GGT	Thr ACG	Gln C AA	Ala GCT	Ala GCT	Ser AGC	Lys AAA	Val GTC	Thr ACA	Ala GCA	Leu TTG	Phe TTT	Leu TTG	Pro CCC	Leu CTA	Arg CGI	Glu GAA	Leu TTA	Lys AAG	Asn AAC	Lys AAG	Ile ATC	Thr ACA	Ala GCA	Val GTA	Ala GCT	G1u GAA	Leu CTT
1619	Leu A:	sn (In T	rp Pr	o Ti	hr (Glu	Lys	Gln	Val	Ile	Ala	Glu	Ala	Lys	Thr	Ala	Asp	Leu	Ala	Ser	Val	Leu	Leu	Leu	Thr	Ser	Leu	G1n	Asn	Arg	Ser	Ile	His	Ile	Thr	Gly	Val	Ser
	CTG A	AT (AG T	GG CC	A A	CT (GAA	AAG	CAA	GTG	ATT	GCA	GAA	GCT	AAA	ACT	GCA	GAT	TTG	GCA	TCG	GTT	TTA	TTG	TTA	ACC	TCC	CTT	C AA	AAC	AGA	TCT	ATT	CAC	ATT	ACT	GGT	CTT	TCG
1658	Asn Lj	ys (lu A	sp Le	u A	la l	Leu	Ile	Met	Thr	Val	Lys	Ala	Lys	Asp	Pro	Arg	Val	Thr	Cys	Asp	Val	Asn	Ile	Tyr	Ser	Leu	Phe	lle	Ala	G1th	Asp	Asp	fyr	Pro	Glu	Ala	Val	Phe
	AAT A	AG (AA G	AT TT	A G	CT 1	ITG	ATC	ATG	ACG	GTC	AAG	GCA	AAA	GAC	CCT	AGA	GTG	ACT	TGT	GAT	GTC	AAT	ATT	TAT	TCT	TTG	TTT	ATT	GCC	CAA	GAT	GAC	TAT	CCG	GAG	GCA	GTT	TTC
1697	Leu P	ro 1	hr L	ys Gl	u A	sp (G1n	Glu	Phe	Phe	Trp	Asn	Asn	Leu	Asp	Ser	Ile	Asp	Ala	Phe	Ser	Val	Gly	Ala	Leu	Pro	Val	Ala	Leu	Ala	Asn	Val	Thr	Gly	Asn	Lys	Val	Asp	Val
	TTG C	77 /	CT A	AG GA	A G	AT (C AA	GAG	TTT	TTC	TGG	AAC	AAC	CTT	GAT	AGT	ATC	GAT	GCT	TTC	TCT	GTC	GGT	GCT	CTT	CCT	GTC	GCC	TTG	GCA	AAT	GTC	ACA	GCT	AAC	AAG	GTT	GAT	GTT
1736	Gly M	et (Iy I	le Ly	s A	sp :	Ser	Leu	Pro	Leu	Leu	Leu	Ala	Ala	Val	Glu	Glu	G1y	Lys	Leu	Thr	Ile	Asp	Asp	Ile	Val	Leu	Arg	Leu	His	Asp	Asn	Pro	Ala	Lys	Ile	Phe	Asn	Ile
	GGT A	IG (GT A	IC AA	A C	AT :	TCA	TTA	CCA	CTA	TTG	TTG	GCT	GCT	GTT	GAA	GAA	GCT	AAA	TTA	ACC	ATT	GAT	GAT	ATC	GTC	CTT	CCT	CTG	CAT	GAC	AAT	CCC	GCT	AAA	ATT	TTC	AAC	ATC
1775	Pro T CCT A	hr (CT (AG G	sp Se AC TC	r V. AG	al ' TI (Val GTC	Glu GAA	Ile ATT	Asp GAT	Leu TTC	Asp GAT	Tyr TAT	Ser TCT	Phe TTC	Arg AGA	Arg CGT	Asn AAT	Lys AAG	Arg AGA	Trp TGG	Ser TCA	Pro CCA	Phe TTC	Asn AAC	Lys AAA	Asp GAT	Met ATG	Asn AAC	G1y G GT	Gly GGT	lle ATT	Glu GAG	Arg CGT	Val GTT	Val GTC	Tyr TAT	Asn AAT	Gly GGC
1814	Glu T	hr I	eu V	al Le	u S	er (Gly	Glu	Leu	Val	Ser	Pro	Giy	Ala	Lys	G1y	Lys	Cys	lle	Val	Àsn	Pro	Ser	Pro	Ala	Ser	Ile	Thr	Ala	Ser	Ala	Glu	Leu	G1n	Ser	Thr	Ser	Ala	Lys
	GAA A	CA I	TA G	FT TT	GA	GC (GGŤ	GAA	TTA	GTT	TCA	CCA	GGC	GCC	AAA	GGA	AAA	TGC	ATT	GTT	AAT	CCA	AGT	CCA	GCT	TCC	ATA	ACT	GCT	TCC	GCA	GAG	CTC	C AA	TCT	ACT	AGT	GCT	AAA
1853	Arg A	rg F	he S	er Il	e T	hr (Glu	Glu	Ala	Ile	Ala	Asp	Asn	Leu	Asp	Ala	Ala	Glu	Asp	Ala	lle	Pro	Glu	Gln	Pro	Leu	Glu	Gln	Lys	Leu	Met	Ser	Ser	Arg	Pro	Pro	Arg	Glu	Leu
	AGA A	GG 1	TC T	CG AT	CA	CG (GAA	GAA	GCA	ATC	GCT	GAT	AAT	TTA	GAT	GCC	GCG	GAG	GAT	GCA	ATT	CCA	GAA	CAG	CCT	TTG	GAA	CAA	AAA	TTG	ATG	TCT	TCA	AGG	CCA	CCA	AGA	GAA	CTT
1892	Val A	la H	ro G	ly Al	a I	le (Gln	Asn	Leu	Ile	Arg	Ser	Asn	Asn	Pro	Phe	Arg	Gly	Arg	His	lle	Leu	Ser	Ile	Lys	Gln	Phe	Lys	Arg	Ser	Asp	Phe	His	Val	Leu	Phe	Ala	Val	Ala
	GTT G	CT (CA G	GC GC	CA	TC (CAG	AAT	TTG	ATC	CGT	AGT	AAC	AAT	CCA	TTC	CGC	GCA	AGA	CAT	ATA	TTG	TCT	ATC	AAA	CAA	TTC	AAA	CCT	TCT	GAT	TTC	CAT	GTG	TTG	TTT	GCT	GTT	GCA
1931	Gln G	lu I	eu A	rg Al	a A	la '	Val	Ala	Arg	Glu	Gly	Val	Leu	Asp	Leu	Met	Lys	Gly	His	Val	Ile	Thr	Thr	Ile	Phe	Phe	Glu	Pro	Ser	Thr	Arg	Thr	Cys	Ser	Ser	Phe	Ile	Ala	Ala
	CAA G	AA (TA A	GG GC	A G	CT (GTC	GCA	AGA	GAA	GGT	GTC	TTA	GAT	TTA	ATG	AAA	GGC	CAC	GTT	ATT	ACT	ACA	ATT	TTC	TTT	GAA	CCA	TCT	ACT	CGT	ACT	IGT	TCT	TCA	TTC	ATT	GCT	GCT
1970	Met G ATG C	lu / AA (GT T	eu G1 TG GC	y G T G	ly . GT .	Arg AGA	lle ATT	Val GTA	Asn AAT	Val G T T	Asn AAT	Pro CCA	Leu TTG	Val GTG	Ser TCT	Ser TCT	Val GTC	Lys AAG	Lys AAA	G1y GGT	Glu GAA	Thr ACC	Leu CTT	G1n CAA	Asp GAT	Thr ACT	lle ATC	Arg AGA	Thr ACT	Leu TTG	Ala GCT	Cys TGT	Tyr TAC	Ser AGT	Asp GAT	Ala GCC	Ile ATT	Val GTC
2009	Met A	rg H	lis S	er Gl	u G	lu	Met	Ser	Val	His	Ile	Ala	Ala	Lys	Tyr	Ser	Pro	Val	Pro	Ile	Ile	Asn	Gly	Gly	Asn	Gly	Ser	Arg	Glu	His	Pro	Thr	Gln	Ala	Phe	Leu	Asp	Leu	Phe
	ATG C	GT (CAT T	CA GA	A G	AA	ATG	TCT	GTT	CAT	ATC	GCC	GCT	AAA	TAT	TCT	CCT	GTT	CCA	ATT	ATT	AAT	GGT	GGT	AAT	GGT	TCT	CGC	GAG	CAT	CCT	ACG	CAG	GCC	TTC	TTG	GAT	TTG	TTT
2048	Thr I	le /	rg G	lu G1	u I	le	Cly	Thr	Val	Asn	G1y	Ile	Thr	Val	Thr	Phe	Met	Gly	Asp	Leu	Lys	His	G1y	Arg	Thr	Val	His	Ser	Leu	Cys	Arg	Leu	Leu	Met	His	Tyr	Gln	Val	Arg
	ACG A	TT (GT G	AA GA	A A	TC (CCT	ACT	GTT	AAT	GGT	ATT	ACT	CTT	ACT	TTC	ATG	GGT	GAT	CTC	AAA	CAT	GCT	AGA	ACC	GTA	CAT	TCA	TTG	TGT	CGT	TTG	TTA	ATG	CAC	TAT	CAA	GTC	AGA
2087	Ile A	sn I	.eu V	al Se	r P	ro	Pro	Glu	Leu	Arg	Leu	Pro	Glu	Gly	Leu	Arg	Glu	Glu	Leu	Arg	Lys	Ala	G1y	Leu	Leu	G1y	Val	Glu	Ser	lle	Glu	Leu	Thr	Pro	His	Ile	Ile	Ser	Lys
	ATT A	AT (CTT G	TT TC	T C	CT	CCG	GAA	TTG	AGG	TTA	CCA	CAA	GGA	T TA	AGA	GAA	GAG	CTA	AGA	AAA	CCT	GGC	TTA	CTT	GCT	GTT	GAG	AGC	ATT	GAA	TTA	ACC	CCT	CAT	ATC	ATC	TCA	AAG
2126	Thr A	sp V	al L	eu Ty	r C	ys	Th≃	Arg	Val	Gln	Glu	Glu	Arg	Phe	Asn	Ser	Pro	Glu	G1u	Tyr	Ala	Arg	Leu	Lys	Asp	Thr	Tyr	Ile	Val	Asp	Asn	Lys	Ile	Leu	Ala	His	Ala	Lys	Glu
	ACC G	AT (TT C	TG TA	T T	Gl	.ĸCA	AGG	GTC	CAA	GAA	GAA	AGA	TTC	AAT	AGC	CCT	GAA	GAA	TAT	GCA	CGT	CTG	AAG	GAT	ACT	TAT	ATC	GTG	GAC	AAC	AAC	ATC	TTG	GCA	CAC	GCC	AAA	GAA
2165	Asn M	et A	Ala I	le Me	t H	is	Pro	Leu	Pro	Arg	Val	Asn	Glu	Ile	Lys	Glu	Glu	Val	Asp	Tyr	Asp	His	Arg	Ala	Ala	Tyr	Phe	Arg	Gln	Met	Lys	Tyr	Gly	Leu	Phe	Val	Arg	Met	Ala
	AAT A	TG (SCT A	TC AT	G C	AT	CCA	TTC	CCT	CGT	GTA	AAT	GAA	ATC	AAA	GAG	GAA	GTG	GAC	TAC	GAT	CAT	CGT	GCT	GCT	TAC	TTC	AGA	C AA	ATG	AAG	TAT	GGT	TTG	TTC	GTC	AGA	ATG	GCT

2204 Leu Leu Ala Met Val Met Cly Val Asp Met XXXX TTG TTG GCC ATG GTC ATG GCT GTT GAT ATG TGA

FIG. 2. Nucleotide sequence of the 3' part of URA2 locus starting from the restriction site Bg/II(e) (Fig. 1). The deduced amino acid sequence encoded by the 946 3'-terminal codons open reading frame is shown. The *numbers* on the *left side* correspond to the position of the amino acids in the polypeptide encoded by the open reading frame of the whole URA2 locus nucleotide sequence (Souciet *et al.*, 1989). The *ClaI* site used for the cloning of the *ClaI-Bam*HI(c) fragment expressed in MN1 strain as well as the *Bg/II(e)* and *Bg/II(f)* sites marking limits of the D region are bracketed, and the ATG codon assumed to initiate the translation of the *ClaI-Bam*HI fragment is indicated by an *arrow*. xxx, stop codon.

discoideum, D. melanogaster, and the known part of the hamster carbamylphosphate synthetase-aspartate carbamyltransferase-dihydroorotase protein. Alignments were first generated by the ALIGN computer program and then adjusted by eye to maximize homology within the ATCase family. It can be seen in Table I that all these proteins show a degree of similarity varying from 28 to 60%. In the case of hamster carbamylphosphate synthetase-aspartate carbamyltransferase-dihydroorotase protein these percentages were calculated on the basis of the partially known sequence, that is the Cterminal half. The most striking observation which emerges from these comparisons is the complete conservation in S.

8368

BSATC ECATC SCATC DDATC DMATC		• * • • • • • • • • • • • • • • • • • •	T E E I K D L L E T A Q E I R D D L N L V L A T A A K I R S D F H V L F A V A Q E I R K Q L H A L F G I A H E M K D H L N D I F N L A Q L I 20 30	• •
BSATC ECATC SCATC DDATC DMATC	* * R F S F E V A R L S F E T S C S S F I A A Q C S F T A A P C S F A A A 60	EKKLGMNVLN SMHRLGASVVG MERLGGRIVN MCRLGGSVVT MCRLGGSVVT MLRLGGRVIS 70	L D G T - S T S V - Q K G E F S D S A N T S L G K K G E V N P L V S - S V - K K G E M D N I T S - S V - K K G E 80	Image: Constraint of the sector of the se
BSATC ECATC SCATC DDATC DMATC CAD	L V S Q V A T E F S G - A A K Y S P - A I Q V A T T F - S Q S 120	V N - I P I L N A G D - N - V P V L N A G D V P I I N G G N - K K - P I I N A G D - R L - S M P A	G C G Q H P T Q S L L D L M G S N Q H P T Q S L L D L F G S R E H P T Q A F L D L F G V G E H P T Q A L L D V F C R R E H S H Q A L L D I F	• •
BSATC ECATC SCATC DDATC DMATC CAD	A E V L T R L T Q A L A K F C R L L M H Y V R L L A N Y P P A D P V E A C L L T Q Y 1	L – G A R V L F S G – F D G N R F Y F I A P (Q V – R I N L V S P (Q V – K I D Y V S P 5 C E P A V – M W R R (R V – S L R Y V A P 80		★ C N T F G T - Y V S M D E A V E S - S D V V M L L R D E K G I A W S L H S S I E E V M V E V D I L Y M T R C R K A G L L - G V - E S I E L T P H I I S K T D V L Y C T R N E K G I E Q K E Y T N I E S I L P T T N V L Y V T R V H Q R G V K Q L F A R D L K E C A A R H G C A L H D S H S R V A S R G T K Q E E F E S I E E A L P D T D V L Y M T R 210 220 210
BSATC ECATC SCATC DDATC DMATC CAD	* * I Q N E R H Q V Q K E R F L D V Q K E R F Q V Q K E R F Q E R F D I Q K E R F G	* * 2 S A V S Q E G Y L N 0 - P S E - Y A N N S P E E - Y A R 2 S I E E - Y A R 5 S T Q E - Y E A 240	• • • • • • • • • • • • • • • • • • •	• • • • • • • • • • • • • • • • • • •
BSATC ECATC SCATC DDATC DMATC	KSRIFKQ HAWYFQQ RAAYFRQ RAAYFRQ	M K N G V F I R M A A G N G I F A R Q A M K Y G L F V R M A M E N G L Y V R M S	• • VIQCALQTNVKR LLALVLNRDLVL LLAMVMGVDM LLALVPGAGV	

FIG. 3. Comparison of the amino acid sequences of ATCases of *B. subtilis* (BSATC) (Lerner and Switzer, 1986), *E. coli* (ECATC) (Schachman *et al.*, 1984), *S. cerevisiae* (SCATC), *D. discoideum* (DDATC) (Faure *et al.*, 1989), *D. melanogaster* (DMATC) (Freund and Jarry, 1987), and the available portion of Syrian hamster carbamylphosphate synthetase-aspartate carbamyltransferase-dihydroorotase (Shigesada *et al.*, 1985). Catalytic site residues (Krause *et al.*, 1987) are enclosed in *boxes*. The putative NLS are denoted by grey boxes. \star , identical in all proteins; \bullet , similar (conservative homology; Kanehisa *et al.*, 1984) in all proteins; -, gaps introduced for alignment. The ALIGN computer program was used to align these sequences for maximum homology.

TABLE I
Degrees of homology between the amino acid sequences of ATCases

from our tous or guittsms													
	BSª	EC	SC	DD	DM	CAD							
BS	-	41	38	41	28	31							
EC		-	42	44	32	45							
SC			-	55	32	50							
DD				-	36	60							
DM					-	51							

^a BS, B. subtilis; EC, E. coli; SC, S. cerevisiae; DD, D. diocoideum; DM, D. melanogaster; CAD, hamster carbamylphosphate synthetase-aspartate carbamyltransferase-dihydroorotase.

cerevisiae of the amino acid residues which are known from crystallographic data to constitute the catalytic site of the E. coli ATCase catalytic subunit (Krause et al., 1987; Volz et al., 1986). These are Ser-52, Thr-53, Arg-54, Thr-55, Ser-80, Lys-84, Arg-105, His-134, Arg-167, Arg-229, Glu-231, and Leu-267 which are indicated in the boxes on Fig. 3. The same is true for the other organisms except for D. melanogaster in which 2 of these residues are not conserved. One conservative change is observed for leucine 267 which is replaced by methionine in the hamster ATCase and by alanine in B. subtilis. It is interesting to note that Gln-137 and Pro-266 whose interaction with the amino group of aspartate was predicted on the basis of molecular modeling (Gouaux et al., 1987) are both conserved in all the ATCases compared in Fig. 3. Furthermore, Glu-50, Lys-164, Arg-234, Glu-239, and Tyr-240 which were shown in the E. coli enzyme to be important for the activity although they are not in direct contact with the substrates (Ladjimi and Kantrowitz, 1988; Kantrowitz and



and of S. cerevisiae (URA2) (algorithm of Levin et al., 1986) ATCases. S, β sheets; H, α -helix.

 TABLE II

 Residues involved in intra- and interchain interactions in E. coli catalytic subunits

 that are concerned in S. complete ATCase

Links	Interactions within the polar domain ^a	Interactions within the equatorial domain ^a	Interactions between polar and equatorial domains	Interchain interactions
Polar	Ser-11-Asp-141 Glu-50-Arg-105 Ser-52-Thr-55 Glu-86-Thr-91 Thr-87-Asp-90 Thr-136-Arg-296	Asp-278–Arg-269 Asp-278–Tyr-285	Gln-137–Gln-288 Asp-141–Arg-234 Asp-141–Gln-288	Gln-86–Arg-54 Asp-90–Arg-269 Asp-100–Arg-65 Glu-239–Lys-164
Nonpolar			Leu-142–Leu-172 Ile-145–Leu-176	Gly-85–Val-270 Ile-12–Ser-171 Leu-142–Leu-172 Ile-145–Leu-176

 a The catalytic chain of *E. coli* ATCase is described as made up of two structural domains, so-called polar (N terminus region) and equatorial (C terminus region).



FIG. 5. Immunoblots of wild type and truncated products of URA2 locus. Analysis was performed with lysates of the following strains: *lane 1*, MN1; *lane 2*, SP903-31; *lane 3*, wild type FL100; *lane 4*, JLS133.

TABLE III												
Specific ATCase activity of cells and protoplasts of different strains												
Specific ATCose estivity												

Ct	Specific A	ATCase activity ^a	
Strains	Cells	Protoplasts	
FL100	2.5	2.6	
GRF-18	2.4	2.2	
JLS-133	2.0	1.9	
SP903-31	1.2	0.9	
MN1	0.12	0.10	

^{*a*} Micromoles ureidosuccinate formed $mg^{-1} h^{-1}$.

Lipscomb, 1988) are also present in the yeast enzyme.

Secondary Structure—We used the algorithm for secondary structure determination of proteins based on sequence similarity (Levin *et al.*, 1986) to predict the secondary structure of yeast ATCase in comparison with the known crystallographic structure of the *E. coli* ATCase catalytic subunits (Kim *et al.*, 1987). The result presented in Fig. 4 shows that this predicted yeast ATCase structure is strikingly similar to that known for the *E. coli* enzyme, especially in terms of the α -helices which are all common. This similarity suggests that the two proteins might have a very close tertiary structure.

Tertiary and Quaternary Structures—The catalytic subunit of *E. coli* ATCase is a trimer of catalytic chains whose assembly involves both intra- and interchain interactions of ionic or hydrophobic nature (Honzatko *et al.*, 1982). Although the quaternary structure of the URA2 product is not known, it is interesting to examine the degree of conservation of the amino



FIG. 6. Localization of the AT-Case staining. Mutant strains: a, JLS133; b, SP903-31; c, MN1. Control strain: d, FL100 trp1 ura2 (9-15-30). \times 14,000.

FIG. 7. Schematic diagram of the polypeptide of Fig. 2 bearing the putative NLS. The 946 amino acid residues of the sequence in Fig. 2 are represented by the *stippled box*. The peptides encoded by the *D* and *MN1* fragments as well as the ATCase activity-bearing peptide are shown. The putative NLS are represented by a one-letter code in *boxes*, and their positions are presented by *black bars*. The *numbers* are those of the residues in Fig. 2.

acid side chains which, in *E. coli*, are involved in these interactions. As far as the intrachain interactions involved in the establishment of the tertiary structure of an *E. coli* catalytic chain are concerned, 13 pairs of amino acid side chains involved in polar and nonpolar interactions are conserved out of 34 (Table II). However, this number raises to 22 if the conservative changes are taken into account, giving a 65%conservation of interacting pairs. Thus it appears that the conservation of the residues involved in intrachain links is considerably higher than the homology calculated over the entire ATCase sequence (42%).

Regarding the catalytic-catalytic interchain links involved in the *E. coli* catalytic subunits, 8 pairs of amino acid side chains out of 13 are entirely conserved (Table II). The possible meaning of this feature will be discussed further.

Regions Involved in the Nuclear Accumulation of URA2 Product

The URA2 protein is known to be concentrated in the nucleus of *S. cerevisiae*. As a first step in the identification of

putative NLS, we have attempted to get some idea of the regions of this very large protein which could be involved in its nuclear targeting. Toward this end, we constructed strains deleted in different regions of URA2 locus and determined their intracellular location by electron microscopy using the *in situ* precipitation of orthophosphate ions liberated specifically by the ATCase reaction (Gomori staining).

Protein Lacking the Amino Terminus of the Carbamylphosphate Synthetase-ATCase Complex—The strain JLS-133 is FL100 ura2 transformed by the multicopy plasmid pJLS1 which carries the 6.4-kilobase pair fragment BamHI(b)-BamHI(c) of the URA2 locus. The protein expressed in this strain lacks 190 residues at the N terminus (Souciet *et al.*, 1987), having a predicted molecular mass of 224 kDa. This size was confirmed by the position of this protein on the immunoblot shown in Fig. 5 (*lane 4*). The ATCase activity in this strain is similar to that of the wild type FL100 (Table III) and thus allows the cytochemical staining. As shown in

	. 1	445	_	_	_	-	_	_			_		~			_	_	_
S. cerevis	:1 <i>ae</i>	R	F.	к	ĸ	5	A	5	Y	v	5	ĸ	G	Y	ĸ	т	R	ĸ
		199	•	*	*	★	•	*	•	•	×	•	×	×			×	×
D. discoid	leum	K	Y	R	R	Ρ	S	S	F	М	S	R	G	Y	S	L	R	R
- ansis cour	nuetain	244		Б	P	5							~		72		Б	7/
p28°1° 55V	protein	R	v	ĸ	к	Р	Ρ	-	-	-	-	ĸ	G	-	ĸ	н	R	ĸ
		*		\star	★	★	٠					\star	★		\star		\star	٠

FIG. 8. Comparison of the proposed NLS of the D region with the corresponding sequence of *D. melanogaster* rudimentary and *D. discoideum* Dd PYR1-3B products aligned for maximum homology as well as with the NLS identified in $p28^{sis}$ protein of simian sarcoma virus. The *numbers* above the 1st residue refer to the position in the amino acid sequence of the proteins. \star , identity; \bullet , conservative homology; -, gaps introduced for alignment.

Fig. 6a, the stain is confined in the nucleus, as was previously observed for the wild type enzyme (Nagy *et al.*, 1982). This result shows that the 190 residues at the N terminus of the native protein do not contain any determinant necessary for its nuclear localization.

Protein Lacking the Internal Region—The strain SP903-31 carries a deletion of the URA2 locus extending from BgIII(e) to BgIII(f) (zone D in Fig. 1). The protein encoded by this gene would have a molecular mass of 202 kDa, and this was confirmed by the immunoblot shown in Fig. 5 (lane 2). Its ATCase activity is about 50% that of the GRF18 wild type (Table III).

The Gomori stain in this strain has a more complex distribution: although it is still concentrated in the nucleus, a large part of it remains in the cytoplasm and an additional stain deposit is visible on the nuclear envelope (Fig. 6b). Such staining of the nuclear envelope was not seen in any of the other strains studied, including one which overproduced the wild type protein (Nagy *et al.*, 1982). In this last strain the ATCase activity was distributed equally between the nucleus and the cytoplasm, but showed no accumulation on the surface of the nucleus.

Protein Encoded by the 3' Region of the URA2 Locus, Corresponding to Aspartate Transcarbamylase Activity-The 2.2-kilobase pair fragment ClaI-BamHI(c) cloned in the pMC2010 vector restored the uracil prototrophy of the recipient ura2 strain. The first ATG on the ClaI-BamHI(c) fragment cloned in the MN1 strain is located 72 base pairs upstream from the ClaI restriction site (Figs. 1 and 2). As this codon has the favorable sequence context described by Kozak (1981), a purine at position -3 and a guanosine residue at position +4, it presumably initiates the translation of the ClaI-BamHI(c) fragment. Starting from this methionine residue, the MN1 polypeptide would have 477 amino acid residues and a molecular weight of 53,347 (Fig. 7); again this is compatible with the immunoblot result (Fig. 5, lane 1). The ATCase activity measured in MN1 cells was low (Table III), and therefore the Gomori staining was less intense than in the strains described above. It was nevertheless clearly localized in the nucleus (Fig. 6c), and its specificity was confirmed by the absence of staining in the recipient cells (Fig. 6d).

Intracellular Stability of the Truncated Proteins—A criticism often made of studies on protein compartmentation using truncated and/or hybrid proteins is the possibility of a differential degradation of these anormal proteins in one of the cellular compartments involved. Immunoblot analysis in Fig. 5 shows that the truncated proteins described here have the expected molecular weights and were therefore not modified by a proteolytic process. The absence of an uncontrolled proteolysis is also inferred from the stability of ATCase activity during the critical period of accelerated proteolysis, *i.e.* between the cell harvest and the transformation into protoplasts (Table III).

Examination of the Amino Acid Sequence for the Presence of Putative Nuclear Location Signals-The subcellular distribution of ATCase in the strains described above points to two regions involved in nuclear accumulation (Fig. 7). The first is the C-terminal ATCase region as expressed in the strain MN1. The sequence of this protein, which accumulates in the nucleus, contains several stretches which could be good candidates for the role of NLS in that they have 2 basic residues flanking 3 or 4 hydrophobic residues. Such peptides have been shown to be able to target proteins to the nucleus (Hall et al., 1984). Two of them (Fig. 7) are conserved in the eukaryotes which accumulate ATCase in their nucleus, but are absent in E. coli and B. subtilis (grey boxes in Fig. 3). The second region corresponds to the D fragment (Fig. 7) whose deletion seems to impair the transport of the URA2 product through the nuclear membrane. It contains a 17-residue sequence which is similar to the NLS identified in the nuclear form of the transforming protein p28sis of simian sarcoma virus (Lee et al., 1987). Again, this sequence is highly conserved in the corresponding region of D. melanogaster and D. discoideum (Fig. 8).

DISCUSSION

Structure-Function Relationship in Yeast ATCase-The results presented here show that all the residues involved in the catalytic function of E. coli ATCase are conserved in the yeast protein, indicating that the structural features required for ATCase activity remained unchanged between the E. coli polypeptide and the multifunctional yeast protein. This result is consistent with the notion that the multifunctional yeast enzyme is organized into independent functional domains, as demonstrated for the homologous hamster carbamylphosphate synthetase-aspartate carbamyltransferase-dihydroorotase protein by controlled proteolysis (Mally et al., 1981). It is well established that the E. coli ATCase catalytic sites are located at the interface between catalytic chains and involves amino acid residues which belong to the two partners (Wente and Schachman, 1987; Krause et al., 1987). The fact that all these residues are conserved in the yeast multifunctional protein indicates that in its oligomeric structure the ATCase catalytic sites should be formed in the same way. In addition, most of the amino acid pairs which are involved in the stabilization of the tertiary and quaternary structure of the catalytic subunit of E. coli are conserved in the URA2 product, suggesting that these residues play also a role in the establishment of the oligomeric structure of the yeast multifunctional enzyme. An additional constraint in the organization of the multifunctional system studied here lies in the channeling of carbamylphosphate from the carbamylphosphate synthetase catalytic site to that of ATCase where it is used as a substrate implying that the spatial organization should ensure a close proximity of these two catalytic sites. The organization must further account for the concomitant feedback inhibition of the two catalytic activities by UTP.

Nuclear Localization of URA2 Product—The two initial models proposed for nuclear accumulation of proteins were that of selective binding, which assumed the free diffusion of all proteins into the nucleus with selective retention of some of them (Feldherr and Ogburn, 1980), and a model assuming the selective transport of nuclear proteins across the nuclear envelope (De Robertis et al., 1978). Davey et al. (1985) proposed a combination of these two models and suggested that both mediated transport and selective binding could play a role in the process and could involve different regions of the protein. Actually, an increasing number of nuclear proteins have been shown to contain multiple NLS (Richardson et al., 1986; Welsh et al., 1986; Picard and Yamamoto, 1987; Hall and Johnson, 1987; Kleinschmidt and Seiter, 1988). In addition, it has been shown that the rate, but not the final nuclear protein concentration, increases with the number of peptide signals (Landford et al., 1986; Roberts et al., 1987; Goldfarb et al., 1986). Moreover, it has been demonstrated recently that the process of nuclear migration can be separated into two steps: a rapid, signal-dependent binding at the nuclear envelope, followed by a slower, energy-dependent translocation through nuclear pores (Richardson et al., 1988; Newmeyer and Forbes, 1988).

We have shown previously that the yeast carbamylphosphate synthetase-ATCase complex is able to accumulate in the nucleus. By means of the MN1 strain we show here that the information responsible for this localization is present in the C-terminal part of the multienzyme protein, and two sequences were noticed as candidate NLS (Fig. 7). It is probably significant that these two sequences are conserved in eukaryotes in which ATCase is located in the nucleus, but not in the prokaryotes *E. coli* and *B. subtilis* (Fig. 3). We believe that despite the relatively small size of the MN1 protein, its nuclear accumulation is unlikely to be accounted for by passive diffusion for the following reasons:

1) All the MN1 protein synthesized during the 3 h doubling time of this strain is concentrated in the nucleus, while model studies have shown that only 5% of a protein of M_r 45,000 can diffuse into the nucleus within 24 h (Bonner, 1978).

2) The results of sucrose gradient centrifugations suggest that the native URA2 product is an oligomer of molecular mass of about 1000 kDa (Denis-Suphil *et al.*, 1981). It is probable that the 53 kDa large MN1 protein exists also *in vivo* in an oligomeric form. It was shown by Aitken *et al.* (1973) that the isolated C-terminal region of the URA2 product forms oligomers of high molecular weight which retain the ATCase activity. Such an oligomer could not diffuse through the nuclear envelope. This conclusion is also supported by the complete conservation of the catalytic site amino acid residues as compared to the *E. coli* ATCase in which the catalytic activity requires subunit association.

3) We have shown previously that the URA2 protein escapes from isolated nuclei, indicating that it is not tightly bound to an intranuclear structure (Nagy *et al.*, 1982).

4) It has been demonstrated in several cases that the ability to bind DNA is not sufficient for the nuclear accumulation of proteins with altered NLS (Paucha *et al.*, 1985; Silver *et al.*, 1986).

The subcellular distribution of the protein truncated in the internal region shows that there is a partial accumulation of this protein in the nuclear envelope membranes. This may be due to a reduction in the rate of translocation through the nuclear envelope, and a 17-amino acid sequence is proposed whose absence could account for this phenomenon. This sequence is highly conserved in the two analogous multifunctional proteins of *D. melanogaster* and *D. discoideum* (Fig. 8). It is interesting that this sequence which is located at the C-end of the carbamylphosphate synthetase domain (Souciet *et al.*, 1989) is lacking in the evolutionary related arginine specific carbamylphosphate synthetase of yeast and rat which

are localized, respectively, in cytoplasm and mitochondria (Urrestarazu *et al.*, 1977; Nyunoya *et al.*, 1985). The fact that the large SP903-31 protein of M_r 202,000 is still able to migrate into the nucleus indicates that the potential signal encoded by the D fragment is partially dispensable and that another NLS is operating in this protein. Since it appears that the binding to the nuclear membrane and the translocation through the pores are two separable steps, the behavior of the SP903-31 strain suggests that this region would be rather involved in the second of these two steps.

Acknowledgments—We thank Yolande Kerjan for her advice on strain construction, Jean Garnier for help in the secondary structure prediction, Gérard Batelier for help in the ATCase sequence alignments, Moncef Ladjimi for stimulating discussion, Anthony Else for improving the manuscript, and Jocelyne Mauger for its editing.

REFERENCES

- Aitken, D. M., Bhatti, A. R., and Kaplan, J. G. (1973) Biochim. Biophys. Acta. 309, 50-57
- Aitken, D. M., Lue, P. F., and Kaplan, J. G. (1975) Can. J. Biochem. 53, 721–730
- Belkaïd, M., Penverne, B., and Hervé, G. (1988) Arch. Biochem. Biophys. 262, 171-180
- Bernhardt, S. A., and Davis, R. H. (1972) Proc. Natl. Acad. Sci. U. S. A. 69, 1868–1872
- Bonner, W. M. (1978) in *The Cell Nucleus* (Busch, H., ed) pp. 97-148, Academic Press, New York
- Bradford, M. M. (1976) Anal. Biochem. 72, 248-254
- Burnett, W. N. (1981) Anal. Biochem. 112, 195-203
- Chaparian, M., and Evans, D. R. (1987) Fed. Proc. 146, 2215
- Chevallier, M. R., Bloch, J. C., and Lacroute, F. (1980) Gene (Amst.) 11, 11-19
- Dale, R. M. K., McClure, B. A., and Houchins, J. P. (1985) Plasmid 13, 31–40
- Davey, J., Dimmock, N. J., and Colman, A. (1985) Cell 40, 667-675
- Denis-Duphil, M., Mathien-Shire, Y., and Hervé, G. (1981) J. Bacteriol. 148, 659–669
- De Robertis, E. M., Longthorne, R., and Gurdon, J. B. (1978) *Nature* **272**, 254–256
- Dingwall, C., and Laskey, R. A. (1986) Annu. Rev. Cell Biol. 2, 367– 390
- Exinger, F., and Lacroute, F. (1979) Mol. Gen. Genet. 173, 109-113
- Faure, M., Camonis J. H., and Jacquet, M. (1989) *Eur. J. Biochem.* **179**, 345-358
- Feldherr, C. M., and Ogburn, J. A. (1980) J. Cell Biol. 87, 589-593
- Freund, J. N., and Jarry, B. P. (1987) J. Mol. Biol. 193, 1-13
- Goldfarb, D. S., Gariepy, J., Schoolnik, G., and Kornberg, R. D. (1986) Nature **322**, 641-644
- Gouaux, J. E., Krause, K. L., and Lipscomb, W. N. (1987) Biochem. Biophys. Res. Commun. 142, 893-897
- Hall, M. N., and Johnson, A. D. (1987) Science 238, 1007-1012
- Hall, M. N., Hereford, L., and Herskowitz, I. (1984) Cell 36, 1057– 1065
- Honzatko, R. B., Crawford, J. L., Monaco, H. L., Ladner, J. E., Ewards, B. F. P., Evans, D. R., Warren, S. G., Wiley, D. C., Ladner, R. C., and Lipscomb, W. N. (1982) J. Mol. Biol. 160, 219-263
- Kalderon, D., Richardson, W. D., Markham, A. T., and Smith, A. E. (1984a) *Nature* **311**, 33–38
- Kalderon, D., Roberts, B. L., Richardson, W. D., and Smith, A. E. (1984b) Cell **39**, 499-509
- Kanehisa, M., Klein, P., Greif, P., and Delisi, C. (1984) Nucleic Acids Res. 12, 417–428
- Kantrowitz, E. R., and Lipscomb, W. N. (1988) Science 241, 669-674
- Kaplan, J. G., Duphil, M., and Lacroute, F. (1967) Arch. Biochem. Biophys 119, 541-551
- Kim, K. H., Pan, Z., Honzatko, R. B., Ke, H. M., and Lipscomb, W. N. (1987) J. Mol. Biol. 196, 853–875
- Kleinschmidt, J. A., and Seiter, A. (1988) EMBO J. 7, 1605-1614
- Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252
- Krause, K. L., Volz, K. W., and Lipscomb, W. N. (1987) J. Mol. Biol. 193, 527–553
- Ladjimi, M. M., and Kantrowitz, E. R. (1988) Biochemistry 27, 276– 283

- Laemmli, U. K. (1970) Nature 227, 680-685
- Lanford, R. E., and Butel, J. S. (1984) Cell 37, 801-813
- Lanford, R. E., Kanda, P., and Kennedy, R. C. (1986) Cell 46, 575-582
- Lee, B. A., Mahler, D. W., Hannink, M., and Donoghue, D. J. (1987) Mol. Cell. Biol. 7, 3527-3537
- Lerner, C. G., and Switzer, R. L. (1986) J. Biol. Chem. 261, 11156-11165
- Levin, J. M., Robson, B., and Garnier, J. (1986) FEBS Lett. 205, 303-308
- Mally, M. I., Grayson, D. R., and Evans, D. R. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 6647–6651
- Moreland, R. B., Nam, H. G., Hereford, L. M., and Fried, H. M. (1985) Proc. Natl. Acad. Sci. U. S. A. 82, 6561-6565
- Nagy, M., Laporte, J., Penverne, B., and Hervé, G. (1982) J. Cell Biol. 92, 790-794
- Newmeyer, D. D., and Forbes, D. J. (1988) Cell 52, 641-653
- Nyunoya, H., Broglie, K. E., Widgren, E. E., and Lusty, C. J. (1985) J. Biol. Chem. 260, 9346-9356
- Paucha, E., Kalderon, D., Richardson, D. W., Harvey, R. W., and Smith, A. E. (1985) EMBO J. 4, 3235-3240
- Penverne, B., and Hervé, G. (1983) Arch. Biochem. Biophys. 225, 562-575
- Perbal, B., and Hervé, G. (1972) J. Mol. Biol. 70, 511-529
- Picard, D., and Yamamoto, K. R. (1987) EMBO J. 6, 3333-3340
- Potier, S., Souciet, J. L., and Lacroute, F. (1987) Mol. Gen. Genet. 209, 283-289
- Richardson, W. D., Roberts, B. L., and Smith, A. E. (1986) Cell 44, 77 - 85
- Richardson, W. D., Mills, A. D., Dilworth, S. M., Laskey, R. A., and

- Dingwall, C. (1988) Cell 52, 655-664
- Roberts, B. L., Richardson, W. D., and Smith, A. E. (1987) Cell 50, 465 - 475
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 77, 5463-5467
- Schachman, H. K., Pauza, C. D., Navre, M., Karels, M. J., Wu, L.,
- and Yano, Y. R. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 115-119 Shigesada, K., Stark, G. R., Maley, J. A., Niswander, L. A., and
- Davidson, J. N. (1985) Mol. Cell. Biol. 5, 1735-1742 Silver, P. A., Keegan, L. P., and Ptashne, M. (1984) Proc. Natl. Acad.
- Sci. U. S. A. 81, 5951-5955
- Silver, P. A., Brent, R., and Ptashne, M. (1986) Mol. Cell. Biol. 6, 4763-4766
- Smith, A. E., Kalderon, D., Roberts, B. L., Colledge, W. H., Edge, M., Gillett, P., Markham, A., Paucha, E., and Richardson, W. D. (1985) Proc. R. Soc. Lond. B Biol. Sci. 266, 43-58
- Souciet, J. L., Hubert, J. C., and Lacroute, F. (1982) Mol. Gen. Genet. 186, 385-390
- Souciet, J. L., Potier, S., Hubert, J. C., and Lacroute, F. (1987) Mol. Gen. Genet. 207, 314-319
- Souciet, J. L., Nagy, M., Le Gouar, M., Lacroute, F., and Potier, S. (1989) Gene (Amst.), in press
- Urresterazu, L. A., Vissers, S., and Wiame, J.-M. (1977) Eur. J. Biochem. **79**, 473–481 Volz, K. W., Krause, K. L., and Lipscomb, W. N. (1986) Biochem.
- Biophys. Res. Commun. 136, 822-826
- Welsh, J. D., Swimmer, C., Cocke, T., and Shenk, T. (1986) Mol. Cell. Biol. 6, 2207-2212
- Wente, S. R., and Schachman, H. K. (1987) Proc. Natl. Acad. Sci. U. S. A. 84, 31-35