Curr Genet (1991) 19:255–260

Current Genetics

© Springer-Verlag 1991

Sequence analysis of the *ARG7* gene of *Schizosaccharomyces pombe* coding for argininosuccinate lyase

Expression of the gene in *Saccharomyces cerevisiae*

Roland Loppes [1], Reiner Michels [1], Isabelle Decroupette [1], and Bernard Joris [2]

[1] Laboratory of Molecular Genetics, Botany Institute, B22, and [2] Laboratory for Protein Engineering, Chemistry Institute, B6, University of Liège, Sart Tilman, B-4000 Liège, Belgium

Received January 3, 1991

Summary. The complete nucleotide sequence of the *ARG7* gene, coding for argininosuccinate lyase (EC 4.3.2.1), in the fission yeast (*Schizosaccharomyces pombe*) has been determined. It consists of an open reading frame of 461 codons. The deduced protein has a molecular weight of 51 200 Da. The gene is devoid of introns which is confirmed by the fact that it is expressed in *Escherichia coli* after spontaneous insertion of a bacterial sequence probably bearing a prokaryotic promoter. A perfect "TATA" box is found at –72 and the major transcription initiation site in *Saccharomyces cerevisiae* is located at –11 as shown by primer extension experiments. Comparison of the *S. pombe* lyase with related proteins from other organisms reveals an important degree of conservation except in the carboxyterminal part of the polypeptide. Additionally, a deletion removing 66 amino acids of the carboxy terminus yields an enzyme exhibiting some biological activity. A unique 1 500 b transcript was found in *S. cerevisiae* when the intact gene was present, but the deleted version of the gene gave rise to at least three transcripts of 1 800, 2 800 and 3 900 b.

Key words: *Schizosaccharomyces pombe* – *Saccharomyces cerevisiae* – Argininosuccinate lyase – Sequence

Introduction

Argininosuccinate lyase (ASL, EC 4.3.2.1) catalyses the cleavage of argininosuccinate into fumarate and arginine, the last step in the biosynthesis of this amino acid. The enzyme also plays a key role in the urea cycle leading to the removal of ingested nitrogen in the mammalian liver. It is highly analogous to the δ-crystallin found exclusively in reptiles and birds (Piatigorsky et al. 1988).

Argininosuccinate lyase has been purified from beef (Lusty and Ratner 1972) and human (O'Brien and Barr 1981) liver and from microorganisms such as *Saccharomyces cerevisiae* (Schweitzer 1982) and *Chlamydomonas reinhardtii* (Farrell and Overton 1987). In every case, the enzyme was shown to be a homotetramer with subunits of about 50 000 Da. In rat (Amaya et al. 1988) and man (O'Brien et al. 1986; Matuo et al. 1988) the amino acid sequence has been determined from cDNA clones. In yeast, the complete sequence (Beacham et al. 1984), and in *Chlamydomonas* a partial sequence (Debuchy et al. 1989), was obtained from clones obtained from genomic libraries. In *Escherichia coli*, only the first 151 bases of the gene are known (Parsot et al. 1988).

The *ARG4* gene of *S. cerevisiae*, coding for ASL, is not under a specific control system mediated by arginine but is subject to the general control of amino acid biosynthesis (Delforge et al. 1975; Messenguy and Dubois 1983). A deletion analysis of the *ARG4* promoter has recently been carried out (Thiry-Blaise and Loppes 1990). This study confirmed the role of the putative UAS and revealed another interesting sequence (a stretch of 14 dA residues lying between –124 and –137 from the initiation codon) probably acting as a constitutive promoter (Struhl 1985).

Nothing is known about the structure of the corresponding gene (*ARG7*) in the fission yeast *Schizosaccharomyces pombe*. This gene has been recently cloned (Remacle et al. 1988) by complementation of an *arg4* mutant of *S. cerevisiae*. In this study, we report the complete sequence of the *ARG7* gene and the mapping of the transcription initiation sites in *S. cerevisiae*. The argininosuccinate lyase of *S. pombe* displays a high degree of similarity to the corresponding *S. cerevisiae*, human and rat proteins and to the chicken lens structural protein δ-crystallin.

Materials and methods

Strains and media. Saccharomyces cerevisiae UL1 (*arg4 his3 leu2*; Remacle et al. 1988) was used for selecting arginine prototrophs. It was grown on YNB minimal medium with 50 mg/l of each appropriate supplement or on YPD rich medium. *Escherichia coli* C600 *argH* (lacking argininosuccinate lyase, ▲*argBH* ▲*pro-argF-lac argI thi* hsr_k hsd_k) was used for the complementation assay of the

Offprint requests to: R. Loppes

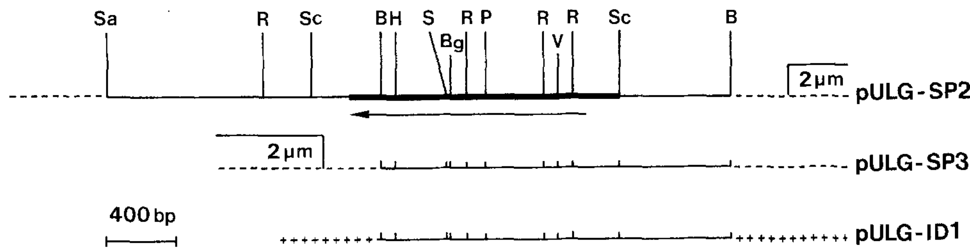

Fig. 1. Plasmids used in this study: pULG-SP2 (3.6 kb *Sau*3A fragment inserted at the YEp13 *Bam*HI site; one *Bam*HI site is regenerated), pULG-SP3 (2 kb *Bam*HI fragment in YEp13) and pULG-ID1 (2 kb *Bam*HI fragment in pUC19). The sequenced region is marked by *black bar.* The *arrow* shows the orientation of the gene in the three constructions. ——, insert; ----, YEp13; ++++, pUC19. *B, Bam*HI; *Bg, Bgl*II; *H, Hin*dIII; *P, Pvu*II; *R, Eco*RI; *S, Sac*I; *Sa, Sau*3A; *Sc, Sca*I; *V, Eco*RV

argH mutation on M9 medium supplemented with proline (50 mg/l), thiamine (1 mg/l) and citrulline (50 mg/l). C600 *pyrF* (▲*pro-argF-lac argI thi pyrF* hsr_k hsd_k) was used as a host for most plasmids. Both C600 strains were provided by M. Crabeel (V. U. Brussels). JM 105 (*thi rpsL endA sbcB15* hsr_k hsd_k ▲ *lac proAB* [F′ *traD36 proAB lacIq* ▲ *M15*]) served as host for M13 phages and plasmid pUC19.

Plasmids. The *S. pombe ARG7* gene is located on a 3.6 kb *Sau*3A fragment cloned at the *Bam*HI site of YEp13 to give pULG-SP2 (Remacle et al. 1988) (see Fig. 1). Plasmids pULG-SP3 and pULG-ID1 were obtained by cloning the 2 kb *Bam*HI-*Bam*HI fragment into YEp13 and pUC19 respectively.

DNA sequencing. DNA fragments were cloned into M13mp18 and M13mp19 and sequenced by the dideoxy chain termination method (Sanger et al. 1977) using the Klenow fragment of DNA polymerase I according to a procedure provided by Bethesda Research Laboratories, Gaithersburg, USA (Focus 9, 3, 1987) and a discontinuous buffer concentration gradient during electrophoresis (Biggin et al. 1983).

RNA preparations and Northern blots. Total *S. cerevisiae* RNA was prepared according to a procedure adapted from Nicolet et al. (1985). About 5×10^9 cells were collected during exponential growth in YNB medium and resuspended in 4 ml of cold 200 mM Tris-HCl pH 7.5, 500 mM NaCl, 1% SDS, 10 mM EDTA in a 30 ml Corex (Du Pont, Wilmington, USA) tube. Two milliliters of glass beads (0.5 mm diameter) and 4 ml phenol-chloroform were added and the cells were disrupted by fast vortex mixing. After centrifugation (15 min at 8000 *g*) the aqueous phase was extracted three times with phenol-chloroform, once with ether and the nucleic acids were precipitated overnight at −20 °C with ethanol, washed with 70% ethanol, dried and dissolved in 250 μl of water. About 0.5 mg RNA in 100 μl 50 mM Tris-HCl pH 7.8, 5 mM $MgCl_2$ were treated with 40 μg/ml RNAse-free DNAse for 30 min at 25 °C. After two extractions with phenol-chloroform, RNA was precipitated with ethanol, washed, dried and dissolved in 100 μl 10 mM Tris-HCl pH 8.0, 1 mM EDTA and kept at −20 °C. RNA samples (20 μg) were denatured with glyoxal (Sambrook et al. 1989) and fractionated by electrophoresis in 1.2% agarose gels in 10 mM sodium phosphate buffer pH 7.0 at 4 °C (3 V/cm) with continuous buffer recirculation. After treatment of the gel for 10 min in 5 mM NaOH, RNA was transferred to Zeta-probe (Bethesda Research Laboratories) by blotting with 5 mM NaOH as described by Vrati et al. (1987). The membranes were briefly rinsed in 2 × SSC (1 × SSC = 150 mM NaCl, 15 mM trisodium citrate), 0.1% SDS and kept at −20 °C. Membranes were prehybridized at 48 °C for 5 h in 50% formamide, 1.5 × SSPE (1 × SSPE = 180 mM NaCl, 10 mM NaH_2PO_4, 1 mM EDTA, pH 7.4), 1% SDS, 0.5% Blotto and 0.5 mg/ml denatured herring sperm DNA then hybridized overnight at 48 °C in the same solution containing 10% dextran sulfate and 10^5 cpm/cm^2 of the probe labelled by nick-translation (Rigby et al. 1977) and finally washed for 15 min in 2 × SSC at 25 °C, 15 min in 0.1 × SSC, 0.1% SDS at 25 °C and 30 min in 0.1 × SSC, 1% SDS at 60 °C.

Primer extension. Primer extension reactions were carried out using total RNA. An oligonucleotide (5′-CATCAGTGGATCAGTAGCT-3′) complementary to bases +48 to +66 of the *ARG7* gene was labelled with T4 polynucleotide kinase and γ-[^{32}P]ATP and purified by chromatography on Sephadex G-10 (Pharmacia, Uppsala, Sweden) equilibrated with 50 mM ammonium bicarbonate pH 7.8 (Zoller and Smith 1982). Twenty-five μg of RNA were resuspended in 7.5 μl of 250 mM KCl, 10 mM Tris-HCl pH 8.3 and annealed with 2 p mole of labeled primer (1.6×10^7 cpm/p mole). The extension reaction was then allowed to proceed for 1 h at 37 °C in a total volume of 25 μl (50 mM Tris-HCl pH 7.5, 10 mM DTT, 3 mM $MgCl_2$, 75 mM KCl, 100 μg/ml BSA, 500 μM each of dATP, dTTP, dGTP and dCTP, 50 μg/ml actinomycin D and 500 units M-MLV reverse transcriptase). The elongated cDNA was precipitated by ethanol, resuspended in sequencing loading buffer and denatured for 3 min at 100 °C before loading on the gel.

Computer analysis. The GCG sequence analysis system was utilized (Devereux et al. 1984). In this package, the "distances" algorithm requires aligned sequences and compares them residue by residue. The final score represents the number of matches divided by the length of the shortest sequence excluding the gaps. In the present analysis, a match was scored when the substitution had a value larger than, or equal to 1.5 in the table of Gribskov and Burgess (1986) which is a normalized form of the Dayhoff (1979) table.

Results

Location of the ARG7 gene within the cloned fragment

The *ARG7* gene of *S. pombe* has previously been shown to be localized on a 3.6 kb *Sau*3A fragment and on a 2 kb *Bam*HI fragment (Remacle et al. 1988). When these fragments were subcloned into YEp13, the resulting plasmids (pULG-SP2 and pULG-SP3 respectively) complemented an *arg4* mutant of *S. cerevisiae* lacking argininosuccinate lyase activity (Fig. 1). The colonies harbouring pULG-SP3, however, grew much more slowly than those harbouring pULG-SP2, which suggested that the gene was slightly altered in pULG-SP3.

The position of the gene in the 3.6 kb fragment and the direction of transcription were indirectly inferred from the study of its expression in *E. coli* C600 *argH.* This strain, harbouring pULG-SP2, was strictly dependent on arginine. However, colonies appeared after 3 days at 37 °C on plates lacking arginine at a frequency of about 1×10^{-8}. The plasmid (pULG-SP2-3) extracted from one of these clones was shown to transform C600 *argH* to arginine independence indicating that this property was specified by the plasmid rather than by the host genome. Restriction of pULG-SP2-3 and pULG-SP2 with various combinations of endonucleases revealed the presence in pULG-SP2-3 of a 1.2 kb insertion in the *Bam*HI-*Bam*HI

```
                                                                  -181
                                                        AGT ACTTGCTATC
                                                                  -121
CACGGATGTT ATTTTGACCC ATATTCGTAT AGTGTTATCT AACTAGAGAA AAAGCTAAAA
                                                                   -61
TGATAAGTGC GAATATCTGA GGAAGAAAAA GTCATTTCTT CACGCTATTA TATAAGTAAA
                                                                    -1
GGAAAATTAG ACCATCATTT GTAGACTGAA AATAATATTG AATAAAAGCA TCTTGGCACT
                                                        * *
   1  Met Ala Glu Lys Ser Ser Lys Lys Leu Trp Gly Gly Arg Phe Ser Gly    16
   1  ATG GCA GAA AAA TCA AGC AAA AAA CTA TGG GGA GGT AGA TTT TCA GGA    48

  17  Ala Thr Asp Pro Leu Met Ala Glu Phe Asn Lys Ser Ile Tyr Ser Gly    32
  49  GCT ACT GAT CCA CTG ATG GCA GAA TTC AAC AAA TCC ATC TAT AGT GGA    96

  33  Lys Glu Met Cys Glu Glu Asp Val Ile Gly Ser Met Ala Tyr Ala Lys    48
  97  AAG GAA ATG TGC GAA GAA GAT GTT ATT GGT TCC ATG GCG TAC GCA AAA   144

  49  Ala Leu Cys Gln Lys Asn Val Ile Ser Glu Glu Glu Leu Asn Ser Ile    64
 145  GCC TTG TGC CAG AAA AAT GTG ATA TCT GAA GAA GAG CTG AAT AGC ATC   192

  65  Leu Lys Gly Leu Glu Gln Ile Gln Arg Glu Trp Asn Ser Gly Gln Phe    80
 193  CTA AAA GGA TTG GAA CAA ATT CAA AGA GAA TGG AAT TCG GGT CAA TTC   240

  81  Val Leu Glu Pro Ser Asp Glu Asp Val His Thr Ala Asn Glu Arg Arg    96
 241  GTT TTG GAA CCA TCC GAC GAA GAT GTT CAC ACA GCA AAC GAG CGC CGA   288

  97  Leu Thr Glu Ile Ile Gly Asp Val Ala Gly Lys Leu His Thr Gly Arg   112
 289  TTA ACT GAG ATA ATC GGT GAT GTT GCT GGC AAG CTA CAT ACT GGC AGA   336

 113  Ser Arg Asn Asp Gln Val Thr Thr Asp Leu Arg Leu Trp Leu Cys Arg   128
 337  AGT CGT AAT GAC CAA GTT ACC ACC GAT TTG CGT TTA TGG CTA TGC AGA   384

 129  Lys Ile Lys Glu Val Glu Val Tyr Val Ile Asn Leu Leu Lys Val Phe   144
 385  AAA ATC AAA GAG GTT GAA GTC TAT GTC ATT AAC TTG CTT AAA GTT TTT   432

 145  Thr Asn Arg Ala Glu Met Glu Ile Asp Val Ile Met Ser Gly Tyr Thr   160
 433  ACC AAC AGA GCT GAG ATG GAG ATT GAT GTA ATA ATG TCA GGT TAT ACG   480

 161  His Leu Gln Arg Ala Gln Pro Val Arg Trp Ser His Phe Leu Met Ser   176
 481  CAT TTA CAA AGG GCT CAG CCT GTT CGT TGG TCC CAT TTT CTC ATG TCT   528

 177  His Ala Leu Pro Leu Leu Gly Asp Leu Gly Arg Leu Arg Gln Leu Tyr   192
 529  CAC GCC TTG CCT TTA TTA GGT GAC CTT GGC AGA CTT CGT CAG CTG TAT   576

 193  Thr Arg Val Ser Gln Leu Thr Ala Gly Ala Gly Ala Leu Ala Gly Lys   208
 577  ACT CGT GTA AGT CAA CTT ACC GCT GGT GCT GGT GCT TTA GCT GGC AAA   624

 209  Pro Phe Asn Val Asp Arg Glu Phe Leu Pro Lys Glu Leu Gly Phe Glu   224
 625  CCT TTC AAC GTC GAT CGC GAG TTC CTT CCT AAA GAG CTT GGA TTC GAA   672

 225  Gly Ile Ile Met Asn Ser Met Asn Ala Val Gly Asp Arg Asp Phe Val   240
 673  GGC ATT ATC ATG AAT TCC ATG AAT GCT GTT GGT GAT CGT GAT TTT GTC   720

 241  Ile Glu Phe Met Phe Trp Ala Gly Met Val Met Leu His Ile Ser Arg   256
 721  ATC GAA TTT ATG TTT TGG GCA GGC ATG GTA ATG CTT CAC ATT TCT CGC   768

 257  Phe Ala Glu Asp Leu Ile Ile Tyr Ser Ser Ser Glu Phe Gly Phe Val   272
 769  TTT GCT GAA GAT CTT ATC ATA TAT TCG AGC TCG GAA TTT GGA TTC GTC   816

 273  Thr Leu Ser Asp Ala Tyr Ser Thr Gly Ser Ser Ile Met Pro Gln Lys   288
 817  ACA CTC TCC GAT GCG TAT TCT ACG GGA AGT AGT ATT ATG CCC CAA AAA   864

 289  Lys Asn Pro Asp Ser Leu Glu Leu Leu Arg Gly Lys Ser Gly Arg Val   304
 865  AAG AAC CCT GAT TCT TTA GAG CTA CTT CGG GGT AAG AGC GGT CGT GTT   912

 305  Leu Gly Asp Met Ile Gly Leu Met Ile Thr Val Lys Gly Thr Pro Thr   320
 913  TTA GGT GAT ATG ATT GGC CTC ATG ATA ACT GTT AAA GGC ACA CCT ACA   960

 321  Thr Tyr Asn Lys Asp Leu Gln Glu Asp Lys Glu Pro Leu Phe Asp Ala   336
 961  ACC TAT AAC AAA GAT TTG CAA GAA GAC AAG GAA CCA CTA TTT GAT GCC  1008

 337  Phe Lys Thr Val Ser Asp Ser Leu Gln Ile Leu Thr Gly Val Val Ser   352
1009  TTT AAG ACC GTC TCT GAC TCT TTG CAA ATT TTG ACT GGC GTT GTC TCA  1056

 353  Thr Leu Thr Ile Asn Pro Thr Lys Ile Ala Glu Ser Leu Thr Pro Asp   368
1057  ACC CTT ACC ATC AAT CCT ACA AAG ATT GCC GAA AGC TTG ACC CCC GAT  1104

 369  Leu Leu Ala Ser Thr Asp Leu Ala Glu Tyr Leu Val Arg Lys Gly Leu   384
1105  TTA CTA GCT AGC ACT GAT TTG GCT GAG TAT CTT GTT CGT AAA GGT CTT  1152
                                         BamH1
 385  Pro Phe Arg Gln Thr His His Ile Ser Gly Ser Ala Val Arg Met Ala   400
1153  CCA TTT CGC CAA ACT CAT CAT ATT TCG GGA TCC GCA GTT CGC ATG GCT  1200

 401  Glu Glu Arg Asn Thr Thr Leu Asp Lys Leu Ser Val Ser Asp Leu Gln   416
1201  GAA GAG AGA AAC ACT ACT CTC GAC AAG TTA TCA GTT TCT GAT TTG CAA  1248

 417  Ser Leu His Pro Leu Phe Asp Glu Asp Val Ser Lys Val Phe Asn Tyr   432
1249  TCA TTG CAT CCT TTA TTT GAT GAA GAC GTT TCA AAA GTC TTT AAT TAC  1296

 433  Glu Glu Ser Val Glu Lys Arg Cys Ser Ile Gly Gly Thr Ala Lys His   448
1297  GAA GAA AGT GTT GAA AAA AGA TGT TCA ATT GGT GGT ACT GCT AAG CAT  1344

 449  Cys Val Gln Asp Asn Arg Ala Tyr Thr Ile Ser Asn Ser ***           461
1345  TGT GTT CAA GAC AAT CGA GCA TAT ACG ATC AGC AAT TCT TAA ACT GTC  1492
```

Fig. 2. Nucleotide and deduced amino acid sequences of the *ARG7* gene. The nucleotide sequence is numbered from the first nucleotide of the presumed initiation codon. At the 5′ end, the *solid underline* indicates the putative "TATA" box. *Asterisks* at −11 and −10 indicate the major transcriptional start points. At the 3′ end, the *solid underline* shows the deleted amino acids in pULG-SP3

fragment, more precisely between the *Eco*RV and *Sca*I sites (data not shown). This observation suggested that the expression of *ARG7* in *E. coli* results from the insertion near the 5′ end of the gene of a bacterial IS element bearing a prokaryotic promoter. In this case, and considering that the *S. cerevisiae ARG4* (1 389 bp) and the *S. pombe ARG7* genes probably have similar sizes, the origin of transcription should be localized between *Sca*I and *Eco*RV and the direction of transcription should be from *Sca*I to *Eco*RV.

Nucleotide sequence of ARG7

The complete nucleotide sequence of *ARG7* is shown in Fig. 2. An open reading frame of 1 383 bp starts at an ATG codon located 193 bp downstream from the *Sca*I site and terminates at a TAA codon downstream from the *Bam*HI site. This sequence encodes a polypeptide of 461 amino acid residues. The translation initiation codon is the only one (see other ATG codons in the same reading frame at positions +64 and +103) for which the surrounding region (ACTATGG) is in good agreement with the *S. cerevisiae* A/GXXATGG consensus (Beacham et al. 1984). This choice seems also to be correct on the basis of amino acid comparisons with other eukaryotic genes coding for argininosuccinate lyase (see Fig. 3). The first stop codon (TAA) is found at position 1 384. The predicted molecular weight for the polypeptide chain is 51 200 Da. Within the 193 bp of the 5′ untranslated region, the percentage of A+T is 69% against 59% in the coding region. The high A+T content is common to a number of *S. cerevisiae* (Beacham et al. 1984) and *S. pombe* (Kikuchi et al. 1988; Szankosi et al. 1988; Russel and Hall 1982; Mc Leod et al. 1987) promoters. Note in particular a 19 bp region (−14 to −32) containing 18 (dA+dT) residues. A perfect TATA box (TATATAA) (Breatnach and Chambon 1981) is seen at position −72 in the *ARG7* sequence. The bias in codon usage is relatively low: only two (CCG, GGG) of the 61 amino acid coding triplets are not used at all and only small differences are found in the usage patterns between the two yeast species.

Amino acid sequence comparison

The complete or partial amino acid sequences of ASL and δ-crystallins are aligned in Fig. 3. Only short gaps were introduced to obtain a consensus sequence with 168 identities (36%) and six highly analogous stretches (boxed in Fig. 3). There is a remarkable lack of similarity in the carboxy-terminal region of the protein where only 16 amino acids are shared from residues 388 to 461. This part of the protein most probably plays a minor role in catalysis. This hypothesis is confirmed by the fact that pULG-SP3, in which the ASL gene is interrupted at the *Bam*HI site (Fig. 2), complements a *S. cerevisiae arg4* mutant.

The genetic distances between four aligned sequences have been determined (Table 1). A value approaching 1 reflects a close relationship at the evolutionary level. It can be seen that the two higher eukaryotes are very close to each other (0.8937) whereas the two yeast species are both far from each other (0.5466) and from rats and humans.

Hydrophobicity, charge plots and secondary structure predictions comparing *S. cerevisiae* and *S. pombe* enzymes show that the two proteins are very similar over most of their sequence (data not shown).

```
                  1                                                49
S. pombe          .MAEKSSKKL WGGRFSGATD PLMAEFNKSI YSGKEMCEED VIGSMAYAKA
S. cerevisiae     .MSD.GTQKL WGGRFTGETD PLMHLYNASL PYDYKMYKAD LEGTKVYTAG
Rat               .MASE.SGKL WGGRFAGSVD PTMDKFNSSI AYDRHLWNVD LQGSKAYSRG
Man               .MASE.SGKL WGGRFVGAVD PIMEKFNASI AYDRHLWEVD VQGSKAYSRG
Chicken Cd1       .MATE.GDKL LGGRFVGSTD PIMEILSSSI STEQRLTEVD IQASMAYAKA
Chicken Cd2       .MASE.GDKL WGGRFSGSTD PIMEMLNSSI ACDQRLSEVD IQGSMAYAKA
C. reinhardtii    AAPADNTKKL WGGRFTAKTD PLMEKFNESL PFDKRLWAED IK
E.coli            .MA......L WGGRFTQAAD QRFKQFNDSL RFDYRLAEQD IVGSVAWSKA
Beef                                                                G
Consensus         -M------KL -GGRF----D --------S- ---------D ----------

                  50                                               99
S. pombe          LCQKNVISEE ELNSILKGLE QIQREWNSGQ FVLEPSDEDV HTANERRLTE
S. cerevisiae     LQKLGLLTET ELAKIHEGLA EIKKEWDADK FVRHPNDEDI HTANERRLGE
Rat               LEKAGLLTKA EMQQILQGLD KVAEEWAQGI FKLYPNDEDI HTANERRLKE
Man               LEKAGLLTKA EMDQILHGLD KVAEEWAQGT FKLNSNDEDI HTANERRLKE
Chicken Cd1       LEKASILTKT ELEKILSGLE KISEESSKGV LVMTQSDEDI QTAIERRLKE
Chicken Cd2       LEKAGILTKT ELEKILSGLE KISEEWSKGV FVVKQSDEDI HTANERRLKE
C. reinhardtii                          VAEEWKAGA FVINAGDEDI HTANERRLTE
E.coli            LVTVGVL
Beef              LEKAGLLTK
Consensus         L--------- E---I--GL- ----E----- -----|DED- HTA-ERRL-E
                                                             *
                100                                              146
S. pombe          IIG.DVAGKL HTGRSRNDQV TTDLRLWLCR KIKEVEVY.. VINLLKVFTN
S. cerevisiae     LIGREIAGKV HTGRSRNDQV VTDLRIY.CR DIVNDTLFPA LKGLVEVLIK
Rat               LIG.EAAGKL HTGRSRNDQV VTDLRLWMRQ TYSKLSTF.. LKVLIEAMVD
Man               LIG.ATAGKL HTGRSRNDQV VTDLRLWMRQ TCSTLSG..L LWELIRTMVD
Chicken Cd1       LIG.DIAGKL QTGRSRNEQV VTDLKLLLKS SISVISTH.. LLQLIKTLVE
Chicken Cd2       LIG.DIAGKL HTGRSRNDQV VTDLKLLLKS SISVISTH.. LLQLIKTLVE
C. reinhardtii    LVG.AVGGKL HTGRSRNDQ. .......... .......... ..........

Consensus         --G---AGK- HTGRSRN-QV -TDL|----- ---------- ---L------
                             *
                147                                              196
S. pombe          RAEMEIDVIM SGYTHLQRAQ PVRWSHFLMS HALPLLGDLG RLRQLYTRVS
S. cerevisiae     RAEGEIDVLM PGYTHLQRAQ PIRWSHWLSS YATYFTEDYK RLGQILHRLN
Rat               RAEAECEVLF PGYTHLQRAQ PIRWSHWILS HAVALTRDLE RLKEVQKRIN
Man               RAEAERDVLF PGYTHLQRAQ PIRWSHWILS HAVALTRDSE RLLEVRKRIN
Chicken Cd1       RAAIEIDIIM PGYTHLQKAL PIRWSQFLLS HAVALTRDSE RLGEVKKRIT
Chicken Cd2       RAATEIDVIM PGYTHLQKAL PIRWSQFLLS HAVALIRDSE RLGEVKKRMS
C. reinhardtii
Consensus         RA--E----- |GYTHLQ-A- P-RWSH|--S -A-----D-- RL-----R--
                                *          *
                197                                              246
S. pombe          QLTAGAGALA GKPFNVDREF LPKELGFEGI IMNSMNAVGD RDFVIEFMFW
S. cerevisiae     QSPLGAGALA GHPYGIDREF LAEGLGFNSV IGNSLVAVSD RDFIVELMFW
Rat               VLPLGSGAIA GNPLGVDREF LCAELNFGAI TLNSMDATSE RDFVAEFLFW
Man               VLPLGSGAIA GNPLGVDREL LRAELNFGAI TLNSMDATSE RDFVAEFLFW
Chicken Cd1       VLPLGSGALA GNPLEIDREL LRSELDMTSI TLNSIDAISE RDFVVELISV
Chicken Cd2       VLPLGSGALA GNPLEIDREL LRSELDFASI SLNSMDAISE RDFVVELLSV
C. reinhardtii                                          VSD REVVIETVFA

Consensus         ----G-GA-A G-P---DRE- L---L----- --NS--A--- R-F--E----

                247                                              296
S. pombe          AGMVMLHISR FAEDLIIYSS SEFGFVTLSD AYSTGSSIMP QKKNPDSLEL
S. cerevisiae     GTLFMNHISR FAEDLIIYCT AEFGFIQLSD AYSTGSSLMP QKKNADSLEL
Rat               ASLCMTHLSR MAEDLILYGT KEFNFVQLSD AYSTGSSLMP QKKNPDSLEL
Man               RSLCMTHLSR MAEDLILYCT KEFSFVQLSD AYSTGSSLMP QKKNPDSLEL
Chicken Cd1       ATLLMIHLSK LAEDLIIFST TEFGFVTLSD AYSTGSSLLP QKKNPDSLEL
Chicken Cd2       ATLLMIHLSK LAEDLIIFST TEFGFVTLSD AYSTGSSLLP QKKNPDSLEL
C. reinhardtii    ASLLCVHLSR WAEDLIIYSS GPFGYVQCSD AYATGSSLMP QKKNPDALEL

Consensus         ----M|H-S- -AEDLI---- --F-----SD AYSTGSS--P QKKNPD-LEL
                       *
                297                                              346
S. pombe          LRGKSGRVLG DMIGLMITVK GTPTTYNKDL QEDKEPLFDA FKTVSDSLQI
S. cerevisiae     LRGKSGRVFG DLTGFLMSLK GIPSTYDKDM QEDKEPLFDC LTTVEHSMLI
Rat               IRSKARRVFG RCAGLLMTLK GLPSTYNKDL QEDKEAVFEV SDTMTAVLQV
Man               IRSKAGRVFG RCAGLLMTLK GLPSTYNKDL QEDKEAVFEV SDTMSAVLQV
Chicken Cd1       IRSKAGRVFG RLAAILMVLK GIPSTFSKDL QEDKEAVLDV VDTLTAVLQV
Chicken Cd2       IRSKAGRVFG RLAAVLMVLK GLPSTYNKDL QEDKEAVFDV VDTLTAVLQV
C. reinhardtii    I

Consensus         -R-K--RV-G| --------|K G-P----KD- QEDKE|---- --T-------

                347                                              396
S. pombe          LTGVVSTLTI NPTKIAESLT PDLLASTDLA EYLVRKGLPF RQTHHISGSA
S. cerevisiae     ATGVISTLTV NKEKMEAALT MDMLA.TDLA DYLVRKGVPF RETHHISGEC
Rat               ATGVISTLQI HRENMAQALS PDMLA.TDLA YYLVRKGMPF RQAHEASGKA
Man               ATGVISTLQI HQENMGQALS PDMLA.TDLA YYLVRKGMPF RQAHEASGKA
Chicken Cd1       ATGVISTLQV NKENMEKALT PELLS.TDLA LYLVRKGMPI RQAQTASGKA
Chicken Cd2       ATGVISTLQV NKENMEKALT PELLS.TDLA LYLVRKGMPF RQAHVASGKA
C. reinhardtii                           LS ADMLA.TDLA EYLVRKGVPF RETHHH

Consensus         -TGV-STL|- --------L- ---|L--TDLA -YLVRKG-P- R--H|-SG--
                                                                *
                397                                              445
S. pombe          VRMAEERN.T TLDKLSVSDL QSLHPLFDED VSKVFNYEES VEKRCSIGGT
S. cerevisiae     VATAERLGLS GIDKLTLEQY QKIDSRFGQD LFETFNFEQS VERRDATGGT
Rat               VVVAEMKGV. ALNQLSLQEL QTVSPLFSSD VNLVWDYSHS VEQYTALGGT
Man               VFMAETKGV. ALNQLSLQEL QTISPLFSGD VICVWDYGHS VEQYGALGGT
Chicken Cd1       VHLAETKGI. TINNLTLEDL KSISPLFASD VSQVFSVVNS VEQYTAVGGT
Chicken Cd2       VHLAETKGI. AINKLTLEDL KSISPLFASD VSQVFNIVNS VEQYTAVGGT
C. reinhardtii                                       VWDFNRS AEMRDTEGGT

Consensus         V--AE----- ----L----- ------F--D ---------S -E-----GGT

                446              461
S. pombe          AKHCVQDNRA YTISNS.... ...
S. cerevisiae     AKSAVLKQLD NLKSQLN... ...
Rat               AQSSVEWQIS QVRALLQM.. ...
Man               ARSSVDWQIR QVRALLQAQQ A..
Chicken Cd1       AKSSVTAQIE QLRELLKKQK EQA
Chicken Cd2       AKSSVTAQIE QLRELLKKQK EQA
C. reinhardtii    SKRSVLEQVQ KMRTYLAAEG QH.

Consensus         A---V----- ---------- ---
```

Fig. 3. Alignments of the amino acid sequences of argininosuccinate lyases and of chicken δ-1 and δ-2 crystallins. The human sequence is from Matuo et al. (1988). The numbering corresponds to the *S. pombe* sequence without gaps

Table 1. Comparison matrix of argininosuccinate lyases. The values were calculated on the basis of the "distances" algorithm and represent the fractions of identical residues (see Materials and methods)

Species	*S. pombe*	*S. cerevisiae*	Rat	Man
S. pombe	1.0000	0.5466	0.5141	0.5184
S. cerevisiae		1.0000	0.5401	0.5335
Rat			1.0000	0.8937
Man				1.0000

pH rate profiles and chemical modifications of beef ASL indicate that a histidine and a carboxylate group are essential for catalytic activity (Garrard et al. 1985). On the other hand, experiments with site-directed agents on the same enzyme suggest that a lysine residue (lysine 52 in Fig. 3) plays a key role in the binding of argininosuccinate (Lusty and Ratner 1987). The position of this amino acid is not strictly conserved in *S. pombe*. In addition, no lysine residue is present in that region of the *E. coli* sequence, and we thus propose that it is not essential for the binding of the substrate.

The introduction of chicken δ-crystallins in the comparison does not greatly affect the consensus sequence. Delta-1 (CD1) and delta-2 (CD2) crystallins are very similar proteins (91% identity) but only CD2 has retained ASL activity. The analysis of substitutions of the CD1 sequence versus the consensus sequence provides interesting information. Acidic residues are well conserved or else display conservative substitutions (asp→glu, glu→asp) except in the case of glutamine 149 which is substituted by an alanine in both crystallins. Six histidine residues are conserved in the four ASL enzymes at positions 90, 109, 161, 172, 253 and 390. Four of them are replaced by glutamine (at positions 90, 109, 172 and 390) in CD1 but only histidine 390 is replaced by glutamine in fully active CD2. Accordingly, an essential difference between the inactive CD1 and the fully active ASLs is the absence of the histidine residues 90, 109 and 390 in CD1, which confirms the role of histidine residues in catalytic activity.

Northern blot and primer extension analyses

The expression of the *ARG7* gene at the transcriptional level was examined by Northern analysis. RNAs were extracted from *S. cerevisiae* UL1, harbouring plasmid pULG-SP2 (full gene) or pULG-SP3 (gene lacking the C-terminal part). Total RNA was electrophoresed, blotted and probed with ^{32}P-labeled pULG-ID1. No signal was visualized in the control (host strain UL1 without plasmid). A single hybridization band was detected in pULG-SP2 at a position (1 500 b) in agreement with the length of the gene. At least three transcripts of about 1800, 2800 and 3900 b were observed in pULG-SP3 (Fig. 4). As the orientiation of the truncated gene in pULG-SP3 is opposite to the orientation of the gene in pULG-SP2 (Fig. 1), the transcripts must find their 3′ ends in the 2 µm part of the plasmid.

The transcription initiation sites have been determined by primer extension using as a primer a 19 b oligonucle-

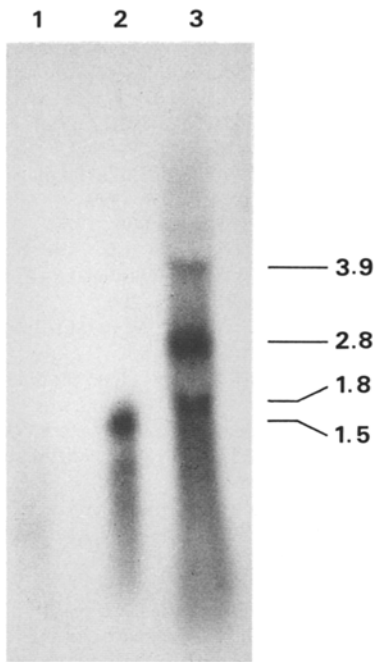

Fig. 4. Northern blots of RNA prepared from *S. cerevisiae* strain UL1 without plasmid (*lane 1*) and with plasmids pULG-SP2 (*lane 2*) or pULG-SP3 (*lane 3*). The probe was [^{32}P]-labeled (nick-translation) pULG-ID1 (1.14×10^8 cpm/μg DNA). Autoradiography was performed at $-70\,°C$ with intensifying screen for 24 h

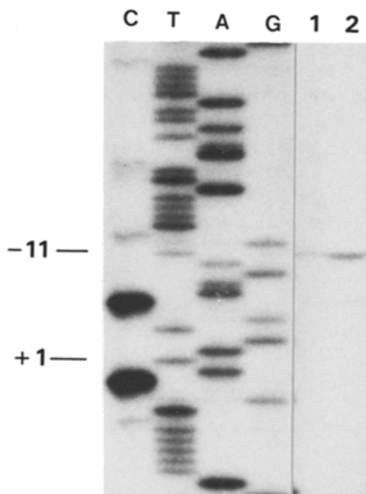

Fig. 5. Initiation site determination of the *ARG7* mRNAs by primer extension and DNA sequence analysis; *lane 1*, pULG-SP3; *lane 2*, pULG-SP2. In parallel, DNA sequencing of the *Sac*I-*Eco*RI fragment (see Fig. 1) was performed by the method of Sanger using the same oligonucleotide as that used for primer extension

otide complementary to the region +48 to +66 from the first translation codon (Fig. 5). Extended products were run in parallel with the products of the dideoxy sequencing of the non-transcribed strand (in the region *Sca*I-*Eco*RI, Fig. 1) using the primer extension oligonucleotide. In pULG-SP2, as well as in pULG-SP3, the main 5′ terminus was found at position −11. Two other initiation sites were present at −10 and −6. Accordingly, the distance between the TATA box (−72) and the major transcription initiation site is 61 bases, a value which is very close to that found in the *S. cerevisiae ARG4* gene (TATA box at −119, initiation site at −57) and which is thus in agreement with the location of the initiation "window" in budding yeast (Furter-Graves and Hall 1990).

Discussion

The *ARG7* gene of *Schizosaccharomyces pombe* is naturally expressed in *Saccharomyces cerevisiae* (Remacle et al. 1988). In *Escherichia coli*, the gene is inactive but can be turned on by a short DNA sequence inserted close to its 5′ end. The identity of this element has not been determined but could correspond to IS2 which is able to reactivate yeast genes originally silent in *E. coli* (Walz et al. 1978; Harashima et al. 1981) owing to the presence in this element of a bacterial promoter (Charlier et al. 1982). The fact that *ARG7* complements an *E. coli* mutant lacking argininosuccinate lyase indicates that this gene is devoid of introns. The analysis of the *ARG7* coding sequence confirms this assumption.

In the 5′ non-coding region of *ARG7* we have detected a perfect "TATA" box (Breatnach and Chambon 1981) located 61 bp upstream of the major transcription initiation site, a distance which is favourable to the expression of the gene in *S. cerevisiae* (Furter-Graves and Hall 1990). It should be stressed, however, that the initiation sites were determined in *S. cerevisiae* and that they do not necessarily correspond to those operating in *S. pombe* (Russel 1983). How *ARG7* is regulated in *S. pombe* is unknown and the upstream activator sequence 5′-TGACTC-3′ involved in the general control of *S. cerevisiae ARG4* gene has not been found in the short (193 bp) upstream sequence analyzed. Anyway, these activator sequences are dispensable and the deletions removing the *ARG4* UAS do not completely abolish the expression of this gene in *S. cerevisiae* (Thiry-Blaise and Loppes 1990).

The *ARG7* protein displays good similarity with other related proteins except in their carboxy-terminal part. In the course of this work, a deletion removing the 66 C-terminal amino acids was examined and shown to retain some enzymatic activity, which indicates that this part of the protein is not directly involved in catalysis. Such a situation has been described, for example, in the *S. pombe mei3* gene (Mc Leod et al. 1987) where a deletion removing 20 C-terminal amino acids of the protein (148 amino acids) retains full *mei3* activity. In the *ARG7* truncated gene present in pULG-SP3 (Fig. 1), the protein is synthesized up to serine 393. Translation should go on briefly through pBR322 in the direction *Bam*HI to *Hin*dIII and stop (TGA) after incorporation of seven amino acids.

The signal for termination of the wild-type *ARG7* gene is unknown but it should not be very far from the stop codon at position +1387 since the mRNA transcript is about 1500 bases long. The truncated gene in pULG-SP3 gives rise to at least three transcripts of higher molecular weight which should terminate at various points of the 2 μm fragment. It will probably be difficult to identify the corresponding signals since no clear consensus for termi-

nation of transcription in yeast has been established so far (Zaret and Sherman 1982; Henikoff et al. 1983).

Finally, it is interesting to note that, on the basis of the genes coding for argininosuccinate lyase, *S. pombe* does not appear to be more closely related to *S. cerevisiae* than to mammals and that, in contrast, fission and budding yeasts are equally distant from these even if several features described in the literature (Käufer et al. 1985; Toda et al. 1984) support the hypothesis that *S. pombe* is more closely related to higher eukaryotes than is *S. cerevisiae*.

Acknowledgements. This work was supported by a grant from the Belgian F.R.F.C. (2.4521.89). The authors are grateful to Dr. J.-M. Frère for revising the manuscript and to F. Hayet for expert typing. R.L. and B.J. are Research Associates of the National Foundation for Scientific Research (Belgium).

References

Amaya Y, Matsubasa T, Takiguchi M, Kobayashi K, Saheki T, Kawamoto S, Mori M (1988) J Biochem 103:177–181
Beacham IR, Schweitzer BW, Warrick HM, Carbon J (1984) Gene 29:271–279
Biggin MD, Gibson TJ, Hong GF (1983) Proc Natl Acad Sci USA 80:3963–3965
Breathnach R, Chambon P (1981) Annu Rev Biochem 50:349–383
Charlier D, Piette J, Glansdorff N (1982) Nucleic Acids Res 10:5935–5948
Dayhoff MO (1979) Atlas of protein sequence and structure. Vol 5, suppl 3, National Biochemical Research Foundation, Silver Spring, Md., USA
Debuchy R, Purton S, Rochaix JD (1989) EMBO J 8:2803–2809
Delforge J, Messenguy F, Wiame JM (1975) Eur J Biochem 57:231–239
Devereux J, Haeberli P, Smithies O (1984) Nucleic Acids Res 12:387–395
Farrell K, Overton S (1987) Biochem J 242:261–266
Furter-Graves E, Hall BD (1990) Mol Gen Genet 223:407–416
Garrard LJ, Bui QTN, Nygaard R, Raushel FM (1985) J Biol Chem 260:5548–5553
Gribskov M, Burgess RR (1986) Nucleic Acids Res 14:6745–6763
Harashima S, Sidhu RS, Toh-e A, Oshima Y (1981) Gene 16:335–341
Henikoff S, Kelly JD, Cohen EH (1983) Cell 33:607–614
Käufer NF, Simanis V, Nurse P (1985) Nature 318:78–80
Kikuchi Y, Katazawa Y, Shimatake H, Yamamoto M (1988) Curr Genet 14:375–379
Lusty CJ, Ratner S (1972) J Biol Chem 247:7010–7022
Lusty CJ, Ratner S (1987) Proc Natl Acad Sci USA 84:3176–3180
Matuo S, Tatsuno M, Kobayashi K, Saheki T, Miyata T, Iwanaga S, Amaya Y, Mori M (1988) FEBS Lett 234:395–399
Mc Leod M, Stein M, Beach D (1987) EMBO J 6:729–736
Messenguy F, Dubois E (1983) Mol Gen Genet 189:148–156
Nicolet CM, Chenevert JM, Friedberg EC (1985) Gene 36:225–234
O'Brien WE, Barr R (1981) Biochemistry 20:2056–2060
O'Brien WE, Mc Innes R, Kalumuck K, Adcock M (1986) Proc Natl Acad Sci USA 83:7211–7215
Parsot C, Boyen A, Cohen GN, Glansdorff N (1988) Gene 68:275–283
Piatigorsky J, O'Brien WE, Norman BL, Kalumuck K, Wistow GJ, Borras T, Nickerson JM, Wawrousek EF (1988) Proc Natl Acad Sci USA 85:3479–3483
Remacle JE, Breyer D, Loppes R (1988) Curr Genet 14:381–385
Rigby PWJ, Dieckmann M, Rhodes C, Berg P (1977) J Mol Biol 113:237–251
Russell PR (1983) Nature 301:167–169
Russell PR, Hall BD (1982) Molec Cell Biol 2:106–116
Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Sanger F, Nicklen S, Coulson AR (1977) Proc Natl Acad Sci USA 74:5463–5467
Schweitzer BW (1982) Ph.D Thesis, University of California, Santa Barbara, Calif., USA
Struhl K (1985) Proc Natl Acad Sci USA 82:8419–8423
Szankosi P, Heyer WD, Schuchert P, Kohli J (1988) J Mol Biol 204:917–925
Thiry-Blaise LM, Loppes R (1990) Mol Gen Genet 223:474–480
Toda T, Adachi Y, Hiraoka Y, Yanagida M (1984) Cell 37:233–242
Vrati S, Mann DA, Reed KC (1987) Mol Biol Reports (Bio-Rad Laboratories) 1, 3:1–4
Walz A, Ratzkin B, Carbon J (1978) Proc Natl Acad Sci USA 75:6172–6176
Zaret K, Sherman F (1982) Cell 28:563–573
Zoller MJ, Smith M (1982) Nucleic Acids Res 10:6487–6500

Communicated by K. Wolf