The *SSA1* and *SSA2* genes of the yeast *Saccharomyces cerevisiae*

Michael R.Slater+ and Elizabeth A.Craig*

Department of Physiological Chemistry, University of Wisconsin-Madison, Madison, WI 53706, USA
Submitted September 20, 1988 Accession nos X12926, X12927

The *SSA1* and *SSA2* genes of *Saccharomyces cerevisiae* encode 70 kiloDalton proteins related to the hsp70 of other eucaryotes (1). These hsp70-related proteins may play a role in the transport of polypeptides both across mitochondrial membranes and into the endoplasmic reticulum (2,3). *SSA1* (formerly called YG100) is located on the left arm of chromosome I, about 7 kb from the centromere (4). The 3' end of *SSA1* is distal to the centromere. A gene encoding a $tRNA^{Pro}$, *trn1* (5), is 248 bp downstream from the *SSA1* termination codon and is transcribed convergently with respect to *SSA1* (3). There is a divergently transcribed gene of unknown function, *FUN15*, upstream of *SSA1* (6). *SSA2* (formerly called YG102) is genetically unlinked to *SSA1*, but its map position has not yet been determined (Jacobsen, K. and E.A. Craig, unpublished data).

The protein coding regions of these two genes are nearly identical; there are 64 nucleotide changes resulting in 15 amino acid substitutions, and a 9 bp deletion resulting in the deletion of 3 amino acids near the carboxy terminus in *SSA2* relative to *SSA1* (Fig. 1). 6 of the amino acid substitutions are conservative in nature. Both proteins are 33% divergent from the *Drosophila* hsp70 protein (4,7).

The two proteins are less than 3% divergent, yet their non-protein coding regions are dissimilar. This observation is consistent with the different regulation shown by the two genes (4,6,8). However, the abrupt divergence in non-protein coding regions is difficult to explain without postulating some difference in function between the *SSA1* and *SSA2* proteins upon which selection could act. For example, if gene conversion were keeping the two protein coding regions similar then one would expect the regions contiguous to the coding regions to be similar as well. Or, if these two genes had simply diverged after a duplication then one would expect a similar amount of divergence in both the coding and non-coding regions. Therefore, we propose that there is a functional difference between the *SSA1* and *SSA2* proteins.

```
                                                GTCGACAAATTGTTACGTTGTGCTTTGATTTCTAAAGCGCTTCTTCACCTGCAGGTTCTGAGCCCTA   67

AGAAAAAAAATTTCCTTGGTTGAAAATGGCGGAAAAAAAAAATTCAGAAAAAGAAATAAAGCACGTGTGCGCGGTGTGTGGATGATGGTTTCATCATTGT  167

CAACGGCATTTTCGTTCTTGTGGATTGTTGTAAACTTTCCAGAACATTCTAGAAAGAAAGCACACGGAACGTTTAGAAGCTGTCATTTGCGTTTTTTCTC  267

CAGATTTTAGTTGAGAAAGTAATTAAATTATTCTTCTTTTTCCAGAACGTTCCATCGGCGGCAAAAGGGAGAGAAAGAACCCAAAAAGAAGGGGGGCCAT  367

TTAGATTAGCTGATCGTTTCGAGGACTTCAAGGTTATATAAGGGGTGGATTGATGTATCTTCGAGAAGGGATTGAGTTGTAGTTTCGTTTCCCAATTCTT  467
 A  C  C AG T TTGACGA ATTTCATTCTTAAGC CT T   AA G GTGGACGAA  GTTAT TAT AGCC CAA T GGCT GG TTTCTCCAAA   99

ACTTAAGTTGTTTTATTTTCTCTATTTGTAAGATAAGCACATCAAAAGAAAAGTAATCAAGTATTACAAGAAACAAAAATTCAAGTAAATAACAGATAAT  567
 AA GTTGAAAA ATAA A    C  AT T AG T CTT TATTCTTC  TT ATTAATTCC AC G TC  G  G TT  AT CAG  AT TTT   CA  199

ATGTCAAAAGCTGTCGGTATTGATTTAGGTACAACATACTCGTGTGTTGCTCACTTTGCTAATGATCGTGTGGACATTATTGCCAACGATCAAGGTAACA  667
    T                         T C    C            CT           T               C              299

GAACCACTCCATCTTTTGTCGCTTTCACTGACACTGAAAGATTGATTGGTGATGCTGCTAAGAATCAAGCTGCTATGAATCCTTCGAATACCGTTTTCGA  767
            C  T G       T                C          C             C  AG T  C  T            399

CGCTAAGCGTTTGATCGGTAGAAACTTCAACGACCCAGAAGTGCAGGCTGACATGAAGCACTTCCCATTCAAGTTGATCGATGTTGACGGTAAGCCTCAA  867
                           T          C A G   T                                          A     499

ATTCAAGTTGAATTTAAGGGTGAAACCAAGAACTTTACCCCAGAACAAATCTCCTCCATGGTCTTGGGTAAGATGAAGGAAACTGCCGAATCTTACTTGG  967
                                                                                              599

GAGCCAAGGTCAATGACGCTGTCGTCACTGTCCCAGCTTACTTCAACGATTCTCAAAGACAAGCTACCAAGGATGCTGGTACCATTGCTGGTTTGAATGT 1067
 T                                                                                            699

CTTGCGTATTATTAACGAACCTACCGCCGCTGCCATTGCTTACGGTTTGGACAAGAAGGGTAAGGAAGAACACGTCTTGATTTTCGACTTGGGTGGTGGT 1167
                                                                                              799

ACTTTCGATGTCTCTTTGTTGTTCATTGAAGACGGTATCTTTGAAGTTAAGGCCACCGCTGGTGACACCCATTTGGGTGGTGAAGATTTTGACAACAGAT 1267
                     C                                                                        899

TGGTCAACCACTTCATCCAAGAATTCAAGAGAAAGAACAAGAAGGACTTGTCTACCAACCAAAGAGCTTTGAGAAGATTAAGAACCGCTTGTGAAAGAGC 1367
                                                                                    T         999
```

```
CAAGAGAACTTTGTCTTCCTCCGCTCAAACTTCCGTTGAAATTGACTCTTTGTTCGAAGGTATCGATTTCTACACTTCCATCACCAGAGCCAGATTCGAA 1467
                                                                                                      1099

GAATTGTGTGCTGACTTGTTCAGATCTACTTTGGACCCAGTTGAAAAGGTCTTGAGAGATGCTAAATTGGACAAATCTCAAGTCGATGAAATTGTCTTGG 1567
                                                                    T                                 1199

TCGGTGGTTCTACCAGAATTCCAAAGGTCCAAAAATTGGTCACTGACTACTTCAACGGTAAGGAACCAAACAGATCTATCAACCCAGATGAAGCTGTTGC 1667
                                                                                                      1299

TTACGGTGCTGCTGTTCAAGCTGCTATTTTGACTGGTGACGAATCTTCCAAGACTCAAGATCTATTGTTGTTGGATGTCGCTCCATTATCCTTGGGTATT 1767
                                                                                                      1399

GAAACTGCTGGTGGTGTCATGACCAAGTTGATTCCAAGAAACTCTACCATTTCAACAAAGAAGTTCGAGATCTTTTCCACTTATGCTGATAACCAACCAG 1867
                                                C   T    A C  AG T  C  T          C                   1499

GTGTCTTGATTCAAGTCTTTGAAGGTGAAAGAGCCAAGACTAAGGACAACAACTTGTTGGGTAAGTTCGAATTGAGTGGTATTCCACCAGCTCCAAGAGG 1967
                                                                                                      1599

TGTCCCACAAATTGAAGTCACTTTCGATGTCGACTCTAACGGTATTTTGAATGTTTCCGCCGTCGAAAAGGGTACTGGTAAGTCTAACAAGATCACTATT 2067
                                                                                                      1699

ACCAACGACAAGGGTAGATTGTCCAAGGAAGATATCGAAAAGATGGTTGCTGAAGCCGAAAAATTCAAGGAAGAAGATGAAAAGGAATCTCAAAGAATTG 2167
                                                                                                      1799

CTTCCAAGAACCAATTGGAATCCATTGCTTACTCTTTGAAGAACACCATTTCTGAAGCTGGTGACAAATTGGAACAAGCTGACAAGGACACCGTCACCAA 2267
                                                                GC A  G              G T     T        1899

GAAGGCTGAAGAGACTATTTCTTGGTTAGACAGCAACACCACTGCCAGCAAGGAAGAATTCGATGACAAGTTGAAGGAGTTGCAAGACATTGCCAACCCA 2367
          A     G                          T C                 C A       A        GG                1999

ATCATGTCTAAGTTGTACCAAGCTGGTGGTGCTCCAGGTGGCGCTGCAGGTGGTGCTCCAGGCGGTTTCCCAGGTGGTGCTCCTCCAGCTCCAGAGGCTG 2467
          A                          AA    AAAAAAAAAA       T                                    A     2090

AAGGTCCAACCGTTGAAGAAGTTGATTAAGCCA  2500   SSA1
         T  C               TT  2123   SSA2
```

Figure 1. Comparison of the *SSA1* and *SSA2* DNA sequences.

The *SSA1* DNA sequence is displayed completely and all differences in the *SSA2* sequence are indicated below the *SSA1* sequence. No significant similarities were found in the promoter regions and they have been aligned without gaps relative to the initiation codons for the two genes. Nucleotides are numbered to the right of the sequences. *SSA1* encodes a predicted 642 amino acid protein of 69,749 m.w. and *SSA2* encodes a predicted 639 amino acid protein of 69,451 m.w. The protein coding region for *SSA1* is from nucleotide 568 to 2496 and for *SSA2* from 200 to 2119. Nucleotide changes in *SSA2* relative to *SSA1* that result in amino acid substitutions are indicated by underlining the variant nucleotide in the *SSA2* DNA sequence. The single gap inserted in *SSA2* to align the protein coding regions is indicated with Δ's (amino acids 616-618 in the *SSA1* protein). The nucleotide sequences were determined on both DNA strands by the chemical cleavage method and all overlaps were obtained. Four conflicts in the overlapping sequence of *SSA1* (nucleotides 2337 to 2500) and pGKN-1 (a plasmid containing *trn1* DNA(5)) were found and checked on the *SSA1* sequencing gels to confirm our assignments.

*To whom correspondence should be addressed

+Present address: Department of Molecular Genetics and Cell Biology, University of Chicago, 1103 E. 57th St., Chicago, IL 60637, USA

REFERENCES

1. Craig, E.A. (1985) Crit. Rev. Biochem. **18**, 239-280.
2. Deshaies, R.J., Koch, B.D., Werner-Washburne, M., Craig, E.A., and Schekman, R. (1988) Nature **332**, 800-805.
3. Chirico, W.J., Waters, M.G., and Blobel, G. (1988) Nature **332**, 805-810.
4. Slater, M.R. (1986) Ph.D. Thesis, University of Wisconsin — Madison.
5. Ogden, R.C., Lee, M-C, and Knapp,G. (1984) Nucl. Acids Res. **9**, 5075-5092.
6. Craig, E.A., Slater, M.R., Stone, D.E., Park, H.O., and Boorstein, W.R. (1987) In Reznikoff, W.S. *et al.* (eds), RNA Polymerase and the Regulation of Transcription, Elsevier Science, New York, pp. 267-278.
7. Ingolia, T.D., Slater, M.R., and Craig, E.A. (1982) Mol. Cell. Biol. **2**, 1388-1398.
8. Ellwood, M.S. and Craig, E.A. (1984) Mol. Cell. Biol. **4**, 1454-1459.