Volume 15 Number 8 1987 Nucleic Acids Research

Compilation and comparison of the sequence context around the AUG startcodons in *Saccharomyces cerevisiae* mRNAs

Robert Hamilton[1], Colin K.Watanabe[2] and Herman A.de Boer[3]

[1]Department of Fermentation Research and Process Development, [2]Department of Scientific Computing and [3]Department of Cell Genetics, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080, USA

Received November 25, 1986; Revised and Accepted March 16, 1987

Abstract

The nucleotide sequence of the translation initiation regions of 96 Saccharomyces cerevisiae mRNAs was compiled and compared. The entire 5' untranslated sequence of most mRNAs is very rich in A-residues. G-residues are underrepresented in the untranslated region. The AUG startcodon context appeared to be distinctly different from that of animal mRNAs, although an A-residue at -3 also occurs very frequently (81 percent) in yeast mRNAs. The prevailing codon 3' adjacent to the AUG is the UCU serine codon.

All these features are more extreme in the highly expressed genes. Fifty percent of all highly expressed genes use the UCU serine codon as second triplet. In this group G-residues are completely absent in the 7 bases preceding the startcodon and an A-residue occurs at position -1 and -3 at a frequency of 89 percent and 100 percent, respectively. The abundance of A-residues throughout the leader suggests that unstructured mRNA is required for efficient translation initiation in yeast. The consensus sequence for the AUG context in highly expressed genes can be summarized as follows:

$$\overset{}{} \quad \overset{-1\ +1}{} \quad \overset{}{}$$

$$\ldots\ \underset{U}{A}\ A\ \underset{C}{A}\ A\ \underset{C}{A}\ A\ A\ U\ G\ U\ C\ \underset{C}{U}\ \ldots$$

INTRODUCTION

In animal cells, protein synthesis starts at the 5'-proximal AUG codon in about 93 percent of all mRNAs (1-3). It is generally accepted that, after binding of the 40S ribosomal subunit to the 5' Cap-site, the subunits scan the leader until the first AUG codon is encountered at which point protein initiation takes place (4). However, the context of the AUG is also recognized by the ribosomal subunit as an important signal to trigger protein initiation events. The importance of the startcodon context became evident from a compilation of the mRNA initiation sites used by higher eukaryotes (1,2). The sequence 5'-CC$\overset{A}{G}$CCAUGG emerged as a consensus protein initiation site for mammalian mRNAs (5).

Subsequent mutational analysis confirmed Kozak's initial observations (6-8). She showed that replacing the purine at position -3 decreased the mRNA translation efficiency about 20-fold (6). The -3 purine appears to

have a dominating role; nucleotide replacements elsewhere in the consensus do affect initiation only in the absence of a purine at -3 (6).

In cases where two AUG codons are found, or introduced by means of mutagenesis (5,9), in the 5' proximal region, the AUG used most frequently for protein initiation is the one placed in the most optimal context, i.e. the sequence most resembling the consensus. An upstream located AUG in a poor context has less effect on initiation of a downstream AUG in an optimal context than an upstream located AUG in optimal context (5,9).

In the compilation of Kozak (1,2), lower eukaryotes were excluded. In this paper we compare 99 translation initiation sites of the lower eukaryote Saccharomyces cerevisiae. This comparison reveals that initiation sites in this yeast are distinctly different from those in mammalian cells and also from those in plant cells. The entire untranslated sequence preceding the startcodon is very rich in A-residues. T- and C-residues appear at close to normal frequency, but G-residues are rare.

RESULTS

In the following compilation only sequences of Saccharomyces cerevisiae are included that were present in Genbank dated August 4, 1986. Thus, 96 sequences could be compared. A window of 100 bases around the startcodon was analyzed (Table Ia and Ib). Note that the transcribed sequences upstream of the AUG start are often longer than the 47 base "window" and have been truncated to make the table more readable. It should also be noted that several yeast genes are now known to have multiple transcription start sites (10-13); we have used the sequences for the upstream transcribed regions as found in the Genbank database. For the genes SUC2 and SUC7 tandem AUG triplets are found at the start position and the second AUG has been aligned as the actual start site as referenced in the Genbank database. For some ribosomal protein genes and the actin gene, the startcodon is on a different exon than the body of the coding sequence. In such cases the separating intron was omitted and the actual translated mRNA sequence is used. All sequences use AUG as startcodon; no exceptions were encountered. Table Ia shows that the sequence immediately preceding the startcodon is extremely rich in A-residues, at most positions an A-residue is present in more than 40 percent of all cases. This is in contrast to mammalian initiation sequences where A prevails only at -3, whereas C-residues are abundant at -1 and -2 and to a lesser extent also at -4 and -5. In yeast, the highest occurrence of A-residues is, as in mammalian mRNAs, at position -3 (in 81 percent of all cases).

Table I[a] Nucleotide sequences preceding the start codon

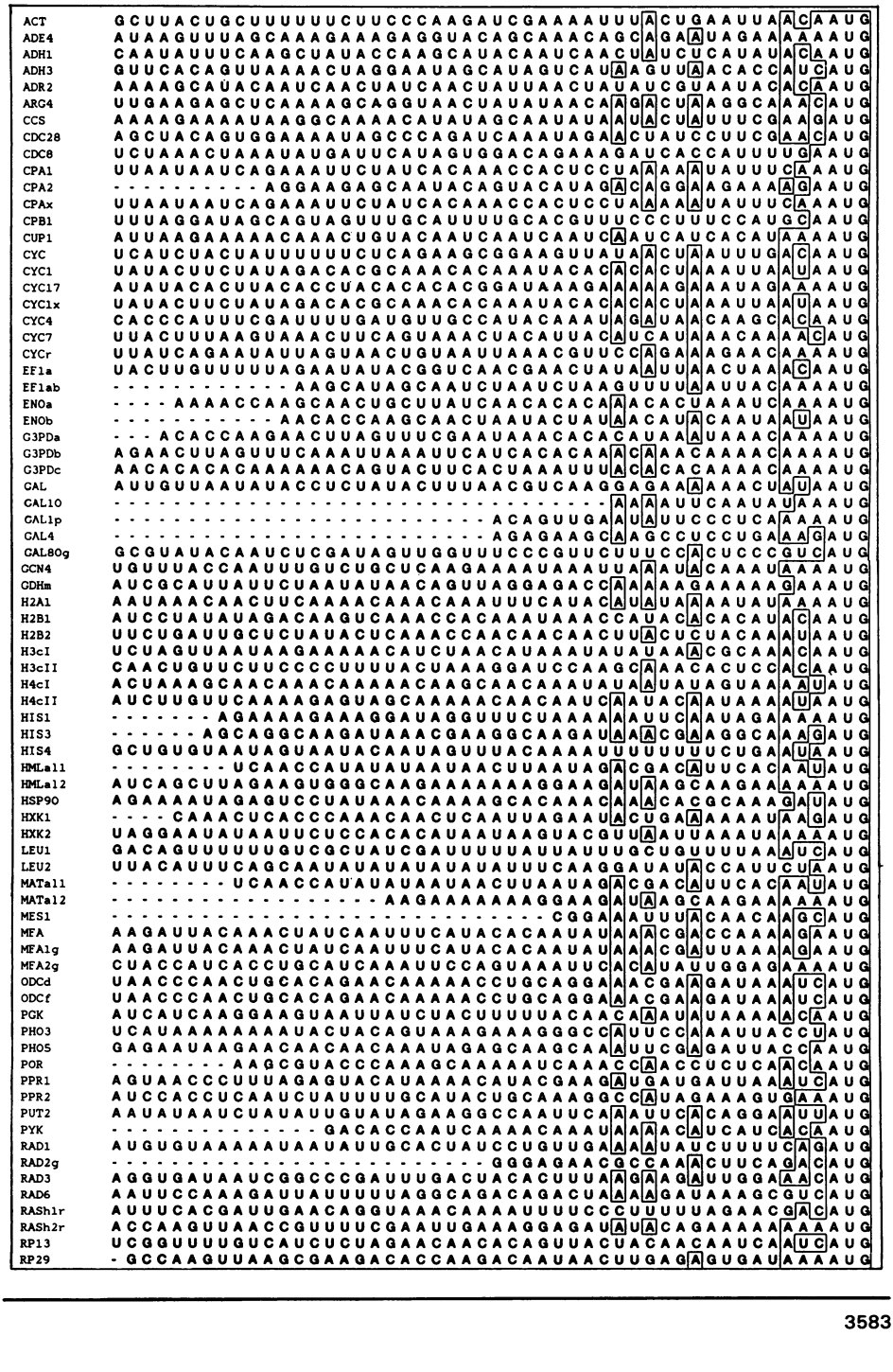
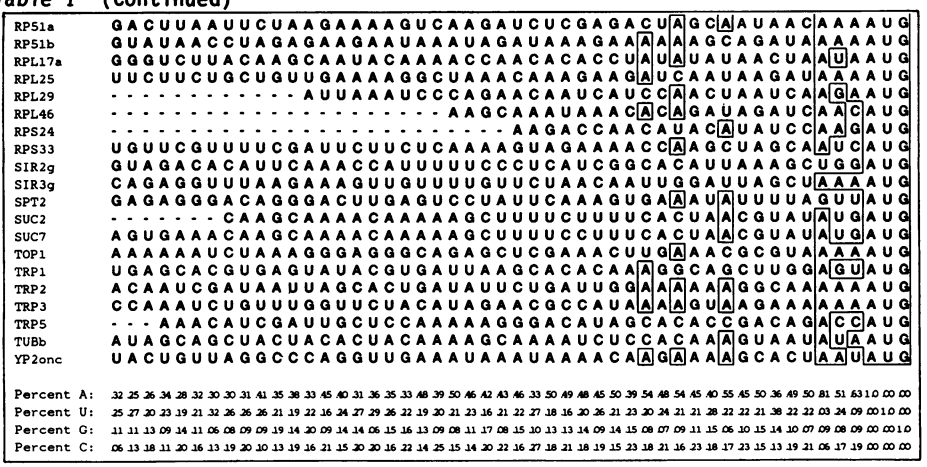


Table I[a] (Continued)



Alignment of the nucleotides preceding the startcodon of 96 yeast genes from Genbank dated August 4, 1986.

The "RNA", strand is shown in this and all other tables. All 96 sequences were arranged alphabetically and aligned at the startcodon shown to the right. The frequency of occurrence of each base at each position is shown. The nucleotides that occur at a frequency of 51 percent or greater are boxed. Throughout the paper the first base of the startcodon is numbered 1. The full name of each gene shown here is given in Table Ic. Several genes have fewer than 48 bases of untranslated sequence. In these cases the most leftward base shown is the 5' end of the mRNA.

Table I[b] Nucleotide sequences following the start codon

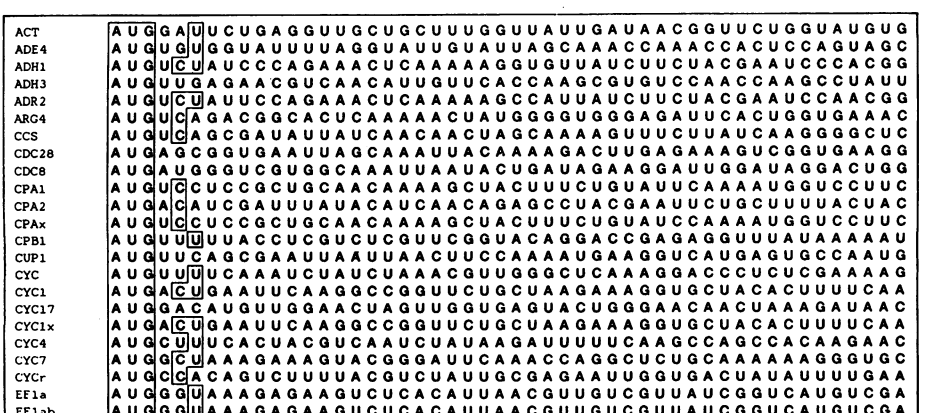
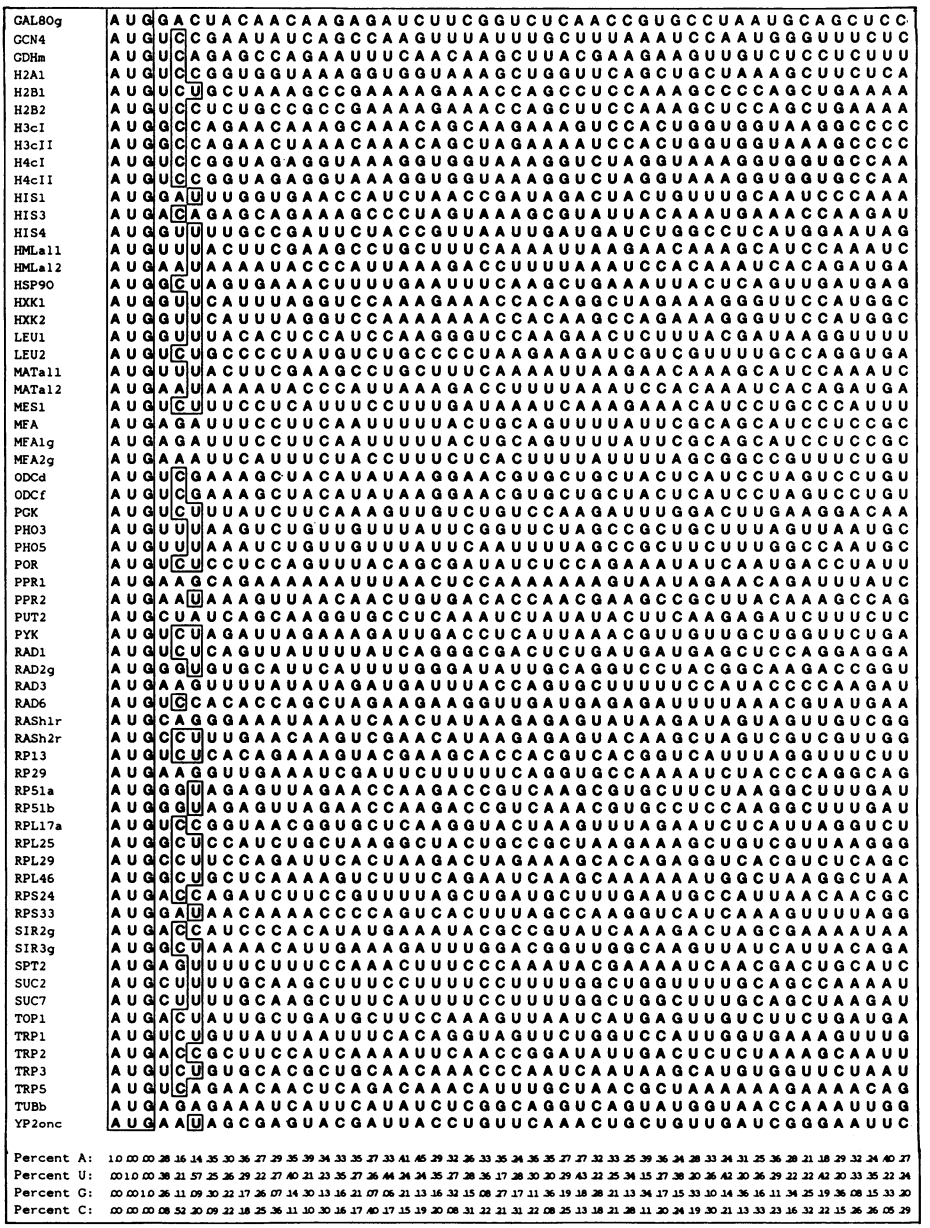


Table I[b] (Continued)



Alignment of the 47 nucleotides following the startcodon. The same genes as shown in Table Ia are aligned in the same order. In this table nucleotides occurring at a frequency of 51 percent or greater are boxed.

Table I[c] Name and reference of genes listed in table I[a] and I[b].

ACT	actin, Gallwitz (1980) *PNAS 77,* 2546; Ng (1980) *PNAS 77,* 3912 (1980); Domdey (1984) *Cell 39,* 611; Nellen (1981) *J Mol Appl Genet 1,* 239
ADE4	amidophosphoribosyltransferase, Maentsaelae (1984) *JBC 259,* 8478
ADH1	alcohol dehydrogenase, Bennetzen (1982) *JBC 257,* 3018
ADH3	alcohol dehydrogenase III, Pilgrim (1985) Unpublished, Biochem Dept, U Washington, Seattle WA 98195
ADR2	alcohol dehydrogenase II Russell (1983) *JBC 258,* 2674
ARG4	argininosuccinate lyase, Beacham (1984) *Gene 29,* 271
CCS	citrate synthase, Suissa (1984) *EMBO J 3,* 1773
CDC28	cell division control protein, Loerincz(1984) *Nature 307,* 183
CDC8	CDC8 gene, Birkenmeyer (1984), *Mol Cell Biol 4,* 583
CPA1	carbamyl phosphate synthetase small subunit, Nyunoya (1984) *JBC 259,* 9790
CPA2	arginine-specific carbamyl phosphate synthetase large subunit, Lusty (1983) *JBC 258,* 14466
CPAx	carbamoyl-phosphate synthetase small subunit, Werner (1985) *Eur J Biochem 146,* 371
CPB1	CBP1 gene, Dieckmann (1984) *JBC 259,* 4732
CUP1	copper chelatin, Karin (1984) *PNAS 81,* 337; Butt (1984) *PNAS 81,* 3332
CYC	cytochrome c1 precursor (nuclear), Sadler (1984) *EMBO J 3,* 2137
CYC1	CYC1 gene promoter region, McNeil (1985) *Mol Cell Biol 5,* 3545
CYC17	17-kd subunit of ubiquinol-cytochrome c reductase (nuclear), Van Loon (1984) *EMBO J 3,* 1039
CYC1x	iso-1-cytochrome c, Smith (1979) *Cell 16,* 753-761; Boss (1981) *JBC 256,* 12958
CYC4	cytochrome c oxidase subunit IV, Maarse (1984) *EMBO J 3,* 2831
CYC7	iso-2-cytochrome c, Montgomery (1980) *PNAS 77,* 541-545; Montgomery (1982) *JBC 257,* 7756
CYCr	ubiquinol-cytochrome c reductase 14 kd subunit, De Haan (1984) *Eur J Biochem 138,* 169
EF1a	elongation factor EF-1 alpha (TEF1), Schirmaier (1984) *EMBO J 3,* 3311
EF1ab	EF-1-alpha-* (elongation factor 1-alpha), Cottrelle (1985) *JBC 260,* 3090
ENOa	enolase (clone peno46), Holland (1981) *JBC 256,* 1385; Holland (1983) *JBC 258,* 5291
ENOb	enolase (clone peno8), *ibid*
G3PDa	glyceraldehyde-3-phosphate dehydrogenase, Holland (1979) *JBC 254,* 9839
G3PDb	glyceraldehyde-3-phosphate dehydrogenase, Holland (1980) *JBC 255,* 2596
G3PDc	glyceraldehyde-3-phosphate dehydrogenase, Holland (1983) *JBC 258,* 5291
GAL1	GAL1 inducible promoter, Johnston (1984) *Mol Cell Biol 4,* 1440
GAL10	GAL10 inducible promoter, *ibid*
GAL1p	GAL7 gene, transcript initiation region, Nogi (1983) *NAR 11,* 8555
GAL4	positive regulator of galactose inducible genes, Laughon (1984) *Mol Cell Biol 4,* 260
GAL80	GAL80 regulatory gene, Nogi (1984) *NAR 12,* 9287
GCN4	GCN4 gene, Hinnebusch (1984) *PNAS 81,* 6442
GDHm	NADPH-dependent glutamate dehydrogenase, Moye (1985) *JBC 260,* 8502
H2A1	histone h2a-1, Choe (1982) *PNAS 79,* 1484
H2B1	histone H2B-1, Wallis (1983) *Cell 22,* 799
H2B2	histone H2B-2, *ibid*; Wallis (1983) *Cell 35,* 711
H3cI	histone copy-I H3, Smith (1983) *JMB 169,* 663
H4cI	histone copy-I H3, *ibid*
H3cII	histone copy-II H3, *ibid*
H4cII	histone copy-II H3, *ibid*
HIS1	atp phosphoribosyltransferase, Hinnebusch (1983) *JBC 258,* 5238
HIS3	imidazoleglycerolphosphate dehydratase, Struhl (1981) *JMB 152,* 553
HIS4	HIS4 gene, Farabaugh (1980) *Nature 286,* 352-356; Donahue (1982) *Gene 18,* 47
HMLal1	mating-type locus HML-alpha-1, Nasmyth (1980) *Cold Spring Harb Symp Quant*

Table I[c] Continued

	Biol 45, 961; Astell (1981) *Cell 27*, 15
HMLal2	mating-type locus HML-alpha-2, *ibid*
HSP90	heat shock-inducible gene, Farrelly (1984) *JBC 259*, 5745
HXK1	hexokinase P-I, Kopetzki (1985) *Gene 39*, 95
HXK2	hexokinase PII, Froehlich (1985) *Gene 36*, 105
LEU1	isopropylmalate-1 (IPM-1), Hsu (1984) *JBC 259*, 3714
LEU2	beta-isopropylmalate (beta-IPM) dehydrogenase, Andreadis (1982) *Cell 31*, 319; Andreadis (1984) *JBC 259*, 8059
MATal1	mating-type locus MAT-alpha-1, Nasmyth (1980) *Cold Spring Harb Symp Quant Biol 45*, 961; Tatchell (1981) *Cell 27*, 25
MATal2	mating-type locus MAT-alpha-2, *ibid*
MES1	methionyl-trna synthetase, Walter (1983) *PNAS 80*, 2437
MFA	pheromone MF-alpha, Kurjan (1982) *Cell 30*, 933
MFA1g	pheromone MF-alpha-1, Singh (1983) *NAR 11*, 4049
MFA2g	pheromone MF-alpha-2, *ibid*
ODCd	OMP decarboxylase, Rose (1984) *Gene 29*, 113
ODCf	OMP decarboxylase, *ibid*
PGK	3-phosphoglycerate kinase, Hitzeman (1982) *NAR 10*, 7791
PHO3	acid phosphatase, Bajwa (1984) *NAR 12*, 7721
PHO5	acid phosphatase, *ibid*
POR	porin, Mihara (1985) *EMBO J 4*, 769
PPR1	pyrimidine pathway regulatory 1 (PPR1) gene, Kammerer (1984) *JMB 180*, 239
PPR2	PPR2 gene, regulating dihydrooratase production, Hubert (1983) *EMBO J 2*, 2071
PUT2	P5C dehydrogenase, Krzywicki (1984) *Mol Cell Biol 4*, 2837
PYK	pyruvate kinase, Burke (1983) *JBC 258*, 2193
RAD1	RAD1 protein, Yang (1984) *Mol Cell Biol 4*, 2161
RAD2g	RAD2 protein, Nicolet (1985) *Gene 36*, 225
RAD3	RAD3 protein, Naumovski (1985) *Mol Cell Biol 5*, 17
RAD6	RAD6 protein, Reynolds (1985) *PNAS 82*, 168
RASh1r	ras-H related protein c-ras-sc-1, Dhar (1984) *NAR 12*, 3611
RASh2r	ras-H related protein c-ras-sc-2, *ibid*
RP13	ribosomal protein l3 (tcm1), Schultz (1983) *J Bacteriol 155*, 8
RP29	ribosomal protein 29, Mitra (1984) *JBC 259*, 9218
RP51a	ribsomal protein 51A, Teem (1983) *PNAS 80*, 4403
RP51b	ribosomal protein 51B, Abovich (1984) *Mol Cell Biol 4*, 1871
RPL17a	ribosomal protein L17a, Leer (1984) *NAR 12*, 6685
RPL25	ribosomal protein L25, *ibid*
RPL29	ribosomal protein L29, gene CYH2, Kaeufer (1983) *NAR 11*, 3123
RPL46	ribosomal protein L46, Leer (1985) *NAR 13*, 701
RPS24	ribosomal protein S24, *ibid*
RPS33	ribosomal protein S33, Leer (1983) *NAR 11*, 7759
SIR2g	silent information regulator protein, Shore (1984) *EMBO J 3*, 2817
SIR3g	silent information regulator protein, *ibid*
SPT2	SPT2 gene encoding regulatory protein, Roeder (1985) *Mol Cell Biol 5*, 1543
SUC2	invertase, Carlson (1983) *Mol Cell Biol 3*, 439
SUC7	invertase, Sarokin (1985) *NAR 13*, 6089
TOP1	topoisomerase I, Thrash (1985) *PNAS 82*, 4374
TRP1	trp1 (n-(5'-phosphoribosyl)-anthranilate, Tschumper (1980) *Gene 10*, 157
TRP2	anthranilate synthase, component I, Zalkin (1984) *JBC 259*, 3985
TRP3	anthranilate synthase, component II, *ibid*
TRP5	tryptophan synthase, Zalkin (1982) *JBC 257*, 1491
TUBb	beta-tubulin, Neff (1983) *Cell 33*, 211
YP2onc	YP2 protein proto-oncogene (human c-has/bas), Gallwitz (1983) *Nature 306*, 704

Abbreviation and name of the genes used in Table Ia and Ib and their references describing the DNA sequence.

Optimal AUG context in yeast and mammalian mRNAs

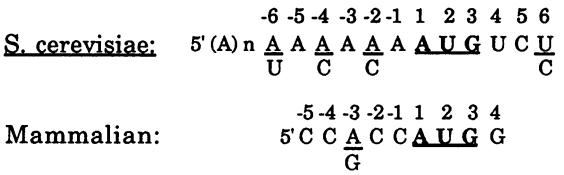


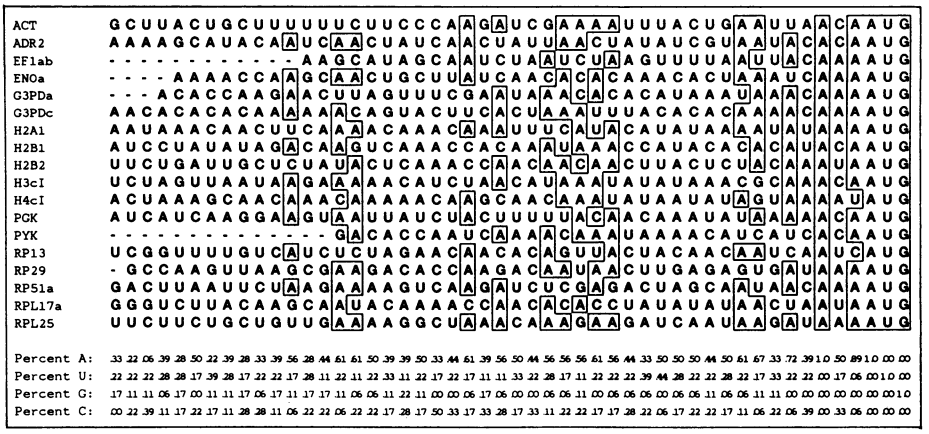
Figure 1. Comparison of the optional context around the start AUG derived for S. cerevisiae with that for mammalian mRNAs (5). With the exception of the A-residue at -3, the S. cerevisiae consensus is clearly different from that derived for mammalian mRNAs.

A strong bias exists for the first three bases downstream of the startcodon as well. Here, a U-residue prevails in the +4 and +6 positions (38 and 57 percent, respectively). G-residues are found at average frequency at +4 (G is the consensus nucleotide at position +4 in Kozak's rule; see figure 1). C-residues appear to be avoided (8 percent) at +4. Position +5 is occupied by a C-residue in 52 percent of all cases. Thus, the A/U/GCU type codons (Ser, Thr and Ala) are used most frequently with a preference for the UCU serine codon. The sequences further downstream have little or no biased nucleotide distribution.

The alphabetically listed sequences in Table I represent all translated mRNAs. Since the translation initiation efficiency is expected to be maximal in highly expressed genes, we analyzed a group of such genes separately. This group comprises the genes encoding glycolytic enzymes, ribosomal proteins, the elongation factor EF-1, and histones. Their gene products are abundantly produced in yeast cells and accordingly their coding sequence have a characteristic biased codon usage profile, i.e. they almost exclusively use 25 out of the 61 possible codons (14-16).

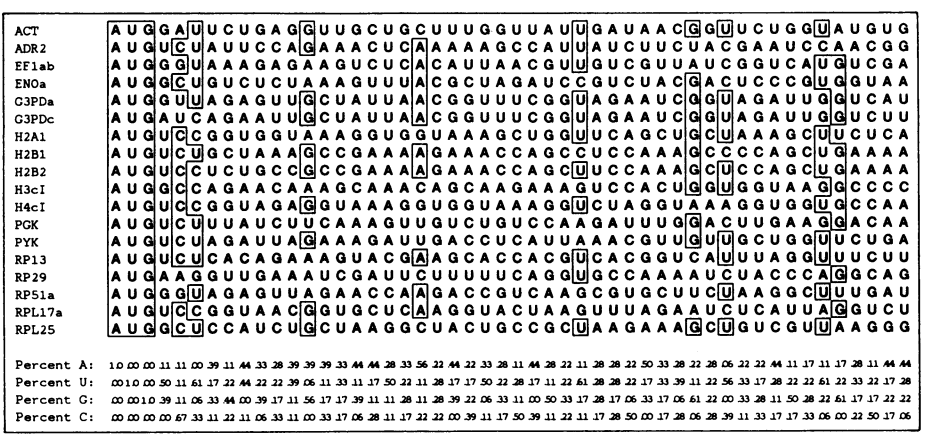
The sequences from the highly expressed genes are shown in Table IIa and IIb. Strikingly, the startcodon is preceded by an A-residue in almost all cases (with the exception of H4c1 and RP13) . At position -2 an A-residue occurs in 50 percent of all cases. Position -3 is occupied by an A-residue in all cases. An A-residue occurs in 72 percent and 67 percent of all cases at positions -5 and -7, respectively. In all other positions of the untranslated region, A-residues prevail also strongly. The frequency of A-residues gradually decreases towards the 5' end. This biased nucleotide usage in the 5' untranslated region is more pronounced in the highly expressed genes shown in Table IIa than in all genes considered as a group

Table II[a] Nucleotide sequences preceding the start codon of highly expressed genes



Alignment of nucleotide sequences before the startcodon of highly expressed genes. Nucleotides occurring in more than 50 percent of all cases are boxed.

Table II[b] Nucleotide sequences following the start codon of highly expressed genes



Alignment of nucleotide sequences following the startcodon of the same highly expressed genes shown in Table IIa. Nucleotides occurring in more than 50 percent of all cases are boxed.

(Table Ia). G-residues are rare throughout the leader and are lacking in the five positions preceding the AUG. It is interesting to note that a string of ten G-residues just prior to the AUG startcodon has recently been shown to profoundly affect the expression of the Hepatitis virus coat protein gene in S. cerevisiae. When the G-residues preceding the AUG were

removed, the protein levels increased about 100-fold (17-18; Loren Schulz, personal communication).

After the startcodon of the highly expressed genes, the UCU and UCC serine codons also prevail as second codon at a frequency that is somewhat higher than that of all genes taken together. In Figure 1, the consensus sequence for mammalian mRNAs is compared with that from highly expressed yeast mRNAs.

DISCUSSION

The compilation presented here shows that AUG is used in all cases as protein initiation triplet. The question whether ribosomes of this yeast are absolutely restricted to the use of AUG for protein initiation is somewhat controversial. Sherman and Stewart (19) reported that no iso-1-cytochrome c is made when the starting AUG was mutated to GUG, AUA, CUG, AGG or AAG. However Zitomer *et al.* (20) studying fusions of the same CYC1 initiation region to the *E. coli* galactokinase gene, showed that the triplets AUA, UUG can be used at low efficiency provided that they are preceded by an A-residue at position -3 (20). No initiation occurred at AUA or UUG when position -3 was occupied by a U-residue.

Our compilation confirms the conclusion drawn earlier by Ammerer *et al.* (21) who compiled 20 initiation sequences. From the limited sequences available at that time, they also concluded that the untranslated region is rich in A-residues; that G-residues occur rarely in the 7 bases preceding the startcodon, and that position +6 is occupied frequently by a U-residue.

Although the nonrandom nature of the AUG context is clear, its purpose with respect to the mechanism of protein synthesis initiation is not. The extremely high frequency of A-residues in the 5' untranslated part suggests that absence of RNA structure is crucial for the scanning activity of the 40S subunit; the high A-content may prevent strong interaction between the leader and the rRNA within the subunit. It is possible that the distinct difference in nucleotide bias before and after the startcodon can be read by the subunit as a signal to terminate scanning and to initiate protein synthesis at the proper AUG.

The features of the translation initiation sites of yeast mRNAs are not only distinctly different from mammalian mRNAs but also from that of plant mRNAs (22). The untranslated region of plant mRNAs is A-U rich with a moderate preference for an A-residue at -3, but this A-residue is not flanked by C-residues as is the case in mammalian mRNAs. The preference of

Table III. Amino acid frequencies at the first 10 N-terminal positions

Amino Acid	Position 1	2	3	4	5	6	7	8	9	10
A	0	9	7	7	12	8	5	8	11	14
C	0	1	0	0	0	1	1	0	0	0
D	0	5	1	0	1	3	3	6	4	2
E	0	0	7	10	6	2	3	5	2	2
F	0	7	6	2	3	8	5	10	6	4
G	0	5	8	2	6	2	6	10	1	10
H	0	0	3	2	2	2	2	0	0	0
I	0	1	4	9	6	8	5	7	5	9
K	0	5	12	3	18	8	17	9	10	14
L	0	5	8	9	8	3	10	8	8	7
M	96	1	1	0	1	1	0	0	0	1
N	0	4	1	5	3	1	7	1	2	3
P	0	3	2	11	3	6	1	1	9	2
Q	0	1	3	6	1	5	6	2	4	3
R	0	3	10	5	5	4	1	4	10	7
S	0	29	12	12	10	12	9	9	7	8
T	0	11	3	5	5	12	6	12	6	1
V	0	6	6	6	4	5	3	2	8	7
W	0	0	0	0	0	0	1	0	2	0
Y	0	0	2	2	2	5	5	2	1	2

The frequency of occurrence of each of the first 10 amino acids of the mRNAs from Table I.

a G-residue at +4 coincides with Kozak's rule. This preference is accounted for by the extremely high frequency of occurrence of an Alanine codon as second triplet in the plant mRNA sequenced thus far (see below).

The prevalence of the UCU/C codons at the second position in yeast mRNAs may have a few interesting possible explanations. Bachmair et al. (23) showed that the N-terminal amino acid of β-galactosidase determines to a great extent its stability in yeast. They infer that this is likely to be true for all deblocked non-compartmentalized proteins. Thus, they divide the amino acids into a stabilizing and destabilizing group. The stabilizing amino acids are Met, Ser, Ala, Gly, Thr and Val. Table III shows that these amino acids occur in 63 percent of all cases at the second position. When the frequency of occurrence at the second position is compared with that at the following nine positions it appears that only serine (and possibly threonine) is used preferentially at the second position. This high frequency of serine usage may be related to its protein stabilizing role. It should be noted that the codons for the stabilizing amino acids Ser, Ala and Thr all have a C-residue in their central position and in most cases (i.e. the major codons of each family) have a C- or U-residue at the third position. This accounts for the occurrence of C- and U-residues at

positions +5 and +6 in the yeast consensus initiation site (Fig. 1) . The U-residue at the +4 position in the consensus sequence is accounted for by the high frequency of serine. It is interesting to note that plant mRNAs have in almost all cases, with three exceptions, at the second position also a codon for a a stabilizing amino acid, namely alanine (see the compilation of 29 sequences by Heidecker and Messing, Ref. 23).

The high frequency of the serine codons UCU and UCC at the second position of yeast mRNAs may also play a role in mRNA translatability. This is the case for E. coli mRNAs. A mutational analysis of the second codon in the lacZ mRNA of E. coli showed that the nature of the second codon affects expression over a 20-fold range (24,25). In E. coli, serine is the second most frequently used amino acid for the second position (Ala is used most frequently, followed by serine, followed by lysine). Therefore, it is possible that the frequent use of the serine codon at the second position might likewise play a role in determining the translation initiation frequency in yeast mRNAs. Whether this is indeed the case remains to be proven experimentally. Other factors related to enzymatic requirements for the removal of the N-terminal methionine by methionine amino peptidase may also play a role in second amino acid selection (26).

ACKNOWLEDGEMENTS

We like to thank Dr. R. Hitzeman for helpful suggestions and comments. We thank Socorro Cuisia for typing the drafts and the final manuscript.

REFERENCES

1. Kozak, M. (1984) Nucleic Acids Research 12, 857-872.
2. Kozak, M. (1983) Microbiol. Rev. 47, 1-45.
3. Kozak, M. (1981) Curr. Top. Microbiol. Immunol. 93, 81-123.
4. Kozak, M. (1980) Cell 22, 7-8.
5. Kozak, M. (1984) Nucleic Acids Research 12, 3873-3893.
6. Kozak, M. (1986) Cell 44, 283-92.
7. Kozak, M. (1984) Nature 308, 241-246.
8. Kozak, M. (1981) Nucleic Acids Research 9, 5233-5262.
9. Liu, C.C., Simonsen, C.C. and Levinson, A.D. (1984) Nature 309, 82-85.
10. McNeil, J.B. and Smith, M. (1985) Molecular and Cellular Biology 5, 3545-3551.
11. Hsu, Y.P. and Schimmel, P. (1984) J. Biol. Chem. 259, 3714-3719.
12. Johnston, M. and Davis, R.W. (1984) Molecular and Cellular Biology 4, 1440-1448.
13. Zalkin, H., Paluh, J.L., van Cleemput, M., Moye, W.S. and Yanofsky, C. (1984) J. Biol. Chem. 259, 3985-3992.
14. de Boer, H.A. and Kastelein, R. (1986) in Maximizing Gene Expression, pp. 225-285, J. Davies, W. Reznikoff, and L. Gold, Eds., Butterworths Publishing Co., Boston.
15. Bennetzen and Hall (1982) J. Mol. Biol. 158, 573-597.

16. Ikemura (1982) J. Mol. Biol. 158, 573-597.
17. Kniskern, P.J., Hagopian, A., Montgomery, D.L., Burke, P., Dunn, N.R., Hoffman, K.J., Miller, W.J. and Ellis, R.W. (1986) Gene 46, 135-141.
18. Valenzuela, P., Gray, P., Quiroya, M., Zaldivar, J., Goodman, H.M. and Rutter, W.J. (1980) Nature 280, 815-819.
19. Sherman, F. and Stewart, J. W. (1982) In: The molecular biology of the yeast Saccharomyces, Metabolism and Gene Expression. Eds. J. N. Strathern, E. W. Jones and J. R. Broach, pp. 301-333.
20. Zitomer, R.S. et al. (1984) Molecular and Cellular Biology 4, 1191-1197.
21. Ammerer, G., Hitzeman, R., Hagie, F., Barta, A. and Hall, B.D. (1981) In: Recombinant DNA. Proceeding of the third Cleveland Symposium on Macro molecules, Cleveland, Ohio. Elsevier Scientific Publishing Co., Amsterdam, The Netherlands, pp. 185-197.
22. Heidecker, G. and Messing, J. (1986) Ann. Rev. Plant Physiol. 37, 439-466.
23. Bachmair, A. et al. (1986) Science 234, 179-186.
24. Looman, C., van Knippenberg, P.H. and de Boer, H.A. (1986) Ph.D. Thesis, pp. 70-80, University of Leiden, The Netherlands.
25. Looman, C., van Knippenberg, P.H., Eaton, D., Jhurani, P. and de Boer, H.A. EMBO J. (in press).
26. Stewart, J. W. et al. (1971) J. Biol. Chem. 246, 7429-7459.