Influence of the three nucleotides upstream of the initiation codon on expression of the *Escherichia coli lacZ* gene in *Saccharomyces cerevisiae*

Alfred C.Looman* and Jan Albert Kuivenhoven^{1,†}

Institut für Gärungsgewerbe und Biotechnologie, Seestraße 13, D 13353 Berlin and

¹Technische Universität Berlin, Abt. Mikrobiologie und Genetik, Gustav-Meyer-Allee 25, D 13355 Berlin, Germany

Received May 28, 1993; Revised and Accepted August 4, 1993

ABSTRACT

By introducing synthetic oligonucleotides into a *lacZ*-yeast expression vector a set of 47 plasmids (out of 64 possible) was generated, differing only in the three bases immediately upstream of the AUG initiation codon of the *Escherichia coli lacZ* gene. Expression of the β -galactosidase fusion protein encoded by the different plasmids was determined in *Saccharomyces cerevisiae* by immunogel electrophoresis. Among the clones tested we found a factor 3 difference in expression. A slight nucleotide preference was found in positions $-3(A>G>C=U)$ and $-2(G>C=U>A)$. The choice of the nucleotide at position -1 immediately 5' of the AUG did not effect translation efficiency. Increasing homology to the yeast consensus sequence (AAAAAAUGUCU) was not concomitant with an increased translation efficiency. Our results indicate that the choice of nucleotides immediately preceding the initiation codon in yeast does not dramatically influence translation efficiency, as in prokaryotes or higher eukaryotes.

INTRODUCTION

Translation initiation in the lower eukaryote *S.cerevisiae* is in general accordance with the 'scanning' model for translation initiation in eukaryotes, as was postulated in 1978 (1). The model predicts that the small ribosomal subunit (together with the initiator Met-tRNA^{Met} and initiation factors) binds to the capped 5'-end of the mRNA and migrates in the 3'-direction until the first AUG codon is reached. Here the large subunit joins the initiation complex and translation goes into the elongation phase. This mechanism sufficiently explains 95% of the translation initiation events in eukaryotes, including yeast, based upon mRNA sequences to date. Additions to this model have been postulated to account for the few reported cases where mRNA

sequences were found to contain AUG codons upstream of the initiation codon (2, 3). These additions include a favourable sequence context for an initiation codon (in vertebrates: GCCA^A/_GCCAUGG) and the regulatory influence of the mRNA secondary structure, both upstream and downstream of the initiation codon. Statistical sequence analysis and experimental data have shown that yeast may have a special place among eukaryotes as far as these details of translational modulation are concerned (4–9). The statistical consensus sequence for translation initiation sites in vertebrates and yeast differs in all except four nucleotides. The yeast consensus sequence AAAAAAUGUCU (9) shares only the AUG initiation codon and the A at position -3 with the vertebrate consensus sequence GCCA^A/_GCCAUGG (10). In fact, the yeast sequence upstream of the initiation codon is more common to the plant consensus sequence AACAAUGGC (11). The nucleotide at position -3 was found to influence translation initiation efficiency in yeast by about a factor 2 (12–14), whereas the other nucleotides surrounding the initiation codon had only a minor influence on initiation efficiency (9, 12, 15). The nucleotide choice at position -3 in vertebrate cells influences translation efficiency up to a factor of 20 and has a dominant effect (16). In higher eukaryotes the statistical consensus sequence of initiator regions coincides with the sequence that was experimentally deduced as being optimal for translation initiation (2). The statistical yeast consensus sequence does not always reflect the optimal initiation sequence in the systems that were tested (17, 19). The optimal vertebrate initiation sequence CACCAUGG, placed in front of the yeast *HIS4* gene was not superior to the original start sequence AAUAAUGG (12). In some cases where yeast mRNA contain 'upstream' initiation codons followed by short open reading frames these uORFs are involved in the regulation of gene expression such as in *MOD5* (19), *GCN4* (20–23), or *CPA1* (13). The translational modulation of the upstream open reading frames resulting in 'leaky scanning' is essential for their

* To whom correspondence should be addressed at: Technische Universität Berlin, Abt. Mikrobiologie und Genetik, Gustav-Meyer-Allee 25, D 13355 Berlin, Germany

† Present address: Academic Medical Centre, Lipid Research Group, Meibergdreef 9, NL 1105 AZ Amsterdam, The Netherlands

regulatory function (22). The nucleotide sequence context of the initiation codon can be part of this modulation (5, 13, 21). Codon context is also involved in the recognition of non-AUG initiation codons in yeast (24). Experiments designed to systematically determine the sequence requirements for translation initiation in yeast have only considered a limited set of mRNA: *CYC1* (14, 15), *HIS4* (12), *CPA1* (13). In this paper we study the *in vivo* effect in *S. cerevisiae* of nucleotide sequence changes at positions -1, -2, and -3 in a mRNA composed of the yeast *ADHI* 5'-leader sequence of 15 bases, a synthetic insert of 22 bases containing the translation initiation signal, and the *E. coli lacZ* coding sequence from the 8th codon onwards in phase with the AUG initiation codon. The sequence of the synthetic insert has been previously tested for its effect on translation initiation in *E. coli* and yeast (17, 25). In these studies the effect of codon choice in positions +4, +5, +6 was analysed. UCU was chosen as the second codon for the present study: UCU being the statistical yeast consensus sequence. The total leader length is 22 bases which fits into the normal range of yeast leader lengths (14). Computer predictions of the local secondary structure of this leader sequence have shown that it is relatively unstructured ($\Delta G > -2.5$ kcal/mol) and that the nucleotide choice of the bases surrounding the AUG codon does not change the structure to such an extent that inhibiting structures arise (17). To inhibit translation in yeast, hairpin structures with a free energy of -20 kcal/mol are required (12, 14).

MATERIALS AND METHODS

Strains, plasmids and molecular methods

Strains used were *E. coli* K12 KMBL1164 ($\Delta lac-pro$, *thi*, *supE*); *E. coli* K12 SF8 *recB12*, *recC22*, *lop-11*, *tonA1*, *thr-1*, *leuB6*, *thi-1*, *lacY1*, *supE44*, *hsdM⁻*, *hsdR⁻*; *S. cerevisiae* S3: a triploid wild-type baker's yeast strain. The expression plasmid pPTK400 (Fig. 1) has previously been described (17). It contains the *E. coli lacZ* gene without a translational start sequence positioned between the yeast *ADHI* promoter and *TRP1* terminator sequence. Oligonucleotides containing an ATG codon in phase with the *lacZ* coding sequence can be introduced into a multiple cloning site between the promoter and the *lacZ* gene. Transformation of *E. coli* was done after bivalent cationic treatment (26). Transformants were selected on LC medium (1% trypton, 0.5% yeast extract, 0.8% NaCl, 1 mM Tris-HCl pH 7.0) supplemented with 50 μ g/ml kanamycin, 50 μ g/ml ampicillin and 40 μ g/ml X-gal. Transformation of yeast was performed after treatment of logarithmically growing cells with Li-Acetate (27). Transformants were selected on YE medium (2% glucose, 0.5% yeast extract, pH 6.3) supplemented with 250 μ g/ml G418 (Gibco). Further molecular genetical methods as described in (28) were used.

Oligonucleotides and hybridisation condition

One 22 base oligonucleotide was synthesised to define the coding strand of the translation initiation signal (I). During synthesis of the three bases preceding the ATG a mixture of all four nucleotides was used. This oligonucleotide was then hybridised to four 22-base oligonucleotides (II-G,-A,-T,-C) matching the coding strand. During the synthesis of these oligonucleotides the base opposite of position -3 was chosen A, G, C or T and the bases opposite position -2 and -1 were synthesised using a mixture of all four nucleotides. After hybridisation the double-stranded fragments contained overhanging ends for *EcoRI* at the 5'-end and *BamHI* at the 3'-end:

-3-2-1+1

I: 5'-AATTNNNATGTCTGACTTACCG-3'
II-G: 3'-GNNTACAGACTGAATGGCCTAG-5'
II-A: 3'-ANNTACAGACTGAATGGCCTAG-5'
II-T: 3'-TNNTACAGACTGAATGGCCTAG-5'
II-C: 3'-CNNTACAGACTGAATGGCCTAG-5'

The non-phosphorilized oligonucleotides II-G,-A,-T,-C were dissolved in 0.07 mM EDTA, 0.7 mM Tris-HCl, pH 7.4 and mixed with oligo I in a 1:1 ratio. The oligonucleotide mixtures were heated to 60°C for 5 min and slowly cooled down to room temperature. Thus hybridised, the four oligonucleotide mixtures were ligated to *EcoRI*-*BamHI* restricted vector DNA.

Transformation, segregation and sequencing

Ligation mixtures were transformed to *E. coli* KMBL1164, and blue, ampicillin-, kanamycin-resistant colonies on X-gal plates were selected. Colonies were streaked to single colonies on selective plates to ensure segregation of clones. Isolated plasmid DNA was sequenced using an oligonucleotide priming on the *ADHI* sequence upstream of the translation initiation signal (29). In this way a set of 47 different plasmids was isolated differing in the three nucleotides upstream of the initiation codon. These were then transferred to *S. cerevisiae* S3. The DNA from G418 resistant yeast colonies was isolated (30) and re-transformed to *E. coli* SF8. The plasmid isolates from the *E. coli* transformants were analysed by restriction for changes in the DNA. No deletions or insertions in the plasmids were observed.

Quantification of β -galactosidase fusion protein

The amount of β -gal. fusion protein was measured in logarithmically growing yeast in YE medium supplemented with 150 μ g/ml G418. Cells were rapidly cooled down and lysates were prepared by disruption with glass beads in a 0.1 M K-Phosphate buffer, pH 6.6, 1 mM PMSF. The amount of β -gal. fusion protein in the clear lysates was determined by immunogel electrophoresis, as previously described, (17) with pure *E. coli* β -galactosidase (Sigma) as a standard and related to the total amount of protein in the lysates as determined by the Bradford assay (31). Expression of the different constructs is expressed as % β -gal. fusion protein of total protein.

Quantification of *lacZ* specific mRNA

In several clones showing high, middle and low β -gal. expression, the amount of *lacZ*-specific mRNA was determined by Northern blot analysis. Total RNA was isolated from logarithmically growing yeast transformants and separated on formaldehyde containing agarose gels (32). The RNA was blotted onto nitrocellulose filters. Filters were hybridised to an excess of ³²P-labelled DNA fragments from plasmid pMC1403 (33) to quantify *lacZ* mRNA and from plasmid pMY60 (34) to quantify 26S rRNA as an internal standard. The *lacZ*-mRNA and 26S rRNA appeared as single bands and the ratio of both RNA species was the same in all tested clones.

RESULTS AND DISCUSSION

The insertion of translation initiation signals in pPTK400 (Fig. 1) resulted in the expression of the *lacZ* gene in *E. coli* transformants. Plasmid DNA from 147 segregated *lac⁺* transformants was sequenced and 47 (out of 64 possible) different constructs were isolated. The 17 missing triplets were unevenly composed with

a high content of G (16) and A (11) versus C (4) and T (3). This was either due to insufficient expression of these sequences in *E. coli* resulting in non-chosen *lacZ*⁻ colonies or to an uneven concentration of nucleotides during the random synthesis of the oligonucleotides in positions -1 and -2. Since analysing the effect of the bases upstream of the AUG codon in *E. coli* (35) showed that the triplets missing in our study are expressed in *E. coli*, the latter explanation seems more likely. Transferred into yeast, all 47 constructs expressed the *lacZ* gene. The data in Table

Figure 1. Expression vector pPTK400. The plasmid contains the *E. coli lacZ* gene from the 8th codon onwards and a truncated form of the *lacY* gene. Translation initiation signal containing oligonucleotides can be inserted into the unique *EcoRI*-*Bam*HI site to restore translation of *lacZ*. Transcription of the *lacZ* gene is initiated on the yeast *ADHI* promoter sequence and terminated by the yeast *TRP1* termination sequence. Replication in yeast is provided by the 2 μm origin, in *E. coli* by the pBR322 origin of replication. Selection markers for *E. coli* are the ampicillin and kanamycin (Tn903) resistance genes. In *S. cerevisiae* the kanamycin gene confers resistance to G418 (17). Above the figure the nucleotide sequence of the 5'-part of the mRNA is shown with the oligonucleotide insert described under materials and methods. NNN depict the nucleotides whose effects on translation are described in Table 1.

1 show that the differences in expression between constructs varies by a factor 3 with a mean expression of 2.5% of total protein. Our results agree with the order of magnitude found for other initiation codon context mutants in yeast (*CYC1*: factor 2.5 (14), *HIS4-lacZ*: factor 1.6 (12)) and show that the effects are small compared to both mammalian cells (16) and *E. coli* (25). They can be amplified (factor 10) by measuring 'upstream' initiation codon context effects on the expression of the subsequent gene like in *CYC-362-Q*, (15), *CPA1* (13) or *MOD5* (19).

In the constructs GUA (1.5%), AAG (1.5%), CUC (2.5%), AGU (4.5%) and AUC (4.5%), which differ by a factor 3 in expression, the *lacZ*-mRNA was measured in comparison to the 26S rRNA. The ratios were the same in all clones (0.011, 0.012, 0.012, 0.013 and 0.012). This largely excludes the possibility that variation in mRNA concentration explains the differences in the expression found.

To find an order of preference for a certain nucleotide at the positions -1, -2 and -3, the average expression of the constructs in Table 1 was determined. At position -1 there is no clear preference order: NNG (2.6%) = NNA (2.6%) = NNU (2.6%) > NNC (2.4%). In revertants of *cyc1* with changes in position -1 of its uORF (15) the preference for the nucleotide at position -1 of the upstream AUG was G = A > C > U.

At position -2 the preference order is: NGN (2.8%) > NUN (2.7%) > NCN (2.5%) > NAN (2.2%). Changes in position -2 in other studies were found to have a negligible effect (9). At position -3 we find the order of preference: ANN (2.8%) > GNN (2.6%) > CNN (2.5%) > UNN (2.3%). Although the order of magnitude of the effect of the nucleotide choice in position -3 differs slightly, other studies confirm the positive effect of an A in position -3 on translation in yeast and the order of preference (9, 12-14, 19, 23, 24). In vertebrate cells the nucleotide at position -3 is dominant in determining the efficiency of translation initiation. A variation of a factor 20 was observed and the order of preference was the same as in yeast (16). Effects of neighbouring nucleotides are not evident from our data with the exception that in the setting ANN, nucleotide

Table 1. Effect of the nucleotide sequence preceding the initiation codon on the expression of the *lacZ* gene in *S. cerevisiae*

Seq. ^a	Expr. ^b	Seq.	Expr.	Seq.	Expr.	Seq.	Expr.
UUU	3	UCU	1.5	UAU	2.5	UGU	—
UUC	2	UCC	2	UAC	1.5	UGC	2
UUA	2	UCA	— ^c	UAA	2.5	UGA	3
UUG	3.5	UCG	—	UAG	2.5	UGG	—
CUU	2.5	CCU	2.5	CAU	2	CGU	—
CUC	2.5	CCC	2	CAC	2.5	CGC	—
CUA	3	CCA	3.5	CAA	—	CGA	—
CUG	2.5	CCG	2.5	CAG	2	CGG	—
AUU	3	ACU	2	AAU	3	AGU	4.5
AUC	4.5	ACC	3	AAC	2	AGC	1.5
AUA	2.5	ACA	2.5	AAA	—	AGA	—
AUG	2 ^d	ACG	3	AAG	1.5	AGG	—
GUU	2.5	GCU	2.5	GAU	2	GGU	3
GUC	—	GCC	2.5	GAC	2.5	GGC	3
GUA	1.5	GCA	2.5	GAA	—	GGA	—
GUG	2.5	GCG	3.5	GAG	—	GGG	—

—3-2-1

^a Nucleotide sequence from -3 to -1 in the context AAUUNNNAUG UCU.

^b Expression in % of total protein in lysates of logarithmically growing transformants with a standard error of 0.4%.

^c The sequence was not found among the sequenced constructs.

^d The sequence AUG was not considered in further calculations since it shifts the initiation codon into a different context.

choice at position -1 and to lesser extent at position -2 influence expression more than in other settings. The statistical preference for an A in position -3 relative to the initiation codon is preserved in all organisms from prokaryotes over yeast to plants and vertebrates (10, 11, 18) and is the only common nucleotide found in translation initiation consensus sequences. In vertebrate cells the statistical consensus sequence coincides with the optimal sequence for translation initiation (2), in yeast this is not always the case. The codon UCU following the initiation codon AUG was not superior to other sequences tested in the expression of *lacZ* (17). In the case of the regulatory uORF of the yeast *MOD5* gene, the yeast consensus sequence performed better than the original sequence (19). In this study the sequences AUA, ACA, AAU, AAC or AAG are not superior in comparison to sequences with less homology to the yeast consensus sequence AAA.

The nucleotide sequence context of the initiation codon in yeast plays a less dominant role in the modulation of gene expression than in prokaryotes or vertebrate cells. In searching for *in vivo* mutations that alter the translation initiation region of *HIS4* it was found that none of them affected the initiation codon context (36). Only in the more sensitive system of regulation by an uORF was a revertant of *cyc1-362* found with a mutation in the -3 position of the upstream AUG (15). The scanning model for translation in eukaryotes attributes a special role to the initiation codon context in case of 'leaky scanning' of upstream ORFs (2). In the case of the best-described yeast gene where 4 uORFs regulate gene expression, (*GCN4*) the availability and activity of initiation factors play a more important role than initiation codon context (20). The part of the ribosome itself and its initiation factors in the modulation of gene expression in yeast could in fact be the more dominant (4, 22, 37-39).

ACKNOWLEDGEMENTS

J.A.K. was supported by the European Community Erasmus Program for student exchange. The authors acknowledge the support of Prof. Dr. U. Stahl of the Dept. of Microbiology and Genetics at the Technische Universität Berlin and Prof. Dr. H.J.W. van den Broek of the Dept. of Genetics of the Agricultural University of Wageningen, The Netherlands. We thank Mrs. M. Strattmann of the Dept. Willmitzer of the Institut für Genbiologische Forschung in Berlin for preparation of the oligonucleotides and Dr. Ch. Lang-Hinrichs for helpful discussions and exchange of plasmids and methods. We are obliged to Mrs. R. Bensmann for correction of the manuscript.

REFERENCES

- Kozak, M. (1978) *Cell*, **15**, 1109-1123.
- Kozak, M. (1989) *J. Cell. Biol.*, **108**, 229-241.
- Kozak, M. (1991) *J. Biol. Chem.*, **266**, 19867-19870.
- Linder, P. (1992) *Ant. v. Leeuwenhoek*, **62**, 47-62.
- Yoon, H. and Donahue, T.F. (1992) *Molec. Microbiol.*, **6**, 1413-1419.
- Merrick, W.C. (1992) *Microbiol. Rev.*, **56**, 291-315.
- Donahue, T.F. and Cigan, A.M. (1990) *Meth. Enzym.*, **185**, 366-372.
- Müller, P.P. and Trachsel, H. (1990) *Eur. J. Biochem.*, **191**, 257-261.
- Laz, T., Clements, J. and Sherman, F. (1989) In Ilan, J. (ed.), *Translational regulation of gene expression*, Plenum Publishing Corp., New York, pp. 413-429.
- Cavener, D.R. and Ray, S.C. (1991) *Nucleic Acids Res.*, **19**, 3185-3192.
- Lütcke, H.A., Chow, K.C., Mickel, F.S., Moss, K.A., Kern, H.F. and Scheele, G.A. (1987) *EMBO J.*, **6**, 43-48.
- Cigan, A.M., Pabich, E.K. and Donahue, T.F. (1988) *Mol. Cell. Biol.*, **8**, 2964-2975.
- Werner, M., Feller, A., Messenguy, F. and Piérard, A. (1987) *Cell*, **49**, 805-813.
- Baim, S.B. and Sherman, F. (1988) *Mol. Cell. Biol.*, **8**, 1591-1601.
- Pinto, I., Jong, G.N., Sherman, F. and Hampsey, M. (1992) *Genetics*, **132**, 97-112.
- Kozak, M. (1986) *Cell*, **44**, 283-292.
- Looman, A.C., Laude, M. and Stahl, U. (1991) *YEAST*, **7**, 157-165.
- Cigan, A.M. and Donahue, T.F. (1987) *Gene*, **59**, 1-18.
- Slusher, L.B., Gillman, E.C., Martin, N.C. and Hopper, A.K. (1991) *Proc. Natl. Acad. Sci. U.S.A.*, **88**, 9789-9793.
- Abastado, J.-P., Miller, P.F., Jackson, B.M. and Hinnebusch, A. (1991) *Mol. Cell. Biol.*, **11**, 486-496.
- Hinnebusch, A. and Liebman, S.W. (1991) In Broach, J.R., Pringle, J.R. and Jones, E.W. (ed.), *The molecular and cellular biology of the yeast Saccharomyces. Genome dynamics, protein synthesis, and energetics*. Cold Spring Harbor Laboratory Press, New York, pp. 627-735.
- Lindahl, L. and Hinnebusch, A. (1992) *Curr. Op. Genet. Dev.*, **2**, 720-726.
- Müller, P.P., Jackson, B.M., Miller, P.F. and Hinnebusch, A.G. (1988) *Mol. Cell. Biol.*, **8**, 5439-5447.
- Zitomer, R.S., Walthall, D.A., Raymond, B.C. and Hollenberg, C.P. (1984) *Mol. Cell. Biol.*, **4**, 1191-1197.
- Looman, A.C., Bodlaender, J., Comstock, L.J., Eaton, D., Jhurani, P., de Boer, H.A. and van Knippenberg, P.H. (1987) *EMBO J.*, **6**, 2489-2492.
- Himeno, M., Shibata, T., Kawahara, Y., Hanaoka, Y. and Komano, T. (1984) *Agric. Biol. Chem.*, **48**, 657-662.
- Ito, H., Fukuda, Y., Murata, K. and Kimura, A. (1983) *J. Bacteriol.*, **153**, 163-168.
- Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: A laboratory manual*, Cold Spring Harbor University Press, Cold Spring Harbor, New York.
- Chen, E.Y., and Seeburg, P.H. (1985) *DNA*, **4**, 165-170.
- Hoffman, C.S. and Winston, F. (1987) *Gene*, **57**, 267-272.
- Bradford, M.M. (1976) *Anal. Biochem.*, **72**, 248-264.
- Davis, L.G., Dibner, M.D. and Battey, J.F. (1986) *Methods in molecular biology*, Elsevier, New York, pp. 143-146.
- Casadaban, M.J., Chou, J. and Cohen, S.N. (1980) *J. Bacteriol.*, **143**, 971-980.
- Verbeet, M.Ph., Klootwijk, J.K., van Heerikhuizen, H., Fontijn, R., Vreugdenhil, E. and Planta, R.J. (1983) *Gene*, **23**, 53-63.
- Hui, A., Hayflick, J., Dinkelspiel, K. and de Boer, H.A. (1984) *EMBO J.*, **3**, 623-629.
- Donahue, T.F. and Cigan, A.M. (1988) *Mol. Cell. Biol.*, **8**, 2955-2963.
- Nygård, O. and Nilsson, L. (1990) *Eur. J. Biochem.*, **191**, 1-17.
- Thomas, A.A.M., Scheper, G.C. and Voorma, H.O. (1992) *The New Biologist*, **4**, 404-407.
- Rhoads, R.E. (1993) *J. Biol. Chem.*, **268**, 3017-3020.