ANALYSIS OF THE EFFECT OF dG · dC HOMOPOLYMER TAILS ON EXPRESSION OF A MOUSE α -AMYLASE cDNA GENE IN YEAST

by

OLE OLSEN

Department of Physiology, Carlsberg Laboratory,
Gamle Carlsbergvej 10, DK-2500 Copenhagen Valby

Keywords: Recombinant DNA, secretion, homopolymer cDNA tails, DNA sequence analysis

Yeast expression plasmids with an α -amylase cDNA gene inserted behind the *ADHI* gene promoter direct different levels of amylase secretion to the culture medium. The difference among these plasmid constructions is the length of the 5' and 3' nontranslated flanking regions of the cDNA gene. The nucleotide sequences of these regions were determined and it was found that a stretch of 17 dG · dC bp in front of the amylase gene reduces secretion substantially compared to a stretch of 3 or 8 dG · dC bp. The length of the 3' dC · dG homopolymer tail does not influence the secretion level. Whether the 5' homopolymer nucleotide region reduces transcription or translation is not known, but a G stretch in the 5' end of an mRNA may adopt a secondary structure which reduces the translatability. Problems related to nucleotide sequence determination of homopolymer regions are discussed.

1. INTRODUCTION

Differential rates of translation of mRNAs in *Saccharomyces cerevisiae* and other eukaryotic cells may be an important factor in regulation. Consequently an important question concerns the nature of structural features of mRNAs and it is now known that secondary structures of eukaryotic mRNAs influence the efficiency of translation (14, 15, 23, 29, 30).

Secretion in *Saccharomyces cerevisiae* cells is mediated by the same set of organelle intermediates employed by other eukaryotes (28). This paper presents a continuation of the analysis of expression and secretion of mouse salivary α -amylase (32). Plasmids carrying an α -amylase cDNA gene inserted in the EcoRI cloning site of the yeast expression vector pMA56 directed different levels of amylase secretion when intro-

duced in yeast cells (Figure 1). The only difference among these plasmid constructions was the length of the 5' and 3' nontranslated flanking regions of the cDNA gene.

In an effort to analyse the basis for the observed differences in secretion activity, the nucleotide sequences flanking the α -amylase cDNA and the vector, pMA56, were determined. The results demonstrate that the secretion level is reduced by the presence of a long G tail (originating from the cDNA construction (10, 22)) in the transcribed 5' noncoding region of an mRNA. In contrast, the length of the 3' homopolymer tail does not influence secretion activity.

The sequence of a 5' (dC)_n-(dA)₂₆ 3' region was obtained by standard sequencing methods, whereas it was not possible to sequence the

Abbreviations: ADHI (PR) = Alcohol dehydrogenase I (promoter); bp = basepair(s).

Figure 1. Physical map of α -amylase expression plasmids. Mouse salivary α -amylase cDNAs of different length inserted adjacent to the yeast *ADHI* gene promoter. The plasmid also contains a yeast *TRP1* gene for selection and the origin of replication from the yeast 2μ circular plasmid. The pMS plasmids contain most of the *E. coli* plasmid pBR322 including both the ampicillin resistance gene and the origin of replication. Cleavage sites for the restriction enzymes Apal (A), BamHI (B), ClaI (C), EcoRI (E), and HindIII (H) are shown.

complementary homopolymer region 5' (dT)₂₆-(dG)_n 3'. These difficulties in sequencing homopolymer stretches are probably caused by secondary structures of DNA.

2. MATERIALS AND METHODS

2.1. Strains and media

E. coli K-12 strain JM101 (20) was used to propagate M13mp phage vectors as described by MESSING (21). Otherwise the *E. coli* K-12 strain JM83 (20) was used as host in bacterial cloning experiments as described by VIEIRA and MESSING (33). The *Saccharomyces cerevisiae* strain DBY746 (32) was grown in either YPD or SC media (24).

2.2. Enzymes and chemicals

Radioactive nucleotides and a Maxam-Gilbert DNA sequencing kit were purchased from

from New England Nuclear. All enzymes as well as deoxy- and dideoxynucleotides were from Boehringer Mannheim.

2.3. DNA purification

Plasmid DNA was prepared by the method of BIRNBOIM and DOLY (3) which for large scale preparations was followed by caesium chloride equilibrium centrifugation. Purification of specific DNA fragments from low melting temperature agarose gels was carried out as described by MANIATIS et al. (18).

2.4. Transformations

Transformation of *E. coli* with plasmid DNA was performed as described by COHEN et al. (6). Yeast cells were transformed according to the method of ITO et al. (12) using lithium acetate to induce competence. Yeast transformants were selected on SC medium without tryptophan, and cells secreting α -amylase were detected on similar medium supplied with 1% starch (32).

2.5. DNA sequence determination

The nucleotide sequence of the α -amylase cDNA gene and its 5' and 3' flanking regions was determined by the dideoxynucleotide chain termination method of SANGER et al. (25, 27) except that the elongation reaction temperature was 55 °C, or by the chemical degradation method of MAXAM and GILBERT (19). For Maxam-Gilbert sequencing restriction fragments were end-labelled by filling in 5' protruding ends with DNA polymerase I (Klenow fragment) in the presence of an appropriate α -³²P-dNTP (18).

3. RESULTS

3.1. Nucleotide sequence analysis of cDNA 5' regions

Specific restriction fragments flanking the 5' region of α -amylase cDNA inserts were sequenced by the M13-dideoxynucleotide chain termination method (Figure 2). It was necessary

Figure 2. Analysis of cDNA 5' flanking region.

Horizontal arrows represent direction of sequence information obtained from each M13 subclone. (dG)_n and ATG show location of homopolymer nucleotide tail and initiator codon for protein synthesis, respectively. The star indicates location (+65) of a single dC·dG bp insertion in plasmid pMS27. The small AluI-EcoRI fragment represents a length of 22 bp.

to raise the elongation reaction temperature to 55 °C to obtain the sequences of the homopolymer stretches in front of the α -amylase cDNA.

In the plasmid pMS12, which directs the highest level of amylase secretion, 3 dG·dC bp are located 5' to the α -amylase cDNA whereas in pMS29 a homopolymer tail of 17 dG·dC bp is present in the homopolymer region. The latter plasmid directs secretion of relatively low amounts of amylase.

The plasmid pMS27 does not direct synthesis and secretion of α -amylase. Sequence analysis of the 5' flanking region revealed a homopolymer region of 8 dG·dC bp in front of the amylase gene. However, analysis of the coding sequence showed a single dC·dG bp insertion at position 65 from the initiation codon causing a reading frame shift.

To analyse the effect on α -amylase synthesis and secretion of 8 dG·dC bp in the 5' homopolymer region plasmid pMS12-27 was constructed. This was done by replacing the small ApaI-ClaI fragment from pMS27 by the analogous fragment from pMS12. This plasmid contains the flanking regions from pMS27 including a homopolymer region of 8 dGs in front of the α -amylase cDNA, and an intact coding region partly derived from pMS12. The level of α -amylase secretion directed by pMS12-27 is similar to that directed by pMS12 (Figure 3).

Figure 3. Yeast cells secreting α -amylase. Colonies from individual transformed yeast cells containing plasmids (1: pMS12, 2: pMS12-27, 3: pMS27-12, 4: pMS29) were transferred to SC medium without tryptophan supplied with a 1% starch. After incubation overnight the plates were stained with iodine vapour. A halo shows that the colony secretes α -amylase.

3.2. Nucleotide sequence analysis of cDNA 3' regions

The sequence of the 3' end of the α -amylase cDNA in pMS12, pMS27, and pMS29 (Figure 4) was also determined by the dideoxynucleotide chain termination method. The results obtained confirmed the previously published nucleotide sequence (10), but using the anti-sense strand as template it was not possible to determine the extent of the presumed homopolymer region. A compression in the sequence ladder covering the distal part of the 3' noncoding end of the cDNA and the proximal part of the 5' end of the 2 μ region made it impossible to read a sequence. A similar result was obtained by Maxam-Gilbert sequencing of the anti-sense strand of the ClaI-HindIII fragment (Figure 4). However, when either M13 or Maxam-Gilbert techniques were employed in sequencing the sense strand, the primary structure of the entire homopolymer region could be determined. All of the plasmids analysed have a dA·dT tail of 26 bp following the cDNA. This dA·dT tail is followed by the expected dC·dG region which is

Figure 4. Analysis of cDNA 3' flanking regions.

Horizontal arrows with black dot show where sequence information was obtained by Maxam-Gilbert sequencing of either the sense strand (M-G (s)) or the anti-sense strand (M-G (a)). Other horizontal arrows indicate where sequence information was obtained by Sanger sequencing using either the sense strand as template (S (s)) or the anti-sense strand as template (S (a)). Regions impossible to sequence are indicated by a zigzag line. α -Amylase protein translation stop codon TAA is shown as well as the position of the homopolymer nucleotide stretch. The small EcoRI-HindIII fragment represents a length of 103 bp.

of different length in the plasmids. The results of the entire analysis are summarized in Figure 5.

3.3. Exchange of 3' dC·dG cDNA tails

The effect of expression level of the 3' noncoding homopolymer dC tail was analysed by exchanging unique BamHI-ClaI fragments be-

tween plasmids. All pMS plasmids have a unique BamHI site at the extreme 5' end of the promoter and a unique ClaI site in the 3' end of the amylase gene (Figure 1). The small BamHI-ClaI fragment from pMS12 was ligated into the larger ClaI-BamHI fragment from plasmid pMS27. Thereby the 3' dC·dG tail of pMS12 is changed from 14 to 5 bp. Transformation of competent yeast cells with this plasmid construction, pMS27-12, gives rise to a secretion level similar to that obtained from the pMS12 construction (Figure 3) demonstrating that the secretion level from a plasmid with 14 dC·dG bp in 3' end is indistinguishable from one having 5 dC·dG bp in the 3' end when the 5' end is unchanged.

4. DISCUSSION

The expression of mouse α -amylase in yeast has previously been described (32). Amylase cDNA was treated with Bal31 exonuclease and inserted behind the *ADHI* gene promoter in the yeast expression plasmid pMA56. Introduced in yeast the resulting plasmids direct different levels of α -amylase secretion. In the present investigation, we have analysed the effect on gene expression and secretion level of homopolymer nucleotide tails flanking the cDNA gene in the yeast expression plasmid. From the results presented it is concluded that the structural basis for the variation in secretion is the length of the

Figure 5. Summary of the nucleotide sequence analysis. Sequence parts of the *ADHI* promoter (ADHI-PR), α -amylase coding region (CR) and 2 μ region are shown together with the nucleotide sequence of the EcoRI linkers (L), α -amylase untranslated region (UT), and the homopolymer regions (H). The α -amylase coding region consists of 1548 bp. Plasmids pMS27 and pMS12-27 contain the same flanking regions. However, pMS27 does not direct synthesis of α -amylase because of a dC·dG bp insertion in position +65 of the coding sequence.

homopolymer dG·dC tail preceding the cDNA. 3 dG·dG bp (pMS12) or 8 dG·dC bp (pMS12-27) have similar – if any effect on expression level, whereas 17 dG·dC bp (pMS29) reduce the amount of α -amylase secreted into the growth medium considerably. In contrast, the length of the homopolymer dC·dG region flanking the 3' end of the cDNA does not influence gene expression since secretion levels are indistinguishable whether 5 dC·dG bp (pMS27-12) or 14 dC·dG bp (pMS12) are flanking the 3' end of the cDNA.

Determination of nucleotide sequences was complicated by possible structural features of the 5' and 3' homopolymer regions flanking the cDNAs. Sequencing of the 3' regions was only possible if the sense strand (5' (dA)₂₆-(dC)_n 3') was used as template for Sanger sequencing or as the labelled strand in Maxam-Gilbert sequencing. When either sequencing technique was used in the analysis of the anti-sense strand representing the same region (5' (dG)_n-(dT)₂₆ 3') a compression in the sequence ladder was observed in the corresponding area of the sequencing gels.

In solution poly(dT) does not exhibit any ordered conformation (26), whereas poly(dG) can adopt at least single-, double-, and triple-stranded secondary structures (7, 26). Apparently, the sequence 5' (dG)_n-(dT)₂₆ 3' adopts a conformation which has severe impact both on DNA transcription (when the nucleotide stretch acts as template for the Klenow fragment of DNA polymerase I) and on the reactions employed in Maxam-Gilbert sequencing. The sequence appears to adopt a stable conformation of increased stability compared to the complementary sequence 5' (dC)_n-(dA)₂₆ 3' since the latter sequence was obtainable by both sequencing techniques.

Poly(dG) in front of the α -amylase cDNA also seems to form a secondary structure, but apparently not at 55 °C since poly(dG) may act as template for DNA polymerase I (Klenow fragment) at that temperature.

The site for transcription initiation of the *ADHI* gene is present in the expression plasmid pMA56, and it is believed that transcription of the amylase gene is initiated within the *ADHI* promoter (1). Consequently an oligo(G) stretch becomes part of the mRNA 5' region. It is not

known how far transcription proceeds, but it has been reported that it continues into the 2 μ region of the vector in the absence of a transcription termination signal (1).

If both the 5' and 3' homopolymer regions are parts of the mature mRNA intramolecular G·C basepairing could affect the translational potential in a similar way as found by SPENA et al. (30) who showed that translational efficiency of zein mRNA is reduced by hybrid formation between inverted repeats of the 5' and 3' nontranslated regions. The low secretion level directed by pMS29 could thereby be explained by a thermodynamically stable secondary structure between the 5' 17 Gs and 3' 13 Cs of the transcript preventing efficient translation. However, since the efficiency of translation seems to be independent of the cDNA 3' homopolymer dC stretch, it is unlikely that intramolecular G·C base pairing of the transcript is the reason for low translation efficiency. Similar results were obtained by GALILI et al. (8). They transcribed dG/dC tailed zein cDNA with the SP6 promoter, translated the mRNA in cell-free systems, and found that 20 Gs almost completely blocked mRNA translation, while even 8 Gs had an inhibitory effect compared to a situation where the entire G stretch had been removed. The 3' oligo(C) had no effect on translation efficiency.

In the plasmid constructions reported on here, the polyadenylation signal for the native mouse α -amylase mRNA is present in the cDNA gene. If that signal is recognized in yeast the poly(C) region may be removed post-transcriptionally. This may be the reason why the length of the cDNA oligo(C) region has no influence on the level of expression.

It is now generally believed that differential rates of translation initiation play an important role in regulation of protein synthesis in eukaryotic cells (4, 9, 14, 16, 23) and that the relevant information for preinitiation complex formation is found in the 5' noncoding region of the mRNA (31). The exact nature of this information is not known and therefore it is of interest to find which structural features of an mRNA influence the efficiency of binding to ribosomes. It has been found that the presence of secondary structures near the cap structure reduces the translatability of the mRNA (15, 23), and that

cDNA/mRNA hybrids upstream from the AUG initiator codon inhibit the translation of rabbit globin mRNA (17), but the mere length of the 5' noncoding region has no regulatory effect (5, 23).

A computer analysis of the nucleotide sequences of the α -amylase cDNA 5' and 3' nontranslated regions did not reveal any inverted repeats which could be candidates for formation of secondary structure(s) reducing the translatability.

However, it is known that a poly(G) sequence is capable of forming secondary structures. It appears likely that the stable form of poly(G) in solution is a four stranded helix (2, 11, 34), but poly(G) also exist as ordered single- and double-helical structures (26). It is therefore likely that even if the AUG initiator codon appears in a favourable nucleotide context (13) a long G tail in the nontranslated region of an mRNA can form a secondary structure which may be antagonistic to efficient translation.

ACKNOWLEDGEMENTS

KARL KRISTIAN THOMSEN is gratefully acknowledged for encouragement, stimulating discussion, and critically reading of the manuscript. KARINA BOSERUP is thanked for excellent technical assistance and NINA RASMUSSEN for preparing figures. DITER VON WETTSTEIN is thanked for support and encouragement and for critical reading of the manuscript. This work was supported by a grant no. 1984-133/001-84.127 from Teknologistyrelsen of the Ministry of Industry to DITER VON WETTSTEIN and M.C. KIELLAND-BRANDT.

REFERENCES

1. AMMERER, G.: Expression of genes in yeast using the *ADCI* promoter. *Methods Enzymol.* 101, 192-201 (1983)
2. ARNOTT, S., R. CHAUDRASEKARAN & C. M. MARTILLA: Structures for polyinosinic and polyguanylic acid. *Biochem. J.* 141, 537-543 (1974)
3. BIRNBOIM, H. C. & J. DOLY: A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Res.* 7, 1513-1523 (1979)
4. BECERRA, S. P., J. A. ROSE, M. HARDY, B. M. BAROUDY & C. W. ANDERSON: Direct mapping of adeno-associated virus capsid proteins B and C: A possible ACG initiation codon. *Proc. Natl. Acad. Sci. USA* 82, 7919-7923 (1985)
5. BRENDLER, T., T. GODEFROY-COLBURN, S. YU & R. E. TACH: The role of mRNA competition in regulating translation. *J. Biol. Chem.* 256, 11755-11761 (1981)
6. COHEN, S. N., A. C. CHANG & L. HSU: Nonchromosome antibiotic resistance in bacteria. Genetic transformation of *Escherichia coli* by R-factor DNA. *Proc. Natl. Acad. Sci. USA* 69, 2110-2114 (1972)
7. DUGAICZYK, A., D. L. ROBBERSON & A. ULLICH: Single-stranded poly(deoxyguanylic acid) associates into double- and triple-stranded structures. *Biochemistry* 19, 5869-5873 (1980)
8. GALILI, G., E. K. KAWATA, R. E. CUELLAR, L. D. SMITH & B. A. LARKINS: Synthetic oligonucleotide tails inhibit *in vitro* and *in vivo* translation of SP6 transcripts of maize zein cDNA clones. *Nucleic Acids Res.* 14, 1511-1524 (1986)
9. GODEFROY-COLBURN, T. & R. E. TACH: The role of mRNA competition in regulating translation. *J. Biol. Chem.* 256, 11762-11773 (1981)
10. HAGENBÜCHLE, O., R. BOVERY & R. A. YOUNG: Tissue-specific expression of mouse α -amylase genes: nucleotide sequence of isoenzyme mRNAs from pancreas and salivary gland. *Cell* 21, 179-187 (1980)
11. HOWARD, F. B., J. FRAZIER & H. T. MILES: Stable and metastable forms of poly(G). *Biopolymers* 16, 791-809 (1977)
12. ITO, H., Y. FUKUDA, K. MURATA & A. KIMURA: Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* 153, 163-168 (1983)
13. KOZAK, M.: Comparison of initiation of protein synthesis in prokaryotes, eukaryotes, and organelles. *Microbiol. Rev.* 47, 1-45 (1983)
14. KOZAK, M.: Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283-292 (1986)
15. LAWSON, T. G., B. K. RAY, J. T. TODDS, J. A. GRIFCO, R. D. ABRAHAMSON, W. C. MERRICK, D. F. BETSCH, H. L. WEITH & R. E. TACH: Influence of 5' proximal secondary structure on the translational efficiency of eukaryotic mRNAs and on their interaction with initiation factor. *J. Biol. Chem.* 261, 13979-13989 (1983)
16. LEE, K. A. W., D. GUERTIN & N. SONENBERG: mRNA secondary structure as a determinant in cap recognition and initiation complex formation. *J. Biol. Chem.* 258, 707-710 (1983)

17. LIEBHABER, S. A., F. E. CHASH & S. H. SHAKIN: Translationally associated helix-destabilizing activity in rabbit reticulocyte lysate. *J. Biol. Chem.* 259, 15597-15602 (1984)
18. MANIATIS, T., E. F. FRITSCH & J. SAMBROOK: In: *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1982)
19. MAXAM, A. M. & W. GILBERT: Sequencing end-labeled DNA with base-specific chemical cleavage. *Methods Enzymol.* 65, 499-560 (1980)
20. MESSING, J.: A multi-purpose cloning system based on the single-stranded bacteriophage M13. *Recombinant DNA Technical Bulletin, NIH Publications* No. 79-99, 2, No. 2, 43-48 (1979)
21. MESSING, J.: New M13 vectors for cloning. *Methods Enzymol.* 101, 20-79 (1983)
22. PEACOCK, S. L., C. M. McIVER & J. J. MONAHAN: Transformation of *E. coli* using homopolymer-linked plasmid chimeras. *Biochim. Biophys. Acta* 665, 243-250 (1981)
23. PELLETIER, J. & N. SONENBERG: Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. *Cell* 40, 515-526 (1985)
24. PETERSEN, J. G. L., M. C. KIELLAND-BRANDT, S. HOLMBERG & T. NILSSON-TILLGREN: Mutational analysis of isoleucine-valine biosynthesis in *Saccharomyces cerevisiae*. Mapping of *ilv2* and *ilv5*. *Carlsberg Res. Commun.* 48, 21-34 (1983)
25. RASMUSSEN, S. K., H. E. HOPP & A. BRANDT: Nucleotide sequences of cDNA clones for B1 hordein polypeptides. *Carlsberg Res. Commun.* 48, 187-199 (1983)
26. SAENGER, W.: In: *Principles of Nucleic Acid Structure*. Springer Verlag, pp. 315-320 (1984)
27. SANGER, F., S. NICKLEN & A. R. COULSON: DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463-5467 (1977)
28. SCHEKMAN, R., P. NOVIC, S. FERRO-NOVIC, B. ESMON, W. HANSEN, T. ETEHEVERRY & T. STEVENS: Protein secretion and organelle assembly in yeast. *Rec. Adv. Yeast Mol. Biol.* 1, 143-155 (1982)
29. SHAKIN, S. H. & S. A. LIEBHABER: Destabilization of messenger RNA/complementary DNA duplexes by the elongating 80S ribosome. *J. Biol. Chem.* 261, 16018-16025 (1986)
30. SPENA A., E. KRAUSE & B. DOBBERSTEIN: Translation efficiency of zein mRNA is reduced by hybrid formation between the 5'- and 3'-untranslated region. *EMBO J.* 4, 2153-2158 (1985)
31. SONENBERG, N., D. GUERTIN, D. CLEVELAND & H. TRACHSEL: Probing the function of the eukaryotic 5' cap structure using a monoclonal antibody directed against cap-binding proteins. *Cell* 27, 563-572 (1981)
32. THOMSEN, K. K.: Mouse α -amylase synthesized by *Saccharomyces cerevisiae* is released into the culture medium. *Carlsberg Res. Commun.* 48, 545-555 (1983)
33. VIEIRA, J. & J. MESSING: The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19, 259-268 (1982)
34. ZIMMERMAN, S. B., G. H. COHEN & D. R. DAVIES: X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *J. Mol. Biol.* 92, 181-192 (1975)

Accepted by E. LUND