Volume 11 Number 22 1983 Nucleic Acids Research

Complete structure of the porcine pro-opiomelanocortin mRNA derived from the nucleotide sequence of cloned cDNA

Guy Boileau, Claire Barbeau, Lucie Jeannotte, Michel Chrétien and Jacques Drouin*

Institut de Recherches Cliniques de Montréal, 110 Avenue des Pins Ouest, Montréal, Québec H2W 1R7, Canada

Received 18 August 1983; Revised and Accepted 19 October 1983

ABSTRACT

Polyadenylated RNA isolated from porcine pituitary neurointermediate lobes was used to construct a cDNA library. The library was screened with a rat genomic DNA fragment specific for pro-opiomelanocortin sequences. Two positive clones, pJA-19 and pJA-20, containing respectively 850 bp and 550 bp were characterized. Sequence analysis of the cDNA inserts revealed the complete structure of the porcine pro-opiomelanocortin mRNA. This mRNA would include 129 5'-untranslated nucleotides, 801 nucleotides coding for the 267 amino acids precursor and 162 3'-untranslated nucleotides. Comparison with pro-opiomelanocortin mRNA sequences from other species shows regions of high homology not only in the coding sequences but also in the 5' untranslated region where the first 50 nucleotides are over 80% purines.

INTRODUCTION

Adrenocorticotropic hormone (ACTH), β-lipotropic hormone (β-LPH) and β-endorphin (β-END) are synthetized in the pituitary gland from a common glycoprotein precursor called pro-opiomelanocortin (POMC). This was first suggested from "pulse" and "pulse-chase" experiments performed in the rat pars intermedia and in the mouse pituitary tumor cell line AtT-20 (1-3). These results were confirmed when Nakanishi et al. (4) deduced the complete amino acid sequence of the bovine precursor from the nucleotide sequence of a cloned cDNA. This work also revealed the presence of a MSH-like sequence in the N-terminal portion of the precursor and the presence of pairs of basic amino acid residues at the processing sites in the precursor. Cloning of genomic DNA fragments showed that the gene coding for POMC is formed of three coding regions (exons) interrupted by two large intervening sequences (introns) (5-10).

We report in this paper the nucleotide sequences of two cDNA clones encoding the porcine POMC precursor. These two clones overlap in such a way that the complete structure of the porcine POMC mRNA could be deter-

mined. This work confirms the protein sequences of the N-terminal peptide (11) and of the "joining peptide" (12) that we recently published and shows, once more, the strong homology of the POMC N-terminal peptide between different species. This important sequence conservation suggests a biological role for the N-terminal glycopeptide of POMC.

METHODS

Poly(A)RNA purification

Total cellular RNA was isolated from porcine pitultary neurointermediate lobes essentially as described by Chirgwin et al. (13) and the poly(A) containing species purified on oligo(dT) cellulose (P.L. Biochemicals) as detailed by Aviv and Leder (14).

cDNA synthesis

Single stranded cDNA (sscDNA) was synthetized using AMV reverse transcriptase. A standard reaction mixture contained: 50 mM Tris.HCl pH 8.3, 60 mM KCl, 8 mM $MgCl_2$, 0.5 mM DTT, 1 mM each dATP, dGTP and dTTP, 0.5 mM dCTP, 30 µCi (^{32}p) dCTP (New England Nuclear) and 50 µg/ml each of poly (A^+) RNA and oligo $(dT)_{12-18}$ (Collaborative Research). The reaction mixture was heated at 68°C for 2 min and rapidly quenched on ice. The synthesis was started by the addition of 2.5U of AMV reverse transcriptase per µg of poly (A^+) RNA and proceeded at 42°C for 90 min. The reaction was stopped with 20 mM EDTA, the mixture extracted with an equal volume of phenol/chloroform and the sscDNA precipitated with ethanol.

The sscDNA was dC-tailed at its 3' end with terminal deoxynucleotidyl transferase (Boehringer Mannheim Co.) in the presence of $CoCl_2$ as described by Deng and Wu (15) and the second strand was synthetized using DNA polymerase large fragment (Bio Labs) and oligo$(dG)_{12-18}$ (Collaborative Research). A typical reaction mixture contained: 50 mM $K_2H\ PO_4$ pH 7.4; 6.5 mM $MgCl_2$; 180 µM each of dATP, dCTP, dGTP and dTTP; 1 mM DTT; 50 µCi of (^{32}p) dCTP; 20 µg/ml each of oligo$(dG)_{12-18}$ and dC-tailed sscDNA previously heated to 90°C and quenched on ice. Second strand synthesis was initiated by the addition of 10U DNA polymerase large fragment per µg of dC-tailed sscDNA. The reaction was stopped with 20 mM EDTA and extracted with phenol/chloroform. The dscDNA was precipitated with ethanol and its 3' ends were dC-tailed as described above.

Cloning of the cDNA and isolation of the POMC clones.

The dC-tailed dscDNA was annealed to an equimolar amount of PstI-cut pBR327, dG-tailed in the presence of $MnCl_2$ (15) and the hybridization

mixture used to transform E. coli DH-1 strain (16). Recombinants were selected on tetracycline and screened by colony hybridization (17) using as probe (18) the 1.6 kb XhoI-HindIII fragment containing the rat POMC exon 3 (7). Positive clones were isolated and their plasmid DNA purified (19). The cDNA inserts were sequenced (20-23) as outlined in Figure 1.

RESULTS

Pig neurointermediate poly A^+ RNA (12,5 µg) was used to synthesize about 1 µg dC-tailed double-stranded cDNA. When 20 ng cDNA was annealed with 50 ng dG-tailed pBR327 and used to transform E. coli DH1, the 107 transformants obtained were screened by filter hybridisation (17) with a 1.6 Kb XhoI-HindIII rat genomic DNA fragment containing exon 3 of the POMC gene (7). Plasmid DNA was prepared (19) from recombinant colonies showing strong hybridization and digested with PstI. The fragments were electrophoresed on agarose gel and transferred to nitrocellulose paper (24). Hybridization with the 1.6 Kb POMC probe revealed cDNA inserts of approximately 850 and 550 bp in two recombinants, pJA-19 and pJA-20, respectively (not shown).

Preliminary restriction mapping of both inserts resulted in maps with more differences than would be expected from two cDNAs overlapping over a stretch of 550 bp (not shown). These differences could be explained if the two cDNAs encoded different regions of the porcine POMC mRNA. Further analysis was done by DNA sequence determination.

The nucleotide sequence of the cDNA inserts present in pJA-19 and pJA-20 was determined by both the dideoxy chain termination method (20) using single stranded DNA templates (21) produced by subcloning in the bacteriophage M13 mp8 (22) and the chemical technique of Maxam and Gilbert (23) according to the strategy depicted in Figure 1. The nucleotide sequences revealed that clone pJA-19 covers the region of the mRNA from the 5'-untranslated segment to the γ-LPH sequence and clone pJA-20 covers the regions of the mRNA from the ACTH sequence to the poly(A) tail. The two clones overlap over a stretch of about 300 nucleotides in the ACTH/γ-LPH region. The exact length of the overlap is uncertain since neither the 3' end of pJA-19 nor the 5' end of pJA-20 have been fully sequenced and the boundaries shown in Figure 1 are approximate. The sequenced portion of both clones is identical in the overlapping region suggesting that both inserts were synthesized from similar mRNA templates.

The nucleotide sequence corresponding to the porcine POMC mRNA was

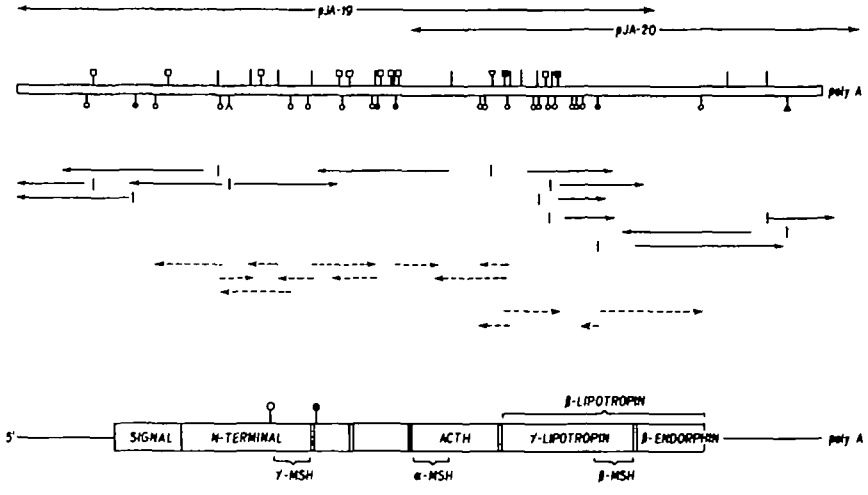

Figure 1. Schematic representation of cDNA clones pJA-19 and pJA-20 encoding pig POMC. The sequences included in each clone are represented at the top of the Figure followed by a map of the restriction sites used for chemical (23) sequencing (solid arrows) or for subcloning in M13 mp8 (21,22) and dideoxy (20) sequencing (broken arrows). The following symbols are used for Bgl II ( ), Xhol ( ), Eco RV ( ), Apa I ( ), Hae III ( ), Alu I ( ), FnudD II ( ) and Msp I ( ). The major peptides produced by processing of POMC are illustrated in the bottom diagram showing their respective position in the protein precursor and mRNA. Double vertical bars indicate pairs of basic amino acids whereas open and filled-in lollipops indicate O- and N-glycosylation, respectively (12).

deduced from the combined nucleotide sequences of pJA-19 and pJA-20. This sequence is presented in Figure 2 along with its translation product. The porcine POMC mRNA would contain 1092 transcribed nucleotides divided into 129 5'-untranslated nucleotides, 801 coding nucleotides and 162 3'-untranslated nucleotides. This latter region contains the AAATAAAA sequence thought to be important for the addition of the poly(A) tail to the 3' end of the mRNA (25). pJA-20 contains a poly(A) tail of approximately 75 residues (not shown).

Translation of the nucleotide sequence into amino acid residues from the first ATG codon at position 130 reveals an open reading frame ending with the TGA codon at position 931. This 801 nucleotides stretch codes for a 267 amino acid long polypeptide which has the expected structure for pig pre-POMC including a 26 amino acids signal peptide. The deduced amino acid sequence is indeed in complete agreement with those reported for pig POMC-derived peptides (11-12, 26-28). As in other species (4-10, 29) active

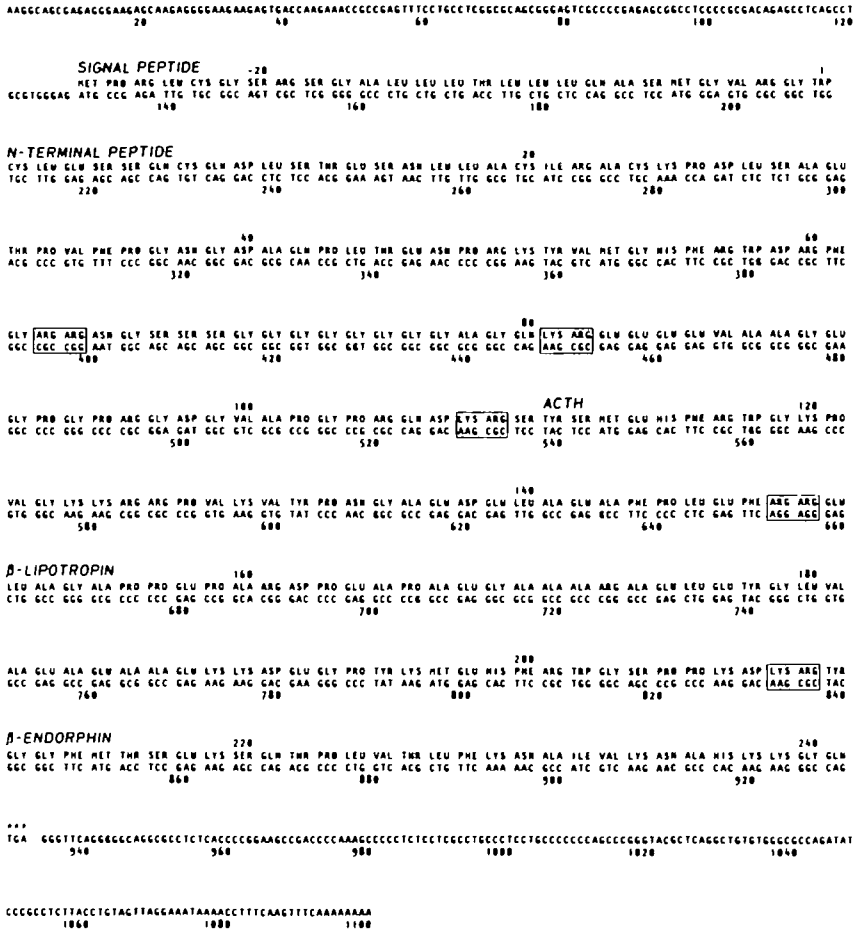

Figure 2. Combined DNA sequences of pJA-19 and pJA-20 cDNA inserts. The complete (± 2 nucleotides) mRNA sequence is shown together with it's POMC translation product. Signal peptide amino acids are numbered negatively and boxed residues indicate processing sites in the anterior pituitary.

peptides are linked by the -Lys-Arg- dipeptide except, here, between ACTH and β-LPH where the sequence Arg-Arg is found.

Comparison between the nucleotide sequences of bovine (5) and human (8-10) POMC genomic DNA and the sequence of the porcine POMC mRNA presented in Figure 2 suggest that the AA doublet located at the 5'-end of our sequence is very close if not itself the mRNA capping site. Indeed, as shown in Figure 3, the homology between these three 5'-non coding segments is very

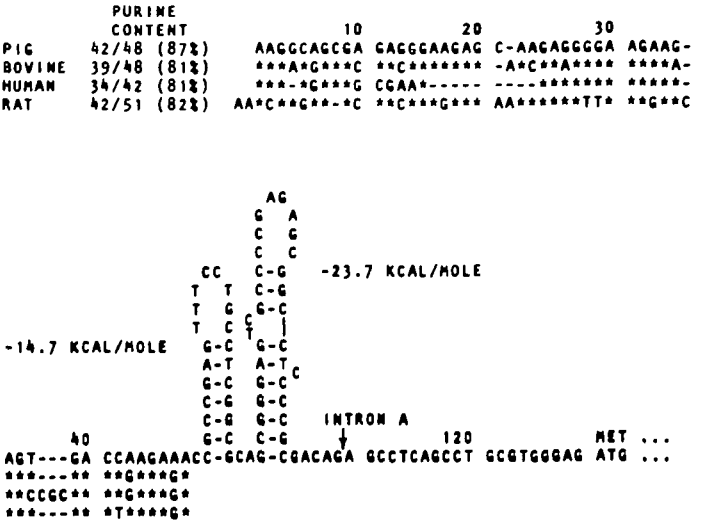

Figure 3. Sequence comparison of purine-rich stretch in 5'-untranslated region of various POMC mRNAs and postulated secondary structure in pig POMC 5'-untranslated region. Purine content in the first 50 or so nucleotides of each mRNA is shown at the beginning of each sequence. Asterisks and dashes indicate homology and gap in sequences, respectively. The human and bovine sequences are from ref. 8-10 and 30 whereas the rat sequence is from our own work (submitted). The hairpin-loop structures are the most stable that could be computed (33) for the pig Exon 1 (5'-untranslated) sequence Presumed intron A splice position and initiator ATG triplet are shown.

good, in particular around the mRNA capping site. This site was precisely located for the bovine mRNA (30). Furthermore, sequencing of primer extended reverse transcripts (E. Oates and E. Herbert, personal communication) map the 5'-end of pig POMC mRNA at this position (± 2 nucleotides). These results and the presence of a poly(A) tail in pJA-20 strongly suggest that we have characterized the entire structure of the porcine POMC mRNA. This interpretation is in agreement with the size of pig POMC mRNA as determined by Northern blots (31) and hybridization of pig anterior and neurointermediate lobes poly A^+ RNA (Figure 4).

The sequence of the 5'-untranslated region (which is largely encoded by exon 1 in the gene, ref. 5-10) has two noteworthy particularities that are illustrated in Figure 3. First, nucleotides 1 to 49 are very rich in purines (86%). Second, nucleotides 50 to 104 could form two relatively stable hairpin-loop structures (ΔG^o of -14,7 and -23,7 Kcal/mole calculated as in Ref. 32). Of the different possible base-paired structures that can be

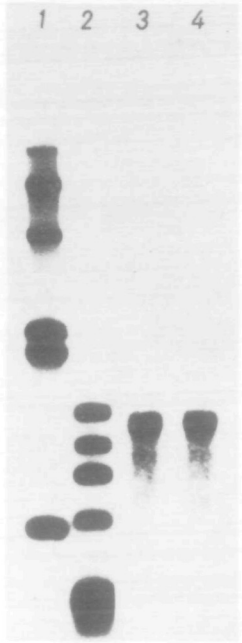

Figure 4. Detection of pig POMC mRNA by Northern blot (31) and hybridization with a DNA insert containing a complete cDNA copy produced by recombination of pJA19 and pJA-20. Pig neurointermediate (0,5 µg, lane 3) and anterior (2 µg, lane 4) pituitary poly A^+ RNA were glyoxylated and electrophoresed on a 1.1% agarose gel (38). Fragment size markers are shown in lane 1 (Hind III-digested λ DNA) and lane 2 (Hae III-digested ØX 174 DNA: four larger fragments are 1353, 1078, 872 and 603 nucleotides). The autoradiogram was exposed overnight at room temperature.

computed (33) for the 5'-untranslated region (exon 1), these close-range interactions lead to the highest gain in free energy. The comparative alignment of the first 50 nucleotides of POMC mRNA in Figure 3 is meant to illustrate sequence homology between the different species rather than to imply that these sequences cannot form secondary structures with other parts of POMC mRNA.

DISCUSSION

We have presented in this paper the characterization of two cDNA clones encoding porcine pituitary neurointermediate lobe POMC. Sequencing of the two DNA inserts lead to the determination of the complete structure of porcine POMC mRNA. Comparison of the coding sequence with that of other species (4-10,29) revealed a strong homology in three portions of the precursor, namely, in the N-terminal glycopeptides, ACTH and β-MSH/β-endorphin. In contrast, the regions corresponding to the "joining peptide" and to γ-LPH have diverged substantially. The sequence of the mRNA encodes eight pairs of basic amino acids which could be recognized by specific proteases for the processing of the POMC precursor protein. Purification and characterization of POMC related peptides from porcine pituitaries has shown cleavages of the

peptide chain at seven of these eight pairs (12,26-28). Only the pair located in front of the γ-MSH sequence (amino acids 51-61) has not been shown to be cleaved in porcine pituitaries. However, it has been shown to occur in the rat pituitary (34).

Comparison of the 5'-untranslated regions of POMC mRNAs from various species (4,10,29) reveals in all of them a purine-rich (over 80%) stretch of about 50 nucleotides and the potential for hairpin-loop formation. These hairpin-loop structures vary in their stability, their position and their sequence. They may play a role in mRNA translation efficiency (35) but their sequence variability across species argues against specificity of sequence being involved in this putative role. The 5' purine-rich stretch is a fairly well conserved sequence in the various POMC mRNAs. It is more conserved than the remainder of the 5'-untranslated region or the 3'-untranslated region except for sequence around the AAUAAA. This conclusion might stand to a larger number of species comparisons and reflect a true selective advantage but more comparative and experimental data will be needed to establish it. Similar purine-rich stretches are not found in the 5'-untranslated regions of mRNAs encoding other pituitary hormones, other enkephalin precursors or other steroid-regulated transcriptional units. These sequences could play a role in regulation of mRNA utilization and/or turnover; for example, it has been shown a long time ago that the half-life of the ovalbumin and conalbumin mRNAs increase after exposure to estrogens (36,37) but the structural basis for this effect is unknown. Such hypotheses have been difficult to approach experimentally but they can now be tested by directed mutagenesis of a well characterized transcription unit.

ACKNOWLEDGEMENTS

Present address for G. Boileau is Département de Biochimie, Université de Montréal, C.P. 6128, Succ. "A", Montréal, Québec H3C 3J7. We are grateful to Ed Oates and Ed Herbert for communicating their results before publication, to Michael Zuker for allowing us to use his RNA structure programs and to Nahun Sonenberg for his comments on the manuscript. We are thankful for the secretarial assistance of Nicole Valiquette and the financial support of the Medical Research Council and the National Cancer Institute of Canada.

*To whom correspondence should be addressed

REFERENCES

1. Roberts, J.L. and Herbert, E. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 4826-4830.
2. Mains, R.E., Eipper, B.A. and Ling, N. (1977) Proc. Natl. Acad. Sci.

U.S.A. 74, 3014-3018.
3. Crine, P., Gianoulakis, C., Seidah, N.G., Gossard, F., Pezalla, P.D., Lis, M. and Chrétien, M. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 4719-4723.
4. Nakanishi, S., Inoue, A., Kita, T., Nakamura, M., Chang, A.C.Y., Cohen, S.N. and Numa, S. (1979) Nature (London) 278, 423-427.
5. Nakanishi, S., Taranishi, Y., Noda, M., Notake, M., Watanabe, Y., Kadidani, H., Jingami, H. and Numa, S. (1980) Nature 287, 752-755.
6. Chang, A.C.Y., Cochet, M. and Cohen, S.N. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 4890-4894.
7. Drouin, J. and Goodman, H.M. (1980) Nature (London) 288, 610-613.
8. Takahashi, H., Teranishi, Y., Nakanishi, S. and Numa, S. (1981) FEBS Letters, 135, 97-102.
9. Cochet, M., Chang, A.C.Y. and Cohen, S.N. (1982) Nature 297, 335-339.
10. Whitfeld, P.L., Seeburg, P.H. and Shine, S. (1982) DNA 1, 133-143.
11. Larivière, N., Seidah, N.G. and Chrétien, M. (1981) Int. J. Pept. Protein Res. 18, 487-491.
12. Boileau, G., Larivière, N., Hsi, K.-L., Seidah, N.G. and Chrétien, M. (1982) Biochemistry 21, 5341-5346.
13. Chrigwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. (1979) Biochemistry 18, 5294-5299.
14. Aviv, H. and Leder, P. (1972) Proc. Natl. Acad. Sci. U.S.A. 69, 1408-1412.
15. Deng, G.-R. and Wu, R. (1981) Nucleic Acids Res. 9, 4173-4188.
16. Hanahan, D. (1983) J. Mol. Biol. 166, 557-580.
17. Benton, W.D. and Davis, R.W. (1977) Science 196, 180.
18. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
19. Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193-197.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467.
21. Schreier, P.H. and Cortese, R. (1979) J. Mol. Biol. 129, 169-172.
22. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
23. Maxam, A.M. and Gilbert, W. (1980) Meth. Enzym. 65, 499-560.
24. Southern, E.M. (1975) J. Molec. Biol. 98, 503-517.
25. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
26. Bradbury, A.F., Smyth, D.G. and Snell, C.R. (1976) Nature (London) 260, 165-167.
27. Graf, L., Szukely, J.I., Ronai, A.Z., Dunai-Kovaes, Zs. and Bajurz, S. (1976) Nature (London) 263, 240-242.
28. Riniker, B., Sirber, P., Rittel, W. and Zuber, H. (1972) Nature New Biol. 235, 114-115.
29. Uhler, M. and Herbert, E. (1983) J. Biol. Chem. 258, 257-261.
30. Inoue, A., Nakamura, M., Nakanishi, S., Hidaka, S., Miura, K.-I. and Numa, S. (1981) Eur. J. Biochem. 113, 531-539.
31. Thomas, P.S. (1980) Proc. Natl. Acad. Sci. U.S.A. 77(9), 5201-5205.
32. Salser, W. (1977) Cold Spring Harbor, Symp. on Quant. Biol. 42, 985-1002.
33. Zuker, M. and Stiegler, P. (1981) Nucl. Acids Res. 9, 133-148.
34. Browne, C.A., Benneth, H.P.J. and Solomon, S. (1981) Biochem. Biophys. Res. Commun. 100, 336-343.
35. Kozak, M. (1982) In: Protein Biosynthesis in Eukaryotes (Perez-Bercoff Ed.) Plenum Press, pp. 167-197.
36. Palmiter, R.D. (1973) J. Biol. Chem. 248, 8260-8270.
37. Palmiter, R.D. and Carey, N.H. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 2357-2361.
38. McMaster, M.K. and Carmichael, G.G. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 4835-4838.