Vol. 175, No. 1, 1991
February 28, 1991

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
Pages 110-116

CLONING AND NUCLEOTIDE SEQUENCE OF A BOVINE PANCREATIC PREPROCARBOXYPEPTIDASE A CDNA

Isabelle Le Huërou, Paul Guilloteau*, René Toullec*
Antoine Puigserver+ and Catherine Wicker

Centre de Biochimie et de Biologie Moléculaire du Centre National de la Recherche Scientifique, 31 Chemin Joseph-Aiguier, 13402 Marseille Cedex 9, France

* Laboratoire du Jeune Ruminant de l'Institut National de la Recherche Agronomique, 65 rue de Saint Brieuc, 35042 Rennes Cedex, France

Received January 3, 1991

A full-length cDNA clone encoding bovine pancreatic preprocarboxypeptidase A was isolated and sequenced. The 1405-base pair insert contains a 26-nucleotide 5'-noncoding region, a 1260-nucleotide open reading frame and a 76-nucleotide 3'-noncoding fragment plus a poly(A) tail of at least 43 nucleotides. The open reading frame encodes a protein of 419 amino acids, including the 16 amino acid signal peptide. The mature enzyme (309 residues) has two additional C-terminal amino acids, as compared with the amino acid sequence of the protein which was reported more than 20 years ago. In addition, four residues deduced from the nucleotide sequence differed from those identified in the reported amino acid sequence from their net charge: Asp-89, Asp-114, Gln-122, and Asp-185 instead of Asn-89, Asn-114, Glu-122, and Asn-185, respectively. A high degree of identity exists between the nucleotide sequences (81.3%), on the one hand, and the amino acid sequences (78.3%), on the other hand, of bovine preprocarboxypeptidase A and rat preprocarboxypeptidase A1.

Pancreatic carboxypeptidase A (EC 3. 4. 17. 1.) is a zinc exopeptidase that catalyzes the release of carboxyl-terminal amino acids in peptides and proteins provided that the side chains are aromatic or aliphatic but uncharged. The proenzyme is characterized by its ability to form complexes with another or two other proteins in the acinar cell, mostly in ruminant species (1-5) but also in hog (6,7), whale (8) and human (9). Bovine procarboxypeptidase A forms mainly a ternary complex with a sedimentation coefficient of 6S known as proCPA-S6 in which the zymogen itself is noncovalently associated with chymotrypsinogen C (10,11) and with a third protein formerly known as subunit III which, despite its poor enzyme activity (12,13), was recently found to be a protease E (14). Two binary complexes with a sedimentation coefficient of approximately 5S have been found to consist of procarboxypeptidase A and chymotrypsinogen C (15) or of procarboxypeptidase A and the zymogen E (4) in the bovine pancreas.

+To whom correspondence should be addressed.

0006-291X/91 $1.50

Bovine carboxypeptidase A has been purified in two allelomorphic forms (16). Using electrophoretic techniques (17), two isoenzyme forms have been found to exist in guinea-pig, rabbit, and human pancreatic glands, and three forms in the dog. A single and two separate forms of monomeric carboxypeptidase A were isolated from the pig and human pancreas (18,19), respectively, using high-performance liquid chromatography.

The amino acid sequence of bovine carboxypeptidase A (20,21) and that of its activation peptide (22,23) have been determined by Edman degradation while the primary structures of rat preprocarboxypeptidases A1 and A2 have been deduced from nucleotide sequencing of the corresponding cDNAs (24-26). Moreover, the cDNA clones made it possible to investigate more closely the role of some putative functional residues using site-directed mutagenesis (27,28).

In this study, the nucleotide sequence of a full-length cDNA of bovine preprocarboxypeptidase A and the deduced amino acid sequence were determined and compared with those of rat preprocarboxypeptidases A1, A2 and B.

MATERIALS AND METHODS

Materials. [α-^{32}P] dCTP (>110 TBq/mmol) and [^{35}S α] dATP (>37 TBq/mmol) were from Amersham Corp. (Les Ulis, France). Restriction enzymes, Klenow fragment of *Escherechia coli* DNA polymerase and T4 DNA ligase were purchased from Boehringer (Mannheim, Germany). The sequencing kit was from Pharmacia (Saint-Quentin en Yvelines, France).

Library screening. Preparation of poly(A) RNA and construction of the bovine pancreatic cDNA library using the plasmid pUC 9 as a cloning vector have previously been described (29). Screening was performed using a rat pancreatic carboxypeptidase A cDNA from our laboratory, radiolabelled by nick-translation (30). The filters were hybridized at 42°C as described in (31) and washed three times at 65°C for 45 min each with a solution containing 0.5 x SSC (75 mM NaCl, 7.5 mM sodium citrate, pH 7.0) and 0.1% sodium dodecyl sulfate.

Nucleotide sequence analysis. DNA was sequenced by the dideoxy chain termination procedure (32) using a sequencing kit from Pharmacia. The plasmid containing the preprocarboxypeptidase A cDNA insert was digested with various restriction enzymes (Fig. 1) and the resulting cDNA fragments were subcloned into M13 mp18 digested with the appopriate enzymes (33). The nucleotide sequence was determined in the case of both orientations and across all the restriction sites used in subcloning.

RESULTS AND DISCUSSION

The cDNA library from bovine pancreas was screened with a rat carboxypeptidase A cDNA which was previously ^{32}P-labelled. Only 14 hybridization-positive clones were isolated from 1500 transformants, what is rather surprising since procarboxypeptidase A is one of the main proteins in bovine pancreatic secretion (34). A single clone appeared to be full-length with an insert size of 1405 nucleotides including a poly(A) tail of 43 nucleotides. The strategy used for entirely sequencing the cloned cDNA is shown in Fig. 1.

The nucleotide sequence of bovine preprocarboxypeptidase A mRNA and the corresponding amino acid sequence are shown in Fig. 2. The open reading frame extends from the initiation codon

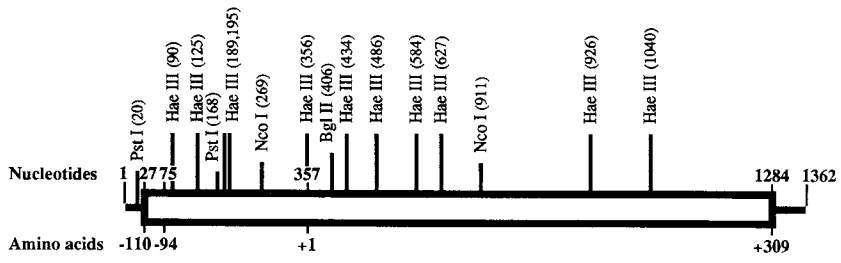

Figure 1. *Restriction endonuclease map of bovine pancreatic preprocarboxypeptidase A cDNA used for sequencing.* Horizontal rectangle delineates the coding region and lines extending from it represent the 5'- and 3'-noncoding regions, not including the poly(A) tail. The position of each restriction site is indicated.

ATG to the termination codon TGA and encodes a preproenzyme of 419 amino acids with a calculated molecular weight of 47,029. The 5'-noncoding region of bovine preprocarboxypeptidase A has 26 bases, whereas those of rat preprocarboxypeptidases A1, A2 and B contain 11, 28, and 19 nucleotides, respectively (Fig. 3). A high degree of identity exists in the 5'-noncoding segments of bovine preprocarboxypeptidase A and rat preprocarboxypeptidase A1 if a gap extending from nucleotide -5 to nucleotide -14 is introduced into the rat sequence (Fig. 3). In addition, the 3 nucleotides preceding the putative capping site in the rat preprocarboxypeptidase A sequence are identical to those occupying positions -22 to -24 in the bovine cDNA. Still, in the 5'-noncoding region, the presence of an adenosine at position -3 and two cytidines at positions -1 and -4 is in agreement with the consensus sequence for eukaryotic initiation sites (35). By contrast, the first nucleotide after the initiation codon is a cytidine instead of a purine as in bovine anionic trypsinogen mRNA (36), but contrary to what occurs in 83% of the eukaryotic messengers (37).

After the 5'-untranslated fragment, the open reading frame including the stop codon TGA is 1260 nucleotides in length like the rat preprocarboxypeptidase A1 mRNA (24), but in sharp contrast with rat preprocarboxypeptidase A2 (26) and B (25) mRNAs, in which the open reading frame consists of 1254 and 1248 nucleotides, respectively. Table 1 shows that the nucleotide sequences of bovine procarboxypeptidase A mRNA and rat procarboxypeptidase A1 mRNA had the highest degree of identity (81.3%). Beyond the stop codon, the 76-base-containing 3'-noncoding region includes a consensus polyadenylation signal AATAAA located 16 nucleotides upstream from a poly(A) tail at least 43 nucleotides in length.

The bovine preprocarboxypeptidase A mRNA encodes a protein containing a signal peptide of 16 amino acids, an activation peptide of 94 amino acids and a mature carboxypeptidase A of 309 amino

Figure 2. *Nucleotide sequence and deduced amino acid sequence of bovine pancreatic preprocarboxypeptidase A.* Signal and activation peptides extend from residue -110 to residue -95 and from residue -94 to residue -1, respectively. The N-terminal amino acid of mature carboxypeptidase A was numbered as 1. The two additional amino acids at the C-terminus and the four separate residues which were found to be different from those already reported (20,21) are circled and boxed, respectively. The polyadenylation signal (AATAAA) is underlined.

```
          AGCTGACCTTCCCAACTGACTGCAGC ATG CAG GGG CTG CTG ATT TTG AGT GTG CTG CTG
                                     met gln gly leu leu ile leu ser val leu leu
                                    -110                                   -100
GGG GCT GCC CTT GGC AAA GAG GAC TTT GTG GGC CAC CAG GTG CTC CGA ATC ACT GCC GCC
gly ala ala leu gly lys glu asp phe val gly his gln val leu arg ile thr ala ala
                    -94             -90                                    -80
GAT GAG GCC GAG GTG CAG ACG GTG AAG GAG CTG GAG GAC CTG GAG CAC CTG CAG TTG GAC
asp glu ala glu val gln thr val lys glu leu glu asp leu glu his leu gln leu asp
                                    -70                                    -60
TTC TGG AGG GGC CCT GGC CAG CCA GGC TCC CCC ATC GAC GTC CGA GTG CCC TTC CCC AGC
phe trp arg gly pro gly gln pro gly ser pro ile asp val arg val pro phe pro ser
                                    -50                                    -40
CTC CAG GCT GTT AAA GTC TTC CTG GAA GCC CAT GGC ATC AGA TAC AGG ATC ATG ATC GAG
leu gln ala val lys val phe leu glu ala his gly ile arg tyr arg ile met ile glu
                                    -30                                    -20
GAC GTG CAG TCC CTG CTA GAC GAG GAG CAG GAG CAG ATG TTC GCC TCC CAG AGC CGG GCC
asp val gln ser leu leu asp glu glu gln glu gln met phe ala ser gln ser arg ala
                                    -10                                -1   1
CGC AGC ACC AAC ACA TTT AAC TAC GCC ACC TAC CAC ACC CTG GAT GAG ATC TAT GAC TTC
arg ser thr asn thr phe asn tyr ala thr tyr his thr leu asp glu ile tyr asp phe
                                   10                                  20
ATG GAC CTG CTG GTG GCC GAG CAC CCA CAG CTT GTC AGC AAA CTC CAG ATT GGC AGA AGC
met asp leu leu val ala glu his pro gln leu val ser lys leu gln ile gly arg ser
                                   30                                  40
TAT GAA GGC CGT CCC ATC TAC GTG CTG AAG TTC AGC ACT GGG GGA AGC AAC CGT CCA GCC
tyr glu gly arg pro ile tyr val leu lys phe ser thr gly gly ser asn arg pro ala
                                   50                                  60
ATC TGG ATC GAC TTA GGC ATC CAT TCC AGG GAG TGG ATC ACC CAG GCC ACT GGG GTC TGG
ile trp ile asp leu gly ile his ser arg glu trp ile thr gln ala thr gly val trp
                                   70                                  80
TTT GCA AAG AAG TTC ACA GAA GAC TAT GGC CAG GAC CCG AGT TTC ACC GCC ATT CTT GAC
phe ala lys lys phe thr glu asp tyr gly gln asp pro ser phe thr ala ile leu asp
                                   90                                 100
AGC ATG GAC ATA TTC TTG GAG ATT GTC ACC AAC CCT GAT GGT TTT GCC TTC ACC CAC AGC
ser met asp ile phe leu glu ile val thr asn pro asp gly phe ala phe thr his ser
                                  110                                 120
CAG AAT CGA TTG TGG CGC AAG ACT CGA TCT GTC ACG TCA AGC TCC CTC TGT GTT GGG GTG
gln asn arg leu trp arg lys thr arg ser val thr ser ser ser leu cys val gly val
                                  130                                 140
GAC GCC AAC CGG AAC TGG GAT GCC GGC TTT GGG AAG GCA GGA GCC AGC AGC AGC CCC TGC
asp ala asn arg asn trp asp ala gly phe gly lys ala gly ala ser ser ser pro cys
                                  150                                 160
TCG GAG ACT TAT CAT GGC AAG TAT GCC AAT TCT GAA GTG GAG GTC AAG TCC ATC GTG GAC
ser glu thr tyr his gly lys tyr ala asn ser glu val glu val lys ser ile val asp
                                  170                                 180
TTT GTG AAA GAC CAT GGG AAC TTC AAG GCC TTC CTC TCC ATC CAC AGC TAC TCC CAG CTC
phe val lys asp his gly asn phe lys ala phe leu ser ile his ser tyr ser gln leu
                                  190                                 200
CTC CTC TAT CCC TAT GGC TAC ACA ACA CAA TCA ATC CCT GAC AAG ACT GAG CTG AAT CAG
leu leu tyr pro tyr gly tyr thr thr gln ser ile pro asp lys thr glu leu asn gln
                                  210                                 220
GTG GCT AAG TCC GCT GTT GAG GCC CTG AAG TCT CTG TAT GGG ACC AGC TAC AAG TAT GGC
val ala lys ser ala val glu ala leu lys ser leu tyr gly thr ser tyr lys tyr gly
                                  230                                 240
AGC ATC ATC ACA ACA ATT TAC CAA GCC AGT GGA GGC AGC ATT GAC TGG TCC TAC AAC CAA
ser ile ile thr thr ile tyr gln ala ser gly gly ser ile asp trp ser tyr asn gln
                                  250                                 260
GGC ATC AAG TAC TCC TTC ACC TTT GAA CTC CGG GAC ACG GGG CGC TAT GGC TTC CTG CTG
gly ile lys tyr ser phe thr phe glu leu arg asp thr gly arg tyr gly phe leu leu
                                  270                                 280
CCA GCC TCC CAG ATC ATC CCC ACA GCC CAG GAG ACG TGG CTG GGG GTT CTG ACC ATC ATG
pro ala ser gln ile ile pro thr ala gln glu thr trp leu gly val leu thr ile met
                                  290                                 300
GAG CAC ACG TTG AAT AAC CTC TAC TGA CCCAGCCCTCCAGCACTCTTTTCCTCCTCCTCTTCAGCCCTAC
glu his thr leu asn asn leu tyr
                              309
CCAGGCAACCAAATAAAGTCTGAGAGTACATGGAAAAAA.....
```

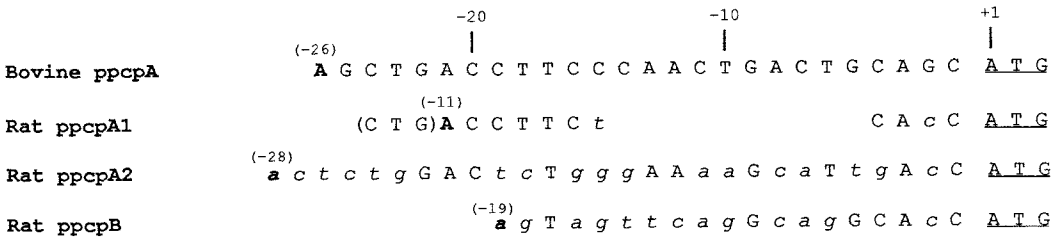

Figure 3. *Comparison of the 5'-noncoding regions of four pancreatic preprocarboxypeptidase (ppcp) mRNAs.* A gap was introduced into the rat ppcp A1 sequence to optimize sequence identity. Nucleotides which are unchanged in the 4 mRNAs are in capital letters, only. Transcription initiation sites are indicated by thick letters, and the numbers just above indicate the length of the corresponding 5'-noncoding region. References: ppcpA1 (24,25), ppcpA2 (26) and ppcpB (25).

acids. The signal peptide terminates with the glycine residue at position -95, as compared with the rat preprocarboxypeptidase mRNA sequences, and as a result of the presence of a lysine residue at position -94 which is the N-terminal residue of the bovine procarboxypeptidase A activation peptide (22,23). Moreover, glycine is known to be the most frequent C-terminal residue in eukaryotic signal peptides (37). The amino acid sequence of the activation peptide, as deduced from the nucleotide cDNA sequence, is identical to that previously reported (22,23). By contrast, the mature enzyme contains two additional amino acids at the C-terminus (Leu-308 and Tyr-309) and four amino acids with different charges (Asp-89, Asp-114, Gln-122, and Asp-185 replacing Asn-89, Asn-114, Glu-122, and Asn-185, respectively), as compared with the amino acid sequence reported about 20 years ago (20,21). Thus, bovine procarboxypeptidase A and carboxypeptidase A have molecular weights of 45,480 and 37,721 daltons, respectively, and the same number of residues as rat carboxypeptidase A1 (24), whereas rat carboxypeptidase A2 contains 307 amino acids only (26). The overall amino

TABLE 1. *Nucleotide and amino acid identity between rat and bovine preprocarboxypeptidases.* The percentage of nucleotide and amino acid identity existing between each pair of preprocarboxypeptidases is given in the upper right and in the lower left portions of the table, respectively. Each nucleotide or amino acid deletion was counted as a difference. These values were obtained assuming the presence of 1260 nucleotides and 419 amino acids in the case of preprocarboxypeptidases A (ppcp A) comparisons, and of 1275 nucleotides and 424 amino acids in the case of preprocarboxypeptidase A / preprocarboxypeptidase B (ppcp B) comparisons. References: ppcp A1 (24), ppcp A2 (26) and ppcp B (25).

(Nucleotides) / *(Amino acids)*		Bovine ppcp A	Rat ppcp A1	Rat ppcp A2	Rat ppcp B
Bovine	ppcp A	_	81.3	62.6	49.1
Rat	ppcp A1	78.3	_	61.7	49.7
Rat	ppcp A2	61.6	59.7	_	52.5
Rat	ppcp B	38.9	40.1	40.6	_

acid sequence corresponding to the bovine preprocarboxypeptidase A exhibits 78.3% identity with that of the rat preprocarboxypeptidase A1 (Table 1). In a similar manner, the signal peptide and activation peptide of bovine preprocarboxypeptidase A share 62.5% and 74.5% identity with those of rat preprocarboxypeptidase A1, respectively. On the other hand, the degree of identity is only 37.5% and 31.3% between the signal peptide of bovine preprocarboxypeptidase A and those of rat preprocarboxypeptidase A2 and B, respectively.

In conclusion, bovine preprocarboxypeptidase A is more closely related to rat preprocarboxypeptidase A1 as regards the number of amino acids as well as both the nucleotide and amino acid sequences. The identity existing between bovine and rat preprocarboxypeptidases A is stronger than that existing between preprocarboxypeptidases A1 and A2 in the rat. In addition, the existence of a high degree of identity in the 5' noncoding sequence between bovine and rat preprocarboxypeptidases A may be consistent with comparable, if not similar, interactions occuring between these regions and ribosomes during protein synthesis.

ACKNOWLEDGMENT

This work was partly supported by a research grant from the Conseil Régional de Bretagne.

REFERENCES

1 Keller, P.J., Cohen, E., and Neurath, H. (1958) J. Biol. Chem. 230, 905-915.
2 Yamasaki, M., Brown, J.R., Cox, D.J., Greenshields, R.N., Wade, R.D., and Neurath, H. (1963) Biochemistry 2, 859-866.
3 Puigserver, A., Vaugoyeau, G., and Desnuelle, P. (1972) Biochim. Biophys. Acta 276, 519-530.
4 Kobayashi, Y., Kobayashi, R., and Hirs, C.H.W. (1981) J. Biol. Chem. 256, 2466-2470.
5 Kerfelec, B., Chapus, C., and Puigserver, A. (1985) Eur. J. Biochem. 151, 515-519.
6 Kobayashi, R., Kobayashi, Y., and Hirs, C.H.W. (1978) J. Biol. Chem. 253, 5526-5530.
7 Martinez, M.C., Avilés, F.X., Sansegundo, B., and Cuchillo, C.M. (1981) Biochem. J. 197, 141-147.
8 Yoneda, T. (1980) Comp. Biochem. Physiol. 67B, 81-86.
9 Pascual, R., Burgos, F.J., Salva, M., Soriano, F., Mendez, E., and Avilés, F.X. (1989) Eur. J. Biochem. 179, 609-616.
10 Peanasky, R.J., Gratecos, D., Baratti, J., and Rovery, M. (1969) Biochim. Biophys. Acta 181, 82-92.
11 Keil-Dlouha, V., Puigserver, A., Marie, A., and Keil, B. (1972) Biochim. Biophys. Acta 276, 531-535.
12 Puigserver, A., and Desnuelle, P. (1975) Proc. Natl. Acad. Sci. USA 72, 2442-2445.
13 Kerfelec, B., Cambillau, C., Puigserver, A., and Chapus, C. (1986) Eur. J. Biochem. 157, 531-538.
14 Pascual, R., Vendrell, J., Avilés, F.X., Bonicel, J., Wicker, C., and Puigserver, A. (1990) FEBS Lett. in press.
15 Brown, J.R., Yamasaki, M., and Neurath, H. (1963) Biochemistry 2, 877-886.
16 Walsh, K.A., Ericsson, L.H., and Neurath, H. (1966) Proc. Natl. Acad. Sci. USA 56, 1339-1344.
17 Scheele, G. (1986) In The Exocrine Pancreas: Biology, Pathobiology, and Diseases (Go, V.L.W *et al.*, eds.) pp 185-192. Raven press, New York.

18 Vilanova, M., Vendrell, J., Lopez, M.T., Cuchillo, C.M., and Avilés, F.X. (1985) Biochem. J. 229, 605-609.
19 Burgos, F.J., Pascual, R., Vendrell, J., Cuchillo, C.M., and Avilés, F.X. (1989) J. Chromatogr. 481, 233-244.
20 Bradshaw, R.A., Ericsson, L.H., Walsh, K.A., and Neurath, H. (1969) Proc. Natl. Acad. Sci. USA 63, 1389-1394.
21 Pétra, P.H., Hermodson, M.A., Walsh, K.A., and Neurath, H. (1971) Biochemistry 10, 4023-4025.
22 Wade, R.D., Hass, G.M., Kumar, S., Walsh, K.A., and Neurath, H. (1988) Biochimie 70, 1137-1142.
23 Foglizzo, E., Bonicel, J., Kerfelec, B., Granon, S., and Chapus, C. (1988) Biochim. Biophys. Acta 954, 183-188.
24 Quinto, C., Quiroga, M., Swain, W.F., Nikovits Jr., W.C., Standring, D.N., Pictet, R.L., Valenzuela, P., and Rutter, W.J. (1982) Proc. Natl. Acad. Sci. USA 79, 31-35.
25 Clauser, E., Gardell, S.J., Craik, C.S., Mac Donald, R.J., and Rutter, W.J. (1988) J. Biol. Chem. 263, 17837-17845.
26 Gardell, S.J., Craik, C.S., Clauser, E., Goldsmith, E.J., Stewart, C.-B., Graf, M., and Rutter, W.J. (1988) J. Biol. Chem. 263, 17828-17836.
27 Gardell, S.J., Craik, C.S., Hilvert, D., Urdea, M.S., and Rutter, W.J. (1985) Nature 317, 551-555.
28 Gardell, S. J., Hilvert, D., Barnett, J., Kaiser, E.T., and Rutter, W. J. (1987) J. Biol. Chem. 262, 576-582.
29 Le Huërou, I., Wicker, C., Guilloteau, P., Toullec, R., and Puigserver, A. (1990) Biochim. Biophys. Acta 1048, 257-264.
30 Rigby, P.W.J., Dieckman, M., Rhodes, C., and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
31 Mueckler, M.M., Merrill, M.J., and Pitot, H.C. (1983) J. Biol. Chem. 258, 6109-6114.
32 Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
33 Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982) Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Labor, Cold Spring Harbor, NY.
34 Keller, P.J., Cohen, E., and Neurath, H. (1958) J. Biol. Chem. 233, 341-349.
35 Kozak, M. (1981) Nucleic Acids Res. 9, 5233-5252.
36 Le Huërou, I., Wicker, C., Guilloteau, P., Toullec, R., and Puigserver, A. (1990) Eur. J. Biochem., in press.
37 Watson, M.E.E. (1984) Nucleic Acids Res. 13, 5145-5164.