Biochem. J. (1992) 287, 299–303 (Printed in Great Britain)

cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1

Lluís CATASÚS,* Virtudes VILLEGAS,* Roger PASCUAL,* Francesc X. AVILÉS,*
Catherine WICKER-PLANQUART† and Antoine PUIGSERVER†‡
*Departament de Bioquímica (Facultat de Ciències) and Institut de Biologia Fonamental, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain, and †Centre de Biochimie et de Biologie Moléculaire du CNRS, 31 Chemin Joseph-Aiguier, 13402 Marseille Cedex 9, France

Using polyclonal antibodies raised against human pancreatic procarboxypeptidases, a full-length cDNA coding for an A-type proenzyme was isolated from a λgt11 human pancreatic library. This cDNA contains standard 3′ and 5′ flanking regions, a poly(A)$^{+}$ tail and a central region of 1260 nucleotides coding for a protein of 419 amino acids. On the basis of sequence comparisons, the human protein was classified as a procarboxypeptidase A1 which is very similar to the previously described A1 forms from rat and bovine pancreatic glands. The presence of the amino acid sequences assumed to be of importance for the zymogen inhibition by its activation segment, primarily on the basis of the recently reported crystal structure of the B form, further supports the proposed classification.

INTRODUCTION

The fact that pancreatic procarboxypeptidases A and B are the precursors of active enzymes with separate specificities has now been clearly established [1]. Moreover, the existence of further isomorphism within each of the zymogen forms has been demonstrated at the sequence level in the case of only a few species [2–4]. Three different native forms of human pancreatic procarboxypeptidase A have been isolated by means of anion-exchange h.p.l.c., and partly characterized by *N*-terminal extended sequence analysis [4]. The A1 and A2 forms are monomeric procarboxypeptidases, whereas A3 is a non-covalent association of procarboxypeptidase A and proproteinase E. In view of the differences between their *N*-terminal sequences, ability to hydrolyse synthetic peptide substrates, activation processes and susceptibility to thermal denaturation, the enzymes generated from the two corresponding monomeric zymogens A1 and A2 were actually assigned to separate subfamilies of carboxypeptidases A, the properties of the A2 form being midway between those of the A1 and B forms. However, the exact degree of structural identity among the three A isoforms is still a point which remains to be elucidated by further structural studies.

Nucleotide sequencing of the procarboxypeptidase cDNAs is a convenient method for obtaining the complete amino acid sequence of these rather high-molecular-mass pancreatic proteins (over 400 residues), since for evolutionary reasons their native forms are difficult to isolate. This approach has been successfully applied by Rutter and co-workers [3,5] for characterizing procarboxypeptidases A1 and A2 from the rat pancreas and, on the basis of enzymic and sequence data on the corresponding enzymes, for classifying them in the appropriate protein subfamilies. The above-mentioned isomorphism in human and rat procarboxypeptidases A is in contrast with the fact that, in bovine pancreatic cDNA libraries, a single preprocarboxypeptidase A cDNA, which is more like the A1 than the A2 type, has been cloned and sequenced [6]. The primary structure of bovine carboxypeptidase A isolated from autolysed pancreatic tissue was described more than 20 years ago [7], and only two forms sharing very similar sequences and enzymic properties have been described in the case of this enzyme [8]. As regards the other mammalian procarboxypeptidases, such as the extensively studied porcine procarboxypeptidases A, no evidence has been obtained so far that isomorphism of this kind occurs. All of the biochemical and biophysical studies published on porcine procarboxypeptidases A were carried out on the A1 form [9,10], including its preliminary three-dimensional crystal structure [11], the topology of which is similar to that reported for the related B form [12].

In this paper we describe the amino acid sequence of a human pancreatic preprocarboxypeptidase A which is derived from a full-length cDNA clone and can definitely be said to be a member of the A1 subfamily. Along with evolutionary considerations, the present data allowed us to identify more closely the structural factors determining the active site of these zinc-containing exopeptidases, as well as those responsible for proenzyme activation and subsequent inhibition of the resulting enzyme by the activation segment.

MATERIALS AND METHODS

Materials

Affinity-purified goat anti-rabbit IgG (H+L)-horseradish peroxidase conjugated human IgG was purchased from Bio-Rad (Madrid, Spain). [α-^{35}S]dATP (> 37 TBq/mmol) was from Amersham Corp. (Les Ullis, France). Restriction endonucleases, Klenow fragment of *Escherichia coli* DNA polymerase and T4 DNA ligase were from Boehringer (Mannheim, Germany). The human pancreatic cDNA library constructed in bacteriophage λgt11 was generously provided by M. E. Lowe (Washington University, U.S.A.).

Preparation of antibodies

Anti-(human pancreatic procarboxypeptidase) sera were obtained by immunizing rabbits (strain NZW) with 500 μg of purified human procarboxypeptidases A and B, previously isolated as described in [4]. The resulting IgGs were purified by ammonium sulphate precipitation followed by chromatography

‡ To whom correspondence should be addressed.
The nucleotide sequence data reported have been submitted to the EMBL, DDBJ and GenBank database under accession no. X67318.

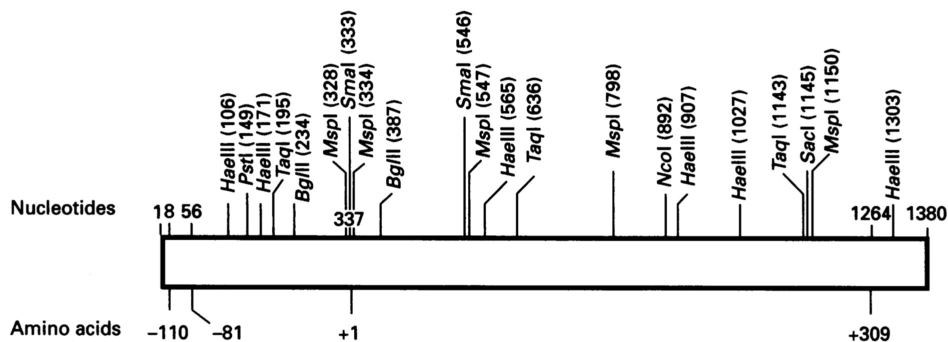

Fig. 1. Restriction endonuclease map of human pancreatic preprocarboxypeptidase A1 cDNA used for sequencing

The position of each restriction site is indicated. The following 5′→3′ fragments (1–195, 149–423, 387–570, 546–816, 636–927, 892–1089, 1027–1230 and 1145–1380), and 3′→5′ fragments (1380–1101, 1145–892, 907–735, 798–547, 565–315, 333–156 and 195–1) were sequenced.

on a DEAE-Sepharose column equilibrated with 20 mM-potassium phosphate buffer (pH 6.8).

Library screening

After the λgt11 human pancreatic cDNA library had been grown on a lawn of *E. coli* Y1090 R^- [Δlac u169 $proA^+$ Δlon ara D139 strA supF (Trp C22::Tn10) (pMC9)], screening was performed with antibodies directed against procarboxypeptidases using the procedure described in [13]. Briefly, the plates were incubated at 42 °C for 3.5 h, then lacZ-directed gene expression was induced by placing an isopropylthiogalactoside-soaked nitrocellulose membrane over them. After an additional incubation at 37 °C for 3.5 h, the membrane was removed, rinsed, and any still-unsaturated protein binding sites were blocked by further incubation with BSA. The membrane was then incubated with the primary antibody against human procarboxypeptidases A and B, followed by a washing step and another incubation with horseradish peroxidase-conjugated second antibody. After a final wash, the colour development substrate (4-chloronaphthol) was added. The positive clones were located and removed from the original plates, and the recombinants were subsequently purified by repeating the same screening procedure until all the clones were positive.

Nucleotide sequence analysis

The λgt11 DNA recombinants were subcloned into the plasmid pUC9. Nucleotide sequencing was performed by means of the dideoxy chain-termination procedure [14] using an appropriate kit from Pharmacia. The plasmid containing the preprocarboxypeptidase A1 cDNA insert was digested with various restriction endonucleases, and the resulting cDNA fragments were subcloned into M13mp18. The strategy used for subcloning is illustrated in Fig. 1. The nucleotide sequence was determined in both orientations and across the limits of all cDNA pieces analysed.

RESULTS AND DISCUSSION

The full-length preprocarboxypeptidase A cDNA isolated from the human pancreatic library contains a 7-nucleotide 5′-non-coding region, including the consensus sequence for eukaryotic transcription initiation sites [15] (two C at positions −1 and −4, and one A at position −3), followed by an open reading frame of 1260 nucleotides coding for a protein of 419 amino acids (Fig. 2). A coding region 1260 nucleotides in length was also found to exist in both the rat preprocarboxypeptidase A1 cDNA [16] and the bovine preprocarboxypeptidase A cDNA [6]. It is worth stressing here that, in contrast with most eukaryotic messengers [17], the fourth nucleotide of the coding region is a cytidine, as in bovine preprocarboxypeptidase A [6] and anionic pretrypsinogen [18]. In the 3′-non-coding region, a 54-nucleotide segment extending from the stop codon to the consensus polyadenylation site (AATAAA) was detected in both the human and the bovine precursor cDNAs, while the $poly(A)^+$ tail is located 36 nucleotides downstream in the former, compared with 16 nucleotides in the latter.

Table 1 shows that 12 amino acids out of 16 (75%) in the signal peptide of human A1 and bovine A preprocarboxypeptidases are the same. This degree of similarity increases to 85% in the activation region (80 amino acids out of 94) and to 81% in the carboxypeptidase region (250 amino acids out of 309).

Comparisons between the amino acid sequences corresponding to the regions which are thought to be functionally important in procarboxypeptidases provided some clues as to the identity of the human proenzyme studied here, as well as on the degree of allowed variability of these regions in the polypeptide chain. This comparison is presented in Fig. 3. Of particular interest is the very high degree of identity existing around the residues involved in enzyme–substrate interactions which are thought to be important for the enzyme specificity, such as Ser/Gly-253, Thr/Ser-254 and Thr/Ala-268 between the A1 and A2 forms [5], and Ile/Asp-255 between the A and B forms [3,5,12]. The presence of an Ile residue at position 255 in the amino acid sequence of the human procarboxypeptidase studied here clearly indicates that the pancreatic zymogen belongs to the A type and not to the B type, in which an Asp residue is located at the same position. In addition, the presence of Ser-253, Thr-254 and Thr-268 instead of Gly, Ser and Ala at the corresponding positions is consistent with the substrate specificity of the A1 form, which is directed towards smaller amino acids, rather than with that of the A2 form, which is directed towards bulkier amino acids [3]. The enzymic experiments previously carried out with the corresponding active enzyme [4] strengthened this hypothesis.

Sequence comparisons between the activation regions also indicate that the human A1 form actually resembles the rat A1 rather than the rat A2 form, and differs considerably from the B form. The main features of the recently determined crystal structure of porcine procarboxypeptidase B [12], as well as the preliminary report on the crystal structure of porcine procarboxypeptidase A [11], were of great help in making these comparisons. Thus Asp-A41 (where A denotes a residue in the activation segment), a key residue in the inhibition of procarboxypeptidase B by establishing a salt link with Arg-145 of the enzyme region [12] and therefore blocking the anchorage of peptide substrates to the bottom of the active site, is replaced by an apolar residue

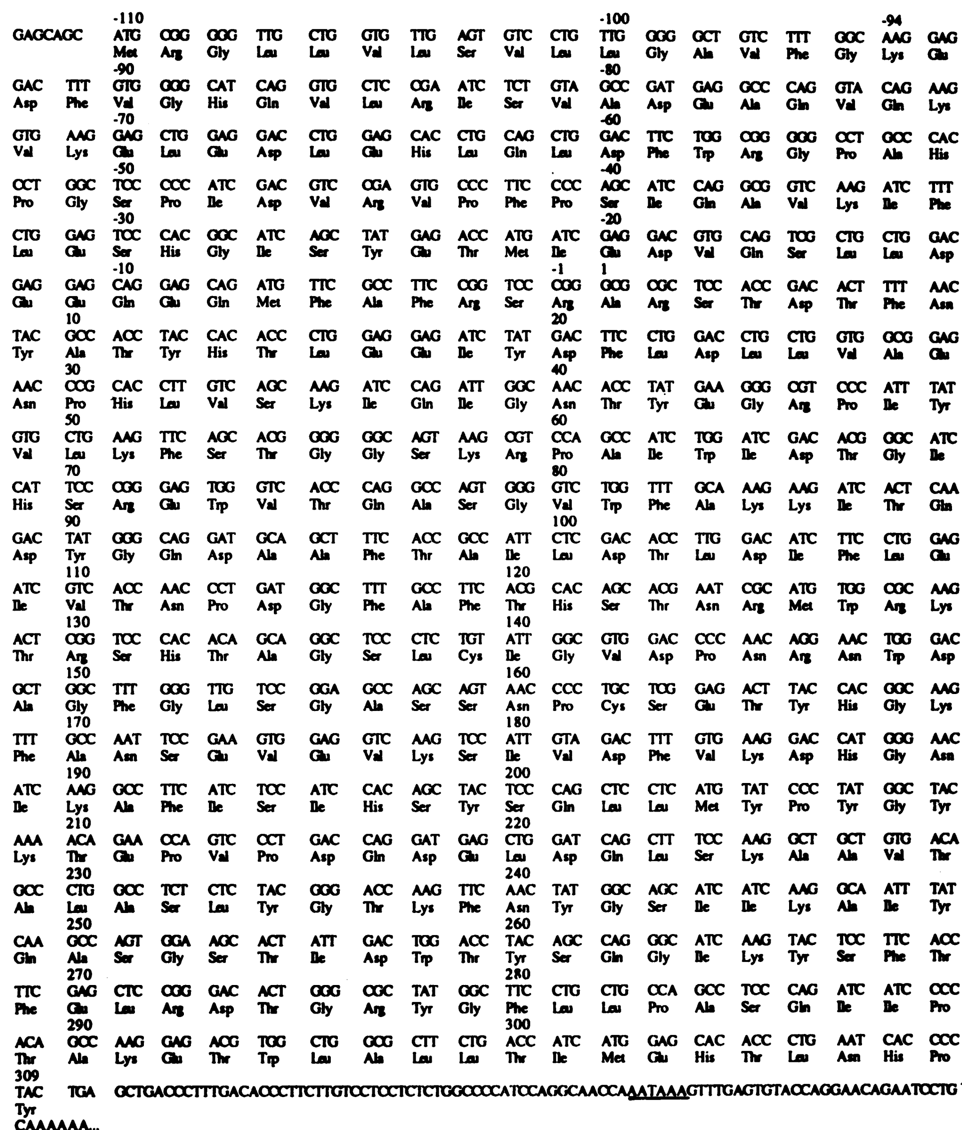

Fig. 2. Nucleotide sequence and derived amino acid sequence of human pancreatic preprocarboxypeptidase A1

The signal and activation regions extend from residues −110 to −95 and from residue −94 to −1, respectively. The polyadenylation signal (AATAAA) is underlined.

Table 1. Amino acid identity between human and bovine preprocarboxypeptidases as deduced from the nucleotide sequences

The comparison was between the amino acid sequences of human preprocarboxypeptidase A1 (this study) and bovine preprocarboxypeptidase A [6].

Amino acid sequence	Number of residues	Identity (%)	Charge change* (number)	Net charge difference (number)
Signal peptide	16	75	1	1
Activation peptide	94	85	6	1
Active enzyme	309	81	18	2
Proenzyme	403	82	24	3
Preproenzyme	419	82	25	4

* Residues at the same position with distinct charges.

in both the rat and human A1 forms. Other non-conservative substitutions involve Gln-A45 and Lys-A47, two residues of procarboxypeptidase B containing side chains which are involved in hydrogen bonding with carbonyls 249 and 245/247 of the enzyme, thus keeping the Tyr-248 side chain in an 'up' position. By contrast, the triad Asp-A36, Phe-A37 and Trp-A38, which is assumed to be important for binding of the activation segment to the enzyme region [11,12], is conserved in all procarboxypeptidase A and B amino acid sequences described so far.

The conservation in the physico-chemical characteristics of the residues located at the above-mentioned potentially important positions in the amino acid sequence of rat, bovine and human A1 procarboxypeptidase strongly supports their proposed role in the inhibition of these proenzymes. This is the case with the hypothesis that mainly attributes to the presence of an acidic residue in position 41 of the activation segment (Asp-A41) the lack of activity of porcine procarboxypeptidase B towards small peptide substrates, compared with the intrinsic activity exhibited by porcine procarboxypeptidase A [11,12]. Previous enzymic studies have indicated that similar activity differences also occur with isolated human proenzymes [4].

On the other hand, it is worth mentioning that human

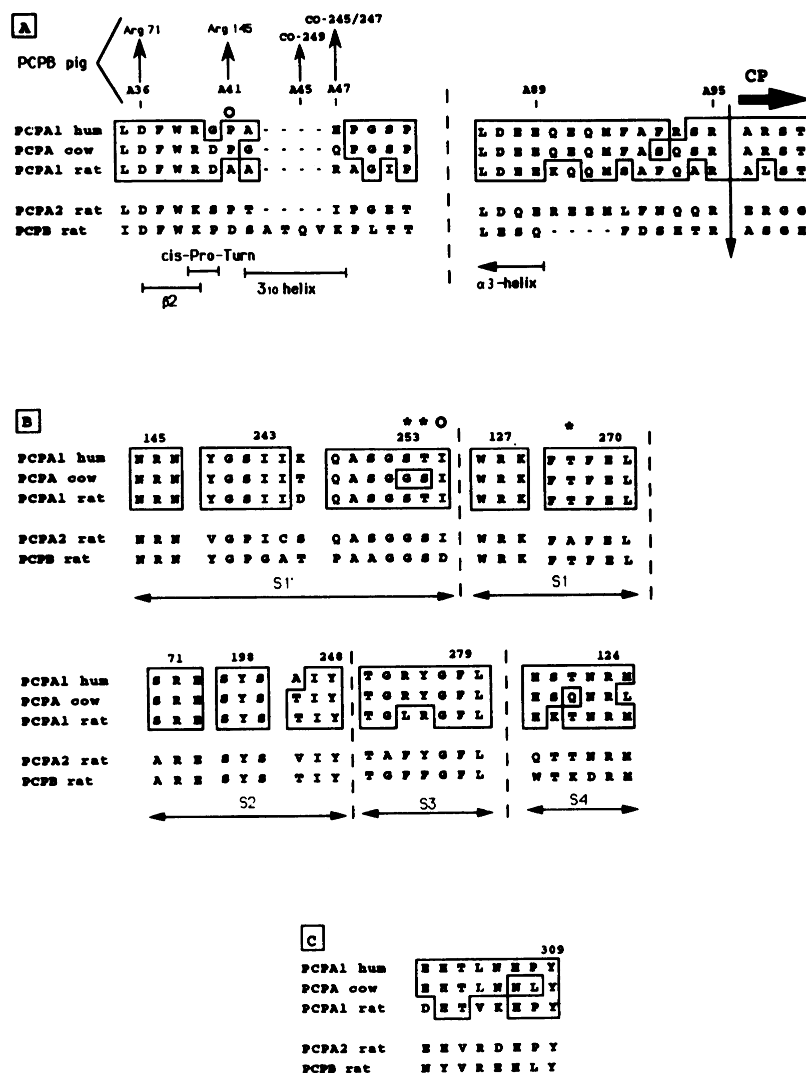

Fig. 3. Comparison between the deduced amino acid sequence of human procarboxypeptidase A1 and critical regions of other pancreatic procarboxypeptidases

Identical sequences have been boxed for internal comparisons among the A1 forms only. (*a*) Activation regions showing the sequences which seem to be essential for the inhibition of these zymogens during their proteolytic activation. Some residues of the enzyme region involved in the porcine procarboxypeptidase B (PCPB) inhibition mechanism, as deduced from its crystal structure [12], are indicated at the top. The secondary structures observed along the corresponding region of the same form are depicted at the bottom. The numbering of the porcine B proenzyme is that used. (*b*) Sequences around the residues which may form the various active centre subsites (S'1, S1, S2, S3). (*c*) *C*-Terminal region. * indicates significant substitutions correlated with the A1/A2 divergence. ° Indicates significant substitutions correlated with the A/B divergence. The references for the previously determined sequence data are as follows: PCPA cow [6], PCPA1 rat [5], PCPA2 [3] and PCPB rat [5], (where PCP = procarboxypeptidase).

procarboxypeptidase A contains a connecting region between the activation domain and the enzyme which is much more like that of the bovine A and the rat A1 forms than either the rat A2 or B forms (Fig. 3). The amino acid sequence differences in this region known to be important for proteolytic activation of the zymogen appear to be quite extensive among the B, A2 and A1 forms, although the two latter forms show a considerable internal evolutionary preservation. The differences observed between the amino acid sequences in this connecting region, and the corresponding effects on their local tertiary structure, probably account for the great differences between the proteolytic activation of the A1, A2 and B forms [4,19,20].

Finally, the *C*-terminal region of these proenzymes also deserves a comment. On the basis of initial sequence determination data using chemical methods, 307 residues were assigned to bovine carboxypeptidase A [7]. Recently, two additional residues were found to exist at the *C*-terminus of the enzyme by analysis of the corresponding cDNA nucleotide sequence [6]. Since a similar length of sequence was obtained in the present study in the case of human procarboxypeptidase A1 (Fig. 3), the *C*-terminal sequence of this proenzyme is clearly very similar to that of the bovine A form and the rat A1 form, and differs from the rat A2 form and particularly the rat B form.

We are indebted to Mr. M. E. Lowe, Washington University, for kindly supplying us with the human cDNA library. This study was partly supported by Action Intégrée Franco-Espagnole 1990–70, by Acción Integrada Hispano-Francesa 1990–219, and by CICYT grant B1088-0456 (Ministerio de Educación, Spain). LL.C., V.V. and R.P. are the recipients of PFPI fellowships (Ministerio de Educación, Spain).

REFERENCES

1. Puigserver, A., Chapus, C. & Kerfelec, B. (1986) in Molecular and Cellular Basis of Digestion (Desnuelle, P., Sjöström, H. & Norén, O., eds.), pp. 235–247, Elsevier, Amsterdam

2. Scheele, G. (1986) in The Exocrine Pancreas: Biology, Pathobiology, and Diseases (Go, V. L. W., Brooks, J. P., Di Magno, E. P., Gardner, J. D., Lebenthal, E. & Scheele, G. A., eds.), pp. 185–192, Raven Press, New York
3. Gardell, S. J., Craik, C. S., Clauser, E., Goldsmith, E. J., Stewart, C.-B., Graf, M. & Rutter, W. J. (1988) J. Biol. Chem. **263**, 17828–17836
4. Pascual, R., Burgos, F. J., Salva, M., Soriano, F., Mendez, E. & Avilés, F. X. (1989) Eur. J. Biochem. **179**, 609–616
5. Clauser, E., Gardell, S. J., Craik, C. S., McDonald, R. J. & Rutter, W. J. (1988) J. Biol. Chem. **263**, 17837–17845
6. Le Huërou, I., Guilloteau, P., Toullec, R., Puigserver, A. & Wicker, C. (1991) Biochem. Biophys. Res. Commun. **175**, 110–116
7. Bradshaw, R. A., Ericsson, L. H., Walsh, K. A. & Neurath, H. (1969) Proc. Natl. Acad. Sci. U.S.A. **63**, 1389–1394
8. Petra, P. H., Bradshaw, R. A., Walsh, K. A. & Neurath, H. (1969) Biochemistry **8**, 2762–2768
9. Vendrell, J., Avilés, F. X., Genescà, E., SanSegundo, B., Soriano, F. & Mendez, E. (1986) Biochem. Biophys. Res. Commun. **141**, 517–523
10. Conejero-Lara, F., Sanchez-Ruiz, J. M., Mateo, P. L., Burgos, F. J. Vendrell, J. & Avilès, F. X. (1991) Eur. J. Biochem. **200**, 663–670
11. Guasch, A., Coll, M., Avilés, F. X. & Huber, R. (1992) J. Mol. Biol. **224**, 149–157
12. Coll, M., Guasch, A., Avilés, F. X. & Huber, R. (1991) EMBO J. **10**, 1–9
13. Mierendorf, R. C., Percy, C. & Young, R. A. (1987) Methods Enzymol. **152**, 458–469
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
15. Kozak, M. (1981) Nucleic Acids Res. **9**, 5233–5252
16. Quinto, C., Quiroga, M., Swain, W. F., Nikovits, W. C., Jr., Standring, B. N., Pictet, R. L., Valenzuela, P. & Rutter, W. J. (1982) Proc. Natl. Acad. Sci. U.S.A. **79**, 31–35
17. Watson, M. E. E. (1984) Nucleic Acids Res. **13**, 5145–5164
18. Le Huërou, I., Wicker, C., Guilloteau, P., Toullec, R. & Puigserver, A. (1990) Eur. J. Biochem. **193**, 767–773
19. Vendrell, J., Cuchillo, C. M. & Avilés, F. X. (1990) J. Biol. Chem. **265**, 6949–6953
20. Burgos, F. J., Salva, M., Villegas, V., Soriano, F., Mendez, E. & Avilés, F. X. (1991) Biochemistry **30**, 4082–4089

Received 30 December 1991/25 March 1992, accepted 30 March 1992