

cDNA encoding a 59 kDa homolog of ribosomal protein S6 kinase from rabbit liver

Beate Harmann and Manfred W. Kilimann

Institut für Physiologische Chemie, Abteilung für Biochemie Supramolekularer Systeme, Ruhr-Universität Bochum, D-4630 Bochum 1, FRG

Received 16 August 1990; revised version received 3 September 1990

We have isolated cDNA molecules encoding a protein with the characteristic sequence elements that are conserved between the catalytic domains of protein kinases. This protein is apparently a serine/threonine kinase and is most closely related to the amino-terminal half of the ribosomal protein S6 kinase II first characterized in *Xenopus* eggs (42% overall identity and 56% identity in the predicted catalytic domain). However, it clearly differs from S6 kinase II in that it has only one, rather than two predicted catalytic domains and a deduced molecular mass of 59 109 Da. We propose that it may be more related to, or identical with, the mitogen-inducible S6 kinase of molecular mass 65–70 kDa described in mammalian liver, mouse 3T3 cells and chicken embryos. Remarkable structural features of the cDNA-encoded polypeptide are a section rich in proline, serine and threonine residues that resemble the multiphosphorylation domains of glycogen synthase and phosphorylase kinase α subunit, and a characteristic tyrosine residue in the putative nucleotide-binding glycine cluster which, by analogy to cdc2 kinase, is a potential tyrosine phosphorylation site.

S6 kinase; Protein phosphorylation; Cell cycle; Mitogen; Insulin; Growth factor

1. INTRODUCTION

Protein phosphorylation is a major mechanism of signal transduction in eukaryotic cells, and protein kinases with broader or narrower substrate specificities that act on one another can be linked to form intricate signal cascades or networks. The fact that a number of growth-factor receptors and oncogene products are tyrosine kinases emphasizes a crucial role of protein phosphorylation at tyrosine residues in the control of cell growth and proliferation, but the signal pathways involved in their action are poorly understood.

The protein kinases form a large family of enzymes (with ca 100 known members at present) which display significant structural similarity in their catalytic domains of ca 250 amino acids [1]. In addition to enzymological and genetic approaches, these structural similarities therefore have been employed to identify and characterize new protein kinases by homology screening of DNA libraries. DNA sequences encoding different protein kinases may cross-hybridize under low-stringency conditions.

In the course of a low-stringency screen of a rabbit liver cDNA library with a phosphorylase kinase γ subunit cDNA [2], we obtained a weakly hybridizing clone encoding a novel protein kinase. Sequence comparison with other protein kinases indicates that this is

apparently a serine/threonine kinase which is most closely related to, yet distinct from, the ribosomal protein S6 kinase II originally cloned from *Xenopus* ovaries [3,4].

2. EXPERIMENTAL

Size-fractionated cDNA (larger than 1800 nucleotides (nt)) was synthesized as described [5] and inserted into λ gt10 with *Eco*RI linkers. A primary library of 700 000 plaques from rabbit liver was screened under reduced stringency with the phosphorylase kinase γ subunit cDNA (complete coding sequence plus 228 nt of 3'-noncoding sequence) from mouse muscle (donated by J. Chamberlain [2]). Hybridization was carried out at 25°C in 50% formamide/5 \times SSPE/5 \times Denhardt's solution/0.1% NaDodSO₄/boiled sonicated herring DNA (100 μ g/ml) and washing was performed at 40°C in 2 \times SSC/0.1% NaDodSO₄ (1 \times SSPE = 150 mM NaCl/10 mM sodium phosphate/1 mM EDTA/pH 7.4; 1 \times Denhardt's solution = 0.2 mg of Ficoll per ml/0.2 mg of polyvinylpyrrolidone per ml/0.2 mg of bovine serum albumin fraction V per ml; 1 \times SSC = 150 mM NaCl/15 mM sodium citrate, pH 7.0). DNA sequencing according to Sanger was carried out in both directions on restriction fragments subcloned into M13 vectors.

RNA from various rabbit tissues was isolated by the LiCl/urea method [6], subjected to oligo(dT) cellulose chromatography, denatured by glyoxal, resolved by agarose gel electrophoresis and transferred to a nylon membrane (Pall Biodyne). The RNA blot was hybridized to the complete cDNA probe labelled by nick translation at 42°C in 50% formamide/5 \times SSPE/5 \times Denhardt's solution/0.1% NaDodSO₄/100 μ g/ml herring DNA/50 μ g/ml yeast tRNA/50 μ g/ml poly(A) and washed at 60°C in 0.2 \times SSC/0.1% NaDodSO₄. The same hybridization and washing conditions (minus tRNA and poly(A)) were used for the rescreening of two additional liver libraries with the G3 cDNA we obtained from the first library.

Correspondence address: M.W. Kilimann, Institut für Physiologische Chemie, Ruhr-Universität Bochum, D-4630 Bochum 1, FRG

3. RESULTS AND DISCUSSION

Originally searching for liver isoforms of the catalytic γ subunit of phosphorylase kinase, we probed a rabbit liver cDNA library with a mouse muscle phosphorylase kinase γ subunit cDNA. Out of 700 000 primary clones containing size-selected cDNA (representing a whole library complexity of ca 3×10^6) we obtained, besides several phosphorylase kinase clones, a single weakly cross-hybridizing clone of 1794 nucleotides (nt) arbitrarily designated G3 (Fig. 1). Subsequent screening of two analogously prepared, but independent libraries of 250 000 clones each with this probe yielded 10 additional clones of very similar length.

A long open reading frame extends from the 5'-end of the G3 cDNA sequence to nucleotide 1602. We have assumed the first ATG beginning at nucleotide 28 as the translation start codon. Its sequence vicinity is in reasonable agreement with the Kozak consensus, CC(G,A)CCATG(G,A) [7]. It appears that the cDNA sequence is nearly complete at the 5' end and that there is no translation start site further upstream. The 11 clones that we have isolated all have inserts of 1700-2000 nt as judged from gel electrophoresis. The five longest inserts were sequenced at their 5' ends with the result that four are almost identical in length (differing at most by 8 nt) while one is 200 nt shorter. The 5' terminal 27 nt of the cDNA are very G + C rich (85%) like the 5'-untranslated leader sequences of many mRNAs. Premature termination of cDNA synthesis by a strong secondary structure upstream from nucleotide 1 of course remains a possibility.

The 3' end of the long open reading frame is shortly followed by several stop codons in all reading frames. Although the 3 clones that we have sequenced 3' terminally end with oligo(A) tracts of either 14 or 16 residues in identical positions, a polyadenylation signal near the 3' end of the cDNA is not evident. The actual polyadenylation site may therefore lie further downstream, accounting for the larger size of the major mRNA apparent from the Northern blot.

RNA blot analysis (Fig. 2) demonstrates that the mRNA represented by G3 is moderately abundant in all tissues tested. Abundance varies to some degree between tissues (highest in muscle and lowest in testis) but remains within the same order of magnitude. The major mRNA species has an apparent length of ca 2400 nt in all tissues, while minor bands correspond to molecular sizes of 2000, 3300, 3800, 4800 and 6000 nt.

The polypeptide encoded by the long open reading frame of G3 displays the characteristic consensus sequence elements of a protein kinase. It probably is a serine/threonine kinase according to the criteria defined by Hanks et al. [1] in subdomains VI and VIII (consensus sequences: amino acids 218-223, DLKPEN; and 255-263, GTXXYXAPE). When its sequence is compared to other protein kinases, similarity is highest to

| | |
|--|------|
| GCCTGTGGCTCGCGCGGTCGCGGCCATGAGGCCAGCAGCGAGCGCGGCGCTTTTACCAGCGCGCTGACTTC | 75 |
| M R R R R R R R D G F Y P A P D F | 16 |
| CGAGACAGGAAAGCTGAGACATGCTGGAGTGTGTGACATAGACCTGGCCAGCCAGGAGCGCGCGCTGTAG | 150 |
| R D R E A E D M A G V F D I D L D Q P E D A G S E | 41 |
| GATGAGCTGGAGGGGGGTCACTTAAATGAAAGCATGGACCATGGGGAGTTGGACCATACCAACTTGGCATG | 225 |
| D E L E E G G Q L N E S M D H G G V G P Y E L G M | 66 |
| GAACATGTGAGAAATTTGAAATCTCAGAACTAGTGTGACAGAGCGCCAGAAAAATCAGACAGAAATGTTTT | 300 |
| E H C E K F E I S E T S V N R G P E K I R P E C F | 91 |
| GAGTACTTCCGGTACTGTTGTAAGGGGGCTATGGAAAGTTTTCAGTACGGAAATACAGGAGCAAAACT | 375 |
| E L L R V L G K G G Y G K V F Q V R K V T G A N T | 116 |
| GGGAAATATTTGCTATGAAAGTCTTAAAGGCAATGATGTAAGAAATGCTAAGGATACAGCTCATGACAAA | 450 |
| G K I F A M K V L K K R A M I V R N A K D D T A H T K | 141 |
| GCAGAAGGAAATTTCTGGAGAGTAAAGCATCCCTCATTTGGATTTAATTTATGCTTCAGACCGGTGGA | 525 |
| A E R N I L E E V K H P F I V D L I Y A P Q T G G | 166 |
| AACTCTACCTCATCTTGAAGTACTCAGTGGAGGAGAATTTATTCAGTGTAGAAAGAGAGGGAATATTATG | 600 |
| K L Y L I L E Y L S G G E L F M Q L E R E G I F M | 191 |
| GAGATACAGCTGCTTTTACTGGCAGAAATCTCCATGGCTTTGGGGCATTACATCAAAAGGGATCATCTAC | 675 |
| E D T A C P Y L A E I S M A L G H L H Q K G I Y | 216 |
| AGAGACCTGAAGCCGAGAATATCATCTTAAATCAGGAGTCTGTAAGAACTAACAGACTTTGGACTCATGAAA | 750 |
| R D L K P E N I M L N H Q G H V K L T D F G L C K | 241 |
| GAATCCATTCATGTAAGCAGTCCACACACATTTTCCGGAACAATAGAAATACATGGCCCTGAAATCTTGATG | 825 |
| E S I H D G T V T H T F C G T I E Y M A P E I L M | 266 |
| AGAAATGCCACATCTGTCGGTGTGATGTGGAGTTTGGAGCATTAATGATGAGATGACTGGAGAGCC | 900 |
| R S G H N R A V D W W S L G A L M Y D M L T G A P | 291 |
| CCATTCAGTGGGGAATAGAAAGAAACAATTCAGCAAAATGTAAGAACTCAATTTGCTCCCTGACTC | 975 |
| P F T G E N R K K T I D K I L K C K L N L P P Y L | 316 |
| ACACAAGAGCCAGAGATCTCTTAAAAAGCTGCTGAAAAGAAAGCTTCTCAGCTTGGAGCTGGTCTGG | 1050 |
| T Q E A R D L L K K L L K R A A N N L G A G P G | 341 |
| GATGCTGGAGGTTCAAGCTCACCCATCTTCAGACACATTAATGGGAAAGCTTCTAGCTCGAAAGGTGGAG | 1125 |
| D A G E V I Q A H P F F R H I N W E E L L A R K V E | 366 |
| CGCCCTTTAAACCTCTGTTGCAATCTGAAGAGGATGTGAGTCAGTTTGAATCAAAAGTTTACAGCTCAGACACT | 1200 |
| P P F K P L L Q S E E D V S Q F D S K F T R O T P | 391 |
| GTTGACAGCCAGATGACTCAACTCTCAGTGAAGTGCACAGGCTCTCTGGGTTTACATATGTGGCTCCA | 1275 |
| V D S P D D S T L S E S A N Q V L F G F T Y V A P | 416 |
| TCTGTACTGAAAGTGTGAAAGAAAGTTTCTCTTGAACCAAAAATCGATCACCTCGAAGATTATTTGGCAGC | 1350 |
| S V L E S V K E K F S F E P K I R S P R R F I G S | 441 |
| CGACGACAGCTGTCAGCCGATCAAAATTTCTCTGGGATTTCTGGGAGAGAGTCTTCAGCCAGGAGCCCA | 1425 |
| P R T P V S P V K F S P G D F W G R G A S A S T A | 466 |
| AACTCTCAGACCTGTGAAATACCAATGGAACAAGTGGATGAGATGGATGTGACGACAAGTGGGGAA | 1500 |
| N P Q T P V E Y P M E T E T S G I E Q M D V T T S G E | 491 |
| GCATCAGCTCCACTCCGATACGACAGCCGAACTCCGGACCATCAAAAAACAAGCTTTTCCATGATCTGAAA | 1575 |
| A S A P L P I R Q P N S G P Y K K Q A F P M I S K | 516 |
| CGCCAGACAGCTCGTATGAAATCTATGACAGAGCAAGCTTCCAAATGATATGAGGTGAAAAAGGGGAGGGA | 1650 |
| R P E H L R M N L | 525 |
| GGTGTGAGCTCTGCAAAAGTGAACAAGACTCAAAATGACAGTCTCAGAGCTCAGTGTGTTACACAGAAC | 1725 |
| ACTTTGGCCAGGAGAAACAACATAGATTTTAAACTCAATGCTGCAAAAAAATAAAAAA | 1794 |

Fig. 1. Nucleotide and predicted amino acid sequences of G3. Prominent potential phosphorylation sites are marked (see text). This sequence is being deposited in the EMBL/GenBank data base under access number X54415.

the ribosomal protein S6 kinase II originally cloned from *Xenopus* ovaries and later from chicken and mouse embryonal fibroblasts [3,4] by Erikson and colleagues (42% overall identity with the chicken sequence). Substantial similarity is also apparent with the various forms of protein kinase C (36% overall identity with the ϵ type from rabbit brain, [8]). The 3 sequences are compared in Fig. 3. Similarity with the phosphorylase kinase γ subunit is rather low (20%). Detection of this clone with the phosphorylase kinase probe was quite coincidental; the longest contiguous sequence match of 21 nt responsible for hybridization with the mouse phosphorylase kinase cDNA is disrupted by differential codon usage in the rabbit phosphorylase kinase cDNA.

S6 kinase II has a characteristic dimeric structure predicting two catalytic domains. The protein encoded by the G3 cDNA is obviously distinct because it has only one catalytic domain. The G3-encoded kinase is clearly more similar to the N-terminal than to the C-

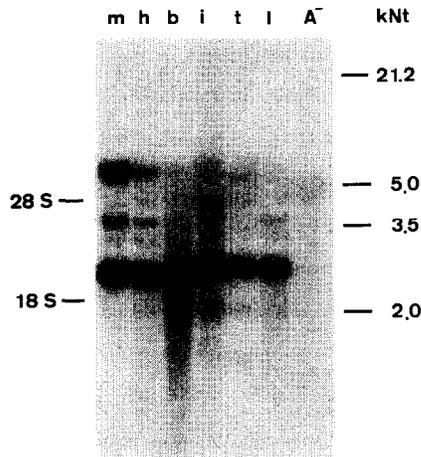


Fig. 2. RNA blot analysis. Poly(A)⁺ RNA (15 μg each) from rabbit skeletal muscle (m), heart (h), brain (b), small intestine (i), testis (t) and liver (l), and 15 μg poly(A)⁻ RNA from rabbit brain were resolved on a 1% agarose gel, transferred to a nylon membrane and probed with the complete G3 cDNA (see section 2). Film exposure with intensifying screens was for 15 h. Positions of glyoxylated rRNA standards (left) and double-stranded DNA standards (right) are indicated.

terminal domain of S6 kinase II (42% as compared to 23% overall identity) and, according to this alignment, is longer at the N-terminus and much shorter at the C-terminus. Moreover, the S6 kinase II sequences from frog, chicken and mouse are much more similar to each other (more than 80% identity) than to the G3-encoded kinase from rabbit.

While the protein kinase encoded by G3 is distinct from S6 kinase II, its sequence similarity suggests that it may be related to or identical with a different mitogen-inducible S6 kinase that has been described by several groups in mammalian liver, mouse 3T3 cells and chicken embryos but not yet structurally characterized [9–11] and references therein). Phosphorylation of ribosomal protein S6, accompanied by an increase in the rate of translation, is an early obligatory response of quiescent cells to agents that stimulate cell division, e.g. insulin and peptide growth factors, phorbol esters, or oncogene products (discussed in more detail e.g. in [4,11–13]). This other S6 kinase differs from S6 kinase II according to enzymological criteria and apparent molecular mass (65–70 kDa vs 90 kDa). The size of the G3-encoded kinase (59 kDa) comes close to that of the liver S6 kinase. The remaining discrepancy could be explained by the phosphate content of the native kinase, which is known to reduce its electrophoretic mobility [14]. The apparent molecular mass of S6 kinase II in gel electrophoresis (90–92 kDa) is also substantially higher than the cDNA-deduced molecular mass (80–83 kDa depending on the species [3,4,15]).

Comparison of the G3 kinase sequence with its two most closely related counterparts, S6 kinase II (S6KII) and protein kinase C, ϵ type (PKC), reveals a pattern of distinct domains (Fig. 3). Within the catalytic domain

of 260–270 amino acids, all 3 kinases are very similar (G3/S6KII: 56% identity; G3/PKC: 43% identity). Homology extends downstream from the catalytic domains for ca 65 amino acids (G3/S6KII: 42% identity; G3/PKC: 33% identity until the C-terminus of protein kinase C and the beginning of the second catalytic domain of S6 kinase II. Further downstream, G3 kinase and S6 kinase II sequences lose significant similarity (17% identity). Upstream from the catalytic domain towards the N-terminus, G3 kinase displays similarity only with S6 kinase II (26% identity) but not with protein kinase C (15% identity), and it extends beyond the amino terminus of S6 kinase II. This domain pattern is emphasized by a comparison with other PKC types (α , β I and γ): no similarity in the N-terminal domain (4–9% identity) but pronounced similarity in the catalytic domain (42–43%) and further downstream until the C-terminus of PKC (31–38%). The limits of these structural domains are marked by arrows in Fig. 3.

Many protein kinases are themselves regulated by phosphorylation, often on multiple sites. This holds also for the S6 kinases, upon which multiple signal pathways involving several serine/threonine and tyrosine kinases (directly or indirectly) seem to converge (reviewed in [16]). We have therefore inspected the G3-encoded sequence for potential phosphorylation sites. A prominent potential site for cAMP-dependent protein kinase [17] is not evident. There are a few potential target sites for casein kinases I or II with clusters of acidic residues in their vicinity [18], e.g. serines 40, 375, 380, 394 and 398. Serine 434 is surrounded by an excellent consensus for a cdc2 kinase recognition site [19]. These residues are marked in Fig. 1. The C-terminal PST domain (below) contains 8 potential target sites (Ser/Thr-Pro) for the growth factor activated, proline-directed protein kinase described recently [20].

A striking structural feature of higher order is an extended domain studded with proline (12.4%), serine (12.4%) and threonine (6.3%) residues (PST domain) that is reminiscent of the large multiphosphorylation domains in glycogen synthase [21] and phosphorylase kinase α subunit [22,23]. Whereas the N-terminal and catalytic domains of G3 kinase are rather poor in serine and proline, this changes abruptly at the carboxy-terminal edge of the catalytic domain (Fig. 3). The PST domain comprises sequences conserved between G3 kinase, S6 kinase II and protein kinase C as well as G3-specific sequences and might accommodate multiple phosphorylation sites.

The glycine cluster that forms part of the ATP binding site (subdomain I [1]) contains tyrosine (Y 102) in a position where most protein kinases carry phenylalanine. Among the minority of kinases that have tyrosine in this site is cdc2 kinase that plays a role in cell cycle regulation, and this tyrosine residue has recently been shown to be subject to phosphorylation

Acknowledgements: We thank Jeff Chamberlain for providing the phosphorylase kinase γ subunit cDNA, Helmut Meyer for help with the sequence data evaluation, and Ludwig Heilmeyer for bringing tyrosine 102 to our attention. This work was supported by the Deutsche Forschungsgemeinschaft.

REFERENCES

- [1] Hanks, S.K., Quin, A.M. and Hunter, T. (1988) *Science* 241, 42-52.
- [2] Chamberlain, J.S., VanTuinen, P., Reeves, A.A., Philip, B.A. and Caskey, C.T. (1987) *Proc. Natl. Acad. Sci. USA* 84, 2886-2890.
- [3] Jones, S.W., Erikson, E., Blenis, J., Maller, J.L. and Erikson, R.L. (1988) *Proc. Natl. Acad. Sci. USA* 85, 3377-3381.
- [4] Alcorta, D.A., Crews, C.M., Sweet, L.R., Bankston, L., Jones, S.W. and Erikson, R.L. (1989) *Mol. Cell. Biol.* 9, 3850-3859.
- [5] Zander, N.F., Meyer, H.E., Hoffmann-Posorske, E., Crabb, J.W., Heilmeyer Jr, L.M.G. and Kilimann, M.W. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2929-2933.
- [6] Auffray, C. and Rougeon, F. (1980) *Eur. J. Biochem.* 107, 303-314.
- [7] Kozak, M. (1986) *Cell* 44, 283-292.
- [8] Ohno, S., Akita, Y., Konno, Y., Imajoh, S. and Suzuki, K. (1988) *Cell* 53, 731-741.
- [9] Blenis, J., Kuo, C.J. and Erikson, R.L. (1987) *J. Biol. Chem.* 262, 14373-14376.
- [10] Tabarini, D., Garcia de Herreros, A., Heinrich, J. and Rosen, O.M. (1987) *Biochem. Biophys. Res. Commun.* 144, 891-899.
- [11] Kozma, S.C., Lane, H.A., Ferrari, S., Luther, H., Siegmann, M. and Thomas, G. (1989) *EMBO J.* 8, 4125-4132.
- [12] Sweet, L.J., Alcorta, D.A. and Erikson, R.L. (1990) *Mol. Cell. Biol.* 10, 2787-2792.
- [13] Meier, K.E., Weiel, J.E., Bloom, J.T. and Krebs, E.G. (1990) *J. Biol. Chem.* 265, 4635-4645.
- [14] Ballou, L.M., Siegmann, M. and Thomas, G. (1988) *Proc. Natl. Acad. Sci. USA* 85, 7154-7158.
- [15] Chen, R.-H. and Blenis, J. (1990) *Mol. Cell. Biol.* 10, 3204-3215.
- [16] Kozma, S.C., Ferrari, S. and Thomas, G. (1989) *Cell. Signal.* 1, 219-225.
- [17] Cohen, P. (1985) *Eur. J. Biochem.* 151, 439-448.
- [18] Kuenzel, E.A., Mulligan, J.A., Sommercorn, J. and Krebs, E.G. (1987) *J. Biol. Chem.* 262, 9136-9140.
- [19] Moreno, S. and Nurse, P. (1990) *Cell* 61, 549-551.
- [20] Hall, F.L., Mitchell, J.P. and Vulliet, P.R. (1990) *J. Biol. Chem.* 265, 6944-6948.
- [21] Picton, C., Aitken, A., Bilham, T. and Cohen, P. (1982) *Eur. J. Biochem.* 124, 37-45.
- [22] Kilimann, M.W., Zander, N.F., Kuhn, C.C., Krabb, J.W., Meyer, H.E. and Heilmeyer Jr, L.M.G. (1988) *Proc. Natl. Acad. Sci. USA* 85, 8381-9385.
- [23] Meyer, H.E., Meyer, G.F., Dirks, H. and Heilmeyer Jr, L.M.G. (1990) *Eur. J. Biochem.* 188, 367-376.
- [24] Steele, R.E. (1990) *Trends Biochem. Sci.* 15, 124-126.
- [25] Dayhoff, M.O., Schwartz, R.M. and Orcutt, B.C. (1978) in: *Atlas of Protein Sequence and Structure* (Dayhoff, M.O., ed) Vol. 5, Suppl. 3, pp. 345-352, Natl. Biomed. Res. Found., Silver Spring, MD.