Volume 264, number 2, 275–278 FEBS 08448 May 1990

A single amino acid difference distinguishes the human and the rat sequences of stathmin, a ubiquitous intracellular phosphoprotein associated with cell regulations

Alexandre Maucuer, Valérie Doye and André Sobel

INSERM U153, CNRS URA614, 17 rue du Fer à Moulin, 75005 Paris, France

Received 31 March 1990

Stathmin is a ubiquitous phosphoprotein proposed to play a general role as an intracellular relay integrating diverse regulatory signals of the cell's environment. We used a rat stathmin probe to isolate two classes of cDNAs coding for the human protein and corresponding to the usage of different polyadenylation sites. Compared to the rat sequences, they displayed a very high conservation both at the nucleic acid and the deduced protein sequence levels, with a single conservative amino acid difference. Further analysis of the protein sequence revealed novel putative phosphorylation sites, as well as internal repeated sequences which might reflect structural features involved in the molecular mechanisms by which stathmin fulfills its biological functions. The extreme conservation of the entire stathmin sequence further stresses the essential and general role of stathmin in cell regulations.

Stathmin; Phosphoprotein; Phylogenetic conservation; cDNA; Amino acid sequence

1. INTRODUCTION

Stathmin [1], also designated P19 [2], pp17 or prosolin [3] and p18 [4], is a ubiquitous protein whose phosphorylation in vivo was observed in numerous biological systems in response to extracellular regulatory agents. The regulations of its phosphorylation in diverse tissues and cell types, as well as its ontogenic expression profile, suggested that stathmin is physiologically involved in both 'developmental' and 'functional' regulations [5]. It may thus play a general role as an intracellular relay integrating the diverse second messenger pathways triggered by the various signals of the cell's environment [1].

Stathmin ($M_r \approx 19000$, $pI \approx 6.2–5.6$) is most abundant in neurons [6] and was purified from rat [1] and bovine [2] brain. Two isoforms, α and β, were identified [7], differing only by posttranslational modifications [8]. The corresponding unique rat cDNA was recently cloned from libraries of both the PC12 cell line [8] and testis [9]. The deduced amino acid sequence displayed a strong homology (65% amino acid identity) [9] with the developmentally regulated neuronal protein SCG10 [10].

On the basis of two-dimensional electrophoretic patterns and immunological cross-reactivity, stathmin appears to be well conserved throughout the evolution of vertebrates [5]. Partial protein sequencing of the bovine brain protein also suggested a very high conservation among mammals [9].

In order to further investigate stathmin conservation and to determine its potentially functionally important conserved regions, we cloned stathmin cDNAs from a human testis library. We show here that stathmin is extremely well conserved among mammals, since the rat and human sequences differ only by a single conservative amino acid change. Such a high conservation is likely related to the very general and ubiquitous role of stathmin in the biological regulations of cells.

2. MATERIALS AND METHODS

2.1. *cDNA library screening*

500000 phages of a λgt11 human testis cDNA library (Clontech) were screened with the multiprime labeled [11] |1–809| *Eco*RI-*Dra*I fragment of the rat stathmin cDNA pS62b clone [8].

Hybridizations were carried out at 51°C for 14 h in a solution containing 5 × SSC, 0.5% SDS, 20% formamide, 25 mM sodium phosphate, 2 mM EDTA and 100 μg/ml salmon sperm DNA. Final washes were performed at 55°C in 0.2 × SSC, 0.1% SDS.

2.2. *Subcloning and cDNA sequencing*

The DNA of 14 out of 70 positive recombinant phages was purified, digested with *Eco*RI, and analyzed by Southern blotting [12] with the rat cDNA probe. In all cases two bands were revealed indicating the existence of an internal *Eco*RI site. The cDNA inserts were extracted after either partial or total *Eco*RI digestion, and various inserts were subcloned in both directions into the plasmid vector Bluescript SK$^+$ (Stratagene, San Diego, CA).

Single-stranded DNA templates were prepared by infection of cultures with M13 helper phage K07. Nucleotide sequences were determined with the dideoxy chain terminator method [13] using syn-

Correspondence address: A. Maucuer, INSERM U153, 17 rue du Fer à Moulin, 75005 Paris, France

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/90/$3.50

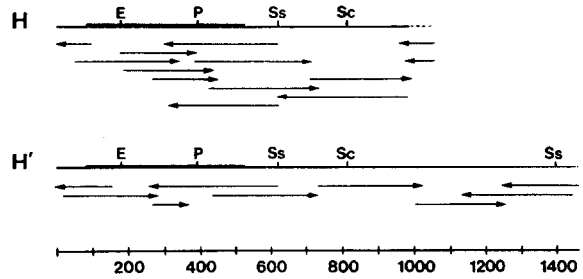

Fig. 1. Strategy for sequencing the cDNAs encoding human stathmin. Arrows correspond to the sequences determined with three 'short' cDNA inserts (H) and with two 'long' ones (H'). The open reading frame (dark bar) polyA tails (shaded line) and relevant restriction sites (E, *Eco*RI; P, *Pfl*mI; Ss, *Ssp*I; Sc, *Sca*I) are indicated.

thetic primers and a modified T7 DNA polymerase (Sequenase, US Biochemical Corp., Cleveland, OH).

2.3. *Southern blot*

Ten μg of *Bam*HI- or *Eco*RI-digested human genomic DNA were electrophoresed on 0.7% agarose gels, transferred to nylon membrane (Zeta Bind) and probed with the 1393–8171 *Pfl*mI-*Sca*I ^{32}P-labeled fragment of human stathmin cDNA as described [12].

3. RESULTS AND DISCUSSION

3.1. *Cloning of human stathmin cDNAs*

Testis being the tissue where stathmin is most expressed after brain, we screened a human testis cDNA library with a rat probe originally cloned from a PC12 cell library [8] and known to be identical to the corresponding rat testis probe [9].

At high stringency, positive clones were found at a frequency of about 1/10000, in good agreement with the expected expression of stathmin in testis as compared to brain [5,14]. Among the positive clones, two classes were found according to the size of their inserts: long ones, approximately 1450 bp, and short ones, approximately 1050 bp (Fig. 1).

```
R  GGTCCGAGCCGCCTGGCTTAGGCGGAGGCGACGGCAGGGTAGGGAGCAGCCGGGAACGCGGCGAGCCCCTAGCCGGGGACACGGGACGGGCGGAGACCTCGAGCCCTGGGCTTTCCTT-  118
H                                                                                                   GGCGGCTCGGACTGAG----------CAGGACTTTCCTTA  30

R  -GCCAGTGGATTGTGTAGAGTGTACAGCCAGTCTCTTGTCTTCTGTCCAACATGGCATCTTCTGATATTCAGGTGAAAGAGCTGGAGAAGCGAGCTTCCGGCCAGGCTTTTGAGCTGATT  237
H  TCCCAGTTGATTGTGCAGAATACACTGCCTGTCGCTTGTCTTCTATTCACCATGGCTTCTTCTGATATCCAGGTGAAAGAACTGGAGAAGCGTGCCTCAGGCCAGGCTTTTGAGCTGATT  150
                                                    M  A  S  S  D  I  Q  V  K  E  L  E  K  R  A  S  G  Q  A  F  E  L  I  23

R  CTCAGCCCTCGATCAAAAGAATCTGTCCCCGAGTTCCCCCTTTCCCCCCCAAAGAAGAAGGATCTTTCCCTGGAGGAAATTCAGAAGAAATTAGAAGCTGCAGAAGAAAGACGCAAGTCT  357
H  CTCAGCCCTCGGTCAAAAGAATCTGTTCCAGAATTCCCCCTTTCCCCTCCAAAGAAGAAGGATCTTTCCCTGGAGGAAATTCAGAAGAAATTAGAAGCTGCAGAAGAAAGACGCAAGTCC  270
   L  S  P  R  S  K  E  S  V  P  E  F  P  L  S  P  P  K  K  K  D  L  S  L  E  E  I  Q  K  K  L  E  A  K  E  E  R  A  K  S  63

R  CATGAAGCAGAAGTCTTGAAGCAGCTCGCTGAGAAGCGGGAGCATGAAAAAGAAGTGCTCCAGAAAGCCATTGAGGAGAACAACAACTTCAGCAAAATGGCAGAGGAGAAACTGACCCAC  477
H  CATGAAGCTGAGGTCTTGAAGCAGCTGGCTGAGAAACGAGAGCACGAGAAAGAAGTGCTTCAGAAGGCAATAGAAGAGAACAACAACTTCAGTAAAATGGCAGAAGAGAAACTGACCCAC  390
   H  E  A  E  V  L  K  Q  L  A  E  K  R  E  H  E  K  E  V  L  Q  K  A  I  E  E  N  N  N  F  S  K  M  A  E  E  K  L  T  H  103

                                                                                                    Val
R  AAAATGGAGGCTAACAAAGAGAACCGGGAGGCGCAAATGGCTGCCAAGCTGGAGCGTTTGCGAGAGAAGGACAAGCACGTTGAAGAGGTGCGGAAGAACAAAGAATCCAAAGACCCCGCG  597
H  AAAATGGAAGCTAATAAAGAGAACCGAGAGGCACAAATGGCTGCCAAACTGGAACGTTTGCGAGAGAAGGATAAGCACATTGAAGAAGTGCGGAAGAACAAAGAATCCAAAGACCCTGCT  510
   K  M  E  A  N  K  E  N  R  E  A  Q  M  A  A  K  L  E  R  L  R  E  K  D  K  H  I  E  E  V  R  K  N  K  E  S  K  D  P  A  143

R  GACGAGACCGAGGCTGACTAAGTTGTTCCGAGAACTGACTTTCTCCCCGACCCCTTCCTAAATATCCAAAGACTGTACTGGCCAGTGTCA-TTTACTTTTTCCTCCCTGACAAATATTCT  716
H  GACGAGACTGAAGCTGACTAATTTGTTCTGAGAACTGACTTTCTCCCCATCCCCTTCCTAAATATCCAAAGACTGTACTGGCCAGTGTCATTTTATTTTTTCCCTCCTGACAAATATTTT  630
   D  E  T  E  A  D                                                                                                              149

R  AGAAGCTGATGTAGGAACCGTATAGGTAGATCCAGACCGTGA---------GATGTTTTAGGGGCTCAAGGGGAGAAACTGAAAGTGTTTTGCTCTTTTT-TAAAGTGTTGGTCTTTCTA  826
H  AGAAGCTAATGTAGG-ACTGTATAGGTAGATCCAGATCCAGACTGTAAGATGTTGTTTTAGGGGCTAAAGGGGAGAAACTGAAAGTGTTTTACTCTTTTTCTAAAGTGTTGGTCTTTCTA  749

R  ACGTAGCTATTTTTCTCGTTGCATCTTTTCCTCTCGGGCACACTCGGTGTGCTGGGTTAATGGCTAGTGCTGTATTGACTGTGGA------AGACGTTCGTGAAGAGTATGTAGTGGCTTC  941
H  ATGTAGCTATTTTTCTTGTTGCATCTTTTCTACTTCAGTACACTTGGTGTACTGGGTTAATGGCTAGTACTGTATTGGCTCTGTGAAAACATATTTGTGAAAAGAGTATGTAGTGGCTTC  869

R  TTCCAACCCATTAGATGCTGAGTATCTGTTCACTTTGCGATCCCAATTCTGTCCCAATCTCACCAGATGCTACTGTACTGGAATGGTTAATAAA-CTGCACAGTGCTGTTGGTC pA  1054
H  TTTTGAACTGTTAGATGCTGAATATCTGTTCACTTTTCAATCCCAATTCTGTCCCAATCTTACCAGATGCTACTGGACTTGAATGGTTAATAAAACTGCACAGTGCTGTT pA  979

H'                                                                                                  GGTGGCAGTG  989
H' ACTTCTTTTGAGTTAGGTTAATAAATCAAGCCATAGAGCCCCTCCTGGTTGATACTTGTTCCAGATGGGGCCTTTGGGGCTGGTAGAAATACCCAACGCACAAATGACCGCACGTTCTCT  1109
H' GCCCCGTTTCTTGCCCCAGTGTGGTTTGCATTGTCTCCTTCCACAATGACTGCTTTGTTTGGATGCCTCAGCCCAGGTCAGCTGTTACTTTCTTTCAGATGTTTATTTGCAAACAACCAT  1229
H' TTTTTGTTCTGTGTCCCTTTTAAAAGGCAGATTAAAAGCACAAGCGTGTTTCTAGAGAACAGTTGAGAGAGAATCTCAAGATTCTACTTGGTGGTTTGCTTGCTCTACGTTACAGGTGGG  1349
H' GCATGTCCTCATCCTTTCCTGCCATAAAAGCTATGACACGAGAATCAGAATATTAATAAAACTTTATGTACTGCTGTAGT pA  1429
```

Fig. 2. Compared polynucleotide and deduced amino acid sequences of human and rat stathmin cDNA clones. The polynucleotide sequences of the 'short' (H) and the further extension of the 'long' (H') human clones are shown in comparison with the sequence of the rat pS62b clone (R) [8]; polyadenylation signals are underlined. The single amino acid residue difference between the human and rat sequences is indicated (boxed). Presumed phosphorylation sites previously identified [8,9] for the cAMP-dependent protein kinase (serines-16, -46 and -63) are circled, together with putative phosphorylation sites (serines-25 and -38) for the proline-directed serine/threonine kinase whose surrounding sequences are overlined (see text). Brackets underline the short HE(A/K)EVL repeat, whereas sequences corresponding to the long, 40%-conserved repeat (see Fig. 4) are shaded.

Sequencing of the two classes of clones revealed that they differ only by their 3′ ends, the short ones being polyadenylated a few bases after the first polyadenylation signal, whereas the long ones extend about 450 bp further, with two other polyadenylation signals (Fig. 2). We did not find any clone corresponding to the second polyadenylation signal, at only 46 bp from the first one. This might be due to a relatively low abundance of corresponding messengers, or to the proximity of the two polyadenylation sites, making the resulting clones difficult to distinguish by size. Due to the difficulty of getting human testis mRNA in good condition, RNA blots were unfortunately not of sufficient resolution to allow a direct detection of distinct stathmin mRNAs. However, since several independent clones, with slightly different polyA lengths, were identified for each of the two clone types, it is clear that they correspond to mRNAs actually expressed in the tissue of origin. The low-intensity mRNA bands hybridizing with the rat probe in rat brain and testis might thus correspond to such alternatively processed transcripts [9], although their sizes are different in rat and human. In any case, the expression of at least two mRNAs, which may have different stabilities in the cell, most likely reflects a possible level of regulation of the expression of stathmin itself.

Southern blots of human genomic DNA revealed one or two positive restriction fragments with *Eco*RI or *Bam*HI digestions (Fig. 3), a pattern compatible with a single gene for stathmin. However, more bands were detected at lower stringency (not shown), possibly corresponding to genes coding for other members of the family of stathmin-related proteins [5,9]. The additional low-stringency bands might, however, also reflect the existence of pseudogenes, as suggested by preliminary rat genomic cloning results (V. Doye, unpublished results).

3.2. *Extreme conservation of stathmin among mammals*

The nucleic sequences of the short clones are overall extremely homologous – 88% with a few gaps – to the corresponding region of the rat pS62b clone [8] (Fig. 2). They have a shorter 5′-untranslated sequence, whereas their 3′ sequence covers the whole length of the rat cDNA. The degree of homology is even higher in the coding (91%) than in the 5′ (74%) and 3′ (86%) non-coding areas. In the 5′-end region where the two published rat sequences [8,9] diverge, the homology of the human sequence is best with the pS62b clone [8].

The amino acid sequence of human stathmin deduced from the cloned cDNAs (Fig. 2) is identical to the sequence of the rat protein with the exception of an isoleucine replacing a valine at position 129. This difference is very conservative, since isoleucine is the closest possible analogue of valine. The conservation appears therefore almost total between rat and human stathmin. Interestingly, the same amino acid difference exists between the rat and bovine sequences, which may differ, however, by at least two additional residues according to the partial bovine protein sequence available as determined by direct peptide sequencing [9].

We showed recently that stathmin is immunologically well conserved through the evolution of vertebrates [5]. Our present results show more precisely that, at least throughout mammalian evolution, the selection pressure must have been extremely high for the conservation of stathmin, suggesting that its nearly entire sequence must be important to fulfill its biological functions.

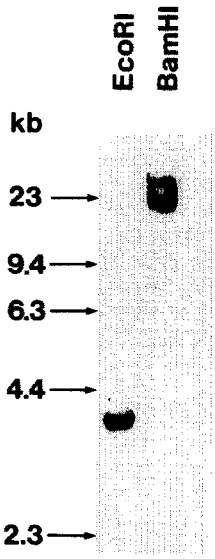

Fig. 3. Genomic Southern blot analysis. Human placental DNA was digested with the indicated endonucleases and hybridized with a 415-bp probe located downstream the internal *Eco*RI site of the cDNA inserts (see Fig. 1).

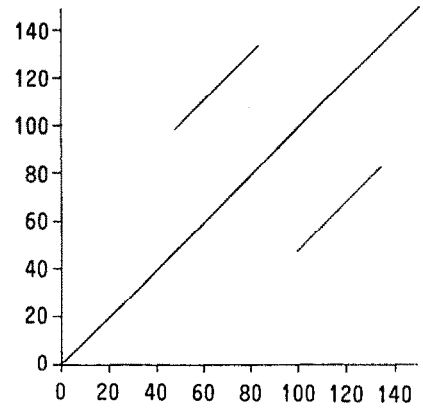

Fig. 4. Long internal repeat in the stathmin amino acid sequence. Dot matrix analysis was performed with a 40% conservation score and a window of 35, revealing the long internal repeat between regions |48–82| and |99–134|. Numbers indicate amino acid positions in the stathmin sequence.

3.3. *Phosphorylation sites and internal repeats*

We previously identified two consensus sequences around serines-16 and -63 corresponding to potential phosphorylation sites for the cAMP-dependent protein kinase [8], and a third less stringent one around serine-46 was also proposed [9]. Further examination of the sequence reveals the existence of a potential site, around serine-38, for the proline-directed serine/threonine kinase recently identified in rat pheochromocytoma and PC12 cells [15]. The sequence in stathmin (PLSPPKKK) fits the typical characteristics of the sequences identified as good substrates for this enzyme: Ser-Pro-X_{1-3} followed by basic residues, usually in a proline-rich environment [15]. Ser-25 (ILSPRSK) might also be a phosphorylation site, although slightly less typical for the same kinase. These observations are interesting in view of the fact that stathmin phosphorylation is also stimulated in PC12 cells, where this kinase was originally identified, in response to NGF and other extracellular regulatory agents ([16], and Doye, V., Boutterin, M.C. and Sobel, A., submitted for publication).

Computer analysis of the amino acid sequence of stathmin revealed the existence of two types of internal repeats. The first one is a short, 6-residue motif (64–69: HEAEVL), repeated with a single substitution (78–83: HEKEVL). Due to the low probability for such a repeat to occur within a 149-residue protein, it is possible that the sequence HE(A/K)EVL corresponds to some specific site whose functional or structural role remains to be determined.

The other internal repeat concerns a more extended 35 amino acid sequence, repeated with 40% identity between residues |48–82| and |99–134| (Figs 2, 4). These relatively large repeats might correspond to functional and/or evolutionary duplicated domains of the protein. Interestingly, the |48–82| sequence covers almost entirely the region previously identified in rat stathmin as a region of homology with proteins of the intracellular (cytoplasmic and nuclear) matrix [8]. Although we proposed that this domain might be a region of self-association of stathmin or of its interaction with cell matrix components [8], further studies will be necessary to understand the structural and/or functional importance of this long repeat in the molecular mechanisms by which stathmin fulfills its biological functions.

In conclusion, the characterization of human clones for stathmin has brought new insights regarding its molecular properties and the regulation of its expression, and it mostly demonstrated the extreme conservation of this protein among mammals, a feature stressing the most likely essential and general role of stathmin in cell regulations.

Acknowledgements: We wish to thank Professor M. Fardeau for his constant support and to gratefully acknowledge the generous gifts of biological samples and of the human testis cDNA library by Drs F. Soubrier and A.L. Lattion. This work was supported by funds from the Institute National de la Santé et de la Recherche Médicale, the Centre National de la Recherche Scientifique, the Ministère de la Recherche et de la Technologie, and the Association Française contre les Myopathies. A.M. is a bursar of the Ecole Polytechnique.

REFERENCES

[1] Sobel, A., Boutterin, M.C., Beretta, L., Chneiweiss, H., Doye, V. and Peyro-Saint-Paul, H. (1989) J. Biol. Chem. 264, 3765–3772.
[2] Schubart, U.K., Alago, W. Jr. and Danoff, A. (1987) J. Biol. Chem. 262, 11871–11877.
[3] Braverman, R., Bhattacharya, B., Feuerstein, N. and Cooper, H.L. (1986) J. Biol. Chem. 261, 14342–14348.
[4] Hanash, S.M., Strahler, J.R., Kuick, R., Chu, E.H.Y. and Nichols, D. (1988) J. Biol. Chem. 263, 12813–12815.
[5] Koppel, J., Boutterin, M.C., Doye, V., Peyro-Saint-Paul, H. and Sobel, A. (1990) J. Biol. Chem. 265, 3703–3707.
[6] Chneiweiss, H., Beretta, L., Cordier, J., Boutterin, M.C., Glowinski, J. and Sobel, A. (1989) J. Neurochem. 53, 856–863.
[7] Beretta, L., Houdouin, F. and Sobel, A. (1989) J. Biol. Chem. 264, 9932–9938.
[8] Doye, V., Soubrier, F., Bauw, G., Boutterin, M.C., Beretta, L., Koppel, J., Vandekerckhove, J. and Sobel, A. (1989) J. Biol. Chem. 264, 12134–12137.
[9] Schubart, U.K., Das Banerjee, M. and Eng, J. (1989) DNA 8, 389–398.
[10] Stein, R., Mori, N., Matthews, K., Lo, L.C. and Anderson, D. (1988) Neuron 1, 463–476.
[11] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
[12] Southern, E.M. (1975) J. Biol. Chem. 98, 503.
[13] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[14] Schubart, U.K. (1988) J. Biol. Chem. 263, 12156–12160.
[15] Vulliett, P.R., Hall, F.L., Mitchell, J.P. and Hardie, D.G. (1989) J. Biol. Chem. 264, 16292–16298.
[16] Doye, V., Boutterin, M.C. and Sobel, A. (1987) Ann. Endocrinol. 48, 19N.