Volume 289, number 2, 133–137 FEBS 10117 September 1991

ADONIS 001457939100833G

cDNA sequence of a *Drosophila melanogaster* gene, D*fur*1, encoding a protein structurally related to the subtilisin-like proprotein processing enzyme furin

Anton J.M. Roebroek[1], Ilse G.L. Pauli[1], Young Zhang[1] and Wim J.M. van de Ven[1,2]

[1]*Molecular Oncology Section, Center for Human Genetics, University of Leuven, Herestraat 49, B-3000 Leuven, Belgium* and [2]*Molecular Oncology Section, Department of Biochemistry, University of Nijmegen, Adelbertusplein 1, 6525 EK Nijmegen, The Netherlands*

Received 11 June 1991

Screening a genomic library of *Drosophila melanogaster* DNA with a human *fur* cDNA probe resulted in the isolation of DNA clones that apparently belonged to two different DNA regions of the *Drosophila* genome. Subsequently, corresponding *Drosophila* cDNA clones were isolated. Nucleotide sequence analysis indicated that these cDNA clones originated from two different genes, which were called D*fur*1 and D*fur*2. From overlapping D*fur*1 cDNA clones, a composite cDNA could be constructed and analysis of its nucleotide sequence revealed the coding sequence for a protein of 899 amino acid residues. This protein, designated Dfurin1, exhibited striking sequence homology to human furin and contained the same protein domains except for the cysteine-rich region. Furthermore, unlike human furin, Dfurin1 possessed an extended amino-terminal region in which a potential transmembrane anchor was present.

Subtilisin-like proprotein processing enzyme; Furin; Dfurin 1; *Drosophila melanogaster*

1. INTRODUCTION

Biosynthesis of a variety of bioactive proteins including neuropeptides, peptide hormones and particular plasma proteins involves endoproteolytic cleavage of inactive proproteins [1–4]. Initial endoproteolytic cleavage most often occurs at the carboxy-terminal side of paired basic amino acid residues or, less frequently, at single basic amino acid residues. The involved processing enzymes remained unknown for long, except for the Kex2 enzyme of *Saccharomyces cerevisiae*, which is a membrane-associated, Ca^{2+}-dependent, subtilisin-like serine protease with cleavage selectivity for paired basic amino acid residues (for review see [4]). Recently, however, a number of mammalian cDNA sequences with striking sequence similarity to yeast KEX2 have been described [5–10]. The deduced proteins of these candidate processing enzymes are known as furin [6], PC1 [7,9,10] and PC2 [8,9]. The furin protein is encoded by the *fur* gene [11,12], which is located immediately upstream of the *fes/fps* proto-oncogene in the genome of a number of species. The deduced amino acid sequences of human [6], mouse [13] and rat [14] furin have been reported now. In contrast to the somewhat restricted expression pattern of PC1 and PC2 [7,9], *fur* gene expression is found in a variety of tissues [13,15] and, interestingly, it is strongly elevated in a significant percentage of primary non-small cell lung carcinomas but not in small cell lung carcinomas [15]. Based on comparative nucleotide sequence analysis [5,6] and computer-assisted molecular modelling [16], structural features of furin have been predicted, including a 'pro' domain, a subtilisin-like catalytic domain which is immediately preceded by a potential (auto)proteolytic cleavage site (K-R-R-T-K-R), a middle domain, a cysteine-rich region, a potential transmembrane domain and a cytoplasmic domain. In contrast to furin, PC1 and PC2 do not possess a cysteine-rich region and a hydrophobic region that could function as a transmembrane anchor [7–10]. A three-dimensional model for the catalytic domain of furin has also been proposed and it revealed characteristics, that predicted the protein to possess an endoproteolytic cleavage selectivity at paired basic residues [16]. Recently, experimental evidence was provided in support of this as correct cleavage by furin was shown of the precursor of von Willebrand factor (pro-vWF) [16,17] and β-nerve growth factor [18]. Similar analysis of a cleavage mutant of pro-vWF (pro-vWFgly763), having the arginine residue adjacent to the proteolytic cleavage site (R-S-K-R) replaced by glycine (R-S-K-G), revealed that provWFgly763 is not processed by furin. Recently, it was shown that PC1 and PC2 also have proprotein processing activity with selectivity for paired basic amino acid residues when it was demonstrated

Correspondence address: A.J.M. Roebroek, Molecular Oncology Section, Center for Human Genetics, University of Leuven, Herestraat 49, B-3000 Leuven, Belgium. Fax: (32) (16) 215997.

that they were capable of processing the prohormone proopiomelanocortin at distinct pairs of basic residues [19]. Based upon present knowledge of the three proprotein processing enzymes, one could tentatively divide them in two subclasses. One consisting of furin, the other of PC1 and PC2.

As an approach to define protein structures of furin that might be functionally relevant, we planned to search for evolutionary conserved structural features by comparing the human furin protein to that of a distantly related species. As second species for such a comparative study, we selected *Drosophila melanogaster*. Here, we report about the discovery of two *fur*-like genes in *Drosophila melanogaster*, D*fur*1 and D*fur*2, and the molecular characterization of a D*fur*1 cDNA clone that appeared to contain the complete coding sequences for a furin-like protein, Dfurin1. Common and different structural features of human furin and Dfurin1 are discussed.

2. MATERIALS AND METHODS

2.1. *Molecular cloning of* fur-*like genetic sequences of* Drosophila melanogaster

In order to isolate *fur*-like genetic sequences of *Drosophila melanogaster* (Oregon R strain), an EMBL4 phage library of genomic DNA was screened under highly reduced stringency conditions with the complete human *fur* cDNA as molecular probe. The library was kindly provided by Dr Michael Hoffmann. Hybridization experiments on nitrocellulose membranes were performed as previously described [11], except that 35% formamide instead of 50% formamide was used.

2.2. *Isolation of* fur-*like cDNA clones*

To isolate *fur*-like cDNA clones for *Drosophila melanogaster*, embryonic cDNA libraries were screened under normal hybridization conditions using subfragments of the *fur*-like genomic clones. The embryonic *Drosophila melanogaster* (Oregon R strain) libraries (3 to 12 h in λgt10 and 0 to 16 h in λgt11) were kindly provided by Dr Roel Nusse and Dr Michael Hoffmann.

2.3. *RNA isolation and Northern blot analysis*

Drosophila melanogaster (Tübingen strain) embryos (2 to 8 h) were collected according to the procedure described by Brand and Hennig [20]. Total cellular RNA was isolated from these embryos using the lithium–urea procedure described by Auffray and Rougeon [21]. 15 μg of total RNA was glyoxylated and size-fractionated on a 1% agarose gel and transferred to Hybond-N (procedure as recommended by Amersham Corp.). For hybridization on nylon membranes the method of Church and Gilbert [22] was used.

2.4. *Nucleotide sequence analysis*

Nucleotide sequences were determined according to the dideoxy chain termination method using the T7 polymerase sequencing kit of Pharmacia/LKB. The DNA fragments to be sequenced were subcloned in the pGEM-3Zf(+) vector and sequenced using standard primers and primers synthesized based upon newly obtained sequences. The sequences were obtained from both strands and analyzed using the sequence analysis computer programs Genepro (Riverside Scientific) and Intelligenetics (IntelliGenetics, Inc.).

3. RESULTS AND DISCUSSION

The objective of this study was to detect and characterize *fur*-like genetic sequences of *Drosophila melanogaster* and, by comparison of these to the known human *fur* gene sequences, to identify evolutionary conserved structural features of the protein domains. Such an approach might reveal protein motifs that are functionally relevant for the furin-type proprotein processing enzymes. As an initial approach to detect *fur*-like genetic sequences in genomic DNA of *Drosophila melanogaster*, Southern blot experiments were performed with a 4.3 kpb *Eco*RI/*Eco*RI human *fur* cDNA fragment as molecular probe. This cDNA contained the complete coding region for human furin. Using hybridization conditions of highly reduced stringency, as outlined in section 2, weak hybridization signals were observed (data not shown). Under similar hybridization conditions, a phage library of genomic DNA of *Drosophila melanogaster* was screened and four hybridizing clones, designated λYZ2–λYZ5, were isolated. Southern blot analysis of these clones revealed that three of them, λYZ3–λYZ5, contained overlapping inserts. The DNA inserts of λYZ3–λYZ5 all contained a 2.1 kbp *Bam*HI/*Bam*HI DNA fragment which hybridized relatively strongly with the human *fur* cDNA probe. The DNA insert of clone λYZ2 appeared to be different, indicating that it corresponded to another region of the genome of *Drosophila*. It did not contain the 2.1 kbp *Bam*HI/*Bam*HI hybridizing fragment but instead, a 1.2 kbp *Eco*RI/*Kpn*I DNA fragment that hybridized relatively strongly with the human probe. Nucleotide sequence data of the 2.1 kpb *Bam*HI/*Bam*HI and the 1.2 kbp *Eco*RI/*Kpn*I fragments were obtained and analysis of these revealed a certain degree of sequence homology, as might be expected from the hybridization results. Comparison of the nucleotide sequence of each of the two fragments to the known human *fur* cDNA sequence revealed that parts of both fragments exhibited sequence homology (65%) to the human *fur* region that encodes the subtilisin-like catalytic domain of furin. These results suggested that there are at least two different genetic regions in the genome of *Drosophila melanogaster* that contain genetic sequences homologous to human *fur*. We have named these D*fur*1 and D*fur*2, implicating that the phage clones λYZ3–λYZ5 correspond to D*fur*1 and clone λYZ2 to D*fur*2.

It was tempting to speculate that the D*fur*1 and D*fur*2 sequences belong to *fur*-like genes of *Drosophila* and that these encode proteins with a subtilisin-like catalytic protein domain as in human furin. To find out whether this was the case, we tried to identify and molecularly clone such *fur*-like transcripts. In pilot Northern blot analysis of mRNA isolated from embryos of *Drosophila melanogaster*, transcripts could be detected both with D*fur*1 and D*fur*2 probes. These results suggested that D*fur*1 and D*fur*2 are expressed during early development and, therefore, we screened cDNA libraries constructed with mRNA of *Drosophila* embryos. Screening of the cDNA libraries with a D*fur*1 DNA probe resulted in the isolation of two cDNA clones. With a D*fur*2

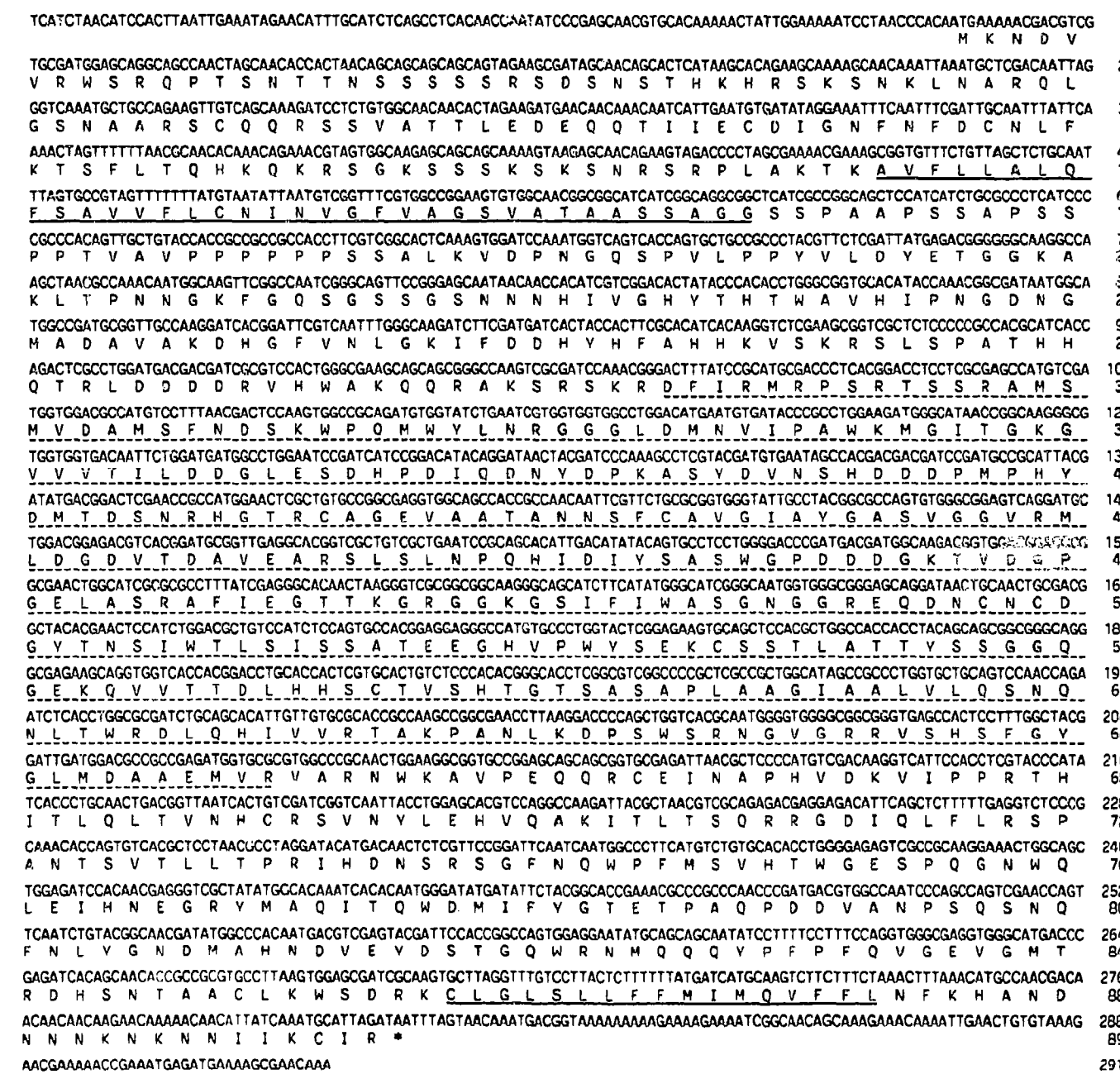

Fig. 1. Nucleotide sequence of a D*fur*1 cDNA and the predicted amino acid sequence of Dfurin1. Numbering of the amino acid residues and the nucleotides is indicated at the end of each line. Potential transmembrane domains in the amino- and carboxy-terminal region of the deduced protein are underlined by solid lines. The presumptive subtilisin-like catalytic domain is underlined by a broken line. Essential amino acid residues of the active site of it are D (372), H (413) and S (587) and the oxyanion hole N (514).

probe, a single cDNA clone was obtained. The nucleotide sequence of the DNA inserts of the two D*fur*1 cDNA clones was determined. Analysis of these revealed that the cDNA inserts were overlapping. With the two overlapping inserts, a composite cDNA could be constructed and its nucleotide sequence is presented in Fig. 1. The nucleotide sequence of the D*fur*1 cDNA of *Drosophila melanogaster* will be identified in the EMBL Data Library under accession number X59384. Within the overlapping parts of the cDNAs, a difference between the clones was noticed. The 3'-end cDNA clone lacked 21 nucleotides that were present in the 5'-end clone (nucleotides 2598–2618). These nucleotides correspond to amino acid residues 832–838 in the deduced protein (Fig. 1). Analysis of the corresponding genomic sequence revealed in what way this difference could be explained. The 21 bp appeared to be encoded by the 5'-end of an exon, that has two alternative intron/exon boundaries. The composite cDNA appeared to consist of 2915 nucleotides. This is smaller than each of the

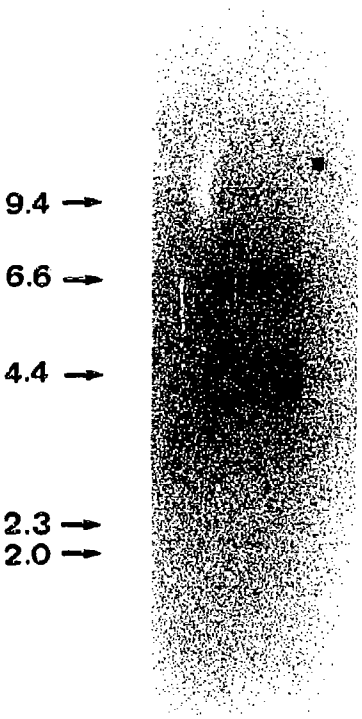

Fig. 2. Northern blot analysis of D*fur*1 gene expression in embryos (2–8 h) of *Drosophila melanogaster*. 15 μg of total RNA was subjected to agarose gel electrophoresis as described in section 2 and probed with the composite D*fur*1 cDNA clone. Molecular weight markers (kb) are indicated.

three mRNA species (about 4.0, 4.5 and 6.5 kb) that were detected with the composite cDNA as molecular probe in Northern blot analysis of RNA isolated from embryos of *Drosophila melanogaster* (Fig. 2). It is possible that, like the 4.3 kb human *fur* transcript, the D*fur*1 transcripts also contain a substantial amount of noncoding sequences. However, this fact alone may not explain the observed size differences of the three D*fur*1 mRNAs. The molecular basis for this remains to be established. Analysis of the nucleotide sequence of the composite cDNA revealed the complete coding region for a protein of 899 amino acid residues. This protein shared a number of characteristics with furin, PC1 and PC2, the only three mammalian proprotein processing enzymes presently known. Of these three, the closest resemblance seemed to exist with furin, as discussed below, and, therefore, we have named the *Drosophila* protein Dfurin1. In Fig. 3, the protein domains in human furin and Dfurin1 are schematically represented. Dfurin1 contained a domain with striking sequence homology (69%) to the subtilisin-like catalytic domain of human furin. Compared to human furin, this presumptive subtilisin-like catalytic domain appeared to have an amino-terminal insertion of 17 amino acid residues (Fig. 3). Furthermore, the protein also contained a domain corresponding to the 'pro' domain of human furin (sequence homology 51%) and one corresponding to the middle domain (sequence homology 39%). Like human

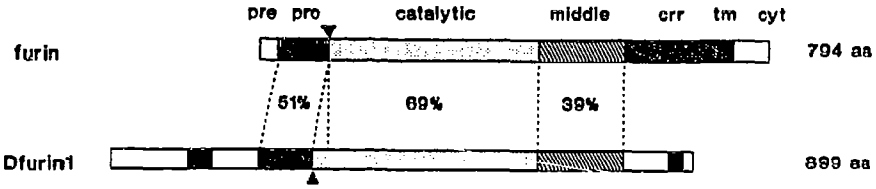

Fig. 3. Schematic representation of human furin and Dfurin1. Corresponding regions in human furin and Dfurin1 are connected by dotted lines and percentages of sequence homology are given. The potential (auto)proteolytic cleavage sites are indicated by black triangles. The potentital membrane anchor regions are represented as black boxes. crr, cysteine-rich region; tm, transmembrane domain; cyt, cytoplasmic domain.

furin, Dfurin1 also had a potential transmembrane domain in the carboxy-terminal region of the protein and a potential cleavage site (R-S-K-R) (Fig. 1, aa 306–309) for (auto)processing immediately upstream of its presumptive catalytic domain. However, unlike human furin, the deduced protein had no cysteine-rich region. Furthermore, it possessed an extended amino-terminal region with a hydrophobic region that might function as a transmembrane anchor. In that respect it resembles the product of the *sevenless* gene of *Drosophila* [23]. In Table I, the amino acid sequence homologies between the 'pro', the subtilisin-like catalytic and the middle domain of Dfurin1 and those of human furin, and the two other mammalian subtilisin-like proprotein processing enzymes are summarized. These data point towards a somewhat stronger homology of the *Drosophila* protein to human furin than to PC1 or PC2. Furthermore, unlike PC1 and PC2, Dfurin1 possesses a potential transmembrane anchor in its carboxy-terminal region. However, like PC1 and PC2, Dfurin1 does not have a cysteine-rich region. The question remains whether Dfurin1 indeed has proprotein processing activity with specificity for paired basic amino acid residues. If this can be demonstrated, the observed substrate preference of it might further resolve the question as to which of the two subclasses Dfurin1 might belong.

The nucleotide sequence of the single cDNA clone that was isolated with the D*fur*2 probe was also determined (data not shown). Analysis of reading frames revealed that this clone was incomplete. Comparison of the

Table I

Amino acid sequence homologies between protein domains of Dfurin1, Dfurin2, human furin, mouse PC1 and human PC2

	Dfurin1		
	'Pro' domain	Catalytic domain	Middle domain
Furin (human)	51%	69%	39%
PC1 (mouse)	42%	59%	34%
PC2 (human)	33%	51%	37%
Dfurin2	46%	62%	

deduced partial amino acid sequence to human furin and Dfurin1 sequences revealed that it was homologous to the 'pro' domain, the subtilisin-like catalytic domain and part of the middle domain as found in Dfurin1 and human furin (see also Table I). Furthermore, it appeared that, like in Dfurin1, there were also coding sequences for a potential amino-terminal anchor region upstream of the 'pro' domain. Finally, a potential cleavage site (R-R-K-R) for (auto)processing was also preceding the subtilisin-like catalytic domain. These data strongly suggested that, in addition to Dfurin1, another furin-like protein (Dfurin2) is expressed during early development of *Drosophila melanogaster* embryos.

In conclusion, the results of our studies suggest that *Drosophila melanogaster* possesses at least two *fur*-like genes, D*fur*1 and D*fur*2, the products of which structurally resemble the mammalian subtilisin-like proprotein processing enzyme furin and may be involved in proprotein processing during *Drosophila* embryogenesis.

Acknowledgements: We thank G. Doucet for synthesizing primers and R. Brand for providing embryos of *Drosophila melanogaster*. This work was supported by the Inter-University Network for Fundamental Research sponsored by the Belgian Government and by the Dutch Cancer Society (KWF).

REFERENCES

[1] Docherty, K. and Steiner, D.F. (1982) Annu. Rev. Physiol. 44, 625–638.
[2] Douglas, J., Civelli, O. and Herbert, E. (1984) Annu. Rev. Biochem. 53, 665–715.
[3] Sossin, W.S., Fisher, J.M. and Scheller, R.H. (1989) Neuron 2, 1406–1417.
[4] Fuller, R.S., Sterne, R.E. and Thorner, J. (1988) Annu. Rev. Physiol. 50, 345–362.
[5] Fuller, R.S., Brake, A.J. and Thorner, J. (1989) Science 246, 482–486.
[6] Van den Ouweland, A.M.W., Van Duijnhoven, J.L.P., Keizer, G.D., Dorssers, L.C.J. and Van de Ven, W.J.M. (1990) Nucleic Acids Res. 18, 664.
[7] Seidah, N.G., Gaspar, L., Mion, P., Marcinkiewicz, M., Mbikay, M. and Chretien, M. (1990) DNA Cell Biol. 9, 415–424.
[8] Smeekens, S.P. and Steiner, D.F. (1990) J. Biol. Chem. 265, 2997–3000.
[9] Seidah, N.G., Marcinkiewicz, M., Benjannet, S., Gaspar, L., Beaubien, G., Mattei, M.G., Lazure, C., Mbkikay, M. and Chretien, M. (1991) Mol. Endocrinol. 5, 111–122.
[10] Smeekens, S.P., Avruch, A.S., LaMendola, J., Chan, S.J. and Steiner, D.F. (1991) Proc. Natl. Acad. Sci. USA 88, 340–344.
[11] Roebroek, A.J.M., Schalken, J.A., Bussemakers, M.J.G., Van Heerikhuizen, H., Onnekink, C., Debruyne, F.M.J., Bloemers, H.P.J. and Van de Ven, W.J.M. (1986) Mol. Biol. Rep. 11, 117–125.
[12] Roebroek, A.J.M., Schalken, J.A., Leunissen, J.A.M., Onnekink, C., Bloemers, H.P.J. and Van de Ven, W.J.M. (1986) EMBO J. 5, 2197–2202.
[13] Hatsuzawa, K., Hosaka, M., Nakagawa, T., Nagase, M., Shoda, A., Murakami, K. and Nakayama, K. (1990) J. Biol. Chem. 265, 22075–22078.
[14] Misumi, Y., Sohda, M. and Ikehara, Y. (1990) Nucl. Acids Res. 18, 6719.
[15] Schalken, J.A., Roebroek, A.J.M., Oomen, P.P.C.A., Wagenaar, Sj.Sc., Debruyne, F.M.J., Bloemers, H.P.J. and Van de Ven, W.J.M. (1987) J. Clin. Invest. 80, 1545–1549.
[16] Van de Ven, W.J.M., Voorberg, J., Fontijn, R., Pannekoek, H., Van den Ouweland, A.M.W., Van Duijnhoven, J.L.P., Roebroek, A.J.M. and Siezen, R.J. (1990) Mol. Biol. Rep. 14: 265–275.
[17] Wise, R.J., Barr, P.J., Wong, P.A., Kiefer, M.C., Brake, A.J. and Kaufman, R.J. (1990) Proc. Natl. Acad. Sci. USA 87, 9378–9382.
[18] Bresnahan, P.A., Leduc, R., Thomas, L., Thorner, J., Gibson, H.L., Brake, A.J., Barr, P.J. and Thomas, G. (1990) J. Cell Biol. 111, 2851–2859.
[19] Benjannet, S., Rondeau, N., Day, R., Chretien, M. and Seidah, N.G. (1991) Proc. Natl. Acad. Sci. USA 88, 3564–3568.
[20] Brand, R.C. and Hennig, W. (1989) Mol. Gen. Genet. 215, 469–477.
[21] Auffrey, C.H. and Rougeon, F. (1980) Eur. J. Biochem. 107, 303–314.
[22] Church, G.M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. USA 81, 1991–1995.
[23] Basler, K. and Hafer, E. (1988) Cell 54, 299–311.