Volume 309, number 1, 15–19 FEBS 11459 August 1992

The gene structure of tetranectin, a plasminogen binding protein

Lars Berglund and Torben E. Petersen

Department of Molecular Biology, University of Aarhus, C.F. Møllers Allé 130, DK-8000, Aarhus C, Denmark

Received 29 June 1992; revised version received 15 July 1992

The gene for human tetranectin was isolated from a genomic library with a mixture of degenerate oligonucleotide probes. The gene is about 12 kbp and contains two intervening suquences. The gene encodes a protein of 202 amino acid residues, with a signal peptide of 21 amino acid residues, followed by the tetranectin sequence of 181 amino acid residues. Northern blot analysis revealed that tetranectin mRNA was present in all eight tissues tested with the highest concentration in lung. Southern blot analysis showed hybridization to two genes. Further investigations are needed to determine whether the genes are allelic or non-allelic.

Tetranectin; Cloning; Expression; Oligonucleotide probe; Plasminogen binding

1. INTRODUCTION

Tetranectin is a protein isolated from human plasma [1]. Tetranectin antigen has been shown to be present in several cells and tissues. Tetranectin may therefore have a ubiquitous cellular distribution [2–4]. In invasive breast cancer, the distribution of tetranectin is radically changed [5]. The extracellular matrix of the desmoplastic cells surrounding the tumor cells shows a high concentration of tetranectin whereas the tumor tissue and the normal breast tissue have no or very little tetranectin in the extracellular matrix. The plasma concentration of tetranectin is significantly reduced in patients with various malignancies [6].

Tetranectin is a trimeric protein with four identical and non-covalently bound polypeptide chains each of 181 amino-acid residues. The amino-acid sequence has been characterized, and the three intrachain disulfide bonds localized [7]. Tetranectin does not include carbohydrates [7]. It has a specific binding affinity for sulphated polysaccharides [8] and the kringle 4 of plasminogen [1]. These binding properties may be essential for the unknown biological function of tetranectin. The binding of tetranectin to plasminogen and its stimulating effect on tissue-type plasminogen activator (t-PA) catalysed plasminogen activation in vitro, indicate that tetranectin may participate in processes where plasminogen activators are involved like tissue degradation, involution, extracellular proteolysis and cell migration.

Sequence homologies have been found to the regions of the asialoglycoprotein receptor family, from human, chicken and rat that are thought to be extracellular. In addition tetranectin has homologies to the C-terminal globular domain of the cartilage proteoglycan core proteins from chicken and rat [7], the IgE receptor from human B lymphocytes [9], canine lung surfactant factor [10], human thrombomodulin and pancreatic store protein [11]. A protein has been isolated from shark cartilage which shows 45% identity to the tetranectin sequence [12].

In order to elucidate how the synthesis of tetranectin is regulated in normal cells and tumor cells, it is necessary to isolate and analyze the gene and its regulatory sequences. We here report the isolation and characterization of a human tetranectin gene. The DNA sequence shows that the gene has three exons encoding the 202 residues pretetranectin.

Abbreviations: kbp, kilo base pairs; nt, nucleotides.

Correspondence address: L. Berglund, Department of Molecular Biology, University of Aarhus, C.F. Møllers Allé 130, DK-8000 Århus C, Denmark. Fax: (45) 86196226.

2. MATERIALS AND METHODS

2.1. *Enzymes and reagents*

Restriction endonucleases, T4 DNA ligase, polynucleotide kinase and nick translation kits were from Boehringer Mannheim. DNA sequence reactions were performed with a Sequenase kit (United States Biochemical Corporation). [γ-^{32}P]ATP was from ICN and [α-^{32}P]dATP was from Amersham. The human genomic λ-phage library was from Stratagene and the multiple tissue Northern blot and the human genomic Southern blot were from Clontech.

2.2. *Screening of the genomic λ-phage library*

The human genomic λ-phage library was constructed in vector λFIX™II with *Sau*3A partially-digested caucasian male placenta DNA. About 500,000 plaques were screened by in situ hybridization essentially as described by Benton and Davis [13]. A mixture of seven degenerate oligonucleotides labeled at the 5′ end with ^{32}P was used as hybridization probe (see below).

Hybridization was performed in 6 × SSC (1×SSC = 150 mM NaCl, 15 mM trisodiumcitrate dihydrate), 10 × Denhardt's solution (100 × Denhardt's solution = 2% bovine serum albumin, 2% Ficoll-

400, 2% plyvinylpyrrolidone), 0.1% SDS and denatured salmon sperm DNA (50 μg/ml) overnight at 37°C. The nitrocellulose filters were washed for 15 min at room temperature and 2 × 10 min at 45°C in 6 × SSC, 0.1% SDS.

Phage DNA prepared from positive plaques was digested with restriction endonucleases, and characterized by Southern blot analysis [14].

2.3. *Oligonucleotide probes*

The known amino-acid sequence of tetranectin [7] was used for the design of seven degenerate oligonucleotides (Tet1–7) which were used as probes and primers in screening and characterization of the tetranectin gene. (X = deoxy Inosine)

Tet1: CA(AG)AA(AG)CC(AGCT)AA(AG)AA(AG)AT.
Tet2: TCXTCAAACATXTTXGTXTTXACXACATCXTTXTT.
Tet3: AA(CT)AC(AGCT)AA(AG)ATGTT(CT)GA(AG)GA.
Tet4: AA(AG)AC(AGCT)TT(CT)CA(CT)GA(AG)GC.
Tet5: TC(AGCT)GT(CT)TCCCA(AG)TT(CT)TT(AG)TA.
Tet6: TT(AG)TC(AG)AACCA(CT)TT(AGCT)CC(AG)TT.
Tet7: CC(AG)AA(CT)TG(AG)CAXAT(AG)TA.

2.4. *DNA sequence analysis*

Phage DNA from one clone which hybridized both to oligonucleotide probes corrsponding to the N-terminal region and to oligonucleotides corresponding to the C-terminal region of tetranectin, was subcloned in plasmid pUC19 [15]. The degenerate oligonucleotides Tet1, Tet3, Tet4 and Tet7 were used as primers in double-strand sequencing of subclones with the dideoxy chain-termination method [16]. The sequences were extended by using complementary primers to the determined DNA sequences.

2.5. *Northern and Southern blot analyses*

The multiple tissue Northern and the genomic Southern blots were hybridized with a nick translated probe containing an 850 bp *Xba*I–*Eco*RI fragment from λTN1. This fragment contains the 5′ exon and sequences from the promoter and 5′ intron (see Fig. 1).

3. RESULTS

3.1. *Isolation and analysis of human genomic tetranectin DNA*

A human genomic library was screened with a mixture of seven degenerate oligonucleotides (Tet1–7) as probes. Twelve positive clones were purified and one clone λTN1, hybridized to all of the degenerate oligonucleotides except Tet2. Later the genomic sequence showed that oligonucleotide Tet2 spans the 5′ intron which explains the lack of hybridization.

An extensive restriction map was made of the λTN1 clone (Fig. 1). The fragments with the tetranectin coding sequences were localized, subcloned in plasmid pUC19 and sequenced. The sequence showed that λTN1 encodes pre-tetranectin, a protein of 202 amino-acid residues with a putative translational start in position 254 (Fig. 2). The N-terminal signal peptide is 21 amino-acid residues long and has, like other signal peptides, a central core of hydrophobic amino acids. The N-terminal of plasma tetranectin is in agreement with the predicted cleavage site for the signal peptide [17].

The mature tetranectin encoded by the gene in λTN1 is identical with the amino-acid sequence of plasma tetranectin except for one amino acid. The gene encodes a serine in amino-acid position 106 (Fig. 2) whereas a glycine was found by protein sequencing [7]. The gene sequence predicts the same C-terminal residue as was found in plasma tetranectin.

The coding sequence is interrupted by two introns. The intron border sequences GT/AG are present in both introns. The 5′ intron is about 3.5 kbp and the 3′ intron is about 6.5 kbp as estimated from restriction fragment analysis. A polyadenylation signal AATAAA is located 117 bp downstream for the stop codon. A putative cap site in position 132 (Fig. 2) is located 34 bp downstream for the TATA-box-like sequence TATT in position 99 (Fig. 2). This putative cap site is 122 bp upstream for the ATG start codon.

3.2. *Northern blot and genomic Southern blot*

The Northern blot with 2 μg of poly(A) RNA in each lane contained RNA from eight different human tissues. It was hybridized with the 0.8 kb *Xba*I/*Eco*RI fragment, from the 5′ end of the gene, as probe. One hybridizing band was seen in each lane within a size range of 900 to 1,100 nucleotides (Fig. 3). The size of the tetranectin gene exons in λTN1 was calculated to be about 875 nucleotides. Addition of a poly(A) tail results in tetranectin mRNA molecules with a size which is in agreement with the Northern blot. The hybridization signal was strong in lanes with mRNA from lung, muscle, heart, and placenta, compared to the weak signal seen in lanes with mRNA from brain, liver, kidney, and pancreas.

The genomic Southern blot was hybridized with the same nick translated probe as was used for the Northern blot. One hybridizing band was observed in *Eco*RI,

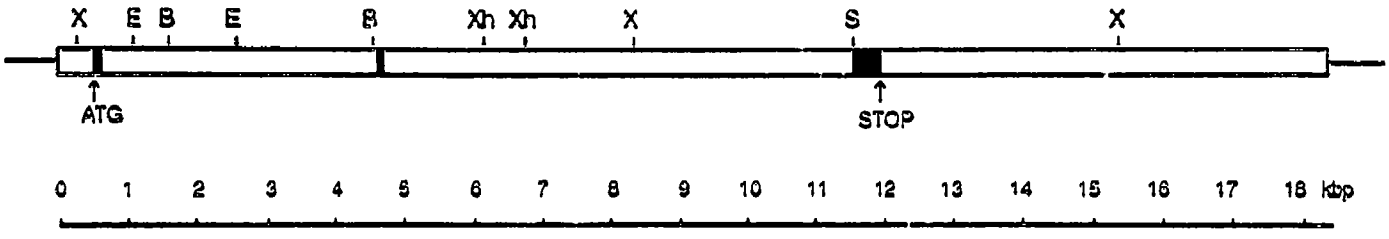

Fig. 1. Map of the tetranectin gene in the λTN1 clone. The boxed area represents the insert of the λ clone, with the coding region shown in black. E, *Eco*RI; B, *Bam*HI; S, *Sal*I; X, *Xba*I; Xh, *Xho*I.

THE HUMAN TETRANECTIN GENE

```
tctagaaggcagggccaggatgctgcgtgggtgttccctggcagaggaggctctgtggccctcaggacttcatccttctgcaggcttgtgt   90
gtttgtttaatttttttggcacgcctgcagccaagttggggagggtttcctggacagaggtcctttggctgctgccttaagacgtgcagc  180
ctgggccgtggctgtcactgcgttcggacccagacccgctgcaggcagcagcagccccgccgcgcagcagcATGGAGCTCTGGGGGGC  270
                                                                    M  E  L  W  G  A      6

CTACCTCCTCCTCTGCCTCTTCTCCCTCCTGACCCAGGTCACCACCGAGCCACCAACCCAGAAGCCCAAGAAGATTGTAAATGCCAAGAA  360
 Y  L  L  L  C  L  F  S  L  L  T  Q  V  T  T  E  P  P  T  Q  K  P  K  K  I  V  N  A  K  K   36

AGgtaaggaggggggacagagccctgtgccatcttccagggagcaggtccccctctccttagtgtgtcctcatgctcctttccttcatttt  450
 D                                                                                            37

cctccttttgagtcatctaggggacttgagtaaagtgggctttgggggcttgggggagttctctcatagcagagttggttgtcagaatcat  540
gagggtccctgacccttagagttgggaaga                                                               570
                                    Intron 3.5 kbp

gtggttaattagactaccaccagcaaggatagagagagagaaagtttcccaagatatcaaggttgaggggagtcttttcctctcatcccc  660
ctgcagctaaggtaggatccttcctgctcagtcccctgccccacccctgacatatttcctcctgcttcagATCTTCTGAACACAAAGATGTT  750
                                                                    V  V  N  T  K  M  F     44

TGAGGAGCTCAAGAGCCGTCTGGACACCCTGGCCCAGGAGGTGGCCCTGCTGAAGGAGCAGCAGGCCCTGCAGACGGgtgagtgcaggca  840
 E  E  L  K  S  R  L  D  T  L  A  Q  E  V  A  L  L  K  E  Q  Q  A  L  Q  T  V              70

gtagcctctctgggcaggagcgtctgagagaagggccc                                                       878
                                    Intron 6.5 kbp

atcaggacccctttttgcctgggtctaaggggctgtcaggactgatgggataaacgggatggatggaggacggtgggcttggcctgggcgg  968
ttggctctttgcctgggggaacaggggggacttcggtgacagccatttctccccactcccAGTCTGCCTGAAGGGGACCAAGGTGCACATGA 1058
                                                        C  L  K  G  T  K  V  H  M  K     80

AATGCTTTCTGGCCTTCACCCAGACGAAGACCTTCCACGAGGCCAGCGAGGACTGCATCTCGCGCGGGGGCACCCTGAGCACCCCTCAGA 1148
 C  F  L  A  F  T  Q  T  K  T  F  H  E  A  S  E  D  C  I  S  R  G  G  T  L  S  T  P  Q  T  110

CTGGCTCGGAGAACGACGCCCTGTATGAGTACCTGCGCCAGAGCGTGGGCAACGAGGCCGAGATCTGGCTGGGCCTCAACGACATGGCGG 1238
 G  S  E  N  D  A  L  Y  E  Y  L  R  Q  S  V  G  N  E  A  E  I  W  L  G  L  N  D  M  A  A  140

CCGAGGGCACCTGGGTGGACATGACCGGCGCCCGCATCGCCTACAAGAACTGGGAGACTGAGATCACCGCGCAACCCGATGGCGGCAAGA 1328
 E  G  T  W  V  D  M  T  G  A  R  I  A  Y  K  N  W  E  T  E  I  T  A  Q  P  D  G  G  K  T  170

CCGAGAACTGCGCGGTCCTGTCAGGCGCGGCCAACGGCAAGTGGTTCGACAAGCGCTGCCGCGATCAGCTGCCCTACATCTGCCAGTTCG 1418
 E  N  C  A  V  L  S  G  A  A  N  G  K  W  F  D  K  R  C  R  D  Q  L  P  Y  I  C  Q  F  G  200

GGATCGTGTAGccggcggggcgggggccgtgggggggcctggaggagggcaggagccgcgggaggccgggaggagggtggggacccttgcag 1508
 I  V  *                                                                                     202

ccccatcctctccgtgcgcttggagcctctttttgcaaataaagttggtgcagcttcgcggagaggagaggcgctgcagtctgtgctgt 1598
gtccgct                                                                                      1605
```

Fig. 2. Nucleotide sequence of the human tetranectin gene. The translated regions are in capital letters with the derived amino-acid sequence shown below the first letter of each codon. The start codon for translation (ATG), the N-terminal glutamic acid (E) of the native protein, the intron border sequences (GT/AG), the polyadenylation signal (AATAAA), and the putative TATA-box sequence (TATT) are underlined. The stop codon for translation is marked with an asterisk (*).

*Pst*I, and *Bgl*II digests, while two bands were observed in the *Bam*HI and *Hin*dIII digested DNA.

4. DISCUSSION

This study shows that a gene encoding tetranectin has been isolated from a genomic library screened with a mixture of seven degenerate oligonucleotide probes. The probes were designed on the basis of the known amino-acid sequence for tetranectin [7]. Each of the seven oligonucleotides represents up to 64 different sequences. The problem with false positive signals, in the screening of a large genomic library with oligonucleotide probes, was minimized by using a mixture of seven degenerate oligonucleotides as probe. The true positive clones with the entire gene or large fragments of the gene hybridized with several of the different probes in the hybridization mixture, whereas the false positive clones hybridized with fewer of the oligonucleotide probes. The clones with the strongest hybridization signals were purified, and rescreened with the seven oligonucleotide probes one by one. One clone, λTN1, which hybridized to oligonucleotides representing both the N-terminal and C-terminal parts of tetranectin, was characterized by sequencing.

The λTN1 clone contains the entire gene of tetranectin (Fig. 1). The gene is about 12 kbp and contains three exons. This gene encodes a pre form of tetranectin with 202 amino-acid residues. The signal peptide of 21 amino-acid residues is followed by the 181 amino-acid

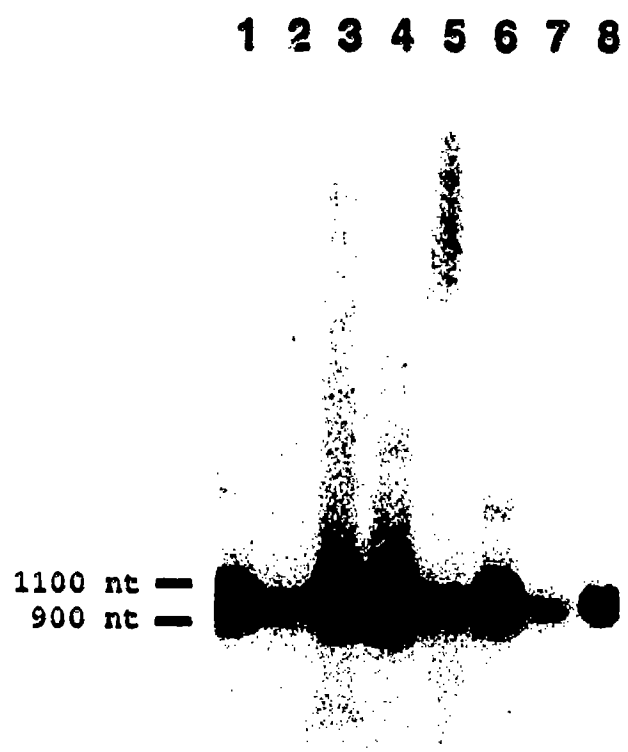

Fig. 3. Northern blot analysis of human tetranectin in various normal tissues. Each lane contains 2 μg of poly(A) RNA from normal healthy tissues. Lanes: 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas. The probe was a nick translated 0.8 kbp *Xba*I/*Eco*RI fragment from the 5' end of the tetranectin gene.

long sequence of tetranectin (Fig. 2). The only difference between the sequences of the gene-encoded tetranectin and the plasma tetranectin is the Ser found in position 106 of the gene-encoded protein which corresponds to a Gly in the amino-acid sequence of the plasma protein. The difference may be caused by the presence of another tetranectin gene in the genome, since two genes were detected in the genomic Southern blot.

The Northern blot shows that tetranectin may be a ubiquitous protein in normal tissue since all eight tissues tested showed hybridization with the tetranectin probe (Fig. 3). This is in agreement with previous immunohistochemical studies [2,3]. One band with a size range of 900 to 1,100 nucleotides was seen in each lane on the Northern blot. This size is in accordance with the size of the tetranectin gene in the isolated λTN1 clone. It has previously been discussed whether another tetranectin protein of 80,000 Da exists [4,5]. With the probe containing only the first exon a mRNA for such a protein was not detected.

The differences in hybridizing intensities are caused by differences in gene expression, since equal amounts of poly(A) RNA were loaded in each lane. The tissues used for RNA isolation were from normal healthy persons, which minimises the probability of changed tetranectin gene expression caused by cancer.

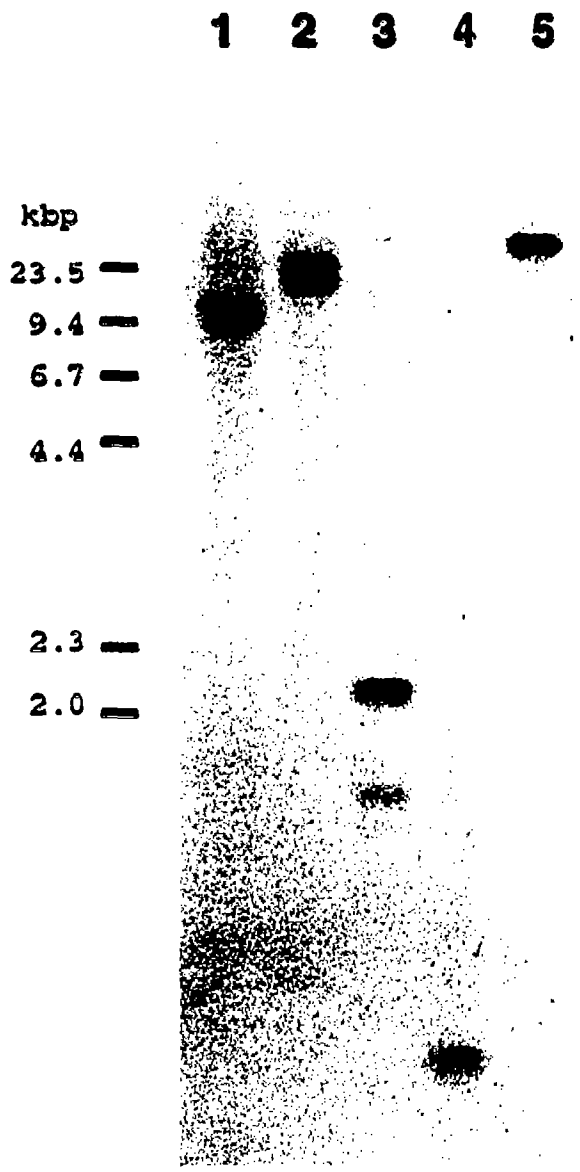

Fig. 4. Genomic Southern blot analysis of the human tetranectin gene. Each lane contains 8 μg DNA digested with various enzymes. Lanes: 1, *Eco*RI; 2, *Hin*dIII; 3, *Bam*HI; 4, *Pst*I; 5, *Bgl*II. The probe was a nick translated 0.8 kbp *Xba*I/*Eco*RI fragment from the 5' end of the tetranectin gene.

The genomic Southern blot does not give a conclusive answer about the number of tetranectin genes (Fig. 4). One gene is expected from the *Eco*RI and *Bgl*II digests with one hybridizing band in each lane. Two bands are seen in the *Bam*HI and *Hin*dIII lanes. *Bam*HI and *Hin*dIII do not cut the probe region. The two hybridizing bands in each of these lanes therefore represent two genes. These results can be explained as an allelic polymorphism caused by an insertion/deletion in the tetranectin gene. If the two genes are non-allelic, these genes have to be closely located, since both the *Eco*RI and *Bgl*II band appear as a single band.

The *Pst*I enzyme cuts the probe fragment once. The

finding of only one and not two bands in the *Pst*I-digested DNA is caused by *Pst*I sites close to the probe fragment, which therefore results in two small fragments where only the one is bound to the blot.

Further investigations are needed to determine whether there are two allelic or non-allelic genes for tetranectin. The finding of heterogeneities in positions 34 and 37 in the protein sequence of plasma tetranectin [7] could be a result of two genes.

The function of tetranectin is still not known. The extracellular localization in the desmoplastic tissue, may be a result of activation of the transcription as well as the transport from the cytoplasm to the extracellular compartment. This activation may be caused by growth factors secreted by the tumor cells. The isolation of the tetranectin gene improves the ability to investigate this activation.

Recently a cDNA has been isolated from a human placenta library [18]. Data from this clone is in agreement with the genomic results presented in this paper.

Acknowledgements: The authors wish to thank Torsten Kristensen for helpful discussions, critical reading of the manuscript and access to his laboratory facilities. We also thank Inge Clemmensen for her interest and help. This research was supported by The Danish Biotechnology Programme and The Danish Cancer Society.

REFERENCES

[1] Clemmensen, I., Petersen, L.C. and Kluft, C. (1986) Eur. J. Biochem. 156, 327–333.
[2] Christensen, L., Johansen, N., Jensen, B.A. and Clemmensen, I. (1987) Histochemistry 87, 195–199.
[3] Christensen, L. and Clemmensen, I. (1989) Histochemistry 92, 29–35.
[4] Borregaard, N., Christensen, L., Bjerrum, O.W., Birgens, H.S. and Clemmensen, I. (1990) J. Clin. Invest. 85, 408–416.
[5] Christensen, L. and Clemmensen, I. (1991) Histochemistry 95, 427–433.
[6] Jensen, B.A. and Clemmensen, I. (1988) Cancer 62, 869–872.
[7] Fuhlendorff, J., Clemmensen, I. and Magnusson, S. (1987) Biochemistry 26, 6757–6764.
[8] Clemmensen, I. (1989) Scand. J. Clin. Lab. Invest. 49, 719–725.
[9] Ludin, C., Hofstetter, H., Sarfari, M., Levy, C.A., Suter, U., Alaimo, D., Kilcherr, E., Frost, H. and Delespresse, G. (1987) EMBO J. 6, 109–114.
[10] Drickamer, K., Dordal, M.S. and Reynolds, L. (1986) J. Biol. Chem. 261, 6878–6887.
[11] Petersen, T.E. (1988) FEBS Lett. 231, 51–53.
[12] Neame, P.J., Young, C.N. and Treep, J.T. (1992) Protein Science 1, 161–168.
[13] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[14] Southern, E. (1975) J. Mol. Biol. 98, 503–517.
[15] Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.
[16] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
[17] von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.
[18] Wewer, U.M. and Albrechtsen, R. (1992) Lab. Invest., in press.