Volume 213, number 2, 254–260 FEB 04511 March 1987

Molecular cloning and characterization of a full-length cDNA clone for human plasminogen

Margaretha Forsgren, Benny Råden, Marianne Israelsson, Kerstin Larsson and Lars-Olof Hedén

KabiGen AB, Department of Molecular Biology, S-112 87 Stockholm, Sweden

Received 9 January 1987

A human liver cDNA library enriched for full-length clones was screened for plasminogen cDNA using a synthetic 24-nucleotide probe derived from a reported partial cDNA sequence. 12 positive clones were identified and one of these was characterized in detail. The 2.7 kb insert contains the complete coding region. At 5 positions, it gives residues different from those reported in a previous amino acid sequence analysis of the protein. The present results show an extra Ile at position 65, Gln instead of Glu at positions 53 and 342, Asn at position 88 instead of Asp, and Asp at position 453 rather than Asn. In the 3′-non-coding region an extension of 29 bases is found which does not contain any structure compatible with a known polyadenylation signal. Instead, the consensus signal AATAAA is placed at a distance of 46 bases upstream of the poly(A)-tail.

cDNA; Plasminogen; Plasma protein; (Human)

1. INTRODUCTION

The glycoprotein plasminogen is a zymogen involved in the final steps of fibrinolysis [1]. The amino acid sequence of human plasminogen has been determined [2–5]. The one-chain proenzyme is converted to the active two-chain molecule plasmin by cleavage of the peptide bond between Arg-560 and Val-561 [6]. This specific cleavage is mediated by plasminogen activators, e.g. tissue plasminogen activator and urokinase. The serine protease function of the molecule is located in the carboxy-terminal part of the original protein (giving the light chain after the activation cleavage). This region shows considerable amino acid sequence homology with the serine proteases involved in blood coagulation, as well as those involved in other physiological processes such as the digestive enzymes of the pancreas.

Correspondence address: L.-O. Hedén, University of Lund, Dept of Microbiology, S-223 62 Lund, Sweden

In the amino terminal part of the original molecule (the heavy chain) five tandem repeats, called kringles, are present. These structures, containing about 80 amino acids each are homologous to the two kringles present in the amino-terminal region of tissue plasminogen activator and prothrombin, as well as the single kringle present in the amino-terminal portion of urokinase and factor XII of the blood coagulation (review [7]). The exact role of the kringles has not yet been elucidated, but at least in plasminogen and prothrombin some of the kringles have been shown to mediate fibrin binding.

Based on the amino acid sequence homology between the serine proteases of coagulation and fibrinolysis Patthy [7,8] among others has proposed a hypothesis for the evolution of these enzymes based on the assumption of exon-shuffling. To allow a more thorough and stringent comparison of the different domains in these proteins knowledge of the complete nucleotide sequences of the respective genes and their genomic organization is essential.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/87/$3.50

Plasminogen, and many other enzymes in the coagulation and fibrinolysis systems, are synthesized in the liver. Recently a partial cDNA clone for plasminogen from human liver mRNA has been described [9]. In the present communication, a full-length cDNA clone for human plasminogen is reported.

2. MATERIALS AND METHODS

2.1. *Restriction enzymes and isotopes*

Restriction enzymes, [α-^{35}S]dCTP (1100 Ci/mmol), and [γ-^{32}P]ATP (3000 Ci/mmol) were obtained from Amersham.

2.2. *Construction and isolation of cDNA clones*

In order to increase the probability to isolate a full-length cDNA clone for plasminogen, the method described by Okayama and Berg [10,11] was used with some modifications. The vector and linker fragments were both prepared from the plasmid pT$_4$ (kindly provided by G. Gross, Gesellschaft für Biotechnologische Forschung, Braunschweig) as described in fig.1. The use of the plasmid pT$_4$ in preparing cDNA libraries has a great advantage in that the same vector can be used in the preparations of the vector and linker fragments, and inserts can easily be cleaved out. Some 10 bp downstream of the *Kpn*I site there is a unique *Pst*I site. Preparation of the linker fragment as outlined in fig.1 will recreate the *Kpn*I site. From 1 μg vector DNA and 5 μg human liver mRNA about 1.5×10^5 transformants were obtained. The transformants were pooled and frozen in glycerol at -70°C. From the published partial cDNA sequence [9], a 24-nucleotide probe was synthesized [12] and labelled with [^{32}P]ATP as described [13].

2.3. *Plasmid DNA preparation and restriction enzyme analysis*

Plasmid DNA was isolated by density gradient centrifugation of cleared lysates [13] or by preparation of minilysates [14]. Restriction enzyme analysis of plasmid DNA was done as described [13].

Agarose gel electrophoresis was done in 0.8% DNA grade agarose (Bio-Rad Laboratories) in Tris-acetate buffer [13], 100 V, at room temperature. Low gelling temperature agarose gels were run in the same buffer at $+4$°C.

Transfer of restriction DNA fragments separated on agarose gel to nitrocellulose filter for hybridization was done by the Southern procedure [13].

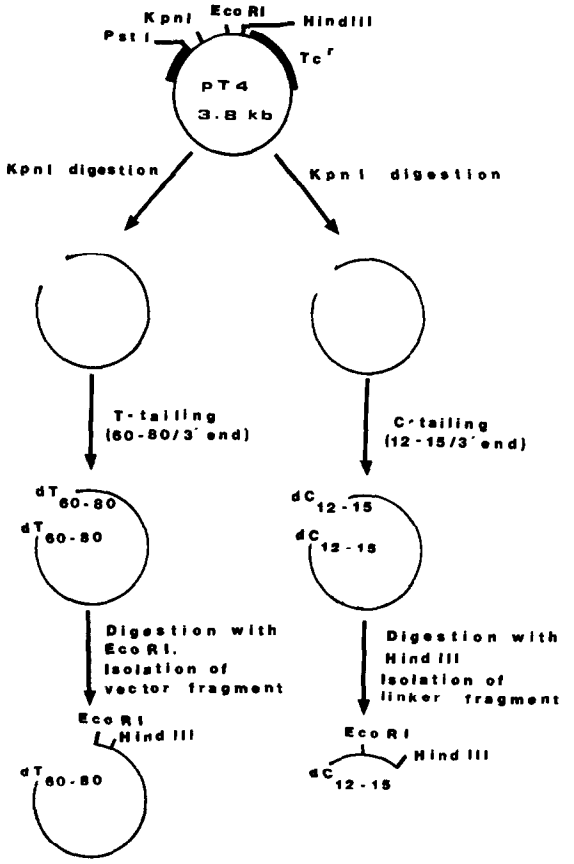

Fig.1. Schematic description of the vector pT$_4$ and the preparation of vector and linker fragments for cDNA cloning according to the Okayama-Berg procedure [10,11]. Only restriction enzyme sites relevant to the text are indicated.

2.4. *Nucleotide sequence determination*

Nucleotide sequence determination was done by the Sanger dideoxy chain-termination method [15] using ^{35}S-labelled dCTP. Subcloning of appropriate restriction enzyme fragments in M13mp18 and M13mp19 was done by separating the DNA fragments on a 0.8% low gelling temperature agarose gel (Bio-Rad Laboratories) and subsequent ligation to the vectors as described

Table 1

Oligonucleotide probe derived from the partial cDNA sequence [9] for the screening of the cDNA library

	292							299
Amino acid sequence	Ser-	Gly-	His-	Thr-	Cys-	Gln-	His-	Trp
DNA sequence	5'- TCC	GGG	CAC	ACC	TGT	CAG	CAC	TGG -3'
	3'- AGG	CCC	GTG	TGG	ACA	GTC	GTG	ACC -5'
DNA probe	5'- TCC	GGG	CAC	ACC	TGT	CAG	CAC	TGG -3'

The amino acid numbers given are with reference to the sequence in fig.4

by Crouse et al. [16]. For confirmation of the DNA sequence around the restriction enzyme sites used in the subcloning, synthetic oligonucleotide primers were used.

3. RESULTS

About 10000 transformants in the human liver cDNA library were screened for plasminogen cDNA using a ^{32}P-labelled oligonucleotide probe (see table 1 for the sequence). Twelve positive clones were identified and, after rescreening, minilysates were analysed by agarose gel electrophoresis. One of the clones was chosen for a detailed analysis. An initial restriction enzyme analysis of plasmid DNA from this clone showed the size of the insert to be about 2.7 kb (fig.2A) which is a size expected for a full-length clone. Southern analysis (fig.2B) and DNA sequencing of fragments hybridizing to the oligonucleotide probe confirmed the presence of a plasminogen coding sequence. For the complete nucleotide sequence

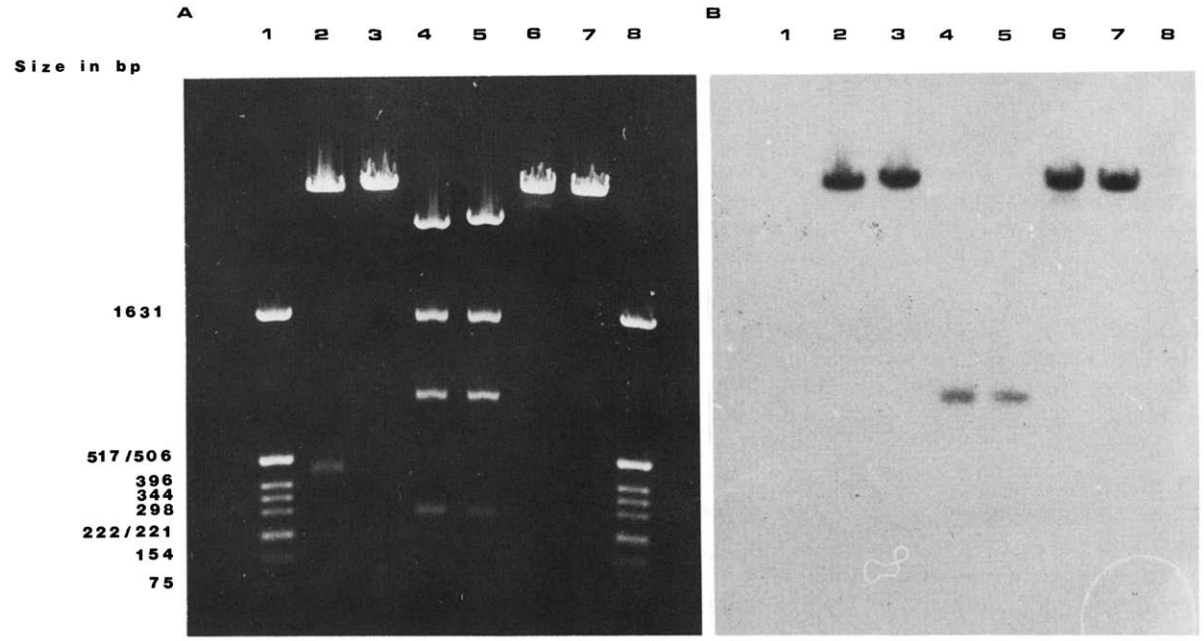

Fig.2. Restriction enzyme analysis and Southern analysis of the human plasminogen cDNA clone pPLGKG. Lanes 1 and 8, pBR322 digested with *Hin*fI as a molecular size marker. Lanes 2–7: pPLGKG digested with *Eco*RI (lane 2), *Bam*HI (3), *Kpn*I and *Pst*I (4), *Pst*I (5), *Kpn*I (6), and *Hin*dIII (7). (A) Ethidium bromide staining and (B) autoradiogram after hybridization with the DNA probe used for the colony hybridization.

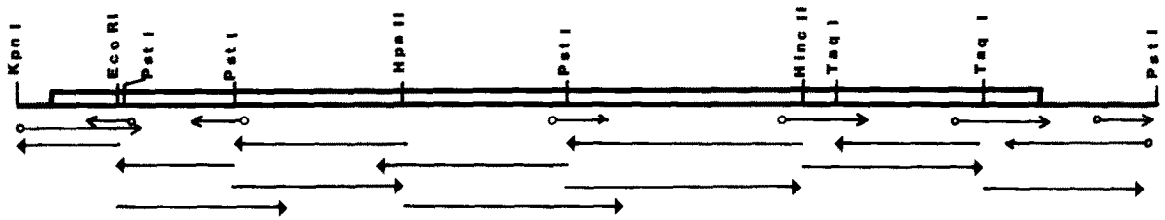

Fig.3. Partial restriction enzyme map for the human plasminogen cDNA clone pPLGKG. Only restriction enzyme sites used in the DNA sequence determination are shown. The boxed region shows the coding part. The arrows represent subcloned fragments and the direction of the sequence determination. ○⟶ represents synthetic primers.

determination, restriction enzyme fragments were subcloned in M13 as outlined in fig.3. The DNA sequence was determined on both strands. For the 5′-*Kpn*I-*Eco*RI and the 3′-*Taq*I-*Pst*I fragments internal synthetic oligonucleotide primers were used due to difficulties determining the sequence over the G/C and A/T tails.

In fig.4 the complete nucleotide sequence is presented. A 5′-non-coding region of 64 bp is present and as most eukaryotic mRNA usually have a 5′-non-coding region ranging from 40 to 80 nucleotides [17] this indicates that the clone obtained may be of full length. The 253 bp 3′-non-coding region is an extension by 29 bp in comparison to the partial cDNA clone described by Malinowski et al. [9]. This extension does not contain a sequence known to function as a polyadenylation signal in eukaryotic mRNA [18]. A consensus AATAAA sequence is located at a position 46 nucleotides upstream of the polyadenylation site. The extension was further confirmed by sequencing using synthetic primers as indicated in figs 3 and 4. The coding region contains an amino-terminal sequence of 19 amino acids with the characteristics of a signal sequence [19].

The sequence for the mature protein shows the presence of an extra isoleucine at position 85 (fig.4) in comparison to the sequence published by Sottrup-Jensen et al. [5], giving a total of 791 amino acids for human plasminogen. The present cDNA sequence deviates from the published amino acid sequence at the following positions as also observed by Malinowski et al. [9]: Gln instead of Glu at position 342 and Asp instead of Asn at position 453 (fig.4). Beside these discrepancies, the cDNA sequence now obtained predicts Gln instead of Glu at position 53 and Asn instead of Asp at position 88. The total amino acid composition of mature human plasminogen as well as the codon usage for the non-processed protein are given in tables 2 and 3, respectively. The calculated molecular mass for the processed non-glycosylated protein is 88400 Da, and for the heavy and the light chains 63200 and 25200 Da, respectively, as compared to an estimated 92 kDa for the glycosylated form and 57 and 25 kDa, respectively, for the heavy and light chains [20].

Table 2

Total amino acid composition of mature plasminogen (the total number of amino acids is 791)

Amino acid	Number of residues	Mol%
Ala	36	4.55
Arg	42	5.31
Asn	39	4.93
Asp	36	4.55
Cys	48	6.07
Gln	30	3.79
Glu	54	6.83
Gly	60	7.59
His	23	2.91
Ile	22	2.78
Leu	42	5.31
Lys	48	6.07
Met	10	1.26
Phe	20	2.53
Pro	69	8.72
Ser	55	6.95
Thr	61	7.71
Trp	19	2.40
Tyr	30	3.79
Val	47	5.94

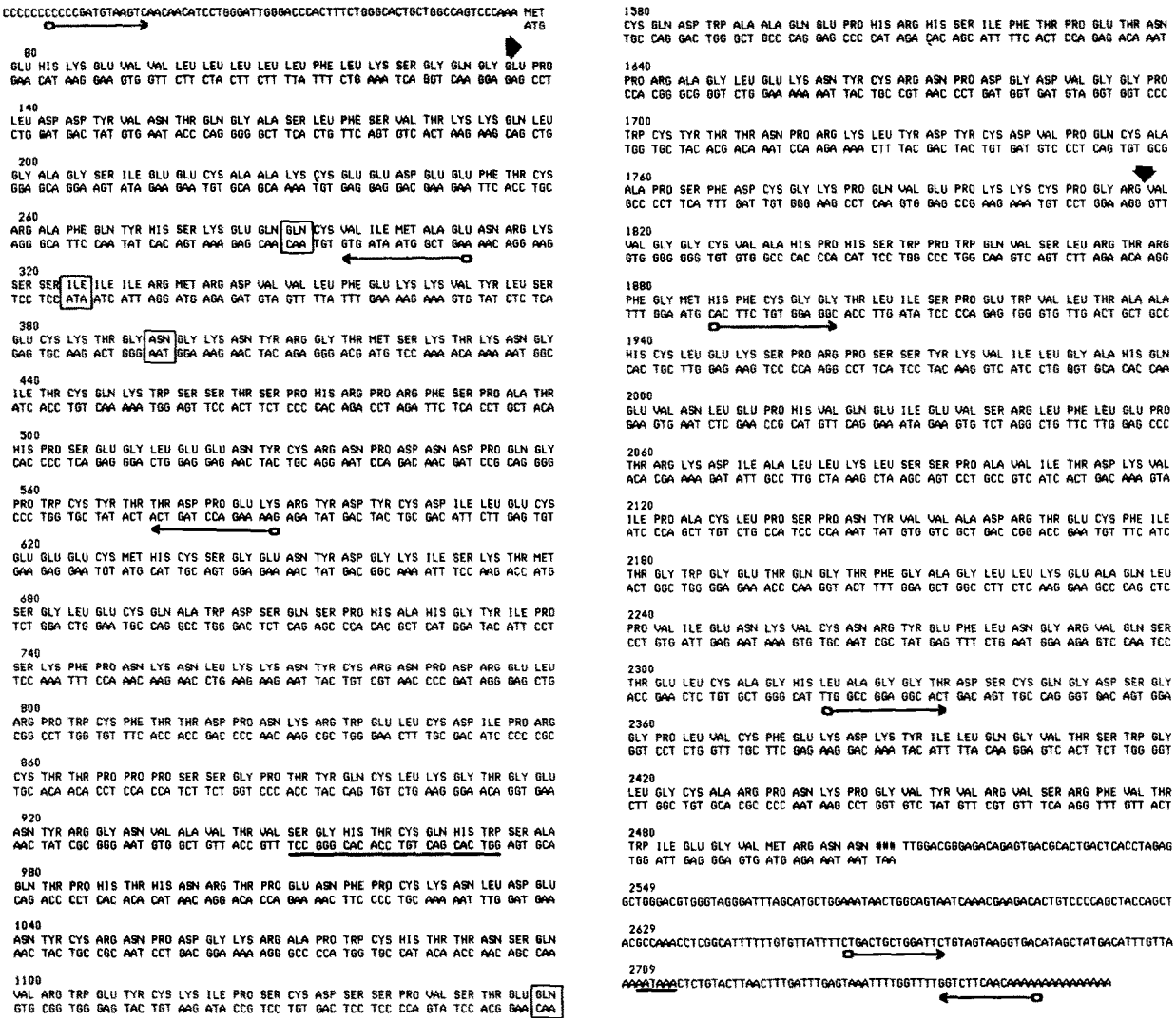

Fig.4. cDNA sequence for pPLGKG. The arrows ▶ and ▼ indicate the start of the mature protein and the site of cleavage of the proenzyme, respectively. Boxed amino acids represent amino acid deviations in comparison to the published amino acid sequence [5]. Underlining in the coding region of the DNA sequence indicates the segment used for preparation of the oligonucleotide probe. ○⟶ represents synthetic oligonucleotide primers. In the 3′-non-coding region, the consensus polyadenylation site is underlined.

Table 3

Codon usage for human plasminogen including the signal sequence

Total number of amino acids 811

Phe		Ser		Tyr		Cys	
TTT	10	TCT	9	TAT	10	TCT	25
TTC	11	TCC	20	TAC	20	TGC	23
Leu		TCA	10	Stop		Stop	
TTA	3	TCG	0	TAA	1	TGA	0
TTG	8			TAG	0	Trp	
						TGG	19
Leu		Pro		His		Arg	
CTT	10	CCT	23	CAT	10	CGT	3
CTC	5	CCC	16	CAC	14	CGC	6
CTA	4	CCA	25	Gln		CGA	3
CTG	18	CCG	5	CAA	13	CGG	5
				CAG	18		
Ile		Thr		Asn		Ser	
ATT	9	ACT	16	AAT	22	AGT	11
ATC	7	ACC	22	AAC	18	AGC	6
ATA	6	ACA	19	Lys		Arg	
Met		ACG	4	AAA	22	AGA	10
ATG	11			AAG	27	AGG	15
Val		Ala		Asp		Gly	
GTT	13	GCT	13	GAT	15	GGT	14
GTC	12	GCC	11	GAC	21	GGC	11
GTA	6	GCA	9	Glu		GGA	24
GTG	17	GCG	4	GAA	31	GGG	13
				GAG	25		

4. DISCUSSION

The amino acid sequence for human plasminogen predicted from the nucleotide sequence now presented shows the following discrepancies, besides those reported by Malinowski et al. [9], when compared to the sequence published by Sottrup-Jensen et al. [5]: an extra Ile in the amino-terminal region of the mature protein, a Gln at position 53 instead of Glu, and Asn at position 88 instead of the reported Asp. The presence of the extra isoleucine residue can be explained by polymorphism or more likely by being overlooked in the protein sequencing since complete liberation of isoleucine, upon hydrolysis, can require extensive hydrolysis. All other deviations in the amino acid sequence are most likely due to difficulties in assignments of amidated or non-amidated forms of the amino acids in the protein sequence. However, in view of recent work [21] a polymorphism cannot be ruled out. The NMR studies on kringle 1 of human plasminogen which suggest discrepancies in the number of methionine, phenylalanine, tyrosine, and histidine residues [20] cannot be confirmed from the predicted amino acid sequence of the cDNA clone now isolated. For detection of any heterogeneity, in kringle 1 due to allelic variations, further analysis of cDNA clones from more human individuals are required in order for this to be determined. The now presented amino acid sequence therefore essentially confirms the sequence presented by Sottrup-Jensen et al. [5].

Unlike many plasma proteins synthesized in the liver, human plasminogen does not seem to be synthesized with a pre-pro leader sequence. It is unlikely that the nucleotide sequence (fig.4) does not represent a full-length clone. The signal sequence predicted from the DNA sequence is a strong argument for the presence of a full-length clone.

The discrepancy in molecular mass found for the heavy chain is most likely due to the absence of the activation peptide in the heavy chain reported to have a molecular mass of 57 kDa. The overall codon usage is fairly random except that one codon, TCG for Ser, is not used at all.

A polyadenylation signal (AATAAA) is generally found within 6–26 bases before the polyadenylation site in most eukaryotic genes characterized so far [18]. Compared to the partial plasminogen cDNA described by Malinowski et al. [9] our clone contains an additional 29 bases at the 3′-end. No typical polyadenylation sequence is contained within these 29 bases. The consensus polyadenylation sequence AATAAA is thus located 46 bases upstream from the poly-A tail. A variation in the length of the 3′-non-coding region has been found previously, e.g. in cDNA clones corresponding to the β-subunit of alcohol dehydrogenase which were isolated from the same cDNA library [22].

Sadler et al. [23] have presented a partial genomic organization of kringles 4 and 5, and part of the light chain in human plasminogen. The cDNA clone for human plasminogen now

presented will facilitate isolation of genomic clones for a further characterization of the genomic organization of the human gene for plasminogen and its evolutionary relationship to other serine proteases.

ACKNOWLEDGEMENTS

We would like to thank Hans Jornvall and Lars Rutberg for valuable comments. This work was supported by KabiGen AB.

REFERENCES

[1] Robbins, K.C. (1982) in: Hemostasis and Trobosis (Colman, R.W. et al. eds) pp.623–639, J.B. Lippincott, Philadelphia.
[2] Wiman, B. and Wallen, P. (1975) Eur. J. Biochem. 50, 489–494.
[3] Wiman, B. and Wallen, P. (1975) Eur. J. Biochem. 58, 539–547.
[4] Wiman, B. (1977) Eur. J. Biochem. 76, 129–137.
[5] Sottrup-Jensen, L., Petersen, T.E. and Magnusson, S. (1978) in: Atlas of Protein Sequence and Structure (Dayhoff, M.O. ed.) vol.5, suppl.3, p.91, National Biomedical Research Foundation, Silver Spring, MD.
[6] Robbins, K.C., Summaria, L., Hsich, B. and Shah, R.J. (1967) J. Biol. Chem. 242, 2333–2342.
[7] Patthy, L. (1985) Cell 41, 657–663.
[8] Patthy, L., Trexler, M., Vali, Z., Banyai, L. and Varadi, A. (1984) FEBS Lett. 171, 131–136.
[9] Malinowski, D.P., Sadler, J.E. and Davie, E.W. (1984) Biochemistry 23, 4243–4250.
[10] Okayama, H. and Berg, P. (1982) Mol. Cell. Biol. 2, 161–170.
[11] Okayama, H. and Berg, P. (1983) Mol. Cell. Biol. 3, 280–289.
[12] Josephson, S., Lagerholm, E. and Palm, G. (1984) Acta Chem. Scand. B38, 539–545.
[13] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning. A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[14] Holmes, D.S. and Quigley, M. (1981) Anal. Biochem. 114, 193–197.
[15] Sanger, S., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[16] Crouse, G.F., Frischauf, A. and Lehrach, H. (1983) Methods Enzymol. 101, 78–89.
[17] Kozak, M. (1983) Microbiol. Rev. 47, 1–45.
[18] Wickens, M. and Stephenson, P. (1984) Science 226, 1045–1051.
[19] Von Heijne, G. (1983) Eur. J. Biochem. 133, 17–21.
[20] De Marco, A., Hochschwender, S.M., Laursen, R.A. and Llinas, M. (1982) J. Biol. Chem. 257, 12716–12721.
[21] Nishimukai, H., Shinmyozu, K. and Tamaki, Y. (1986) Hum. Hered. 36, 137–142.
[22] Hedén, L.-O., Hoog, J.-O., Larsson, K., Lake, M., Lagerholm, E., Holmgren, A., Vallee, B.L., Jornvall, H. and Von Bahr-Lindstrom, H. (1986) FEBS Lett. 194, 327–332.
[23] Sadler, J.E., Malinowski, D.P. and Davie, E.W. (1985) in: Progress in Fibrinolysis VII (Davidson, J.F. et al. eds) pp.201–204, Churchill Livingstone, Edinburgh.