# Identification of a novel human serpin gene; cloning sequencing and expression of leupin

Ruth C. Barnes, D. Margaret Worrall\*

Department of Biochemistry, University College Dublin, Belfield, Dublin 4, Ireland
Received 12 July 1995; revised version received 25 August 1995

Abstract A novel serpin gene has been isolated, cloned and sequenced. A PCR amplified fragment of the gene was originally identified from human genomic DNA, and the full-length cDNA was subsequently isolated from HeLa cells and sequenced. The novel serpin is very high in protein sequence similarity (91.8%) to the squamous cell carcinoma antigen (SCCA), but contains substantial differences in the reactive site loop sequence, including a different amino acid (leucine) in the  $P_1$  position. The gene product, named leupin, is expressed in HeLa cells, SKGIIIa cells and human placenta. The protein has a predicted  $M_r$  of 44,857 and an isoelectric point of 6.04 which is consistent with the more acidic form of SCCA associated with squamous cell carcinomas.

Key words: Serpin; Squamous cell carcinoma antigen; Leupin; cDNA sequence; HeLa cell

## 1. Introduction

The serpins (serine protease inhibitors) are a superfamily of proteins comprising over 60 members from a wide range of organisms, and includes well-characterised plasma proteins such as  $\alpha_1$ -antitrypsin and antithrombin III [1,2]. The inhibitory specificity of serpins is largely determined by residues at the P<sub>1</sub>-P<sub>1</sub>' positions within the reactive site loop region which acts as a pseudo-substrate for the target protease [3]. Serpins also include proteins which lack protease inhibitory activity such as chicken ovalbumin, angiotensinogen and hormone binding globulins. A sub-family of serpins related to ovalbumin, the ov-serpins, has been identified based on higher sequence identity (45-55%), gene organisation, and lack of an identifiable N-terminal signal sequence [4]. Human proteins belonging to this ov-serpin family include plasminogen activitor inhibitor-2 (PAI-2) [5], leucocyte elastase inhibitor (LEI) [6], placental thrombin inhibitor (PTI) [7], and squamous cell carcinoma antigen (SCCA) [8].

SCCA is a component of the T-4 antigen protein which was isolated from human uterine cervical squamous cell carcinoma tissue and is used as a serum tumour marker for the management of squamous cell carcinomas [9,10]. The antigen is also expressed in normal squamous epithelial cells [11] and is associ-

\*Corresponding author. Fax: (353) (1) 283-7211.

Abbreviations: SCCA, squamous cell carcinoma antigen; PAI-2, plasminogen activator inhibitor-2; LEI, leukocyte elastase inhibitor; PTI, placental thrombin inhibitor; RT-PCR, reverse transcription-polymerase chain reaction.

The sequence reported in this paper has been deposited in the EMBL database (Accession number X89015, mRNA for leupin).

ated with skin disorders such as psoriasis and eczema [12]. SCCA isolated from carcinoma of the cervix demonstrates heterogeneity on isoelectric focussing, with normal squamous epithelia containing a predominantely neutral form and carcinoma tissue containing both neutral and acidic forms [13]. The gene encoding for SCCA has been cloned and sequenced from SKGIIIa cells, a uterine cervical carcinoma cell line [8]. The cleavage site of SCCA is Ser–Ser which is unique to this serpin, and the protein has recently been shown to have inhibitory activity against cathepsin L and papain which are cysteine rather than serine proteases [14]. This cross-class inhibitory activity has previously been seen in the viral serpin crmA which inhibits the cysteine protease interleukin-1  $\beta$ -converting enzyme (ICE) [15,16].

This study set out to look for novel ov-serpins by amplifying PCR fragments between two conserved regions in the human ov-serpins. Degenerate oligonucleotide primers to conserved sequences flanking the variable reactive site loop region were used. We describe a novel serpin gene, leupin which has been isolated from HeLa cell cDNA, and which is high in sequence similarity to the squamous cell carcinoma antigen.

## 2. Materials and methods

#### 2.1. Materials

Oligonucleotide primers were synthesized and HPLC purified by Genosys, UK. MMLV reverse transcriptase and Taq polymerase were obtained from Promega. Dideoxy sequencing was performed using Sequenase, version 2.0, United States Biochemicals. SKGIIIa cells, a uterine cervical cancer epidermoid cell line [17] were kindly given by Dr. Shiro Nozawa, Keio University.

2.2. PCR amplification of serpin fragments from human genomic DNA Human genomic DNA was prepared from whole blood [18], and used as a template for amplification of serpin fragments. The following degenerate oligonucleotide primers complementary to SCCA, PAI-2 and EI sequences were synthesized: primer A: 5'-GGGGATCCCCNCGGTTCAAA(G/C)TNGAAGAG-3' (sense, corresponding to nucleotides 853-873 of SCCA and incorporating a BamHI restriction site); primer B: 5'-AAAAAGCTTCGGNGANGAA/GAATCTNCC-3' (antisense, corresponding to nucleotides 1153-1170 in SCCA, with HindIII site incorporated at 5' end). Amplification was carried out in a volume of 100 µl with 1.5 mM MgCl<sub>2</sub>, for 30 cycles of (94°C×1.5 min, 45°C×1 min, and 72°C×2 min). The product of 315 bp containing a mixture of ov-serpin fragments was digested with BamHI and HindIII restriction enzymes and cloned into pBluescript KS\*. Individual inserts were subjected to dideoxy sequencing [19] using M13 forward and reverse primers.

2.3. Detection of leupin gene expression

Expression of the leupin gene was examined by RT-PCR. Total RNA was isolated from a range of human cell lines and tissues using a phenol/guanidinium thiocyananate extraction method [20]. RNA (1 µg) was reverse transcribed with MMLV-reverse transcriptase using random hexanucleotide primers [21]. PCR amplification of this first strand cDNA was performed with a specific oligonucleotide to leupin,

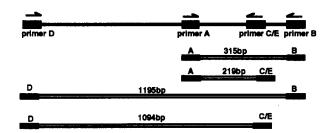


Fig. 1. PCR amplification strategy for isolation of ov-serpin fragments and for amplification of specific leupin and SCCA products. Primers A and B are degenerate sense and antisense primers based on the sequences of PAI-2, LEI and SCCA. Primers C and E are antisense primers to unique reactive site loop sequences of leupin and SCCA, respectively. Primer D is a sense primer to the 5' open reading frame sequence of SCCA and will also amplify from a leupin template.

primer C: 5'-GGGGGTACCTGAAGGAGATGATAATCGAC-3' (antisense primer corresponding to nucleotides 1054–1071 in leupin, and incorporating a *Kpn*I restriction endonuclease site at the 5' end). Initial studies to detect expression used primer A with primer C which yielded a product of 219 bp (Fig. 1).

## 2.4. Isolation of full-length cDNA from HeLa cell RNA

HeLa cell cDNA, prepared as described above, was used as a template to isolate the 5' end of the gene sequence. Primer D, 5'-GGGG-GATCCATGAATCACTCAGTGAAG-3' (sense primer corresponding to the 5' end of the SCCA open reading frame sequence, incorporating a BamHI site) and primer B were used to amplify the full 1195 bp product, and primers D and C were used to amplify a specific 1094 bp leupin product (see Fig. 1). These fragments were restriction digested, cloned into Bluescript KS<sup>+</sup>, and dideoxy sequencing of both strands was carried out. To facilitate internal sequencing, products were further digested with HindIII (sites at positions 259 and 766), and fragments were subcloned into Bluescript and subjected to dideoxy sequencing.

M N S L S E A N T K F M F D L F Q Q F R K S ATG AAT TCA CTC AGT GAA GCC AAC ACC AAG TTC ATG TTC GAT CTG TTC CAA CAG TTC AGA AAA TCA 66 K E N N I F Y S P I S I T S A L G M V L L G A K
AAA GAG AAC AAC ATC TTC TAT TCC CCT ATC AGC ATC ACA TCA GCA TTA GGG ATG GTC CTC TTA GGA GCC AAA 138 D N T A Q Q I S K V L H F D Q V T B N T T B K A GAC AAC AAC ACA ATT AGC AAG GTT CTT CAC TTT GAT CAA GTC ACA GAG AAC ACC ACA GAA AAA GCT 210 A T Y H V D R S G N V H H Q F Q K L L T E F N K
GCA ACA TAT CAT GTT GAT AGG TCA GGA AAT GTT CAT CAC CAG TTT CAA AAG CTT CTG ACT GAA TTC AAC AAA 282 S T D A Y E L K I A N K L F G E K T Y Q F L Q E
TCC ACT GAT GCA TAT GAG CTG AAG ATC GCC AAC AAG CTC TTC GGA GAA AAG ACG TAT CAA TTT TTA CAG GAA 354 Y L D A I K K F Y Q T S V E S T D F A N A P E E
TAT TTA GAT GCC ATC AAG AAA TTT TAC CAG ACC AGT GTG GAA TCT ACT GAT TTT GCA AAT GCT CCA GAA GAA 426 S R K K I N S W V E S Q T N E K I K N L F P D G
AGT CGA AAG AAG AAT AAC TCC TGG GTG GAA AGT CAA ACG AAT GAA AAA ATT AAA AAC CTA TTT CCT GAT GGG 498 T I G N D T T L V L V N A I Y F K G Q W B N K F ACT ATT GGC AAT GAT ACG ACA CTG GTT CTT GTG AAC GCA ATC TAT TTC AAA GGG CAG TGG GAG AAT AAA TTT 570 K K E N T K B E K F W P N K N T Y K S V Q M M R Q Y N S F N F A L L E D V Q A K V L E I P Y K G CAA TAC AAT TCC TTT AAT TTT GCC TTG CTG GAG GAT GTA CAG GCC AAG GTC CTG GAA ATA CCA TAC AAA GGC 714 K D L S M I V L L P N E I D G L Q K L E E K L T
AAA GAT CTA AGC ATG ATT GTG CTG CTG CCA AAT GAA ATC GAT GGT CTG CAG AAG CTT GAA GAG AAA CTC ACT 786 A E K L M E W T S L Q N M R E T C V D L H L P R GCT GAG AAA TTG ATG GAA AGT TTG CAG AAT ATG AGA GAG ACA TGT GTC GAT TTA CAC TTA CCT CGG 858 D A D L S G M T W S H G L S V S K V L H K A F V GAT GCA GAC CTC TCA GGC ATG ACC TGG AGC CAC GGT CTC TCA GTA TCT AAA GTC CTA CAC AAG GCC TTT GTG 1002 E V T E E G V E A A A A T A V V V V E L S S P S GAG GTC GCA GCT GCC ACC GCT GTA GTA GTA GTA GTA GTA TCA TCT CCT TCA 1074 T N E E F C C N H P F L F F I R Q N K T N S I L ACT AAT GAA GAG TTC TGT TAT CAC CCT TTC CTA TTC TTC ATA AGG CAA AAT AAG ACC AAC AGC ATC CTC 1146 F Y G R F S S P
TTC TAT GGC AGA TTC TCA TCC CCG 1170
B

## 2.5. Expression of SCCA and leupin

Further expression studies on human tissue were performed which detected a larger transcribed product of both SCCA and leupin. RT-PCR was carried out using the specific primer for leupin as described above or with a specific primer corresponding to the unique SCCA reactive site loop region sequence [8], primer E: 5' CCCCCGGGTA-CCTACGGGGATGAATCTG 3' (antisense to nucleotides 1054–1074 in SCCA). In each case, amplification with the sense primer D to the 5' open reading frame sequence was performed to yield a 1094 bp product.

## 3. Results

Sequence analysis of amplified ov-serpin fragments from human genomic DNA gave sequences of the known PAI-2, leucocyte elastase inhibitor and SCCA genes. The SCCA and LEI genes must therefore also contain this 315 bp 3' end of the open reading frame within the last exon as seen in the gene organisation of ovalbumin, gene Y and PAI-2 [22,23]. The fourth recognized ov-serpin, placental thrombin inhibitor was not amplified by the PCR primers used.

In addition to the recognized sequences, a SCCA-like 315 bp sequence was also obtained which contained major differences with SCCA, particularly in the predicted reactive site loop protein sequence. In order to look for expression of this novel gene and to isolate the full-length cDNA, an oligonucleotide primer to the unique reactive loop sequence was used. RT-PCR with this primer (primer C) and with primer A produced the expected 219 bp sequence from HeLa cell RNA. To isolate the full-length predicted ov-serpin sequence, it was assumed that the high sequence similarity with SCCA might also extend to the 5' end of the gene. Thus a primer corresponding to the first 18 nucleotides of the open reading frame of SCCA [8] was synthesized (primer D). Using the HeLa cell cDNA as a template, this yielded the expected product of 1195 bp with primer B, and nested PCR with primer C gave a 1094 bp product. Sequencing of these products gave the predicted open reading frame of the gene and protein sequence of the novel serpin (Fig. 2). The nucleotide sequence is 95.3% identical to the SCCA gene (Accession no. S66896) and the protein sequence shows 91.8% identity to SCCA (Fig. 3). The most concentrated region of divergence is the reactive site loop sequence which contains 6 amino acid differences including the  $P_1$  amino acid. The novel protein has been named leupin due to the presence of a leucine residue at the P<sub>1</sub> position.

Previous studies on SCCA expression used a nucleotide probe to detect the transcribed gene [8]. Digoxygenin labelled probes for leupin and SCCA were synthesised from the individ-



Fig. 3. Protein sequence alignment of leupin with SCCA. Amino acids which differ from the SCCA sequence are boxed. Leupin and SCCA share 95.3% nucleotide identity and 91.8% protein sequence identity. The reactive site loop sequence contains 6 amino acid differences within the 7 amino acids from  $P_4$  to  $P_3'$  (351–357) including a leucine residue (L354) in the predicted  $P_1$  position. Specific primers to this region of both genes were used for RT-PCR expression studies.

ual PCR products for use in Northern and Southern blot analysis. Southern blots using high stringency conditions showed crossreactivity of the probes with both cDNA species (data not shown). Therefore RT-PCR with specific primers was used as the method of choice for examining individual expression of these genes. Oligonucleotide primers corresponding to the unique reactive site loop regions of leupin and SCCA were synthesized in order to detect such expression by RT-PCR. Both SCCA and leupin are expressed in HeLa cells, SKGIIIa cells and human placenta (Fig. 4). However, the relative levels of expression are different with greater amounts of SCCA expressed in SKGIIIa cells (from which this gene was originally cloned) and with leupin showing higher levels of expression in HeLa cells and placenta than SCCA.

## 4. Discussion

A new serpin gene, leupin, has been identified using PCR between conserved regions of the ov-serpin subfamily of proteins. The gene product is expressed in human placenta and in SKGlla and HeLa cell lines. The predicted protein shows high sequence similarity to the squamous cell carcinoma antigen, is

Fig. 2. cDNA sequence and deduced protein sequence of leupin. Underlined regions indicate positions of PCR primers (from top, primer D, A, C and B). An original fragment of the gene was identified by amplification from genomic DNA between regions A and B using degenerate primers based on the conserved sequences of SCCA, PAI-2 and LEI in these regions. Expression of the gene was examined using a specific oligonucleotide (primer C) to the unique reactive site loop sequence of leupin. The full-length gene was obtained by amplification between primer B and primer D (based on SCCA sequence) using HeLa cell cDNA as a template. (Note that the given sequence of the underlined regions B and D is that of the incorporated oligonucleotide primers, and so the actual sequence may contain minor differences in these two regions.)

identical in length to SCCA (390 aa) and has a predicted  $M_{\rm r}$  of 44,857. There are 6 amino acid differences in the reactive loop sequence including the P<sub>1</sub> position which is a leucine residue in leupin, compared to a serine residue in SCCA. These key differences imply that leupin is not merely as isoform of SCCA, but is likely to have a different target protease or profile of target proteases. The two serpins represent a closer evolutionary divergence than previously seen in mammalian serpins which generally show 30–50% sequence identity [24], and must represent a relatively recent gene duplication event.

Alignment of the reactive site loop of leupin with other serpins (Fig. 5) shows a shared cleavage site with human  $\alpha_1$ -antichymotrypsin, an inhibitor of chymotrypsin and cathepsin G [25]. This would suggest that leupin may inhibit a chymotrypsin-like protease. However, SCCA has recently been found to inhibit the cysteine proteases cathepsin-L and papain and showed no inhibitory activity against a wide range of serine proteases [14], so it is perhaps more likely that leupin will also show inhibitory activity against a cysteine protease.

Purified SCCA has been shown to divide into two subgroups, an acidic form with pI<6.25, and a neutral form with a pI > 6.25. Monoclonal antibodies against seven epitopes recognised both forms of the antigen and one monoclonal antibody detected the acidic form only [10]. Further separation of SCCA by 2-D electrophoresis has shown 4 distinct spots [28]. This heterogeneity was assumed to be due to posttranslational modification since Southern blot analysis of genomic DNA originally suggested that only one copy of the gene was present [8], and there is some evidence that the most acidic spot is phosphorylated. However, our leupin sequence, which is also expressed in Hela and SKGIIIa cells, may explain the presence of heterogeneity in SCCA protein preparations. Indeed, the protein sequences obtained from endopeptidase fragments which were used for primer design in isolating the original SCCA gene sequence [8], are identical in the leupin sequence. The higher levels of SCCA expression found by us in SKGIIIa

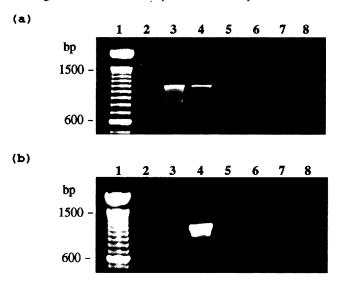


Fig. 4. Expression of (a) leupin and (b) SCCA in human tissues. Total RNA was isolated and RT-PCR was performed using either antisense primers C (leupin) or primer E (SCCA) with sense primer D. Products were separated on 1% agarose and detected by ethidium bromide staining. Lane 1 = 100 bp ladder; lane 2 = ECV 304 endothelial cells; lane 3 = HeLa cells; lane 4 = SKGIIIa cells; lane 5 = liver; lane 6 = placenta; lane 7 = small intestine; lane 8 = no template control.

	P14 P											Pı							P7'					
ŀ	ł	1	-1	- 1	1	1	-1	ł	-	1	ł	1	ł	į	1	ŧ	-1	-	1	1	1	-	1	
G	Α	E	A	A	A	A	T	A	٧	v	V	V	E	L	s	s	P	s	Т	N	E	B	F	Leupin
G	A	E	A	A	A	Α	т	A	V	v	G	F	G	S	S	P	A	s	T	N	E	Е	F	SCCA
G	$\mathbf{T}$	E	A	s	A	A	T	A	V	K	I	Т	L	L	s	A	L	v	Ε	Т	R	T	I	AChT
G	$\mathbf{T}$	E	A	A	A	G	T	G	G	٧	M	T	G	R	$\mathbf{T}$	G	Н	G	G	P	Q	-	F	PAI-2
G	S	Е	A	A	A	s	т	A	V	V	Ι	A	Ģ	R	S	L	N	P	N	R	V	T	F	ATIII
G	R	E	V	V	G	s	A	E	A	G	V	D	A	Α	S	v	-	s	-	-	E	E	F	Oval
	> <																							

Fig. 5. Alignment of reactive site loop sequences of leupin with other serpins; SCCA = squamous cell carcinoma antigen [8], AChT = human  $\alpha_1$ -antichymotrypsin [25], PAI-2 = plasminogen activator inhibitor-2 [5], ATIII = antithrombin III [26], Oval = chicken ovalbumin [27]. The putative  $P_1\text{-}P_1'$  cleavage site of leupin deduced from this alignment is Leu–Ser, which is shared by human  $\alpha_1$ -antichymotrypsin.

cells (Fig. 4) would favour the amplification and isolation of this cDNA species over that of leupin. The predicted isoelectric points of the unmodified proteins show leupin to have the lower pI of 6.04 which is consistent with the acidic form, whereas SCCA has a predicted pI of 6.63 consistent with the neutral form. Phosphorylated and dephosphorylated forms of both proteins could explain the presence of a total of 4 spots on 2-D electrophoresis. If this is the case, then previous biochemical and clinical studies on SCCA including the inhibition of cysteine proteases could be attributed to a mixture of SCCA and this second gene product, leupin.

Acknowledgements: We wish to thank Dr. Geraldine Butler and Dr. Luke O'Neill for helpful advice. This research was supported by the Alltech Newman Scholarship in Biotechnology and by the Health Research Board of Ireland.

#### References

- Hunt, L.T. and Dayhoff, M.O. (1980) Biochem. Biophys. Res. Commun. 95, 864–871.
- [2] Carrell, R.W. and Boswell, D.R. (1986) in: Proteinase Inhibitors (Barrett, A.J. and Salvesen, G. eds.) pp. 403–420, Elsevier, Amsterdam
- [3] Huber, R. and Carrell, R.W. (1989) Biochemistry 28, 8951-8966.
- [4] Remold-O'Donnell, E. (1993) FEBS Lett. 315, 105-108.
- [5] Antalis, T.M, Clark, M.A, Barnes, T, Lehrbach, P.R., Devine, P.L., Schevzov, G., Goss, N.H., Stephens, R.W. and Tolstoshev, P. (1988) Proc. Natl. Acad. Sci. USA 85, 985–989.
- [6] Remold-O'Donnell, E., Chin, J. and Alberts, M. (1992) Proc. Natl. Acad. Sci USA 89, 5635–5639.
- [7] Coughlin, P. Sun, J., Cerruti, L., Salem, H.H. and Bird, P. (1993) Proc. Natl. Acad. Sci. USA 90, 9417–9421.
- [8] Suminami, Y, Kishi, F., Sekiguchi, K. and Kato, H. (1991) Biochem. Biophys. Res. Commun. 181, 51-58.
- [9] Kato, H. and Torigoe, T. (1977) Cancer 40, 1621-1628.
- [10] Kato, H (1992) in: Serological Cancer Markers (Sell, S. ed.) pp. 437-451, Humana Press, Totowa, NJ.
- [11] Crombach, G., Scharl, A., Vierbuchen, M., Wurez, H. and Bolte, A. (1989) Cancer 63, 1337–1342.
- [12] Duk, J.M., van Voorst Vader, P.C., Ten Hoor, K.A., Hollema, H., Doeglas, H.M.G. and de Bruijn, H.W.A. (1989) Cancer 64, 1652– 1656.
- [13] Kato, H., Suehiro, Y. Morioka, H., Torigoe, T., Myoga, A., Sekiguchi, K. and Ikeda, I. (1987) Gann 78, 1246-1250.
- [14] Takeda, A., Yamamoto, T., Nakamura, Y., Takahashi, T. and Hibino, T. (1995) FEBS Lett. 359, 78–80.
- [15] Ray, C.A., Black, R.A., Kronheim, S.R., Greenstreet, T.A., Sleath, P.R., Salvesen, G.S. and Pickup, D.J. (1992) Cell 69, 202– 208.
- [16] Komiyama, T, Ray, C.A., Pickup, D.J., Howard, A.D., Thornberry, N.A., Peterson, E.P. and Salvesen, G.S. (1994) J. Biol. Chem. 269, 19331–19337.

- [17] Nozawa, S., Udagawa, Y., Ohta, H., Kurihara, S. and Fishman, W.H. (1983) Cancer Res. 43, 1748–1760.
- [18] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a Laboratory Manual (2nd edn.), Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.
- [19] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
- [20] Chomczynski, P. (1993) BioTechniques 15, 532-537.
- [21] Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1994) Current Protocols in Molecular Biology, Wiley, New York.
- [22] Heilig, R., Muraskowsky, R., Kloepfer, C. and Mandel, J.L. (1982) Nucleic Acids Res. 10, 705-712
- [23] Ye, R.D. Ahern, S.M., Le Beau, M.M., Lebo, R.V. and Sadler, J.E. (1989) J. Biol. Chem. 264, 5495–5502.
- [24] Marshall, C.J. (1993) Phil. Trans. Royal Soc. 342, 101-119.
- [25] Rubin, H., Wang, Z.M., Nickbarg, E.B., McIarney, S., Naido, N., Schoenberger, O.L., Johnson, J.L. and Cooperman, B.S. (1990) J. Biol. Chem. 265, 1199–1207.
- [26] Bock, S.C., Wion, K.L., Vehar, G.A. and Lawn, R.M. (1982) Nucleic Acids Res. 10, 8113–8215.
- [27] McReynolds, L., O'Malley, B.W., Nisbet, A.D., Fothergill, J.E., Givol, D., Fields, S., Robertson, M. and Brownlee, G.G. (1978) Nature 273, 723-728.
- [28] Abe, H., Okuno, N., Takeda, O., Suminami, Y., Kato, H. and Nakamura, K. (1994) Electrophoresis 15, 988-991.