

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 P₁ 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 α_1 -antitrypsin and antithrombin III [1,2]. The inhibitory specificity of serpins is largely determined by residues at the P₁–P₁' 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 activator 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-

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 predominantly 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 β -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'-GGGGGATCCCC-NCGGTTCAAA(G/C)TNGAAGAG-3' (sense, corresponding to nucleotides 853–873 of SCCA and incorporating a *Bam*HI restriction site); *primer B*: 5'-AAAAAGCTTCGGNGANGAA/GAATCTNCC-3' (antisense, corresponding to nucleotides 1153–1170 in SCCA, with *Hind*III site incorporated at 5' end). Amplification was carried out in a volume of 100 μ l with 1.5 mM MgCl₂, for 30 cycles of (94°C \times 1.5 min, 45°C \times 1 min, and 72°C \times 2 min). The product of 315 bp containing a mixture of ov-serpin fragments was digested with *Bam*HI and *Hind*III 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 thiocyanate extraction method [20]. RNA (1 μ g) 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,

*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).

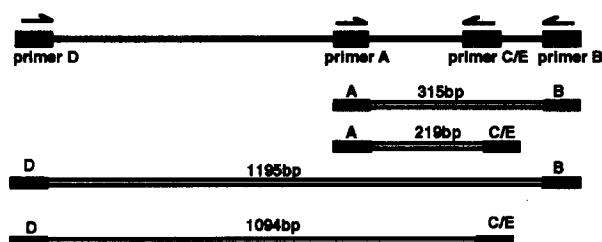


Fig. 1. PCR amplification strategy for isolation of ov-serpin fragments and for amplification of specific leupin and SCCA products. Primers A and B are degenerate sense and antisense primers based on the sequences of PAI-2, LEI and SCCA. Primers C and E are antisense primers to unique reactive site loop 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 *KpnI* 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⁺, 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
D

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 E N T T E K A
GAC AAC ACT GCA CAA CAA 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 ATT 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 E 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 E E K F W P N K N T Y K S V Q M M R
AAA AAA GAA AAC ACT AAA GAG GAA AAA TTT TGG CCA AAC AAG AAT ACA TAC AAA TCT GTA CAG ATG ATG AGG 642

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 TGG ACA AGT TTG CAG AAT ATG AGA GAG ACA TGT GTC GAT TTA CAC TTA CCT CGG 858

F K M E E S Y D L K D T L R T M G M V N I F N G
TTC AAA ATG GAA GAG AGC TAT GAC CTC AAG GAC ACG TTG AGA ACC ATG GGA ATG GTG AAT ATC TTC AAT GGG 930
A

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 ACT GAG GAG GGA GTG GAA GCT GCA GCT GCC ACC GCT GTA GTA GTA GTC GAA TTA TCA TCT CCT TCA 1074
C

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 TGT AAT 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 P₁ amino acid. The novel protein has been named leupin due to the presence of a leucine residue at the P₁ position.

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

1	M	N	S	L	S	E	A	N	T	K	F	M	F	D	L	F	Q	O	F	R	K	S	K	E	N	SCCA
1	M	N	S	L	S	E	A	N	T	K	F	M	F	D	L	F	Q	O	F	R	K	S	K	E	N	Leupin
26	N	I	F	Y	S	P	I	S	I	T	S	A	L	G	H	V	L	L	G	A	K	D	N	T	A	SCCA
26	N	I	F	Y	S	P	I	S	I	T	S	A	L	G	H	V	L	L	G	A	K	D	N	T	A	Leupin
51	Q	Q	I	K	K	V	L	H	F	D	Q	V	T	E	N	T	T	G	K	A	A	T	Y	H	V	SCCA
51	Q	Q	I	K	K	V	L	H	F	D	Q	V	T	E	N	T	T	G	K	A	A	T	Y	H	V	Leupin
76	D	R	S	G	N	V	H	H	Q	F	Q	K	L	L	T	E	F	N	K	S	T	D	A	Y	E	SCCA
76	D	R	S	G	N	V	H	H	Q	F	Q	K	L	L	T	E	F	N	K	S	T	D	A	Y	E	Leupin
101	L	K	I	A	N	K	L	F	G	E	K	T	Y	L	F	L	Q	E	Y	L	D	A	I	K	K	SCCA
101	L	K	I	A	N	K	L	F	G	E	K	T	Y	L	F	L	Q	E	Y	L	D	A	I	K	K	Leupin
126	F	Y	O	T	S	V	E	S	V	D	F	A	N	A	P	E	E	S	R	K	K	I	N	S	W	SCCA
126	F	Y	O	T	S	V	E	S	V	D	F	A	N	A	P	E	E	S	R	K	K	I	N	S	W	Leupin
151	V	E	S	O	T	N	E	K	I	K	N	L	I	P	E	G	N	I	G	S	N	T	T	L	V	SCCA
151	V	E	S	O	T	N	E	K	I	K	N	L	I	P	E	G	N	I	G	S	N	T	T	L	V	Leupin
176	L	V	N	A	I	Y	F	K	G	Q	W	E	K	K	F	N	K	E	D	T	K	E	E	K	F	SCCA
176	L	V	N	A	I	Y	F	K	G	Q	W	E	K	K	F	N	K	E	D	T	K	E	E	K	F	Leupin
201	W	P	N	K	N	T	Y	K	S	I	O	M	H	R	O	Y	T	S	F	H	F	A	S	L	E	SCCA
201	W	P	N	K	N	T	Y	K	S	I	O	M	H	R	O	Y	T	S	F	H	F	A	S	L	E	Leupin
226	D	V	O	A	K	V	L	E	I	P	Y	K	G	K	D	L	S	H	I	V	L	L	P	N	E	SCCA
226	D	V	O	A	K	V	L	E	I	P	Y	K	G	K	D	L	S	H	I	V	L	L	P	N	E	Leupin
251	I	D	G	L	Q	K	L	E	E	K	L	T	A	E	K	L	M	E	W	T	S	L	Q	N	H	SCCA
251	I	D	G	L	Q	K	L	E	E	K	L	T	A	E	K	L	M	E	W	T	S	L	Q	N	H	Leupin
276	R	E	T	R	V	D	L	H	L	P	R	F	K	V	E	E	S	Y	D	L	K	D	T	L	R	SCCA
276	R	E	T	R	V	D	L	H	L	P	R	F	K	V	E	E	S	Y	D	L	K	D	T	L	R	Leupin
301	T	M	G	M	V	D	I	F	N	G	D	A	D	L	S	G	M	T	G	S	R	G	L	V	L	SCCA
301	T	M	G	M	V	D	I	F	N	G	D	A	D	L	S	G	M	T	G	S	R	G	L	V	L	Leupin
326	S	G	V	L	H	K	A	F	V	E	V	T	E	E	G	A	E	A	A	A	T	A	V	V	SCCA	
326	S	G	V	L	H	K	A	F	V	E	V	T	E	E	G	A	E	A	A	A	T	A	V	V	Leupin	
351	G	F	G	S	S	P	A	S	T	N	E	E	F	H	C	N	H	P	F	L	F	F	I	R	Q	SCCA
351	G	F	G	S	S	P	A	S	T	N	E	E	F	H	C	N	H	P	F	L	F	F	I	R	Q	Leupin
376	N	K	T	N	S	I	L	F	Y	G	R	F	S	S	P											SCCA
376	N	K	T	N	S	I	L	F	Y	G	R	F	S	S	P											Leupin

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 P₄ to P₃' (351–357) including a leucine residue (L354) in the predicted P₁ 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 SKGIIa 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.)

- [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., Klopfer, 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., McLarney, 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.