Volume 207, number 1 FEBS 4108 October 1986

A novel member of the serpin superfamily is encoded on a circular plasmid-like DNA species isolated from rabbit cells

C. Upton, R.W. Carrell* and G. McFadden

Department of Biochemistry, University of Alberta, Edmonton, Alberta, T6G 2H7 Canada and **Haematology Department, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, England*

Received 3 August 1986

A novel member of the serpin family of serine protease inhibitors is presented. A plasmid-like DNA was isolated from rabbit cells by its homology to the genome of Shope fibroma virus (SFV), a tumorigenic poxvirus of rabbits, and was shown elsewhere to encode a serpin-like protein [(1986) Mol. Cell. Biol. 6, 265–276]. Although significant DNA homology exists between the rabbit plasmid serpin open reading frame and the SFV terminal inverted repeat DNA there is no intact serpin counterpart encoded by this region of the SFV genome. The alignment of the novel plasmid-borne polypeptide with the serpin family of proteins confirms its status within this group.

Shope fibroma virus *Poxvirus* *spc DNA* *Serpin* *Serine protease inhibitor*

1. INTRODUCTION

Members of the serpin (serine protease inhibitors) superfamily (review [1,2]) comprise a variety of distinct but undoubtedly related proteins, not all of which have a known protease inhibitor function. Recently several of the blood plasma protease inhibitors have attracted considerable attention in terms of the therapeutic use of genetically engineered serpins to treat genetic deficiencies, notably that of plasma protein α_1-antitrypsin [3]. Here we describe the serendipitous discovery of a novel member of the serpin superfamily from rabbit cells. This serpin-like protein is derived from the translation of an open reading frame (ORF) on an extrachromosomal plasmid-like DNA species isolated from rabbit cells [4]. This plasmid DNA species was cloned by virtue of its DNA sequence homology to the genome of Shope fibroma virus (SFV), a tumorigenic poxvirus of rabbits [4]. The presence of small polydisperse extrachromosomal circular (spc) DNAs in eukaryotic cells has been observed for a number of years (review [5,6]). The size distribution and number of spc DNAs is found to vary with development, growth state and mitotic capacity of the cell, but the function of these molecules is poorly understood [5–7]. SFV is a tumorigenic leporipoxvirus which is capable of inducing fibromas in rabbits, its natural host [8–10]. Like all poxviruses, SFV replicates in the cytoplasm of infected cells and its large (160 kb) linear ds DNA genome [10] probably encodes all of the proteins required for the replication and transcription of the viral DNA [11,12]. The relationship between SFV and spc DNAs of the host cell is unclear but recombination with these host DNAs may be one way by which the cytoplasmically replicating poxviruses are able to enhance their pool of genetic information [4].

The rabbit sequence described here was discovered when the cloned *Bam*HI fragments of the SFV genome were examined for the presence of host-related sequences which might be implicated in the tumorigenic phenotype of this virus [4]. Unexpectedly, it was found that a 4.8 kb covalently closed circular (CCC) species of DNA, present in both total and Hirt preparations of rabbit (SIRC cell) DNA, hybridized to recombinant plasmids derived from a defined region of the terminal inverted repeats (TIR) of SFV [4]. Analysis of the

Published by Elsevier Science Publishers B.V. (Biomedical Division)

DNA sequence of one clone (pSIC9) which contains 1.9 kb of the plasmid sequence confirmed the homology with SFV TIR sequences found by Southern blotting and revealed the presence of one intact and one truncated ORF [4]. The truncated ORF (ORF-2) was found to be identical to the N-terminus of ORF-T8 in the SFV TIR, but the function of this viral/plasmid gene product has not yet been clarified. The intact ORF (ORF-1) from pSIC9 has no complete counterpart in the SFV TIR, but significant DNA homology exists between the two sequences in this region, although several small gaps plus one of 819 bases must be inserted in order to align these homologies. Translation of this rabbit plasmid ORF-1 yields a 361 amino acid polypeptide which has some homology with human α_1-antichymotrypsin, as determined by computer search through the National Biochemical Research Foundation protein database [4]. Here we report the complete amino acid sequence of the rabbit plasmid ORF-1 and confirm that it represents a novel bona fide member of the serpin superfamily.

2. MATERIALS AND METHODS

The DNA sequence for pSIC9 has been reported [4]. Analysis of the DNA sequence, protein database searches and alignment of the polypeptides were performed using the core library programs of the BIONET computer resource (IntelliGenetics Inc.). A final alignment of the ORF-1 polypeptide to other members of the serpin family utilized the programme of A.M. Lesk, M. Levitt and C. Chothia [1] which uses the Needleman-Wunsch algorithm modified to penalise insertion of gaps in regions of secondary structure based on that of α_1-antitrypsin.

3. RESULTS AND DISCUSSION

The complete amino acid and DNA sequences with the 5′- and 3′-flanking DNA regions of the rabbit plasmid ORF-1 are shown in fig.1. It should

```
        10         20         30         40         50         60         70
ACAGGCAAGA CACGTTGGAT CGGTTGAAGG AACGACTCGT GGGACGCGTG ATTAAAACGC GAGTCGTTCG

        80         90        100        110        120        130        140
CGCGGACGGT TTATACGTGG ACCTGCGACG TTTTTTTTGA GGGTTAAATG AAGTATCTGG TCCTCGTCTT

                155                 170                 185
ATG TTT AAC GTC GTG CGC GTT CGA GAT ATC GGA CTA TGG ACG TTC CGA TAC GTC
MET Phe Asn Val Val Arg Val Arg Asp Ile Gly Leu Trp Thr Phe Arg Tyr Val

    200                 215                 230                 245
TAC AAC GAA AGC GAC AAC GTC GTG TTC TCA CCG TAC GGC TTG ACC TCC GCG TTG
T/r Asn Glu Ser Asp Asn Val Val Phe Ser Pro Tyr Gly Leu Thr Ser Ala Leu

            260                 275                 290
TCC GTG TTA CGG ATC GCG GCG GGC GGT AAC ACG AAA CGA GAA ATA GAC GTC CCC
Ser Val Leu Arg Ile Ala Ala Gly Gly Asn Thr Lys Arg Glu Ile Asp Val Pro

305                 320                 335                 350
GAA TCC GTC GTG GAG GAC TCC GAC GCC TTT CTC GCG TTA CGG GAG TTG TTC GTA
Glu Ser Val Val Glu Asp Ser Asp Ala Phe Leu Ala Leu Arg Glu Leu Phe Val

        365                 380                 395                 410
GAC GCA TCC GTT CCG TTA CGT CCC GAG TTT ACG GCG GAG TTC TCC TCG CGA TTC
Asp Ala Ser Val Pro Leu Arg Pro Glu Phe Thr Ala Glu Phe Ser Ser Arg Phe

                425                 440                 455
AAT ACC TCC GTG CAA CGC GTG ACG TTT AAC TCG GAG AAC GTC AAA GAC GTC ATT
Asn Thr Ser Val Gln Arg Val Thr Phe Asn Ser Glu Asn Val Lys Asp Val Ile

    470                 485                 500                 515
AAC TCG TAC GTT AAG GAT AAG ACG GGA GGA GAC GTC CCA CGC GTA TTG GAC GCC
Asn Ser Tyr Val Lys Asp Lys Thr Gly Gly Asp Val Pro Arg Val Leu Asp Ala

            530                 545                 560
TCC CTA GAC CGA GAT ACT AAA ATG CTG CTA TTG AGC TCC GTT CGT ATG AAG ACG
Ser Leu Asp Arg Asp Thr Lys Met Leu Leu Leu Ser Ser Val Arg Met Lys Thr

575                 590                 605                 620
AGC TGG AGA CAC GTA TTC GAC CCT TCG TTC ACG ACG GAT CAA CCT TTT TAT TCC
Ser Trp Arg His Val Phe Asp Pro Ser Phe Thr Thr Asp Gln Pro Phe Tyr Ser

        635                 650                 665                 680
GGA AAC GTC ACA TAC AAG GTA CGT ATG ATG AAT AAA ATA GAT ACG TTG AAA ACG
Gly Asn Val Thr Tyr Lys Val Arg Met Met Asn Lys Ile Asp Thr Leu Lys Thr

                695                 710                 725
GAG ACG TTT ACG CTT AGA AAC GTG GGA TAC TCC GTA ACG GAA CTG CCG TAT AAA
Glu Thr Phe Thr Leu Arg Asn Val Gly Tyr Ser Val Thr Glu Leu Pro Tyr Lys

    740                 755                 770                 785
CGG CGT CAA ACG GCC ATG TTG CTC GTC GTT CCG GAC GAC TTG GGA GAG ATC GTG
Arg Arg Gln Thr Ala Met Leu Leu Val Val Pro Asp Asp Leu Gly Glu Ile Val
            800                 815                 830
CGG GCC CTC GAT CTT TCT CTA GTA CGC TTC TGG ATA CGC AAC ATG AGG AAA GAC
Arg Ala Leu Asp Leu Ser Leu Val Arg Phe Trp Ile Arg Asn Met Arg Lys Asp

845                 860                 875                 890
GTG TGT CAG GTG GTA ATG CCC AAG TTC TCC GTC GAA TCG GTC CTG GAT CTG AGG
Val Cys Gln Val Val Met Pro Lys Phe Ser Val Glu Ser Val Leu Asp Leu Arg

        905                 920                 935                 950
GAC GCC CTC CAG AGA CTG GGG GTG CGA GAC GCG TTC GAT CCA TCC CGG GCG GAC
Asp Ala Leu Gln Arg Leu Gly Val Arg Asp Ala Phe Asp Pro Ser Arg Ala Asp

                965                 980                 995
TTC GGT CAG GCG TCC CCG TCG AAC GAT CTA TAC GTC ACG AAG GTG TTA CAG ACG
Phe Gly Gln Ala Ser Pro Ser Asn Asp Leu Tyr Val Thr Lys Val Leu Gln Thr

   1010                1025                1040                1055
TCC AAG ATA GAG GCG GAC GAA CGG GGA ACG ACG GCG TCG AGC GAC ACA GCC ATC
Ser Lys Ile Glu Ala Asp Glu Arg Gly Thr Thr Ala Ser Ser Asp Thr Ala Ile

           1070                1085                1100
ACC CTC ATC CCC AGG AAC GCC CTC ACG GCG ATC GTG GCG AAC AAA CCG TTT ATG
Thr Leu Ile Pro Arg Asn Ala Leu Thr Ala Ile Val Ala Asn Lys Pro Phe Met

1115               1130                1145                1160
TTT CTC ATC TAT CAC AAG CCT ACA ACG ACC GTG TTG TTT ATG GGA ACG ATA ACA
Phe Leu Ile Tyr His Lys Pro Thr Thr Thr Val Leu Phe Met Gly Thr Ile Thr

       1175                1190                1205                1220
AAG GGT GAA AAA GTA ATA TAC GAT ACG GAG GGT CGA GAT GAT GTC GTA TCC TCT
Lys Gly Glu Lys Val Ile Tyr Asp Thr Glu Gly Arg Asp Asp Val Val Ser Ser

              1236       1246       1256       1266       1276       1286
GTA TAA ACTCTTTTTG AAGGGTAAAC TATGCGACGT CGAAATCGTC GCGGAAGGCA AAAGCATCCG
Val  .
```

Fig.1. Amino acid sequence of ORF-1, translated from the rabbit plasmid clone pSIC9. The full DNA sequence of clone pSIC9 (complementary strand) is presented in fig.6 of [4] where nucleotide no.1678 corresponds to no.1 of this figure.

```
Rabbit ORF-1                                                                               MFNVVRVRDIGLWTFRYVYNES---DNVVF
Human α1-AT                                                         EDPQGDAAQKTDTSHHDQDHPTFNKITPNLAEFAFSLYRQLAHQSNS-TNIFF
Mouse α1-AT
Mouse contrapsin
Human α1-AC                                                          PNSPLDEENLTQENQDRGTHVDLGLASANVDFAFSLYKQLVLKALD-KNVIF
Human AT3                                  HGSPVDICTAKPRDIPMNPMCIYRSPEKKATEDEGSEQKIPEATNRRVWELSKANSRFATTFYQHLADSKNDNDNIFL
Chicken ovalbumin                                                                          MGSIGAASMEFCFDVFKELKVHHAN-ENIFY
Chicken gene Y                                                                             MDSISVTNAKFCFDVFNEMKVHHVN-ENILY
Rat angiotensinogen  DRVYIHPFHLLYYSKSTCAQLENPSVETLPEPTFEPVPIQAKTSPVDEKTLRDKLVLATEKLEAEDRQRAAQVAMIANFMGFRMYKMLSEARGVASGAVL
                                                                                        [-----------------------]   [-
                                                                                                Helix A             Strand B6

Rabbit ORF-1         SPYGLTSALSVLRIAAGGNTKREIDVPESVV-ED----------------------------------SDAFLALRELFVDASVPLRPEFTAEFSSRFNT
Human α1-AT          SPVSIATAFAMLSLGTKADTHDEILEGLNFN-LTEI---------PEAQIHEGFQELLRTLNQPDSQ--LQLTTGNGLFLSEGLKLVDKFLEDVKKLYHS
Mouse α1-AT
Mouse contrapsin
Human α1-AC          SPLSISTALAFLSLGAHNTTLTEILKASSSP-HGDL---------LRQKFTQSFQHLRAPSISSSDE--LQLSMGNAMFVKEQLSLLDRFTEDAKRLYGS
Human AT3            SPLSISTAFAMTKLGACNDTLQQLMEVFKFDTISEK---------TSDQIHFFFAKLNCRLYRKANK-SSKLVSANRLFGDKSLTFNETYQDISELVYGA
Chicken ovalbumin    CPIAIMSALAMVYLGAKDSTRTQINKVVRFDKLPGFGDSIEAQCGTSVNVHSSLRDILNQITKPNDV--YSFSLASRLYAEERYPILPEYLQCVKELYRG
Chicken gene Y       CPLSILTALAMVYLGARGNTESQMKKVLHFDSITGAGSTTDSQCGSSEYVHNLFKELLSEITRPNAT--YSLEIADKLYVDKTFSVLPEYLSCARKFYTG
Rat angiotensinogen  SPPALFGTLVSFYLGSLDPTASQLQVLLGVPVKEGDCTSRLDGH-KVLTALQAVQGLLVTQGGSSSQTPLLQSTVVGLFTAPGLRLKQPFVESLGPFTPA
                     -][-------------][----------]                [----------------]      [-----------]    [-----------][-
                            Helix B      Helix C                        Helix D             Strand A2         Helix E

                                                                                                                  PT
Rabbit ORF-1         S-VQRVTFN---SENVKDVINSYVKDKTGGDVPRVLD-ASLDRDTKMLLLSSVRMKTSWRHVFDPSFTTDQPFY-SGNVTYKVRMMNKIDTL-KTFTLRN
Human α1-AT          E-AFTVNFG-D-TEEAKKQINDYVEKGTQGKIVDLV--KELDRDTVFALVNYIFFKGKWERPFEVKDTEEEDFHVDQVTTVKVPMMKRLGMF-NIQHCKK
Mouse α1-AT                                                         SPANYILFKGKWKKPFDPENTEEAEFHVDESTTVKVPMMTLSGML-DVHHCST
Mouse contrapsin                                                   VVLVNYIYFKGKWKISFDPQDTFESEFYLDEKRSFKVPMMKMKLLTTRHFRDEE
Human α1-AC          E-AFATDFQ-D-SAAAKKLINDYVKNGTRGKITDLI--KDPDSQTMMVLVNYIFFKAKWEMPFDPQDTHQSRFYLSKKKWVMVPMMSLHHLTIPYFRDEE
Human AT3            K-LQPLDFK-ENAEQSRAAINKWVSNKTEGRITDVIPSEAINELTVLVLVNTIYFKGLWKSKFSPENTRKELFYKADGESCSASMMYQEGKF-RYRRVAE
Chicken ovalbumin    G-LEPINFQ-TAADQARELINSWVESQTNGIIRNVLQPSSVDSQTAMVLVNAIVFKGLWEKAFKDEDTQAMPFRVTEQESKPVQMMYQIGLF-RVASMAS
Chicken gene Y       G-VEEVNFK-TAAEEARQLINSWVEKETNGQIKDLLVSSSIDFGTTMVFINTIYFKGIWKIAFNTEDTREMPFSMTKEESKPVQMMCMNNSF-NVATLPA
Rat angiotensinogen  IFPRSLDLSTD-PVLAAQKINRFVQAVTGWKMNLPL--EGVSTDSTLFFNTYVHFQGKM-RGFSQL-TGLHEFWVDNSTSVSVPMLSGTGNF--QHWSDA
                     -------]    [---------------]                [-----------]                     [----------]  [----]
                      Strand A1        Helix F                      Strand A3                         Strand C3  Strand B1

                                         Y                                                                 R
Rabbit ORF-1         VGYSVTELPYKR-RQTAMLLVPD----DLGEIVRALDLSLVRFWIRNM--RKDVCQVVMPKFSVESVLDL-RDALQRLGVRDAFDPSAD-FGQASPSN--
Human α1-AT          LSSWVLLMKYLG-NATAIFFLPD--EGKLQHLENELTHDIITKFLENE--DRRSASLHLPKLSITGTYDL-KSVLGQLGITKVFSNGAD-LSGVTEEA--
Mouse α1-AT          LSSWVLLMDYAG-NATAVFLLPD--DGKMQHLEQTLSKELISKFLLNR--RRRLAQIHFPRLSISGEYNL-KTLMSPLGITRIFNNGAD-LSGITEENA-
Mouse contrapsin     LSCSVLELKYTG-NASALLILPD--QGRMQQVEASLQPETLRKWRKTLF-PSQIEELNLPKFSIASNYRLEEDVLPEMGIKEVFTEQAD-LSGIIETK--
Human α1-AC          LSCTVVELKYTG-NASALFILPD--QDKMEEVEAMLLPETLKRWRDSLEF-REIGELYLPKFSISRDYNL-NDILLQLGIEEAFTSKAD-LSGITGAR--
Human AT3            -GTQVLELPFKGDDITMVLILPKP-EKSLAKVEKELTPEVLQEWLDEL--EEMMLVVHMPRFRIEDGFSL-KEQLQDMGLVDLFSPEKSKLPGIVAEGRD
Chicken ovalbumin    EKMKILELPFASGTMSMLVLLPDE-VSGLEQLESIINFEKLTEWTSSNVMEERKIKVYLPRMKMEEKYNL-TSVLMAMGITDVFSSSAN-LSGISSAE--
Chicken gene Y       EKMKILELPYASGDLSMLVLLPDE-VSGLERIEKTINFDKLREWTSTNAMAKKSMKVYLPRMKIEEKYNL-TSILMALGMTDLFSRSAN-LTGISSVD--
Rat angiotensinogen  QNNFSVTRVPLG-ESVTLLLIQPQCASDLDRVEVLVFQHDFLTWIKNP--PPRAIRLTLPQLEIRGSYNL-QDLLAQAKLSTLLGAEAN-LGKMGDTN--
                      [--------]   [-------]   [--------][------]        [------][----------][------]
                       Strand B2   Strand B3    Helix G  Helix H         Strand C2 Strand A6  Helix I

                                                                                    TITKGE
                                                                                      \/
Rabbit ORF-1         DLYVTKVLQTSKIEADERGTTASSDTAITLIP-RNALTA----IVANKPFMFLIYHKPTTTVLFMGKVIYDTEGRDDVVSSV
Human α1-AT          PLKLSKAVHKAVLTIDEKGTEAAGAMFLEAIP-MSIPPE----VKFNKPFVFLMIEQNTKSPLFMGKVVNPTQK
Mouse α1-AT          PLKLSQAVHKAVLTIDETGTEAAAVTVLLAVP-YSMPPI----LRFDHPFLFIIFEEHTQSPLFVGKVVDPTHK
Mouse contrapsin     KLSVSQVVHKAVLDVAETGTEAAAATGVIGGIRKAILPA----VHFNRPFLFVIYHTSAQSILFMAKVNNPK
Human α1-AC          NLAVSQVVHKVVSDVFEEGTEASAATAVKITL-LSALVETRTIVRFNRPFLMIIVPTDTQNIFFMSKVTNPSKPRACIKQWGSQ
Human AT3            DLYVSDAFHKAFLEVNEEGSEAAASTAVVIAG-RSLNPN-RVTFKANRPFLVFIREVPLNTIIFMGRVANPCVK
Chicken ovalbumin    SLKISQAVHAAHAEINEAGREVVGSAEAGVDA-ASVS-EE---FRADHPFLFCIKHIATNAVLFFGRCVSP
Chicken gene Y       NLMISDAVHGVFMEVNEEGTEATGSTGAIGNIKHSLELEE---FRADHPFLFFIRYNPTNAILFFGRYWSP
Rat angiotensinogen  P-RVGEVLNSILLEL-QAGEEEQPTESAQQ-P-GSPEVLD---VTLSSPFLFAIYERDSGALHFLGRVDNPQNVV
                      [--------------][-------------]   [-----------][--------][---------]
                          Strand A5      Strand A4            Strand C1  Strand B4  Strand B5
```

Fig.2. Alignment of the rabbit ORF-1 amino acid sequence with eight members of the serpin superfamily.

be noted that this sequence is translated from a single continuous pSIC9 ORF DNA sequence. In comparison, introns are present within the genes of ovalbumin [13], α_1-antitrypsin [14], angiotensinogen [15] and antithrombin III [16,17], although only in angiotensinogen and antithrombin III are they similarly arranged [15]. The absence of introns from ORF-1 indicates that this rabbit plasmid species may have been originally derived by reverse transcription. In fig.2 the protein sequences of the rabbit plasmid ORF-1, human α_1-antitrypsin (α_1-AT) [18], mouse α_1-AT [19], human α_1-antichymotrypsin (α_1-AC) [20], mouse contrapsin [19], human antithrombin III (AT3) [21], chicken ovalbumin [22], chicken gene Y [13] and rat angiotensinogen [15] have been aligned to minimize the insertion of gaps within regions of secondary structure based on that of α_1-AT. There are 18 residues which are conserved through all the serpins shown and in 133 positions the amino acid is conserved in at least 50% of the proteins. As expected from the alignment of the other serpins there is considerable variation at the N-terminus of the rabbit plasmid ORF-1. An unexpected finding is the complete deletion of the D helix, confirmed by the presence of the conserved sequence in the C helix and the A2 strand that flank each side of the missing D helix. Evidence that the overall conformation is retained at the reactive center is provided by the homologies of the A4 strand region (fig.3). This strand forms the loop that must be present in the native serpins [1] to join the P_1 and P_1' residues of the reactive center. There is strong homology throughout the family at the hinge region of this loop with the typical sequence Glu:polar:Gly:Thr/Ser:Glu

| Hinge | Exposed loop | Reactive centre | Leaving peptide → |

Residue number based on α_1-antitrypsin	342	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	-	58	59	60	61	62	63	-	-	-	-	64	65	66	67	68	69	70	71	72	Score
	A5								A4									P$_1$	P$_1$'						C$_1$										B$_4$		(19)
Rabbit plasmid ORF-1	E	R	G	T	T	A	S	S	D	T	A	I	T	L	I	P	-	R	N	A	L	T	A	-	-	-	-	I	V	A	N	K	P	F	M	F	12
Human α_1 AT	E	K	G	T	E	A	A	G	A	M	F	L	E	A	I	P	-	M	S	I	P	P	E	-	-	-	-	V	K	F	N	K	P	F	V	F	18
Baboon α_1 AT	E	K	G	T	E	A	A	G	A	M	F	L	E	A	I	P	-	M	S	I	P	P	E	-	-	-	-	V	K	F	N	K	P	F	V	F	18
Mouse α_1AT	E	T	G	T	E	A	A	A	V	T	V	L	L	A	V	P	-	Y	S	I	P	P	E	-	-	-	-	L	R	F	D	H	P	F	L	F	18
Mouse α_1 AC (contrapsin)	E	T	G	T	E	A	A	A	A	T	G	V	I	G	G	I	R	K	A	I	L	P	A	-	-	-	-	V	H	F	N	R	P	F	L	F	17
Human α_1 AC	E	E	G	T	E	A	S	A	A	T	A	V	K	I	T	L	-	L	S	A	L	V	E	T	R	T	I	V	R	F	N	R	P	F	L	M	18
Human AT-III	E	E	G	S	E	A	A	A	S	T	A	V	V	I	A	G	-	R	S	L	N	P	N	-	R	V	T	F	K	A	N	R	P	F	L	V	18
Chicken Ovalb.	E	A	G	R	E	V	V	G	S	A	E	A	G	V	D	A	·	A	S	V	S	-	E	E	-	-	-	F	R	A	D	H	P	F	L	F	14
Chicken Gene Y	E	E	G	T	E	A	T	G	S	T	G	A	I	G	N	I	K	H	S	L	E	L	E	E	-	-	-	F	R	A	D	H	P	F	L	F	18
Rat AngioT	Q	A	G	E	E	E	Q	P	T	E	S	A	Q	Q	-	P	-	G	S	P	E	V	L	D	-	-	-	V	T	L	S	S	P	F	L	F	13
Human AngioT	E	A	-	D	E	R	E	P	T	E	S	T	Q	Q	-	L	-	N	K	P	E	V	L	E	-	-	-	V	T	L	N	R	P	F	L	F	11
Barley Z Prot.	E	E	G	T	E	A	G	A	A	T	V	A	M	G	V	A	-	M	S	M	P	L	K^{V}	D	L	V	D	F	V	A	N	H	P	F	L	F	16
Human α_2 AP	E	V	G	V	E	A	A	A	A	T	S	I	A	M	S	-	-	R	M	S	L	S	S	-	-	-	-	F	S	V	N	R	P	F	L	F	17
Human PA-I	E	S	G	T	A	S	S	S	T	A	V	I	V	S	A	-	R	M	A	P	E	E															
Human Hep Cof-II															P	P	-	L	S	T	Q	V	R	-	-	-	-	F	T	V	D	R	P	F	L	F	
Human C_1 Inhib	E	T	G	V	E	A	A	A	A	S	A	I	-	S	V	A	-	R	T	L	L	V	-	-	-	-	-	F	E	V	Q	Q	P	F	L	F	18
Conserved	E	⊕	G	T/S	E	A	-	A A/T T	S	S/	-	-	-	-	-	-	-	-	S/T	-	-	-	-	-	-	-	-	⊖	⊕	⊖	⊕	⊕	P	F	⊖	F	19
Consensus Sequence	E	E	G	T	E	A	A	A	A	T	S	V	V	X	V	P	(R)	X	S	L	X	V	E	E	-	-	-	F	R	L	N	R	P	F	L	F	

⊕ represents hydrophilic ; ⊖ represents hydrophobic

Fig.3. Alignment of the reactive centers of 16 members of the serpin superfamily [1]. The normal cleavage site for the proteinases inhibited by the serpins is after the P_1 residue. Conserved and consensus sequences are indicated.

which is conserved in ORF-1 as Glu:Arg: Gly:Thr:Thr. Furthermore, the ORF-1 sequence has the typical distance of 16 residues between the hinge and the putative P_1 reactive center residue, with the implication that the protein retains the stretched loop conformation necessary for inhibitory function. The common tertiary structure of the serpins is also supported by the even distribution of conserved residues throughout their sequence which can be correlated with the structural features of the α_1-AT model. Examples of this are the conservation of glycine residues at interstrand hinges and salt bridges linking areas of secondary structure. Two Glu-Lys bridges have been maintained in most serpins (fig.2; G7-$B_5$9 and $A_5$16-$A_6$1) with the exception of angiotensinogen, barley protein Z and ORF-1, each of which has lost one complete bridge.

The alignment of the reactive center region of the serpins is shown in fig.3 where more distantly related proteins are included [1,23–27] in addition to a conserved and consensus sequence. The structure of each of the serpins provides a clue to its inhibitory activity and likely target proteinase. The P_1 residue of the reactive center acts as a potential cleavage site for the target enzyme with a specificity that is further increased by the P_2 residue. This is of course an approximate rule but it does allow an estimate as to the target enzymes of serpins of dubious inhibitory activity. Thus the rabbit plasmid ORF-1 fits the description of an Argserpin and if it does have an inhibitory function its target is likely to be a serine protease cleaving after arginine, probably at a Pro:Arg:X sequence.

DNA homology between the SFV TIR and pSIC9 extends across the full 1.9 kb plasmid insert. However, it is obvious that only one of the ORFs (ORF-2, equivalent to SFV ORF-T8) present on pSIC9 has been maintained by SFV [4]. Thus it appears that SFV may have originally acquired a large fraction of the rabbit plasmid sequences but those genes which did not confer a selectable advantage to the virus (including the serpin ORF) were not faithfully maintained in the SFV genome.

ACKNOWLEDGEMENTS

This work was supported by the Alberta Heritage Foundation for Medical Research (AHFMR), the National Cancer Institute of Canada and the Alberta Cancer Board. Computer costs of the BIONET resource were funded by Public Health Service grant 1-441-RR01685-01 from the National Institutes of Health. G.M. is an AHFMR scholar and C.U. a postdoctoral fellow. R.W.C. was supported by the Medical Research Council of New Zealand.

REFERENCES

[1] Carrell, R.W. and Boswell, D.R. (1986) in: Proteinase Inhibitors (Barrett, A. and Salvesen, G. eds) pp.403–420, Elsevier Science Publishers, Amsterdam, New York, in press.
[2] Travis, J. and Salvesen, G.S. (1983) Annu. Rev. Biochem. 52, 655–709.
[3] Boswell, D.R. and Carrell, R.W. (1986) TIBS 11, 102–103.
[4] Upton, C. and McFadden, G. (1986) Mol. Cell. Biol. 6, 265–276.
[5] Esser, K., Kuck, U., Lang-Hinrichs, C., Lemke, P., Osiewacz, H.D., Stahl, U. and Tudzynski, P. (1986) Plasmids of Eukaryotes, Fundamentals and Applications, Springer.
[6] Yamagishi, H. (1986) Bioessays 4, 218–221.
[7] Smith, C.A. and Vinograd, J. (1972) J. Mol. Biol. 69, 163–178.
[8] Shope, R.E. (1932) J. Exp. Med. 56, 593–802.
[9] Wills, A., Delange, A.M., Gregson, C., Macaulay, C. and McFadden, G. (1983) Virology 130, 403–414.
[10] Delange, A.M., Macaulay, C., Block, W., Meuller, T. and McFadden, G. (1984) J. Virol. 50, 408–416.
[11] Moss, B. (1985) in: Virology (Fields, B.N. eds.) pp.685–703, Raven, New York.
[12] Pogo, B.G.T. and Dales, S. (1981) Biology of Poxviruses, Springer, Berlin.
[13] Heilig, R., Muraskowsky, R., Kloepfer, C. and Mandel, J.L. (1982) Nucleic Acids Res. 10, 4363–4382.
[14] Leicht, M., Long, G.L., Chandra, T., Kurachi, K., Kidd, V.J., Mace, M., jr, Davie, E.W. and Woo, S.L.C. (1982) Nature 297, 655–659.
[15] Tanaka, T., Ohkubo, H. and Nakanishi, N. (1984) J. Biol. Chem. 259, 8063–8065.
[16] Jagd, S., Vibe-Pedersen, K. and Magnusson, S. (1985) FEBS Lett. 193, 213–216.
[17] Prochownik, E.V., Clark Bock, S. and Orkin, S.H. (1985) J. Biol. Chem. 260, 9608–9612.
[18] Carrell, R.W., Jeppsson, J., Laurell, C., Brennan, S.O., Owen, M.C., Vaughan and Boswell, D.R. (1982) Nature 298, 329–334.
[19] Hill, R.E., Shaw, R.K., Boyd, P.A., Baumann, H. and Hastie, N.D. (1984) Nature 311, 175–177.

[20] Chandra, T., Stackhouse, R., Kidd, V.J., Robson, K.J.H. and Woo, S.L.C. (1983) Biochemistry 22, 5055–5061.

[21] Chandra, T., Stackhouse, R., Kidd, V.J. and Woo, S.L.C. (1983) Proc. Natl. Acad. Sci. USA 80, 1845–1848.

[22] 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.

[23] Hejgaard, J., Rasmussen, S.K., Brandt, A. and Svendsen, I. (1985) FEBS Lett. 180, 89–94.

[24] Griffith, M.J., Noyes, C.M. and Church, F.C. (1985) J. Biol. Chem. 260, 2218–2225.

[25] Salvesen, G.S., Catanese, J.J., Kress, L.F. and Travis, J. (1985) J. Biol. Chem. 260, 2432–2436.

[26] Kurachi, K., Chandra, T., Friezner Degan, S.J., White, T.T., Marchioro, T.L., Woo, S.L.C. and Davie, E.W. (1981) Proc. Natl. Acad. Sci. USA 78, 6826–6830.

[27] Wiman, B. (1986) in: Proteinase Inhibitors (Barrett, A. and Salvesen, G. eds) pp.477–488, Elsevier Science Publishers, Amsterdam, New York, in press.