Journal of Computer-Aided Molecular Design, 8 (1994) 175–191
ESCOM

J-CAMD 227

Common prevalence of alanine and glycine in mobile reactive centre loops of serpins and viral fusion peptides: Do prions possess a fusion peptide?

Isabelle Callebaut[a,b,c], Anne Tasso[a], Robert Brasseur[d], Arsène Burny[b], Daniel Portetelle[c] and Jean Paul Mornon[a,*]

[a]*Département des Macromolécules Biologiques, Laboratoire de Minéralogie-Cristallographie, CNRS URA09, Universités P6 et P7, Tour 16, Case 115, 4 Place Jussieu, F-75252 Paris Cedex 05, France*
[b]*UER de Biologie moléculaire et Physiologie Animale and* [c]*UER de Microbiologie, Faculté des Sciences Agronomiques, 2 Passage des Déportés, B-5030 Gembloux, Belgium*
[d]*Laboratoire de Chimie Physique des Macromolécules aux Interfaces, Université Libre de Bruxelles, 206/2, Blvd. du Triomphe, B-1050 Brussels, Belgium*

Received 18 January 1993
Accepted 22 July 1993

Key words: Serpin; Fusion; Envelope; Prion; Hydrophobic cluster analysis; Amino acid; Mobility

SUMMARY

Serpin reactive centre loops and fusion peptides released by proteolytic cleavage are particularly mobile. Their amino acid compositions reveal a common and unusual abundance of alanine, accompanied by high levels of glycine. These two small residues, which are not simultaneously abundant in stable helices (standard or transmembrane), probably play an important role in mobility. Threonine and valine (also relatively small amino acids) are also abundant in these two kinds of peptides. Moreover, the known 3D structures of an uncleaved serpin reactive centre and a fusion peptide are strikingly similar. Such sequences possess many small residues and are found in several signal peptides and in PrP, a protein associated with spongiform encephalopathies and resembling virus envelope proteins. These properties may be related to the infection mechanisms of these diseases.

INTRODUCTION

Serpins (SERine Protease INhibitors) are a widely distributed family of proteins that includes serine protease inhibitors of human plasma as well as noninhibitory homologues such as hormone-binding globulins, angiotensinogen and egg-white ovalbumin [for reviews, see Refs. 1–3]. The serpin reactive centre, which acts as a pseudosubstrate for the target protease, is remarkable in terms of structural (secondary structure) and spatial mobilities. Indeed, in most serpins, the

*To whom correspondence should be addressed.

0920-654X/$ 10.00 © 1994 ESCOM Science Publishers B.V.

cleavage occurring after interaction with the target protease is accompanied by an exceptional conformational change: the new C-terminus moves as much as 60 Å from the new N-terminus to the opposite pole of the molecule [4,5]. The crystallographic structure of uncleaved ovalbumin [6] (Fig. 1A) shows that the reactive centre, before cleavage, protrudes from the main body of the molecule in an unexpected labile α-helical structure. In the cleaved form of inhibitory serpins [7] (Fig. 1B), the cleaved loop is incorporated into the core of the A sheet to form the fourth strand. Partial inclusion of the reactive centre within the A sheet appears to modulate biological activities. The structural diversity and spatial mobility of these peptides [8–10] is further exemplified by the inactive and uncleaved latent form of plasminogen activator inhibitor-1 [11].

We compared this particular behaviour to that of viral envelope glycoproteins. These proteins, once cleaved, also possess new terminal segments called 'fusion peptides' which are remarkable for their structural and spatial mobility [see Ref. 12 for review]. It has been hypothesized that fusion peptides act as 'sided insertional helices' which locally destabilize membranes [13–16]. However, circular dichroism and infrared spectroscopy reveal α, but also β structures in a lipid environment [17,18]. The only known 3D structure of a fusion peptide is that of the influenza A haemagglutinin (HA-19) which, outside the membrane in a water environment, exhibits both extended structure and reverse turns in a pseudohelical conformation. These data suggest secondary structure versatility for these segments. Spatial mobility is also required, since (i) the HA2 fusion peptide position is 21 Å from the new C-terminus of HA, indicating a major reorganization after activation by cleavage; and (ii) this position is also clearly far away from the area interacting with the cellular receptor, or from the viral membrane (at least dozens of angstroms – Fig. 1C).

Analysis of the mean amino acid composition of these two peptide families (reactive centre loops and fusion peptides) shows that the percentages alanine and glycine are significantly higher than in any other peptide in the protein sequence databanks. Valine and threonine also appear to be abundant in these peptides. This particular composition may account for the mobile character of such segments.

MATERIALS AND METHODS

Sequences of fusion peptides and serpin reactive centre loops

Two sequence sets (Table 1) were extracted from the MIPS (Martiensrieder Institut für Protein Sequenzen – March 1992 release) protein sequence databank according to literature data [see Refs. 1 and 13].

We chose only *distinct* enveloped viruses, known to mediate membrane fusion processes, and avoided inclusion in the data set of any variants of a given virus. In this way, 28 envelope glycoproteins of paramyxoviruses, influenza virus (orthomyxovirus) and retroviruses (lentiviruses and oncoviruses) were selected. From these 28 sequences, we defined a fusion peptide as a sequence including only hydrophobic or uncharged amino acids (V,I,L,F,M,Y,W,C,A,G,S,T) after a cleavage signal (stretch of five amino acids, rich in positively charged residues). Charged amino acids in fusion peptides were allowed only if there were no such residues in the following five amino acids. We excluded from our study fusion peptides from alpha-viruses such as Semliki Forest virus [20] and from PH30 involved in sperm-egg fusion [21]; these peptides are not liberated by proteolytic cleavage and do not constitute new terminal segments.

We also selected *distinct* members of the serpin family. From 20 independent sequences of

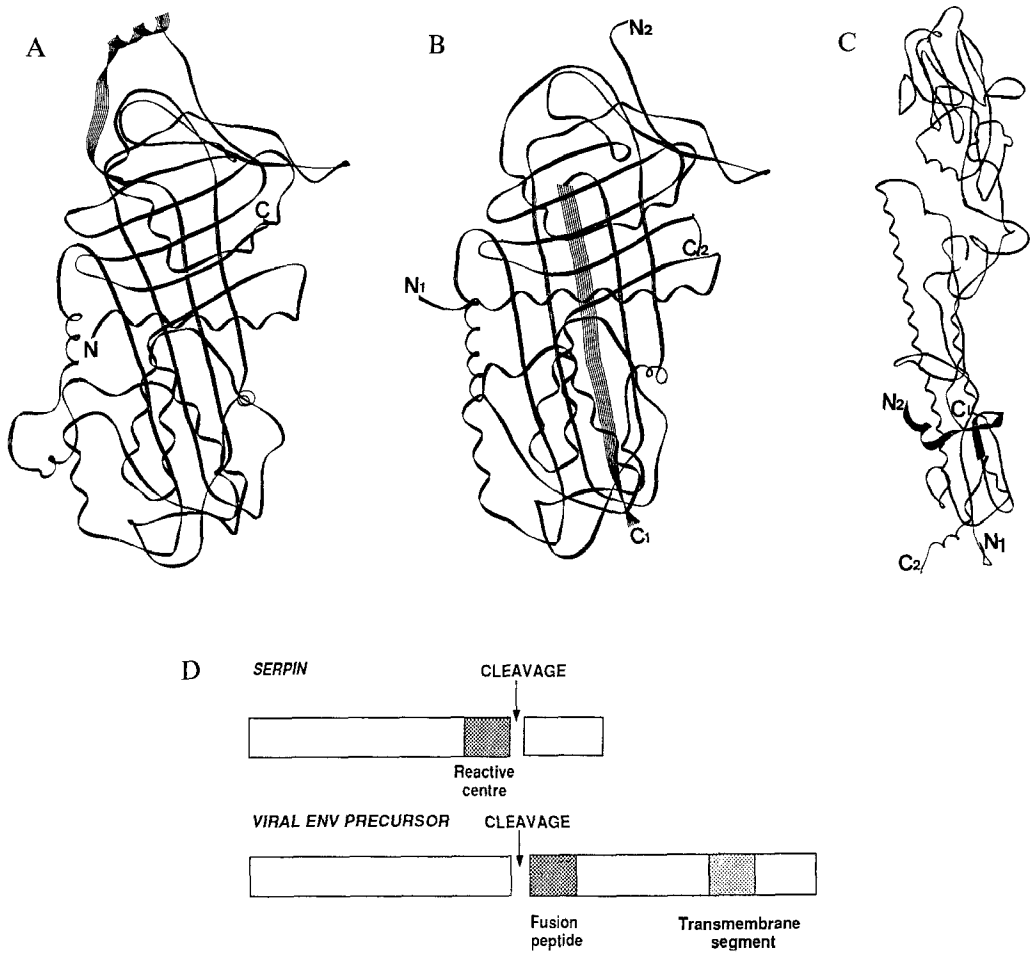

Fig. 1. (A) 3D crystal structure of ovalbumin (PDB: 1OVA), model for uncleaved serpins. (B) 3D crystal structure of α1 antitrypsin (PDB: 9API), model for cleaved serpins. The reactive centre loop, after cleavage corresponding to the new C-terminus and defined as described in the Materials and Methods section, is highlighted as a wide ribbon. The rest of the molecule is shown as a narrow ribbon. Note the insertion, after cleavage, of the reactive centre loop within the middle of the main A β-sheet as a supplementary strand (ovalbumin: four long strands, α1 antitrypsin: five long strands). (C) 3D crystal structure of influenza A haemagglutinin (PDB: 2HMG). The fusion peptide, after cleavage corresponding to the new N-terminus and defined as described in the Materials and Methods section, is highlighted as a wide ribbon. The rest of the molecule (after cleavage, HA1 and HA2) is shown as a narrow ribbon. (D) Schematic representation of the serpin/viral envelope precursor sequences and localization of their mobile (reactive centre and fusion peptide) and transmembrane segments.

inhibitory serpins, we considered segments of fixed length (15 amino acids) corresponding to the reactive centre loops, beginning at a conserved glycine (P15) which is located in the α1 antitrypsin structure in a loop essential for the mobility of the fourth strand of the A sheet. The last residue of this segment corresponds, in the α1 antitrypsin sequence, to the P1 residue (elastase cleavage site) [10].

Some of the fusion peptides and reactive centre loops show relatively good sequence conservation compared to the rest of the protein from which they were extracted. For example, fusion

TABLE 1A
INDIVIDUAL AMINO ACID COMPOSITION OF SERPIN REACTIVE CENTRE LOOPS[a]

Serpin	Amino acid																			
	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Σ A+G
HA1AT	4	0	0	2	1	2	0	1	0	1	2	0	1	0	0	0	1	0	0	0	6
PPAI	3	0	0	1	0	5	0	0	0	0	1	0	0	0	1	0	3	1	0	0	8
MAP1	2	0	1	0	0	1	0	2	0	1	0	0	1	0	1	2	4	0	0	0	3
SILK	6	0	0	1	1	1	0	0	0	0	1	1	0	0	1	1	2	0	0	0	7
HTBG	3	0	1	3	0	1	0	0	0	1	0	0	1	1	0	1	1	2	0	0	4
HCBG	1	0	1	0	0	3	0	0	0	2	0	1	0	0	0	1	4	2	0	0	4
A1AC	4	0	0	1	0	1	0	1	0	2	0	0	0	0	1	1	3	1	0	0	5
AP	5	0	0	1	0	1	0	1	0	0	2	0	0	0	1	2	1	1	0	0	6
AT	5	0	0	1	0	2	0	1	0	0	0	0	0	0	1	2	1	2	0	0	7
CC1I	6	0	0	1	0	1	0	1	0	0	0	0	0	0	1	2	1	2	0	0	7
EPAI	3	0	0	0	0	1	0	1	0	0	0	0	0	0	1	3	3	3	0	0	4
HCF2	1	0	0	0	1	2	0	0	0	1	1	0	1	1	0	0	5	2	0	0	3
PCI	4	0	0	0	2	2	0	1	0	0	0	0	0	0	2	0	4	0	0	0	6
NEXIN	5	0	0	0	0	1	0	2	1	1	0	0	0	0	1	1	3	0	0	0	6
GENE Y	2	0	0	1	0	4	0	2	1	0	0	1	0	0	0	1	3	0	0	0	6
GHRPI	4	0	0	1	0	2	0	1	0	0	0	0	0	0	1	0	4	2	0	0	6
COWPOX	6	2	1	1	0	0	0	0	0	1	0	0	0	0	0	0	2	1	0	1	6
BARLEY	5	0	0	1	0	3	0	0	0	0	2	0	0	0	0	0	2	2	0	0	8
MA1AT	4	0	0	1	0	1	0	0	0	2	0	0	1	0	0	0	2	3	0	1	5
CONTRA	4	0	0	1	0	4	0	2	0	0	0	0	0	0	1	0	2	1	0	0	8

[a] Lengths of serpin reactive centre loops are constant (15 amino acids) and are defined as described in the Materials and Methods section. Sequence accession numbers: HA1AT: human α1-antitrypsin (ITHU); PPAI: placental plasminogen activator inhibitor type II – human (A26553); MAP1: MAP1 protein – myxoma virus (B36418); SILK: antitrypsin-silkworm (JX0130); HTBG: human thyroxin-binding globulin (A26479); HCBG: human corticosteroid-binding globulin (A28321); A1AC: α1 antichymotrypsin – human (ITHUC); AP: α2-antiplasmin – human; AT: antithrombin-III – human (XHHU3); CC1I: complement C1 inhibitor – human (ITHUC1); EPAI: endothelial plasminogen activator inhibitor – human (ITHUP1); HCF2: heparin cofactor II – human (A28704); PCI: protein C inhibitor (A26139); NEXIN: glia-derived neurite promoting factor (nexin) – human (A26061); GENE Y: gene Y protein – chicken (DYCH); GHRPI: growth hormone-regulated proteinase inhibitor 1 – rat (A29131); COWPOX: antithrombin-III homolog 3-vaccinia virus (WMVZHI); BARLEY: serpin gene – barley (S13822); MA1AT: mouse α1-antitrypsin (A25495); CONTRA: contrapsin – mouse (B25420). Additionally, outside this set, four distinct noninhibitor serpins have been considered: OVA: ovalbumin – chicken (CHKOVA); HSP47: heat shock protein 47 – chicken (S16673); COLLAGEN: collagen-binding glycoprotein – rat (A40968); ANGIO: angiotensinogen – rat (ANRT).

peptides of MPMV (Mason Pfizer Monkey Virus) and AREV (Avian Reticuloendotheliosis Virus) exhibit 64% sequence identity, whereas the intact glycoprotein precursors from which they were extracted share only 43% sequence identity. The same observation holds for serpin reactive centre loops; this suggests a functional and/or structural role for these segments.

Amino acid composition – comparison with different distributions

Each residue type was enumerated for each sequence and for the two complete data sets of the two peptides. For each data set, the percentages (%fusion and %serpin) of each amino acid were calculated. These percentages were compared in different ways to amino acid distributions in 'random' peptides (%random), similar in length to fusion peptides (average length of 27 residues) and to reactive centre loops (15 residues). These distributions contain 10 362 peptides which were

TABLE 1B
INDIVIDUAL AMINO ACID COMPOSITION OF FUSION PEPTIDES[a]

Virus	Length	Amino acid																			
		A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y	Σ A+G
BIV	27	6	0	0	0	1	1	0	3	0	6	1	0	0	0	0	2	3	4	0	0	7
VISNA	24	7	0	0	0	0	5	0	5	0	3	1	0	0	0	0	0	0	3	0	0	12
BLV	27	5	0	0	0	0	3	0	1	0	6	0	1	1	0	0	4	2	4	0	0	8
FIV	29	8	0	0	0	1	4	1	3	0	2	2	0	0	0	0	1	4	3	0	0	12
HIV2	27	6	0	0	0	3	5	0	0	0	5	1	0	0	0	0	3	2	2	0	0	11
SIVcpz	29	8	0	0	0	3	6	0	0	0	5	1	0	0	0	0	1	3	2	0	0	14
SIVagm	23	5	0	0	0	4	5	0	0	0	3	1	0	0	0	0	2	1	2	0	0	10
HIV1	28	5	0	0	0	2	6	0	1	0	4	2	0	0	0	1	2	3	2	0	0	11
INFA	24	2	0	1	1	2	8	0	3	0	1	1	0	0	0	0	0	1	0	2	2	10
HTLV1	31	7	0	0	0	0	6	0	1	0	3	2	0	1	0	0	4	1	5	1	0	13
HTLV2	31	7	0	0	0	0	5	0	2	0	4	0	0	2	0	0	4	2	4	1	0	12
MuLV	33	5	0	0	1	0	7	0	3	0	6	2	0	1	0	0	1	6	1	0	0	12
MCFFV	33	5	0	0	1	0	7	0	1	0	7	1	0	1	0	0	1	6	3	0	0	12
FeLV	27	3	0	0	1	0	7	0	2	0	4	1	0	1	0	0	1	4	3	0	0	10
MMTV	27	7	0	0	0	2	1	0	6	0	3	0	0	0	0	0	2	3	3	0	0	8
SMV	28	5	0	0	0	0	5	0	1	0	5	0	0	1	0	0	1	4	6	0	0	10
MPMV	29	4	0	0	0	1	5	0	4	0	4	0	0	1	1	0	2	4	3	0	0	9
AREV	28	3	0	0	0	1	7	0	2	0	5	0	0	1	1	0	1	3	4	0	0	10
SRV	29	3	0	0	0	1	5	0	4	0	4	0	0	1	1	0	2	5	3	0	0	8
SENDAI	26	7	0	0	0	2	4	0	4	0	2	0	0	0	1	0	1	3	2	0	0	11
NDV	26	9	0	0	0	1	4	0	5	0	2	0	0	0	1	0	0	2	2	0	0	13
MEASLE	25	9	0	0	0	1	4	0	2	0	3	0	0	0	1	0	0	2	3	0	0	13
PINF2	26	9	0	0	0	1	3	0	2	0	2	0	0	0	1	0	0	2	6	0	0	12
PINF3	26	6	0	0	0	2	4	0	3	0	2	0	0	0	1	0	1	3	4	0	0	10
MUMPS	26	9	0	0	0	1	3	0	3	0	2	0	0	0	1	0	1	2	4	0	0	12
RSV	19	3	0	0	0	2	4	0	1	0	3	0	0	0	0	0	3	0	3	0	0	7
EIAV	29	10	0	1	0	1	1	0	4	0	1	1	0	0	0	0	3	3	3	0	1	11
RREP	26	7	0	0	0	2	2	0	4	0	3	1	0	0	0	0	0	3	4	0	0	9

[a] Lengths of fusion peptides are variable (between 19 and 33 amino acids, average length: 27.25) and are defined as described in the Materials and Methods section. Sequence accession numbers: BIV: bovine immunodeficiency virus (VCLJBT); VISNA: visna lentivirus (VCLJVS); BLV: bovine leukemia virus (VCLJB); FIV: feline immunodeficiency virus (VCLJFP); HIV2: human immunodeficiency virus 2 (VCLJG2); SIVcpz: simian immunodeficiency virus strain cpz (VCLJSI); SIVagm: simian immunodeficiency virus strain agm (VCLJS2); HIV1: human immunodeficiency virus 1 (VCLJLV); INFA: influenza A virus (HMIV); HTLV1: human T lymphotropic virus type 1 (VCVWH); HTLV2: human T lymphotropic virus type 2 (VCLJH2); MuLV: murine leukemia virus (VCVWEM); MCFFV: mink cell focus forming virus (VCVWFS); FeLV: feline leukemia virus (VCMVSA); MMTV: mouse mammary tumor virus (VCMVM); SMV: squirrel monkey virus (VCLJHD); MPMV: mason–pfizer monkey virus (VCLJMP); AREV: avian reticuloendotheliosis virus (VCVDAR); SRV: simian AIDS retrovirus (VCLJSA); SENDAI: sendai virus (VGNZSV); NDV: newcastle disease virus (VGNZVN); MEASLE: measle virus (VGNZMV); PINF2: parainfluenza 2 (A36423); PINF3: parainfluenza 3 (VGNZH3); MUMPS: mumps virus (VGNZMU); RSV: respiratory syncytial virus (VGNZA2); EIAV: equine infectious anemia virus (VCLJEV); RREP: retrovirus related env polyprotein (VCHUER). Rous sarcoma virus (VCFVER) is considered outside this set, because its fusion peptide could not be unambiguously defined.

extracted randomly from a nonredundant version of the MIPS databank (NR-MIPS, deduced from the MIPS databank, release 30), constructed by eliminating any sequence containing more than 10 consecutive amino acids identical with another sequence [22]. Comparisons can be made by computing : (i) *R values* (%fusion (or %serpin) / %random), indicating amino acid preferences

for fusion peptide or serpin reactive centre loop; (ii) mean *Z values* of each individual peptide, calculated as (%fusion (or %serpin) – %random) / σrandom, where σrandom is the standard deviation of the random distribution mean.

Amino acid distributions in segments belonging exclusively to a particular structure (α (484 representants), β (667 representants) or coil (835 representants)) and of defined length can also be deduced from a nonredundant version of the Protein Databank [NR-PDB; Ref. 23]. This databank, compiled from the July 1991 Protein Databank release [Ref. 24; 492 entries], contains 154 fragments or chains from 125 proteins and is similar to others produced during 1991 and 1992 [25–28]. Structures were defined in this databank using a consensus assignment [23]. Such distributions were used as described above to calculate R and Z values.

Observed distributions (those of fusion peptides and of serpin reactive centre loops) can be compared with theoretical distributions found in the entire databank, or in particular structural segments (α, β or coil structures). This is achieved using a χ^2 test, by summing the values $(oi - ci)^2/ci$ computed for each class i (individual amino acid or group of amino acids if the number of representants is less than five), where oi is the number of observed occurrences and ci the number of calculated occurrences, derived from theoretical distributions.

Searching in NR-MIPS for segments harbouring fusion peptide sequence similarities

Segments of amino acid composition similar to fusion peptides were sought, using a FORTRAN program running on a VAX computer, considering the same criteria as those used to define fusion peptides (vide supra).

We also used Hydrophobic Cluster Analysis (HCA) [29–32], a 2D method of protein sequence analysis which is very sensitive in selective detection of similar secondary and tertiary structures, independent of the sequence identity level and structural class. Typical applications are described in the references mentioned above and in Refs. 33–35.

3D studies

3D molecular studies were performed using the polyvalent MANOSK software [36] on an Evans & Sutherland PS390 display.

RESULTS

Amino acid composition of serpin reactive centre loops and fusion peptides

Table 1 lists the individual amino acid compositions of all the sequences belonging to our two data sets (fusion peptides and reactive centre loops of serpins). These peptides were selected according to literature references, carefully excluding any peptide of the same protein belonging to another variant or species (see the Materials and Methods section).

Table 2 lists the overall amino acid composition (%serpin or %fusion) of the two data sets, as well as the preference of individual residues to be in a fusion peptide or in a serpin reactive centre loop (R values). This preference is the ratio of a given amino acid composition in such peptides (%serpin or %fusion) and that of distributions of 'random' peptides (%random). These random peptides have lengths similar to those of the peptides considered (15 residues for serpin and 27 for fusion) and were selected randomly from the NR-MIPS databank described above. It is clear that alanine is a desirable constituent of these peptides, as indicated by preferences greater than

2 (3.30 and 2.89 in reactive centre loops and in fusion peptides, respectively). Glycine and threonine are also considered favourable (1.81 and 2.40 for glycine; 2.98 and 1.77 for threonine in reactive centre loops and in fusion peptides, respectively). Curiously, as illustrated with haemagglutinin influenza A, locally found lower percentages of alanine are apparently compensated by higher percentages of glycine (Table 1).

As shown in Table 2, the alanine composition (%serpin and %fusion) has a value outside the calculated distributions, as shown by the mean (Z_m) of individual Z scores, corresponding to ((%serpin – %random) / σrandom), where %random is the distribution mean and σ its standard deviation. Glycine and threonine individually do not have significant Z_m values (<1.96). However, it should be stressed that these Z_m values are significant when calculated for the sums alanine + glycine, alanine + threonine and alanine + glycine + threonine (Table 2). In these particular cases, the standard deviations, when expressed as percentages of the distribution means, are lower than for individual counts. In this regard, it should be pointed out that alanine is more frequently encountered in α-helices (the first occurrence), glycine in coil structures (the first occurrence) and threonine in β-strands (the fourth occurrence after V, I and F). The three kinds of residues are

TABLE 2
COMPARISON OF OBSERVED DISTRIBUTIONS WITH THOSE OF RANDOM PEPTIDES[a]

Amino acids	Serpin reactive centre loops					Viral fusion peptides				
	%serpin	%random	σrandom	R	Z_m	%fusion	%random	σrandom	R	Z_m
A	25.67	7.77	8.09	3.30	2.21	22.28	7.72	6.49	2.89	2.24
C	0.67	2.05	3.60	0.33	–0.38	0	2.06	3.60	0	–0.57
D	1.33	5.28	6.14	0.25	–0.64	0.26	5.27	4.77	0.05	–1.05
E	5.67	6.08	6.93	0.93	–0.06	0.52	6.09	5.50	0.09	–1.01
F	1.67	3.94	5.43	0.42	–0.42	4.46	3.95	4.15	1.13	0.12
G	12.67	6.98	7.30	1.81	0.78	16.64	6.92	5.85	2.40	1.66
H	0	2.31	4.27	0	–0.54	0.13	2.24	3.21	0.06	–0.66
I	5.33	5.61	6.38	0.95	–0.04	9.17	5.62	4.92	1.63	0.72
K	0.67	5.96	6.96	0.11	–0.76	0	5.93	5.59	0	–1.06
L	4.00	9.07	7.98	0.44	–0.63	13.11	9.08	6.22	1.44	0.65
M	3.00	2.04	3.81	1.47	0.25	2.36	2.05	2.94	1.15	0.11
N	1.00	4.31	5.58	0.23	–0.59	0.13	4.48	4.53	0.03	–0.96
P	1.67	5.03	6.56	0.33	–0.51	1.44	5.02	5.32	0.29	–0.67
Q	0.67	3.86	5.44	0.17	–0.59	1.18	3.93	4.35	0.30	–0.63
R	4.33	5.69	6.91	0.76	–0.20	0.13	5.65	5.58	0.02	–0.99
S	5.67	7.07	7.27	0.80	–0.19	5.64	7.13	5.83	0.79	–0.26
T	17.0	5.70	6.36	2.98	1.78	10.09	5.76	5.00	1.75	0.87
V	8.33	6.65	6.74	1.25	0.25	11.53	6.51	5.11	1.77	0.98
W	0	1.31	3.08	0	–0.43	0.52	1.33	2.38	0.39	–0.34
Y	0.67	3.26	4.86	0.21	–0.53	0.39	3.25	3.76	0.12	–0.76
A+G	38.34	14.75	10.87	2.63	2.17	38.92	14.64	8.89	2.66	2.73
A+T	42.67	13.47	9.94	3.17	2.94	32.37	13.48	7.96	2.40	2.37
A+G+T	55.34	20.25	12.12	2.71	2.90	49.01	20.40	9.83	2.40	2.91

[a] Means (%serpin and %fusion) are compared with amino acid distributions in random peptides (10 326 sequences) of a nonredundant version of the MIPS databank (NR-MIPS), whose lengths are similar to that of serpin loops (15 amino acids) or to that of fusion peptides (27 amino acids). These 'random' distributions are characterized by means (%random) and by standard deviations of their means (σrandom). R values (%serpin / %random and %fusion / %random) and Z_m values (means of individual Z values, calculated as (%serpin – %random) / σrandom and (%fusion – %random) / σrandom) are indicated. Differences between compared distributions are significant ($P_\alpha = 0.05$) if $Z > 1.96$.

TABLE 3
COMPARISON OF OBSERVED DISTRIBUTIONS WITH DATABANKS: χ^2 VALUES[a]

Amino acid		NR-MIPS		NR-PDB		NR-PDBα		NR-PDBβ		NR-PDBc	
	o	c	χ^2	c	χ^2	c	χ^2	c	χ^2	c	χ^2
Serpin reactive centre loops											
A	77	24	119	26	100	38	40	20	162	22	140
T	51	18	61	19	55	14	104	24	30	20	51
G	38	19	19	24	8	16	31	15	35	37	<0.1
V	25	18	3	21	0.8	19	2	38	4	15	8
S	17	22	1	21	0.8	15	0.3	19	0.2	25	2
E	17	21	0.8	17	0	24	2	11	3	9	7
I	16	18	0.2	16	0	17	0.1	25	3	10	3
R	13	18	1	12	0.1	15	0.3	11	0.4	11	0.4
L	12	26	4	25	6	33	13	29	10	18	2
M	9	7	0.6	6	2	8	0.1	7	0.6	4	6
F	5	11	3	12	4	13	5	16	8	9	2
P	5	15	7	15	6	6	0.2	13	5	20	12
Q+N	5	25	16	24	15	24	15	18	9	28	19
C+D+H+K+W+Y	10	51	33	63	45	61	43	56	38	68	49
Σ			269		243		256		308		301
Fusion peptides											
A	170	61	198	66	163	97	55	51	278	55	236
G	127	48	129	62	68	30	307	40	189	93	12
V	88	45	41	54	21	50	30	96	0.7	38	65
T	77	46	22	48	18	34	53	61	4	51	14
I	70	46	13	40	22	44	15	63	0.8	27	68
L	100	67	16	63	22	84	3	73	10	42	80
M	18	19	0.1	15	0.6	21	0.4	17	0.1	9	9
F	34	28	1	30	0.5	34	0	40	0.9	23	5
S	43	55	3	52	2	38	1	48	0.5	73	12
P	11	39	20	37	18	15	1	34	16	47	28
Q+N	10	65	47	54	36	62	44	45	27	70	51
D+E+H+K+R+C	8	211	195	195	179	218	202	150	134	204	188
W+Y	7	33	20	38	25	37	24	49	36	34	21
Σ			705		575		738		697		789

[a] χ^2 values (see the Materials and Methods section for calculations), where observed distributions are those of serpin reactive centre loops and fusion peptides and 'theoretical' distributions are those encountered in several databanks, i.e., NR-MIPS, NR-PDB α-structures, NR-PDB β-structures and NR-PDB coil structures (NR = nonredundant). All members of the χ^2 sums are indicated in order to obtain information about values exceeding significance cutoff. Differences between compared distributions are significant ($P_\alpha = 0.05$) if $\chi^2 > 22.4$ (serpin reactive centre loops: 13 degrees of freedom) or > 21.0 (fusion peptides: 12 degrees of freedom).

thus not simultaneously abundant in peptides of defined length; their sum along random peptides covering different kinds of structures is therefore more constant than their individual values. The simultaneous abundance of alanine, glycine and threonine encountered in fusion peptides and reactive centre loops is thus worth noting, even if individual counts of glycine (1.81) and threonine (1.77) are just below the significant value, i.e. 1.96, for a probability of 5%.

χ^2 values (Table 3) further assess the significant differences in observed mean alanine composition between our peptide sets and databanks (NR-MIPS, NR-PDB, NR-PDBα, NR-PDBβ, NR-PDBcoil). It is interesting to note that alanine χ^2 values are significant in each case considered,

even when calculated in relation to the amino acid composition of α-helices where alanine is the most abundant residue (12.69%). In contrast, χ^2 values are not significant for glycine and threonine in structures where these residues are prevalent (coil for glycine and β-strands for threonine).

An observed difference between fusion peptides and reactive centre loops is the higher percentages of hydrophobic residues in fusion peptides, as these peptides interact with the membrane through the fusion process. Transmembrane segments are also rich in such hydrophobic residues. However, alanine and glycine %fusion values are clearly outside the means of the amino acid distribution of clear transmembrane segments (373 representants) found in the complete NR-MIPS databank (Z_m values for alanine and glycine are 3.23 and 2.54, respectively). This observation supports the assumption that alanine, glycine and threonine might play a particular role in fusion peptides. Interestingly enough, it appears from Table 2 that valine is the preferred hydrophobic residue, in fusion peptides as well as in serpin reactive centre loops, whereas longer hydrophobic residues, such as leucine and isoleucine, are preferentially encountered in transmembrane segments.

Serpin reactive centre loops and fusion peptides can be divided into two distinct segments

Independently, likely alignments of the two peptide sets (Fig. 2) reveal that small residues (A,G,T) are more abundant in the inner half of the peptides (the N-terminal half of the serpin reactive centres, including the 'loop hinge' [11] (73.8%) and the C-terminal half of viral fusion peptides (58.6%)), whereas large hydrophobic residues (V,I,L,F,M,Y,W) are more abundant in their outer half (43.6% and 52.1% for reactive centre loops and fusion peptides, respectively). Known 3D structures of a serpin (ovalbumin) and a fusion peptide (influenza A haemagglutinin) show that their outer halves, containing more hydrophobic residues, are associated with α-helical or pseudo-α-helical structures, whereas their inner halves correspond to extended coil structures. In order to compare the amino acid compositions of these observed structures to those in the NR-PDB databank, we computed several distributions of the amino acid compositions of peptides of lengths similar to those encountered in the two halves of peptides (between 5 and 15 amino acids in order to get sufficient representants) and belonging exclusively to protein loops and turns (coil), α-helices or β-strands. Alanine, alanine + glycine and alanine + threonine compositions of fusion peptides and of reactive centre loops (%fusion and %serpin) are located significantly outside the calculated distributions, as shown by the mean of the individual Z scores (Z_m). The particularly high Z_m score of alanine vs. coil structure for the inner peptide half (associated with coil structures in the 3D structure of ovalbumin and haemagglutinin – Table 4) is noteworthy.

Comparison with other terminal segments liberated after cleavage

It is worth noting that, unlike the preceding new C-terminus (reactive centre – Fig. 1D), the new N-termini of serpins, generated after proteolytic cleavage at the P1-P1' bond, do not exhibit mobility. Their composition (also over 15 residues) does not show any prevalence of alanine and/or glycine, as is also the case for the new C-terminus of viral envelope glycoproteins (the C-terminus of the extramembrane glycoprotein, having the same length as fusion peptides, see Table 1), that is therefore probably not mobile.

Comparison with the Kunitz serine protease inhibitors

The Kunitz serine protease inhibitor family, whose 3D structures are known, do not exhibit

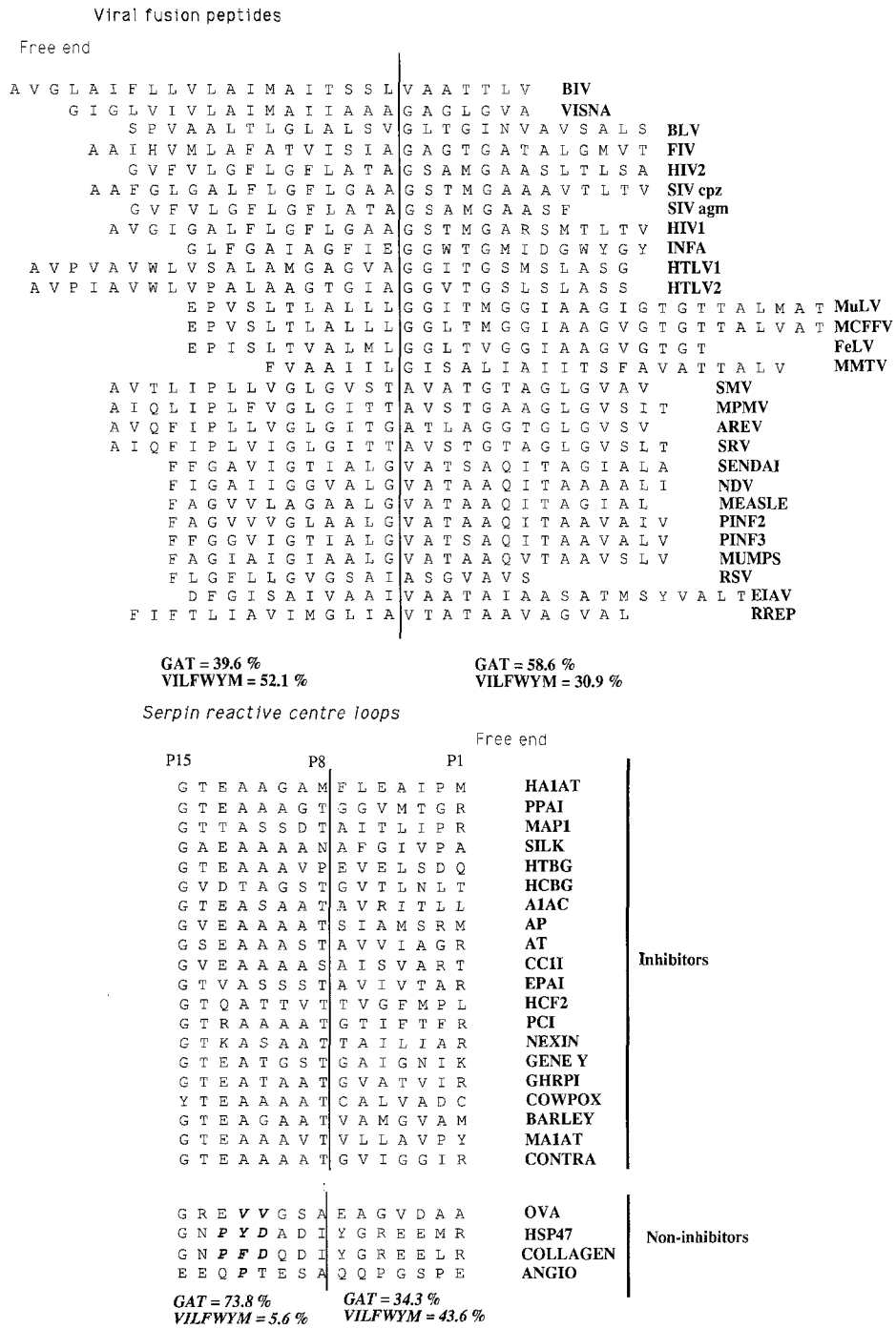

Fig. 2. Sequence alignments of serpin reactive centre loops and tentative alignments for virus fusion peptides. Separations indicate regions where small amino acids such as alanine, glycine and threonine are abundant. The noninhibitory serpins, which do not undergo large conformational changes even after cleavage, possess between P8 and P15 ('hinge loop') large amino acids (V,Y,F) and/or proline, which are thought to be responsible for their lack of mobility [Ref. 9 – OVA, HSP47, COLLAGEN, ANGIO].

particular mobility in any segments liberated by proteolytic cleavage. The amino acid compositions of these segments do not show any particular prevalence of glycine or alanine.

Screening of the whole databank for segments with fusion peptide characteristics

We have written a program allowing the screening of the MIPS protein sequence databank for fusion peptides and analogous segments (criteria identical to those used for the definition of fusion peptides – see the Materials and Methods section). Apart from viral envelope glycoproteins that include fusion peptides, few proteins contain such segments, which often correspond to signal peptides. Interestingly, we also singled out the prion protein (PrP) thought to represent the transmissible agent for spongiform encephalopathies such as scrapie in animals and Creutzfeldt–Jakob disease in man [37,38]. Indeed, among 137 candidates, it exhibits the highest A+G proportion (58%), ahead of a signal peptide (54%), the first fusion peptide (50%) being that of Visna lentivirus.

HCA comparison (Fig. 3) of entire viral transmembrane proteins (and particularly that of human T-lymphotropic virus 1 (HTLV-1)) with whole PrP protein revealed a similar folding pattern, not found in any of the other proteins selected by our screening program. Indeed, HCA analysis clearly shows similarities in: (i) the length of the segments included between the putative fusion peptides and the C-terminal transmembrane segments; (ii) the secondary structures of these segments, which are principally helicoidal; (iii) the presence of a basic stretch of sequences preceding the putative fusion peptides; and (iv) the general features of the putative fusion peptides. Moreover, all amino acids present in the PrP putative fusion peptide (Fig. 4) are among the eight most frequent amino acids of the consensus for fusion peptides (Table 1B: A,G,V,T,L,F,M,S), with similar, high frequencies for A and G.

DISCUSSION

This study highlights the particular role that small amino acid residues (especially alanine, and other small residues such as glycine and threonine) may play in conferring mobility on peptides generated after proteolytic cleavage.

Although our results resemble those reported by Argos on linker peptides [39], there are several discrepancies which show that fusion peptides or serpin reactive centre loops cannot be likened to 'linker' sequences. Indeed, alanine is not as abundant in linker peptides (R = 1.05, whereas R values for serpins reactive centre loops and fusion peptides are 3.30 and 2.89, respectively, the three values being calculated for comparable protein compositions). Similarly, threonine and glycine R values are also much higher in fusion peptides and reactive centre loops (glycine: 1.81 and 2.40 vs. 1.25 in linker peptides; threonine: 2.98 and 1.77 vs. 1.55 in linker peptides). Threonine and serine are both prevalent in linker peptides, whereas only threonine is noticeable in our data set. No standard deviations are given for the amino acid distributions in the linker data set, and the statistical significance of these differences therefore cannot be dicussed further.

Alanine percentages are much higher in fusion peptides and serpin reactive centre loops than in any other segment of similar length, even in α-helices where alanine is the most abundant residue. This would suggest that the inner half of these peptides (at the opposite side of the cleavage site and where alanine residues are concentrated) has a strong disposition for α-helical structures. However, what is unusual in α-helices is the pronounced presence of glycine, probably

TABLE 4
COMPARISON OF OBSERVED DISTRIBUTIONS WITH DISTRIBUTIONS OF PEPTIDES BELONGING TO PARTICULAR STRUCTURES (α, β AND COIL)[a]

Amino acid	Serpin reactive centre loops								Fusion peptides							
	Inner half (P15-P8) → coil				Outer half (P7-P1) → α-helix				Outer half → pseudo helical				Inner half → coil			
	%	Z_m α	Z_m β	Z_m coil	%	Z_m α	Z_m β	Z_m coil	%	Z_m α	Z_m β	Z_m coil	%	Z_m α	Z_m β	Z_m coil
A	35.63	1.94	3.03	2.71	15	0.23	0.89	0.73	19.27	0.59	1.33	1.14	25.33	1.09	1.96	1.72
C	0	−0.33	−0.39	−0.33	1.43	0.01	−0.14	−0.03	0	−0.33	−0.39	−0.33	0	−0.33	−0.39	−0.33
D	1.25	−0.59	−0.26	−0.67	1.43	−0.57	−0.23	−0.65	0.26	−0.73	−0.42	−0.77	0.26	−0.71	−0.42	−0.77
E	8.75	0.00	0.75	0.42	2.14	−0.67	−0.16	−0.37	1.04	−0.78	−0.31	−0.50	0	−0.89	−0.45	−0.62
F	0	−0.06	−0.61	−0.48	3.57	−0.11	−0.20	0.09	8.33	0.55	0.35	0.85	0.54	−0.54	−0.55	−0.39
G	15	1.74	1.21	0.19	10.71	1.08	0.69	−0.16	15.36	1.80	1.26	0.22	17.94	2.19	1.57	0.42
H	0	−0.41	−0.33	−0.40	0	−0.41	−0.33	−0.40	0.26	−0.35	−0.29	−0.35	0	−0.41	−0.33	−0.40
I	0	−0.69	−0.75	−0.48	12.14	0.91	0.38	1.49	12.24	0.93	0.39	1.51	6.07	0.11	−0.18	0.50
K	0.63	−0.67	−0.52	−0.59	0.71	−0.66	−0.50	−0.58	0	−0.74	−0.58	−0.66	0	−0.74	−0.58	−0.66
L	0	−1.10	−0.88	−0.67	8.57	−0.17	−0.02	0.51	17.97	0.64	0.73	1.54	8.18	−0.31	−0.15	0.34
M	0.63	−0.35	−0.28	−0.15	5	0.45	0.44	0.81	1.56	−0.18	−0.13	0.05	3.17	0.12	0.14	0.41
N	0.63	−0.51	−0.34	−0.59	1.43	−0.38	−0.21	−0.50	0	−0.61	−0.44	−0.66	0.26	−0.57	−0.40	−0.63
P	0.63	−0.30	−0.48	−0.68	3.57	0.33	−0.11	−0.36	2.86	0.18	−0.20	−0.45	0	−0.44	−0.56	−0.75
Q	0.63	−0.59	−0.35	−0.40	0.71	−0.58	−0.34	−0.39	0.78	−0.57	−0.33	−0.38	1.58	−0.45	−0.21	−0.25
R	0.63	−0.57	−0.44	−0.40	7.86	0.38	0.58	0.65	0	−0.66	−0.53	−0.49	0.26	−0.62	−0.50	−0.45
S	7.5	0.28	0.14	−0.10	2.86	−0.28	−0.35	−0.53	3.13	−0.25	−0.32	−0.50	8.18	0.36	0.22	−0.04
T	23.13	2.96	1.45	1.79	8.57	0.67	0.07	0.21	4.95	0.10	−0.28	−0.19	15.3	1.73	0.71	0.94
V	4.38	−0.25	−0.67	−0.06	13.57	0.89	0.05	1.12	11.46	0.63	−0.12	0.85	11.61	0.65	−0.11	0.87
W	0	−0.36	−0.36	−0.31	0	−0.36	−0.36	−0.31	0.52	−0.25	−0.24	−0.19	0.54	−0.25	−0.24	−0.14
Y	0.63	−0.42	−0.48	−0.36	0.71	−0.41	−0.47	−0.34	0	−0.52	−0.56	−0.45	0.79	−0.40	−0.46	−0.33
A+G	50.63	2.55	3.09	2.95	25.71	0.73	1.13	0.55	34.63	1.49	1.83	1.41	43.27	2.02	2.51	2.24
A+T	58.76	3.22	3.26	4.47	23.57	0.54	0.68	0.95	24.22	0.59	0.73	1.01	40.63	1.83	1.93	2.65
A+G+T	73.76	3.69	3.45	4.25	34.28	0.97	0.96	0.68	39.58	1.34	1.30	1.16	58.57	2.64	2.52	2.88

[a] Z_m values calculated between observed amino acid distributions (serpin reactive centre loops and fusion peptides divided into two parts, according to Fig. 2) and α-helices, β-strands and coil structures, where peptide lengths range between 5 and 15 amino acids.

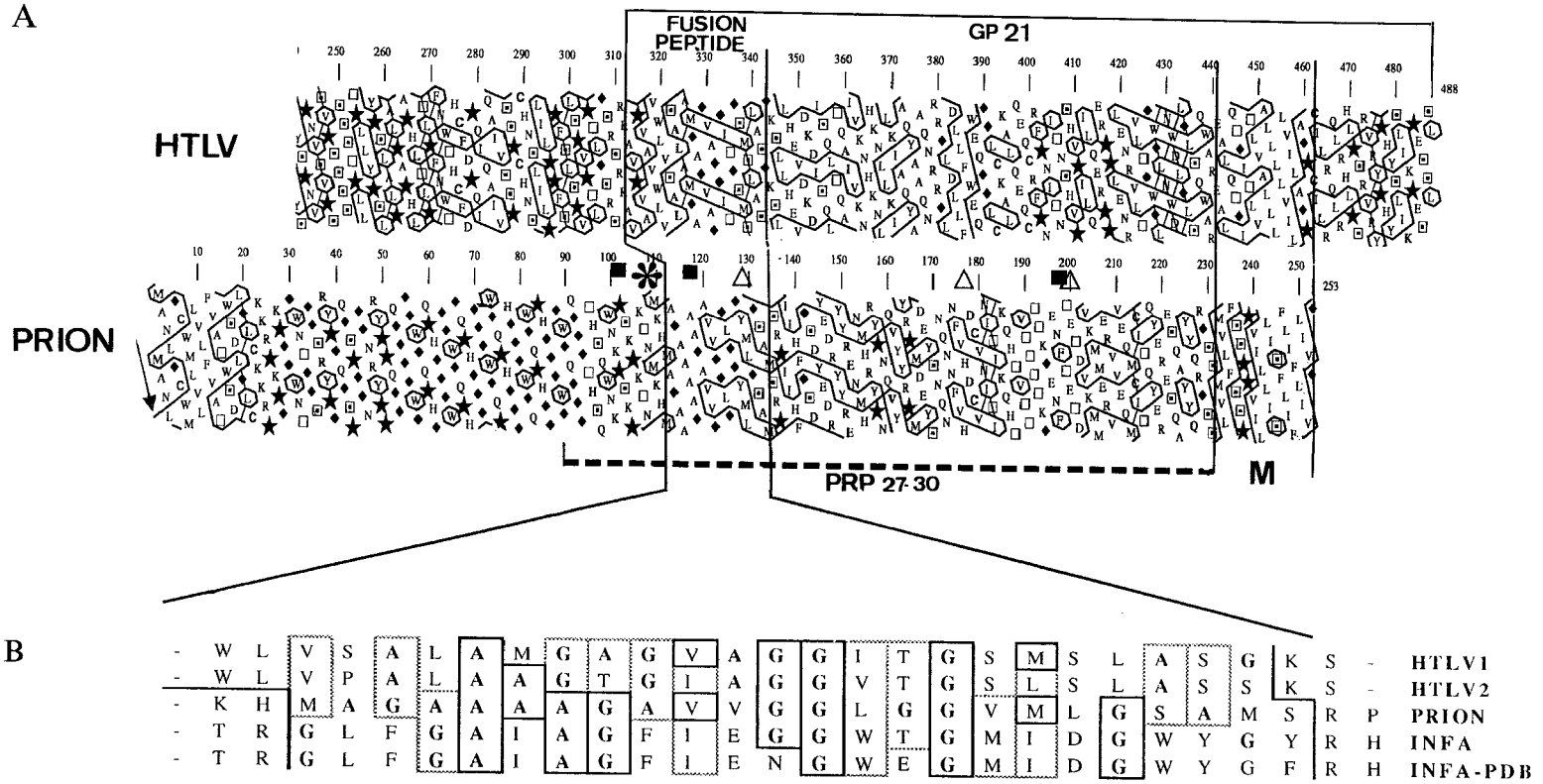

Fig. 3. (A) Putative correspondence in the HCA plot between HTLV1 transmembrane gp21 (starting at Ala^{312}) and human PRION protein. M indicates membrane segment, * highlights basic residues preceding the cleavage point of HTLV1 envelope precursor and the putative fusion peptide of PRION. ■ and Δ indicate crucial mutation points leading to Gerstmann–Sträussler (codons 102, 117, 198) and Creutzfeldt–Jakob (codons 129, 178, 200) diseases, respectively. The linear sequence can be read along the quasi-vertical diagonal (arrow). Four HCA symbols are used for proline (star), glycine (diamond), threonine (open square) and serine (dotted square). (B) Tentative alignment of the putative fusion peptide of PRION with those of HTLV1, HTLV2 and haemagglutinins from two influenza A viruses. INFA PDB is relative to the 3D known haemagglutinin (PDB: 2HMG). Note the identical length for the haemagglutinin fusion peptide and the putative one for prion.

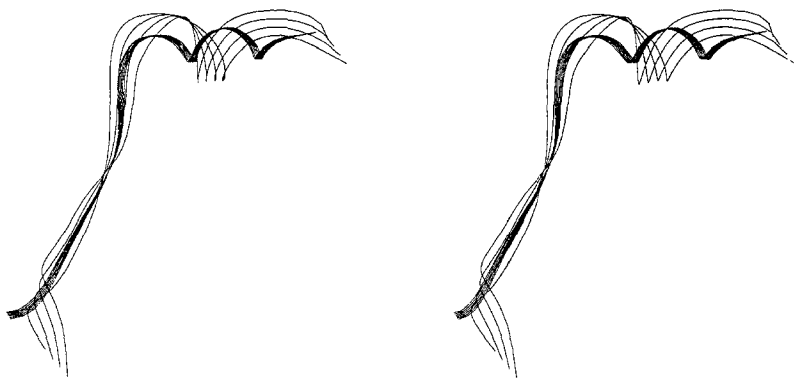

Fig. 4. Superimposition of the reactive centre loop of ovalbumin (sequence GREVVGSAEAGVDAA; narrow ribbon) and fusion peptide of haemagglutinin HA2 (sequence GLFGAIAGFIENGWEGMID; wide ribbon). According to Figs. 1 and 3, the sequence directions are reversed so that the N- and C-termini of the ovalbumin centre loop are superimposed on the C- and N-termini of haemagglutinin fusion peptide.

due to the flexibility of this residue and the absence of a C^{β}, providing an enthalpic interaction with the backbone [40]. Glycine is the first component of coil structures, but alanine is never so abundant in coil structures. Simultaneous prevalence of these two residues in the inner half of fusion peptides and serpin reactive centre loops appears therefore contradictory. Consequently, no clear structural prediction could be made for such segments in this way. We therefore conjectured a particular role for the simultaneous pronounced presence of these two amino acids in mobile peptides.

In recent years, the penetration of fusion peptides into the lipid bilayer was thought to be associated with α-helical structures. Helices supposedly insert obliquely with respect to the lipid/water interface, thereby allowing membrane destabilization [13]. This hypothesis is contradicted by Epand et al. [18], who reported β-structures for measles fusion peptides immersed in lipid media. This type of secondary structure seems plausible in such an environment, as exemplified by porins forming 16-stranded antiparallel β-barrels [41,42]. However, the stability of such barrels is ensured, since main-chain polar atoms are engaged in interstrand hydrogen bonds and therefore are not exposed to the membrane core, fulfilling the same requirements as an α-helix for stabilization. Stabilization of a hairpin β-strand fusion peptide cannot occur in this way and interaction of main-chain polar atoms with those of the lipid bilayer may confer a destabilizing character on these peptides.

Gallaher et al. [17] concluded from these experimental data that viral fusion peptides could adopt several secondary structures. In light of this remark and our sequence analysis, we hypothesize that the destabilizing character of fusion peptides could be the result of the fact that these peptides adopt multiple secondary structures, switching easily between α, β and coil structures, each of them favoured by the particular abundance of residues encountered more particularly in one structure (alanine for helices, glycine for coil, ...). It should also be pointed out that all of the prevalent amino acids, even the strongly hydrophobic ones (valine), are small and therefore enhance the mobility of these segments. This hypothesis reconciles the models described above (the sided insertional helix and the β-strand of measles fusion peptides), each of them accounting for part of the actual mechanism(s), and highlights the special role of dynamics in the biological functions of proteins.

The 3D structures of one intact serpin reactive centre [ovalbumin, Ref. 7], and of one fusion peptide [influenza haemagglutinin A, Ref. 19] are known. They exhibit a hitherto unrecognized striking similarity (Fig. 4). Both possess an inner extended structure, rich in small residues (see the results presented above), and an α-helical tip region of nearly identical length and orientation if the polypeptide path is reversed and aligned with the free end of each peptide (the C-terminus for serpin reactive centre and the N-terminus for fusion peptide). Small root mean square (rms) deviations between corresponding C main-chain atoms (C=O) after superimposing are apparent for each half of the peptides on both sides of the kink (Fig. 4 – 0.42 Å rms for the first six residues in the extended coil moiety and 0.87 Å rms for the first five residues in the helical part). Such striking 3D similarity, accompanied by a reversal of chain path, has already been noted for active sites of DNA-binding proteins and calcium-binding proteins [43], suggesting that evolution selects efficient 3D motifs, whatever the sequence orientation.

Previous theoretical calculations for fusion peptides [13–16, R. Brasseur, unpublished results] agree with this new experimental observation. According to these calculations (using, among other methods, a stereoalphabet [44], a simplex minimization procedure [45], the hypermatrix method [46] and the molecular hydrophobicity potential [14]), the fusion peptide is divided into two main domains: a helix tilt inserted at the lipid–water interface, followed by a turn and an extended structure with the carboxyl terminus in the hydrophilic phase. These convergent data strongly suggest that fusion peptides adopt this structure and probably coexist dynamically with other regular (β-strands) and irregular structures favoured by small amino acids, especially alanine and glycine.

Otherwise, the partition of fusion peptides in two halves (Fig. 3), the first one possibly being instantaneously more stable than the second one, which is especially rich in small amino acids, recalls the situation leading to nematic crystal liquid phases, molecules possessing a rigid part followed by a very flexible one. Thus, one could ask whether a cooperativity within oligomeric (likely trimers) envelope glycoproteins occurs at the level of fusion peptides when they are activated.

The exceptionally high alanine/glycine content within a hydrophobic environment found in a PrP peptide associated with prion, the agent of spongiform encephalopathies (Fig. 3), leads us to ask whether the infectious character of prions is linked to the presence of a fusion peptide and thereby to infection mechanisms related to the viral mechanisms. This hypothesis is supported by: (i) the presence of a basic amino acid sequence preceeding the putative cleavage site; (ii) the localization of crucial mutations within the putative fusion peptide; (iii) the existence of an unidentified posttranscriptional process leading to an abnormal pathogenic protein. Modified properties of the putative cleavage area (codon 102) and/or changes in behaviour of the putative fusion peptide (codons 112 and 129) could lead to abnormal proteins [37,38,47,48]. Our observations therefore suggest that proteolytic activation and interactions with membranes may play a role in PrP infectivity.

NOTE ADDED IN PROOF

After submission of this manuscript, Forloni et al. [Nature, 362 (1993) 543] published results about the effects of several synthetic peptides from the PrP sequence on neuron viability. They showed that only one PrP peptide (PrP 106–126) significantly reduces cell viability. Its sequence (KTNMKH**MAGAAAAGAVVGGLG**) overlaps the sequence which we have detected by search-

ing for segments with fusion peptide characteristics (bold). Our results may allow further overall comprehension of the pathogenic effects, by testing, for example, the membrane interaction potential of this peptide. Interestingly enough, the tested scrambled sequence of PrP 106–126 (NGAKALMGGHGATKVMVGAA) does not induce neurotoxicity. We could hypothesize that the sequence order specificity is due to the succession of a basic segment, followed by an alanine/glycine-rich region lacking any positively charged amino acids (see Table 2). Disorganization of this succession would result in loss of biological activity. We suggest that the integral putative peptide defined in Fig. 3 could be an interesting target for completion of the above biological study.

ACKNOWLEDGEMENTS

We are indebted to M.T. Semertzidis for help with the statistical analysis of databanks. This work was supported by Universités P6 and P7, CNRS, INSERM (contract no. 910912) and the Fondation pour la Recherche Médicale. I.C. and R.B. are research assistant and senior research associate of the Belgian National Fund for Scientific Research (FNRS), respectively. A.T. is a postdoctoral fellow of the French CM2AO Organibio program.

REFERENCES

1 Huber, R. and Carrell, R.W., Biochemistry, 28 (1989) 8951.
2 Bock, S.C., Protein Eng., 4 (1990) 107.
3 Carrell, R.W., Pemberton, P.A. and Boswell, D.R., Cold Spring Harbor Symp. Quant. Biol., 52 (1987) 527.
4 Banzon, J.A. and Kelly, J.W., Protein Eng., 5 (1992) 113.
5 Creighton, T.E., Nature, 356 (1992) 194.
6 Stein, P.E., Leslie, A.G.W., Finch, J.T. and Carrell, R.W., J. Mol. Biol., 221 (1991) 941.
7 Loeberman, H., Tokuoka, R., Deisenhofer, J. and Huber, R., J. Mol. Biol., 177 (1984) 531.
8 Crowther, D.C., Evans, D.L.I. and Carrell, R.W., Curr. Opin. Biotechnol., 3 (1992) 399.
9 Stein, P. and Chotia, C., J. Mol. Biol., 221 (1991) 615.
10 Carrell, R.W. and Evans, D.L.I., Curr. Opin. Struct. Biol., 2 (1992) 438.
11 Mottonen, J., Strand, A., Symersky, J., Sweet, R.M., Danley, D.E., Geoghegan, K.F., Gerard, R.D. and Goldsmith, E.J., Nature, 355 (1992) 270.
12 White, J.M., Science, 258 (1992) 917.
13 Brasseur, R., Cornet, B., Burny, A., Vandenbranden, M. and Ruysschaert, J.M., AIDS Res. Hum. Retroviruses, 4 (1988) 83.
14 Brasseur, R., J. Biol. Chem., 266 (1991) 16120.
15 Horth, M., Lambrecht, B., Chuah Lay Khim, M., Bex, F., Thiriart, C., Ruysschaert, J.M., Burny, A. and Brasseur, R., EMBO J., 10 (1991) 2747.
16 Vonèche, V., Portetelle, D., Kettmann, R., Willems, L., Limbach, K., Paoletti, E., Ruysschaert, J.M., Burny, A. and Brasseur, R., Proc. Natl. Acad. Sci. USA, 89 (1992) 3810.
17 Gallaher, W.R., Segrest, J.P. and Hunter, E., Cell, 70 (1992) 531.
18 Epand, R.M., Cheetham, J.J., Epand, P.F., Yeagle, P.L., Richardson, C.D. and Degrado, W.F., Biopolymers, 32 (1992) 309.
19 Wilson, I.A., Skehel, J.J. and Wiley, D.C., Nature, 289 (1981) 366.
20 Levy-Mintz, P. and Kielian, M., J. Virol., 65 (1991) 4292.
21 Blobel, C.P., Wolfsberg, T.G., Turck, C., Myles, D.G., Primakoff, P. and White, J.M., Nature, 356 (1992) 248.
22 Bourat, G., Thoreau, E. and Mornon, J.P., J. Pharm. Belg., (1994) in press.
23 Colloc'h, N., Etchebest, C., Thoreau, E., Henrissat, B. and Mornon, J.P., Protein Eng., 6 (1993) 377.

24 Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer Jr., E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. and Tasumi, M., J. Mol. Biol., 112 (1977) 535.
25 Heringa, J. and Argos, P., J. Mol. Biol., 220 (1991) 151.
26 Sander, C. and Schneider, R., Proteins, 9 (1991) 56.
27 Boberg, J., Salakoski, T. and Vihinen, M., Proteins, 14 (1992) 265.
28 Hobohm, U., Scharf, M., Schneider, R. and Sander, C., Protein Sci., 1 (1992) 409.
29 Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P., FEBS Lett., 224 (1987) 149.
30 Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. and Mornon, J.P., Biochimie, 72 (1990) 555.
31 Semertzidis, M.T., Thoreau, E., Tasso, A., Henrissat, B., Callebaut, I. and Mornon, J.P., In Pickover, C. (Ed.) Visual Display of Biological Information, World Scientific Publishers, Teaneck, NJ, 1994, in press.
32 Woodcock, S., Mornon, J.P. and Henrissat, B., Protein Eng., 5 (1992) 629.
33 Callebaut, I., Renoir, J.M., Lebeau, M.C., Massol, N., Burny, A., Baulieu, E.E. and Mornon, J.P., Proc. Natl. Acad. Sci. USA, 89 (1992) 6270.
34 Schoentgen, F., Seddiqi, N., Bucquoy, S., Jolles, P., Lemesle-Varloot, L., Provost, K. and Mornon, J.P., Protein Eng., 5 (1992) 295.
35 Thoreau, E., Petridou, B., Kelly, P.A., Djiane, J. and Mornon, J.P., FEBS Lett., 282 (1991) 26.
36 Cherfils, J., Vaney, M.C., Morize, I., Surcouf, E., Colloc'h, N. and Mornon, J.P., J. Mol. Graphics, 6 (1988) 155.
37 Prusiner, S.B., Science, 252 (1991) 1515.
38 Weissmann, C., Nature, 351 (1991) 679.
39 Argos, P., J. Mol. Biol., 211 (1990) 943.
40 Chakrabartty, A., Schellman, J.A. and Baldwin, R.L., Nature, 351 (1991) 586.
41 Weiss, M.S. and Schultz, G.E., J. Mol. Biol., 227 (1992) 493.
42 Cowan, S.W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R.A., Jansonius, J.N. and Rosenbusch, J.P., Nature, 358 (1992) 727.
43 Richardson, J.S. and Richardson, D.C., Proteins, 4 (1988) 229.
44 Liquori, A.M., Q. Rev. Biophys., 2 (1969) 65.
45 Deleers, M. and Brasseur, R., Biochem. Pharmacol., 38 (1989) 2441.
46 Brasseur, R., Killian, J.A., De Kruijff, B. and Ruysschaert, J.M., Biochim. Biophys. Acta, 903 (1987) 11.
47 Hsiao, K., Baker, H.F., Crow, T.J., Poulter, M., Owen, F., Terwilliger, J.D., Westaway, D., Ott, J. and Prusiner, S.B., Nature, 338 (1989) 342.
48 Goldfarb, L.G., Petersen, R.B., Tabaton, M., Brown, P., LeBlanc, A.C., Montagna, P., Cortelli, P., Julien, J., Vital, C., Pendelbury, W.W., Haltia, M., Wills, P.R., Hauw, J.J., McKeever, P.E., Monari, L., Schrank, B., Swergold, G.D., Autilio-Gambetti, L., Gajdusek, D.C., Lugaresi, E. and Gambetti, P., Science, 258 (1992) 806.