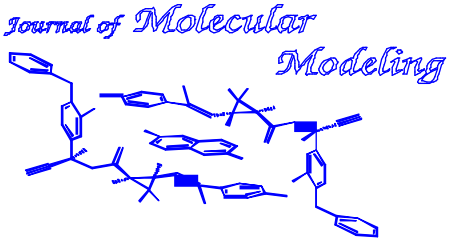
J. Mol. Model. **1997**, *3*, 203 – 215



Are Amphipathic Asymmetric Peptides Ubiquitous Structures for Membrane Destabilisation?

Mehdi Rahman[1,*], Laurence Lins[1], Annick Thomas-Soumarmon[2] and Robert Brasseur[1]

[1]Centre de Biophysique Moléculaire Numérique, Faculté des Sciences Agronomiques de Gembloux, 2, Passage des Déportés, B-5030 Gembloux, Belgium; Tel: +32-81-6225-25, Fax: +32-81-6225-22 (mehdirah@fsagx.ac.be)

[2]Unité de Recherches de Gastro-Entérologie, INSERM 10, Hôpital Bichat, 170 bd Ney, 75018 Paris, France

Received: 8 January 1997 / Accepted: 16 April 1997 / Published: 5 May 1997

Abstract

The fusion of some viruses (SIV, BLV, etc) to host cells implicates short fragments of the fusion protein that are asymmetric amphipathic helices in molecular modelling. The tilted orientation of these fragments at a water/lipid interface is directly related to their fusogenic capacity. On this basis, we have searched for fragments of sequences corresponding to "viral fusion peptides" in other proteins. We have developed a strategy to detect them from primary sequences. Many candidates were detected, especially in transmembrane areas of membranous proteins, in signal sequences and in globular proteins. We suggest that they are involved in the dynamics of lipid-protein interactions

Keywords: Tilted peptides, membrane destabilisation, signal sequences, lipases, membrane proteins

Introduction

The so-called "viral fusion peptides" are short fragments (11 to 18 amino acids) of viral fusion proteins. Fusion proteins are located in the capside spikes of viruses and are responsible for the selective recognition of the host cells [1, 2, 3, 4, 5]. After binding to the host cell or to the endosome, the fusion protein changes its conformation to unmask a fragment of sequence, the "fusion peptide", which inserts in the host membrane and triggers the fusion process [3]. Fusion peptides were initially characterised in Newcastle Disease Virus (NDV [6]), Simian Immunodeficiency Virus (SIV [1]), Bovine Leukaemia Virus (BLV [2]) and Human Immunodeficiency Virus (HIV [7]). More recently, fragments of sequence described as fusion peptide-like were found in non-viral proteins such as the cholesteryl ester transfer proteins (CETP) [8], spermatozoids [3], and signal sequences [9].

Modelling of NDV fusion protein has suggested that, due to its hydrophobicity profile when in an alpha-helical structure, the fusion peptide should not insert perpendicularly to the membrane surface but obliquely [6]. This tilted insertion is likely to destabilise the phospholipid bilayer by disturbing the parallel alignment of acyl chains. Similar modes of insertion were suggested from the modelling of fusion peptides of BLV, SIV and HIV [1, 10]. For BLV and SIV, mutants with different hydrophobicity profiles were calculated, prepared and experimentally tested for their fusogenic capacities [1, 2]. ATR infrared spectroscopy demonstrated that the angle of insertion of these peptides in membranes changed

* *To whom correspondence should be addressed*

as modelling had predicted [11]. When results from infrared spectroscopy and fusion were compared, it was concluded that the fusogenic capacity depended upon the tilted insertion of peptides in bilayers. Analysis of mutants also supported the hypothesis that a tilted insertion of peptides depends upon the gradient of hydrophobicity along the secondary structure rather than upon the nature of the amino acids.

A fusion-peptide like structure was also found by molecular modelling in the C-terminal end of the Alzheimer beta-amyloid peptide. This peptide has been tested experimentally for vesicle fusion and calcein leakage [12, 13]. The percentage of alpha-helical secondary structure has been measured by circular dichroism, which shows that helicity and fusogenic potential raise when trifluoroethanol is added. Those results are similar to those obtained with viral fusion peptide. They suggest that tilted peptides are not limited to viral fusion proteins and could be present in other protein families.

Fusion events and membrane destabilisation occur commonly in cells [3], for example internalisation of proteins in mitochondria and chloroplasts, translocation of proteins from one cytoplasmic compartment to another (endoplasmic reticulum, Golgi [14]), fusion of vesicles as during the excretion of secretory vesicle contents, cell to cell fusions (spermatozoid-egg fusion [15]), effect of lipases on organised substrates, insertion of signal peptides [9, 16], etc. The mechanisms of these events are unclear. Thus, the purpose of this study was to investigate whether fusion peptide-like structures could be involved in these events (for a review, see 17).

We first postulated that oblique peptides should be hydrophobic with an alpha-helical structure and an amphipathic asymmetric distribution (as observed in viral fusion peptides [1, 2, 11]). The hydrophobicity would be responsible for membrane insertion and the asymmetry of that hydrophobicity dictating an oblique angle of insertion. We now refer to such sequences as “tilted peptides”. Following this, we developed a procedure to detect tilted peptides from the primary sequences of proteins. We ran that procedure on protein data banks and reactive proteins were classified according to their biological functions.

Material and methods

Preparing a non-redundant protein data bank.

We used the 2nd May 1995 release of the OWL composite protein sequence database [18, 19] in order to prepare a non-redundant protein data bank. The entries of OWL came from 4 sources: the Swiss Prot release 31 [20], the NBRF release 44 [21], the GENBANK release 88.0 [22] and the NRL release 18.0 [23]. Swiss Prot and NBRF entries are well documented proteins primary sequences; GENBANK entries are translations of a DNA bank, and are poorly documented; NRL are the sequences of proteins whose 3D structure has been determined. We have extracted two sets of sequences: in the set I, sequences whose origin is NRL; in the set II, sequences whose origin is Swiss Prot and NBRF.

Table 1. *Transfer energy in kcal·mol^{-1} for the various atomic types. Negative values are associated to the hydrophobic atoms, positive values to the hydrophilic atoms.*

Atom	E^{tr} (kcal·mol^{-1})
C	-2.436
C (in CH/CH2/CH3 without explicit H)	-4.047
C (involved in a double bond)	-1.513
H (bonded to a C atom)	-0.537
H (bonded to a non C atom)	1.030
N	3.035
O	2.833
S	-2.751

The following algorithm was applied on each set in order to remove redundancy. Each sequence was checked against all the sequences following it in the bank. If the two sequences were too similar (criteria below), the second sequence was discarded. The criteria for redundancy are: first, having a similar size: the ratio shortest to longest must be greater or equal to 0.8. Second is an identity level greater than or equal to 30 %. The alignments have been realised with the pairing alignment program of ClustalW [24], using default values for parameters: gap open penalty = 10, gap extent penalty = 0.1, the BLOSUM homology matrix [25].

Search of amphipathic asymmetric peptides

The goal was to find peptides that are tilted at a lipid-water interface. In order to do this, we displaced a window along each protein sequence, and checked each peptide (Figure 1). We used window sizes ranging from 11 to 18 residues. For each peptide, two criteria must be met.

The first criterion was that the peptide was sufficiently hydrophobic to insert into lipids. Hydrophobicity was defined in terms of transfer energy of atoms from an hydrophobic to an hydrophilic phase (Table 1) [26]. For each of the 20 amino-acids present in proteins, the sum of the transfer energies of atoms with negative (hydrophobic) and positive (hydrophilic) transfer energy was computed (Table 2), giving hydrophobicity and hydrophilicity values for each amino-acid. With those values, for each peptide tested for obliqueness, the sum of negative and positive transfer energies for all its residues was computed, giving hydrophobicity and hydrophilicity values for the peptide. Only peptides where the absolute value of the ratio negative (hydrophobic) to positive (hydrophilic) transfer energy was greater than 2 were kept, as this is the inferior limit observed in viral fusion peptides [27, 28].

Table 2. *Sum of negative (hydrophobic) and positive (hydrophilic) transfer energies of atoms of the 20 amino-acids. $E^{trPhi}{}_i$ and $E^{trPho}{}_i$ are hydrophilic (positive values) and hydrophobic (negative values) energies of transfer respectively. The absolute value of the sum of negative atomic transfer energy is in column 2 and the sum of positive atomic transfer energy is in column 3.*

Amino-acid	abs($\Sigma E^{trPho}{}_i$) (kcal·mol^{-1})	$\Sigma E^{trPhi}{}_i$ (kcal·mol^{-1})	Ratio
A (Ala)	9.61	6.90	1.39
C (Cys)	12.36	7.93	1.55
D (Asp)	11.12	12.56	0.89
E (Glu)	15.17	12.56	1.21
F (Phe)	18.68	6.90	2.71
G (Gly)	5.56	6.90	0.81
H (His)	16.68	15.03	1.11
I (Ile)	21.75	6.90	3.16
K (Lys)	21.75	13.02	1.67
L (Leu)	21.75	6.90	3.16
M (Met)	20.45	6.90	2.96
N (Asn)	11.12	14.83	0.75
P (Pro)	17.70	5.87	3.02
Q (Gln)	15.17	14.83	1.02
R (Arg)	19.21	21.15	0.91
S (Ser)	9.61	10.76	0.89
T(Thr)	13.65	10.76	1.27
V (Val)	17.70	6.90	2.57
W (Trp)	21.71	10.96	1.98
Y (Tyr)	18.68	10.76	1.74

The second criterion was the oblique orientation of the peptide. As the orientation can only be computed on a 3D structure, the program modelled each peptide that passed the first criterion. The first step of the modelling was to build a 3D structure of the peptide, according to bond length and valence angles of the AMBER united atoms force field [29]. This was always done with an alpha-helix secondary structure as known fusion peptide and signal sequence seem to adopt this structure when inserted in lipids [11, 30, 31, 32]. The second step was to optimise the conformation of lateral chains by energy minimisation. This optimisation was carried out residue by residue. Most frequent conformations [33] were imposed on the current residue, and interaction energy computed with neighbouring residues (up to -4 and +4). The conformation that gave the best (minimum) energy was kept, and the next residue treated. Energy was computed according to the AMBER force field.

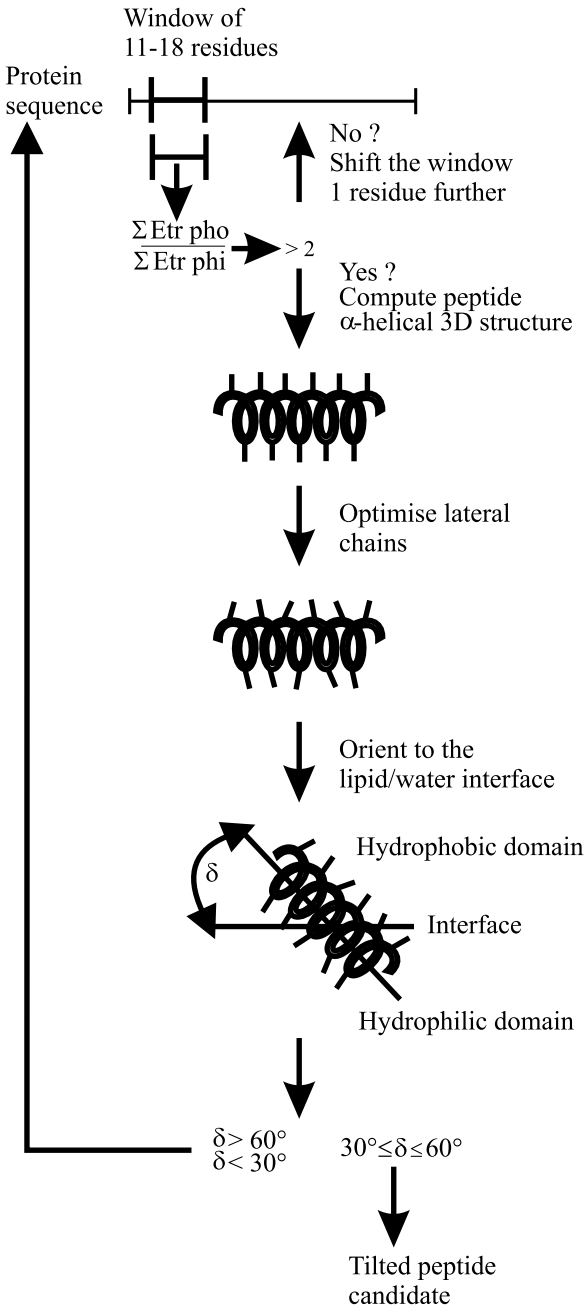


Figure 1. *Representation of the different steps of the procedure to detect asymmetric amphipathic peptides out of the primary sequence of proteins (see Methods).*

The third step was the orientation at the interface and the computation of the insertion angle. Hydrophobic $\vec{C}^{pho}$ and hydrophilic $\vec{C}^{phi}$ centres were computed for the peptide, using Eqs. 1 and 2, respectively. E^{tr} are the transfer energies of atoms (Table 1). $E^{trPhi}{}_i$ and $E^{trPho}{}_i$ are hydrophilic (positive values) and hydrophobic (negative values) transfer energies, respectively. $\vec{r}_i$ is the position vector of the atom. The total number of atoms is m. The position of the interface $\vec{I}$ is given by the Eq. 3 [27, 34].

$$\vec{C}^{phi} = \frac{\sum_{i=1}^{m} E^{trPhi_i} \vec{r}_i}{\sum_{i=1}^{m} E^{tr_i}} \quad (1)$$

$$\vec{C}^{pho} = \frac{\sum_{i=1}^{m} E^{trPho_i} \vec{r}_i}{\sum_{i=1}^{m} E^{tr_i}} \quad (2)$$

$$\frac{\sum_{i=1}^{m} E^{trPhi_i}}{\vec{C}^{phi} - \vec{I}} = \frac{\sum_{i=1}^{m} E^{trPho_i}}{\vec{C}^{pho} - \vec{I}} \quad (3)$$

The orientation of the peptide at the interface was computed as follows (Figure 2). n is the number of alpha carbons in the peptide. $\vec{C}_1$, $\vec{C}_3$, $\vec{C}_{n-2}$ and $\vec{C}_n$ are the position vectors of the first, third, n-2 and last alpha carbons. $\vec{A}_1$ and $\vec{A}_2$ are two vectors computed with Eqs. 4 and 5. A normed vector $\vec{A}$, which is oriented nearly parallel to the helix axis, is computed by Eq. 6. A normed vector $\vec{P}$, whose direction is perpendicular to the interface, is computed by the Eq. 7. As $\vec{A}$ and $\vec{P}$ are normed vectors, their scalar product is the cosine of the angle between them. This angle between the helix and the perpendicular to the interface, γ, is obtained by Eq. 8. The angle δ between the helix axis and the interface is given by the Eq. 9. Peptides were kept when this angle was between 30 and 60 degrees.

$$\vec{A}_1 = \vec{C}_1 + \vec{C}_3 \quad (4)$$

$$\vec{A}_2 = \vec{C}_{n-2} + \vec{C}_n \quad (5)$$

$$\vec{A} = \frac{\vec{A}_2 - \vec{A}_1}{\left|\vec{A}_2 - \vec{A}_1\right|} \quad (6)$$

$$\vec{P} = \frac{\vec{C}^{pho} - \vec{C}^{phi}}{\left|\vec{C}^{pho} - \vec{C}^{phi}\right|} \quad (7)$$

$$\gamma = \arccos\left(\vec{P} \cdot \vec{A}\right) \quad (8)$$

$$\delta = \left|90 - \gamma\right| \quad (9)$$

Representation of the interface

To represent the lipid-water interface, the position of the interface was computed as explained above, and represented as a plane. When there was an experimental structure for the

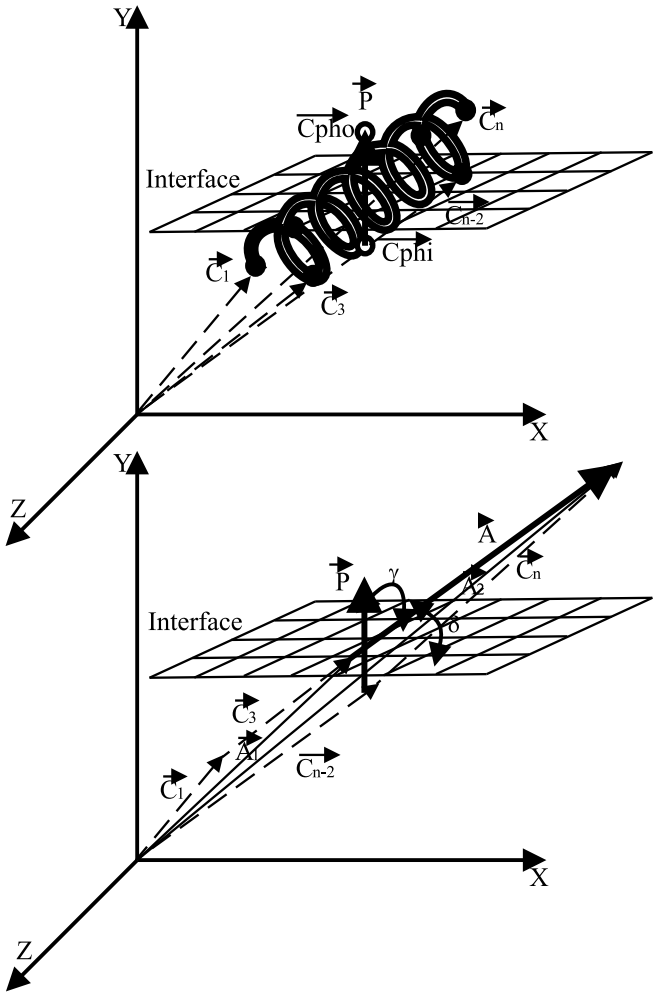


Figure 2. *Representation of the different vectors used to calculate the angle of helix insertion in an hydrophilic/ hydrophobic interface. Etr are energies of transfer of atoms (Table 1). The top graph represents the peptide alpha helix, with position vectors of α carbons (C_1, C_3, C_{n-2}, C_n), and the vector $\vec{P}$ (equation 7) joining hydrophobic and hydrophilic centres (perpendicular to the interface). The bottom graph plots the computed vectors: $\vec{A}_1$ and $\vec{A}_2$ are the two sum vectors (equations 4 and 5), $\vec{A}$ is the difference between $\vec{A}_1$ and $\vec{A}_2$ (equation 6), and is parallel to the helix axis. Computed angles, γ (equation 8) and δ (equation 9), the angle between the helix and the interface, are indicated.*

Table 3. *Proteins with amphipathic asymmetric peptides found in the set I (NRL data bank [23]). Column 1 gives the size of the window used for the search of amphipathic asymmetric peptides. Column 2 gives the total number of proteins found. The three last columns give the percentages of peptides that have the corresponding secondary structure.*

Window size	Proteins	Helix	Sheet	Coil
11	499	37	26	37
12	289	33	20	47
13	239	35	22	43
14	207	28	18	54
15	167	31	20	49
16	138	38	20	42
17	107	32	20	48
18	79	37	19	44

peptide, it was used for the orientation. If hydrogens were not present in this structure, they were added to all but carbon atoms with respect to valence angles and bond lengths of the united-atoms AMBER force field. If no experimental structure existed, peptides were constructed as explained above. The only difference was during the optimisation of lateral chains, where all atoms were taken into account, and two iterations performed. Amino- and carboxy-extremities were not added. The only exception was viral fusion peptides, where the amino-extremity was added: this extremity is present in the active form.

Molecular Hydrophobicity Potentials (MHP)

Hydrophobicity profiles of peptides were calculated using the MHP method [26], on a tridimensional grid, with a precision of 3 Å.

f_{ij}, the portion of atom i covered by an atom j, was computed by Eq. 10 for each atom couple [35], where r_i and r_j are the atoms van der Waals radius, d_{ij} the distance between them, r_{sol} the solvent radius (1.4 Å). f_{ij} was null if the distance between the two atoms was large enough to accommodate a solvent molecule between them.

$$f_{ij} = \frac{r_j^2}{4\left(r_i + r_j\right)^2}\left(1 - \frac{d_{ij} - r_i - r_j}{2r_{sol}}\right) \qquad (10)$$

f_i is the fraction of the surface of the atom i which is accessible to the solvent. f_i was computed for each atom by Eq. 11.

Table 4. *Number of proteins retrieved in the set II (Swiss Prot [20] and NBRF [21] data banks), with the different window sizes. Column 1 gives the size of the window used in the search. Column 2 gives the total number of proteins; the 3, the number of proteins where an amphipathic asymmetric peptide was found in the signal sequence; the 4, the number of proteins where an amphipathic asymmetric peptide was found in a transmembrane area. In columns 2- 4, the first number is the total number for the original non redundant bank.*

Window size	Proteins number	Signal sequence number	Transmembrane number
	1419	168	104
11	854	141	94
12	803	135	93
13	736	127	92
14	678	118	90
15	637	106	90
16	579	94	88
17	528	78	86
18	476	67	82

$$f_i = 1 - \sum_{i=1}^{m} f_{ij} \qquad (11)$$

The Eq. 12 was used to obtain the hydrophobic potential MHP on each point of the grid. d_i is the distance between the point and the atom i, r_i is the van der Waals radius of the atom i, and E^{tr} are the transfer energies of atoms (Table 1).

$$\mathrm{MHP} = \sum_{i=1}^{m}\left[E^{tr_i} f_i \exp\left(-\left|r_i - d_i\right|\right)\right] \qquad (12)$$

Once the potential MHP was computed at each point, lines of hydrophobic (–0.1 kcal·mol^{-1}) and hydrophilic (0.1 kcal·mol^{-1}) isopotentials were calculated and drawn in 3D around the molecule.

Software

Calculations for molecular modelling and manipulations of calculated structures were performed with WinMGM software (Windows Molecular Graphic Manipulation [36]). Peptide search and data banks manipulation has been performed with the WinDNA program. Both programs were run under Windows NT 3.51, on 60 and 90 MHz pentium computers.

Table 5. *Different amphipathic asymmetric peptides found in bacteriorhodopsin (1bct) and photosynthetic reaction centre (2rcr). The first column gives the NRL code of the protein and the position of the peptide. The second column give its sequence, and the third, the angle between the axis of the amphipathic asymmetric peptide helix and the interface plane.*

PDB code and peptide limit	Sequence	Angle between the helix and the interface (°)	Transmembrane
1bct, 177-192	vtvvlwsaypvvwlig	43	yes
1bct, 195-212	gagivplnietllfmvld	35	yes
2rcr, H, 11-28	dlaslaiysfwiflagli	45	yes
2rcr, H, 53-67	qgpfplpkpktfilp	50	no
2rcr, L, 21-37	lfdfwvgpfyvgffgva	56	no
2rcr, L, 40-55	ffaalgiiliawsavl	43	yes
2rcr, L, 63-75	lisvyppaleygl	49	no
2rcr, L, 111-128	lgigyhipfafafailay	54	yes
2rcr, L, 174-189	miaisfffnalalal	57	yes
2rcr, L, 234-251	lllslsavffsalcmiit	35	yes
2rcr, M, 49-66	piylgslgvlslfsglmw	47	mostly
2rcr, M, 90-100	fffsleppape	57	no
2rcr, M, 106-123	aaplkegglwliasffmf	39	yes
2rcr, M, 147-164	awaflsaiwlwmvlgfir	34	yes
2rcr, M, 203-218	glsiaflygsallfam	47	yes
2rcr, M, 268-285	waiwmavlvtltggigil	51	yes

Results and discussion

Data Banks of sequences.

Sequence data banks are not homogenous representations of all families of proteins. Numerous sequences of one family can be present while a single member, or even none, of other families are present. Since our purpose was to test how fusion-like peptides are distributed in all kinds of proteins, we were interested in a non-redundant bank of sequences. In order to decrease the redundancy, each sequence of set I (NRL bank) and each sequence of set II (Swiss Prot plus PIR banks) was aligned with all following sequences of their respective set, using the alignment program ClustalW [24]. If two sequences were similar, the second was discarded. This procedure had a very drastic effect: set I decreased from 5069 to 863 sequences. Set II dropped from 69118 to 1419 sequences.

Detection of tilted peptides in non-redundant data banks.

Set I (NRL bank) contains proteins for which the structure is experimentally known. The procedure to detect amphipathic asymmetric helices was carried out on the sequences of set I as defined in Methods (Table 3). Since true structures are known, predictions as well as true structures are presented in the Table 3. This analysis demonstrate that some peptides are not always helical. In fact some tilted peptides of proteins are beta strands and coiled in structure. These results should not refute the hypothesis, since some viral fusion peptides (e.g. the haemagglutinin [37]) are beta strands when buried inside fusion proteins and convert into alpha helices, when unmasked, to interact with the membrane of the host cell [4]. Therefore, it is important to keep in mind that changes of secondary structures might be possible and linked to a resting/functioning switch of the fragment structure. We suggest therefore that the helix could be the active form of the peptide and the beta strand the quiescent one.

The transmembrane proteins of set I (NRL bank) identified as having tilted peptides were the two photosynthetic reaction centre, prc [38, 39] and rcr [40], and the bacteriorhodopsin bct [41]. All have alpha-helical bundle structures and have at least one tilted peptide located in the transmembrane region. Furthermore, no tilted peptide was detected from the porin sequence (the last transmembrane proteins of set I), which has a beta sheet barrel structure [42, 43]. When the window of analysis of primary sequences was lengthened (Table 3), the transmembrane proteins remained but most globular proteins disappeared. Therefore, tilted peptides of these three transmembrane proteins are longer than tilted peptides of globular proteins.

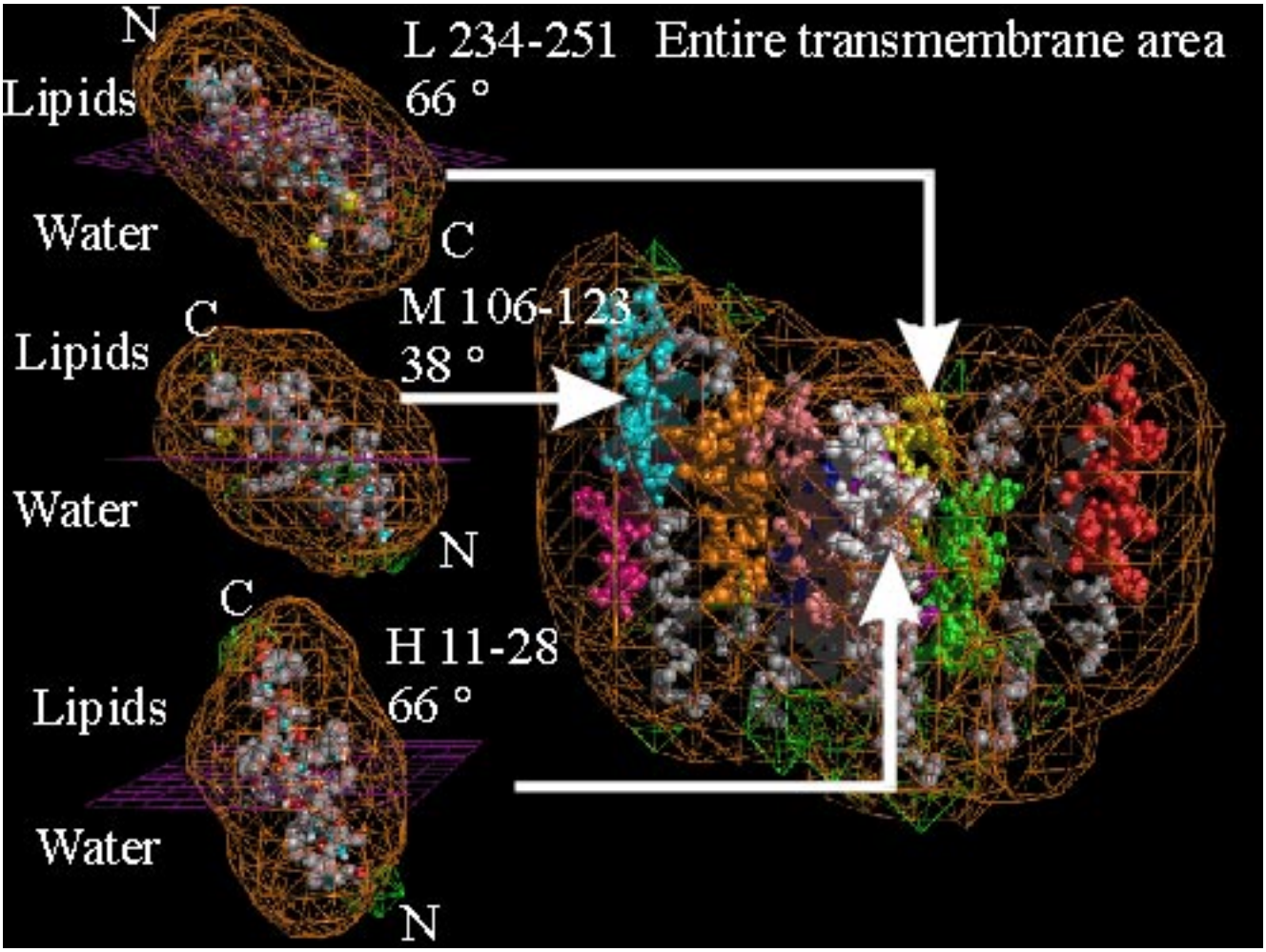


Figure 3. *On the left, a few transmembrane amphipathic asymmetric peptides from 2rcr: up is 234-251 of chain L, middle is 106-123 of chain M, bottom is 11-28 of chain H. Amino- and carboxy-ends, angles between peptide axis and the interface (see Methods) are indicated on the figure. All peptides are taken from the experimental structure, positioned relative to the lipid-water interface (in violet), oriented with the lipid part of the membrane up (lipidic and water phases are indicated). The entire transmembrane area of 2rcr is represented on the right. Their MHP is computed: hydrophobic isopotentials (-0.1 kcal·mol^{-1}) are in orange, hydrophilic ones (0.1 kcal·mol^{-1}) are in green. In the entire transmembrane domain, the different transmembrane amphipathic asymmetric peptides are coloured. From chain H: 11-28 in white; from chain L: 40-55 in red, 111-128 in green, 174-189 in blue, 234-251 in yellow; from chain M: 49-66 in magenta, 106-123 in cyan, 147-164 in orange, 203-218 in purple and 268-285 in pink. The MHP around the entire domain shows it as entirely hydrophobic, and so stable in lipids.*

In general, the soluble proteins selected in set I are not expected to interact with lipids, except for the lipases (1thg [44], 1tca [45], 3tgl [46], 1hpla [47], 1tahb [48]) and apolipoprotein D (2apd [49]). Whether tilted peptides of the proteins that do not interact with lipids such as the t-RNA synthase, the dehydrogenases, hydrolases, etc. interact with other hydrophobic/hydrophilic interfaces, such as that between the hydrophobic core of a protein and its hydrophilic surface [50] remains to be explored.

Set II contains proteins with unknown structure. More protein families are represented in set II than in set I, i.e. soluble proteins are still numerous but the relative representation of transmembrane proteins is higher than in set I (7 % as compared to 0.2 %). The procedure of detection of tilted peptides was carried out (Table 4). Using a window of 11 residues, 60 % of the proteins (854 out of 1419 proteins) were positive. From the analysis of set I, we concluded that tilted peptides were present in transmembrane proteins. From the analysis of set II, we concluded that tilted peptides were in fact frequent in transmembrane proteins, since 92 % (96 out of 104 membrane proteins) of those proteins have at least one tilted peptide. We further conclude that tilted peptides

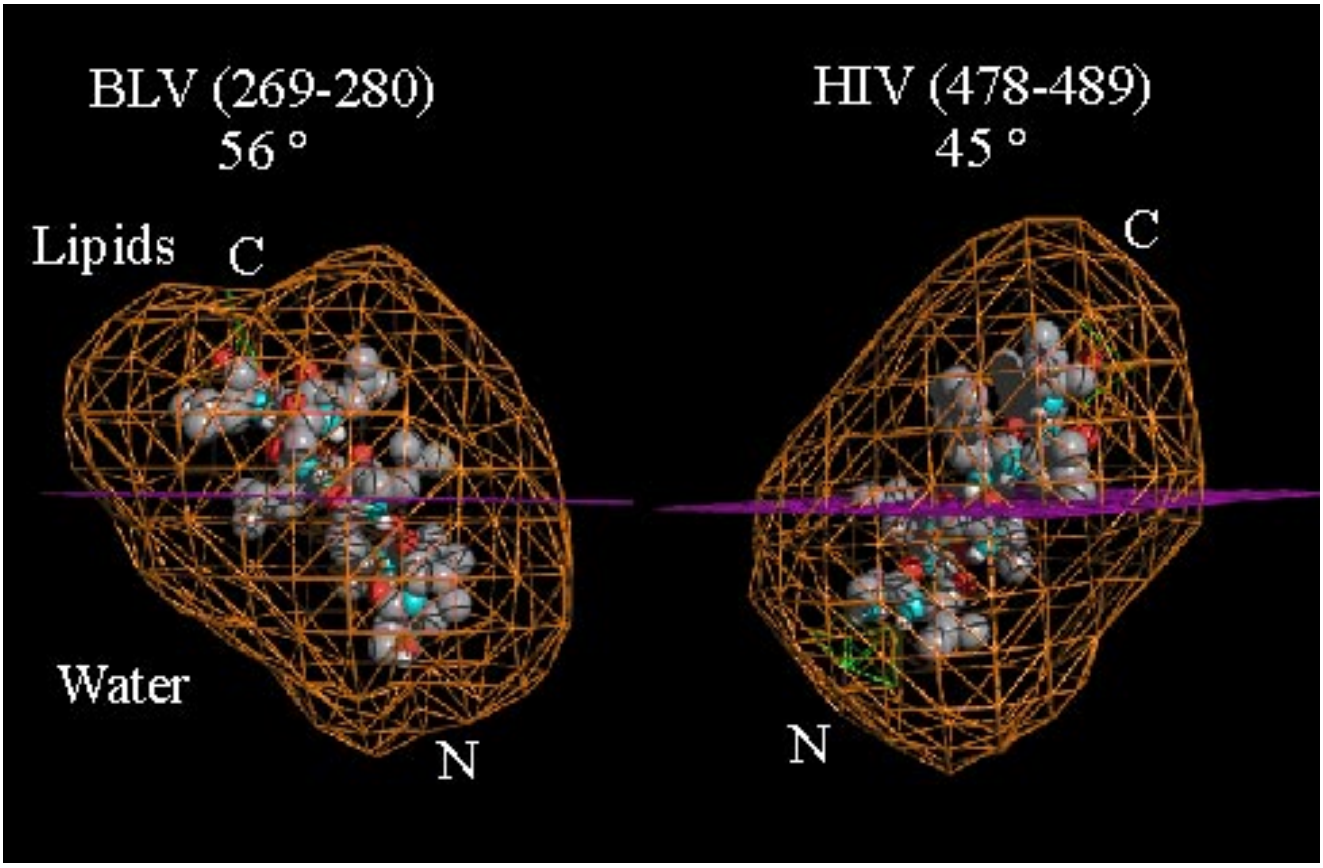


Figure 4. *Viral fusion peptide of BLV on left (269-280) and HIV on right (478-489). The representation of interface, MHP, Amino- and carboxy-end, insertion angle are the same as in Figure 3.*

are more frequently observed in transmembrane proteins than in the whole set (92 % as compared to 60 %). Tilted peptides of transmembrane proteins are mainly located in the intramembrane domains (location of those domains are only predicted since no experimental confirmation is available). Tilted peptides of transmembrane proteins are longer than those of other proteins: lengthening the screening window from 11 to 18 residues favours the selection of transmembrane proteins over the others since the percentage of transmembrane proteins then increases from 11 to 17 % of all selected peptides. This supports the hypothesis that tilted peptides of transmembrane proteins are longer than those of other kinds of proteins. Furthermore, analysis of Table 4 supports the proposal that an optimal window to detect tilted peptides in transmembrane proteins using our algorithm is 18 residues, since 79 % of the tilted peptides of membrane proteins were detected and 17 % of all tilted peptides were from the 7 % of transmembrane proteins.

Signal sequences are involved in the secretion of proteins and could interact with membranes to initiate the protein transfer [16]. This process has analogies with the viral fusion and it seems reasonable to suggest that tilted peptides could be involved in the insertion process as in viral fusion. Tilted peptides are present in 89 % of the documented signal sequences (149 out of 168), in the proteins of set II. The frequency is nearly as high as for transmembrane proteins. Tilted peptides of signal sequences are shorter than those of transmembrane proteins since increasing the screening window from 11 to 18 residues, reduced selection of signal sequences by 52 %, whereas selected transmembrane proteins are decreased by only 13 %. This suggests that the length of tilted peptides is related to their stability in the membrane since signal peptides with shorter tilted peptides cross the membrane whereas transmembrane proteins are stabilised within.

In set II, as in set I, many soluble proteins were selected with tilted peptides in an area for which no function is indicated. It is probable that for some of those proteins, the selected peptide is part of a transmembrane or signal region, which has not been identified. Some of those proteins also interact with lipids, such as apolipoproteins B and D, and lipases.

Transmembrane Proteins

As the 3D structure of most membrane proteins is unknown, we focused our examination on the 3 proteins having a resolved alpha-helical structure: the two photosynthesis reaction centres (1prc [38, 39], 2rcr [40]) and bacteriorhodopsin [41]. Both photosynthesis reaction centres are similar multimeric transmembrane proteins, 1prc contains 4 chains:

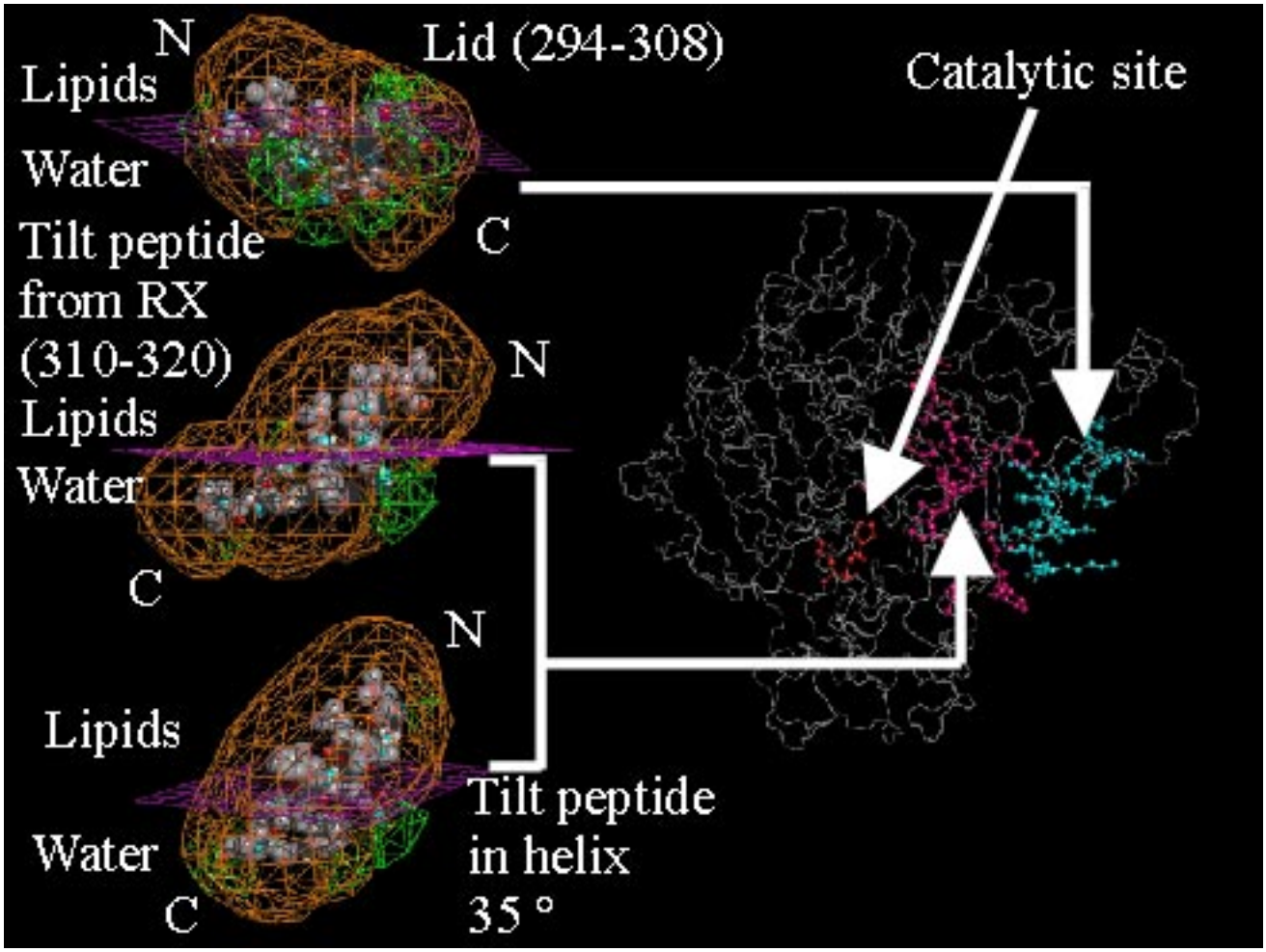


Figure 5. *On the left are the amphipathic asymmetric peptide and the lid from 1thg lipase. The helix of the lid (from the experimental structure) is on the top, the amphipathic asymmetric peptide, with its conformation from the experimental structure is on the middle, the same peptide put in helix at the bottom. The representation of interface, MHP, amino- and carboxy-end, insertion angle are the same as in Figure 3. On the right is the entire protein. The backbone of the protein is represented in grey. The residues of the catalytic site are in red. They are covered by the amphipathic asymmetric peptide, whose residues are in magenta. The helix of the lid is in cyan. Note that the amphipathic asymmetric peptide is far more hydrophobic than the lid, with protruding hydrophobic residues.*

C, H, L and M, 2rcr contains 3 chains: L, H and M. Chains with the same name are homologous between the two proteins. It is worth noting that since the algorithm used removed redundancy, all the chains of photosynthesis reaction centres were not in the data bank searched: only M and C of 1prc (H and L are absent since they were discarded as redundant to the H and L chains of 2rcr) and L and H of 2rcr (M is absent because it was discarded as redundant to the M chain of 1prc) are present. For completeness, we added the missing chains of 2rcr. 1prc, redundant with 2rcr, was discarded.

By increasing the size of the screening window, overlapping fragments of sequences were selected as tilted peptides; the consensus peptides are listed in Table 5. Most tilted peptides are transmembranous, except when specified. The molecular hydrophobicity isopotential profiles (MHP [26]) and the angle of insertion at a hydrophobic/hydrophilic interface of all selected fragments were calculated from the PDB coordinates (Figure 3). All structures have a large hydrophobic envelope with hydrophilic patches. Hydrophobicity isopotential profiles were very similar to those predicted for the virus fusion peptides of BLV (269-280) and HIV (478-489) (Figure 4).

Thus transmembrane peptides, stably inserted quasi-perpendicularly with respect to the membrane surface, when in the protein core, undergo a tilted orientation when alone and have the same hydrophobicity profile as the tilted peptides of virus fusion proteins which are unstable in membranes. In the case of the photosynthetic reaction centres, if MHP is computed out of the entire transmembrane area, the domain is clearly hydrophobic (Figure 3), and inserted perpendicularly to the lipid/water interface. Therefore one hypothesis is that tilted peptides destabilise lipids bilayers during their in-

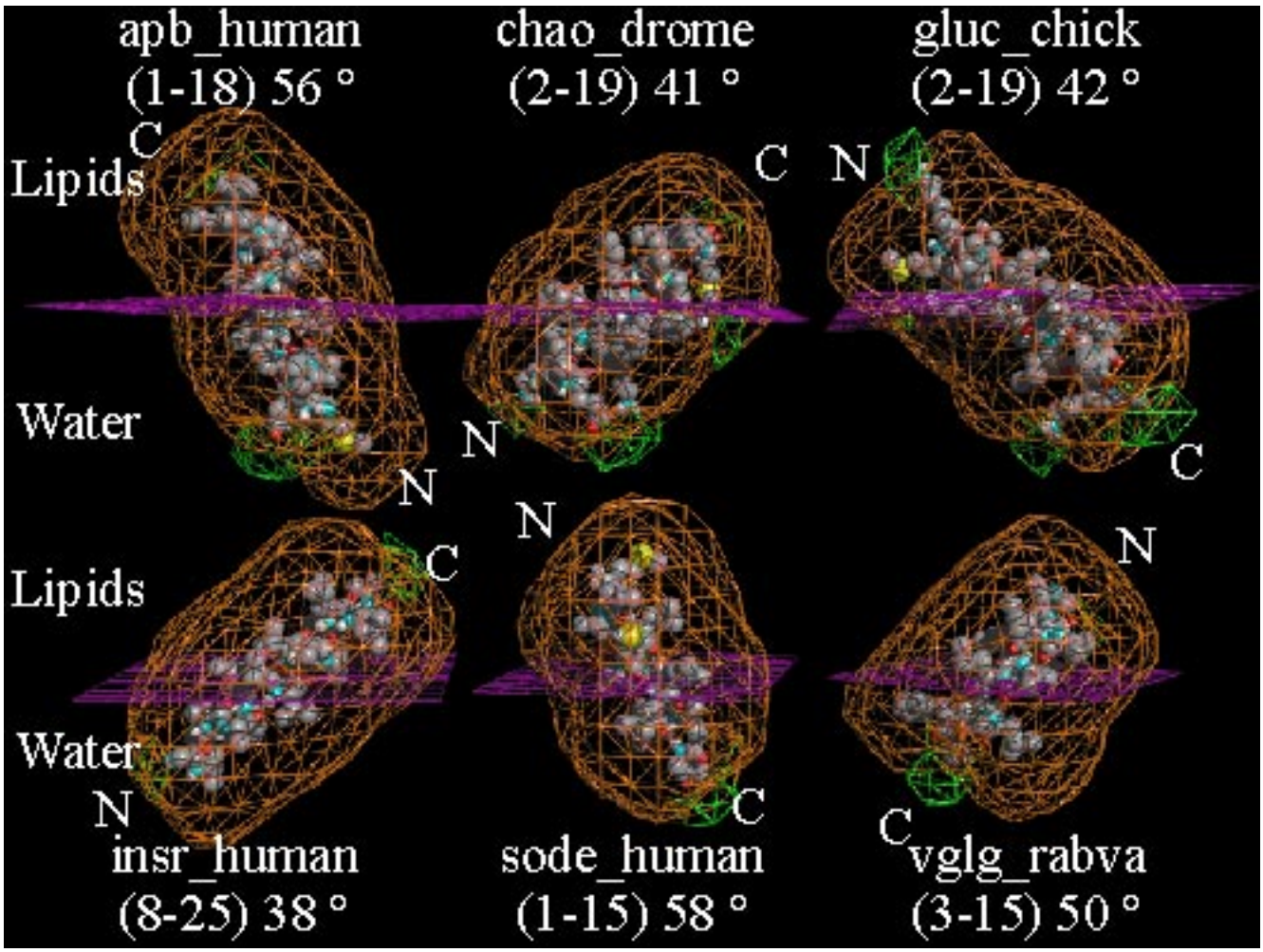


Figure 6. *The 6 amphipathic asymmetric peptides from signal sequences presented in Table 6. The representation of interface, MHP, Amino- and carboxy-end, insertion angle are the same as in Figure 3. Note the variety of the patterns of the hydrophilic patches on the different peptides.*

dividual insertion and that, via helix-helix interactions, hydrophilic domains are buried. This suggests a role for tilted peptides in the insertion of proteins in membranes and interactions between helices.

Lipases

In the lipases, the recognition site for lipids is a very mobile amphipathic helix (the lid) whose hydrophobic side seems to cover the catalytic site, while the hydrophilic side faces the solvent. The catalytic site is unmasked when the lipase reaches a water/lipid interface, undergoing interfacial activation [51, 52, 53]. During lipid hydrolysis, the hydrophobic side of the helix is in contact with acyl chains and holds lipid in a correct orientation to present the ester bond inside the catalytic site.

Due to their putative mechanism of action, we examined whether, in the tridimensional structure, a region close to the recognition site of lipases contains a tilted peptide. In set I, the following lipases have tilted peptides: 1thg (triacylglycerol lipase from Geotrichum candidum [44]) in 23-34, 71-84, 117-134 and 309-323, 1tca (triacylglycerol hydrolase from Candida antarctica, form B [45]) in 27-39, 59-74 and 275-287, 3tgl (triacylglycerol lipase from Rhizomucor miehei [46]) in 45-56, 88-98 and 200-214, 1hpla (triacylglycerol hydrolase, chain a, from Equus caballus [47]) in 16-30 and 207-219, 1tahb (triacylglycerol hydrolase, chain b, from Pseudomonas glumae [48]) in 243-254.

Here, we focus our study on the 309-323 tilted peptide of 1thg. It follows the helix of the lipid recognition site (294-308), and is located just over the catalytic site (Figure 5). The tilted peptide, in the free form, is part of a long exposed loop which joins the helix of the recognition site to the core of the protein. We have calculated the orientation relative to the interface of the lipid recognition site helix (294-308) and of the amphipathic asymmetric peptide (309-323). MHP were also computed on both peptides. According to those results (Figure 5), the helix 294-308, is amphipathic, and lies parallel to interface, while the loop 309-323 lies obliquely (Figure 5).

Table 6. *A few proteins from set II (Swiss Prot [20] and NBRF [21] data banks) with an amphipathic asymmetric peptide in their signal sequence. The first column indicate the id of the protein, the second one, its name, the third, the length of the amphipathic asymmetric peptide, the fourth, the sequence of the signal peptide (the amphipathic asymmetric peptide is in bold), and the last one, the angle formed by the amphipathic asymmetric peptide (modelled as an alpha-helix) and the lipid-water interface.*

Protein	Name	Length	Signal sequence	Angle between the helix and the interface (°)
apb_human	apolipoprotein b-100 precursor	18	**mdpprpalla llalpall**ll llagara	56
chao_drome	chaoptin precursor	18	m**glefffkfg yvfltitlm**i miwmslara	41
gluc_chick	glucagon precursor	18	m**kmksiyfia glllmivqg**s wq	42
insr_human	insulin receptor precursor	18	mgtggrr**gaa aapllvavaa lllga**ag	38
sode_human	extracellular superoxide dismutase precursor	15	**mlallcscll laaga**sda	58
vglg_rabva	spike glycoprotein precursor (rabies virus)	13	mv**pqvllfvl llgfs**lcfg	50

The N-terminal part of this loop contains mainly hydrophobic residues (its sequence is LFGLLP) and is the most exposed region of the sequence in the experimental structure. This exposed hydrophobic area could interact with lipids when the enzyme meets its substrate, and then adopts a helical conformation. In this conformation, the peptide lies obliquely relative to the theoretical lipid/water interface, with FGLLP residues in the lipid part. Its MHP profile is similar to that of transmembrane helices and viral fusion peptides (Figure 5).

Signal sequences

An important fraction of signal sequences present in the bank, 149 out of 168 (89 %), contains a short to medium sized amphipathic asymmetric peptide (Table 4). Those proteins have diverse functions: they are transmembrane proteins, cellular receptors, toxins, enzymes and hormones... The size range of the signal sequences is wide (most have between 18 and 30 residues), while most amphipathic asymmetric peptides found within them are relatively short (11-15 residues).

As most selected signal sequences are indicated as "potential" or "by homology", we selected a few well established signal sequences (Table 6). The related proteins cover various targeting: insertion in the membrane (the insulin receptor and chaoptin), secretion (apolipoprotein B-100 [16], glucagon, extracellular superoxide dismutase), and viral capside protein. We modelled the amphipathic asymmetric peptide as an alpha-helix since helicity of signal peptides increases when inserted into lipids [30], and inserted the peptides into a lipid-water interface and computed their MHP. Results are presented in Figure 6: all amphipathic asymmetric peptides present a similar hydrophobicity profile, mostly hydrophobic, with a few hydrophilic areas which have great homology with those of viral, transmembranous and lipase amphipathic asymmetric peptides (Table 6).

Two observations can be made. Firstly, although almost 90 % of signal sequences have amphipathic asymmetric peptide, the remaining 10 % does not contain one. Thus the presence of an amphipathic asymmetric peptide suggests at least one mechanism by which the targeting can be made, but others may exist. Secondly, the amphipathic asymmetric peptide is part of the signal peptide, but never the entire sequence. We can thus suppose that the activity of the tilted peptide (i.e. specificity) is controlled by the other domains of the signal peptide.

Conclusion

We have developed a strategy to identify amphipathic asymmetric peptides in primary sequences of proteins. This enabled us to screen protein data banks for such peptides in a systematic manner. We found different kinds of amphipathic asymmetric peptides, those in transmembrane proteins, those in lipases, those in signal sequences. We therefore propose that those peptides are implicated in different functions: insertion in the membrane, destabilisation of lipids, and targeting. Thus, in this study we have demonstrated that sequences corresponding to amphipathic asymmetric peptides are very general feature of proteins rather than a specificity of viral fusion proteins.

Acknowledgements: R. Brasseur is Senior Research Associate of the Belgian Fonds National de la Recherche Scientifique. We are grateful to the Association Française de Lutte contre la Mucoviscidose, the Belgian Fonds National de la Recherche Scientifique and the Fonds National pour la Recherche Fondamentale Collective (grant n° 2.4534.95) for financial support. We are grateful to Dr. Philippa Talmud for text reviewing.

References

1. Horth, M.; Lambrecht, B.; Chuah Lay Khim, M.; Bex, F.; Thiriart, C.; Ruysschaert, J.-M.; Burny, A.; Brasseur, R. *The EMBO Journal* **1991**, 10, 2747-2755.
2. Vonèche, V.; Portetelle, D.; Kettmann, R.; Willems, L.; Limbach, K.; Paoletti, E.; Ruysschaert, J.-M.; Burny, A.; Brasseur, R. *Proc. Natl. Acad. Sci. USA* **1992**, 89, 3810-3814.
3. White, J.M. *Science* **1992**, 258, 917-924.
4. Wharton, S.A.; Martin, S.R.; Ruigrok, R.W.H.; Skehel, J.J.; Wiley, D.C *J. Gen. Virol.* **1988**, 69, 1847-1857.
5. Burny, A.; Bex, F.; Brasseur, R.; Chuah Lay Khim, M.; Delchambre, M.; Horth, M.; Verdin, E. *Journal of Acquired Immune Deficiency Syndromes* **1988**, 1, 579-282.
6. Brasseur, R.; Lorge, P.; Goormaghtigh, E.; Ruysschaert, J.-M.; Espion, D.; Burny, A. *Virus Genes* **1988**, 1, 325-332.
7. Gallaher, W.R. Virus *Cell* **1987**, 50, 327-328.
8. Tall, A.R. *J. Lipids Res.* **1993**, 34, 1255-1274.
9. Talmud, P.; Lins, L.; Brasseur, R. *Protein Eng.* **1996**, 9, 317-321.
10. Brasseur, R.; Cornet, B.; Burny, A.; Vandebranden, M.; Ruysschaert, J.-M. *Aids Res. Hum. Retroviruses* **1988**, 4, 83-90.
11. Martin, I.; Dubois, M.-C.; Defrise-Quertain, F.; Saermark, T.; Burny, A.; Brasseur, R.; Ruysschaert, J.-M. *J. Virol.* **1994**, 68, 1139-1148.
12. Pillot, T.; Goethals, M.; Vanloo, B.; Talussot, C.; Brasseur, R.; Vandekerckhove, J.; Rosseneu, M.; Lins, L. *J. Biol. Chem.* **1996**, 271, 28757-28765.
13. Pillot, T.; Goethals, M.; Vanloo, B.; Lins, L.; Brasseur, R.; Vandekerckhove, J.; Rosseneu, M. *Eur. J. Biochem.*; in press.
14. Rothman, J. E.; Orci, L. *Nature* **1992**, 355, 409-415.
15. Primakoff, P.; Hyatt, H.; Tredick-Kline, J. *J. Cell Biol.* **1987**, 104, 141-149.
16. Sturley, S.; Talmud, P.; Brasseur, R.; Culbertson, M.; Humphries, S.; Attie, A. *J. Biol. Chem.* **1994**, 269, 21670-21675.
17. Brasseur, R.; Pillot, T.; Lins, L.; Vandekerckhove, J.; Rosseneu, M. *TIBS* **1997**, *22*, 167 - 171.
18. Bleasby, A.J.; Wootton, J.C. *Prot. Eng.* **1990**, 3, 153-159.
19. Bleasby, A.J.; Akrigg, D.; Attwood, T.K. *Nucleic Acids Res.* **1994**, 22, 3574-3577.
20. Bairoch, A.; Boeckman, B. *Nucleic Acids Res.* **1991**, 19, Suppl.; 2247-2249.
21. George, D.G, Barker, W.C, Hunt, L.T. *Nucleic Acids Research* **1986**, 14, 11-15.
22. Benson.; D, Lipman, D.J. *Nucleic Acids Res.* **1993**, 21, 2963-2965.
23. Namboodiri, K.; Pattabiraman, N.; Lowrey, A.; Gaber, B.; George, D.G.; Barker, W.C. *PIR Newsletter* **1989**, 8, 5.
24. Thompson, J.D.; Higgins, D.G.; Gibson, T.J. *Nucleic Acids Research* **1994**, 22, 4673-4680.
25. Henikoff, S.; Henikoff, J.G. *Proc. Natl. Acad. Sci. USA* **1992**, 89, 10915-10919.
26. Brasseur, R. *J. Biol. Chem.* **1991**, 266, 16120-16127.
27. Brasseur, R.; Deleers, M.; Ruysschaert, J.-M. *J. Col. Interf. Sci.* **1986**, 114, 277-282.
28. Brasseur, R. Molecular *Description of Biological Membranes by Computer Aided Conformational Analysis*. CRC Press.
29. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta Jr.; S.; Weiner, P. *J. Am. Chem. Soc.* **1990**, 106, 765-784.
30. Goormaghtigh, E.; Martin, I.; Vandenbranden, M.; Brasseur, R.; Ruysschaert, J.-M. *Biochem. Biophys. Res. Com.* **1989**, 158, 610-616.
31. Martin, I.; Defrise-Quertain, F.; Mandieau, V.; Nielsen, N. M.; Saemark, T.; Burny, A.; Brasseur, R.; Ruysschaert, J.-M.; Vandenbranden, M. *Biochem. Biophys. Res. Communications* **1991**, 175, 872-879.
32. Carr, C.M.; Kim, P.S. *Cell* **1993**, 73, 823-832.
33. Desmet, J.; de Mayer, M. and Lasters, I. The "dead-end elimination" theorem: a new approach to the side-chain packing problem. In K. Merz, Jr. and S. Legrand (eds), *The Protein Folding Problem and Tertiary Structure Prediction* **1994**.
34. Brasseur, R.; Ruysschaert, J.-M. *Biochem. J.* **1986**, 239, 1-11.
35. Brasseur, R. *J. Mol. Graphics* **1995**, 13, 312-322.
36. Rahman, M.; Brasseur, R. *J. Mol. Graphics* **1994**, 12, 212-218.
37. Weis, W.; Brown, J.H.; Cusack, S.; Paulson, J.C.; Skehel, J.J.; Wiley, D.C. *Nature* **1988**, 333, 426-431.
38. Deisenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. *J. Mol. Biol.* **1984**, 180, 385-398.
39. Deisenhofer, J.; Epp, O.; Miki, K.; Huber, R.; Michel, H. *Nature* **1985**, 318, 618-624.
40. Chang, C.-H.; El-Kabbani, O.; Tiede, D.; Norris, J.; Schiffer, M. *Biochemistry* **1991**, 30, 5352-5360.
41. Henderson, R.; Baldwin, J.M.; Ceska, T.A.; Zemlin, F.; Beckmann, E.; Downing, K.H. *J. Mol. Biol.* **1990**, 213, 899-929.
42. Weiss, M.S.; Wacker, T.; Weckesser, J.; Welte, W.; Schulz, G.E. *FEBS Lett.* **1990**, 267, 268-272.
43. Weiss, M.S.; Schulz, G.E. *J. Mol. Biol.* **1992**, 227, 493-509.

44. Schrag, J.D.; Cygler, M. *J. Mol. Biol.* **1993**, 230, 575-591.
45. Uppenberg, J.; Hansen, M.T.; Patkar, S.; Jones, T.A. *Structure* **1994**, 2, 293-308.
46. Derewenda, Z.S.; Derewenda, A.U.; Dodson, G.G. *J. Mol. Biol.* **1992**, 227, 818-839.
47. Bourne, Y.; Martinez, C.; Kerfelec, B.; Lombardo, D.; Chapus, C.; Cambillau, C. *J. Mol. Biol.* **1994**, 238, 709-732.
48. Noble, M.E.M.; Cleasby, A.; Johnson, L.N.; Egmond, M.R.; Frenken, L.G.J. *FEBS Lett.* **1993**, 331, 123-128.
49. Peitsch, M.C.; Boguski, M.S. *New Biol.* **1990**, 2, 197-206.
50. Lins, L.; Brasseur, R. *FASEB J.* **1995**, 9 535-540.
51. Brady, L.; Brzozowski, A. M.; Derewenda, Z. S.; Dodson, E.; Dodson, G.; Tolley, S.; Turkenburg, J. P.; Christiansen, L.; Huge-Jensen, B.; Norskov, L.; Thim, L.; Menge, U. *Nature* **1990**, 343, 767-770.
52. McLean, J.; Wion, K.; Drayna, D.; Fielding, C.; Lawn, R. *Nucleic Acids Research* **1986**, 14, 9397-9406.
53. Winkler, F.K.; D'Arcy, A.; Hunziker, W. *Nature* **1990**, 343, 771-774.