International Journal of Peptide Research and Therapeutics, Vol. 11, No. 4, December 2005 (© 2005), pp. 249–259
DOI: 10.1007/s10989-005-9393-1

Prediction of Cell-Penetrating Peptides

Mattias Hällbrink,[1] Kalle Kilk,[1] Anna Elmquist,[1] Pontus Lundberg,[1] Maria Lindgren,[1] Yang Jiang,[1] Margus Pooga,[1,2] Ursel Soomets,[1,3] and Ülo Langel[1,4]

(Accepted November 4, 2005)

Cell-penetrating peptides, CPPs, are used as delivery vectors for pharmacologically interesting substances, such as antisense oligonucleotides, proteins and peptides. We present a general principle for designing cell-penetrating peptides derived from naturally occurring proteins as well as from randomly generated polyamino acid sequences. Thereby, we introduce a novel pharmacological principle for identification of cell-penetrating peptides for which the applications can be numerous, including cellular transduction vectors and mimics of intracellular protein–protein interactions. The methods of identifying a CPP comprises assessing the averaged bulk property values of the defined sequence, and ensuring that they fall within the bulk property value interval obtained from the training set. Despite this simplistic approach, the search criteria proved useful for finding CPP properties in either proteins or random sequences. We have experimentally verified cell-penetrating properties of 10–20-mer peptides derived from naturally occurring proteins as well as from random poly-amino acids. We note that since CPPs can be found in part of the protein sequences that may govern protein interactions, it is possible to produce cell-penetrating protein agonists or antagonists.

KEY WORDS: Cellular uptake; cell-penetrating peptides.

INTRODUCTION

During the last 10 years, several peptides have been demonstrated to translocate across the plasma membrane of eukaryotic cells by a seemingly energy-independent or endocytotic pathway. These peptides have been used for trans-membrane delivery of hydrophilic macromolecules. Cellular delivery using cell-penetrating peptides, CPPs, offers several advantages over conventional techniques as it is efficient for a range of cell types and can be applied to cells *en masse*. Thus, CPPs might be used as delivery vectors for pharmacologically interesting substances, such as antisense oligonucleotides, proteins and peptides. For reviews see (Lindgren et al., 2000; Temsamani and Vidal 2004).

Little structural resemblance has been found between the different families of CPPs. However, as all CPPs have a net positive charge, one can assume that initially they should bind to negatively charged head groups of lipids or proteins in the plasma membrane. Here we present a method for finding CPPs in proteins or random peptide sequences. In many peptide quantitative structure activity relationship studies (QSAR), a set of dimensionless variables is used to describe a composite of the

[1] Department of Neurochemistry and Neurotoxicology, Stockholm University, S-106 91, Stockholm, Sweden.
[2] Estonian Biocenter, Tartu, Estonia.
[3] Department of Biochemistry, Tartu University, Tartu, Estonia.
[4] Correspondence should be addressed to: Ülo Langel, Department of Neurochemistry and Neurotoxicology, Stockholm University, S.Arrheniusväg 21A, S-106 91 Stockholm, Sweden. Tel: +46-8-161-793; Fax: +46-8-161-371; e-mail: ulo@neurochem.su.se

Abbreviations: GLP-1, glucagone like peptide-1; AT1AR, angiotensine 1A receptor; D2(long), dopamine 2 receptor (long); CGRP, calcitonine gene related protein; mGluR, metabotrobic glutamate receptor; hGalR1, human galanin receptor type 1.

physical characteristics of the amino acids. Sandberg and colleagues published 5 such descriptor sets, Z_1–Z_5; covering 87 coded and non-coded amino acids (Sandberg et al., 1998).

Here, using the expanded descriptor scales, bulk property values, ($Z\Sigma/n$), were assembled for 24 published cell-penetrating peptides and 17 non-penetrating analogues (the training set); and averaged over the total number of amino acids in the sequence. Here, the Z_3 value, mainly describing polarity, had the highest predicting power. Using the criteria outlined by our approach, about 95% of the known CPP sequences can be predicted.

Protein and random sequences where searched for sequences falling within the bulk property value ($Z\Sigma/n$) interval obtained from the training set. Using these search criteria we selected and tested several novel cell-penetrating sequences derived from natural proteins.

MATERIALS AND METHODS

Peptide Search Criteria

The averaged sums of various amino acid descriptors for published CPPs (Table I) were assembled. In order to further restrict the descriptor intervals, the averaged descriptors of a range of CPP analogues reported not to internalise (Table I) were used. As further controls, non-related sequences of peptide hormones (Table II) were used.

The most relevant descriptors from (Sandberg et al., 1998) were Z1, Z2 and Z3. We found that the bulk of the amino acid side chain, calculated as the number of heavy atoms (C, N, S and O) in the side chains of the amino acids, and net donated hydrogen bonds of the side chain, calculated as the accepted hydrogen bonds of the side chains subtracted from the donated hydrogen bonds, were necessary in order to predict CPPs. The value intervals with the most predictive values are shown in Table III.

Peptide Synthesis

Peptides were synthesised in a stepwise manner in a 0.1 mmol scale on a peptide synthesiser (Applied Biosystems model 431A, USA) using *t*-Boc strategy of solid-phase peptide synthesis. *tert*-

Table I. Cell-Penetrating Peptides and non-functional analogues

Functional CPPs:	
+	GWTLNSAGYLLGKINLKALAALAKKIL
+	RQIKIWFQNRRMKWKK
+	KLALKALKALKAALKLA
+	LLIILRRRIRKQAHAHSK
+	AGYLLGKINLKALAALAKKIL
+	GWTLNSAGYLLGKINLKALAALAKKLL
+	GWTLNPAGYLLGKINLKALAALAKKIL
+	GWTLNPPGYLLGKINLKALAALAKKIL

Table I. Continued

+	LNSAGYLLGKINLKALAALAKKIL
+	GWTLNSAGYLLGKLKALAALAKKIL
+	KLALKLALKALKAALK
+	KLALKLALKAWKAALKLA
+	KITLKLAIKAWKLALKAA
+	KALAKALAKLWKALAKAA
+	KALKKLLAKWAAAKALL
+	KLAAALLKKWKKLAAALL
+	NAKTRRHERRRKLAIER
+	LLIILRRPIRKQAHAHSK
+	LLIILRARIRKQAHAHSK
+	LLIILRRRIRKQAHAHSA
+	TRRNKRNRIQEQLNRK
+	GGRQIKIWFQNRRMKWKK
+	MGLGLHLLVLAAALQGAKKKRKV
+	RQIKIWFPNRRMKWKK
+	KMTRAQRRAAARRNRWTAR
+	RVIRVWFQNKRCKDKK
+	RKSSKPIMEKRRRAR
+	RQIKIWFQNRRMKWKKLRKKKKKH
+	VQAILRRNWNQYKIQ
+	KRPAATKKAGQAKKKKL
+	MDAQTRRRERRAEKQAQWKAAN
+	TAKTRYKARRAELIAERR
+	RQGAARVTSWLGRQLRIAGKRLEGRSK
+	RQGAARVTSWLGRQLRIAGKRLEGR
+	GAARVTSWLGRQLRIAGKRLEGRSK
+	RVTSWLGRQLRIAGKRLEGRSK
+	SWLGRQLRIAGKRLEGRSK
+	GRQLRIAGKRLEGRSK
+	KCRKKKRRQRRRKKLSECLKRIGDELDS
+	KCRKKKRRQRRRKKPVVHLTLRQAGDDFSR
+	AAVALLPAVLLALLAPVQRKRQKLMP
+	ALWKTLLKKVLKA
+	PKKKRKVALWKTLLKKVLKA
+	RQARRNRRRALWKTLLKKVLKA
+	RGGRLSYSRRRFSTSTGR
+	YGRKKRRQRRRSVYDFFVWL
+	YGRKKRRQRRRGTSSSSDELSWIIELLEK
–	GRKKRRQRRRPQ
–	RRRRRRRRR
–	RKKRRQRRR
–	GRKKRRQRRRPPC
–	GRKKRRQRRRC
–	GRKKRRQ
Non-functional CPP-analogues:	
+	KKLSECLKRIGDELDS
+	PVVHLTLRQAGDDFSR
+	EILLPNNYNAYESYKYPGMFIALSK
+	KKKQYTSIHHGVVEVD
+	GWTNLSAGYLLGPPPGFSPFR
+	QNLGNQWAVGHLM
+	RPPGFSPFR
+	LLKTTELLKTTELLKTTE
+	LNSAGYLLGKALAALAKKIL
+	LNSAGYLLGKLKALAALAK
+	GWTLNSAGYLLGKINLKAPAALAKKIL
–	LLKTTALLKTTALLKTTA
–	LRKKKKKH
–	RQIKIFFQNRRMKFKK
–	GWTLNSAGYLLGKFLPLILRKIVTAL
–	KLALKALKAALKLA

+, Denotes sequence correctly predicted within the category; –, denotes failed prediction.

Table II. Peptide Hormones, Predicted to be Non-CPPs

QNLGNQWAVGHLM
RPPGFSPFR
GWTNLSAGYLLGPPPGFSPFR
GWTLNSAGYLLGPHAI
HDEFERHAEGTFTSDVSSYLEGQAAKEFIAWLVKGR
WSYGLRPG
FVPIFTHSELQKIREKERNKGQ
AGCKNFFWKTFTSC
CYFQNCPRG
DFDMLRCMLGRVYRPCWQV

Table III. Averaged Bulk Property Value Intervals of Cell-Penetrating Peptides

Value	Low	High
Bulk	3.2	5.9
$Z_1\Sigma/n$	−1.25	1.92
$Z_2\Sigma/n$	−1.22	1.29
$Z_3\Sigma/n$	−0.5	−1.94
$Z_{\Sigma hdbn}$	0.28	2

Butyloxycarbonyl amino acids (Neosystem, Strasbourg, France) were coupled as hydroxybenzotriazole (HOBt) esters to a *p*-methylbenzylhydrylamine (MBHA) resin (Neosystem, Strasbourg, France) to obtain C-terminally amidated peptide. Fluoresceincarboxylic acid was coupled manually to the N-terminus by adding a five-fold excess of HOBt and *o*-benzotriazole-1-yl-*N*, *N*, *N*′, *N*′-tetramethyluronium tetrafluorborate (TBTU) activated fluorescein (Chemicon, Stockholm, Sweden) in DMF to the peptidyl-resin. The peptide was finally cleaved from the resin with liquid HF at 0°C for 30 min in the presence of *p*-cresol. The purity of the peptides was >98% as demonstrated by HPLC on an analytical Nucleosil 120-3 C-18 RP-HPLC column (0.4 (10 cm) and the molecular mass was obtained by using a MALDI-TOF mass-spectrometer (Voyager DE-STR, Applied Biosystems, USA), and the calculated values were obtained in each case.

Cell Culture

Mouse N2A and Human SHSY-5Y neuroblastoma cells were grown in 10 cm Petri dishes in Dulbecco's Modified Eagle's Media (DMEM) supplemented with 10% foetal calf serum (FCS), 2 mM L-glutamine, 100 U/ml penicillin and 0.1 mg/ml streptomycin at 37°C in a 5% CO_2 atmosphere.

Bowes melanoma cells (American Type Culture Collection CRL-9607) were cultivated in Minimal Essential Medium (MEM, Life technologies, Stockholm, Sweden) with glutamax supplemented with 10% foetal bovine serum, penicillin–streptomycin, non-essential amino acids and (1 mM) sodium pyruvate.

Cellular Internalisation Microscopy Assay of Fluoresceinyl-Labelled Peptides

The cells used for internalisation assay were seeded out on round glass coverslips in 24-well plates (approximately 60,000 cells/well). One day post seeding, the cells were semi confluent, and the medium was changed to serum-free medium. The fluorescein-labelled peptides were added, with a final concentration of 5 μM. After 60 min of incubation at 37°C, the cells were washed 3 times with 1 ml of Hepes–Krebbs–Ringer (HKR) buffer. The cell nuclei were stained with Hoechst 33258 (0.5 μg/ml, Molecular Probes, Holland). The fluorescence was examined using UltraView ERS confocal live cell imager (PerkinElmer Ltd., Upplands Väsby, Sweden) connected to a Axiovert 200 (Zeiss, Göttingen, Germany).

Quantitative Uptake

Human SH-SY5Y and mouse N2A neuroblastoma cells from ATCC were cultured in DMEM supplemented with NEAA, sodium pyruvate and 10% FBS. Bowes human melanoma cells (American Type Culture Collection CRL-9607) were cultured in Eaglés Minimal Essential Medium with Glutamax-I (Life Technologies, Gaithersburg, MD) supplemented with 10% foetal calf serum, 1% non-essential amino acids, 1% sodium pyruvate, 100 (g/ml streptomycin, 100 U/ml penicillin. 100,000 cells/ml/well were seeded 2 days before experiments onto a 12-well-plate. The cells were washed by 2 (1 ml HKR supplemented with 1 g/l glucose (HKRg) then exposed to 300 μl of 5 or 20 μM peptide in HKRg at 37°C for 30 min.

After the indicated time-points, the peptide treated cells were washed and lightly trypsinated using 200 μl trypsin/EDTA (0.025% T/E in HKR) at 37° degrees for 1 min and subsequently washed thoroughly and lysed in 250 μl 0.1% Triton X-100 in HKR, at 4° degrees for 10 min. The cell lysates were transferred to a black plate for fluorescence readout at 492/520 nm. The samples were compared to the fluorescence of the added amount of peptide and normalised against total protein amount.

Determination of Acute Membrane Disturbance

Lactate dehydrogenase leakage was measured as a determination of acute membrane disturbance. The sampling was done as a multiplexing from assays performed under conditions as defined above. After indicated time and peptide or conjugate concentration, a 100 μl sample/well in triplicate were collected and detected by CytoTox-ONE assay (Promega Corporation, Madison, WI) according to manufactures instructions. The data is presented as a percentage of the maximum leakage induced by treating the cells with 0.1% Triton X-100 in HKR.

RESULTS

First we decided to search for CPP sequences in G-protein coupled receptors. Three hundred and seventy-nine human 7TM receptor sequences were downloaded from the Swissprot database. The sequences were searched for CPPs of the indicated length (8, 12 and 17 residues long). Here, a search window size of 17 residues produced the most hits.

The 379 human 7TM receptor sequences contain 16,000 12-mer and 17,000 18-mer CPP sequences. The position of the start of the CPP in the protein, divided by the total length of the protein, is plotted against the fraction of CPPs occurring at that

position (see Fig. 1). The four peaks in evidence correspond to the four intracellular parts of the 7TM receptors, with the largest corresponding to the third intracellular loop (IC3).

Many amino acid sequences, both, in naturally occurring i.e., native proteins and computer-generated random amino acid sequences, contain cell-penetrating peptides of different length according to our search. In 99 human transcription factor sequences downloaded from Swissprot, 4500 12-mer and 4100 18-mers with CPP properties are predicted.

The surprising number of possible sequences is in part due to clustering around short sequences that function as 'transport motors'. A similar behaviour can be seen in searches in random generated sequences. Due to enormous volume of these data they are not presented here. However, they are listed as a part of a patent application (Hällbrink et al., 2002).

Analysis of CPP search among the previously reported cell-penetrating peptides yields controversial results. Most of the reported CPPs fall into the CPP class according to our search criteria. However, e.g., the tat peptide does not. Indeed, including the tat peptides in the training set yields descriptor intervals with little prediction power. Overall prediction success (calculated as correctly predicted positive CPPs subtracted by the percent of wrongly predicted negative CPPs) is 90%.

We experimentally verified the cell-penetrating property of sequences derived from 7 different proteins: the GLP-1, AT1, D2(long), mGlu and galanin receptors; the amyloid precursor protein and presenilin-1 (Table IV). Fluorescence micrographs showed that the peptides did indeed penetrate to the cellular interior of human Bowes melanoma cells (Fig. 2), indicating that the search criteria for cell-penetrating peptides in naturally occurring proteins have a predictive character.

We also searched random peptide sequences to produce novel CPPs. As examples we present two 16-residue sequences: E162 and E165 (Table IV).

In an attempt to compare the uptake efficiencies of the CPPs, quantitative uptake measurements were performed (Fig. 3). Here, the designed CPPs, E162 and E165 were by far the most effective, reaching intracellular concentrations approx 5–10 times that of the positive control peptide penetratin. In contrast, the peptides M448, M630 and M700 only reached concentrations approx 5 times less than the positive control. However, this corresponds to intracellular concentrations that are above the extracellular concentration, thus indicating the diagnostic accumulation seen for CPPs. Possibly, a partial explanation for the varied uptake of the peptides could be differences in degradation as been pointed out in (Hällbrink et al., 2004). In order to further characterise the uptake, the time course of CPP internalisation into Bowes' melanoma cells was investigated (Fig. 4a and b).

The membrane disturbance of the peptides was investigated by examining the concentration and time

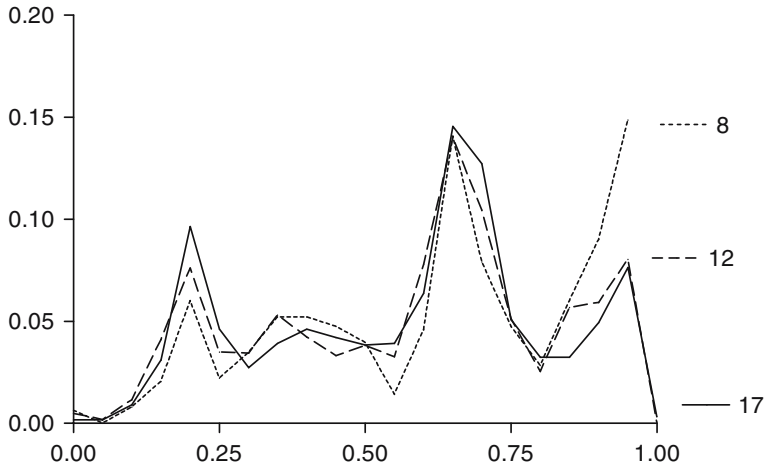

Fig. 1. Position of CPPs in human 7TM receptors. Human 7TM sequences were searched for CPPs of indicated length (8, 12 and 17 aa long). The position of the CPP start in the protein was divided by the total length of the protein, plotted against the frequency of occurrence of CPPs at that position. The four peaks in evidence correspond to the four intracellular parts of the 7TM receptors, with the largest peak corresponding to the third internal loop (IC3). It can be noted that the CPP functionality seems to correlate well, both with the topology of the 7TM receptors, as well as with the proposed G-protein activation sites.

Table IV. Protein-Derived Cell-Penetrating Peptides and Controls

Short name	Protein	Sequence
M 511	r/m AT1R(304–318)	FLGKKFKKYFLQLLK-amide
G53-4	rGLP-1R 3rd intracellular loop analogue	FLIFIRVICIVIAKLKANLMCKT-amide
APP521	hAPP(521–537)	KKAAQIRSQVMTHLRVI-amide
M591	hD2R(213–228), long	YIVLRRRRKRVNTKRS-amide
M593	hD2R(360–370), long	RRKLSQQKEKK-amide
M630	hCGRP(391–405)	VQAILRRNWNQYKIQ-amide
E162	Designed	KTVLLRKLLKLLVRKI-amide
E165	Designed	LLKKRKVVRLIKFLLK-amide
G55-9	r/m GluR1(864–882) 4th intracellular loop	KLPCRSNTFLNIFRRKKPG-amide
M867	mGluR1 (691–706), rat 3rd intracellular loop	KKICTRKPRFMSAWAQ-amide
–	Penetratin	RQIKIWFQNRRMKWKK

Abbreviations: R, receptor; r, rat; h, human; m, mouse.

dependence of LDH leakage from Bowes' melanoma cells (Fig. 5). Here, only 2 peptides, E162 and M511 caused leakage above baseline. This can be explained by the comparatively high amphipatic moment of these peptides.

DISCUSSION

Penetratin, or antennapedia (43–58), is a 16 amino acid long peptide corresponding to the third helix of the *Antennapedia* homeodomain deprived of its N-terminal glutamate (Derossi et al., 1996). The peptide and analogues have been used for transmembrane delivery of a diversity of hydrophilic macromolecules and has therefore been proposed to be a universal intercellular delivery vector (Derossi et al., 1998). Purely synthetic or chimeric peptides have also been designed as reviewed in (Lindgren et al., 2000).

Transportan, a non-natural peptide, is able to deliver an antibody molecule with molecular mass of about 150 kDa over the plasma membrane (Pooga et al., 2001) although transportan itself is only a 3 kDa peptide. Transportan and penetratin were demonstrated to deliver a non-natural DNA analogue, PNA (peptide nucleic acid), into the cytoplasm and nuclei of cells in culture as well as *in vivo* (Pooga et al., 1998).

Another group of peptides, pepducines (Covic et al., 2002; Kuliopulos and Covic 2003), derived from G-protein coupled receptors, was designed by attachment of a hydrophobic moiety to peptides derived from the third intracellular loop of a receptor. These peptides function as cell penetrating, full agonist or antagonist of receptor G-protein signalling. The pepducines are membrane-tethered chimeric peptides and require the presence of their cognate receptor for activity and are highly selective for the receptor type.

Structurally, CPPs can be differentiated into 3 classes: the first are peptides with low amphipathicity where the charge contribution originates mostly from arginine residues. Examples of this class are penetratin and tat. The second class is peptides with a high degree of amphipaticity, where the charge contribution originates mainly from lysine residues. Examples of this class are MAP and transportan. In the third class the charged and hydrophobic residues are separated lengthwise in the chain, exemplified by pVEC and MPG. There is some evidence in the literature that these peptide classes behave differently, especially concerning endocytotic uptake.

Although their astonishing transport capability has put CPPs in focus of scientific interest for the last 10 years, the mechanism(s) of translocation for the different CPPs are unknown. For instance, it is still not known whether any particular secondary structure has to be induced in order to allow a translocation, involving a concomitant transient membrane destabilisation. It is clear, however, that the molecular details of the peptide–membrane interactions must be of fundamental importance for the translocation process. Both, α-helical and β-sheet structures have been suggested as responsible for the CPP properties; however, the exact mechanisms remain to be proven.

The mechanism and requirements for internalisation have been studied on interactions between

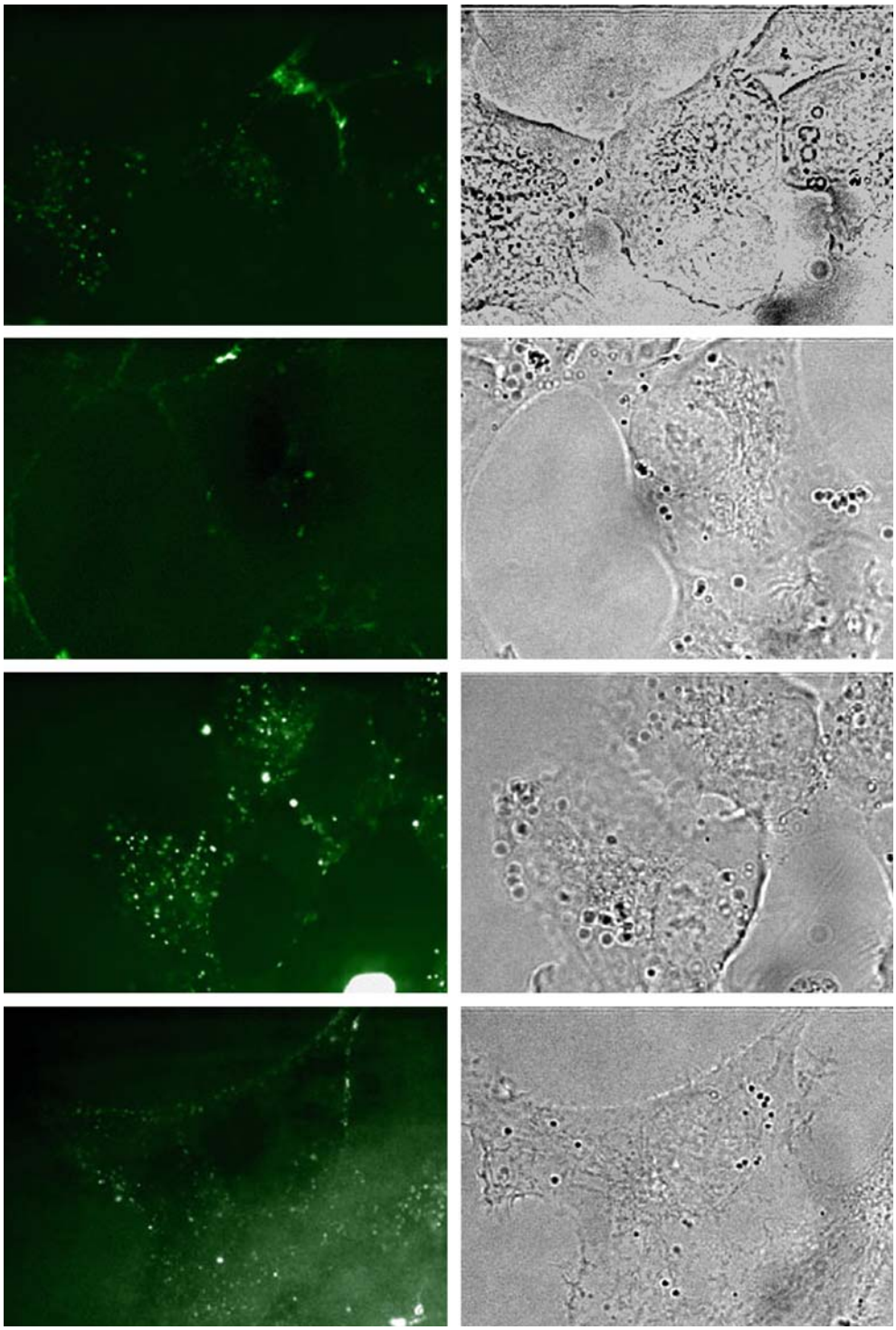

Fig. 2. Fluorescence microscopy fluorescein labelled peptides. Fluorescence microscopy pictures of fluorescein labelled peptides after incubation in 5 μM for 30 min in 37°C in live cells. The left column shows the confocal fluorescence of the peptide, and the right column the corresponding bright field pictures. (a) Fluorescein labelled G55-9 in Bowes cells. (b) Fluorescein labelled E162 in Bowes cells. (c) Fluorescein labelled M591 in Bowes cells. (d) Fluorescein labelled M593 in Bowes cells. (e) Fluorescein labelled App521 in Bowes cells. (f) Fluorescein labelled M630 in Bowes cells. (g) Fluorescein labelled E165 in Bowes cells. (h) Fluorescein labelled G53-4 in Bowes cells. (i) Fluorescein labelled M867 in Bowes cells. (j) Fluorescein labelled scr M511 in Bowes cells. (k) Fluorescein labelled M511 in Bowes cells. (l) Positive control: Fluorescein labeled penetratin in Bowes cells.

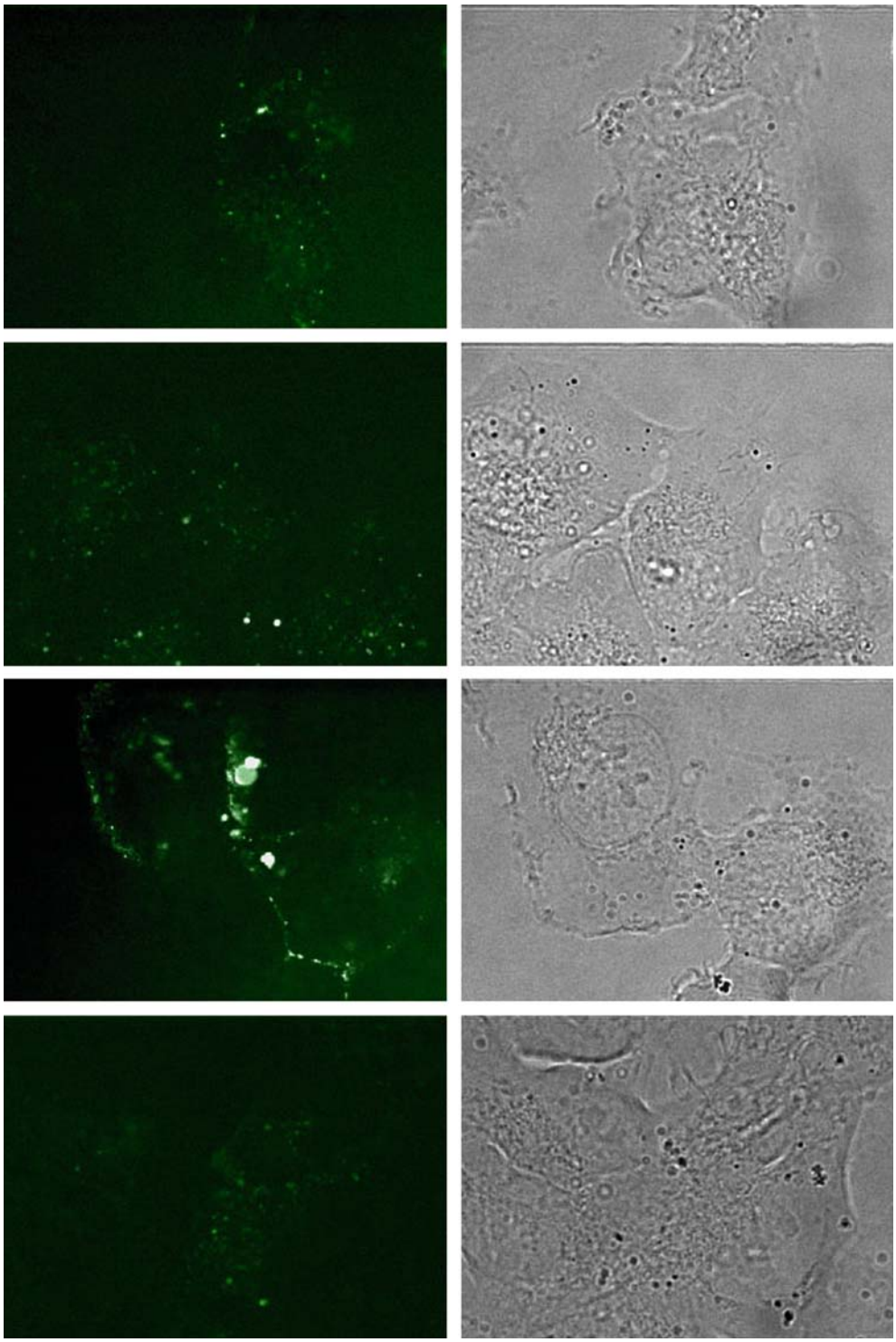

Fig. 2. Continued

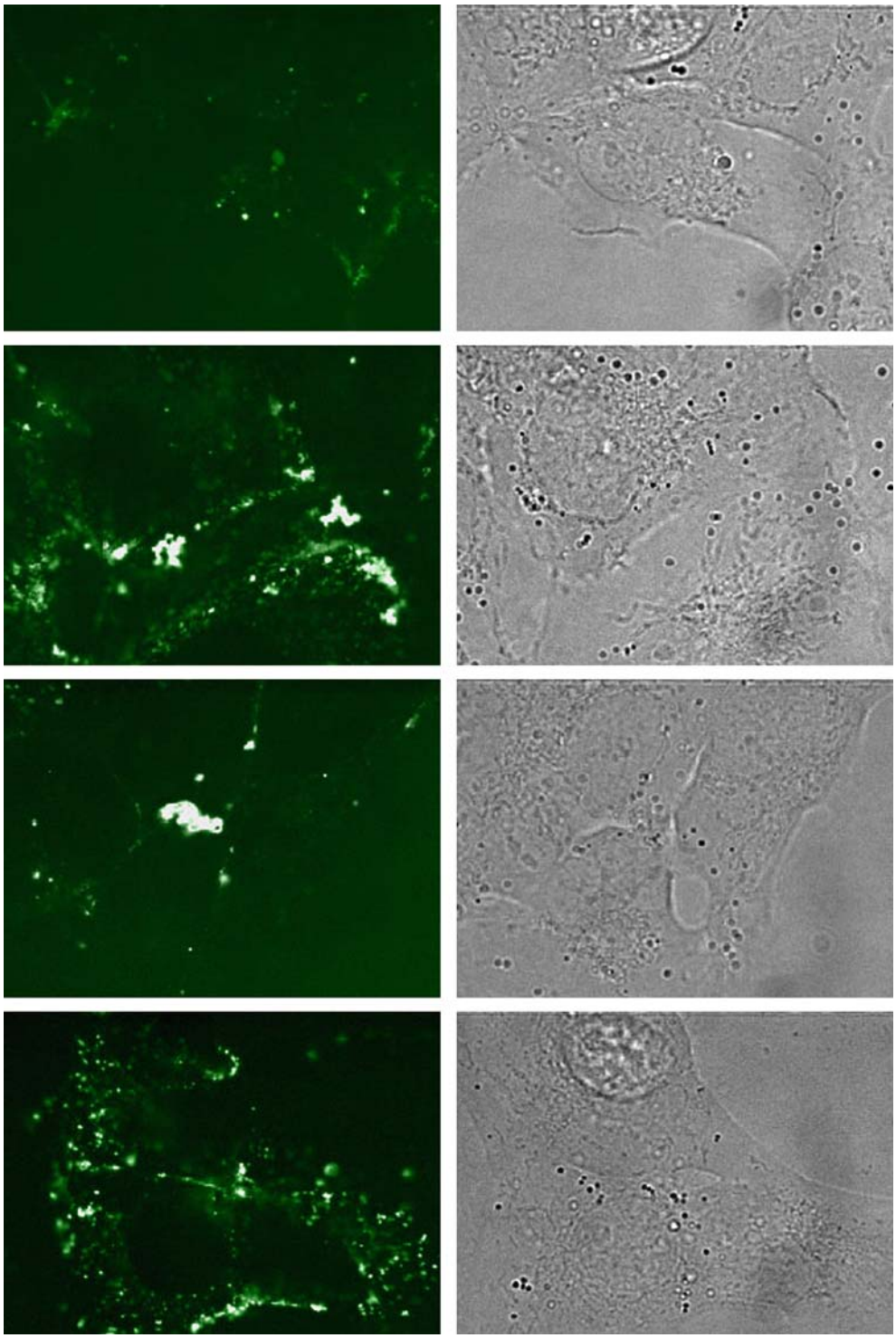

Fig. 2. Continued

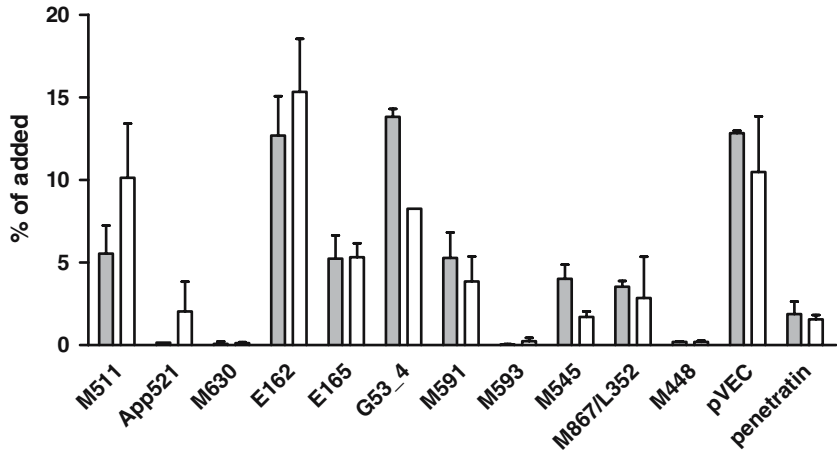

Fig. 3. Quantitative uptake of fluorescein labelled peptides. Human SH-SY5Y (filled bars) and mouse N2A neuroblastoma cells (empty bars) were seeded 2 days before experiments onto a 12-well-plate and washed (2 (1 ml HKRg) then exposed at 300 μl of 5 μM peptide in HKRg at 37°C for 30 min. The peptide treated cells were washed and lightly trypsinated and lysed in 0.1% Triton X-100 The cell lysates were examined for fluorescence readout at 492/520 nm. The samples were compared to the fluorescence of the added amount of peptide and the total protein amount.

amphipathic α-helical peptides and lipid (bi)layers. The results of these studies often suggest tryptophan to be crucial for internalisation of peptides, but although aromatic amino acids may be preferred in CPP sequences, they are not absolutely necessary for cell penetration.

The mechanism by which CPPs enter cells is not clear. However, it has been proposed that CPPs can be in equilibrium over the plasma membrane (Scheller et al., 2000). Most studies on CPPs show that the intracellular concentration of the labelled peptide is significantly higher than in the extracellular media. Recently, we showed that the main mechanism by which CPPs, or rather the cargo attached to the CPPs, accumulate in the cellular interior is by the proteolytic processing of the CPP-cargo conjugate into membrane impermeable products (Hällbrink et al., 2004). Thus, CPP-cargo conjugates can be viewed as a 'pro-drug.' Thus a CPP prediction method should ideally take not only uptake, but also processing of the peptide into account.

Recently, studies focusing on the CPP tat (48–60) have pointed out that several published observations concerning cell uptake may well be flawed due to methodological drawbacks, particularly in the case of the tat peptide (Richard et al., 2003). These studies have pointed out the importance of using methods that avoid bias of measurements of uptake by plasma membrane bound peptides. Thus when searching the literature for CPPs sequences, care must be taken to ensure that they have been evaluated on similar grounds.

Futaki and colleagues have isolated CPPs from proteins on the basis of charge (Futaki et al., 2001). Specifically, Arg rich domains from RNA binding proteins were isolated, in which a correlation with the uptake efficiency and the number of Arg residues were noted. Interestingly, peptides that contain less than 5 Arg residues were much less active. However, this method does not predict CPPs lacking sufficient Arg residues, such as MAP or transportan.

Apart from the cell-penetration capability, little correlation of structure or behaviour has been found between CPPs. Until now, CPPs have thus not been designed in a rational manner, but have been found serendipitously. However, the sequences of CPPs published so far have a positive net charge as a common feature, giving a starting point for the prediction of CPP functionality in a given peptide sequence. Clearly, though, all sequences with a positive net charge cannot be cell penetrating, indicating that further restrictions are needed to select CPPs with any certainty.

Reasoning that the CPP property of the peptide is a bulk property, i.e., that it depends not on specific motives but rather on the net property of the amino acid residues in the peptide, we assembled the averaged sums of various amino acid descriptors for published CPPs (Table II). In order to further restrict the descriptor intervals, we used the averaged

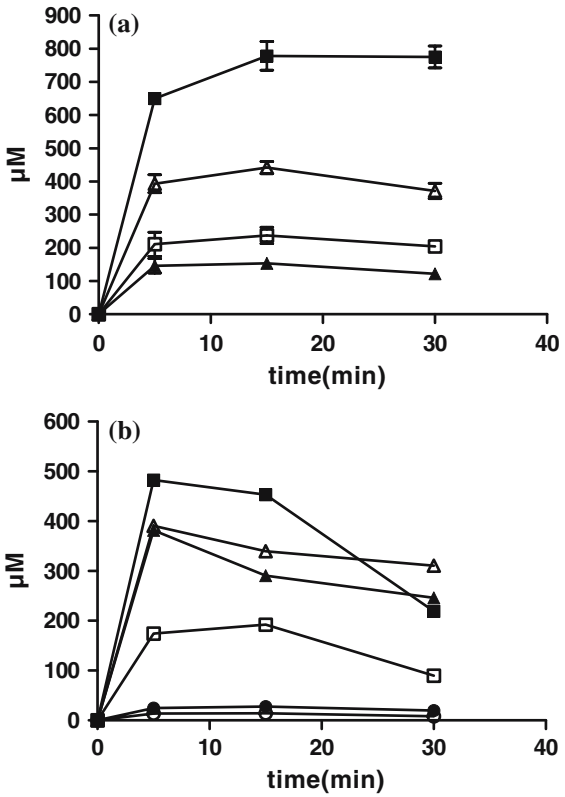

Fig. 4. Time course of CPP uptake in Bowe's melanoma cells. Cells were seeded 2 days before experiments onto a 12-well-plate and washed (2 (1 ml HKRg) then exposed at 300 µl of 5 µM peptide in HKRg at 37°C for 30 min. The peptide treated cells were washed and lightly trypsinated and lysed in 0.1% Triton X-100 The cell lysates were examined for fluorescence readout at 492/520 nm. The samples were compared to the fluorescence of the added amount of peptide and the total protein amount. (a) Peptides internalising using an apparent first order kinetic type. Filled box – E162; empty triangle – E165; empty box – M591; filled triangle – G55-9. (b) Peptides internalizing not using an apparent first order kinetic type. Filled box – G53-4; empty triangle – M867; empty box – M630; filled triangle – penetratin; Filled circle – APP521; empty circle – M593.

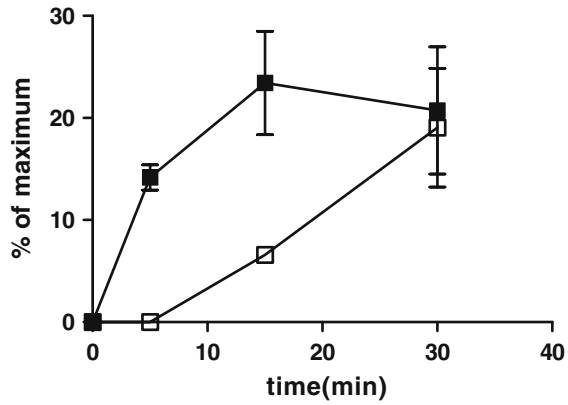

Fig. 5. The membrane disturbance of the peptides was investigated by examining the concentration and time dependence of LDH leakage from Bowe's melanoma cells. Only M511 (filled boxes) and E162 (empty boxes) caused leakage above baseline. The data is presented as a percentage of the maximum leakage induced by treating the cells with 0.1% Triton X-100 in HKR.

descriptors of a range of CPP analogues reported not to internalise (Table II). As controls we used non-related sequences of peptide hormones (Table III). Despite this simplistic approach, the search criteria proved surprisingly useful for finding CPP properties in either proteins or random generated sequences.

Searching either a random- or natural protein sequence, sequences corresponding to CPPs appear clustered in blocks throughout the sequences. This behaviour is due to the existence of 'transport motors,' i.e., shorter sequences with CPP characteristics, in the search window. This has already been demonstrated for several CPPs, e.g., penetratin (Fischer et al., 2000) and transportan (Soomets et al., 2000), using systematic deletions of the sequence. Here we analyse the CPP search results for only a few classes of natural proteins and random amino acid sequences. In particular we were interested in G-protein coupled receptors. If CPPs can be found in part of the protein sequences that govern protein interactions, it is possible to produce cell-penetrating protein agonists or antagonists. Therefore we decided to test potential CPPs derived from CGRP receptor loop iC4, sequence 391–405 (M630); human D2 receptor long version, M591 and M593; mGlu receptor, hGalR1.

From our data, it can be noted that the CPP functionality seems to correlate well, both with the topology of the 7TM receptors, as well as the proposed G-protein activation sites. In a positional plot for CPPs in 7TM receptors, four peaks appear corresponding to the four intracellular parts of the receptors. The largest peak coincides with the position of the third intracellular loop (Fig. 1).

The internalisation experiments demonstrate that APP (521–536), which is derived from the extracellular part of the Alzheimer's disease related amyloid precursor protein (APP), has cell-penetrating abilities (Fig. 2), localising mainly in the cytosol and plasma membrane. This could prove as a putative pathway by which the secretory amyloid precursor protein (sAPP) is internalised. This fact is interesting since this fragment has been shown to protect neurons against hypoglycemic damage and

glutamate neurotoxicity thus acting as neuroprotective agent.

The time course of uptake of the new CPPs indicate that the kinetic profile of the peptides fall within 2 different groups. One, in which the peptides seem to follow first order (equilibrium) kinetics (Fig. 4a). The second, where the intracellular fluorescence quickly reaches a maximum, followed by a steady decrease in signal (Fig. 4b). These different behaviours can be attributed to different degradation propensities of the peptides, both inside and outside the cell. However, the main scope of this paper is not to dissect the various pathways leading to the CPP phenomenon and work to fully characterise the transport and degradation pathways are under way.

Interestingly, only 2 of the examined peptides caused any measurable plasma membrane disturbance (Fig. 5). This indicates that there must exist a large number of useful, non-toxic, CPPs still to be characterised.

Taken together, the uptake and leakage data suggest that several of the peptides examined in this work would be suitable for intracellular delivery of pharmacologically interesting substances. The most interesting peptide, from a practical point-of-view, is the designed E165, which have the highest delivery capacity coupled with negligible plasma membrane toxicity.

In conclusion, we present a simple method for CPP prediction based on bulk properties of the constituent amino acids. It is likely that the mechanism of CPP uptake is somewhat complicated to explain with the bulk properties of the constituents only, in particular concerning peptide degradation. However, this method of CPP prediction can be successfully applied to find novel CPPs, as well as the peptides with non-CPP properties. Peptides containing an active CPP fused to a translocation inhibitory domain via a protease sensitive motif, can be used to deliver cargo selectively *in vivo* (Hällbrink et al., 2002; Jiang et al., 2004). This approach can be applied for design of different drug precursors for medical applications in future.

We hypothesise that the natural protein derived CPP sequences may characterise some functionality of proteins perhaps in the cellular distribution, folding or shuttling, similar to the tat and antennapedia proteins. This tempting hypothesis, however, remains to be proven.

ACKNOWLEDGMENTS

We are grateful to Mr. Joel Ohlsson for the assistance with peptide synthesis and Dr. Carl-Johan Dalsgaard for fruitful discussions. This work was supported by grants from European Community (QLK3-CT-2002-01989), Swedish Research Council VR-NT, CePeP II AB, Stockholm, Sweden. The presented results are protected by the international patent application with international publication number WO03/106491 A2.

REFERENCES

Covic, L., Gresser, A. L., Talavera, J., Swift, S. and Kuliopulos, A.: 2002, *Proc. Natl. Acad. Sci. USA* 99, 643–648.

Derossi, D., Calvet, S., Trembleau, A., Brunissen, A., Chassaing, G. and Prochiantz, A.: 1996, *J. Biol. Chem.* 271, 18188–18193.

Derossi, D., Chassaing, G. and Prochiantz, A.: 1998, *Trends Cell Biol.* 8, 84–87.

Fischer, P. M., Zhelev, N. Z., Wang, S., Melville, J. E., Fahraeus, R. and Lane, D. P.: 2000, *J. Pept. Res.* 55, 163–172.

Futaki, S., Suzuki, T., Ohashi, W., Yagami, T., Tanaka, S., Ueda, K. and Sugiura, Y.: 2001, *J. Biol. Chem.* 276, 5836–5840.

Hällbrink, M., Oehlke, J., Papsdorf, G. and Bienert, M.: 2004, *Biochim. Biophys. Acta* 1667, 222–228.

Hällbrink, M., Kilk, K., Lundberg, P., Soomets, U., Elmquist, A., Zorko, M., Budihna, M., Östenson, C.-G., Gräslund, A., Eriksson, G., Lindgren, M., EL-Andaloussi, S., Meikas, A., Valkna, A., Kogerman, P., Metsis, M., Pooga, M. and Langel, Ü.: 2002, Cell-penetrating peptides. PCT WO 2003106491.

Jiang, T., Olson, E. S., Nguyen, Q. T., Roy, M., Jennings, P. A. and Tsien, R. Y.: 2004, *Proc. Natl. Acad. Sci. USA* 101, 17867–17872.

Kuliopulos, A. and Covic, L.: 2003, *Life Sci.* 74, 255–262.

Lindgren, M., Hällbrink, M., Prochiantz, A. and Langel, Ü.: 2000, *Trends Pharmacol. Sci.* 21, 99–103.

Pooga, M., Kut, C., Kihlmark, M., Raid, R., Hällbrink, M., Fernaeus, S., Hallberg, E., Land, T. and Langel, Ü.: 2001, *FASEB J.* 15(8), 1451–1453.

Pooga, M., Soomets, U., Hällbrink, M., Valkna, A., Saar, K., Rezaei, K., Kahl, U., Hao, J.-X., Xu, X.-J., Wiesenfeld-Hallin, Z., Hökfelt, T., Bartfai, T. and Langel, Ü.: 1998, *Nat. Biotechnol.* 16, 857–861.

Richard, J. P., Melikov, K., Vives, E., Ramos, C., Verbeure, B., Gait, M. J., Chernomordik, L. V. and Lebleu, B.: 2003, *J. Biol. Chem.* 278, 585–590.

Sandberg, M., Eriksson, L., Jonsson, J., Sjostrom, M. and Wold, S.: 1998, *J. Med. Chem.* 41, 2481–2491.

Scheller, A., Wiesner, B., Melzig, M., Bienert, M. and Oehlke, J.: 2000, *Eur. J. Biochem.* 267, 6043–6050.

Soomets, U., Lindgren, M., Gallet, X., Hällbrink, M., Elmquist, A., Balaspiri, L., Zorko, M., Pooga, M., Brasseur, R. and Langel, Ü.: 2000, *Biochim. Biophys. Acta* 1467, 165–176.

Temsamani, J. and Vidal, P.: 2004, *Drug Discov. Today* 9, 1012–1019.