CHAPTER 1

INTRODUCTION

1.1. Statement and significance of the problem

Nowadays, many proteins and peptides have been widely used as therapeutic agents. However, several problems such as low level of absorption and easy to be degraded by digestive enzymes become great challenges for the improvement of bioavailability and efficiency of these therapeutic agents (Amidon and Lee, 1994). Due to these problems, most therapeutic proteins and peptides are administered by parenteral route. Peptide drugs are usually indicated for chronic conditions such as diabetes mellitus. The use of injections on a daily basis of the peptide drugs during long-term treatment has obvious drawbacks. Establishing an oral delivery system for peptides and protein drugs is of great importance because parenteral administration results in poor patient compliance during chronic treatment resulting in limited clinical utility (Saalik et al., 2004). Attempts to improve the oral bioavailability of peptide drugs ranged from the physicochemical properties modifications of peptide molecules to the application of the functional excipients and drug delivery systems such as microemulsions, liposomes and niosomes (Pichayakorn et al., 2003). An enhancement of oral bioavailability of encapsulated insulin in nanoparticles and nanocapsules has been demonstrated in animal studies. However, the high variability of insulin concentration delivered to the blood and the lack of information allowing to adapt these results to humans resulting in no further industrial development and clinical trials (Rieux et al., 2006). One of the important therapeutic peptides is calcitonin (CT). CT is a 32 amino acid peptide hormone produced by the *para-*

follicular cells of the thyroid gland in mammals and the ultimo-branchial glands of birds and fish. The amino acid sequence of CT varies between species but eight residues (near the peptide's N-terminal disulphide bridge) are strictly conserved and integral to the hormone's function. Teleost which is the CT from salmon and eel is up to forty times more potent than that of the mammalians, due to a highly flexible α-helical peptide structure which promotes optimal binding to the receptor. Four forms of CT are used clinically, namely, synthetic human CT, synthetic salmon CT, natural porcine CT, and a synthetic analogue of eel calcitonin. Currently, CT is administered parenterally or nasally (Tanko et al., 2004).

CPPs (cell-penetrating peptides) are short peptides of less than 30 amino acids. These CPPs are able to penetrate cell membranes and translocate different cargoes such as peptide drugs into cells. The only common feature of these peptides appears to be that they are amphipathic and highly positive charged. The mechanism of cell translocation of CPPs is not known but it is apparently receptor and energy independent, although in certain cases, translocation can be partially mediated by endocytosis. Most CPPs are inerted or have very limited side effects. Their penetration into cells is rapid and initially first-order, with half-life from 5 to 20 mins (Richard et al., 2003). The first notification that CPPs could contain sequences responsible for their translocation across cellular membranes originated from the observation that living cells internalized an 86 amino acid long fragment from the HIV-1 Tat protein, activating HIV-LTR driven transcription. It has been discovered that the 60 amino acid homeodomain of the *Antennapedia* protein of *Drosophila* spp. is also able to translocate across cell membranes (Shaw et al., 2007; Dennison et al., 2007; Pujals and Giralt, 2007). Thus, the penetrating and translocating properties of

CPPs can be applied for the development of novel drug delivery systems. The applicative therapeutic potential of CPPs is its attachment with biologically active cargoes such as peptide drugs and being translocated it into cells. Cargoes that are successfully internalized by CPPs range from small molecules to proteins and supramolecular particles. Cargo can be attached to peptide by different methods such as covalent bonding using bifunctional cross linker and recombinant DNA technique. CPPs and peptide cargo can also be synthesized or expressed in tandem as the fusion protein. Only few studies on CPPs-cargoes as novel drug delivery systems have been investigated (Zhou, 1994). Hence, the study in this field still has high potential of applications and innovations in pharmaceutical fields.

Poliovirus (PV) is a human enterovirus that typically acquired through ingestion of contaminated matter and enters tissues through gastrointestinal mucosal surfaces. PV contains four virus-specific protein subunits (VP1, VP2, VP3, and VP4). These polypeptides are constructed into viral capsid and associated in viral-receptor binding mechanism and/or cell entry (Grant et al., 1994). PV can infect follicular-associated epithelial cells as well as cells in the Peyer's patches and tonsils that lie under the epithelial surface. The microfold cells (M-cells), a part of the gastrointestinal epithelium, express CD155 or poliovirus receptor (PVR) on both their apical and basolateral surfaces. M-cells transport poliovirus directly from the gut to the lymphoid follicles of Peyer's patches. Infection is aided by the fact that these epithelial cells have little IgA or mucus secretions, exempting the virus from one of the first lines of host defense. PV infection is initiated by the attachment of capsid protein to specific receptor molecules at the gastrointestinal (GI) tract epithelium (He

et al., 2000). Hence, these capsid proteins might be applicable for ligand-receptor mediated oral drug delivery.

1.2. Objective

The objective of this study was to develop a high efficiency orally active calcitonin delivery system using VP sequence of poliovirus and Tat peptide as a CPP. VP was expected to be a poliovirus receptor binding ligand in gastrointestinal tract and CPP will enhance the translocation efficiency of the delivery system containing therapeutic cargoes. The biotechnology was applied in the delivery system development using green fluorescent protein (GFP) as a model and reporter protein. The application of nanovesicles in the form of liposomes and niosomes both elastic and non-elastic were used to develop a high efficiency orally active calcitonin delivery system..

1.3. Scope of study

This study was divided into 6 steps which were the followings:

1. Construction of GFP, Tat/GFP, VP/GFP and VP/Tat/GFP fusion protein expression plasmid

    1.1) Primer design

    1.2) Amplification of inserted genes

    1.3) Preparation of pET28a(+) expression vector and inserted fragment

    1.4) Ligation

2. Expression and purification of GFP, Tat/GFP, VP/GFP and VP/Tat/GFP fusion protein

2.1) Transformation

2.2) Protein expression

2.3) Purification and analysis of expressed proteins

3. Cellular uptake of GFP, Tat/GFP, VP/GFP and VP/Tat/GFP fusion protein

Cellular uptake experiment was performed in 2 GI-lining cell lines, human adenocarcinoma (HT-29) and human mouth epidermal carcinoma (KB).

4. Cellular uptake efficiency enhancement strategy

4.1) Entrapment in nanovesicles

4.2) Tat/GFP, VP/GFP and VP/Tat/GFP mixture

5. *In vitro* calcitonin activity of Tat/sCT, VP/sCT and VP/Tat/sCT mixtures

5.1) Mixtures preparation

5.2) Physical properties determination of the mixtures

- Sizes and zeta potential determination
- Differential scanning calorimetry (DSC)
- Fourier transform infrared spectroscopy (FT-IR)

6. *In vivo* calcitonin activity of Tat/sCT, VP/sCT and VP/Tat/sCT mixtures

1.4. Literature reviews

1.4.1. Green fluorescent protein (GFP)

1.4.1.1. Introduction

A. Historical perspective

Bioluminescence is the process by which visible light is emitted by an organism as a result of a chemical reaction. The reaction involves the oxidation of

a substrate such as the luciferin by enzyme luciferase. Oxygen is usually the oxidant. Bioluminescent organisms are found in a variety of environments. Common examples are insects, fish, squid, sea cacti, sea pansies, clam, shrimp, and jellyfish. The bioluminescent systems in these organisms are not all evolutionarily conserved, and the genes coding for the proteins involved in bioluminescence are not homologous (Hastings and Morin, 1998). The emitted light commonly has one of three functions: defense, offense, and communication. GFP are found in numerous organisms exclusively in the jellyfish *Aequorea aequorea* (also commonly referred to as *Aequorea victoria* and *Aequorea forskalea* (Shimomura, 1998). It was first reported in 1955 that *Aequorea* fluoresced green when irradiated with ultraviolet light. Two proteins in *Aequorea* are involved in its bioluminescence, aequorin and green fluorescent protein. Aequorin (the luciferase) contains coelenterazine (the luciferin). Upon binding three calcium ions, the aequorin oxidizes the coelenterazine with a protein bound oxygen resulting in a Ca_3-apo-aequorin-coelenteramide complex which in vitro emits blue light. However, *Aequorea* does not emit blue bioluminescence; instead, the aequorin complex undergoes radiationless energy transfer to GFP which gives off green fluorescence (Figure 1). *Aequorea* GFP is a protein of 238 amino acid residues. Its biggest absorbance peak is at 395 nm with a smaller peak at 475 nm. Excitation at 395 nm yields all emission maximum at 508 nm.

GFP is used as a biological marker. It is particularly useful due to its stability and the fact that its chromophore is formed in an autocatalytic cyclization that does not require a cofactor. This has enabled researchers to use GFP in living systems, and it has led to GFP's widespread use in cell dynamics and development

studies. Furthermore, it appears that fusion of GFP to a protein does not alter the function or location of the protein.

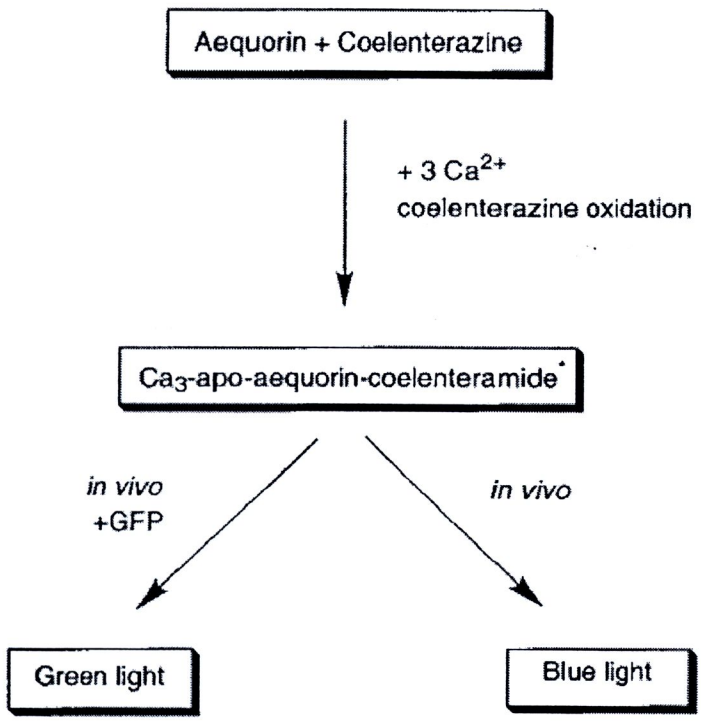


Figure 1 Bioluminescence in *Aequorea victorea* (Zimmer, 2002)

The sequence of wild-type GFP was first determined in 1992 (Prasher et al., 1992) (Table 1). The Protein Data Bank currently lists 22 GFP and GFP mutant crystal structures as well as crystal structures of 2 GFP analogues. Even though many of the mutants have very different spectral properties, their structural features are similar. Although wild-type GFP and most subsequent GFP mutants were crystallized as dimers, some structures of monomeric GFP have been solved. GFP is not an obligate dimer, and dimer formation is very dependent on the crystal growth conditions. GFP has a unique 11 β-sheet barrel-like structure with a diameter of about 24 Å and a height of 42 Å. The β-sheets form the walls of a can, and an α-helix runs diagonally through the can. The chromophore is in the center of the 11 β-

sheets and is linked by the α-helical stretch that runs through the center of the barrel (Figure 2).

Table 1 Nucleotide and amino acid sequence of GFP of wild-type *Aequorea Victoria* (Zimmer, 2002)

a. GFP nucleotide sequence of wild-type *Aequorea*
1 tacacacgaa taaaagataa caaagatgag taaaggagaa gaacttttca ctggagttgt cccaattctt gttgaattag
81 atggtgatgt taatgggcac aaattttctg tcagtggaga gggtgaaggt gatgcaacat acggaaaact taccttaaa
161 tttatttgca ctactggaaa actacctgtt ccatggccaa cacttgtcac tactttctct tatggtgttc aatgcttttc
241 aagataccca gatcatatga aacagcatga ctttttcaag agtgccatgc ccgaaggtta tgtacaggaa agaactatat
321 ttttcaaaga tgacgggaac tacaagacac gtgctgaagt caagtttgaa ggtgataccc ttgttaatag aatcgagtta
400 aaaggtattg attttaaaga agatggaaac attcttggac acaaattgga atacaactat aactcacaca atgtatacat
481 catggcagac aaacaaaaga atggaatcaa agttaacttc aaaattagac acaacattga agatggaagc gttcaactag
561 cagaccatta tcaacaaaat actccaattg gcgatggccc tgtccttta ccagacaacc attacctgc cacacaatct
641 gccctttcga aagatcccaa cgaaaagaga gaccacatgg tccttcttga gtttgtaaca gctgctggga ttacacatgg
721 catggatgaa ctatacaaat aaatgtccag acttccaatt gacactaaag tgtccgaaca attactaaa tctcagggtt
801 ccatggttaaa ttcaggctga gatattattt atatatttat agattcatta aaattgtatg aataatttat tgatgttatt
881 gatagaggtt attttcttat taaacaggct acttggagtg tattcttaat tctatattaa ttacaatttg atttgacttg
961 ctcaaa
b. amino acid sequence of wild-type *Aequorea*
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTL
VTTDFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLV
NRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKKNGIKVNFKIRHNIEDGSVQLAD
HYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYK

By enclosing the chromophore in the can, it may be protected from quenching by oxygen and attack by hydronium ions. Deletion mapping experiments have shown that nearly the entire structure (residues 2-232 or 7-229) is required for

chromophore formation and/or fluorescence. Several polar residues and water molecules comprise a hydrogen-bonding network around the chromophore (Li et al., 1997).

B. Chromophore formation

Folding of GFP into the 11-strand β-barrel shown in Figure 2 is most likely crucial to the formation of the chromophore and its bioluminscene. Early research on GFP and the crystal structures of GFP have shown that the chromophore is formed by an intramolecular cyclization of 65Ser-Tyr-Gly67. Formation is autocatalytic, and the only external requirement is the presence of oxygen (Chalfie et al., 1994). The β-can structure protects the chromophore and is presumably responsible for GFP's stability.

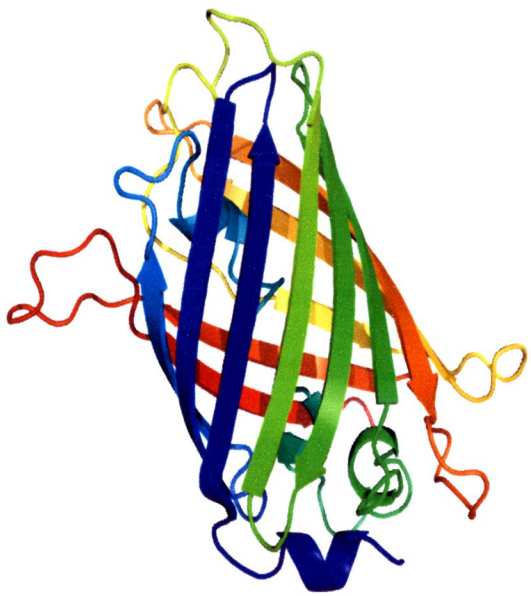


Figure 2 Solid-state structures of GFP, the chromophore is located in the center of the 11-sheet β-barrel (Tsien, 1998)

GFP can be reversibly denatured. Fluorescence is completely lost in the denatured GFP but is regained when the β-can structure is reformed (Bokman and Ward, 1981). The onset of fluorescence can therefore be used as an indicator that the 11-strand β-barrel has been formed. *Aequorea* is found in the cold Pacific Northwest, and mature GFP is most efficiently formed at temperatures well below 37 °C. This has limited the uses of GFP and has led to the search for mutants that mature more efficiently at higher temperatures. Only soluble protein is fluorescent, and it has been suggested that incorrect folding often results in aggregation into insoluble inclusion bodies. Since this misfolding has been blamed for the inefficient maturation, mutants designed for use at higher temperatures have been called folding mutants. The folding mutations can be divided into four groups: buried in close proximity to the chromophore, buried residues located far from the chromophore, surface mutations close to the chromophore, and surface mutations far from the chromophore (Cubbit et al., 1999). Presumably folding mutations located close to the chromophore (e.g., S65A, G, T or L, and F64L and S72A) aid in chromophore formation, while those distant from it (e.g., V163A) are important in forming productive folding intermediates at higher temperatures. Surface mutations (e.g., F99S/M153T/V163A) may aid in decreasing the surface hydrophobicity. Unfortunately not all folding mutations act additively, thereby complicating the design of the most efficient folding mutant. Furthermore, it has been found that some GFP mutants with higher expression efficiencies have increased mRNA transcription and translation efficiencies. One of the most interesting recent findings has been that the folding rates of wild-type GFP and GFP mutants are different in bacteria and mammalian cells. Since GFP most probably folds in the same way in all organisms, this

difference in folding rate is presumably due to the influence of chaperones, which are known to bind GFP. GFP function is based on a chromophore formed through a rarely observed autocatalytic post-translational cyclization of a peptide from its own backbone structure. Initially it was believed that the autocatalytic GFP cyclization was unique, but recent research has indicated that a family of enzymes, including histidine ammonia lyase (HAL) and the closely related phenylalanine ammonia lyase (PAL), also contain post-translational ring formations that occur autocatalytically through the attack of the protein backbone on itself (Donnelly et al., 2001). The detailed mechanism for the formation of the chromophore in GFP is unknown. Molecular mechanical conformational searches based on the wild-type crystal structure of GFP have shown that the chromophore-forming residues of immature GFP are pre-organized in a "tight turn" with the carbonyl carbon of Ser65 less than 2.90 Å from the amide nitrogen of Gly67. Density functional calculations have led to the proposal that dehydration might precede cyclization in GFP (Siegbahn et al., 2001). This led to a proposal of an alternative mechanism where the cyclization is preceded by the dehydration of Tyr66 to dehydrotyrosine. A lot more information about chromophore formation and GFP folding is hidden in the identity and location of the many mutants that result in a total loss of fluorescence. Unfortunately these "negative" results are rarely published.

1.4.1.2. Applications of GFP

A. Fusion Tags

A fusion between a cloned gene and GFP can be created using standard subcloning techniques. The resultant chimera can then be expressed in a cell or organism. In this way, GFP fusion tags can be used to visualize dynamic cellular

events and to monitor protein localization (Lippincott-Schwartz et al., 2001). GFP is ideally suited as a fluorescent fusion protein marker because it does not require the presence of any cofactors or substrates. The chromophore is produced in vivo, and in most cases the resultant chimera does not affect the localization or activity of the tagged protein. For this reason, GFP fusion proteins have been the most common and successful application of GFP in biotechnology. GFP migration from cell-to-cell has been observed; in some cases this was due to nonspecific GFP diffusion. GFP fusion markers have been used in many organisms. The list of successful GFP fusions reported in the literature is very large. Most GFP chimeras have been created by fusing the protein of interest to the amino or carboxyl termini of GFP. Because the amino and carboxyl termini of GFP are located close to each, they can be connected by a peptide linker and circular permuted. To establish why most proteins fused to GFP still retain their function, the motional dynamics of a GFP fusion construct were examined by fluorescence correlation spectroscopy and time-resolved fluorescence anisotropy (Hink et al., 2000). The rotational correlation time of the construct, a single-chain antibody fused to GFP, was too short to be due to a globular rotation of the whole molecule. Instead, the GFP and fusion protein were found to behave independently, in a manner similar to that observed prior to fusion. A fast hinge motion probably occurs between the two fused proteins, and there is no steric interference between the two partners. The length of the polypeptide linker can affect the stability of GFP fusion proteins.

B. The reporter gene

The first application of GFP was as a reporter gene. Gene expression can be monitored by using a GFP gene that is under the control of a

promoter of interest and measuring the GFP fluorescence which directly indicates the gene expression level in living cells or tissue. GFP has been extensively used as a reporter gene, especially in spatial imaging of gene expression in living cells (Naylor, 1999). However, its low sensitivity, due to the fact that there is no signal amplification, because each GFP has only one chromophore, has limited its use. The low sensitivity can be overcome by using sensitive photon counting devices, however, they are too expensive for routine uses. The slow post-translational chromophore formation also limits the use of GFP in the study of fast transcriptional activation processes. Another difficulty in using GFP is the nonlinearity of the fluorescent signal, which necessitates the determination of new calibration curves in each new application. A destabilized form of GFP has been generated for applications in studies that require rapid reporter turnover.

C. Fluorescence resonance energy transfer (FRET)

FRET is a nonradiative exchange of energy from an excited donor fluorophore to an acceptor fluorophore that is within 10-100 Å from the donor. In order for FRET to occur, the emission spectrum of the donor has to overlap with the excitation spectrum of the acceptor. The emission and excitation spectra of BFP and GFP overlap as do the spectra of CFP and YFP, making them good potential FRET pairs. Because any biochemical signal that changes the distance or orientation between the two fluorophores will affect the efficiency of FRET, it is a very useful technique for studying protein-protein interactions in vivo and in vitro. The use of GFP in FRET-based applications has recently been reviewed (Bastiaens and Pepperkok, 2000). Some of the reasons spectral variants of GFP have not been used as FRET partners more frequently are that they fluorescence with relatively low

intensity, the emission spectra of the donor and acceptor pairs are not fully separated, and the chromophore is deeply buried within GFP (~15 Å). The use of FRET to follow the conformational changes involved with metal release from metallathionein is typical of FRET applications based on the concept that a conformation change of a peptide sequence between a FRET pair will change the distance between them and their relative orientations and thereby change the FRET. It will be used to illustrate the technique. The concept has been applied in calcium ion indicators and protease and kinase activity monitoring. Upon reacting with NO (or a secondary product), metallathionein releases zinc or cadmium. A cyan GFP mutant (donor) and a yellow GFP (acceptor) were fused to the amino and carboxy termini of metallathionein. The resultant fusion protein (FRET-MT) retained its ability to bind metal ions. Upon addition of ethylenediamine tetraacetic acid (EDTA) and NaCl, FRET-MT released zinc and the protein unfolded resulting in a drop in the $F535nm/F480nm$ emission ratio. Similar behavior was observed on addition of an aqueous solution of NO to a cell lysate containing FRET-MT. Therefore, the change in conformation upon metal release can be monitored by FRET, which reveals changes in the intramolecular distance and relative orientation of the fluorophores in CFP and YFP (Pearce et al., 2000).

D. Photobleaching

Photobleaching can be used to investigate protein dynamics in living cells. There are two methods based on photobleaching: fluorescence recovery after photobleaching (FRAP) and fluorescence loss in photobleaching (FLIP). By illuminating an area with high intensity illumination (bleaching) and monitoring the recovery of the resultant fluorescence loss (FRAP), the relative mobility of the GFP

chimera can be determined. FLIP can be used to study transport of GFP fusion proteins between different organelles by repeatedly bleaching an area and monitoring the loss of fluorescence from outside the area. Applications of photobleaching GFP in cell biology have been reviewed (White, 1999).

E. Protein-Protein interaction

Because any biochemical signal that changes the distance or orientation between the two fluorophores will affect the efficiency of FRET, it is a very useful technique to study protein-protein interactions in vivo and in vitro. Luminescence resonance energy transfer (LRET) from *Renilla* luciferase to GFP has also been used to study protein-protein interactions in living cells (Wang et al., 2001). Two fusion proteins were expressed in CHO cells: the one linked GFP to insulin-like growth factor while the other joined *Renilla* luciferase to insulin-like growth factor binding protein. Upon addition of coelenterazine, which is required for luciferase luminescence to cells containing the two fusion proteins, LRET occurred indicating that the GFP and luciferase chimeras were in close proximity. Chimeras of GFP linked to aequorin by a protease recognition site linker have been used to demonstrate that chemiluminescence-resonance-energy-transfer (CRET) can be used as a homogeneous assay for proteases and perhaps other proteins. Protein-protein interactions can also be monitored by fluorescence gel retardation, fluorescence polarization assays and affinity capillary electrophoresis (Kiessig et al., 2001). Fluorescence gel retardation is based on the fact that the electrophoretic mobility of a GFP-protein chimera is higher than that of a complex formed by a protein-protein interaction between the GFP-protein chimera and another protein. Some of the drawbacks of fluorescence gel retardation are that it assumes that the protein-protein

interactions remain at equilibrium throughout the electrophoresis and migration though the gel. Furthermore, the conditions during electrophoresis are not always those of interest. Fluorescence polarization assays are based on the fact that a free GFP-protein chimera is likely to rotate more rapidly than a GFP-protein chimera interacting with another protein and will therefore have a lower rotational correlation time than its bound counterpart. Since fluorescence polarization assays can be performed in homogeneous solutions, in which the conditions can be controlled as desired, it is the preferred method over fluorescence gel retardation. A more adventurous method to monitor protein-protein interactions is the production of 2 fusion proteins, protein A and protein B. Protein A is connected to the N terminus of one-half a protein splicing system which in turn is connected to the N terminus of one-half of a EGFP. Protein B is linked to the C terminus of the other half of the splicing system, which is linked to the C terminus of the other half of the EGFP. When proteins A and B interact, the two halves of the protein splicing systems are brought together, correct folding occurs, and the two halves of EGFP are joined by a peptide bond and released. Therefore, the amount of fluorescence, which is proportional to the number of chromophores in mature EGFPs, is proportional to the protein-protein interactions. The interaction between calmodulin and its target peptide M13 has been studied by this method (Ozawa et al., 2000).

F. Other applications

GFP has been used as a marker for tumor cells to illuminate tumor progression and allow for detection of metastases down to the single-cell level and as a whole-body optical imaging system in live mice (Yang et al., 2001). Caution should be used when applying GFP in low oxygen conditions (hypoxia) such as those found

in tumor cells. Histone-GFP fusions have been designed that were sufficiently sensitive to visualize double minute chromosomes in vivo. Double minute chromosomes are paired chromatin bodies found in as many as 50% of human tumors but not found in normal chromosomes. A rapid cell based, functional assay for the screening of chemopreventive agents using GFP as a reporter gene has been developed (Zhu and Fahl, 2000). Using three- and four-color flow cytometry techniques, multiple variables can be evaluated at the same time. The excitation spectra of EYFP, EGFP, ECFP, and DsRed can be analyzed simultaneously by using dual-laser excitation at 458 and 568 nm and multiparameter flow cytometric methods. The spectra of EYFP, EGFP, and ECFP can be simultaneously analyzed by single-laser excitation at 458 nm (Beavis and Kalejta, 1999).

1.4.2. Calcitonin

1.4.2.1. Introduction

Calcitonin (CT), a 32-amino acid peptide hormone secreted by the parafollicular or C cells of the thyroid gland, was originally identified based on its hypocalcemic activity (Copp et al., 1962), which results from its actions in bone and kidney. CT or CT-like substances occur in a wide range of organisms, from unicellular species to mammals. The primary structures of CT from eight species (human, bovine, pig, sheep, rat, chicken, salmon, and eel) have been determined. Six of the first nine residues at the N-terminal are invariant (seven in all species other than chicken), with a disulfide bond between Cys_1 and Cys_7 forming a characteristic ring structure. Gly_{28} and a C-terminal proline amide group are also present in all eight species. The central regions of the peptides are more divergent, although even in this portion of the molecule there are high degrees of identity within groups of two or

three species. Thus, human and rat CT are identical except at positions 16 and 26, eel CT differs from salmon and chicken CT by three and two residues, respectively, and bovine, sheep, and pig CT are closely related. These structural differences are clearly functional significance, because the relative affinities of the different CTs for CT receptors differ by 10-fold or more, with salmon CT binding with a higher affinity than mammalian CTs to CT receptors of all species, even when the mammalian CT and CT receptor are from the same species. The amino acid sequence of salmon calcitonin (sCT) is shown in Figure 3.

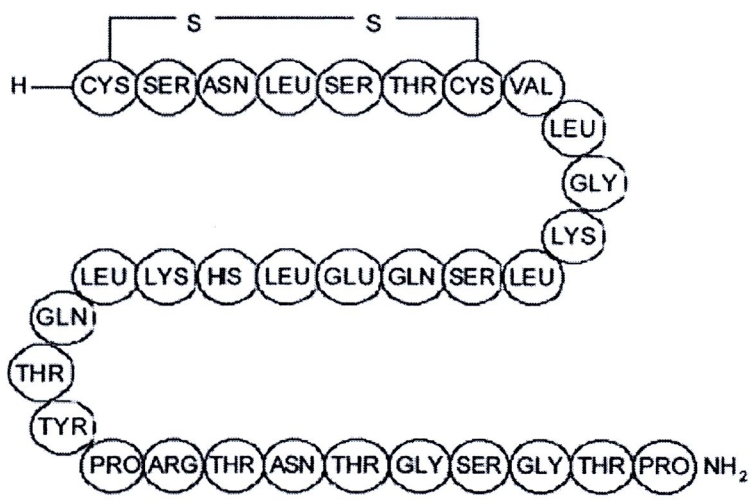


Figure 3 The amino acid sequences of salmon calcitonin (Copp et al., 1962)

In mammals, the primary target organs for CT are bone and the kidney, where it lowers serum calcium concentrations by inhibiting osteoclast-mediated bone resorption and decreasing renal reabsorption of calcium (Azria, 1989). In the kidney, CT also modulates the transport of other ions and water, exerting a potent natriuretic effect (Ardaillou 1975; Bijvoet et al., 1971). CT also affects the central nervous system, where it induces analgesia, gastric acid secretion, and inhibition of appetite (Fischer and Born, 1987). CT is used therapeutically for treatment of a variety of

disorders involving hypercalcemia and/or elevated bone resorption, such as hypercalcemia of malignancy, osteoporosis, and Paget's disease, and has also been used for its analgesic properties (Azria, 1989).

In addition to CT itself having several target organs and diverse effects, the potential complexity of the CT system is further increased by the existence of at least two other structurally related endocrine or paracrine peptides. These are calcitonin gene-related peptide (CGRP), a 37-amino-acid peptide that is also encoded by the CT gene and arises from alternative splicing of the initial CT gene transcript, and amylin, a 37-amino acid peptide encoded by a gene that appears to have evolved from the same ancestral gene as the CT/CGRP gene. The N-terminal domains of CGRP and amylin form disulfide-bonded ring structures similar to that of CT, but the two peptides are otherwise quite dissimilar to CT (16% homology between human CT and human CGRP, 20% homology between amylin and sCT). CGRP is predominantly synthesized in the central and peripheral nervous system, although it is present in and secreted by the thyroid. The presence of CGRP in peripheral nerves that are associated with blood vessels reflects its activity as a vasodilator. It also enhances skeletal muscle contraction and induces increased rate and force of cardiac contraction (Fischer and Born, 1987). In the kidney, it increases blood flow and relaxes mesangial cells effects that are independent of CT receptors and modulates water and electrolyte transport in a manner similar to CT, probably by binding to CT receptors (Kurtz et al., 1989). CGRP has also been recently reported to affect the secretion of interleukin-2 by T lymphocytes. Amylin is secreted by pancreatic β-cells. Its primary effect is in the regulation of carbohydrate metabolism, particularly in muscle, where it inhibits glucose uptake and glycogen synthesis. Both CGRP and

amylin have hypocalcemic effects when administered pharmacologically and inhibit osteoclasts in vitro, probably owing to binding to CT receptors, both on osteoclasts and in the kidney (Zaidi et al., 1990).

1.4.2.2. Physiologic roles of calcitonin

The hypocalcemic effect of CT is primarily due to its inhibitory action on osteoclastic bone resorption, with some additional contribution from increased excretion of Ca^{2+} by the kidney. In bone, actively resorbing osteoclasts seal off an area of bone surface and secrete acid and a variety of acid hydrolases into the resulting space between the cell and the bone via the ruffled border, a specialized, highly invaginated membrane structure that is enriched in vacuolar-type proton pumps and contains a number of proteins characteristic of the endosomal/lysosomal pathway (Baron et al., 1993). Administration of CT induces the rapid internalization of the ruffled border proteins into small intracellular vesicles, thereby halting acid secretion and the resulting demineralization of the bone matrix. In isolated osteoclasts, CT causes reduced osteoclast motility (the Q effect) and retraction of the osteoclast from its usual spread morphology (the R effect) and, possibly as a result of one or both of these activities, inhibits bone resorption as measured by the formation of pits in bone slices (Su et al., 1992). In the kidney, CT enhances the fractional excretion of sodium, chloride, potassium, calcium, and phosphate and stimulates urine flow, as well as influencing the metabolism of vitamin D (Kurokawa et al., 1992). Whereas species differences in the specific effects of CT make the integration of results from different studies problematic, it seems clear that CT acts at a variety of sites throughout the kidney. Thus, CT's natriuretie effect is attributed mostly to an inhibition of sodium reabsorption in the proximal tubule, and CT-induced changes in

phosphate transport and vitamin D metabolism in the rat proximal tubule have also been reported. On the other hand, studies in isolated rabbit and human nephrons have shown that CT activates adenylate cyclase in the medullary and cortical thick ascending limb of the loop of Henle and the distal convoluted tubule, but not in the proximal tubule (Chabardes et al., 1976). The distribution of both immunoreactive CT-like substances (iCT) and CT receptors in human and rat brain has been characterized. Binding sites for ^{125}I-sCT are present in the pituitary and several regions of the diencephalon and the brain stem, including the hypothalamus and the reticular formation (Sexton, 1992). The distribution of iCT is in part congruent with that of the binding sites, with the highest concentrations found in hypothalamus, median eminence, and pituitary. The well-defined distribution of iCT and CT receptors in the central nervous system implies that CT and/or CT-like peptides play specific physiologic roles there. Although such functions have not yet been well characterized, the presence of high levels of both iCT and CT receptors in the hypothalamus, a region of the brain that is involved in pain recognition, feeding, and the regulation of pituitary hormone secretion suggests that CT's antinociceptive and analgesic effects are the results of direct effects of the hormone on cells in this area. CT thus acts at multiple sites to induce a variety of cell-specific effects. The diverse nature of the cellular responses, as well as the apparent discrepancies between the patterns of physiologic response and activation of adenylate cyclase in the kidney, highlights the need to better characterize CT-activated signal transduction in the various target cells and to evaluate the possibility that the presence of multiple CT receptors might underlie the observed physiologic diversity.

1.4.2.3. Indications

A. Osteoporosis

Injectable sCT was first introduced to the market in 1974 and was approved in the United States in 1984. Effects on lumbar spine bone mineral density (BMD) have been reported in a number of smaller controlled clinical trials. Efficacy in vertebral fracture risk reduction was shown in one randomized, controlled clinical trial, risk reduction in hip fractures was reported in the retrospective Mediterranean Osteoporosis study (Kanis et al., 1992). Salmon calcitonin nasal spray (sCT-NS) is currently the most widely used calcitonin formulation, due to its evidence-based efficacy for vertebral fracture prevention and its superior tolerability profile and convenience for daily, long-term administration. The efficacy profile of sCT-NS has been established through results from randomized, controlled clinical studies which have demonstrated a reduction in markers of bone turnover, a moderate effect on BMD, preservation of bone microarchitecture and, most importantly, a reduction in vertebral fracture risk (Chestnut et al., 2000). SCT-NS was approved for treatment of postmenopausal osteoporosis in the United States in 1995.

B. Paget's disease

sCT, which has been used for years in the treatment of Paget's disease, either as injection or nasal spray, is a therapeutic alternative to bisphosphonates. The efficacy and safety of sCT in Paget's disease was established in a number of primarily observational studies initiated in the 1970s and 1980s. Positive clinical effects of sCT included reductions in bone pain, bone turnover markers, improved bone structure assessed by radiography and histology and a reduction in the incidence of pathological fractures (Evans and Slee, 1977). At present sCT constitutes a valid second line treatment option for patients with Paget's disease who

do not tolerate bisphosphonates or for whom bisphosphonates are not effective due to secondary resistance.

C. Analgesic effect

The analgesic potency of CT in bone related pain has been described early on in a number of case series. Numerous CT binding sites have been detected in the CNS, especially in the hypothalamus, and it is therefore assumed that the analgesic effects of sCT are at least to some extent mediated centrally, with serotonergic pathways involved (Colado et al., 1994). In controlled clinical trials, the analgesic efficacy of sCT was demonstrated for various pain-related endpoints, including the reduction in pain scores on the visual analogue scale (VAS), the reduction in concomitant analgesic medication and the improvement of early patient mobilization following acute vertebral fractures. Injectable sCT has been shown to be superior to placebo in acute vertebral fractures, and has shown analgesic effects in metastatic bone disease and in complex regional pain syndrome type I (Perez et al., 2001). In addition a number of non-controlled studies have indicated that injectable CT provides analgesic effects in Paget's disease. For sCT-NS, analgesic effects, especially in pain associated with acute vertebral fractures, have been shown in placebo-controlled clinical trials. Analgesic effects for sCT-NS have also been investigated in the postoperative setting following femoral/hip arthroplasty. Overall, the pain relieving effects of nasal spray and injectable sCT appear comparable (Combe et al., 1997).

1.4.2.4. Pharmaceutical sCT formulations

sCT is commercially available as an injectable formulation for intravenous, intramuscular or subcutaneous use and as a nasal spray. An oral sCT

preparation is currently under clinical development. sCT-NS is recommended at a dose of 200 IU per day for treatment of established post-menopausal osteoporosis to prevent further progression of the disease. Given the comparatively modest effect of sCT-NS on BMD, the clinical monitoring of therapeutic response to sCT-NS is primarily confirmed by preservation of bone mineral density at spine and hip, a significant decrease in markers of bone resorption and the absence of further clinical skeletal fractures. For subcutaneous injection, dosing recommendations range from 50 IU every second day to 100 IU per day depend on the indication and the severity of disease. For emergency treatment of hypercalcemia, sCT has been recommended at a dose of 5-10 IU/kg body weight either as subcutaneous or intramuscular injection or as infusion in 500 ml physiologic saline over at least 6 hours.

1.4.2.5. Calcitonin receptor (CTR) and cell signaling

The activities of CT are mediated by high affinity calcitonin receptors (CTRs). The CTR is a member of a subfamily of the seven transmembrane domain G-protein coupled receptor superfamily that includes several peptides. Members of this family have a similar structure with other seven membrane-spanning domain G-protein coupled receptors. This receptor was characterized by a long NH_2-terminal domain that was extracellular. It was similar to parathyroid/parathyroid hormone related peptide receptor and the secretin receptor (Pondel, 2000). Subsequent cloning of the pCTR gene demonstrated it is approximately 70 kb in length and contains at least 14 exons, 12 of which encode the protein. Different isoforms of CTR resulting from alternative splicing of the gene have been described in various animal species with differential tissue expression transcripts and different signaling properties. Two different isoforms have been described in human giant cell tumor of bone. It is likely

that differential expression of CTR isoforms could be a mechanism of regulation of biological responses to CT. Shift in the predominant CTR isoform could in part explain the variable responsiveness to CT in patients with high turnover metabolic bone disease (Gorn et al., 1995). The principal mechanism of action of CT is due to the ability of its receptor to couple at least two signal transduction pathways. One of the most important pathways is coupled with the cAMP signal transduction. However, CTRs can also couple to the phospholipase C (PLC) enzyme pathway. The PLC pathway, as with the cAMP pathway, can be initiated by the coupling of receptors to multiple G-proteins. Activation of the PLC causes the release of Ca^{2+} from intracellular stores and promotes an influx of external calcium (Pondel, 2000). In addition, CTRs are able to activate the phospholipase D (PLD).

1.4.3. Cell penetrating peptides (CPPs)

1.4.3.1. Introduction

The ability to translocate the cellular membranes and gain access to the cell interior, including the different cellular compartments, still remains a major obstacle in current drug development. Peptide mediated delivery of bioactive molecules appears to be a technology that superior to commonly used delivery agents. Reported high delivery yield, low toxicity and the possibility to add diverse modifications to the peptide backbone make peptides an excellent candidate for future drug delivery platforms. So-called CPPs, also often referred to as protein transduction domains (PTDs), Trojan peptides or membrane translocating sequences (MTS), have in recent years shown great potential in the field of drug delivery. Current publications show that CPPs can deliver a wide range of bioactive molecules such as proteins, peptides, oligonucleotides (ON) and nanoparticles to a variety cell types and

different cellular compartments, both in vivo and in vitro. The peptides named CPPs vary greatly in size, amino acid sequence, and charge, but share the common feature that they have the ability to rapidly translocate the plasma membrane and enable delivery to the cytoplasm or nucleus (Lindgren et al., 2000). The idea to use peptides used as delivery vectors originates from so-called membrane shuttling proteins such as the Drosophila homeobox protein Antennapedia, the HIV-1 transcriptional factor Tat and the capsid protein VP22 from HSV-1. The field started in 1988, when Green et al. showed that the viral protein Tat rapidly translocate over cellular membrane, into the cytoplasm (Green et al., 1988). Later, the same properties were shown for a Drosophila homeobox protein. In 1994, Alain Prochiantz' group demonstrated that a short, 16 amino acid peptide derived from the third loop of the Antennapedia protein was responsible for the cellular translocation of the whole protein. This pioneering work initiated the whole field using peptides as efficient delivery vectors for bioactive compounds, CPPs. Since then, a myriad of peptides has been reported to have cell-penetrating properties. The peptides originate from different classes; either naturally occurring peptide sequences such as virally derived (Tat, VP22), from transcription factors (pAntp), chimeric peptides (transportan, MGP) or synthetic (polyarginines, Pep-1) as shown in Table 2. However, due to difficulties in understanding the true mechanisms of CPP cellular uptake, the classification of CPPs still remains to be clarified. Although great achievements in studies of CPP have been attained, a clear description of their properties is still not defined.

The CPP field has been under constant change during the years and the uptake mechanism still remains ambiguous. Several attempts have been made in order to elucidate the true mechanism of peptide mediated uptake, but the results are

divergent between different reports and experiments. Even when using the same peptide, results vary between different publications. Furthermore, it seems apparent that different peptides utilize different uptake pathways (Thorén et al., 2003).

Table 2 Selection of the commonly used CPPs (Deshayes et al., 2004)

Name	**Sequence**	**Origin**
Penetratin, pAntp	RQIKIWFQNRRMKWKK	*D. melanogaster* transcription factor
HIV Tat peptide (48–60)	GRKKRRQRRRPPQ	Viral transcriptional regulator
HSV-1 VP22 peptide	DAATATRGRSAASRPTE RPRAPARSASRPRRVD	Viral Capsid protein
MAP (Model amphiphilic peptide)	KLALKLALKALKAALKL A-amide	Synthetic
Transportan	GWTLNSAGYLLGKINLK ALAALAKKIL-amide	Chimeric galanin-mastoparan
R7	RRRRRRR	Synthetic
MPG	GALFLGWLGAAGSTMG APKKKRKV	Chimeric HIV-1 gp41-SV40 large T antigen
Pep-1	KETWWETWWTEWSQP KKKRKV	Synthetic

1.4.3.2. Mechanism of translocation

Identification of the mode of action of CPPs is crucial for the design of future generations of CPPs. The present status concerning this aspect is still matter of debate. Investigation of the mechanism of internalization requires identification of several physicochemical properties of the carrier peptides. First, it is crucial to elucidate the type of interaction that the peptide can elicit in the presence of membrane components. It is also necessary to identify the structural criteria, mainly the peptide primary and secondary structures, which can influence the internalization process. Finally, three main entry mechanisms can be examined: direct penetration into the membrane, translocation through formation of a transient structure and endocytosis-mediated entry (Figure 4). With respect to the interaction of CPPs with the plasma membrane, most data stress the role of positive charges, which allow direct electrostatic interactions with phospholipid headgroups. However, some differences can be noticed depending on the nature of the CPP. As an example, penetratin differs from transportan in that the former interacts preferentially with negatively charged membranes, whereas interactions of the latter do not depend on charges (Magzoub et al., 2001). Furthermore, comparison of internalization properties of all-L with those of all-D peptides indicates that this process is not receptor-mediated, as confirmed by the use of giant unilamellar vesicles (Thoren et al., 2000). Other investigations point out the importance of the heparan sulphates present at the cell surface. Indeed, internalization of Tat, penetratin and polyarginines is inhibited upon degradation of the heparan sulphates, by addition of heparin or sulphated polysaccharides or heparan sulphate-deficient cell lines (Tyagi et al., 2001). However, other reports claim that Tat and penetratin can enter cells independent of the presence of heparin sulphates (Silhol et al., 2002). Characterization of structural

criteria associated with efficient CPPs remains to be improved, as only a few studies have been devoted to this issue.

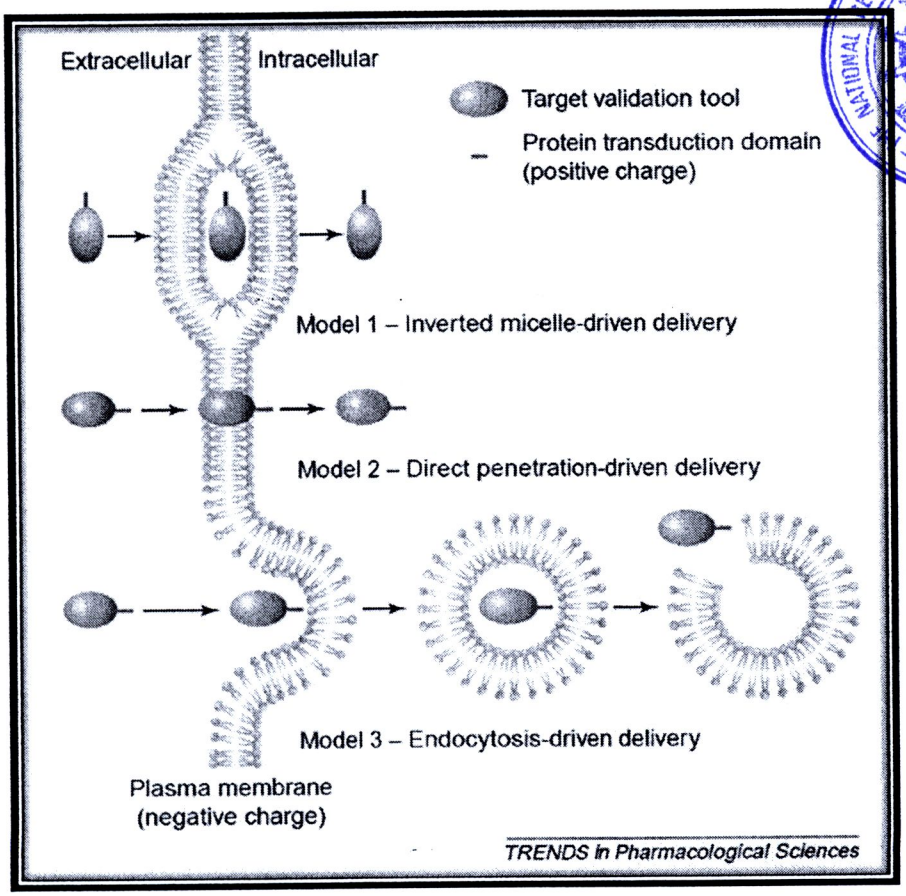


Figure 4 Proposed mechanisms of cellular delivery of cargos mediated by cell penetrating peptides (CPPs) (Tréhin and Merkle, 2004).

Up till now, several mechanisms have been proposed. The first mechanism describes direct penetration through the plasma membrane and has been proposed for Tat. The unfolded fusion protein interacts first with the membrane through electrostatic interactions and then crosses the membrane directly. Once inside the cell, the fusion protein is refolded. However, this mechanism has been recently questioned, and both an internalization mechanism involving macropinocytosis and a clathrin-dependent endocytosis mechanism have been proposed (Richard et al., 2005). A second mechanism accounting for translocation of

penetratin has been proposed based on the formation of inverted micelles (Derossi et al., 1996). In this model, a penetratin dimer interacts with the negatively charged phospholipids, thereby inducing formation of an inverted micelle inside the lipid bilayer. The inverted micelle structure allows the peptide to remain in a hydrophilic environment. However, the mechanism is still matter of debate, since non-symmetric distribution of penetratin between the inner and outer membrane leaflets generating an electric field has been shown. Upon increasing the amount of peptide on the outer leaflet, the electric field reaches a critical value which can generate an electroporation-like phenomenon.

The third mechanism accounting for cellular internalization which has to be considered is endocytosis. Most investigations are based on identification of the cellular localization of CPPs or localization of the cargoes by means of fluorescence or by the use of inhibitors of endocytosis. Unfortunately, artifacts arising during preparation of samples (fixation for example) generate questionable information with respect to endocytosis. Finally, other investigations show that cellular entry of penetratin requires energy and enters by endocytosis, while polyarginines interact with heparan sulphates, followed by endocytosis and Tat is internalized through a macropinocytosis-mediated mechanism. Although these latter investigations clearly indicate that endocytosis is involved in the internalization of CPPs, it appears that different mechanisms could occur simultaneously. This is confirmed by the behaviour reported for penetratin and transportan for which both membrane translocation and endocytosis, occurring simultaneously, have been suggested (Letoha et al., 2003).

The final mechanism was recently suggested for internalization mediated by peptides belonging to the family of primary amphipathic peptides, namely MPG and Pep-1. On the basis of physicochemical investigations, including circular dichroism, Fourier transform infrared and nuclear magnetic resonance spectrometries associated with electrophysiological measurements and investigations dealing with the use of systems mimicking model membranes such as monolayers at the air-water interface and transferred monolayers, two very similar models have been proposed. Both are based on formation of transient pore-like structures. The main difference between the model proposed for MPG and that proposed for Pep-1 is found in the structure giving rise to the pore structure. For MPG, it is formed by a β-barrel structure, while that of Pep-1 depends on association of helices as shown in Figure 5 (Deshayes et al., 2004).

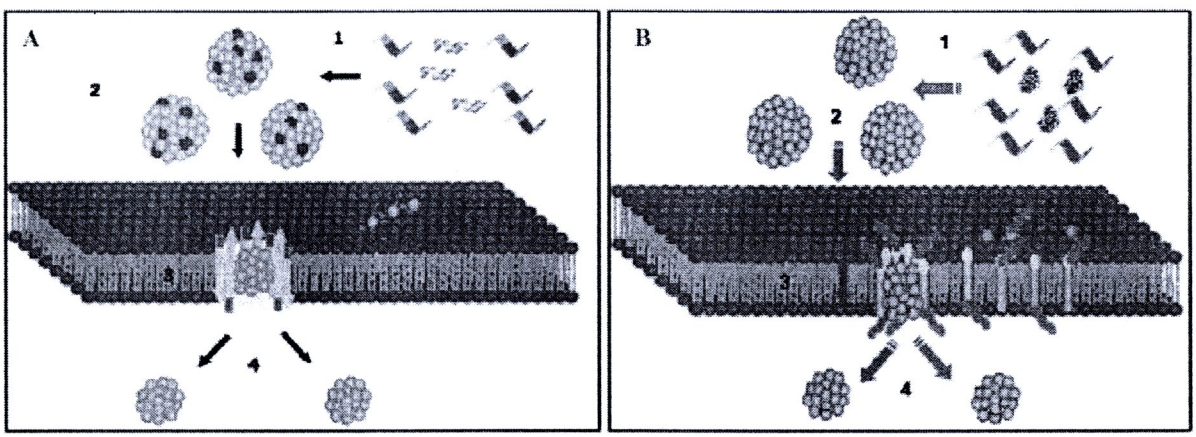


Figure 5 (A) Proposed models for MPG-mediated membrane translocation of nucleic acids. The four steps correspond to formation of the complex; membrane uptake of the complex; translocation through the bilayer; and release into the cytoplasm. (B) Model of Pep-1-mediated transfer of proteins through lipid bilayers. The steps are similar to those of A (Deshayes et al., 2004).

For both peptides strong hydrophobic phospholipid-peptide interactions have been detected, and in both models the folded parts of the carrier molecule correspond to the hydrophobic domain, while the rest of the molecule (linker + NLS) remains unstructured.

1.4.3.3. Applications

A. The targeted and enhanced delivery

Since CPPs penetrate virtually any all cell type both in vivo and in vitro, the delivery seems to utilize a pathway(s) present in all cells. This common feature makes CPP applications complicated in pharmaceutical use. The peptides seem to enter any and every cell they get in contact with, which restricts CPP application as a pharmaceutical tool greatly. Recent results, show promising results in targeted CPP delivery. By exploiting explicit cell features such as extracellular receptors and proteases or addition of cell specific ligands such as small molecules, vitamins, carbohydrates, other peptides or proteins (growth factors or antibodies) to known CPPs may improve targeted delivery (Wagner et al., 2005). The stability of peptide vectors is another problem regarding in vivo delivery. If the vector does not remain intact until it reaches its target, it could do more harm than good. This is most often solved by using the D-form (Brugidou et al., 1995), instead of naturally occurring L-amino acids. The D-form is not degraded by proteases to the same extent and remains intact for a longer time when injected in vivo. Other possibilities are to use peptide mimics such as beta-peptides or peptoids to enhance vector stability (Wender et al., 2000). To increase delivery yield, functional groups can be included in the CPP sequence. For instance, addition of 20 amino acids from the pH-sensitive fusogenic peptide HA2 derived from the N-terminal part of influenza virus

hemagglutinin protein to a Tat-peptide fusion protein enhanced protein uptake significantly. HA2 is a pH-sensitive fusogenic peptide that destabilizes lipid membranes at low pH and thereby enhances endosomal escape which leads to improved delivery (Han et al., 2001).

B. Protein and peptide delivery

Proteins have been evolutionarily selected to perform specific functions. Thus, the ability to deliver a wide variety of full-length, functional proteins has tremendous potential as a biological tool for studying cellular processes as well as for developing novel and potentially very specific therapeutic agents. To date, a growing number of transducible proteins covering a wide range of sizes and functional classes have been successfully used to study intracellular mechanisms. These include PDX-1, BETA2/NeuroD, Ngn3, enhanced green fluorescent protein (EGFP; Cre recombinase, and p53. PTDs have been shown to deliver proteins in excess of 100 KDa into cultured cells and most cells in mammalian model systems (Schwarze et al., 1999). Peptides can be used in a vast range of applications in pharmaceutical research. Several peptides have been fused to PTDs and this could enabling their entry into the cytoplasm. The introduced peptides inhibit protein binding or protein complex formation. Subsequently, several reports have described the delivery of peptides and proteins *in vivo*. Recently, the advantages and versatility of protein transduction over viral transgene delivery has been reported. Retroviral vectors expressing β-galactosidase were compared with results from the injection of Tat-β-galactosidase fusion protein *in vivo* (Barka et al., 2000). Injection of Tat-β-galactosidase by retrograde duct injection resulted in transduction into 100% of the rat

salivary gland cells in a concentration dependent manner, whereas viral delivery could only achieve 30-50% efficiency.

C. Antisense nucleotides delivery

The use of antisense techniques as a pharmaceutical tool is interesting because this technology potentially has the ability to down regulate the expression of virtually any desired gene. Antisense techniques are based on sequence-specific oligonucleotide analogs that after introduction to the cytosol, can hybridize with complementary mRNA strands. This hybridization causes translational arrest or recruitment of RNaseH, thereby altering the gene expression in the cell (Kurreck, 2003). The development of antisense therapy has focused mainly on improving methods for oligonucleotide delivery into cells. Although the cellular uptake of naked oligonucleotides is poor, it is significantly increased by coupling the oligonucleotide to a transporter peptide, leading to decreased expression of the gene product of the targeted gene. Morris et al. describe a potent new strategy for oligonucleotide delivery based on the use of a short peptide vector, termed MPG, which contains a hydrophobic domain derived from the fusion sequence of HIV gp41, and a hydrophilic domain derived from the nuclear localization sequence of SV40 T-antigen (Morris et al., 1997). Tung et al. review the preparation and applications of PTD oligonucleotide conjugates (Tung and Stein, 2000). These strategies of oligonucleotide delivery into cultured cells based on a peptide vector offer several advantages including efficiency, stability and absence of cytotoxicity when compared to other commonly used approaches of delivery. The interaction with PTD strongly increases both the stability of the oligonucleotide to the nuclease and the crossing of the plasma membrane. The rapid degradation of natural oligonucleotides inside the

cell prohibits their use in antisense technology. Many oligonucleotide analogs have been used in antisense techniques which provided varying results. The development of modified oligonucleotides is constantly progressing. Increased stability, enhanced RNA binding affinity and lower toxicity are just some of the aspects to be considered when choosing a suitable oligonucleotide.

D. Peptide nucleic acids (PNAs) delivery

PNAs form stable complexes with DNA and RNA, have low toxicity and, unlike naturally occurring oligonucleotides, are not sensitive to nucleases. These features make them an ideal tool for antisense therapy. As is the case for the majority of large molecules, the cellular uptake of these oligonucleotides is poor, but coupling to PTDs increases their uptake and thus their applicability as tools for the highly specific down regulation of desired gene products. In many cases the PTD-PNA construct is not synthesized as a continuous chain because synthesis of a continuous chain could interfere with the internalization of the construct and the PNA-mRNA interaction. Therefore, the method most frequently employed is coupling of the PNA to the peptide via a disulfide bond (Pooga et al., 1998). The disulfide bond has no significant effect on internalization and, once inside the cell, the conditions of the intracellular environment cause the reduction of the disulfide bond, thereby releasing PNAs, which are then free to interact with the desired target.

E. siRNA delivery

The recent discovery of the RNA interference pathway in a wide variety of eukaryotic organisms has provided a novel means of characterizing gene function in mammalian cells. Short, interfering RNAs (siRNAs) have considerable potential as a powerful tool in molecular biology research and in the future as

pharmaceutical drugs. However, the major drawback with the use of siRNA, similar to most oligonucleotide-based drugs, is the low yield of cellular uptake. By contrast, the use of a mixture of PTD and siRNA directed towards glyceraldehyde 3-phosphate dehydrogenase (GAPDH) mRNA increased the cellular uptake of the siRNA several fold compared with naked siRNA, and the targeted mRNA was down regulated. PTD and siRNA are not covalently linked, but form a complex through electrostatic interactions (Simeoni et al., 2003). The peptide carrier constitutes an excellent tool for the delivery of siRNA into cells.

F. Liposome delivery

The therapeutic delivery of polar compounds is often inefficient owing to the difficulty of crossing the lipid membrane of cells. As a consequence, delivery of these compounds within liposomal carriers has been the focus of increasing attention, but has been hampered by inefficient cellular uptake and consequent degradation through the endocytic pathway. Torchilin et al. demonstrated that relatively large drug carriers, such as 200 nm liposomes, can be delivered into cells by a Tat peptide attached to the liposome surface (Torchilin et al., 2001). Incubation of Tat liposomes with several cells results in intracellular localization of certain liposomes. Confocal analysis showed that the liposomes remained intracellularly intact 1 h after transduction. The approach may have important implications for drug delivery directly into the cell cytoplasm.

G. Plasmid delivery

Although viral methods have several advantages, they also have many drawbacks such as the immunological response provoked by exposure to viral infection. Other currently employed standard methods, including the application of

polycationic agents, electroporation and microinjection, are inefficient for use in vivo. Polycationic agents have toxicity and targeting problems, electroporation causes high cell mortality and microinjection can only be applied to one cell at a time. However, the use of non-viral synthetic vectors minimizes the risk of triggering an immune response in the treated individual because these vectors lack viral components. PTDs can also deliver plasmids. Sakaguchi et al. designed a new system for transporting oligonucleotides into cell nuclei. The vehicle is composed of glutathione-S-transferase, 7 arginine residues, the DNA-binding domain of GAL4 and a nuclear localization signal, which are linked with flexible glycine stretches. The p53-responsive element linked to the GAL4 upstream activating sequence was efficiently transferred by the vehicle protein into the nuclei of primary cultures of neuronal cells, embryonic stem cells, and various human normal cells (Sakaguchi et al., 2005). The peptide sequence PKKKRKV has mostly been used as a nuclear localization signal (NLS), but it has been suggested that this peptide also has cell-penetrating features, and it has been shown to translocate proteins larger than 970 kDa into the nucleus. The linking of the SV40 NLS peptide sequence via PNA to a reporter gene carrying the vector, transfecting cells and PEI, resulted in an up to eight fold increase in the nuclear uptake of EGFP or lacZ-carrying plasmids, compared with only PEI acting as the transfecting agent. Transfection by PTD vectors might use the counter charge between negatively charged plasmid DNA and the positively charged peptide. Hyndman et al. demonstrated that incorporation without covalent linkage of a Tat-PTD peptide into gene delivery lipoplexes improved gene transfer (Hyndman et al., 2004). Furthermore, sterilized R8 (arginine-8) peptides have been shown to transfect

luciferase-containing plasmids into the nucleus with an efficiency that is comparable to Lipofectamine™ (Futaki et al., 2001).

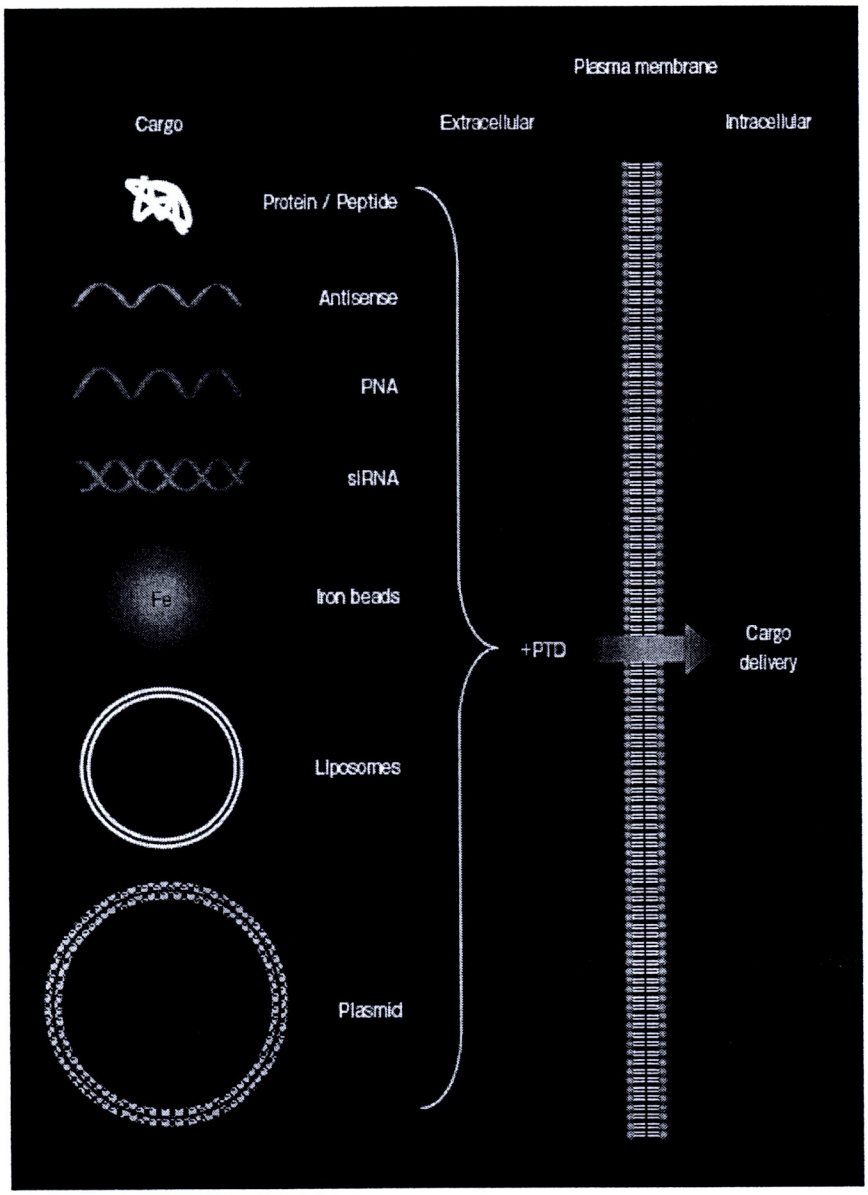


Figure 6 Utility of protein transduction technology. A wide variety of cargo has been covalently linked to PTDs. Currently, most molecules can be transduced into cells when linked to a PTD (Noguchi and Matsumoto, 2006).

To avoid loss of translocation as a result of the PTD interacting with DNA, the PTD can be designed as a branched complex to interact with DNA and

mediate membrane translocation. Eight-branched Tat has been used to transfect cells, with results equivalent to standard Lipofectamine™ strategies. These systems can be a convenient and powerful tool for specifically disrupting the function of DNA-binding proteins.

1.4.3.4. Tat peptide

A. Introduction

Over the last decade, several publications revealed a massive improvement in the cellular delivery of various biologically active molecules upon their attachment to a peptide derived from the HIV-1 Tat protein. This peptide can be reduced to a cluster of basic amino acids containing 6 arginine and 2 lysine residues within a linear sequence of 9 amino acids. Because of the high content of arginine residues within the Tat sequence, various homopolymers of arginine have also been investigated to study the mechanism of entry of various cargoes. Very similar results were obtained with these simple polymers of arginine in terms of transduction efficiency and apparent mechanism of entry compared to the Tat peptide (Futaki et al., 2001). Despite the possibility that the uptake of various entities previously described in the literature could have been artifactual or overestimated, it is unlikely that the efficiency of the Tat-mediated uptake could be disputed, due to the high number of examples of biological activity which have been provided upon Tat peptide mediated cellular delivery of peptides, proteins or nucleic acids (Lindsay, 2002; Gait, 2003). In the large majority of the experiments, the chimera concentration used for obtaining the expected biological responses was not different to those used to assess the cell uptake of the Tat peptide itself. For instance, most of the fused compounds are active at concentrations in the 100 nM range, whereas fluorescence microscopy or

FACS quantification of the uptake were usually performed with 1–10 μM of the peptide (Vives et al., 1997). Despite a possible dose effect causing variations in the efficiency of the uptake, and potentially the cellular pathway induced in uptake, it has been assumed that the entry mechanism of the Tat peptide and of Tat carrying chimera is similar. However, little has been done to unambiguously answer this question by comparing the uptake efficiency of two different entities under the same cellular and experimental conditions. It might be very important to consider the influence of the physicochemical character of the cargo. In spite of the high number of biological applications using these peptides, the precise mechanism of entry still appears controversial and certainly requires further investigations. Contradictory results are still often obtained. They could result from experimental variations in the diversity of the Tat peptide sequence used to promote the translocating activity, the wide variety of cell lines studied, the differing protocols applied to investigate the entry mechanism and the high diversity of cargoes which might influence the behavior of the Tat peptide during the cellular entry process.

B. Tat and cell surface interactions

Because of the highly cationic nature of the Tat peptide, several anionic cellular candidates are available to influence the initial ionic cell surface interactions. These interactions, or this binding to the cell surface can be competitively inhibited with heparin and heparin analogues, such as PPS (pentosan polysulfate) (Rusnati et al., 2001), heparin-binding protein TSP (platelet thrombospondin-1) and other soluble polyanions such as suramin, suramin derivatives dextran sulfate and CS/DS chondroitin/ dermatan sulfates (Sandgren et al., 2002). Many initial studies of internalisation of the Tat peptide, although now known to be

false or compromised in terms of internalisation, can reveal interesting aspects of binding. Where no biological activity is used as a control for effective entry and delivery, externally bound peptide in many cases has been confused with effective delivery. FACS analysis is particularly susceptible to giving artificially high fluorescent values. Given that standard wash techniques in an isotonic buffer, such as the often used PBS, leave residually bound peptide, a more stringent treatment is required before analysis can be performed. The initial association of the Tat peptide with the cell surface membrane occurs independently of temperature, is resistant to mechanical washing with isotonic buffers, such as PBS/EDTA, but is sensitive to treatment with proteases, such as trypsin (Richard et al., 2003). Trypsinisation of externally bound peptide is an often preferred alternative to mechanical washing. Acidic buffers, high salt solutions or competing substances such as heparin which are effective in disrupting ionic interactions has not been quantitatively and comparatively studied for different cell lines. Both the full length GST-Tat protein and fusion proteins containing the basic domain only require a high ionic strength (1.6–1.3 M NaCl) to elute them from bound heparin (Rusnati et al., 2001). In contrast, the substitution of six arginine residues within the basic domain using alanine residues reduced the required ionic strength to only 0.3 M NaCl. This was highlighting the importance of these basic residues in the ionic binding profile of the Tat peptide. Washing the cells with a high salt buffer [20 mM HEPES containing 2 M NaCl (pH 7.4)] produced little difference in the amount of peptides bound to the cell surface compared with PBS alone. In cases where the internalisation of the peptide has not been distinguished from strong extracellular attachment, the data might still be useful with respect to the analysis of binding. In one example, the

uptake of Tat peptide was examined in a range of cells using FACS analysis after washing the cells simply with EDTA, thereby looking at EDTA resistant bound peptide (along with any internalised). Competing anionic compounds, such as heparin and dextran sulfate, caused a significant decrease (60–70%) in cell-surface association of the peptide, whereas other glycosaminoglycans, such as chondroitin sulfate (CS) A, B, and C and hyaluronic acid, had no effect. A similar study looking at GST-Tat-GFP in CHO cells examined only the trypsin resistant fraction of whole protein-GFP and yet came to similar conclusions with regard to the type of glycosaminoglycans (GAG) interacting with Tat. Internalisation of the Tat protein was inhibited by HS, but not by the chondroitin sulfates (Tyagi et al., 2001). Looking at the biological activity of an effectively delivered Cre recombinase, it was shown that heparin was able to confer a total inhibition at low concentrations (2.5 μg/ml), followed by CS-B at only slightly higher amounts (Wadia et al., 2004). CS-C, however, was only able to inhibit recombination by 80% at 20-fold higher concentrations of the inhibitor and CS-A showed only 40% inhibition at all concentrations tested. This study effectively showed the competition for cell surface binding of the Tat-Cre construct with the free GAG and the specificity of this cell surface interaction. It was observed very early on for the full-length Tat protein that uptake and Tat-promoted transactivation of HIV-1 gene expression could be blocked by soluble polyanions (heparin and dextran sulfate) (Mann and Frankel, 1991). Surprisingly, the same study could not show an effect for trypsinisation or heparinase treatment. The Tat protein binds to cell surface heparin sulfate (HS) and heparin. Thus, it can be purified to homogeneity by heparin-affinity chromatography. Neutralization of the positive charges in the basic domain of Tat significantly reduces

its interaction with the GAG. The dissociation constant of heparin to immobilised GST-Tat was observed to be around 0.3 μM (Rusnati et al., 1998).

Work in CHO cells demonstrated that cell uptake and association of Tat constructs containing the basic peptide were effectively blocked by heparin, pre-treatment of HeLa cells with heparinase III or pre-treatment of CHOs with glycosaminoglycan lyases that specifically degrade HS chains (Tyagi et al., 2001). Given the homology of heparin to the surface sulfated glycosaminoglycans, the observed Tat–heparin interaction could reflect the initial cell surface interactions of Tat with exposed surface HS proteoglycans, perhaps serving as the initial point of contact or even a route of entry for Tat, as observed for some other heparin-binding proteins. Sulfated glycosaminoglycans (GAGs), such as HS, increasingly implicated in cell adhesion, are distributed ubiquitously on cell surfaces as the carbohydrate component of proteoglycans. Many microbial and viral particles enter cells using HS receptors via a two-step process, adhering to the cell surface by binding initially to GAGs followed by internalisation. The foot-and-mouth disease virus (FMDV) infects cells in such a way, its primary contact being with a low-affinity HS proteoglycan receptor, followed by transfer to the high-affinity integrin receptor for endocytosis (Jackson et al., 1996). HS facilitates entry of the FMDV and it was found that alteration of the HS affinity had profound consequences for the infectivity of the virus. HSPG serve as cell surface receptors for a number of natural ligands, some of which include matrix proteins, such as laminin, cell adhesion molecules (N-CAM) and growth factors, such as fibroblast growth factor (FGF), insulin-like growth factor-binding protein-2 (IGFBP-2) and vascular endothelial cell growth factor (VEGF). The basic domain of Tat shown to be responsible for the Tat–heparin interaction has

homology with heparin-binding growth factors (Albini et al., 1996). Structurally, there appears to be no conserved conformation for this domain, clusters of basic residues and a heparin-binding capacity using heparin-sepharose chromatography serve as common criteria when labelling a peptide domain as heparin binding. This would potentially provide useful information about the initial binding and entry of the Tat peptide. Binding to HSPGs is often followed by rapid internalisation via endocytosis. There are multiple proposals for the mechanism of internalisation via such proteoglycans, ranging from simple endocytosis via classical clathrin pathways to alternative routes, such as those mediated by the syndecan HSPGs, those independent of coated pits or those utilizing a much slower pathway of internalisation, as was recently described for the perlecans. In the case of syndecan HSPGs, efficient internalisation is triggered by a clustering of transmembrane and cytoplasmic domains and then proceeds via a non-coated pit pathway, possibly caveolae (Fuki et al., 2000).

When CHO cells were pre-treated with chondroitin ABC lyase to eliminate CS/DSPGs or heparitinase/ heparinase to cleave HS chains, all proteolytic treatment resulted in a significant reduction in the uptake of Tat peptide. Likewise, treatment of CHO cells with chlorate (which inhibits GAG sulfation) had a similar inhibitory effect (Sandgren et al., 2002). Mutant cells defective for GAG synthesis show dramatically reduced Tat mediated transmembrane transport. The cell line CHO pgs D-677 which does not produce HSPGs (due to a 10-fold reduction in N-acetylglucosaminyl-transferase and glucuronosyltransferase) produces chondroitin sulfates in excess of about a 3-fold (Rost and Esko, 1997). By contrast, the complete proteoglycan null mutant pgs A-745 (deficient in xylosyl transferase which catalyses the first step in PG assembly/formation) lacks all surface PGs and show a much

severer reduction in Tat uptake of around 80%. CHO 745 cells also show no cell membrane adhesion of the basic peptide or vesicle inclusion using confocal microscopy and show a reduced uptake of Tat peptide/HS complexes, except where the ratios of peptide to anion are particularly high (i.e., high excess of peptide). Thus, the internalisation was if anything more important in the HSPG-deficient cells. In a study examining only EDTA washed cells, the milder D-677 mutant was not observed to have any difference in its binding of Tat when compared to wild type (Console et al., 2003). The severe A-745 mutant showed a reduction of 80-90%. The initial attachment of Tat peptide to GAGs or any other molecule at the cell surface would likely be influenced by an attached or previously bound cargo. The size of the cargo, the overall charge involved, the way in which it was coupled (N- versus C-terminal binding, covalent binding, fusion protein or chemical coupling as well as pure electrostatic interactions) and the degree of exposure of the basic residues would all play a role in influencing these initial cell-surface associations. Aside from the known affinity for HSPGs, other cell surface receptors have also been implicated in Tat binding. Yeast 2-hybrid screens controlled with subsequent GST pull down assays confirmed the binding of the full-length Tat protein to both HSPGs and LRP (low-density lipoprotein receptor family) (Liu et al., 2000). They also confirmed that the domain responsible (amino acids 34–48) was just before the cluster of basic residues (49–57), meaning that any elongation of this minimal sequence to include the α-helical-like structure located just prior to the translocating domain might result in differences in affinity and the pathway of internalisation. Tat and its basic domain have been proposed to bind many cell surface receptors. One study in particular showed the Tat peptide as causing the release of acetylcholine from human and rat

cholinergic terminals. The release was dependent on calcium, effected through voltage sensitive calcium channels and inhibited strongly by cadmium, as well as the mGluR and IP3R antagonists (heparin and xestospongin C). Further studies showed immunoprecipitation of the Tat peptide with various anti-integrin antibodies suggesting that the vitronectin-binding integrin (alphaVbeta5) is the cell surface protein responsible for binding to the basic domain of Tat (Vogel et al., 1993). A natural ligand of this receptor, vitronectin, also contains a related basic peptide sequence (KKQRFRHRNRKG) in its heparin-binding domain, which served to competitively inhibit binding. Another group showed that antibodies to the β-4 integrin subunit were able to inhibit cell attachment to Tat specifically (Weeks et al., 1993). They found a strong relation to a 90-kDa surface membrane protein in both Molt3 and PC12cells. Although HSPGs can resolve at around this size, they vary considerably depending on the saccharide chain length from 12 kDa for the HS chain, to 61 kDa for the core protein, and 90–190 kDa for the intact PG. Rusnati et al. found that a positive correlation existed between the size of heparin oligosaccharides and their capacity to inhibit the internalisation of Tat (Rusnati et al., 1999). Given the size heterogeneity of HSPGs, it is most likely that the observed 90 kDa protein is an additional separate factor in Tat cell membrane adhesion. Taken together, studies on the binding of Tat would implicate more than one component involved in initial cell membrane attachment. A strong argument for the role of GAGs, in particular HS, has been assembled from the data of many independent studies; however, the lack of complete inhibition by mutant, enzymatic digestion or competition studies would tend to preclude their exclusivity.

C. Possible mechanisms of internalisation

Membrane association or binding occurs at any temperature, including the metabolically inhibiting 4°C. In traversing the extracellular membrane, the Tat peptide behaves in an energy-dependent manner requiring temperatures above 4°C and ATP. The initial association is followed by a rapid translocation to the cytosolic side most probably within vesicle-like structures which are acidified according to colocalisation studies with pH markers or inhibition of vesicle acidification. The fluorescence of labeled Tat peptide when observed in live cells is often described as being punctate or vesicular and more rarely as diffuse cytosolic (Ignatovich et al., 2003). It has been observed to be close to the membrane at early time points, progressing to larger aggregations with a more perinuclear type pattern at longer time points and can be observed to continue on into the nucleus. Various drugs have been shown to affect the entry or distribution of the Tat peptide. For instance, Golgi destabilisation of HeLa cells (brefeldinA) converts the punctate vesicular staining to a more cytoplasmic, even distribution, while having no effect on control dextran. Ammonium chloride halted the staining of the nucleus, but appeared to have no effect on the vesicular pattern, leading the authors to conclude that the basic peptide is normally released from vesicles after endosomal uptake by means of a mechanism requiring endosome acidification. Chloroquine, another inhibitor of endosomal acidification appeared likewise to inhibit the release of Tat from vesicles (Fischer et al., 2004), enhancing the vesicular staining for both Arg9 and Tat peptides and reducing/eliminating any diffuse cytoplasmic staining. No effect on Tat was seen with a similar drug, bafilomycinA, nor of wortmannin, a potent PI-kinase inhibitor. The drug monensin, which causes de-acidification of cytoplasmic compartments, resulted in an increase in the fluorescent signal of FITC-labelled Tat, indicating that

the internalised label had been sequestered in acidic compartments, which masked the level of fluorescence. Using the similar Arg9 peptide, it was shown that distribution of the peptide was different according to cell type (vesicular and cytosolic in MC57 cells, while only vesicular in HeLa cells). CytochalasinD is known to depolymerise the actin cytoskeleton causing clustering of caveolae at the cell. A construct of Tat, namely, GST-Tat-EGFP, was found to be restricted to the plasma membrane area following exposure to cytochalasinD, whereas nocodazole treatment which preferentially disrupts the actin microtubules resulted in perinuclear fluorescence similar to untreated controls. Transferrin, often used as a marker for early endosomes following the clathrin-coated vesicular pathway, has been observed to partially colocalise with the Tat peptide (Console et al., 2003). The same has been observed for the non-clathrin markers such as cholera toxin or the SV40 virus, which are internalised via caveolin-cholesterol rich domains. Inhibition of the caveolin pathway by the drug nystatin reduced the Tat peptide reporter β-gal activity by 50% in CHO and HepG2, but showed some cell specificity, having no effect on buffalo green monkey (BGM) cells (Ignatovich et al., 2003). Nystatin also inhibited Tat-phagemediated gene transfer up to 50% of control values, whereas DEAE-dextran-mediated gene transfer remained unaffected. The kinetics of internalisation of a GST-Tat-GFP conjugate and the cholera toxin were reported to be far slower than the comparatively rapid internalisation of transferrin (Fittipaldi et al., 2003). The peptide conjugate showed no evidence of co-localisation with either transferrin, the marker EEA-1 (early endosome antigen-1) or lysotracker dye. Several study concluded that for their conjugate, Tat was internalised via a non-clathrin-dependent route, possibly caveolae. A large number of vesicles containing Tat were positive for Caveolin-1.

Tat uptake was also shown to be inhibited by the sequestration of cholesterol by methyl-hcyclodextrin. It should be noted that uptake of the GST-Tat-GFP conjugates were studied in the presence of 100 μM chloroquine (Tat protein 1 μg/ml), which used as a lysosomal trophic agent to reduce degradation of the parent Tat protein, is nonetheless going to interfere with vesicular recycling and affect the subcellular localization of both control markers and the Tat peptide. The pH neutral environment of caveolae would conflict with data obtained regarding the acidic nature of at least some of the Tat-containing vesicles and the partial colocalisation observed with transferrin for non-conjugated peptide. The data on Tat and Tat-cargoes would tend to preclude the dominance of one exclusive pathway of entry into the cell. The lack of complete inhibition by selective drugs or complete colocalisation with known markers strongly suggests a multiplicity of entry pathways for this sticky basic peptide. Aside from clathrin and caveolae, other mechanisms of crossing the plasma membrane include the non-clathrin/noncaveolin-type pathway(s) (lipid rafts or microdomains, e.g., IL-2 receptor), macropinocytosis (platelet derived growth factor), potocytosis (folate receptor) and phagocytosis (specialised cells only). Perhaps we must also entertain the idea that Tat is simply an opportunistic peptide, adhering strongly to the cell surface on the basis of its charge to any negative offerings, such as lipids or proteins, and then being internalised through natural cell membrane recycling on regions or microdomains, presumably captured by any type of endocytic vesicle. Cell plasma membrane turnover continues constitutively at an estimated rate of ~2%/min (Kilic et al., 2001), in other estimates as fast as 5%/min, meaning 100% of the cells surface is internalised nonspecifically in less than an hour, notwithstanding the faster receptor mediated or stimulated routes of uptake. To this end, most drugs

inhibiting cell membrane recycling will show an effect on Tat uptake, yet not prevent it entirely as long as other possible routes exist. Competitors for cell surface binding, however, would presumably be more effective at reducing the level of peptide internalised as is seen for heparin and the similar glycosaminoglycans.

1.4.4. Poliovirus

1.4.4.1. General introduction

Poliovirus (PV), the causative agent of poliomyelitis, is a member of the genus *Enterovirus* of the Picornaviridae, a large family of small, non-enveloped (+) stranded RNA viruses. Other members of the picornavirus family include the genera of *Rhinovirus* (common cold virus) with over 100 serotypes, *Hepatovirus* including human hepatitis A virus, and two prominent animal virus genera *Cardiovirus* (encephalomyocarditis virus, Theiler's murine encephalomyelitis virus, Mengovirus), and *Apthovirus* (foot and mouth disease virus). Coxsackieviruses and echoviruses, which comprise the majority of the *Enterovirus* genus, have a wide pathological spectrum causing diseases such as meningitis, encephalitis, myocarditis, paralysis, diarrhetic and respiratory diseases (Melnick, 1996). In fact, an estimated one billion human infections per year worldwide attest to the enormous success of enterovirus strategy. As the name implies enteroviruses replicate mainly in the gastrointestinal (GI) tract, often unnoticed or accompanied only by mild symptoms. Spread to secondary sites of replication combined with apparent disease syndromes is relatively rare. However, the vast incidence of human infections translates into large numbers of serious disease complications, sometimes with fatal outcome. The pathology of PV is typical for enteroviruses: the three serotypes replicate very efficiently in the GI tract (up to 60 days post infection). Only at a rate of 10^{-2} (for serotype 1) to 10^{-3} (for

serotypes 2 and 3) does PV cause neurological disease. Thus, the infamous propensity of PV to invade the central nervous system and specifically target motor neurons is rare and accidental and it is neither a prerequisite nor does it present a benefit for its normal life cycle in humans. PV is a neurotropic virus by mistake; its CNS invasion is mostly a chance event and largely independent of the age, gender, or socioeconomic position of the infected person. Although no longer a major public health threat in the developed world, poliovirus continues to be one of most thoroughly studied and best understood model viruses to date. The very stable capsid and ease of virion purification, along with high virus titers and the low bio-safety level requirements make poliovirus a favored target for investigations. At the height of the poliomyelitis epidemics around the middle of the 20th century huge research endeavors that parallel the HIV research of the present day were undertaken to combat poliomyelitis. Many of the basic virological techniques were established during this period as a result of PV research. These include virus purification protocols, crystallization of the virion, an efficient tissue culture system and plaque assay analysis (Dulbecco and Vogt, 1954; Enders et al., 1949). Many millions of Dollars were raised for poliovirus research, which eventually culminated in the independent introduction of two excellent PV vaccines: the inactivated vaccine by Jonas Salk (IPV; Salk et al., 1954) and the live attenuated oral vaccine by Albert Sabin (Sabin, 1957). These vaccines broke the grip of poliomyelitis on society, a success that led the World Health Organization to initiate a global poliovirus eradication program in 1988. The PV genome was the first animal virus RNA genome to be sequenced and genetically mapped. The sequence confirmed the existence of the polyprotein, an intermediate molecule synthesized during expression

of viral genetic information in numerous RNA viruses that was first proposed by Baltimore (Jacobson et al., 1970). The crystal structure of the poliovirion was solved in 1985.

1.4.4.2. The structure of poliovirus

PV contains a single-stranded plus-sense RNA genome within an icosahedral capsid composed of 60 copies each of four coat proteins, VP1, VP2, VP3, and VP4. The virion surface is dominated by star-shaped mesas at the fivefold axes surrounded by deep canyons and three-bladed propellers at the threefold axes separated by saddle-like depressions straddling the twofold axes. The capsid proteins VP1, VP2, and VP3 share the same eight-stranded β-barrel fold but are distinguished from one another by the shape of their loops and N- and C-terminal extensions. The native form of the capsid is stabilized by an intricate network on its inner (RNA-proximal) surface, formed by VP4 and the intertwined N-termini of the other capsid proteins. Much of this network cannot be formed until late in assembly when the precursor, VP0, is cleaved to yield VP4 and VP2 (Basavappa et al., 1994). The poliovirus genome consists of about 7440 nucleotides (Figure 7). A relatively long 5' NTR of 742 nt is followed by a single open reading frame coding for a polyprotein of 2209 amino acids and a 3' NTR of 70 nt (Kitamura et al., 1981).

The first 88 nucleotides at the 5' end of the genome form a clover leaf like secondary structure which has been implicated in the formation of the replication complex during (+) strand RNA synthesis. The 3' NTR contains two stem loops that may interact to form a tertiary pseudoknot structure (Pilipenko et al., 1992). The 3' end of the viral RNA is comprised of a genetically encoded stretch of poly(A) varying around 60 adenylate residues in length. Unlike most eukaryotic mRNAs the

poliovirus genomic RNA does not carry the m7G cap structure essential for translational initiation of normal eukaryotic mRNA. Instead a small viral protein, VPg, is covalently linked to the 5' terminus. Contrary to canonical eukaryotic translational initiation, cap-independent translation of the poliovirus genome is initiated by binding of the host translational machinery to a highly structured RNA *cis*-acting element within the 5' non-translated region of the genome. This element has been termed internal ribosomal entry site (IRES) for its propensity to initiate translation regardless of its location within an RNA molecule (Jang et al., 1989). IRES-dependent translation is now established as a novel mechanism of eukaryotic cap-independent protein synthesis. More and more IRES elements have been discovered not only in viral genomes, but in the 5'untranslated region of many cellular mRNAs.

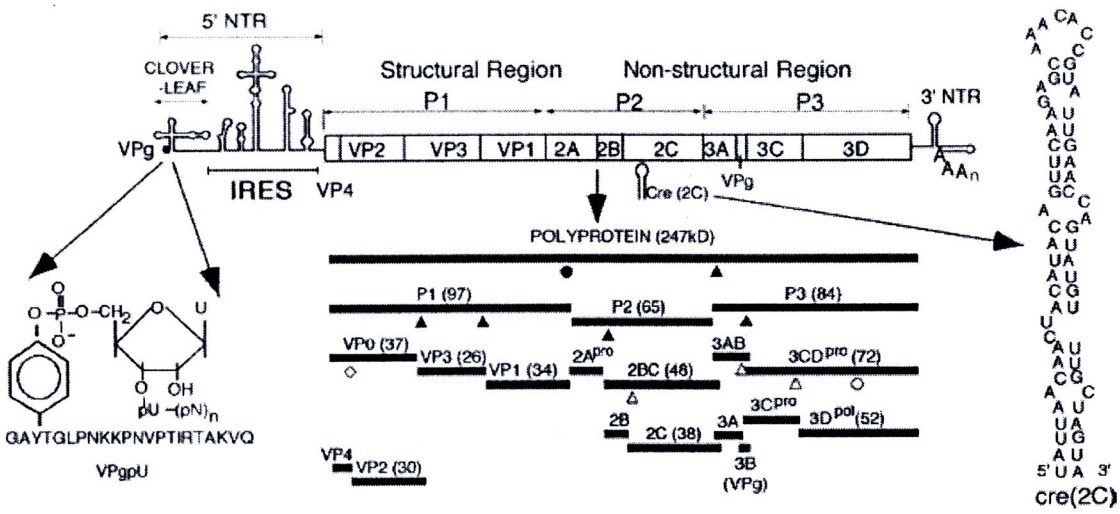


Figure 7 Genomic organization and proteolytic processing of poliovirus. A terminal protein VPg (structure and sequence are indicated on the left) is linked covalently to the 5' end. The 5' NTR consists of several stem loops that form the 5' cloverleaf followed by the IRES element (Mueller et al., 2008).

1.4.4.3. Cellular life cycle

A. Cell entry

PV attachment to the host cell surface is mediated by binding to the N-terminal V-type immunoglobulin like domain of its cognate receptor, CD155. The interaction of PV with its receptor has been thoroughly studied. Mutational and genetic analysis indicated early on that CD155 inserts itself into a surface depression known as the "canyon" which surrounds each of the twelve five-fold axes of the icosahedral capsid (Colston and Racaniello, 1994). This common theme in picornavirus-receptor interaction has subsequently been confirmed through structural analyses by cryo-EM microscopy. It was found that unlike the orthogonal "key in lock" docking of ICAM-1 to rhinovirus, CD155 binds to poliovirus in a rather angled fashion (He et al., 2000). In order for virus uncoating (release of its RNA into the cell cytoplasm) to take place, the extremely stable PV capsid which is resistant to treatment with 2% SDS or pH 2 has to be destabilized. This is thought to occur by CD155-mediated extrusion of the myristoylated capsid protein VP4 and the putative N-terminal amphipathic helix of VP1, both of which are buried inside the virion prior to the transition. These events lead to the formation of 135S particles, also known as "A"-particles. The structural changes during A-particle formation make them hydrophobic which has lead to the theory that A-particles are an intermediate in the uncoating process. According to this hypothesis, following receptor docking the extruded VP1 amphipathic helix and myristoylated VP4 are inserted into the cell membrane (Fricks and Hogle, 1990). This leads to receptor independent attachment of the A-particle to the cell surface and the formation of pores in the membrane through which the virion RNA could be "injected" into the cytoplasm. Thus, the

empty capsid would be left behind at the cell surface and not enter the cell. However, the views are still divided whether A-particles are an obligatory intermediate of virus uncoating or a result of abortive infection. According to another theory PV is taken up by receptor mediated endocytosis although the precise mechanism is just as uncertain. Although PV is clearly found in endosomes, the process is dynamin independent and therefore excludes both classic endocytotic pathways, through either clathrin coated pits or caveoli, as the mode of uptake (Hogle, 2002). Both theories may be reconciled if it is assumed that they actually represent two intermediate steps of the same process. Interestingly, CD155 has two different binding affinities to the virus, arguing in favor of a two-step transition to a proposed "activated" virus-receptor complex, which then proceeds to A-particle formation and uncoating. Whereas under favorable conditions of tissue culture studies the two step conversion to the activated virus-receptor complex may be favored and proceed relatively quickly, the reaction may be much slower when the virus encounters the receptor within a host tissue, such as a motor neuron synapse. Here, virus receptor binding and transition to the activated state could be temporally separated. The virus and receptor may remain loosely attached in a low affinity state while being endocytosed and transported to neuronal cell body, only to undergo the same transition proposed to occur otherwise at the cell surface. In this case, the virus may remain in an endocytic vesicle, i.e. in the extracellular milieu, complexed with the receptor for an undetermined duration and only uncoat, once favorable intracellular conditions trigger the reaction to the high affinity state. Rather than being "injected" at the cell surface the virion RNA could use the same mechanism to "escape" the endocytic vesicle. Since cytoplasmic membranes at a nerve terminal and on cultured cell are vastly

different in composition and function, it should not surprise if it will be found in the future that virus uncoating dynamics may be altered.

B. Polyprotein translation and proteolytic processing

Upon viral infection the virion RNA is released into the cytoplasm of the host cell and serves as a messenger RNA that is translated into a single highly autocatalytic polyprotein, whose proteolytic cleavage products serve as capsid precursors (80 S particles) and replication proteins (Wimmer et al., 1993). Expression of the genomic information in the form of a polyprotein suggests that all viral proteins are produced at an equimolar ratio. However, an intricate cascade of fast and slow proteolytic cleavages by its three proteinases $2A^{pro}$, $3C^{pro}/3CD^{pro}$, allows poliovirus a certain degree of posttranslational control of the amounts of viral proteins produced. All $3C^{pro}/3CD^{pro}$ cleavages occur at a Gln*Gly pair, that of $2A^{pro}$ at a Tyr*Gly pair (Kräusslich and Wimmer, 1988). The individual cleavage recognition signals for $3C^{pro}/3CD^{pro}$ differ slightly from each other in the sequences upstream and downstream of the scissile bond so that the sites are not cleaved with the same efficiency. This leads to a disproportional accumulation of some of the viral proteins and their precursors. Polymerase $3D^{pol}$ is only present in small amounts to synthesize viral (+) and (−) strand RNA templates. Facilitated by a slow cleavage between $3C^{pro}$ and $3D^{pol}$ most of $3D^{pol}$ is therefore present in its precursor form $3CD^{pro}$, the major proteinase employed in polyprotein processing. Interestingly, $3CD^{pro}$ does not posses any polymerase activity while, on the other hand, it is capable of performing almost all the functions of $3C^{pro}$ and plays a major role as an RNA binding protein during genome replication. This ability to encode several functions in the same primary amino acid sequence increases the genomic information while allowing the virus to

keep a small genome. Processing of the polyprotein occurs in three stages once translation is completed. First $2A^{pro}$ carries out a *cis*- cleavage of the Tyr-Gly bond at its own N-terminus to release the capsid precursor P1. The second major processing step is the release of P3 from the remaining protein through *cis*-cleavage by proteinase $3CD^{pro}$ (Lawson and Semler, 1992). In a highly ordered succession of *trans*-cleavage events by 3CDpro the non-structural proteins 2A, 2BC, 3AB, 2B, 2C, 3A, 3B (VPg), $3C^{pro}$, $3D^{pol}$ and the capsid proteins VP0, VP1 and VP3 are then released from their precursors. In the last stage of proteolysis, taking place during assembly of the virus particle, VP0 is cleaved, presumably by an autocatalytic mechanism involving the encapsidating RNA, to yield VP4 and VP2 (Basavappa et al., 1994). This cleavage is required for infectivity of the virus and is therefore referred to as maturation cleavage. In addition to processing of the viral polyprotein $2A^{pro}$ and $3C^{pro}$ are actively involved in inhibition of host cell functions. $2A^{pro}$ is responsible for the cleavage of p220, a component of the cap recognizing complex eIF-4F, thereby shutting off cap-dependent host cell translation to the benefit of the virus' own cap independent IRES driven translation. $3C^{pro}$ has been shown to inactivate transcription factor TFIIIC (Clark et al., 1991) and to cleave the TATA box binding protein, which leads to an inhibition of cellular transcription.

C. RNA replication

RNA replication proceeds via the following general and, admittedly, simplified pathway:

Incoming (+) strand virion RNA → (−) strand synthesis→ RF → (+) strand synthesis → RI → (+) strand RNA,

where RF means replicative form (double stranded RNA), and RI means replicative intermediate [a (−) strand partially hybridized to numerous nascent (+) strands] (Wimmer et al., 1993). However, this simple representation does not do justice to the complex nature of PV genome replication. This and other laboratories have studied the problem of picornavirus RNA replication for over 30 years, and yet more questions are open than have been answered. Countless mutational and genetic studies have shown that all of the non-structural proteins are involved at some step of the genome replication. Naturally, the most direct function lies with the RNA dependent RNA polymerase $3D^{pol}$. This strictly primer dependent enzyme catalyzes both the synthesis of (−) strands and (+) strands. It has two types of catalytic activities, it elongates RNA chains on a template and it uridylylates VPg (Paul et al., 1998). The small protein VPg (3B) serves as a protein primer after being uridylylated on a conserved tyrosine residue to form VPg-pU-pU, representing the first step of both (+) and (−) strand synthesis. It was a long held convention that VPg uridylylation occurs at the poly(A) tail of (+) strands. However, it was unclear how specificity could be conferred to this reaction, in light of the vast excess of cellular poly(A) containing mRNAs. This reaction is strongly enhanced in an in vitro uridylylation assay by viral proteinase $3CD^{pro}$ and is probably dependent on $3CD^{pro}$ in vivo. The need for $3CD^{pro}$ within the uridylylation complex is thought to provide the necessary in vivo specificity to this reaction. It has recently been suggested that cre dependent VPg uridylylation is only involved in (+) strand synthesis but not in (−) strand synthesis (Morasco et al., 2003). RNA replication occurs on rosette-like membranous structures that are induced by the action viral proteins 2C and 2BC and accumulate in the cytoplasm of the infected cell (Cho et al., 1994; Teterina et al.,

1997). The membrane material in these structures is primarily derived from the host cells endoplasmic reticulum. A hydrophobic domain in 3AB is thought to insert the protein into the membranes and recruit the replication complex to the site by 3AB's affinity to $3D^{pol}$ and $3CD^{pro}$. Virus assembly, i.e. encapsidation of progeny genomes into procapsids is poorly understood. Following encapsidation a maturation cleavage of VP0 occurs to form VP4 and VP2, which locks in the RNA and stabilizes the capsid. Since no viral or cellular proteinases seem to be required, the working hypothesis assumes an autocatalytic mechanism, possibly in conjunction with the RNA. Under experimental conditions the poliovirus life cycle is extremely rapid, resulting in death of the host cell approximately 7–8 h post infection. It is commonly held belief that poliovirus exits the cell by lysis of the host cell. However, newly synthesized virus can be detected long before lysis. In addition, poliovirus is able to establish persitent infection without killing the cell. Both observations argue in favor of alternative exit pathways by poliovirus. In vitro studies have shown that PV infections can trigger apoptosis (Agol, 2002). The possible role of this active process of cell death is to limit virus spread. The present data indicate that two nonstructural proteins, the $2A^{pro}$ and $3C^{pro}$ proteases, play an important role in the activation of the apoptotic program in PV-infected cells. On the other hand, poliovirus can also inhibit apoptosis. It has been shown that 2B/2BC, 2CATPase and 3A exhibit an anti-apoptotic activity. Thus, it has become apparent that proapototic and anti-apoptotic functions coexist in PV-infected cells. However, the interplay of these opposite functions in PV-infected cells, particularly in neurons, and the role of this interplay in cell death or survival to polio infection are yet to be elucidated. It can be speculated

that a shift in the balance may be different in different tissue cells and, thus, apoptosis may play a significant in poliovirus pathogenesis.

1.4.4.4. Poliovirus receptor (PVR, CD155) and pathogenesis

PV infection is initiated by binding to a specific receptor, PVR (poliovirus receptor) or CD155. PVR is a member of the nectin family of CAM-like cell surface proteins and consists of an N-terminal ectodomain with three immunoglobulin-like domains, a transmembrane domain and either of two splice variant C-terminal cytoplasmic domains. Cryo-electron microscopy (cryo-EM) reconstructions of virus-receptor complexes demonstrate that the receptors bind with their N-terminal ectodomains inserted into the canyons surrounding the fivefold axes (Belnap et al., 2000). When native virus binds its receptor at physiological temperature, it undergoes an irreversible conformational change, producing the 135S particle. This conformational change, which can also be induced by heating virus in a low-salt buffer, results in a decreased affinity for PVR and externalization of VP4 (which is myristoylated at its N terminus) and the N terminus of VP1. Unlike the native virion, the 135S particle is sensitive to proteases, including the V8 protease of *Staphylococcus aureus*, which specifically cleaves after a glutamate at position 31 of VP1. Removal of the first 31 amino acids of VP1 prevents the 135S particle from associating with membranes or releases it from preformed 135S-membrane complexes, leaving only the 31-amino-acid peptide attached to the membrane (Fricks and Hogle, 1990). Late in infection, the virus undergoes a second irreversible conformational change resulting in release of the viral RNA, leaving an empty particle which sediments at 80S. The trigger for this final stage of the entry process and where it takes place in the cell are not known. Electrophysiology experiments

have implicated the membrane- associated N terminus of VP1 and VP4 in the formation of channels in lipid bilayers (Tosteson and Chow, 1997). In addition, viruses containing mutations in the N terminus of VP1 are defective in RNA release and viruses with a T-G mutation in residue 28 of VP4 are defective in both channel formation and RNA release. Together, these experiments support a model in which these channels facilitate translocation of RNA across cell membranes. Structures of 135S and 80S particles have been determined by cryo-EM at 22-Å and 23-Å resolution. Computational docking of the β-barrels from crystal structures into the 135S density map provided insights into the structural rearrangements that the capsid undergoes during cell entry, which resemble tectonic shifts, i.e., rigid-body movements on a spherical surface. These data were inconsistent with models that envisaged the N terminus of VP1 exiting the virion via the channel at the fivefold axis (Giranda et al., 1992) and, rather, suggested that it exits at the base of the canyon. However, the resolution was insufficient to conclusively identify the point of egress of the N terminus of VP1 or its location in the 135S particle.

Humans and non-human primates are the only natural hosts for PV. Primate research is very expensive and cases of human poliomyelitis due to circulating wild type viruses dwindled due to the great success of the existing PV vaccines. Mice or any other affordable experimental animals were found to be resistant to infection with most poliovirus strains owing to the lack of expression of the poliovirus receptor. An exception is certain strains of poliovirus type 2 that, after intracerebral injection, cause paralysis and death in mice (Gromeier et al., 1995). The development of a mouse model for poliomyelitis that is transgenic for the hPVR makes it much easier to investigate the efficiency of poliovirus dissemination in a

whole organism, although the crucial enteric phase of natural poliomyelitis is impossible to study using this model because of extremely inefficient oral infection, as seen in the monkey model. Nevertheless, the results from the mouse infection experiments using intraspinal (ISp), intracerebral (IC), intravenous (IV) and intramuscular (IM) inoculation routes have indicated new directions for research into poliovirus pathogenesis. Transgenic mice carrying the gene encoding the hPVR are susceptible to PV by IV, IM, IC or ISp inoculation (Ren et al., 1990). As non-transgenic mice are generally not susceptible to PV, the hPVR must be an important determinant of species specificity. Transgenic mice infected via any of these four inoculation routes show signs of paralysis that resemble those of poliomyelitis in humans and monkeys. Furthermore, the results of histocytochemical examinations of the CNS of paralyzed mice are similar to those of humans and monkeys. These data indicate that IV- and IM-inoculated poliovirus can enter the CNS and replicate in neurons. Experimental evidence involving transgenic mice suggests that IV-inoculated poliovirus invades the CNS mainly through permeation of the BBB, and IM-inoculated poliovirus mainly through the neural pathway. Currently, it is unclear why hPVR mediated conformational change was not observed for virion-related particles in the sciatic nerve. Perhaps a few bound hPVRs per virion does not result in viral conformational change but nevertheless induces endocytosis of the virus on the cell surface. PV infection can be established by any of several pathways, such as hPVR-mediated fusion, resulting in release of the viral genome into the cell cytoplasm, and hPVR-mediated endocytosis, followed by uncoating in endosomes (Willingmann et al., 1989). The latter pathway could be the main way that PV enters motor neurons *in vivo*.

Typically, the most frequent manifestation of PV infection in humans is the replication in the gastrointestinal tract and subsequent shedding of the virus in feces. PV initially gains entry through oral route and infects susceptible cells of the mouth, nose, and throat. The incubation period ranges from 7 to 14 days but may vary from 2 to 35 days. Although the precise sites and cell types involved in the primary replication of PV in the gastrointestinal tract are not known, virus could generally be isolated from the lymphatic tissues of the gastrointestinal tract, such as the tonsils, the Peyer's patches of the ileum and mesenteric lymph nodes (Melnick, 1996; Sabin, 1956). Primary (minor) transient viremia occurs in most infected individuals and the virus spreads to the systemic reticuloendothelial tissue without clinical manifestations. In 4-8% of the infected individuals, a second major viremia ensues causing symptoms of "minor illness" (headache, sore throat, fever). A fraction of those with major viremia proceed to develop symptoms of CNS involvement. Thus, neurological symptoms are rare complications of the PV infection and poliomyelitis (the most important and devastating disease causes by PV) affects less than 1% of PV infected individuals (Nathanson and Martin, 1979). Poliomyelitis can therefore be regarded as an "accident" of the enteric infection that is neither an obligatory stage of PV replication nor of any known benefit to the virus. The molecular mechanisms by which PV causes poliomyelitis are poorly understood. This is remarkable considering that PV is one of the most thoroughly investigated viruses of all times. Poliomyelitis is marked by a selective destruction of motor neurons, leading to paralysis and, in severe cases, to respiratory arrest and death. How poliovirus spreads to the central nervous system is poorly understood. Two theories prevail. First, the virus may either directly pass from the blood into the CNS by

crossing the blood brain barrier, independently of its receptor. The second hypothesis, which has been experimentally supported in the CD155-tg mice suggest that the virus is transported by retrograde axonal transport ascending from the muscle to the spinal cord and brain (Figure 8). Both depend on the presence of virus in the blood (viremia).

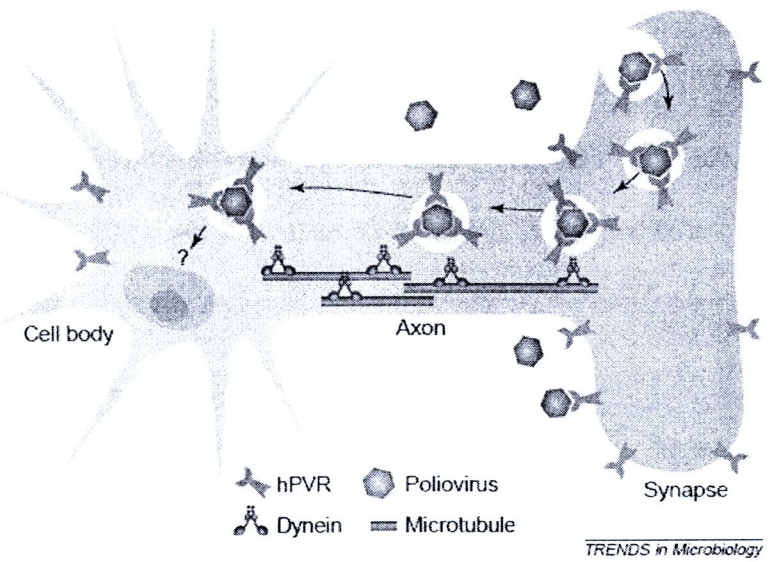


Figure 8 Hypothesis for poliovirus transport through the axon. Human poliovirus receptor (hPVR)-mediated endocytosis occurs at synapses (Ohka and Nomoto, 2001).

1.4.4.5. Interaction of poliovirus receptor to poliovirus

CD155 is membrane-anchored, single-span glycoproteins whose extracellular regions consist of three and five domains, respectively, each with Ig-like folds (Figure 9). The amino-terminal domain, D1, contains the virus recognition site. Hence, virus attachment occurs at a site on the receptor that is distal from the plasma membrane. This property may be important for successful initiation of infection of cells by viruses and may reflect the enhanced ability of the N-terminal Ig domain to penetrate into the picornaviral canyon.

PVR binds at a glancing angle such that its d1 domain penetrates the canyon and makes contact with the capsid surface near the center of the icosahedral asymmetric unit bounded by a 5-fold axis and two 3-fold axes. The major contact points are in a cleft on the ''south rim'' of the canyon and on the side of the mesa on the ''north rim,'' so that the receptor appears to bridge the canyon. By overlaying the virion–PVR reconstruction and the difference map on the atomic-resolution coordinates of the virion, it was found that the PVR-binding site includes many residues in VP1 (102–108, 166–169, 213–214, 222–236, 293–297, 301–302), a few in VP2 (140–144, 170–172), and several in VP3 (58–62, 93, 182–186).

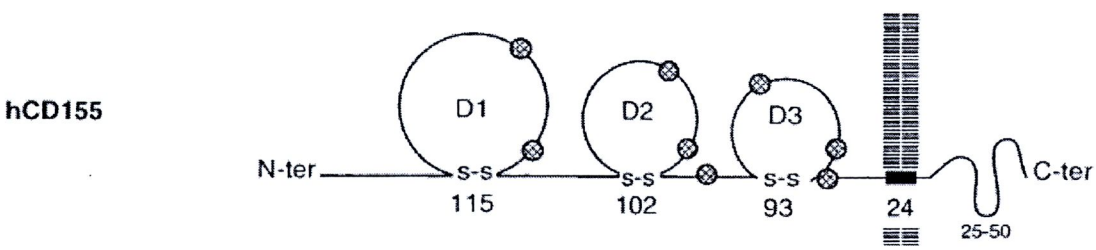


Figure 9 The mature structure of human PV receptor (hCD155) Sites of glycosylation are indicated by circles. The number of amino acids is shown for each domain (He et al., 2000).

All contacts of the receptor with the virus involve domain d1, which approaches the capsid surface from the right. The footprint of CD155 on the surface of PV was established from the atomic model of the PV–CD155 complex by identifying those PV residues with atoms that lie within 4.0 Å from any CD155 atom in the modeled complex (Figure 10). CD155 overlaps the north and south walls as well as floor of the canyon. Indeed, it has been suggested that CD155 may be in contact with residues from two neutralization antigenic sites on either side of the

canyon. Because the D1 domain of CD155 leans toward the virus surface, much of its C, C', C'' face makes additional, extensive contact with the viral surface. With a 1.4-Å-radius probe as a basis for measurement, the CD155 footprint covers 1,300 $Å^2$ of the PV surface. The north rim of the canyon is formed by the VP1 BC loop and the VP1 βE to αB loop. These residues form a hydrophobic surface that interacts with an equally extensive hydrophobic surface on CD155 (residues 29–31 at the amino terminus of the βA strand, residues 54 and 55 at the beginning of the BC loop, and residue 135 near the beginning of the βG strand).

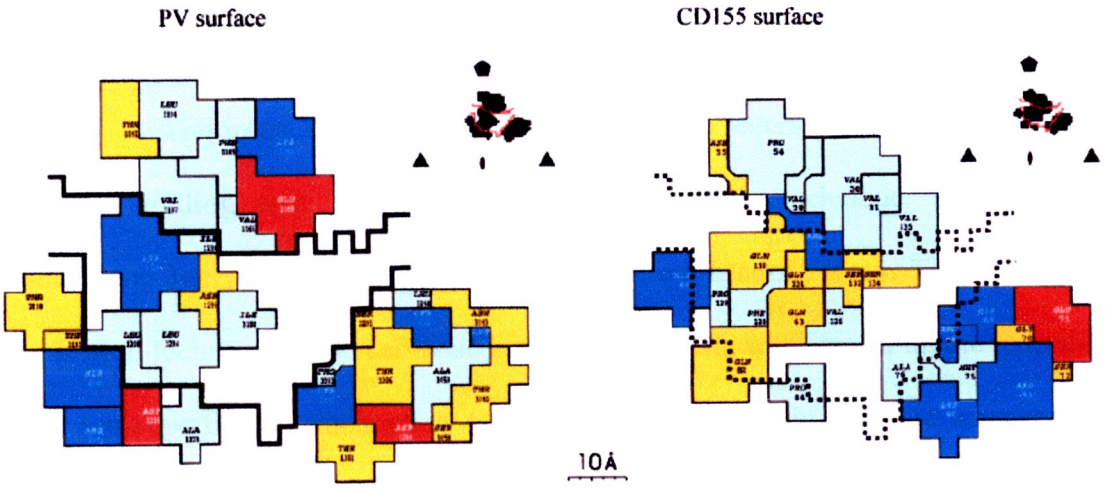


Figure 10 The footprint of the CD155 on the PV surface. (*Left*) The footprint on the virus (the canyon has a black outline). (*Right*) The residues of CD155 in contact with the viral surface (He et al., 2000).

Both PV loops that form the north rim of the canyon had been implicated correctly by mutagenesis studies as being involved in binding of CD155. The south rim of the canyon forms a hydrophilic surface in the ''puff'' (a large insertion in the GH loop of VP1. These residues contact the C'' strand of CD155, consistent with the mutational analysis of PV and of CD155. The floor of the canyon is formed primarily

by residues in the GH loop of VP1, the GH loop of VP3, and the end of the VP1 BC loop. These interact with residues (126–135) in the FG loop of CD155, consistent with the mutational analysis of the virus and of CD155 (He et al., 2000).

The viral surface in contact with the C, C', C'' surface of CD155 domain D1 is formed by the ''knob'' (an insertion in β-strand B of VP3) and the C-terminal region of VP1. Contacts in this region are strengthened by complementary electrostatic interactions between Glu-71 in CD155 and Lys-3062 and between Lys-90 and Arg-91 in CD155 and Asp-1298. Thus, the occurrence of some charge complementarity between PV and CD155 also may play a significant role in the recognition process (He et al., 2000).

The interaction between virus particle and receptor can be determined by receptor binding assay such as filter-based separation technique, fluorescent or radioligand receptor binding assay and competitive binding assay. The filter-based separation technique is the cell-based method. The cells were grown on the extremely low non-specific protein binding filtration plates contain hydrophilic microporous. The bound labeled ligand and free labeled ligand were separately determined. For competitive binding assay, the receptor analog is used to determine the dissociation constant of solid phase immobilized virus-receptor complex (Gambaryan and Matrosovich, 1992).

1.4.4.6. Application of poliovirus capsid proteins

A. Drug design

Attempts to develop antiviral agents against PV have been focused on 3 virus specific targets including proteins involved in viral RNA replication, capsid protein and the 3C protease. Guanidine and the benzimidazole

were shown to be specific inhibitors for RNA replication of picornaviruses, including poliovirus by blocking the initiation of (+)-strand RNA synthesis through its interaction with viral protein 3A (Heinz and Vance, 1995). The most extensively studied picornavirus antiviral class, capsid inhibitors, block virus replication by preventing virus uncoating and the release of viral RNA from the capsid into the cell. By integrating into a specific hydrophobic pocket in virus particles formed principally by capsid protein VP1, the drug increases the stability of the virion to the extent that capsid disassembly is unable to occur. Capsid inhibitors of varying chemicals that have been identified over the years include rhodamine, flavonoids, chalcones, aralkylamino pyridines, pyridazinamines, phenoxyl imidazoles and pyridazinylpiperidines (McKinlay et al., 1992; Carrasco, 1994; Rotbart et al., 1998). For picornavirus 3C protease, it is responsible for processing of the viral polyprotein into mature viral proteins. (Binford et al., 2005).

VP1 is the most-exposed surface protein of PV which plays a major role in the induction of neutralizing antibodies for all three PV serotypes. The antigenic structure of PV has been extensively studied by characterization of neutralizing IgG class monoclonal antibodies (MAbs). Genome sequencing of neutralization-resistant variants selected by MAbs has identified four major epitopes, designated as antigenic sites 1 through 4 (AgS1–AgS4), located on surface-exposed loops in structural proteins VP1, VP2 and VP3 (Minor, 1990). Safe and effective vaccines against poliovirus are available, so there has been little incentive for the development of anti-PV agents, although they may prove to be useful additives for stabilizing the existing Sabin attenuated polio vaccines. Although the anti-viral effects of these drugs are thought to derive primarily from their ability to prevent

conformational changes associated with uncoating, many of them also inhibit receptor attachment. Several crystallographic studies characterizing the interaction of poliovirus with a number of different antiviral compounds have shown that the drugs bind to a hydrophobic pocket in the core of the β barrel of the capsid protein, VP1. One end of the pocket is exposed on the surface of the capsid, at the bottom of the depression that surrounds each of its five-fold axes of icosahedral symmetry. This depression, called the 'canyon', is the site for receptor attachment. Drug binding induces significant alterations in the conformation of VP1 in the vicinity of the drug-binding site. It has been proposed that these conformational changes are responsible for the inhibition of receptor binding by the drug (Smith et al., 1986).

B. Ligand-receptor mediated drug targeting

In the past few years, ligand-receptor mediated cellular events have received major attention in the field of drug delivery. These endogenous ligands/epitopes mediated events could be exploited for designing site-specific and target-oriented delivery systems. Most carrier systems can be used for the site-specific delivery of various bioactives using biorelevant ligands, including antibodies, polypeptides, oligosaccharides (carbohydrates), viral proteins, fusogenic residues, and molecules of endogenous origin. These include blood carbohydrate (lectin) receptors, Fc receptors, complement receptors, interleukin receptors, lipoprotein receptors, transferin receptors, scavenger receptors, receptors/epitopes expressed on tumor cells, and cell adhesion receptors (Vyas et al., 2001). VP1 BC loop and VP1 βE to αB loop were indicated as the important part involved in PVR binding (He et al., 2000).

According to faecal-oral route infection of PV, these peptides might be the applicable tools for oral drug delivery system development based on ligand-receptor interaction.

1.4.5. Protein and peptide delivery systems

1.4.5.1. Barriers of protein and peptide delivery

A. Physical barriers: size, charge and solubility constraints

Paracellular protein transport across the aqueous channels and tight junctions between the epithelial cells is constrained by the physical properties of molecules. Size-dependent transport when charge is kept constant and the dependency on charge when size is the same is vastly indicative of the interplay of these physical properties, such that they all act in unison and in a mutually interdependent manner rather than as independent variables. The observation of higher permeation of net positively charged peptides further indicates an interactive environment wherein the penetrating peptide and protein interacts with lipids/proteins lining the aqueous pores (Pauletti et al., 1997). Lipophilic peptides, of course, would have a higher tendency to follow the transcellular diffusive route through the lipidic membrane.

B. Enzymatic barriers

The enzymatic barrier involves degradation either by hydrolysis of peptide bonds by exopeptidases or endopeptidases or by chemical modification such as oxidation, phosphorylation, or deamidation (Lee, 1991). The hydrolytic cleavage by peptidases is the most dominant degradation pathway, and this irreversible reaction leads to cleavage of peptides and proteins into amino acids and smaller peptide segments. The greatest threat to peptide and protein drugs lies in the small intestine, which contains gram quantities of peptidases secreted from the pancreas as well as

cellular peptidases from the mucosal cells, which are constantly sloughed off the villi. The second major enzymatic barrier is the brush-border membrane of the epithelial cells which can degrade both proteins and peptides. The colon also contains substantial amounts of peptidase activity, largely because of enzymatic production by microorganisms. In line with the primary functions of the GI tract in digestion and absorption of nutrients and exclusion of unwanted material, it secretes large amounts of enzymes specific to different regions of the GI tract (Table 3). Briefly, these include gastric pepsin (secreted as the inactive zymogen pepsinogen); intestinal endopeptidases (trypsin, chymotrypsin and elastase) and exopeptidase (aminopeptidases and carboxypeptidases), which have complementary cleavage type characteristics that enable them to digest permutations and combinations of amino acids found in natural proteins and peptides ingested as food. Endopeptidases hydrolyze the bond internal to the terminal bonds of the peptide chain, while exopeptidases hydrolyze the bond linking the NH_2-terminal or the COOH-terminal amino acid to the peptide chain. For example, luminal digestion in the duodenum by trypsin and chymotrypsin can convert almost half the ingested protein to trichloroacetic acid-soluble material within 10 minutes, 60–70% of which is in the form of small peptides with 2–6 amino acid residues. By the time the content leaves the jejunum, over 50% of ingested protein is digested to constituent amino acids. Another site for metabolism is the brush-border membrane, which contains enzymes such as alkaline phosphatase, sucrase, and a large number of peptidases (Silk et al., 1976). These peptidases also contribute to the low bioavailability of peptide drugs. Brush-border enzyme activity is generally greater in the duodenum and the jejunum than in the ileum, and membrane enzymes are very low in the colon. It is interesting

that the activity of pancreatic proteases is much lower toward small peptides than toward proteins.

Table 3a Bond specificities of secreted peptidases

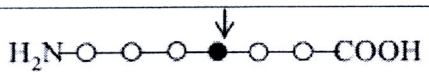
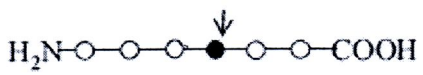
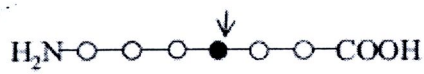
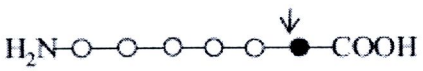
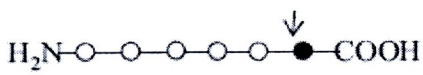
Enzyme		Specificity
Trypsin		Arg, Lys
Chymotrypsin		Phe, Tyr
Elastase		Als, Gly, Ile, Leu, Val, Ser
Carboxypeptidase A		Tyr, Phe, Ile, Thr, Glu, His, Ala
Carboxypeptidase B		Lys, Arg

Table 3b Intestinal brush-border membrane-bound peptidases

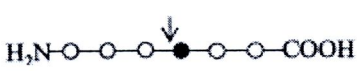
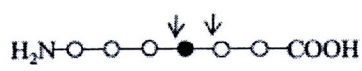
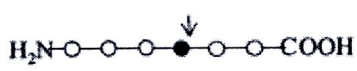
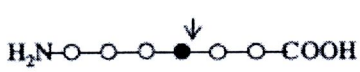
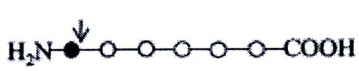
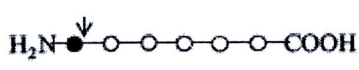
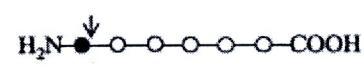
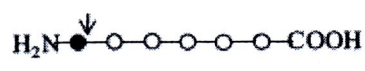
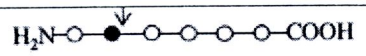
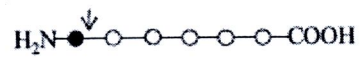
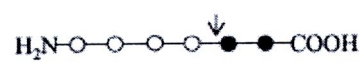
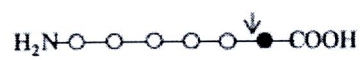
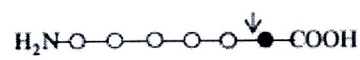
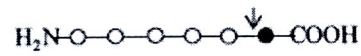
Enzyme		Specificity
Endopeptidases		
Endopeptidase-24.11		Hydrophobic amino acids
Endopeptidase-24.18		Aromatic amino acids
Endopeptidase-3		Arg-Arg?
Enteropeptidase		$(Asp)_4$-Lys
Exopeptidases; Amino-terminus		
Aminopeptidase N		Many
Aminopeptidase A		Asp, Glu
Aminopeptidase P		Pro
Aminopeptidase W		Trp, Tyr, Phe

Table 3b Intestinal brush-border membrane-bound peptidases (continued)

Dipeptidyl peptidase IV		Pro, Ala
γ-Glutamyl transpeptidase		γ-glutamic acid
Exopeptidases; Carboxy-terminus		
Peptidyldipeptidase A		His, Leu
Carboxypeptidase P		Pro, Gly, Ala
Carboxypeptidase M		Lys, Arg
γ-Glutamyl carboxypeptidase		γ-glutamic acid

Note: Specificity indicated by filled circle. (Reproduced with minor modifications from Bernkop-Schnurch and Walker, 1994)

Transporters in the brush-border membrane are designed to transport 2–3 amino acid peptides into the cells, where they are further degraded into constituent amino acids. The peptidase activity characteristics are different between the brush-border membrane and cytosol. While cytoplasmic peptidases hydrolyze di/tripeptidases, oligopeptides of over four amino acids are hydrolyzed primarily by the brush-border membrane peptidases, which are anchored in the apical membrane at their N-terminus such that the active site is in the extracellular environment. Membrane-bound aminopeptidases have been implicated in the cleavage of metkephamid, a pentapeptide enkephalin analog, limiting its absorption. Thus, small di/tripeptides (including β-lactam antibiotics and ACE inhibitors) are transported transcellularly by a carrier-mediated transport process, while tetra- or larger peptides are hydrolyzed by brush-border membrane peptidases to di/tripeptides and amino

acids, which may then be absorbed by simple diffusion or endocytosis (Bai and Amidon, 1992).

Furthermore, degradation within the subcellular organelles becomes important in some special cases such as carrier-mediated uptake, wherein the endosomally internalized peptide drug would be destined to lysosomal degradation. Lysosomes contain a wide variety of enzymes, which are capable of degrading any macromolecules, including peptides and proteins. Lysosomal enzymes include a group of peptidases known as *cathepsins*. Among the group of peptidases, there are at least eight exopeptidases and at least nine endopeptidases. These enzymes act synergistically to degrade a variety of peptide and protein drugs to free amino acids. An interesting example of the importance of oral peptide delivery is the aquaculture industry. Although fish intestine does absorb small amounts of peptides and proteins, systemic bioavailability is compromised because of low mucosal permeability and metabolic barriers. Using bovine serum albumin (BSA) and human and salmon leutinizing hormone releasing hormone (LHRH) as model drugs (Ledger et al., 2002), it was determined that proteolytic activities were different in the different regions of the intestine and that approaches were needed to overcome the proteolytic activity related peptide degradation.

1.4.5.2. Strategy for peptide delivery

A. Chemical modification

Chemical modifications of peptide and protein drugs have been shown to improve their enzymatic stability and/or membrane penetration of peptides and proteins. Chemical modifications can also be used for minimizing immunogenicity. Protein modification can be done either by direct modification of

exposed side-chain amino acid groups of proteins or through the carbohydrate part of glycoproteins and glycoenzymes (Barbaric et al., 1988). The latter approach provides the advantages both of being applicable even when the protein sample is not very pure and of not interfering with the normal structure of the protein. Soluble glycoenzyme derivatives and other modifications of physicochemical and possibly also catalytic properties can thus be achieved. Use of D-amino acids at one or more positions in a peptide drug and the modification of terminal active groups or amino acids are some of the other strategies that have been attempted. The replacement of terminal carboxylic group with an amide group did not result in increase in the stability of tetrapeptide enkephalin analogues in jejunal tissue. Enkephalins are endogenous opioid peptides of significant clinical importance, which also include β-endorphin, dynorphin, and endomorphins. Their clinical importance comes in symptomatic pain modulation. The two types of naturally occurring enkephalins, leucine (Tyr-Gly-Gly-Phe-Leu [YGGFL]) and methionine (Tyr-Gly-Gly-Phe-Met [YGGFM]), derived from the catabolism of α-endorphin, are both orally inactive. Low oral bioavailability of these peptides has been attributed to high peptidase sensitivity. Structural modifications of methionine enkephalin pentapeptide YGGFM, such as NH_2-Tyr-Gly-Gly-Phe-Met-COOH, were done by formation of carbinol analog by alteration of the terminal carboxyl group, substitution of $glycine_2$ by D-alanine, combining these two modifications in the same peptide, creating a sulfoxide terminal instead of the carboxyl, and its *N*-methylation. Only two of these modifications, sulfoxide at C-terminal and sulfoxide modification combined with *N*-methylation, produced significant analgesic activity after oral administration (Roemer et al., 1977). Improved biological responses have also been achieved for enkephalins by designing

its synthetic analog, L-tyrosyl-D-methionyl-L-glycyl-4-nitrophenylaryl-L-prolinamide-S-oxide hydrochloride (BW 942C, nifaltide), which is resistant to intestinal digestion and metabolism (Berschneider et al., 1988).

B. Protease inhibitors co-administration

Co-administration of protease inhibitors provides a viable means to circumvent the enzymatic barrier in achieving the delivery of peptide and protein drugs. The choice of protease inhibitors will depend on the structure of these therapeutic drugs, and the information on the specificity of proteases is essential to guarantee the stability of the drugs in the GI tract (Bernkop-Schnurch, 1998). A number of inhibitors including aprotinin (trypsin/chymotrypsin inhibitor), amastatin, bestatin, boroleucine, and puromycin (aminopeptidase inhibitors) have been reported for this purpose. Yamamoto et al. studied the effect of five protease inhibitors: sodium glycocholate, camostat mesilate, bacitracin, soybean trypsin inhibitor, and aprotinin on the intestinal metabolism of insulin in rats (Yamamoto et al., 1994). Among these peptide inhibitors, sodium glycocholate, camostat mesilate, and bacitracin were more effective in improving the physiological availability of insulin in the large intestine than in the small intestine. However, none of these protease inhibitors were effective in the small intestine, possibly because of the numerous enzymes secreted here. When the inhibitors were used to determine the pH-dependent inhibition of insulin degradation in rat cecal contents, their effectiveness was in the following order: aprotinin=camostat>soyabean trypsin inhibitor>sodium glycocholate. The coadministration of 4-(4-isopropylpiperadinocarbonyl) phenyl 1,2,3,4-tetrahydro-1-naphthoate methanesulphonate (FK-448), a low-toxicity inhibitor of chymotrypsin, significantly increased intestinal absorption of insulin after oral

administration in rats and dogs, resulting in a decrease in blood glucose levels. Similarly, coadministration of FK 448 with insulin showed a sharp drop in blood glucose levels in healthy volunteers. In the presence of peptidase inhibitor amastatin, the pentapeptide enkephalin YAGFL [Tyr-Ala-Gly-Phe-Leu], with D-conformation of alanine and leucine amino acids, was shown to have 22 times higher bioavailability by the oral administration than without the peptidase inhibitor (Lee and Amidon, 2002). However, the use of protease inhibitors may also affect the absorption of other peptides or proteins that normally would be degraded. A major drawback of these inhibitors is their high toxicity especially in chronic drug therapy. In addition, the non-site specific intestinal application of such compounds will markedly change the metabolic pattern in the GI tract because of reduced digestion of food proteins.

C. Absorption enhancers

Because of their large molecular weight and hydrophilic nature, absorption of peptide and protein drugs through transcellular and paracellular routes is severely restricted. One possible way to improve oral bioavailability of these drugs is to make the absorption surface more permeable. Absorption enhancers are formulation components that temporarily disrupt the intestinal barrier to improve the permeation of these drugs. Absorption enhancers can be grouped into chelators, surfactants, bile salts and fatty acids and their derivatives. Possible mechanisms of action for absorption enhancers include an increase in membrane fluidity, a decrease in mucus viscosity, leakage of proteins through membranes and opening of tight junctions. The chelators are believed to exert their action by complex formation with calcium ions and by facilitating permeation of proteins from the membrane. Delivery of peptide and protein drugs to the absorptive surface of the mucosa is increased by

micellar protection or by disruption of the membrane integrity by solubilization of membrane phospholipids. However, permeation enhancers often induce toxic side effects. For example, calcium chelators can lead to Ca^{2+} depletion which can cause global changes in the cells, including disruption of actin filaments, disruption of adherent junctions and diminished cell adhesion (Sood and Panchagnula, 2001). The use of absorption enhancers is often required to improve oral peptide and protein absorption, which was amply exemplified by the oral absorption of human calcitonin in rat. It has been reported that calcitonin administered alone was hardly absorbed to elicit hypocalcemia, whereas its co-administration with surfactants, such as sodium myristate, sodium deoxycholate, sodium taurodeoxycholate, sodium lauryl sulfate, quillajasaponin, and sugar esters, resulted in significant hypocalcemia. One recently developed option for the use of absorption enhancers is to administer protein and peptide drugs in concentrated solutions of so-called "carrier" molecules. These molecules are diverse in structure (Figure 11) and promote protein and peptide drug absorption. The mechanism of action of these agents is still not clear and efforts are being made to explore the same. Leone-Bay et al. reported the use of *N*-acetylated, non-α aromatic amino acids and *N*-acylated, α-amino acids to promote oral delivery of a variety of therapeutic proteins, such as sCT, interferon α, insulin and recombinant human growth hormone (rhGH) (Leone-Bay et al., 2001). Among several compounds tested to facilitate the oral absorption of rhGH in rats, 4-(4-(2-hydroxybenzoyl) aminophenyl) butyric acid (IX) was found to be the most efficacious with a mean peak serum concentration of 55 ng/ml of rhGH in monkeys, while oral administration of rhGH alone showed very low serum concentrations of this protein (Leone-Bay et al., 1996). Similarly, co-administration of *N*-(phenylsulfonyl)-α-amino acids (X), *N*-

benzoyl-α-amino acids (XI) and derivatized phenylglycines (XIII) was found to increase intestinal absorption of sCT and interferon-α in rats and monkeys.

D. Formulation vehicles

To overcome different biological barriers, several formulation strategies including emulsions, microemulsions, hydrogels, nanoparticles, coated liposomes and mucoadhesive polymers are being investigated.

1. Emulsions

Cyclosporine is an atypical peptide in that it possesses significant oral activity as an immunosuppressant. Cyclosporine is resistant to proteolytic degradation and has high hydrophobicity with a partition coefficient of about 991 and aqueous solubility of about 6 μg/ml resulting from seven hydrophobic *N*-methylated amino acid residues in a cyclic structure (Cefalu and Pardridge, 1985). Extensive hepatic metabolism and poor aqueous solubility have been implicated in slow and incomplete absorption and overall low oral bioavailability. Therefore, cyclosporine was formulated with peglicol-5-oleate (Labrafel), olive oil and ethanol at a ratio of 30:60:10, followed by aqueous dilution for emulsification. Reduction in the emulsion droplet size from 4 to 2 μm increased absorption of cyclosporine A by two-fold in rats (Tarr and Yalkowsky, 1989). This can be explained in terms of increased surface area for partitioning of cyclosporine into the surrounding aqueous environment and increased lipase activity resulting in higher solubility in the GI milieu and a higher driving force for penetration across the unstirred aqueous layer of the intestinal mucosa. Absorption of cyclosporine has also been shown to vary with the absorbing site of the GI tract.

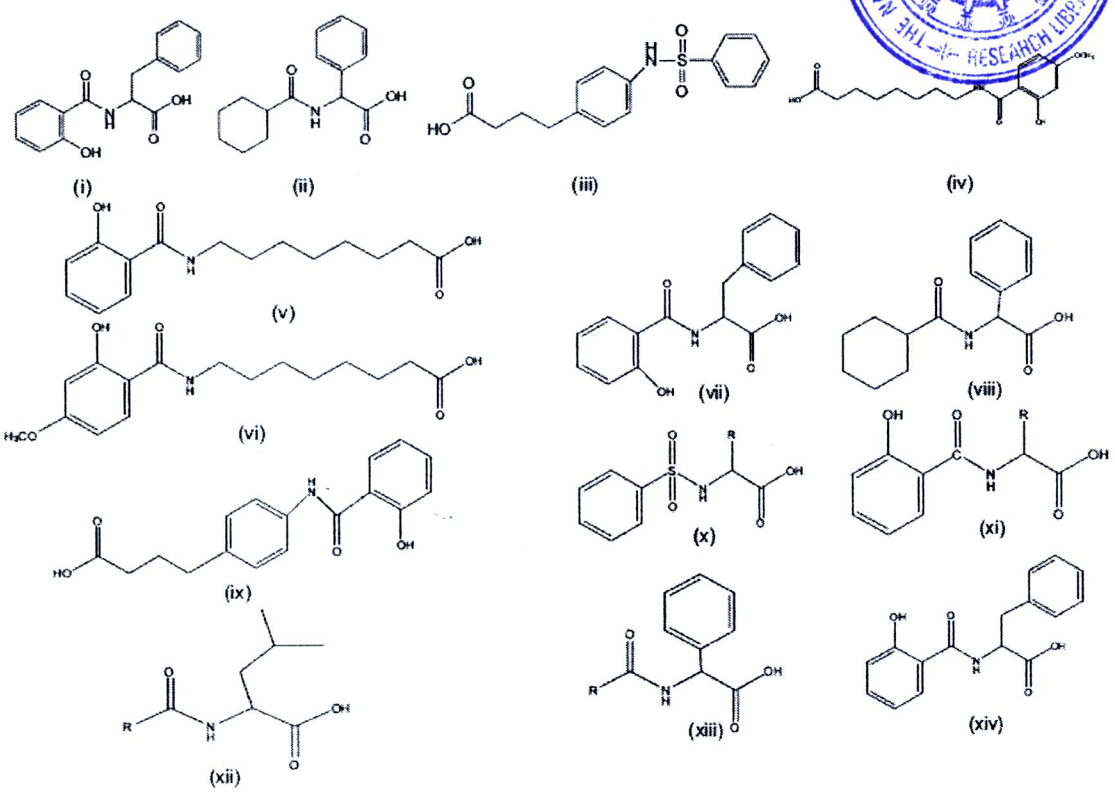


Figure 11 Structures of "carrier" molecules used for enhancing oral bioavailability of protein and peptide drugs (Leone-Bay et al., 2001; Sood and Panchagnula, 2001).

2. Hydrogels

The influence of pH variability through the stomach to the intestine on the oral bioavailability of peptide and protein drugs may be overcome by protecting them from proteolytic degradation in the stomach and upper portion of the small intestine using pH-responsive hydrogels as oral delivery vehicles. Lowman et al., for example, loaded insulin into polymeric microspheres of poly(methacrylic-*g*-ethyleneglycol) and observed oral bioavailability in healthy and diabetic rats. In the acidic environment of the stomach, the gels were unswollen as a result of the formation of intermolecular polymer complexes. The insulin remained in the gel and was protected from proteolytic degradation. While in the basic and neutral environments of the intestine,

the complexes dissociated resulting in rapid gel swelling and insulin release. Within 2 hours of administration of the insulin-containing polymers, strong dose-dependent hypoglycemic effects were observed in both healthy and diabetic rats (Lowman et al., 1999). Numerous pH-sensitive polymers have been investigated for a range of applications.

3. Polymeric Particulate Systems

Biodegradable, aqueous-cored nanoparticles suitable for the entrapment of proteins and peptides can be prepared by interfacial polymerization of water-in-oil (w/o) dispersions using alkylcyanoacrylates. However, to obtain small-size dispersion often requires shear forces which may degrade peptide and protein drugs. This requirement of a high input of energy may be overcome by the use of microemulsions, which are formed spontaneously and are thermodynamically stable, with the dispersed systems of a uniform size of less than 200 nm (Watnasirichaikul et al., 2000).

4. Nanovesicles

Nanovesicular systems such as liposomes and niosomes have been widely used as drug carriers due to their advantages including stability enhancement of the entrapped substances and the ability to cross the cell membrane (Biju et al., 2006). Liposomes (phospholipid vesicles) have been extensively studied over the last three decades as a potential drug delivery vehicle. The popularity of liposomes as drug carriers partly arisen from their biodegradability and the possibility of manipulating their structure to induce specificity and cell targeting. Liposomes have been used as a vehicle for delivering enzymes to defective cells where a specific enzyme deficiency exists (Ryman and Tyrrell, 1980). Also, it is possible to use liposomes to monitor a

slow release of the drug to the system. Other areas of using liposomes for drug delivery include their use as chelating agents as a mean of treating heavy metal poisoning and entrapment of hormones, for example cortisol ester, for intra-articular delivery in arthritis (Bundgaard et al., 1982).

E. Cell penetrating peptides (CPPs)

To use CPPs as vectors for protein delivery can have many advantages over the more traditional approach to endogenously express the desired protein from a plasmid such as difficulties with transfection and varied protein expression in various cell lines. Several proteins have been transported into cells by CPP mediated translocation including proteins involved in apoptosis, cell cycle regulation and DNA recombination (Wadia et al., 2004). By designing a plasmid coding for of a CPP in coding frame with the desired protein, a cell permeable fusion protein is produced. This method has successfully been applied both *in vivo* and *in vitro*. Addition of a highly cationic peptide in the same amino acids chain as a protein may interfere with the proteins tertiary structure and thereby cause loss of biological function. This problem can be overcome by simply co-incubate a CPP and the desired protein and thus facilitate cellular uptake (Morris et al., 2001). The protein thereby keeps its native structure and the proteins biological properties are conserved.

1.4.6. Protein expression

1.4.6.1. Fundamental techniques

A. Polymerase chain reaction (PCR)

The polymerase chain reaction (PCR) is a scientific technique in molecular biology to amplify a single or a few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular

DNA sequence. The method relies on thermal cycling, consisting of cycles of repeated heating and cooling of the reaction for DNA melting and enzymatic replication of the DNA. Primers (short DNA fragments) containing sequences complementary to the target region along with a DNA polymerase (after which the method is named) are key components to enable selective and repeated amplification. As PCR progresses, the DNA generated is itself used as a template for replication, setting in motion a chain reaction in which the DNA template is exponentially amplified. PCR can be extensively modified to perform a wide array of genetic manipulations.

PCR is used to amplify a specific region of a DNA strand (the DNA target). Most PCR methods typically amplify DNA fragments of up to ~10 kilo base pairs (kb), although some techniques allow for amplification of fragments up to 40 kb in size (Cheng et al., 1994). A basic PCR set up requires several components and reagents. These components include:

- *DNA template* that contains the DNA region (target) to be amplified.
- Two *primers* that are complementary to the 3' (three prime) ends of each of the sense and anti-sense strand of the DNA target.
- *Taq polymerase* or another DNA polymerase with a temperature optimum at around 70 °C.
- *Deoxynucleoside triphosphates* (dNTPs; also very commonly and erroneously called deoxynucleotide triphosphates), the building blocks from which the DNA polymerases synthesizes a new DNA strand.

- *Buffer solution,* providing a suitable chemical environment for optimum activity and stability of the DNA polymerase.
- *Divalent cations,* magnesium or manganese ions; generally Mg^{2+} is used, but Mn^{2+} can be utilized for PCR-mediated DNA mutagenesis, as higher Mn^{2+} concentration increases the error rate during DNA synthesis.
- *Monovalent cation* potassium ions.

The PCR is commonly carried out in a reaction volume of 10–200 µl in small reaction tubes (0.2–0.5 ml volumes) in a thermal cycler. The thermal cycler heats and cools the reaction tubes to achieve the temperatures required at each step of the reaction. Thin-walled reaction tubes permit favorable thermal conductivity to allow for rapid thermal equilibration. Most thermal cyclers have heated lids to prevent condensation at the top of the reaction tube. Older thermocyclers lacking a heated lid require a layer of oil on top of the reaction mixture or a ball of wax inside the tube.

Typically, PCR consists of a series of 20-40 repeated temperature changes, called cycles, with each cycle commonly consisting of 2-3 discrete temperature steps, usually three (Figure 12). The cycling is often preceded by a single temperature step (called *hold*) at a high temperature (>90°C), and followed by one hold at the end for final product extension or brief storage. The temperatures used and the length of time they are applied in each cycle depend on a variety of parameters. These include the enzyme used for DNA synthesis, the concentration of divalent ions and dNTPs in the reaction, and the melting temperature (Tm) of the primers (Rychlik et al., 1990).

- *Initialization step*: This step consists of heating the reaction to a temperature of 94–96 °C, which is held for 1–9 minutes. It is only required for DNA polymerases that require heat activation by hot-start PCR.
- *Denaturation step*: This step is the first regular cycling event and consists of heating the reaction to 94–98 °C for 20–30 seconds. It causes DNA melting of the DNA template by disrupting the hydrogen bonds between complementary bases, yielding single-stranded DNA molecules.
- *Annealing step*: The reaction temperature is lowered to 50–65 °C for 20–40 seconds allowing annealing of the primers to the single-stranded DNA template. Typically the annealing temperature is about 3-5 degrees Celsius below the Tm of the primers used. Stable DNA-DNA hydrogen bonds are only formed when the primer sequence very closely matches the template sequence. The polymerase binds to the primer-template hybrid and begins DNA synthesis.
- *Extension/elongation step*: The temperature at this step depends on the DNA polymerase used; Taq polymerase has its optimum activity temperature at 75–80 °C, and commonly a temperature of 72 °C is used with this enzyme. At this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTPs that are complementary to the template in 5' to 3' direction, condensing the 5'-phosphate group of the dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. The extension time depends both on the DNA polymerase used and on the length of the DNA fragment to be amplified. Under optimum conditions, i.e., if there are no limitations due to limiting substrates or reagents, at each

extension step, the amount of DNA target is doubled, leading to exponential (geometric) amplification of the specific DNA fragment.

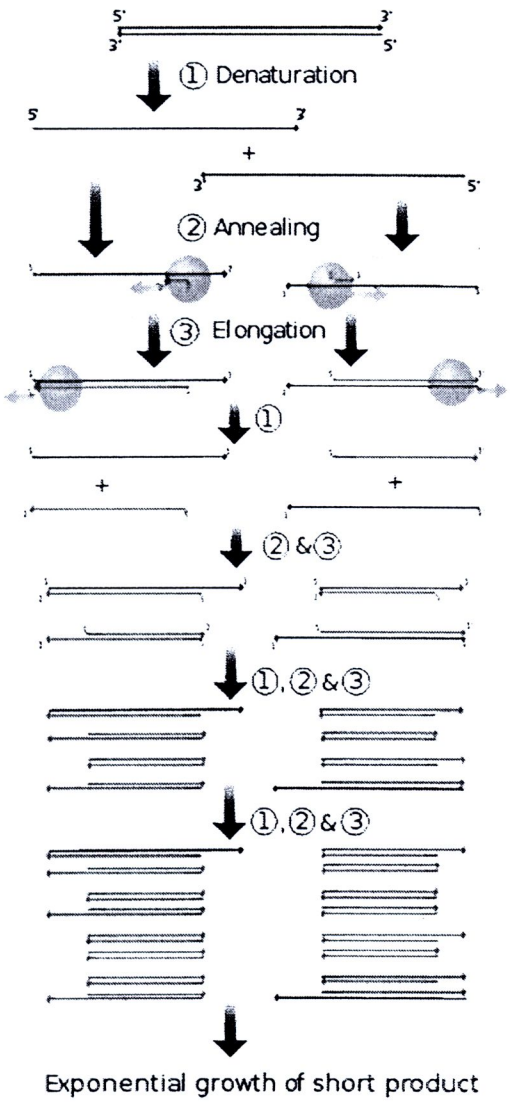


Figure 12 Schematic drawing of the PCR cycle. (1) Denaturing at 94–96 °C. (2) Annealing at ~65 °C (3) Elongation at 72 °C. Four cycles are shown here (Cheng et al., 1994).

- *Final elongation*: This single step is occasionally performed at a temperature of 70-74 °C for 5-15 minutes after the last PCR cycle to ensure that any remaining single-stranded DNA is fully extended.
- *Final hold*: This step at 4-15 °C for an indefinite time may be employed for short-term storage of the reaction.

To check whether the PCR generated the anticipated DNA fragment (also sometimes referred to as the amplimer or amplicon), agarose gel electrophoresis is employed for size separation of the PCR products. The size(s) of PCR products is determined by comparison with a DNA ladder (a molecular weight marker), which contains DNA fragments of known size, run on the gel alongside the PCR products.

B. Agarose gel electrophoresis

Agarose gel electrophoresis is one of several physical methods for determining the size of DNA. In this method, DNA is forced to migrate through a highly cross-linked agarose matrix in response to an electric current. In solution, the phosphates on the DNA are negatively charged and the molecule will therefore migrate to the positive pole. There are three factors that affect migration rate through a gel; size of the DNA, conformation of the DNA, and ionic strength of the running buffer.

Electrophoresis is essentially a sieving process. The larger the fragment of DNA, the more easily it become entangled in the matrix and, therefore, the more slowly will it migrate. Small fragments, therefore, run more quickly than large fragments at a rate proportional to their size. The relationship of size to migration rate is linear throughout most of the gel, except for the very largest

fragments. Large fragments have a more difficult time penetrating the gel and their migration, therefore does not have a linear relationship to size. The matrix can be adjusted by increasing the concentration of agarose (tighter matrix) or by decreasing it (looser matrix). A standard 1% agarose gel can resolve DNA from 0.2 – 30 kb in length (Brody and Kern, 2004).

C. SDS-PAGE

SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis is a technique widely used in biochemistry, forensics, genetics and molecular biology to separate proteins according to their electrophoretic mobility (a function of length of polypeptide chain or molecular weight).

The solution of proteins to be analyzed is mixed with SDS, an anionic detergent which denatures secondary and non–disulfide–linked tertiary structures, and applies a negative charge to each protein in proportion to its mass. Heating the samples to at least 60 degrees C shakes up the molecules, helping SDS to bind (Shapiro et al, 1967). A tracking dye may be added to the protein solution (of a size smaller than protein) to allow the experimenter to track the progress of the protein solution through the gel during the electrophoretic run.

The gels generally consist of acrylamide, bisacrylamide, SDS, and a Tris-Cl buffer with adjusted pH. Ammonium persulfate and TEMED are added when the gel is ready to be polymerized. The separating or resolving gel is usually more basic and has a higher polyacrylamide content than the loading gel. Gels are polymerized in a gel caster. First the separating gel is poured and allowed to polymerize. Next a thin layer of isopropanol is added. Next the loading gel is poured and a comb is placed to create the wells. After the loading gel is polymerized the

comb can be removed and the gel is ready for electrophoresis. The anode buffer usually contains Tris-Cl, distilled deionized water and is adjusted to a higher pH than the cathode buffer. The cathode buffer contains SDS, Tris, Tricine, and distilled deionized water. The electrophoresis apparatus is set up with cathode buffer covering the gel in the negative electrode chamber, and anode buffer in the lower positive electrode chamber. Next, the denatured sample proteins are added to the wells one end of the gel with a syringe or pipette. Finally, the apparatus is hooked up to a power source under appropriate running conditions to separate the protein bands (Figure 13). An electric field is applied across the gel, causing the negatively-charged proteins to migrate across the gel towards the positive (+) electrode (anode). Depending on their size, each protein will move differently through the gel matrix: short proteins will more easily fit through the pores in the gel, while larger ones will have more difficulty (they encounter more resistance). After a set amount of time (usually a few hours depends on the voltage applied across the gel; higher voltages run faster but tend to produce somewhat poorer resolution), the proteins will have differentially migrated based on their size; smaller proteins will have traveled farther down the gel, while larger ones will have remained closer to the point of origin. Therefore, proteins may be separated roughly according to size (and therefore, molecular weight), certain glycoproteins behave anomalously on SDS gels. Following electrophoresis, the gel may be stained (most commonly with Coomassie Brilliant Blue R-250 or silver stain), allowing visualization of the separated proteins, or processed further (e.g. Western blot).

After staining, different proteins will appear as distinct bands within the gel. It is common to run molecular weight size markers of known molecular weight in a separate lane in the gel, in order to calibrate the gel and determine the weight of unknown proteins by comparing the distance traveled relative to the marker. The acrylamide concentration of the gel can also be varied, generally in the range from 5% to 25%. Lower percentage gels are better for resolving very high molecular weight proteins, while much higher percentages are needed to resolve smaller proteins. Determining how much of the various solutions to mix together to make gels of particular acrylamide concentration can be done (Schägger and von Jagow, 1987).

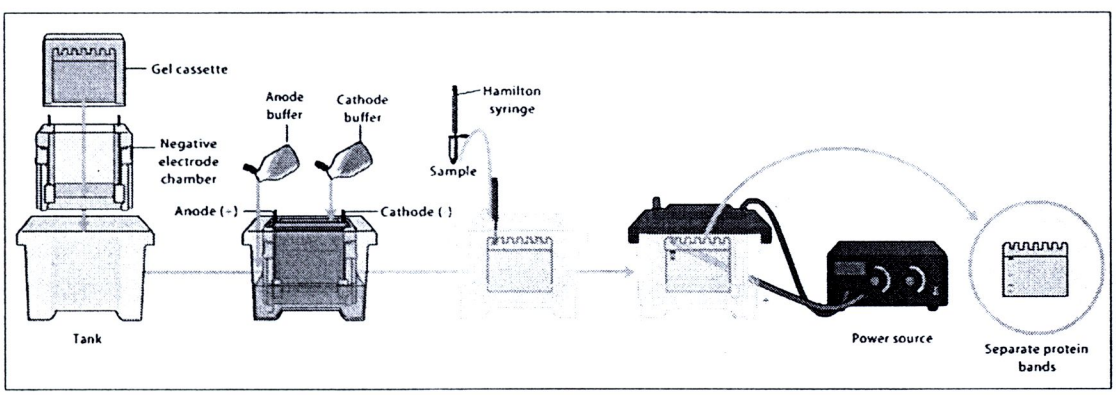


Fig. 13 SDS-PAGE schematic, start from gel preparation until electrophoresis running (Shapiro et al, 1967)

D. Western blot analysis

The Western blot (alternatively, protein immunoblot) is an extremely useful analytical technique used to detect specific proteins in a given sample of tissue homogenate or extract. It uses gel electrophoresis to separate native or denatured proteins by the length of the polypeptide (denaturing conditions) or by the 3-D structure of the protein (native/ non-denaturing conditions). The proteins are

then transferred to a membrane (typically nitrocellulose or PVDF) where they are probed (detected) using antibodies specific to the target protein (Towbin et al., 1979). There are now many reagent companies that specialize in providing antibodies (both monoclonal and polyclonal antibodies) against tens of thousands of different proteins. Commercial antibodies can be expensive, although the unbound antibody can be reused between experiments. This method is used in the fields of molecular biology, biochemistry, immunogenetics and other molecular biology disciplines. Other related techniques include using antibodies to detect proteins in tissues and cells by immunostaining and enzyme-linked immunosorbent assay (ELISA).

1.4.6.2. Generation of expression plasmids

A. Multi-parallel molecular cloning strategies

Standard procedure when setting out to express a recombinant protein is to screen a series of constructs to identify the most appropriate vehicle for the generation of sufficient soluble protein for downstream purposes. This may include expressing full-length proteins or perhaps specific domains or chimeric proteins. A series of fusion partners may also be investigated for their effects on driving enhanced expression or their capacity to capture and purify the target protein quickly with minimal impurities. In some cases, information is known about the target protein from either published literature or a priori knowledge of a specific protein family. In this case, only a few constructs may need to be generated using a single expression system. However, in many more cases where little is known concerning the gene and the expressed protein, the number of possible constructs can rise exponentially. Using traditional cloning methodologies the generation of the many possible combinations and their analysis in different expression systems would

be so labour intensive and time consuming as to make a parallel strategy of expression screening impractical. Therefore, the traditional mantra has been to adopt an intuitive approach and generate only a handful of constructs in the first instance, analyse, and then re-optimise and repeat the process again until the most appropriate conditions are identified. In most instances, *E. coli* would be tried first and then a second system (typically baculovirus) if expression levels were found to be low or solubility of the target protein an issue (in *E. coli*). Exceptions to this approach would be in cases where literature had previously demonstrated the utility of baculovirus or where complex large multi-domain proteins or post-translational modification were desirable. Furthermore, whilst purification and refolding of proteins derived from *E. coli* can also be attempted, this can often prove more time consuming than baculovirus expression. It is also highly unlikely to prove successful with larger more complex proteins. In these cases, the use of *E. coli* becomes impractical and baculovirus becomes the expression system of choice.

However, over the past few years many of the limitations relating to the generation of multiple expression plasmids (and constructs) have been addressed by a number of elegant recombinatorial cloning systems that enable the rapid cloning of potentially hundreds of genes and constructs simultaneously. Developed initially to facilitate analysis of gene pathways in the field of genomics (Hartley et al., 2000), the technology has rapidly proven to be a highly versatile tool in fields where traditional methods of manipulating individual clones have proven to be highly inefficient when working with many genes simultaneously. This technology utilises a recombination process observed in lambda phage to facilitate transfer of heterologous DNA sequences between two vectors. The reaction is catalysed by a mixture of

enzymes that bind to specific signal sequences and covalently recombine the DNA into novel forms (Landy, 1989). The major strength of this system is its high-recombination efficiency, reversibility and directionality.

B. Choice of tag

The use of such a multi-parallel expression strategy to conduct optimisation and screening studies (whether conducted in a prokaryotic or eukaryotic host) necessitates the adoption of a generic purification approach that is quick, simple, and cheap to perform. Several different tags have therefore been developed to facilitate rapid single-step purification of proteins that are commercially available (Stevens, 2000). A list of some of the commonly used tags is shown in Table 4. By far, the most popular are glutathione *S*-transferase (GST) and polyhistidine (His6) tags (Smith and Thorner, 2000), although in both cases there are potential drawbacks to their uses. In the case of GST, the size (27 kDa) can be somewhat prohibitive, whilst the slow binding kinetics of GST to glutathione–sepharose resin make loading of cell extracts extremely time consuming especially when large cell culture volumes are being processed. Furthermore, when using baculovirus-directed insect cell expression, host GST proteins are readily co-purified using standard purification procedures. Polyhistidine tags on the other hand are extremely small and therefore do not in most cases affect the folding of the attached protein. It also has very strong reversible binding attributes allowing for a very rapid and single-step purification. The tag usually consists of six consecutive histidine residues, but can vary in length from 2 to 10. Indeed, there has been great debate with regard to the length of the polyhistidine repeat and their respective capacities to bind metal affinity resins. Recent data revealed that there is no difference between the length of the repeat and

the purification level (Mohanty and Wiener, 2004). Polyhistidine tags can be placed on either the N- or C-termini of recombinant proteins, although the optimal location does vary depending on the folding and biochemical characteristics of the adjacent recombinant protein. Whilst ideal for extraction of most intracellular proteins in a variety of expression systems (*E. coli*, baculovirus, and mammalian), the purification of polyhistidine-tagged proteins secreted from insect cells is somewhat more problematic, since the media require extensive dialysis prior to column purification in order to remove contaminating components which would otherwise prevent the binding of the protein to the resin.

Ubiquitin (Ub) is another fusion partner that has been used for several years to enhance protein expression levels. Ubiquitin is a small eukaryotic protein offering solubility enhancing properties when fused to an adjacent protein, which can be removed by cleavage with a group of proteases known as deubiquitylating enzymes (DUBS). DUBS are highly specific and do not cleave non-specific sequences or leave additional amino acids at the N-terminus of the protein of interest (Baker, 1996).

Table 4 Comparison of commonly used tags for purification and expression enhancement (Smith and Thorner, 2000)

Tag	Residue	Size (kDa)	Use	Matrix/elution
Polyhistidine	2-10	0.84	Purification	Divalent metal/ imidazole or low pH
Maltose binding protein	396	40	Purification and enhanced solubility	Amylose resin, elution with maltose
Glutathione S-transferase	211	26	Purification and enhanced solubility	Glutathione agarore, elution with glutathione

Table 4 Comparison of commonly used tags for purification and expression enhancement (Smith and Thorner, 2000) (continued)

Tag	Residue	Size (kDa)	Use	Matrix/elution
Intein CBD	51	5.59	Purification	Chitin
Strep-tag II	8	1.20	Purification	Strep-Tactin/biotin or desthiobiotin
S-tag	15	1.75	Detection and purification	S-fragment/low pH
NusA	495	54.87	Increased expression and solubility	NA
Trx	109	11.67		
Ubiquitin (Ub)	128	14.73	Solubility	NA
GFP	239	27	Detection	NA

1.4.6.3. Expression system

A. *Escherichia coli*-mediated protein expression

Bacterial expression is perhaps the most commonly employed expression vehicle for the production of recombinant proteins. The organism is relatively simple to manipulate, inexpensive to culture, and the amount of time necessary to generate a recombinant protein is short. However, since it is a prokaryotic based system, heterologously expressed eukaryotic proteins are not modified correctly and it can also be difficult to facilitate the secretion of large amounts of expressed protein. Furthermore, proteins expressed in large amounts can precipitate, forming inclusion bodies, whilst large complex proteins can also be difficult to propagate. Recombinant expression of proteins is normally achieved through the induction of a strong promoter system. Several of them are commonly used including T7, lambda Pl and *ara*B. The most popular in recent years have been

the T7 based pET expression plasmids (commercially available from Merck Biosciences). However, whilst this system leads to the generation of large amounts of mRNA and concomitant protein expression, the high levels of mRNA can cause ribosome destruction and cell death. Furthermore, leaky expression of T7 RNA polymerase may result in plasmid or expression instability. Use of the lac operator and T7 lysozyme (pLysS) can provide an extra level of repression (Makrides, 1996).

Many *E. coli* strains optimised for protein expression purposes are commercially available from suppliers such as Invitrogen, Merck Biosciences and Stratagene. The Origami and Origami B strains (Merck Biosciences) have been developed to express proteins that require disulphide bonds to achieve their active, correctly folded conformation. These strains with mutations in both the thioredoxin reductase (*trx*B) and glutathione reductase (*gor*) genes greatly enhance disulphide bond formation in the cytoplasmic space. Furthermore, strains such as BL21 CodonPlus (Stratagene) and Rosetta-2 derived from BL21 (Merck Biosciences) have been reported to enhance expression of gene sequences that contain rarely used *E. coli* codons in the expression of heterologous protein in *E. coli* (Gustafsson et al., 2004).

B. Baculovirus-mediated insect cell protein expression

Baculovirus-mediated insect cell expression has become one of the most popular vehicles for the production of large quantities of recombinant protein for structural and functional studies of therapeutically relevant bio-molecules. Baculovirus protein expression is a eukaryotic based expression system and thus offers protein modification and processing patterns similar (but not identical) to those in higher eukaryotic cells. In addition, the baculovirus-mediated system uses a helper-independent virus that can be propagated to very high titres and is easily

adapted to suspension culture, making it possible to generate large amounts of recombinant protein with relative ease. To date, many hundreds of proteins from a variety of organisms have been expressed using the system, making it a highly valuable tool in the protein production laboratory. However, the adaptation of the system to multi-parallel expression is much more difficult to achieve than with *E. coli* due to its far more complex nature (Hunt, 2005).

C. Other expression systems

- *Yeast*

Whilst yeast, in particular *Pichia pastoris*, offers a very powerful option in the secretion of proteins, it has lagged somewhat behind other systems in its adoption to many laboratories due to its technical and time-consuming requirements, precluding its use in multi-parallel applications. Nevertheless, the system offers a very powerful alternative to baculovirus or mammalian expression systems, specifically in regard to the generation of large quantities of secreted material (Holz et al., 2002).

- *Mammalian cell expression*

The expression of heterologous proteins in a mammalian background offers many clear advantages to their generation in *E. coli* or insect cells, including correct post-translational modification and folding. However, whilst the use of mammalian cells such as CHO or HEK293 is well documented, the process of creating stable mammalian cell lines can often be laborious and time consuming. Transient expression systems that utilize suspension cell lines therefore may provide a viable alternative, producing high quantities of recombinant protein in a very short period of time. Of particular interest is the use of HEK293–EBNA cells for rapid

transient expression studies (Durocher et al., 2002; Pham et al., 2003). The system utilizes episomally replicating plasmids featuring the Epstein–Barr virus (EBV) *oriP* driven by EBNA-1 protein generated from a gene integrated into the HEK293 genome. In most cases, transient protein expression is driven by the strong CMV promoter. Furthermore, the system is highly amenable to automation and many groups have begun to adapt the system to miniaturised high-throughput strategies in a similar fashion to those described for *E. coli* and baculovirus.

- *Semliki Forest virus*

Another approach to the generation of recombinant proteins with correct post-translational modification has been the extensive use of *Semliki Forest virus* for the production of a wide range of membrane proteins (Lundstrom, 2003). However, due to early biosafety concerns, this expression system has yet to be adopted in many laboratories.

- *Cell-free expression*

Traditional cell-free expression systems, whilst providing an attractive 'quick' route to the production of proteins, have always been marred by low expression levels. However in recent years, there have been seen several improvements that now enable expression of 5–10 mg/ml. Changes include the optimisation of lysate composition, introduction of semi-continuous reactions, and energy regeneration systems (Sawasaki et al., 2002). Cell free systems lend themselves to high-throughput format for the parallel expression of proteins. Indeed, a number of reports have been published on cell-free systems from *E. coli* showing the possibility of producing large amounts of protein for NMR structural studies. Moreover, a recent report describes the utilization of an enhanced wheat-germ cell-

free expression to express 50 human and plant proteins for proteomic analysis, yielding 0.1–2 mg of protein (Endo and Sawasaki, 2003). The use of a eukaryotic based approach coupled with the rapid, cell-free utility offered by the system has a number of clear advantages and offers a viable alternate to traditional approaches. Several commercial cell-free systems are currently available. However, the expression of proteins that require SH-bond formation or membrane expression is not possible in either case. Furthermore, the commercial systems are very expensive and therefore are cost prohibitive for many laboratories.

1.4.6.4 Protein purification

The development of techniques and methods for protein purification has been an essential pre-requisite for many of the advancements made in biotechnology. Protein purification varies from simple one-step precipitation procedures to large scale validated production processes. Often more than one purification step is necessary to reach the desired purity. The key to successful and efficient protein purification is to select the most appropriate techniques, optimize their performance to suit the requirements and combine them in a logical way to maximize yield and minimize the number of step required. Most purification schemes involve some form of chromatography. As a result chromatography has become an essential tool in every laboratory where protein purification is needed. The availability of different chromatography techniques with different selectivities such as hydrophobic interaction chromatography (HIC) and ion exchange chromatography (IEX) provides a powerful combination for the purification of any biomolecule.