The Role of Molecular Simulations in Understanding the Mechanisms of Cell-Penetrating Peptides

Authors: Lauren M. Reid[1,2], Chandra S. Verma[2,3,4], Jonathan W. Essex[1]

Affiliations: [1] School of Chemistry, University of Southampton, Highfield, Southampton, SO17 1BJ, United Kingdom, [2] Bioinformatics Institute (A-STAR), 30 Biopolis Street, Matrix, Singapore 138671, [3] Department of Biological Sciences, National University of Singapore, 16 Science Drive 4, Singapore 11755, and [4] School of Biological Sciences, Nanyang Technological University, 60 Nanyang Dr, Singapore 637551

Corresponding authors: Essex, J. W. J.W.Essex@soton.ac.uk (+4423) 8059 2794

Verma, C. S. chandra@bii.a-star.edu.sg (+65) 9479 7303

Keywords: Cell-Penetrating Peptides; Molecular Dynamics; Simulation; Intracellular Delivery; Membranes; Enhanced Sampling

Teaser: The review discusses the simulation techniques currently being used to study cell-penetrating peptides, along with the mechanistic insights gained and challenges associated with such studies.

Abstract: Cell-penetrating peptides (CPPs) offer an exciting approach to tackle the pharmacokinetic challenges associated with the delivery of large, polar molecules to intracellular targets. Since the discovery of the first CPPs in the early 90s, vast amounts of research have been undertaken to characterise their cellular uptake mechanisms. Despite this, the precise mechanisms of cellular internalisation of many CPPs remain elusive due to inconsistent experimental results. Molecular dynamics (MD) simulations provide an approach to probe specific aspects of the internalisation process and many published CPP studies have incorporated simulation data. This review provides a critical evaluation of the current approaches that are being used to simulate CPPs interacting with artificial lipid bilayers.

Introduction

Cell-penetrating peptides (CPPs) are short, often cationic and/or amphipathic, peptides that are able to enter cells with minimal disruption to the cell membrane [1] [2]. This is a desirable property for intracellular drug delivery systems, which are required to pull therapeutic cargo (such as proteins and siRNA) into the cell to reach their targets. The need for intracellular drug delivery is becoming increasingly evident as there is a growing interest in using biopharmaceuticals and biomimetics to treat intracellular diseases (such as stapled peptides for the inhibition of protein-protein interactions [3]). CPPs have been a topic of interest in this field since the discovery of two proteins that are able to enter cells *in vitro*: Tat, an HIV-1 transactivator protein, discovered in 1988 [4]; and Antennapedia, a homeoprotein of Drosophila melanogaster, discovered in 1991 [5]. Sequence research identified the short peptide sequences within the proteins that were responsible for their cellular uptake, leading to the discovery of the Tat peptide and penetratin [6] [7] (Table 1) . Since then, many other CPPs have been derived from proteins, synthesised or produced chimerically, and have been dubbed `Trojan horse' molecules, due to their ability to pull large, polar cargo into the cell [2] (Table 1). A comprehensive collection of CPPs that have been identified and studied can be found in the CPPsite2.0 database [8].

Insert Table 1

To design new CPPs for intracellular drug delivery, it is important to understand how they enter cells. Many experimental studies have aimed to characterise the cellular uptake of CPPs but, owing to differences in lab protocols, experimental conditions or assay types, conflicting results have emerged, and the topic remains controversial [9] [10] [11] [12]. It is now widely believed that CPPs are internalised via numerous mechanisms and that a single peptide can participate in a number of pathways; this process can be modulated by altering the physicochemical properties of the peptide, peptide concentration, lipid composition or pH [11] [13].

The ability to gain detailed mechanistic insights into these pathways has benefited from developments in the applications of computational methods such as molecular dynamics (MD) simulations. The development of MD methods (Box 1) and biomolecular force fields (Box 1) has provided the means to probe biological mechanisms at the atomistic level [14]. What is more, the ability to control or alter simulation conditions, such as temperature, ion concentration, pH etc., could allow researchers to closely monitor the effects of these parameters and disentangle the controversies surrounding experimental results. However, simulations directed at understanding the interactions of CPPs with models of cellular membranes (CPP-lipid simulations) come with their own set of challenges; including the limited accuracy of biomolecular force fields, the difficulties associated with sampling the full or relevant part of the underlying energy landscape and the simplified models of cell membranes [15] [16] [17]. To obtain valid and useful results from CPP-lipid simulations, these issues must be addressed. As a result, several CPP simulation studies have been published that implement a variety of intelligent techniques to probe their translocation across models of membranes. Each technique comes with its own advantages and challenges.

This review will provide a brief description of CPP properties and their experimentally deduced uptake routes and applications, followed by an in-depth discussion of the various simulation techniques that are being used to probe their membrane penetration mechanisms. Specific attention will be paid to how the aforementioned challenges are being tackled.

Experimental Knowledge of CPPs

Physicochemical properties

CPPs can be classified based on their physicochemical properties, with categories including cationic, amphipathic and hydrophobic peptides (Figure 1, Table 1) [18]. Cationic CPPs contain a stretch of positive charges essential for cellular uptake and are generally unstructured [18]. Amphipathic CPPs can be divided into 4 subcategories:

1. Primary amphipathic peptides, which have defined hydrophobic and cationic domains.
2. Secondary amphipathic peptides, which form α-helices with one hydrophobic face and one hydrophilic face.
3. Amphipathic β-sheet peptides, which have a hydrophobic stretch and a hydrophilic stretch.
4. Proline-rich amphipathic peptides, which form polyproline II (PPII) structures.

The cell penetrating properties of amphipathic peptides can be attributed to their ability to interact with both the polar head and lipid chain regions of cell membranes [19]. The final class of CPPs are the hydrophobic peptides, which contain mainly hydrophobic residues or have a hydrophobic stretch essential for cellular uptake [18]. The uptake mechanisms of hydrophobic peptides have been less well studied; however, it is proposed that they are able to undergo energy-independent translocation across cell membranes due to their favourable interactions with the hydrophobic membrane interior [20] [21]. While the majority of the CPPs documented in the literature are positively charged, a small set of anionic peptides have also been reported [18].

Insert Figure 1

Applications

There are a vast range of potential applications of CPPs and many of these have been studied and reviewed in the literature [22]. An important application is the intracellular delivery of therapeutic agents in the treatment of cancer. For example, CPPs have been effectively conjugated (covalently and non-covalently) to siRNAs, increasing the ability of the siRNA to enter the cell and modulate function [22] [23]. Hyun et al. recently developed moderate RNA binding peptides that can also act as CPPs, meaning they are able to enter the cell and release siRNA into the cytosol [24]. The library of peptides were lysine- and leucine- rich and incorporated hydrocarbon staples, and were demonstrated to chaperone the siRNA into cells [24]. CPPs have the ability to transport many other classes of molecules with anti-cancer properties into cells, such as proteins, plasmid DNA and oligonucleotides, and have often demonstrated selectivity for cancer cells over healthy cells [25] [23]. This selectivity can also be utilised to deliver imaging agents to tumour cells. For example, Du et al. found that the oligoarginine, R11, is able to deliver a fluorescent probe to bladder tumours, producing a more intense signal than that seen in normal bladder tissue [26]. More extensive reviews of the applications of CPPs as anti-cancer therapies and diagnostics can be found in [22] and [23].

CPPs also have applications in other disease areas. Chen et al. demonstrated the ability of the CC12 peptide (Table 1) in delivering antiangiogenesis peptides to the posterior segments of the eye for the treatment of neovascular fundus diseases via topical application [27]. CC12 forms an amphipathic, PPII helix which is able to permeate through the multiple barriers of the eye before reaching the retina. Furthermore, CPPs are being developed as potential vehicles to allow oral delivery of insulin. Fukuoka et al. have studied the effects of co-administering insulin with oligoarginines loaded into hydrogels [28]. They demonstrated that the combination results in improved intestinal absorption of insulin in rats via a non-covalent CPP uptake strategy. More general applications of CPPs can be found in [22].

Finally, it is possible that certain peptides can possess both cell-penetrating and therapeutic properties. Chang et al. published a study that optimised stapled α-helical peptides to work as dual inhibitors of the MDM2 and MDMX p53 regulatory proteins [29]. The lead peptide, ATSP-7041, was a hydrophobic stapled peptide and showed proficiency at entering living cancer cells and inhibiting both MDM2 and MDMX. The combination of favourable pharmacokinetics, cell penetration and target binding resulted in suppression of tumour growth *in vivo*.

Cellular Uptake Mechanisms

Cellular uptake mechanisms of a CPP can vary depending on the experimental protocol and conditions. In fact, it is common to observe both energy-dependent and energy-independent internalisation of the same CPP. The former involves endocytosis and subsequent release from the endosome, while the latter involves direct translocation, pore formation or physical endocytosis (Figure 2). These processes can occur via numerous mechanisms, which will be described here.

Insert Figure 2

To probe endocytotic uptake of CPPs, studies have utilised various endocytosis markers, fluorescent probes, or have been performed in conditions that prohibit endocytosis [10] [30]. Using specific endocytosis markers, Lundin et al. demonstrated the presence of three distinct endocytosis pathways for CPPs: macropinocytosis, caveolae endocytosis or clathrin-mediated endocytosis (CME) [10]. Macropinocytosis is caused by the rearrangement of cytoskeleton leading to ruffling of the cell membrane to form sheet-like extensions that fold around and non-selectively entrap extracellular fluid and particles in macropinosomes [31]. Lundin et al. and Duchardt et al. both showed that cationic CPPs can be internalised via this mechanism [10] [32]. Caveolae endocytosis involves the clustering of caveolin-1 proteins and the formation of flask-shaped invaginations in lipid raft regions of the cell membrane. Phosphorylation of caveolin-1 then leads to internalisation of the vesicle [33] [34]. Caveolae endocytosis was observed by Fittipaldi et al. and Duchardt et al. for cationic peptides, such as Tat and Arg9 (Table 1) [33] [32]. Finally, CME involves recognition of internalisation cargo by membrane bound receptors, resulting in membrane curvature and the formation of an endosome [35]. Lundin et al. observed CME as the main cellular uptake pathway for amphipathic CPPs [10]. For CPPs to reach the cell cytosol, it is important that the peptide is released from the endosome; this can involve endosomal membrane disruption or direct translocation across the endosomal membrane (Figure 2-D and E) [11].

Energy-independent mechanisms (Figure 2-A and B) have also been observed using experiments that utilise artificial lipid membranes or that inhibit endocytosis [13] [30]. These mechanisms include direct penetration (with very little disruption to the membrane), pore formation (with varying degrees of membrane disruption) or physical endocytosis (with vast membrane deformations) [11]. Direct penetration is an attractive route for the design of intracellular delivery vectors, due to the low likelihood of causing cell leakage or cytotoxicity [36]. Swiecicki et al. used large unilamellar vesicles (LUVs) with a combination of anionic/zwitterionic lipids to observe direct translocation of Tat, Arg9 and R6/W3 (Table 1) [12]. This occurred via aggregation of the peptides on the membrane surface, leading to lipid flip-flop, where the anionic lipids chaperone the positive charges in the peptide across the hydrophobic membrane interior. Maniti et al. observed spontaneous translocation of penetratin, Arg9 and RW9 (Table 1) in plasma membrane spheres (PMSs) but also observed endosome-like invaginations in the membrane, along with external budding and tubulation [37] [38]. The authors suggested a physical endocytosis mechanism, which involves membrane manipulation via the formation of large peptide aggregates on the membrane surface, leading to electrostatic and hydrogen bonding interactions that induce negative or positive curvature [37] [38] [39]. Marks et al. also identified a series of peptides, including TP2 (Table 1), observed to spontaneously translocate across artificial and cell membranes [21]. The spontaneous membrane-translocating peptides (SMTPs) are shown to internalise inside cells via non-endocytic mechanisms and with no concentration dependence, suggesting that uptake happens via direct translocation as monomers [21] [30]. What is more, the SMTPs were shown to efficiently pull large, polar cargo inside the cell, meaning they have ideal properties as intracellular drug delivery systems.

Finally, CPPs are also hypothesised to enter cells via pore formation (Figure 2-B), specifically via toroidal pores [13]. Although the structure of the pores are similar to those proposed for antimicrobial peptides (AMPs) [40], CPPs cause less cell leakage and cytotoxicity, as the pores are smaller and shorter lived [41]. Pisa et al. also proposed a carpet model or barrel-stave model resulting in CPP-induced pores; however these models are more commonly associated with AMPs [13]. Islam et al. observed translocation of TP10 (Table 1), followed by TP10-induced pore formation in 80/20 DOPG/DOPC giant unilamellar vesicles (GUVs), using a fluorescent probe and confocal microscopy [42].

In summary, numerous energy-dependent and energy-independent cellular uptake mechanisms have been determined for CPPs. However, the experimental techniques used to characterise these mechanisms do not provide atomistic information and the molecular interactions that occur between CPPs and lipid bilayers remain uncharacterised. MD simulations provide the means to gain insight into these atomistic interactions and probe the various energy-independent mechanisms. Although endocytosis is not currently accessible to simulations, due to the energy-dependence and large scales, MD simulations provide a way to study the interactions of the peptide with the membrane that may be crucial to allow the peptide to escape from the endosome. The next section will discuss the simulation techniques that are currently being used to study CPP translocation. Box 1 describes the various simulation methodologies discussed throughout this review, while Table 2 summarises the example usages, advantages/ disadvantages and software implementations of each technique.

Techniques for CPP Simulation

Force Fields

To simulate CPP-lipid systems, biomolecular force fields (Box 1) have been developed at varying simulation scales; for example the atomistic Amber and CHARMM series [43] [44], the united atom GROMOS series [45] and the coarse-grained (CG) MARTINI and SIRAH series [46] [47]. The CG simulation method is described in Box 1 and discussed in more detail in a later section. The choice of simulation scale depends on the system size, timescales and resolution needed to observe the biological phenomena of interest [48]. The parameterisation of atomistic force fields involves fitting to ab initio and experimental data, while CG force fields can be parameterised using “bottom-up” (fitted to atomistic simulation data) or “top-down” (fitted to experimental data) approaches. More extensive discussions of the development of biomolecular force fields can be found in several excellent references [48] [17].

The ability of a force field to recreate experimental results can greatly depend on the fitting procedure and data utilised, meaning different force fields can often result in different simulation outcomes. Various studies have compared the different protein-lipid force fields and often conclude that the best choice depends on the system and the nature of the calculated property [15] [16]. Piggot et al. performed a comprehensive comparison of 5 lipid force fields from the GROMOS [45], CHARMM [44] and Berger [49] series by simulating DPPC and POPC bilayers and comparing various calculated and experimental physical properties [15]. The study concluded that the calculated physical properties of lipid membranes depends substantially on the choice of force field and, while a `best' force field was not determined, certain ones are better or worse for certain measurements. Sandoval-Perez et al. also published a comparative study of common protein-lipid force fields, specifically calculating protein interactions with lipid bilayers [16]. In this study, GROMOS 54A7 [50], CHARMM36 [51], Amber14SB/ Slipids [52] and Amber14SB/ Lipid14 [53] [43] force fields were used to compare membrane protein secondary structure conservation, transmembrane (TM) peptide positioning and free energy of amino acid adsorption. The study showed that Amber14SB/ Slipids, Amber14SB/ Lipid14 and CHARMM36 performed well in secondary structure conservation and were able to predict TM peptide orientation and insertion depth, while GROMOS 54A7 performed poorly. Furthermore, CHARMM36 was best able to predict adsorption free energies of Wimley White peptide amino acids [54], closely followed by Amber14SB/ Lipid14, while Amber14SB/ Slipids and GROMOS 54A7 each performed significantly worse [16]. In summary, it is advisable for researchers to use multiple force fields to compare simulation and experimental observations, to choose the force field most suited to describe their system.

Insert Box 1

Atomistic MD

The first atomistic MD simulations (Box 1, Table 2) of spontaneous CPP translocation events were published by Herce and Garcia. [55]. Multiple Tat peptides were placed on one side of a DOPC bilayer and atomistic MD was performed using the Berger lipid force field [49]. The peptides were observed to insert under the phospholipid headgroups, making favourable hydrogen bonding/electrostatic interactions with the phosphates. As additional peptides were added to the system, they accumulated under the phospholipid headgroups and the increased positive charge attracted the phosphates from the opposite bilayer, resulting in membrane thinning. Eventually, an arginine side chain was able to reach across the hydrophobic membrane interior and pull a phosphate head group from the opposite bilayer up into the lipid chain region. This resulted in the formation of a toroidal pore, with the peptides at the centre, after only 400 ns of simulation [41] [55]. It should be noted, however, that these simulations were performed in the absence of neutralising counterions, which has been shown to cause membrane thinning and destabilisation due to the charge imbalance across the bilayer [56]. This method is useful as an approach to increase the likelihood of peptide-induced pore formation, allowing the study of pore structure and peptide insertion into the pore, but the observations should be treated with caution due to the artificial charge imbalance.

Self-Assembly Approach

Atomistic MD has also been used as an *a priori* approach to predict the position and orientation of peptides in membranes. For example, Farrotti et al. studied the designed peptide, LAH4 (KKALLALALHHLAHLALHLALALKKA-NH2), which is known to possess both AMP or CPP activity based on pH (in acidic conditions the histidines are cationic (LAH4-c) and the peptide acts as an AMP, while in basic conditions the histidines are neutral (LAH4-n) and the peptide acts as a CPP) [57]. The authors comment on the sampling difficulties associated with peptide insertion into a preformed bilayer and instead utilise an atomistic self-assembly approach. This involved randomly placing the helical peptide, POPC lipid molecules, water and neutralising ions in the system and observing self-assembly of a POPC bilayer with the peptide embedded. An annealing approach, where the system was heated from 300K to 375K and then cooled back to 300K multiple times, was needed to observe bilayer formation in the presence of the peptides. This approach was successful in identifying the different observed experimental behaviours of LAH4-n and LAH4-c, with the neutral histidine species forming a TM helix and the cationic histidine species sitting approximately parallel to the membrane surface and causing more bilayer disruption [57]. The study also involved other simulation methods, namely umbrella sampling (US) and CG MD that are discussed in subsequent sections, to corroborate the results.

Atomistic MD with Initial Configurations Selected using a priori Knowledge

Atomistic simulations carried out with the CPPs placed near the membrane usually result in the peptide making occasional binding interactions with the phospholipid head groups, yet spontaneous pore formation or translocation is unlikely to occur within the timescales currently accessed by MD simulations [56]. However, there have been a number of CPP simulations reported in the literature that probe the membrane translocation mechanism by selecting the initial configuration based on *a priori* knowledge [58] [59] [60] [61] [62]. This requires knowledge derived from experiments or previous simulations.

One method for investigating the ability of CPPs to stabilise membrane pores is to place the peptide inside a pre-formed pore and monitor the rate at which the pore closes. This is a useful technique if the translocation mechanism has been shown to involve pore formation experimentally. Akhunzada et al. used this approach to study the effects of Tat monomers, Tat-TAMRA monomers and Tat-TAMRA dimers on the stabilisation of DOPC pores [58]. First, the configuration of the Tat dimers were taken from a previous study by Macchi et al. that had shown that dimerisation is the main form of Tat aggregation using experimental and computational approaches [63]. In this aggregation study, UV-vis and NMR were used to prove the existence of Tat dimers, while metadynamics (Box 1, Table 2) was used to predict the dimer configurations . The inter-peptide contact number was set as the CV (Box 1) for the metadynamics, effectively pulling the monomers together. High-contact structures were extracted and used to seed MD simulations to assess the stability of the dimers [63]. The stable dimers identified in this study were pulled to the centre of a pre-equilibrated 30 Å DOPC pore, which was formed by inducing surface tension in the membrane and then restraining the pore while simulating the membrane back to equilibrated conditions [58]. Next, unconstrained MD was performed to observe the effect of the peptides on the pore. Monomers were shown to translocate out of the pore after 450 ns, stabilising the pore for much longer than the control system that closed after 30 ns of simulation. The dimer had even stronger interactions with the lipids and stabilised the pore across 1 µs [58]. These simulation studies proved useful in generating Tat dimer configurations and testing the stabilisation of DOPC pores by Tat monomers and dimers, and are a good example of how MD simulations can be used in conjunction with experiment. Nevertheless, it may be interesting to investigate the dependence of dimer configuration prediction on the choice of CV, specifically whether the inclusion of different peptide folding CVs could cause the peptides to fold and dimerise in different ways.

Other simulation studies have also used this approach to investigate the stabilisation of pores by CPPs. For example, Sun et al. simulated Arg8 at the centre of a DPPC bilayer pore by removing the restraint from the last frame of a US window (US is discussed in more detail in a later section and in Box 1) [59]. They found that the interactions between Arg8 and the lipids in the pore stabilised the system for a full 200 ns simulation. This is in contrast to the pore closure observed after 5 ns in a system with no peptides. Sun et al. used this approach again to study the differences in the abilities of Arg8 and Lys8 to stabilise pores in DOPE/DOPS bilayers [60]. They found that a pore with no peptide closed after 30 ns but a pore with an Arg8 peptide at the centre remained open for a full 800 ns simulation. A pore with Lys8 at the centre closed after only 10 ns, consistent with experimental evidence that Lys8 is a much less efficient CPP than Arg8 [64]. This can be attributed to the fact that the guanidinium side chain of arginine is able to form stronger interactions with the phosphates in the lipid head groups than the amino side chain of lysine [65]

Experimental structures of CPPs, when available, are used as initial starting configurations for atomistic MD. For example, Bera et al. derived the three-dimensional structure of the penetratin-derived peptide, DK17 (DRQIKIWFQNRRMKWKK), in the presence of different artificial lipid membranes using NMR, and then used these structures as initial conformations in atomistic simulations [61]. They also used the Orientation of Proteins in Membranes (OPM) server to predict the orientation of the derived peptide structure in the membrane, giving the initial simulation configurations [66]. Briefly, the OPM server allows users to access a database of predicted TM protein spatial arrangements in membranes [66]. The database was built by minimising the transfer free energies from water to a hydrophobic slab of all the TM proteins in the protein data bank (PDB) [67]. Bera et al. performed 100 ns of atomistic simulation and monitored the stability of the secondary structure of the peptide in the membrane, revealing that the peptide had different preferential structures in different membrane compositions [61].

Grasso et al. also recently used experimental NMR structures of the Tat peptide and penetratin, along with the OPM server, to generate initial peptide-embedded DOPC configurations for atomistic simulations [62]. They also studied Arg9, Cady, Pep and MAP (Table 1) embedded in DOPC with initial conformations predicted by the PEP-FOLD 3 server [68]. The CPP-DOPC systems were each simulated without restraints for 200 ns and peptide/membrane properties were monitored, revealing different residing depths and orientations for each CPP in the membrane. They also linked the physicochemical properties of the CPPs to their effects on the membrane bending modulus, finding that more hydrophilic peptides encouraged more water molecules to enter/remain inside the membrane, leading to increased membrane destabilisation [62].

The studies discussed above are examples of how MD simulations can validate experimental results and provide additional insights. However, it is important to note that observations taken from atomistic peptide-lipid MD simulations are likely to depend on the initial conformation/ configuration. It may have been advisable for these studies to include repeat simulations starting from independent starting configurations, to statistically validate the conclusions drawn. Furthermore, to derive valid thermodynamic and kinetic information about a process, multiple, reversible events must occur within the simulation; for example, pore formation and closure must happen several times to ensure statistical convergence. It is very unlikely that atomistic MD is able to achieve the timescales needed to observe this with the computational resources currently available. However, advancements in computer technology, such as the purpose-built Anton supercomputer for MD simulations, promises longer timescales becoming accessible [69].

High Temperature MD (HT-MD)

Ulmshneider et al. have published a series of studies that utilise HT-MD (Table 2) to investigate membrane active peptides (MAPs), such as TM helices and AMPs, and have suggested that the technique could be extended to study CPPs [70] [71] [72]. The principle behind the technique is to simulate peptide-lipid systems at an increased temperature to speed up the kinetics involved with peptide folding on the membrane surface and insertion into the bilayer. TM helices are known to be unstable in aqueous solution but are stabilised by the membrane, as a result of the increased resistance to break favourable backbone hydrogen bonds in a hydrophobic environment [71] [72]. Simulations and circular dichroism (CD) experiments of TM helices at varying temperatures have shown that their structures are thermostable in the membrane, suggesting that increased simulation temperatures can be used without altering the preferred structure of the embedded peptide [71] [72]. Wang et al. used HT-MD in conjunction with various experiments to study the effect of the AMP, maculatin (GLFGVLAKVAAHVVPAIAEHF-NH_2), on lipid bilayers [72]. Using simulation temperatures of up to 150°C, the simulations in this study were initiated with multiple surface-bound helices on a range of phosphatidylcholine lipid membranes and were able to capture transmembrane insertion and multiple pore formation events over 10-20 µs. The simulations revealed a variety of oligomer and pore structures and numerous pore forming mechanisms (Figure 3) [72].

Insert Figure 3

A possible drawback of this approach is the assumption that the force fields remain valid at elevated temperatures, given that most force fields were not parameterised at such high temperatures. Also, while it is shown that the secondary structures of TM peptides are thermostable in the membrane, the nature of the thermodynamic landscape at high temperatures needs more investigation.

Steered Atomistic MD

Steered MD/pulling simulations (Table 2) provide a way to force the peptide across the membrane. This often involves harmonically restraining a part, or the centre of mass (COM), of the peptide to the COM of the lipid bilayer and applying a force that pulls the two groups closer. Alaybeyoglu et al. performed steered-MD of pVEC (Table 1) across a POPE bilayer [73]. The peptide was placed above the membrane in a random coil position, with a β-turn at residues Arg8 to Lys11, and a force was applied to the N terminus C_α atom to pull the peptide to the centre of the membrane. The force needed to pull the peptide through the membrane and the presence of water molecules in the bilayer were monitored. From this study, Alaybeyoglu et al. suggested a translocation mechanism by which pVEC initially binds to the negative phospholipid head groups, via electrostatic interactions of its cationic stretch with the negative phosphates, followed by the formation of a water defect in the membrane that shields the cationic residues from the hydrophobic core [73]. However, the dynamics presented in this publication do not represent equilibrium dynamics and it is likely that mechanistic details taken from these simulations are artefacts of the force applied to the N terminus. The study did not generate a PMF via US and just deduced mechanistic detail by assessing the pulling force needed to keep the peptide moving at a constant rate. In addition, the energy landscape associated with peptide folding and membrane translocation is complex and this simulation may not have been long enough to capture the peptide conformational changes that could occur as it interacts with the membrane.

Atomistic Umbrella Sampling

A more stringent approach to investigate the mechanism and energy associated with peptide translocation is to use snapshots from pulling simulations to perform atomistic US (Box 1, Table 2), as has been shown by several studies [56] [59] [74] [60]. The general approach is to set the US CV as the distance, z, between the COM of a peptide and the COM of a lipid bilayer. Initial configurations along z are generated by pulling the peptide to the centre of the bilayer or by growing the peptide into the membrane at different z values. These initial configurations are then used to seed US simulations, as described in Box 1. The PMFs produced by performing US of CPPs passing through a bilayer often show similar characteristics: an energy minimum is observed when the peptide binds to the membrane surface, followed by a steep increase in energy as the peptide is introduced into the hydrophobic membrane interior and a kink in the PMF is observed if a water pore is formed (Figure 4).

Insert Figure 4

In 2013, Huang and Garcia used US to investigate the movement of cyclic Arg9 across a DOPC bilayer [74]. The study used 2 different approaches to pull the peptide into the membrane:

1. The C_z atom of one arginine residue was restrained to the COM of the bilayer.
2. The C_z atom of one arginine residue was restrained to the centre of weighted mass (COWM) of the lower bilayer phosphates that fell within a cylinder with diameter comparable to the guanidinium group.

They found that using different order parameters resulted in different translocation mechanisms. The first pulling approach did not result in pore formation and the associated PMF showed an increase in energy all the way to the bilayer centre, with a huge energy barrier of 200 kJ mol^{-1}. The second approach pulled the lower leaflet phospholipids closer to the peptide, resulting in the formation of a pore, with a much smaller activation energy of 120 kJ mol^{-1} [74]. The authors comment that the pore formation mechanism is more likely because of the smaller activation energy, however this may be an artefact of the simulation because the lower phosphates were forced into the bilayer centre by the pulling restraint, not through interactions felt with the peptide alone. Surprisingly, adding more unrestrained peptides to the system increased the energy barrier, which is not consistent with experimental evidence that suggests the arginine-rich CPPs form pores by aggregating on the membrane surface. This further emphasises the importance of choosing the right reaction coordinate in getting sensible US data.

Sun et al. also performed pulling and US simulations of Arg8 into the centre of a DPPC bilayer and compared the resulting PMF with that produced from pulling a phospholipid head group into the centre (lipid flip-flop) [59]. They found that both pulling simulations resulted in pore formation, giving characteristic kinks in the PMFs. The energy barrier for Arg8 pore formation was 87.5 kJ mol^{-1}, slightly higher than the energy barrier measured for lipid flip-flop, which was 76 kJ mol^{-1}. As with Huang and Garcia's US simulations, they found that adding more unrestrained peptides resulted in a higher energy barrier for Arg8-induced pore formation and reduced the binding interactions between the peptide and the membrane. Although both US studies discussed here produced results that contradict the experimental data, in the sense that adding more peptide should increase permeation, it is interesting to note the similarities in their results. Furthermore, Sun et al. did observe a slightly smaller energy barrier to lipid flip-flop-induced pore formation with the addition of unrestrained peptides (likely resulting from the increased lipid disorder induced by peptide binding), which could suggest that Arg8 enters the cell by increasing the likelihood of spontaneous lipid flip-flop events [59].

The challenge associated with atomistic pulling and US simulations to study CPP penetration into lipid bilayers is that the resulting PMF is likely to depend on the initial peptide conformation and orientation relative to the membrane. As previously described, the energy landscape associated with peptide conformational changes and membrane translocation is complex and US simulations may only sample one possible energy ‘route’ into the membrane. It is possible that the peptide secondary structure changes as it moves from bulk water to the membrane surface and then again as it enters the hydrophobic interior of the membrane. The simulations described above are of arginine-rich peptides that are found to have random coil structures in water; US simulations may not sample all the possible random coil conformations that the peptides adopt during translocation. When using pulling and US simulations to study peptide insertion into a bilayer, it may be advisable to create repeats with independent initial configurations to assess the dependence on the resulting PMF.

Replica Exchange Methods

To better understand the conformations accessible to peptides in the presence of membranes, replica exchange approaches can be applied to increase conformational sampling (Box 1, Table 2). Nymeyer et al. performed T-REMD (Box 1) on WALP-16 (AWWLALALALALAWWA), a model TM α-helical peptide, in the presence of a DPPC bilayer [75]. An unstructured peptide was placed above the membrane surface and simulated in NVT conditions with the area per lipid big enough to be comparable to that of a previous DPPC/WALP simulation (i.e greater than that of pure DPPC). The study used 38 replicas exponentially spaced between 350-505.8 K, simulated for 1 ns, followed by 64 replicas exponentially spaced between 350-800 K, simulated for 1.6 ns. A planar restraint was added to the C2 atoms of the lipids to prevent the bilayer from disintegrating. The simulation resulted in the peptide inserting into the DPPC bilayer before folding into an α-helix, providing insight into the insertion mechanism. However, the results should be interpreted with caution due to the fact that the DPPC bilayer was simulated with an artificially high area per lipid, meaning that the peptide had more room to insert in between the phospholipid head groups [75]. In reality, it may be necessary for the peptide to fold into a helix upon binding to the membrane, allowing the peptide to interact with the lipids in a way that pushes the head groups apart and facilitates insertion. Also, although T-REMD was applied to enhance the peptide conformational sampling, the simulation times were very short and the peptide secondary structure and depth in the membrane may not have converged.

Smith et al. used REST (Box 1) to enhance the conformational sampling of the Aβ10-40 amyloidogenic peptide in the presence of a DMPC bilayer [76] [77]. They set the peptide and counter-ions as the ‘hot’ solute, while the remaining water, ions and lipids were set as the ‘cold’ solvent, resulting in an increase in the peptide conformational sampling. In the initial configuration, the peptide was bound to the DMPC bilayer through 4 or 5 residues and throughout the course of the simulation, hydrophobic regions were able to insert under the phospholipid heads. The simulations allowed useful comparisons to be made between the secondary and tertiary structures of the peptide in bulk water and on the membrane. Additionally, important binding interactions between the peptide and the lipids were identified. This method was used again by Parikh and Klimov to investigate the effect of including lipopeptides in a DMPC bilayer on the binding and secondary structure propensity of the Aβ10-40 peptide [78]. Both studies highlighted the usefulness of the REST method in identifying peptide-lipid interactions and exploring the conformational space accessible to the peptide while in contact with the membrane; this method could be applied to CPP-lipid systems to investigate the mechanistic role of peptide conformations in membrane translocation.

Coarse-Grained Simulations

Another way to enhance sampling is to use CG simulations (Box 1, Table 2, Figure 5). While this approach decreases the resolution of the model, it vastly increases the computational efficiency and allows for larger systems to be simulated for longer timescales. There have been numerous examples of CG simulations of CPP-lipid systems described in the literature [79] [80] [81] [82] [83].

Insert Figure 5

CG US was used by both Hu and Patel and, Zhang et al. to study the translocation of the SMTP, TP2, and compare it with negative controls [80] [82]. Both studies used the Martini CG force field [46] to describe the peptide and a POPC (or POPC/POPG) bilayer, and performed US to assess the free energy change associated with the peptides penetrating to the centre of the bilayer. Hu and Patel used a growing approach, where they gradually switched on the interactions of the peptide at different insertion depths, separated by 0.1 nm, in the membrane [80]. They then restrained the COM of the peptide at the insertion depths and performed a 300 ns production simulation at each depth. Zhang et al., however, built an initial system with the peptide at the centre of the bilayer and then pulled the peptide into bulk solvent [82]. They also used snapshots every 0.1 nm and performed a 250 ns production simulation for each snapshot. The studies were consistent in predicting a higher energy barrier for penetration of the negative control, ONEG (PLGRPQLRRGQF), compared to TP2 and both observed a stronger and deeper binding interaction of TP2. The studies revealed that TP2 induces less membrane deformation than ONEG, even though it resides deeper in the membrane; this is consistent with experimental evidence showing that TP2 does not cause cell leakage [84]. Hu and Patel also compared their findings with other SMTPs and control peptides, showing that the SMTPs had lower energy barriers for translocation (all SMTPs had barriers of less than 96.2 kJ mol^{-1}, while all the control peptides had barriers of at least 125.5 kJ mol^{-1}). Zhang et al. took the lowest energy structures (when the peptides were bound to the membrane surface) and converted them into atomistic models, using the CHARMM36 force field. This allowed a comparison of the membrane-bound TP2 and ONEG structures at the atomistic level, confirming that TP2 indeed resides deeper in the membrane, likely due to its high hydrophobicity.

Via et al. used the Martini CG force field to study the effects of transmembrane potentials on the ability of Arg9 and R8W3 to translocate across a DOPC bilayer [83]. The simulations involved using well-tempered metadynamics with the CV set as the z distance between the COM of the peptide and the COM of the membrane; this allowed the construction of a PMF from the metadynamics hills. Simulations were run for 6 to 8 µs, providing a less expensive way of constructing a PMF than from US simulations, which require multiple simulations to sample different values of the CV. The authors observed pore formation when the peptide reached the centre of the membrane in all the simulations and were able to relate CPP activity to the presence of a transmembrane potential (applied by adding more ions on one side of the membrane and restraining them from crossing the periodic boundary), showing that CPPs are able to translocate with a lower energy cost in the presence of a charge imbalance. However, it should be noted that this approach may sample configurations along the bilayer normal less thoroughly than US, meaning that the lowest energy translocation path may not be sampled.

An advantage of using CG models is that larger and more realistic cell membranes can be simulated [85], as opposed to the simplistic phospholipid bilayer models accessible to atomistic MD. In reality, eukaryotic cell membranes are complex and contain many different phospholipids, cholesterol, sphingolipids and glycolipids. Also, an asymmetrical distribution of these components means that the upper and lower membrane leaflets have different overall properties, such as charge and curvature, which could affect CPP activity. Marrink et al. recently published a review that discusses the development of more realistic cell membrane models [17]. The next steps in the study of CPPs could involve utilising these complex membrane models to understand the influence of the different membrane components on the cell-penetrating activity.

Although the use of CG models vastly increases the ability of the simulation to sample longer timescales and larger systems, they also lose atomistic detail, which may lead to parameterisation artefacts. For example, the 4-1 mapping of heavy atoms to Martini interaction beads means that lipids with 12 or 16 carbon atoms per acyl chain are split into 3 or 4 interaction beads respectively. However, lipids with 14 carbon atoms cannot be split perfectly, so their length is either over- or under-estimated [86]. Further, CG simulations speed up sampling by smoothening the energy landscape; the kinetics are therefore modified in an unpredictable way and are no longer comparable with real time experiments [86]. In addition, it is not possible to sample peptide conformational transitions, as the description of the backbone is simplified. If the peptide secondary structure is known *a priori*, restraints can be applied in the Martini force field to retain the structure throughout the simulation, resulting in the neglect of any conformational changes associated with the translocation process [86].

Implicit Membranes

Implicit membrane force fields describe the bilayer as a continuous medium, where the hydrophobic membrane interior is modelled as a low dielectric slab with a high dielectric slab on either side, representing the water and polar lipid head groups [87]. This vastly decreases the computational expense, which increases the simulation time accessible to the system, meaning an atomistic peptide is able to sample a wider conformational space. However, simplification of the membrane comes with the problem that the membrane thickness is fixed and simulating water defects is challenging, which could overlook important events associated with CPP translocation [88].

To study the structure of TP2 and ONEG in the presence of a membrane, Lazaridis et al. used a modified version of the IMM1 implicit membrane model that takes into account changes in membrane thickness and water defects [88] [89]. Membrane deformations were accounted for by simulating the system at multiple membrane thicknesses, T, and estimating the free energy of membrane deformation, ΔG^{def}, as a function of T. The optimal T was then chosen as the value that minimised ΔG^{def}. Water defects were accounted for by altering the free energy profile according to all-atom simulations. The study initiated atomistic simulations of a single TP2 peptide, constrained to different secondary structures, near the implicit membrane surface and measured binding. The study also investigated the transition state configurations of helical-constrained TP2 and ONEG passing through the membrane. These simulations involved restraining the peptides at grid values of membrane depth (z) and tilt (θ), to sample data and construct a two-dimensional free energy surface. This produced results consistent with the previously described CG US simulations of TP2 and ONEG [80] [82]; TP2 binds more strongly to lipid bilayers than ONEG and has a smaller free energy barrier to translocation through the bilayer (104.6 kJ mol^{-1} as opposed to 184.1 kJ mol^{-1} reported for ONEG) [88].

Although the use of implicit membranes with atomistic peptides decreases the computational cost and increases the sampling efficiency of the peptide (and Lazaridis et al. address the issues of membrane thinning and water defects with their modified model [88]) it loses the atomistic detail of the membrane. Results from these simulations should therefore be treated with caution as important peptide-lipid interactions required for membrane translocation may have been overlooked, for example, the specific role that arginine could play in binding to the phosphates.

Insert Table 2

Conclusion

There is much interest in the use of CPPs as intracellular drug delivery systems as they are shown to pull large, polar cargo into cells [2]. The experimental literature reveals various energy-dependent and energy-independent cellular uptake mechanisms for the same peptides, which have been discussed here. It is now widely believed that many CPPs enter cells via endocytosis at low concentrations but via energy-independent mechanisms at high concentrations [11]. While the endocytosis mechanisms remain inaccessible to MD simulations due to their energy-dependence and large scales, energy-independent mechanisms can be studied using a variety of simulation techniques and may prove helpful in understanding how CPPs are able to escape endosomes. Examples of simulation studies that have provided useful atomistic insights into CPP translocation have been discussed in this review, along with associated advantages and limitations.

It is generally accepted that conventional MD is unable to reach the timescales needed to observe CPP translocation, which may involve multiple peptide folding, aggregation and insertion events. However, many studies were able to utilise atomistic MD simulations by selecting initial configurations based on knowledge gained from experiment [58] [59] [60] [61] [62]. For example, simulations can probe the ability of a peptide to stabilise pores by placing it in a preformed pore and monitoring its closure [58]. This is a useful method if mechanistic detail is already known; however, biasing simulations in this way could result in incorrect assumptions being made if experimental data is not available. In this review, the inclusion of experimental data in CPP-lipid simulations were discussed in the context of choosing appropriate initial coordinates. In addition, there have been methods reported that directly incorporate experimental data into the simulation by reweighting or biasing the structural ensemble [90]. To the best of our knowledge, these methods have not yet been applied to CPP-lipid simulations but could offer an exciting new approach for the future.

Ulmshneider et al. also demonstrate the usefulness of using elevated simulation temperatures in studying TM peptides, showing that TM helices are thermostable in the membrane and that high temperatures can speed up the kinetics [70]. This technique should, however, be used with caution as the biomolecular force fields were not parameterised at high temperatures, meaning that high temperature simulations could lead to artefacts. Atomistic steered MD and US are commonly used to study CPPs, where the peptide is pulled across the membrane and US is used to produce a PMF [56] [59] [74] [60]. This yields the free energy changes associated with the peptide translocating across the lipid bilayer. However, the free energy barriers calculated for CPPs are often still very large and inconsistent with the fast translocation rates observed experimentally. This could result from missing the lowest free energy translocation path by choosing an incorrect CV or under-sampling the conformational changes adopted by the peptide in contact with the membrane. The result of the simulation is therefore very likely to depend on the initial conformation and orientation of the peptide. It could also be a result of the over-simplification of the membrane; in reality, membranes are likely to consist of multiple lipid types, cholesterol and many membrane proteins. A possible solution to the conformational sampling problem is to use a replica exchange method on the peptide in the presence of the membrane [77] [78]. This could identify important conformational transitions undertaken by the peptide in the translocation process. Finally, CG or implicit membrane simulations can be used to decrease the computational cost associated with sampling large systems and long timescales. Although there are CG US simulations described in the literature that produce PMFs with relative energy barriers consistent with experiment [80] [82], the over-simplification of the system results in important binding interactions and conformational changes being overlooked. The use of implicit membranes also neglects these interactions.

Owing to the complex nature of the energy landscapes associated with peptide folding, aggregation and penetration into the membrane, the choice of initial peptide conformation/ configuration relative to the membrane is important for all simulation techniques. For atomistic MD, it is likely that peptide-lipid simulations will sample only local minima, so it is vital to seed the simulation from experimentally significant conformations. For example, NMR or crystal structures, prediction softwares to determine membrane insertion depth/ orientation or experimental assays may be used to determine if peptide aggregation and/ or membrane pore formation is involved. This is also important when performing atomistic pulling and US, as the CV may not describe the peptide folding or orientation in the membrane. In theory, replica exchange approaches are able to sample from an equilibrated ensemble of structures, so the initial peptide conformation is less important; however multiple simulations should be performed to check convergence. Additionally, CG and implicit membrane simulations allow for longer timescales to be accessed, meaning convergence is more likely to be achieved, but it is again advisable to repeat simulations to validate results.

In conclusion, MD simulations provide a useful approach to complement experiments and to validate experimental results. However, researchers should be aware of the limitations, some of which are discussed in this review. In particular, simulations should be repeated with independent starting conditions and multiple force fields, and simulation convergence must be carefully assessed. Future efforts in CPP simulations should focus on tackling the sampling issues discussed above, including the development of new enhanced sampling approaches and more descriptive CVs for CPP translocation pathways, along with the development and refinement of more accurate force fields. Furthermore, complex membranes and longer simulation times, using CG and mesoscale models [91], could be applied in future CPP studies. It is also important to use experimental knowledge to guide simulation.

Acknowledgements

The authors would like to acknowledge The University of Southampton and A*STAR for ARAP PhD studentship and NSCC, Singapore, for computational support. We would also like to acknowledge Jakob and Martin Ulmschneider for providing Figure 3 and Javier Cáceres-Delpiano for providing Figure 5.

Conflicts of Interest

CSV is the co-founder of Sinopsee Therapeutics; there is no clash with the current work.

References

[1] Falanga, A. *et al*. (2015) Membranotropic Cell Penetrating Peptides: The Outstanding Journey. *Int. J. Mol. Sci.* 16, 25323-25337

[2] Ye, J. *et al.* (2016) CPP-Assisted Intracellular Drug Delivery, What Is Next? *Int. J. Mol. Sci.* 17, 1892-1908

[3] Bruzzoni-Giovanelli, H. *et al.* (2018) Interfering peptides targeting protein–protein interactions: the next generation of drugs? *Drug Discov. Today.* 23, 272-285

[4] Frankel, A. and Pabo, C. (1988) Cellular uptake of the tat protein from human immunodeficiency virus. *Cell.* 55, 1189-1193

[5] Joliot, A. *et al*. (1991) Antennapedia homeobox peptide regulates neural morphogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 88, 1864-1868

[6] Vivès, E. *et al.* (1997) A Truncated HIV-1 Tat Protein Basic Domain Rapidly Translocates through the Plasma Membrane and Accumulates in the Cell Nucleus. *J. Biol. Chem.* 272, 16010-16017

[7] Derossi, D. *et al.* (1994) The third helix of the Antennapedia homeodomain translocates through biological membranes. *J. Biol. Chem.* 269, 10444-10450

[8] Agrawal, P. *et al.* (2016) CPPsite 2.0: a repository of experimentally validated cell-penetrating peptides. *Nucleic Acids Res.* 44, D1098-D1103

[9] Richard, J. *et al.* (2003) Cell-penetrating Peptides. *J. Biol. Chem.* 278, 585-590

[10] Lundin, P. *et al.* (2008) Distinct Uptake Routes of Cell-Penetrating Peptide Conjugates. *Bioconjugate Chem.* 19, 2535-2542

[11] Kauffman, B. *et al.* (2015) Mechanism Matters: A Taxonomy of Cell Penetrating Peptides. *Trends Biochem. Sci.* 40, 749-764

[12] Swiecicki, J. M. *et al.* (2014) The Efficacies of Cell-Penetrating Peptides in Accumulating in Large Unilamellar Vesicles Depend on their Ability To Form Inverted Micelles. *ChemBioChem.* 15, 884-891

[13] Di Pisa, M. *et al.* (2015) Translocation Mechanism(s) of Cell-Penetrating Peptides: Biophysical Studies Using Artificial Membrane Bilayers. *Biochemistry.* 54, 194-207

[14] Karplus, M. and McCammon, A. (2002) Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646-652

[15] Piggot, T. *et al.* (2012) Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study. *J. Chem. Theory. Comput.* 8, 4593-4609

[16] Sandoval-Perez, A. *et al.* (2017) Critical Comparison of Biomembrane Force Fields: Protein-Lipid Interactions at the Membrane Interface. *J. Chem. Theory Comput.* 13, 2310-2321

[17] Marrink, S. J. *et al.* (2019) Computational Modeling of Realistic Cell Membranes. *Chem. Rev.* Jan, A-AQ

[18] Milletti, F. (2012) Cell-penetrating peptides: classes, origin, and current landscape. *Drug Discov. Today.* 17, 850-860

[19] Zaro, J. and Shen, W. C. (2015) Cationic and amphipathic cell-penetrating peptides (CPPs): Their structures and in vivo studies in drug delivery. *Front. Chem. Sci. Eng.* 9, 407-427

[20] Guidotti, G. *et al.* (2017) Cell-Penetrating Peptides: From Basic Research to Clinics. *Trends Pharmacol. Sci.* 38, 406-424

[21] Marks, J. *et al.* (2011) Spontaneous Membrane-Translocating Peptides by Orthogonal High-Throughput Screening. *J. Am. Chem. Soc.* 133, 8995-9004

[22] Gallo, M. *et al.* (2019) 1988–2018: Thirty years of drug smuggling at the nano scale. Challenges and opportunities of cell-penetrating peptides in biomedical research. *Arch. Biochem. Biophys.* 661, 74-86

[23] Borrelli, A. *et al.* (2018) Cell Penetrating Peptides as Molecular Carriers for Anti-Cancer Agents. *Molecules.* 23, 295

[24] Hyun, S. *et al.* (2018) Construction of histidine-containing hydrocarbon stapled cell penetrating peptides for *in vitro* and *in vivo* delivery of siRNAs. *Chem. Sci.* 9, 3820-3827

[25] Tansi, F. L. *et al.* (2019) New generation CPPs show distinct selectivity for cancer and noncancer cells. *J. Cell. Biochem.* 120, 6528-6541

[26] Du, Y. (2018) Novel Application of Cell Penetrating R11 Peptide for Diagnosis of Bladder Cancer. *J. Biomed. Nanotechnol.* 14, 161-167

[27] Chen, C. *et al.* (2017) Anti-angiogenesis through noninvasive to minimally invasive intraocular delivery of the peptide CC12 identified by in vivo-directed evolution. *Biomaterials.* 112, 218-233

[28] Fukuoka, Y. *et al.* (2018) Combination Strategy with Complexation Hydrogels and Cell-Penetrating Peptides for Oral Delivery of Insulin. *Biol. Pharm. Bull.* 41, 811-814

[29] Chang, Y. S. *et al.* (2013) Stapled α-helical peptide drug development: a potent dual inhibitor of MDM2 and MDMX for p53-dependent cancer therapy. *Proc. Natl. Acad. Sci. U.S.A.* 110, no. E3445-E3454

[30] He, J. *et al.* (2013) Direct cytosolic delivery of polar cargo to cells by spontaneous membrane-translocating peptides. *J. Biol. Chem.* 288, 29974-29986

[31] Kerr, M. and Teasdale, R. (2009) Defining macropinocytosis. *Traffic.* 10, 364-371

[32] Duchardt, F. *et al.* (2007) A Comprehensive Model for the Cellular Uptake of Cationic Cell-penetrating Peptides. *Traffic.* 8, 848-866

[33] Fittipaldi, A. *et al.* (2003) Cell membrane lipid rafts mediate caveolar endocytosis of HIV-1 Tat fusion proteins. *J. Biol Chem.* 278, 34141-34149

[34] Pelkmans, L. (2005) Secrets of caveolae- and lipid raft-mediated endocytosis revealed by mammalian viruses. *Biochim. Biophys. Acta. Mol. Cell. Res.* 1746, 295-304

[35] Ramanan, V. et al. (2011) Systems biology and physical biology of clathrin-mediated endocytosis. *Integr. Biol.* 3, 803-815

[36] Bond, P. and Khalid, S. (2010) Antimicrobial and Cell-Penetrating Peptides: Structure, Assembly and Mechanisms of Membrane Lysis via Atomistic and Coarse-Grained Molecular Dynamic Simulations. *Protein Pept. Lett.* 17, 1313-1327

[37] Maniti, O. *et al.* (2012) Metabolic energy-independent mechanism of internalization for the cell penetrating peptide penetratin. *Int. J. Biochem. Cell Biol.* 44, 869-875

[38] Maniti, O. *et al.* (2014) Basic cell penetrating peptides induce plasma membrane positive curvature, lipid domain separation and protein redistribution. *Int. J. Biochem. Cell Biol.* 50, 73-81

[39] Ben-Dov, N. and Korenstein, R. (2015) The uptake of HIV Tat peptide proceeds via two pathways which differ from macropinocytosis. *Biochim. Biophys. Acta.* 1848, 869-877

[40] Sengupta, D. *et al.* (2008) Toroidal pores formed by antimicrobial peptides show significant disorder. *Biochim. Biophys. Acta.* 1778, 2308-2317

[41] Herce, H. and Garcia, A. (2008) Cell Penetrating Peptides: How Do They Do It? *J. Biol. Phys.* 33, 345-356

[42] Islam, M. Z. *et al.* (2014) Entry of cell-penetrating peptide transportan 10 into a single vesicle by translocating across lipid membrane and its induced pores. *Biochemistry.* 53, 386-396

[43] Maier, J. *et al.* (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696-3713

[44] Huang, J. *et al.* (2016) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods.* 14, 71-73

[45] Reif, M. M. *et al.* (2012) New Interaction Parameters for Charged Amino Acid Side Chains in the GROMOS Force Field. *J. Chem. Theory and Comput.* 8, 3705-3723

[46] Marrink, S. J. *et al.* (2007) The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B.* 111, 7812-7824

[47] Darré, L. *et al.* (2015) SIRAH: A Structurally Unbiased Coarse-Grained Force Field for Proteins with Aqueous Solvation and Long-Range Electrostatics. *J. Chem. Theory. Comput.* 11, 723-739

[48] Goossens, K and De Winter, H. (2018) Molecular Dynamics Simulations of Membrane Proteins: An Overview. *J. Chem. Inf. Model.* 58, 2193-2202

[49] Berger, O. *et al.* (1997) Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophys. J.* 72, 2002-2013

[50] Schmid, N. *et al.* (2011) Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophy. J.* 40, 843-856

[51] Huang, J. and MacKerell, A. (2013) CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* 34, 2135-2145

[52] Jämbeck, J. and Lyubartsev, A. (2012) Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids. *J. Phys. Chem. B*. 116, 3164-3179.

[53] Dickson, C. *et al.* (2014) Lipid14: The Amber Lipid Force Field. *J. Chem. Theory. Comput.* 10, 865-879

[54] Wimley, W. C. and White, S. H. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* 3, 842-848

[55] Herce, H. and Garcia, S. (2007) Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20805-20810

[56] Yesylevskyy, S. *et al.* (2009) Alternative Mechanisms for the Interaction of the Cell-Penetrating Peptides Penetratin and the TAT Peptide with Lipid Bilayers. *Biophys. J.* 97, 40-49

[57] Farrotti, A. *et al.* (2015) Molecular dynamics methods to predict peptide locations in membranes: LAH4 as a stringent test case. *Biochim. Biophys. Acta, Biomembr.* 1848, 581-592

[58] Jan Akhunzada M. *et al.* (2017) The role of Tat peptide self-aggregation in membrane pore stabilization: insights from a computational study. *Phys. Chem. Chem. Phys.* 19, 27603-27610

[59] Sun, D. *et al.* (2014) Effect of arginine-rich cell penetrating peptides on membrane pore formation and life-times: a molecular simulation study. *Phys. Chem. Chem. Phys.* 16, 20785-20795

[60] Sun, D. *et al.* (2015) Atomistic Molecular Simulations Suggest a Kinetic Model for Membrane Translocation by Arginine-Rich Peptides. *J. Phys. Chem. B.* 119, 14413-14420

[61] Bera, S. *et al.* (2016) Structural Elucidation of the Cell-Penetrating Penetratin Peptide in Model Membranes at the Atomic Level: Probing Hydrophobic Interactions in the Blood–Brain Barrier. *Biochemistry.* 55, 4982-4996

[62] Grasso, G. *et al.* (2018) Cell penetrating peptide modulation of membrane biomechanics by molecular dynamics. *J. Biomech.* 73, 137-144

[63] Macchi, S. *et al.* (2017) Self-aggregation propensity of the Tat peptide revealed by UV-Vis, NMR and MD analyses *Phys. Chem. Chem. Phys.* 19, 23910-23914

[64] Mitchell, D. J. *et al.* (2000) Polyarginine enters cells more efficiently than other polycationic homopolymers. *J. Pept. Res.* 56, 318-325

[65] Li, L. *et al.* (2013) The Different Interactions of Lysine and Arginine Side Chains with Lipid Membranes. *J. Phys. Chem. B.* 117, 11906-11920

[66] Lomize, M. *et al.* (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nuc. Acids. Res.* 40, D370-D376

[67] Lomize, A. *et al.* (2006) Positioning of proteins in membranes: a computational approach. *Protein Sci.* 15, 1318-1333

[68] Lamiable, A. *et al.* (2016) PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* 44, W449-W454

[69] Shaw, D. E. et al. (2014) Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *IEEE.* 41-53. doi:10.1109/SC.2014.9

[70] Ulmschneider, J. and Ulmschneider, M. (2018) Molecular Dynamics Simulations Are Redefining Our View of Peptides Interacting with Biological Membranes. *Acc. Chem. Res.* vol. 51, 1106-1116

[71] Ulmschneider, J. *et al.* (2011) In Silico Partitioning and Transmembrane Insertion of Hydrophobic Peptides under Equilibrium Conditions. *JACS,* 133, 15487-15495

[72] Wang, Y. *et al.* (2016) Spontaneous formation of structurally diverse membrane channel architectures from a single antimicrobial peptide. *Nat. Commun.* 7, 13535

[73] Alaybeyoglu, B. *et al.* (2018) pVEC hydrophobic N-terminus is critical for antibacterial activity. *J. Pept. Sci.* 24, e3083

[74] Huang, K. and García, A. (2013) Free Energy of Translocating an Arginine-Rich Cell-Penetrating Peptide across a Lipid Bilayer Suggests Pore Formation. *Biophys. J.* 104, 412-420

[75] Nymeyer, H. *et al.* (2005) Folding is not required for bilayer insertion: Replica exchange simulations of an α-helical peptide with an explicit lipid bilayer. *Proteins.* 59, 783-790

[76] Liu, P. *et al.* (2005) Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci.* 102, 13749-13754

[77] Smith, A. *et al.* (2016) Does Replica Exchange with Solute Tempering Efficiently Sample Aβ Peptide Conformational Ensembles? *J. Chem. Theory Comput.* 12, 5201-5214

[78] Parikh, N. and Klimov, D. (2017) Inclusion of lipopeptides into the DMPC lipid bilayers prevents Aβ peptide insertion. *Phys. Chem. Chem. Phys.* 19, 10087-10098

[79] Hu, Y. and Patel, S. (2016) Thermodynamics of cell-penetrating HIV1 TAT peptide insertion into PC/PS/CHOL model bilayers through transmembrane pores: the roles of cholesterol and anionic lipids. *Soft Matter,* 12, 6716-6727

[80] Hu, Y. and Patel, S. (2015) Structural and Thermodynamic Insight into Spontaneous Membrane-Translocating Peptides Across Model PC/PG Lipid Bilayers. *J. Membr. Biol.* 248, 505-515

[81] Hu, Y. *et al.* (2014) Translocation Thermodynamics of Linear and Cyclic Nonaarginine Into Model DPPC Bilayer Via Coarse-Grained Molecular Dynamics Simulation: Implications of Pore Formation and Nonadditivity. *J. Phys. Chem. B*. 118, 2670-2682

[82] Zhang, Y. *et al.* (2017) Structural effects and translocation of spontaneous membrane-translocating peptides with POPC bilayer. *J. Theor. Comput. Chem*. 16, 1750002

[83] Via, M. *et al.* (2018) The interfacial electrostatic potential modulates the insertion of cell-penetrating peptides into lipid bilayers. *Phys. Chem. Chem. Phys.* 20, 5180-5189

[84] Cruz, J. *et al.* (2013) A Membrane-Translocating Peptide Penetrates into Bilayers without Significant Bilayer Perturbations. *Biophys. J.* 104, 2419-2428

[85] Risselada, H. J. et al. (2014) Expansion of the fusion stalk and its implication for biological membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.* 111, 11043-11048

[86] Marrink, S. J. and Tieleman, P. *et al.* (2013) Perspective on the Martini model. *Chem. Soc. Rev.* 42, 6801-6822

[87] Mori, T. *et al.* (2016) Molecular dynamics simulations of biological membranes and membrane proteins using enhanced conformational sampling algorithms. *Biochim. Biophys. Acta, Biomembr.* 1858, 1635-1651

[88] Lazaridis, T. et al. (2014) Implicit membrane treatment of buried charged groups: Application to peptide translocation across lipid bilayers. *Biochim. Biophys. Acta, Biomembr*. 1838, 2149-2159

[89] Lazaridis, T. (2003) Effective energy function for proteins in lipid membranes. *Proteins*. 52, 176-192

[90] Allison, J. R. (2017) Using simulation to interpret experimental data in terms of protein conformational ensembles. *Curr. Opin. Struc. Biol.* 43, 79-87

[91] Chavent, M. *et al*. (2018) How nanoscale protein interactions determine the mesoscale dynamic organisation of bacterial outer membrane proteins. *Nat. Commun.* 9, 2846

[92] Leach, A. (2009) Molecular modelling : principles and applications. 2nd ed. New Jersey: Pearson/Prentice Hall

[93] Barducci, A. *et al.* (2011) Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1, 826-843

[94] Kästner, J. (2011) Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 1, 932-942

[95] Kumar, S. *et al.* (1992) THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13, 1011-1021

[96] Sugita, Y. and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* 314, 141-151

[97] Saunders, M. and Voth, G. (2013) Coarse-Graining Methods for Computational Biology. *Annu. Rev. Biophys.* 42, 73-93

[98] Elmquist, A. *et al.* (2001) VE-Cadherin-Derived Cell-Penetrating Peptide, pVEC, with Carrier Functions. *Exp. Cell Res*. 269, 237-244

[99] Morris, M. C. *et al.* (2001) A peptide carrier for the delivery of biologically active proteins into mammalian cells. *Nat. Biotechnol.* 19, 1173-1176

[100] El-Andaloussi, S. *et al.* (2007) A Novel Cell-penetrating Peptide, M918, for Efficient Delivery of Proteins and Peptide Nucleic Acids. *Mol. Ther.* 15, 1820-1826

[101] Pooga, M. *et al.* (1998) Cell penetration by transportan. *FASEB J.* 12, 67-77

[102] Nekhotiaeva, N. *et al.* (2004) Cell entry and antimicrobial properties of eukaryotic cell-penetrating peptides. *FASEB J.* 18, 394-396

[103] Oehlke, J. *et al.* (1998) Cellular uptake of an alpha-helical amphipathic model peptide with the potential to deliver polar compounds into the cell interior non-endocytically. *Biochim. Biophys. Acta.* 1414, 127-139

[104] Crombez, L. et al. (2009) A New Potent Secondary Amphipathic Cell–penetrating Peptide for siRNA Delivery Into Mammalian Cells. *Mol. Ther.* 17, 95-103

[105] Futaki, S. *et al.* (2001) Arginine-rich Peptides. *J. Biol. Chem.* 276, 5836-5840

[106] Walrant, A. *et al.* (2011) Different membrane behaviour and cellular uptake of three basic arginine-rich peptides. *Biochim. Biophys. Acta, Biomembr.* 1808, 382-393

[107] Delaroche, D. *et al.* (2007) Tracking a New Cell-Penetrating (W/R) Nonapeptide, through an Enzyme-Stable Mass Spectrometry Reporter Tag. *Anal. Chem.* 79, 1932-1938

[108] Abraham, M. J. *et al.* (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX.* 1-2, 19-25

[109] Salomon-Ferrer, R. *et al.* (2013) An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci*. 3, 198-210

[110] Brooks, B. R. *et al.* (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.* 30, 1545-614

[111] Phillips, J. C. *et al.* (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26, 1781-1802

[112] Plimpton, S. (1995) Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys*. 117, 1-19

[113] Tribello, G. *et al.* (2014) PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun*. 185, 604-613

Boxes

Box 1: Simulation methodology descriptions

Molecular dynamics (MD): A simulation technique used to generate the dynamic evolution of atoms in a molecular system through time [92]. MD codes implement iterative algorithms that calculate atomic positions, velocities and accelerations, using Newton's laws of motion.

Force field: The underlying equation and parameters that describe the energy of the system based on the interactions felt by each atom [92]. The equation includes terms that model all the bonded and non-bonded interactions; for example, bond stretching, angle bending, electrostatic interactions etc. The atomic energies provided by the force field are used to derive the forces, which are consequently used in the MD simulation.

Collective Variable (CV): A parameter of the system that describes the process of interest, such as the distance between 2 groups. The purpose of an MD simulation is to sample data across a range of values of the CV; enhanced sampling methods may be required to achieve this (see below).

Metadynamics: A technique used to increase system sampling along a defined CV by applying a biasing force that drives the simulation away from previously visited areas of the energy landscape [93]. The biasing force is in the form of energy Gaussians (or hills), which are deposited throughout the simulation to fill the over-explored energy minima and push the system over energy barriers into different areas of the CV space. As a result, the deposited hills provide a negative imprint of the energy landscape with respect to the CV, much like sand deposited in a sand pit [93].

Umbrella Sampling (US): Another technique that increases sampling along a defined CV. The approach involves running multiple simulations with the system restrained at given values of the CV, in order to collect data at the different values [94]. These data are then reweighted to remove the restraining bias, using methods such as the weighted-histogram analysis method (WHAM) [95], and combined to produce a potential of mean force (PMF) profile, providing the free energy surface associated with different locations along the CV.

Replica Exchange Methods: A series of enhanced sampling techniques that increase conformational sampling. The simplest example is temperature replica exchange MD (T-REMD), where multiple replicas of the system are run in parallel: a base replica at the temperature of interest, and a series of replicas at higher temperatures that are able to access more areas of the conformational landscape [96]. Coordinate swaps between adjacent replicas are attempted frequently throughout the simulation based on a Metropolis acceptance criterion (a criterion that assesses the thermodynamic probability of the move), such that the base replica is able to ‘jump’ over energy barriers and access different energy minima [96]. Other replica exchange methods define the replicas in different ways; for example replica exchange with solute tempering (REST) scales the interaction parameters to create an artificially ‘hot’ solute in a ‘cold’ solvent (the hot solute has more conformational flexibility) [76].

Coarse-Grained (CG) Models: Molecular systems described using interaction beads that represent groups of atoms [97]. CG reduces the number of particles in the system and allows larger MD time steps to be used, as the lower resolution removes the high frequency motions. As a result, computational efficiency is increased and bigger systems/ longer timescales can be accessed. CG models are parameterised to reproduce experimental and/or atomistic MD observations.

Figure Captions

Figure 1: Classification of CPPs based on their physicochemical properties, including examples. Blue residues are cationic, red are anionic and orange are polar.

Figure 2: Internalisation pathways for CPPs: A) spontaneous membrane translocation; B) peptide aggregation leading to direct penetration or pore formation; C) endocytosis; D) endosomal membrane lysis; E) translocation across endosomal membrane; F) degradation/ recycling of endosomal CPP. Image taken from [11] with permission.

Figure 3: Snapshots of the most populated transmembrane oligomeric structures of maculatin in phosphatidylcholine bilayers, formed in a HT-MD simulation. Image taken from [72] with permission.

Figure 4: An example of a PMF produced from atomistic US of penetratin at different Z distances through a DPPC bilayer. The kinks in the PMF associated with pore formation in the membrane are indicated by the red lines. Image taken from [56] with permission.

Figure 5: Mapping of united atom DMPC (A) to SIRAH CG beads [47] (C) with overlaid structures shown in (B). Conventional CPK colouring used for united atom representation. CG bead colouring as follows: dark blue = choline, orange = phosphate, red = ester, light blue = hydrocarbon chain.