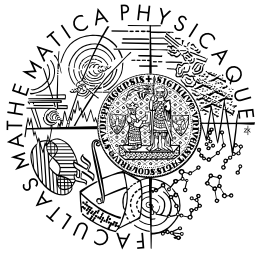


FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Katarína Baxová

Pairing of biologically relevant ions in aqueous solutions

Department of Chemical Physics and Optics

Advisor of the master thesis:	prof. Mgr. Pavel Jungwirth, DSc.
Study programme:	Physics
Study branch:	Biophysics and Chemical Physics

Prague 2018

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In Prague date ..... signature of the author

Title: Pairing of biologically relevant ions in aqueous solutions

Author: Katarína Baxová

Department: Department of Chemical Physics and Optics

Advisor: prof. Mgr. Pavel Jungwirth, DSc., Department of Chemical Physics and Optics

Abstract: Not accounting for the electronic polarizability due to divalent ions such as Ca^{2+} introduces a significant artifacts to force field-based molecular dynamic simulations of biological systems. Two newly developed parameter refinements were used to compute the free energy profile of the Ca^{2+}-Cl^{-} ion dissociation in aqueous solutions, to be compared with a free energy profile obtained from ab-intio molecular dynamics and to data from neutron scattering.

Next, the computational evidence for the existence of a local free energy minimum representing a guanidinium-guanidinium contact ion pair in aqueous solutions is provided suggesting a global preference for a contact ion pair.

Finally, the passive membrane penetration mechanism of oligoarginines was investigated on a cell membrane model systems - lipid vesicles - by fluorescent spectroscopy. In this study, a mechanistic link between membrane penetration and vesicle aggregation and fusion was found.

Keywords: ions, water, molecular dynamics

A big thank you belongs to my advisor Pavel Jungwirth for leading my scientific work and for having me in his group, as well as for his huge help and even bigger patience. I would also like to thank my colleagues Eva Pluhařová, Frank Uhlig and Tomáš Martinek for the help with the simulations as well as the IOCB and MetaCentrum for their computational capacities.

Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures" (CESNET LM2015042), is greatly appreciated.

Christoph Allolio, Piotr Jurkiewicz, Šárka Pokorná and Agnieszka Olżyńska have my gratitude for their help and advice concerning the fluorescence spectroscopy measurements.

Last but not least, I would like to thank my family and close ones for their support - my parents for their love and patience, my sister and her husband with their two beautiful kids for the times of relax, and Tomáš, Roman, Veronika, Alex, Martina and Michaela.

Contents

Introduction

Biological processes within a human body have intrigued people since the earliest recorded times. Not only out of curiosity, but also as a means of healing and disease prevention, our ancestors have tried to understand these processes and the quest continues today. As a matter of fact, we are only beginning to grasp the delicate equilibrium of the complex processes involved in the human body's everyday operation.

The human body is composed of cells. Every mammalian cell is encapsulated in a cell membrane acting primarily as a protector of the cell interior, determining which molecules from the outer space can enter the cell and vice versa.

In this work, as a part of a larger project, the interactions between a model phospholipid membrane and oligoarginines or oligolysine (Figure 1), as well as with calcium dications are investigated by fluorescence spectroscopy (Part 3 of the thesis). In concert, computer simulations allow the study of fundamental properties of Ca^{2+} (Part 1) and of guanidinium ions as a side chain groups of arginine (Part 2) in water solutions.

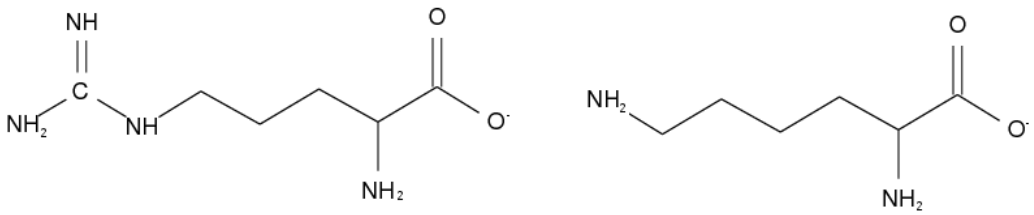


Figure 1: Arginine (left) and lysine (right) amino acids investigated in this work.

Calcium Dication

The vast majority of calcium in the human body is stored in insoluble forms in bones and teeth. The small soluble part is located both in the intra- and extra-cellular space and the differences in calcium concentrations matter hugely. Calcium as a signalling species takes part in neural signal transmission, muscle contraction and many other biological processes.

In order to computationally investigate interactions of calcium dication with lipid bilayers, proteins or big molecular systems in general, a good force field parametrisation is needed. That is in our case obtained as a compromise between fits to experimental results and benchmark ab-initio molecular dynamic (AIMD) simulations.

The first part of this thesis focuses on comparing the results of force field-based molecular dynamic (MD) simulations with AIMD simulations of a system containing two ions - Ca^{2+} and Cl^- -and 64 water molecules in a cubic simulation box as the simplest model system of calcium interactions in an aqueous environment. The chloride anion was chosen as a counterpart due to its simplicity as well as its wide presence in biological systems, mainly in the extracellular environment.

Guanidinium and Fluorescence Experiments

Guanidinium (Gdm^+, Figure 2) is a positively charged quasi-aromatic ion containing three amino groups bound to a central carbon atom. It is present in all biosystems as a part of the arginine amino acid.

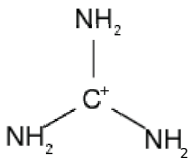


Figure 2: The positively charged guanidinium molecular ion.

The Gdm^+ molecular ion was brought to a particular attention in 1988 with findings of the human immunodeficiency virus type 1 (HIV-1) 86-amino acid long Tat protein entering the cell (Green and Loewenstein [1988], Frankel and Pabo [1988]) using its arginine-rich $Tat_{49\text{-}57}$ nonamer (RKKRRQRRR) (Vivès et al. [1997]). Here R denotes arginine, the crucial component responsible for the uptake of the cell-penetrating peptide sequence (Wender et al. [2000]), K stands for lysine and Q for glutamine.

Later it was found that various arginine-rich peptides (Futaki et al. [2001]) can penetrate the membrane, thus get inside the cell and accumulate in the cytoplasm and nucleus. This penetration occurs not only at 37° C but also at 4° C (e.g. Futaki et al. [2001],Vivès et al. [1997]), suggesting that on top of active endocytosis there is an additional mechanism responsible for the uptake. To present day this passive mechanism is still not entirely clear despite being widely studied for arginine-rich peptides' potential as drug delivery carriers.

It has been proposed that the guanidinium part of arginines plays an important role in the cell penetration processes (Wender et al. [2000]). Several studies have addressed the formation of the guanidinium contact ion pair in water solutions (e.g. Vazdar et al. [2012], Inagaki et al. [2014], Boudon et al. [1990] or No et al. [1997]), where two Gdm^+ molecules can pair despite their like charges. This stacking has been also experimentally reported to occur between two and more arginines (at the Gdm part) or plain guanidiniums (Shih et al. [2013] ,Flocco and Mowbray [1994], Kubíčková et al. [2011]).

The second part of this work is therefore aimed at a simulation study of the Gdm^+ contact ion pair formation by the umbrella sampling method (Kästner [2011]). A force field-based molecular dynamic simulations of two guanidinium ions in aqueous systems were performed using various force fields and water models.

Arginine oligomers of varying length were studied while penetrating cell membranes (e.g. Wender et al. [2000] or Futaki et al. [2001]), showing very low activity at a lengths of 4 or 5 pepride residues, and a high one for longer oligomers, best with $7 - 14$ peptide residues.

Mitchell et al. [2000] showed that lysine and its oligopeptides, another basic amino acid, penetrates the cell membrane at a much lower rate than oligoarginines.

Nona-lysine $K9$ was shown to have the cell-penetrating activity comparable to $R4$ or $R5$.

The calcium dication is known to preferably inhabit the space close to the membrane surface (Magarkar et al. [2017]) and enhance membrane fusion (Papahadjopoulos et al. [1976]). The final part of this work, therefore, focuses on the experimental investigations of calcium, nona-arginine ($R9$) (and tetra-arginine ($R4$) and nona-lysine ($K9$) as controls) peptides in a solution containing a very simple model of a cell membrane - lipid vesicles. Having entrapped a self-quenching dye inside these vesicles, we can measure fluorescence increase after addition of $R9$-, $R4$-, $K9$- or calcium-containing buffers.

1. Molecular Dynamic Simulations

Within the process of investigating the world around us, molecular dynamic simulations stand somewhere between the theory and experiment, closely linked to both. Their basis is rooted in Newton's laws of classical mechanics, which provide the rules on how simulated systems can behave and evolve in time. Experimental results can serve as a test of the simulation models and vice versa. Correctly performed simulations can help scientists to visualise the investigated system with an atomic resolution during a so-called computer experiment, predict its properties and perhaps suggest a new experiment to perform.

The simulational methods can be approached in two ways. We can consider atoms and the bonds between them a group of points of a certain mass ("balls") connected by springs representing bonds with some defined spring constants (elasticity). The initial properties such as spring constants or equilibrium bond lengths are derived from experiments and more precise calculations based on the quantum electronic structure. These kinds of simulations are called "classical" force field-based simulations.

The second method is called "ab-initio" or *from first principles*. Applying Born-Oppenheimer approximation the movement of electrons, treated by a quantum theory methods, is separated from nuclear motions, that are solved classically. These simulations are more computationally demanding than the force field ones due to the necessity of solving the electronic Schrödinger equation, typically by density functional methods.

1.1 Force Field MD Simulations

Classical empirical force field molecular dynamic simulations do not explicitly compute electron positions - they only deal with the movements of atomic nuclei. Molecular dynamics computes forces and moves the atoms according to these forces - i.e., it solves Newton's equations for N interacting atoms.

The force field parameters are fitted experimentally and/or from ab-initio simulations. Each force field uses a slightly different set of parameters and is often aimed at investigating a particular set of molecules such as lipids, proteins or DNA bases.

1.1.1 Energy

The potential energy functions determine the forces acting in the investigated system. The total potential energy in the empirical force field E_{pot} consists of a valence part E_{val} and the non-bonding terms E_{nb}:

$$E_{pot} = E_{val} + E_{nb} \tag{1.1}$$

The valence part 1.2 describes the covalent terms, whereas Coulomb and van der Waals interactions, as well as hydrogen bonds, belong to the non-bonding term 1.3:

$$E_{val} = E_b + E_{ang} + E_{tor} + E_{inv} + E_{UB} \tag{1.2}$$

$$E_{nb} = E_{Coul} + E_{vdW} + E_{Hb} \tag{1.3}$$

Bond Stretch Term

In a harmonic approximation 1.4, the bond strength term is given by a force constant k_b, the equilibrium bond distance r_0 and the actual bond distance r:

$$E_b = \frac{1}{2}k_b(r - r_0)^2, \tag{1.4}$$

A Morse potential can be used instead to obtain a more accurate description of the bonding term.

Angular Term

Similarly as the bond stretch term, the energy term of the angle deformation between three atoms i, j and k can be defined as:

$$E_{ang} = \frac{1}{2}k_{ang}(\theta_{ijk} - \theta_0)^2, \tag{1.5}$$

where the k_{ang} is a force constant, θ_{ijk} the instant angle between particles i, j and k and θ_0 the equilibrium value of the given angle. This term is becomes problematic for angles near 180°, therefore higher powers of the angular deviation can be introduced to the equation, or it can be substituted by some form of cosine expansion for (almost) colinear angular arrangements.

Dihedral Angle Torsion

Let us assume four atoms bonded in a row: i, j, k and l. Atoms i, j and k define one plane and atoms j, k and l define another plane. Dihedral angle torsion describes the angular motion between these two planes. Denoting the equilibrium angle ϕ_0 and the actual one ϕ_{ijkl}, we define the torsional energy term:

$$E_{tor} = \frac{1}{2}k_\phi[1 + \cos(m(\phi_{ijkl} + \phi_0))] \tag{1.6}$$

Here m relates to the periodicity in the potential and k_ϕ defines the rotational barrier.

Inverse Term

An inverse term can describe the deviation from a planar alignment by three possible ways: Umbrella, Amber inversion or Charmm.

Urey-Bradley Term

Sometimes the Urey-Bradley term is added as a potential between the two atoms that are both bonded to a third one, but not to one-another.

Coulomb Interaction

The electrostatic interaction between two atoms of the distance of r_{ij} with (partial) charges q_i and q_j is defined by a Coulomb term

$$E_{Coul} = \frac{1}{4\pi\varepsilon}\frac{q_i q_j}{r_{ij}} \tag{1.7}$$

with ε being the dielectric constant.

Van der Waals Interactions

The van der Waals interactions between atoms i and j of a distance r_{ij} can be described by an exponential term 1.8 (where A, B and C are empirical constants), or by the most commonly used Lennard-Jones potential 1.9. The depth of the potential well is described by ε while σ is the distance, where the potential value is zero and relates to the radii of the described atoms.

$$E_{vdW-e} = A\exp(-Br_{ij}) - Cr_{ij}^{-6} \tag{1.8}$$

$$V_{LJ} = 4\varepsilon\big[\big(\frac{\sigma}{r_{ij}}\big)^{12} - \big(\frac{\sigma}{r_{ij}}\big)^{6}\big] \tag{1.9}$$

In the case of two different atoms, the values of $\sigma = \sigma_{ij}$ and $\varepsilon = \varepsilon_{ij}$ used in the equation 1.9 have to be calculated from σ_i, σ_j and ε_i, ε_j belonging to the specific atoms. There are various methods to perform this calculation, called the combination (or mixing) rules. In this work the Lorentz-Berthelot rules are used:

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \tag{1.10}$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \tag{1.11}$$

Hydrogen Bonds

With a donor-acceptor distance of d_{ij} and empirical constants A and B, hydrogen bond energy term, which is a term sometimes added but most often omitted from the force field, looks like:

$$E_{Hb} = Ad_{ij}^{-12} - Bd_{ij}^{-10} \tag{1.12}$$

1.1.2 Propagation Algorithms

Having the potential, forces acting upon atoms and subsequently their accelerations can be evaluated with Newton's equations. To numerically solve the equations, simulation time has to be discretized. The smallest time unit in which the atoms move is defined as a so-called Δt timestep. Its value for classical all-atom simulations is usually between 0.5 fs and 2 fs.

A propagator is an algorithm which, knowing the positions and velocities of atoms at the time t_0, numerically calculates the positions and velocities at the time $t = t_0 + \Delta t$. This trajectory "slicing" introduces small numerical errors to

the system which results in a non-reversibility of the computer simulation runs in comparison with the microscopic reversibility of the equations of motion.

Performing a Taylor expansion of $r_i(t + \Delta t)$ and $v_i(t + \Delta t)$ to the third and second term respectively, one obtains the Euler propagation algorithm. Similarly, using Taylor expansion of $r_i(t + \Delta t)$ and $r_i(t - \Delta t)$ the Verlet (Verlet [1967]) integrator can be derived. As a result it evaluates $r_i(t + \Delta t)$ as a function of positions in two previous steps but does not directly obtain the velocity $v_i(t)$. The often used Velocity Verlet Integrator, incorporated in a Gromacs software package (Hess et al. [2008]) and used later in this work, yields $r_i(t + \Delta t)$ and $v_i(t + \Delta t)$ at the same time.

1.2 The Techniques Used in Force Field-Based MD as well as in Ab-initio MD Simulations

1.2.1 Periodic Boundary Conditions

Periodic boundary conditions (PBC) is a widely used concept to model bulk liquids and solid phases (crystals) as well as their interfaces. The idea is that instead of simulating computationally demanding larger systems in order to capture bulk effects and their consequences accurately, one encapsulates smaller and computationally less expensive system in a simulation box of a given size LxLxL. When an atom reaches the box border at $L/2$ and should continue further out of the box, it disappears at its original position and a new, identical atom enters the system at the position $-L/2$ (at the other side of the box). This trick allows to effectively simulate an infinite system composed of copies of the small system.

To prevent a particle from interacting with its periodic image, a cutoff range of a value of less than half the box size is introduced. This enables the atom to interact only with atoms within the unit cell cutoff range. In the case of short-range interactions (for example the Lennard-Jones potential), the cutoff does not introduce substantial inaccuracies into the computation. It is however not the right tool for dealing with long-range interactions, such as electrostatics. For these interacions the Ewald summation is used.

A system can be put in a cubic box most often used to investigate, e.g. processes in the liquid environment. However, other shapes such as prismatic cells for liquid slabs are also possible.

1.2.2 Ewald Summation

The Ewald summation decomposes the total electrostatic potential of a system with periodic boundary conditions into two parts. The calculation of the short-ranged potential, performed in real space, converges quickly. The rest, i.e. the long-range potential, is varying slowly. The charges are Gaussian-distributed onto a three-dimensional mesh in real space and subsequently converted into the reciprocal space using the Fourier transform. Here the electrostatic potential and the forces are computed quickly, and they are transformed back to the real space and interpolated at the original point charge locations. The most costly part, the Fourier transform, can be accelerated by using the Fast Fourier Transform (FFT)

algorithm. The popular Ewald summation algorithms with incorporated FFT on a grid include Particle-Mesh Ewald method (Essmann et al. [1995]), used in this work.

1.2.3 Statistical Ensembles

By integrating Newton's equations the constant energy (microcanonical) ensemble is explored. If however, external pressure or a heat bath is introduced, the total energy ceases to be conserved. We distinguish various statistical ensembles depending on which state variables are fixed. The NVE, NVT and NpT belong to the most commonly used statistical ensembles.

Any state of a many-particle system is represented by a point in the phase space called microstate. The time evolution of this state can be represented by a trajectory in the phase space. The ergodic theorem postulates that, in an ergodic system, the trajectory of the system "completely" fills the phase space, so the system visits all its possible microstates. Then the averaging over time is equivalent to the averaging over properties derived as a probability distribution within the ensemble.

Microcanonical Ensemble

The microcanonical ensemble corresponds to an adiabatically isolated system. It is a constant-energy and constant-volume ensemble with a fixed number of particles (NVE). There is no pressure or temperature control. The Newton equations are solved and the total energy surface is explored.

Canonical Ensemble

By fixing the number of particles in the system N, its volume V and the temperature T, the so-called Helmholtz canonical or isothermal (NVT) ensemble is created. It is often used for small systems to avoid rather significant density fluctuations.

Isobaric-isothermal Ensemble

In this case the number of system's particles, its temperature and pressure are controlled creating the isobaric-isothermal (NpT) or the so-called Gibbs canonical ensemble. Laboratory experiments are usually held under similar conditions. This ensemble can also be used during the equilibration phase of a simulation to obtain the correct temperature and pressure.

Grand-canonical Ensemble

In the case of fixed μVT, the number of particles can be changed. The chemical potential μ, volume and temperature of the system are kept fixed.

1.2.4 Barostats

In this and the following sections, several ways of controlling pressure (barostats) and temperature (thermostats) are described. Using these special algorithms

ensures that the values of p or T remain fluctuating around a desired "fixed" one.

Berendsen Barostat

The volume of a cubic system is scaled by a factor of β^3 (1.13) while the atomic positions scale accordingly. Δt is a timestep of the simulation and τ corresponds to a coupling constant. This barostat is quite robust and suitable for equilibrations, but it does not sample the NpT ensemble correctly. p_0 denotes the target pressure while p the actual one.

$$\beta = \left(1 - \frac{\Delta t}{\tau}(p - p_0)\right)^{\frac{1}{3}}, \tag{1.13}$$

Parrinello-Rahman Barostat

Introducing the friction term into the Hamiltonian of an investigated system results in proper ensemble sampling. This barostat is less robust and widely recommended for a production run.

1.2.5 Thermostats

Temperature (as well as pressure) is an intensive thermodynamic quantity that is well-defined only at equilibrium. It is related to the average kinetic energy of system particles through the equipartition theorem, which assigns an energy of $k_B T/2$ to each system's degree of freedom.

The initial temperature of the system is reached by randomly assigning velocities to the system's particles according to a Maxwell-Boltzmann distribution corresponding to the predefined target temperature. The x, y and z components of the velocities are assigned according to the Gaussian distribution.

Controlling the temperature of a system corresponds to coupling with a heat bath of a set temperature. The methods to ensure this include thermostating by scaling velocities, adding stochastic forces and/or velocities or using extended system methods.

Direct Velocity Rescaling

The most basic and quick way to reach the target temperature is to directly rescale the velocities every n-th step by a factor of $\lambda = (T/T_0)^{1/2}$. Here T_0 denotes the target and T the instant temperature of a system at a given time. This method is not suitable for production run since it does not generate a correct canonical ensemble.

Berendsen Thermostat

The Berendsen thermostat does not drastically enforce the target temperature to the system but rather gently pushes it towards T_0 such that the speed of the temperature change is proportional to the difference between T_0 and T. Each particle's velocity scaling takes place every n-th step by a factor of λ:

$$\lambda = \left(1 + \frac{\Delta t}{\tau}(T - T_0)\right)^{\frac{1}{2}}, \tag{1.14}$$

where Δt is the timestep of scaling. The characteristic relaxation time of a thermostat is related to the constant τ, which describes the strength of system's coupling to a heat bath. This thermostat does not strictly generate a canonical ensemble either.

Nosé-Hover Thermostat

The action of the external heat reservoir is incorporated into the calculations as an additional degree of freedom in the system's Hamiltonian, that can be interpreted as a friction between the system and the reservoir. The equations of motion obtained by this modified Hamiltonian ensure that the canonical ensemble is correctly sampled.

Velocity-Rescaling Thermostat

Having the original Berendsen thermostat advantages like stability and speed, by adding a stochastic term the proper kinetic energy distribution is established, therefore the canonical ensemble is sampled correctly by this thermostat (Bussi et al. [2007]).

1.3 Ab-initio MD Simulations

Ab-initio, molecular dynamic simulations use the classical Newton equation for propagating the motion of the atomic nuclei, and subsequently insert the nuclear coordinates to the time-independent Schrödinger equation for electrons. This approach thus does not use empirical force fields, which should result in an increase in accuracy at the expense of dramatically increased computational costs. Therefore, only small systems at short time-scales can be simulated in this way.

The Hamiltonian of a system containing n electrons and M atom cores in a vacuum consists of the kinetic part, electrostatic interaction and a repulsion term:

$$\begin{aligned}\hat{H} = &-\frac{1}{2}\sum_{A=1}^{M}\frac{1}{M_A}\Delta_A - \frac{1}{2}\sum_{i=1}^{n}\Delta_i - \sum_{i=1}^{n}\sum_{A=1}^{M}\frac{Z_A}{\left|\vec{r_i} - \vec{R_A}\right|} \\ &+ \sum_{i=1}^{n}\sum_{j>i}^{n}\frac{1}{|\vec{r_i} - \vec{r_j}|} + \sum_{A=1}^{M}\sum_{B>A}^{M}\frac{Z_A Z_B}{\left|\vec{R_A} - \vec{R_B}\right|},\end{aligned} \tag{1.15}$$

where M_A are masses of atom cores, $\vec{r}$ and $\vec{R}$ describe the positions of electrons and cores and Z is the proton number of a given atom core. This Hamiltonian is missing the interaction with an external field and relativistic effects, that can be omitted in the case of light nuclei. For core electrons, the pseudopotential method is employed. The problem is thus reduced to the calculation of the valence electrons. In this work, the Goedecker et al. [1996] pseudopotentials are used.

1.3.1 Born-Oppenheimer and Adiabatic Approximation

Since atomic nuclei are at least three orders of magnitude heavier than electrons, it is safe to assume that electrons move much faster. In the first approximation, this results in the electrons moving in a field of fixed nuclei and atom cores slowly reacting to the effective field created by electron clouds. Therefore, the Schrödinger equation for this system can be split into two equations: one accounts for the motion of atomic nuclei in the electron cloud, while the other describes the electronic structure around the nuclei. The wavefunction ψ of the system is a product of the nuclear part ϕ and the electronic part φ, parametrically dependent on the positions of the atomic cores:

$$\psi(\vec{r}, \vec{R}, t) = \phi(\vec{R}, t)\varphi_{\vec{R}}(\vec{r}, t) \tag{1.16}$$

Using the separation 1.16 and inserting the time-independent Schrödinger equation for electrons to the equation for the whole system ($\hat{H}\psi = E\psi$), we obtain the Schrödinger equation for the wavefunction of atomic nuclei a.k.a the adiabatic approximation. Further assuming, that changing the positions of the nuclei, the φ electronic wavefunction changes only slightly, we obtain the Born-Oppenheimer approximation. Finally, we assume thet the nuclear motions can be described classically. The searched-for wavefunction can be written as:

$$\psi(\vec{r_i}, \vec{R_j}, t) = \sum_k a_k(R_j, t)\psi_k(\vec{r_i}, \vec{R_j}) \tag{1.17}$$

Here $\psi_k(\vec{r_i}, \vec{R_j}, t)$ are wavefunctions resulting from adiabatic approximation.

1.3.2 Density Functional Theory

The Density Function Theory (DFT) is widely used and popular for its computational cost/accuracy ratio caused by a computationally rather cheap inclusion of the correlation energy. Two Hohenberg-Kohn theorems form the basis of the DFT. The first theorem declares that the potential energy, and hence the total energy of the system is a unique functional of the electron density ρ. The second theorem, in analogy to the variational principle, states that for any functional of the electron density $E[\rho']$:

$$E_0[\rho] \leq E[\rho'] \tag{1.18}$$

Here $E_0[\rho]$ is the energy of the ground state. The density functional can be written as

$$E[\rho] = \int \rho(\vec{r})\nu(\vec{r})d\vec{r} + F[\rho], \tag{1.19}$$

where $\nu(\vec{r})$ is the external potential and $F[\rho]$ is called an universal functional:

$$F_[\rho] = T[\rho] + J[\rho] + E_{xc}[\rho] \tag{1.20}$$

Here $T[\rho]$ is the kinetic energy functional, $J[\rho]$ stands for the coulombic electron repulsion and $E_{xc}[\rho]$ denotes the problematic exchange-correlation term. This term is not known exactly - if it were, it would be possible to compute the

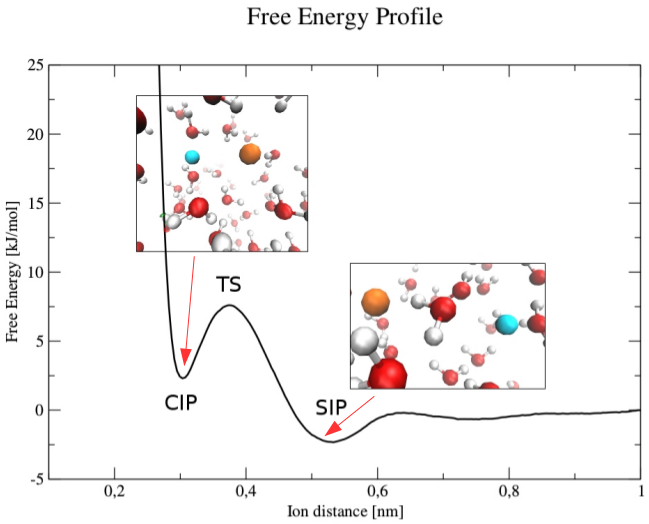


Figure 1.1: The free energy profile of ion association.

exact (non-relativistic) value of the energy. Many approximations were derived to account for this term. In the present work, the widely used BLYP functional (Lee et al. [1988]; Becke [1988]) is used.

1.4 Ion Association

Ion association or ion pairing in a solution occurs when two ions get close to each other (at the order of several Angströms). Their closest conformation, called the Contact Ion Pair (CIP) corresponds to the left-side free energy minimum in the Figure 1.1. It is also depicted in the top left corner showing the ion "balls" (orange and cyan) in the aqueous environment. As the ionic distance increases, the ions reach the metastable Transition State (TS) referring to the interionic distance with the local maximum of the free energy. Separating the ions into another free energy minimum results in a solvent molecule belonging to both ions' first solvation shell - a geometry called the Solvent-Shared Ion Pair (SIP), as can also be seen at the simulation snapshot inserted to the bottom right of the Figure 1.1. Further increasing the distance between the ions results in their first solvation shells occupied by non-shared molecules of a solvent, ergo in a two fully solvated ions (Solvent-separated Ion Pair, SSIP).

The free energy profile describes the ion association as a function of the interionic distance. This profile determines for example whether the ions will form a stable CIP or prefer to stay further away from each other and how stable the ion pair formation is.

1.4.1 Radial Distribution Function and Potential of Mean Force

The radial distribution function (RDF) describes the radial organisation of particles around a chosen (central) particle. It is a static properties analysis tool providing an integrated picture of the system structure and solvation shells as

well as phase information or neighbouring preferences of studied particles. The RDF corresponds to a normalised probability of finding particle i at a distance of r from particle j:

$$g(r) = \frac{1}{4\pi r^2} \frac{1}{N\rho} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \delta(|r_{ij}| - r) \tag{1.21}$$

Here the first term ensures the normalisation as the volume of the spherical shell increases with growing r. The second term relates to the ideal gas normalisation, where N is the number of particles and $\rho = N/V$ the density, so that thr RDF's value is 1 at $r \to \infty$.

The integral over the radial distribution function's peaks refers to the number of atoms in the corresponding solvation shell of a central molecule. These coordination numbers (CNs) play an important role for example in analysing the hydration structure. They are visible as plateaux after integrating the RDF.

The radial distribution function is specific for individual systems. In the case of an ideal gas, it is a constant of 1, while for an ideal crystal system the narrow peaks appear corresponding to the inter-atomic distances in the crystal. An RDF of a Lennard-Jones liquid is zero for interionic distances close to 0, then with increasing the interatomic distance it rises to its global maximum corresponding to the CIP, next goes through the well of the TS to its another, much smaller SIP maximum, after which it smoothly converges to 1. A comparison of the RDF to the neutron scattering or X-ray diffraction experiments is possible through the structure factor via a Fourier transform.

The so-called Potential of Mean Force (PMF, or a free energy profile) provides equivalent information to the radial distribution function:

$$w(r) = \int_{r_0}^{r} \bar{F}(r') \, \mathrm{d}r' + 2k_B T \ln(r) + const, \tag{1.22}$$

where the first part represents the integral of the mean force $\bar{F}$ and the second part is a so-called volume entropy correction - where k_B is Boltzmann constant and T the temperature of the studied system. It holds:

$$w(r) = -k_B T \ln g(r), \tag{1.23}$$

1.4.2 Ways to Obtain PMF or RDF

There are several ways to obtain the PMF or the RDF from computer simulations (Trzesniak et al. [2007]) varying in the computational demands and accuracy. Different methods are suitable for different system conditions.

Calculation from Direct Simulation

Provided a long enough MD simulation generates a converged sampling of the statistical ensemble, we can get the $g(r)$ for the atoms of interest by processing snapshots of a trajectory. This method is inefficient for systems with large potential barriers, where it requires an unfeasibly long simulation to provide a good sampling of all the possible states. As an example relevant for this work, long

simulations become particularly difficult for computationally very demanding ab-initio molecular dynamics.

Umbrella Sampling Simulations

In this method, we insert a set of artificial potentials between the interacting particles (ions in our case):

$$V(r) = \frac{1}{2}k(r - r_0)^2, \tag{1.24}$$

where r is their actual mutual distance, r_0 is the position of a minimum of this newly placed (umbrella) potential and k is a force constant of this potential. Then the simulation is run - the system is set to propagate in time, while thanks to the umbrella potential the ion-ion distance fluctuates around r_0. The simulation with potential 1.24 is called an umbrella window.

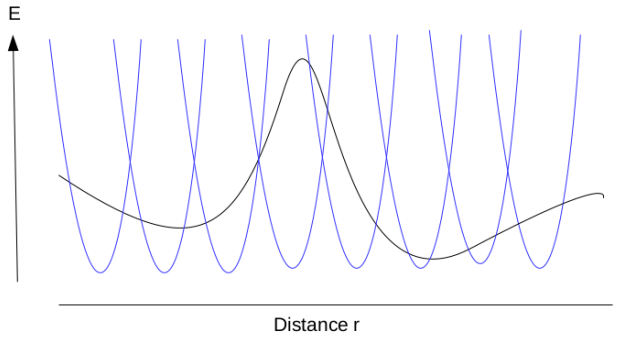


Figure 1.2: Several umbrella potentials (blue) to sample the PMF (black).

The goal is to have a sufficient number of the umbrella windows for different values of r_0 (see Figure 1.2), such that the resulting histograms of ion distance distributions in these simulations overlap and cover the whole interval of interesting interparticle separations (as can be seen in Figure 1.3). Each histogram provides a piece of information about the RDF and thus also the PMF curve around the r_0 distance. Since the histograms are overlapping, it is possible to glue together the whole PMF.

Several algorithms are used to transform the umbrella windows data to the PMF. In this work a tool from Gromacs package g_wham (Hub et al. [2010]) and the Grossfield Lab's (Grossfield [2013]) implementation of the Weighted Histogram Analysis Method (WHAM) is used.

Metadynamic Simulations

This method was introduced relatively recently by Laio and Parrinello [2002]. The authors propose to insert a small biasing potential every metastep $\delta\sigma$ so that the most preferred conformations of the system are later prohibited. The system is therefore forced to overcome barriers and also visit other states on the potential surface, including those that would be otherwise avoided.

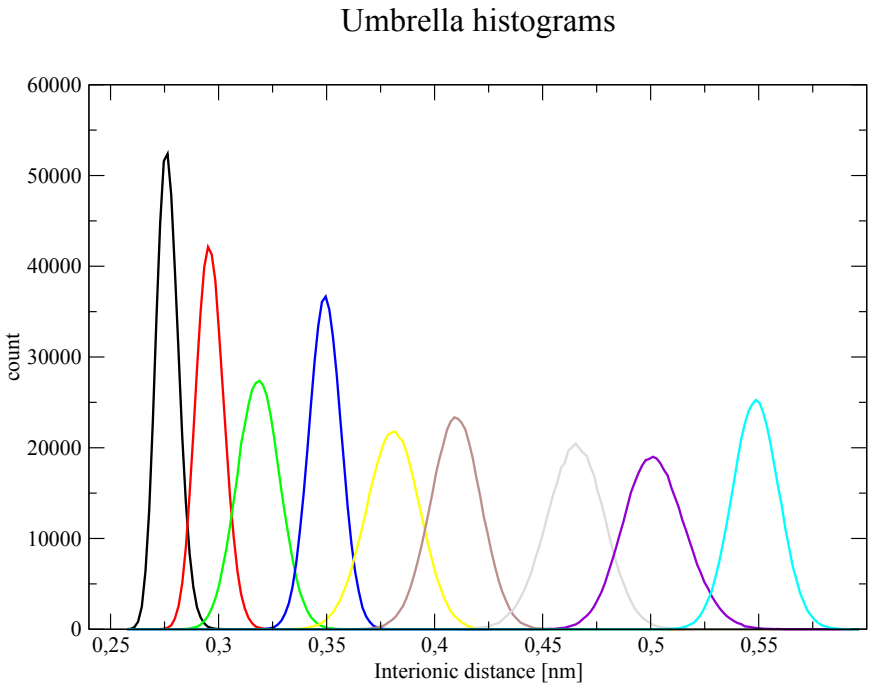


Figure 1.3: Umbrella histogram - full FF, small Ca^{2+}-Cl^{-} system.

Blue-Moon Sampling

Here, the interionic distances are restrained for a number of chosen distance values. For each distance value, the molecular dynamic simulation is performed, and the system propagates in time with the fixed distance of those two atoms. Subsequently, the average force between fixed atoms $\bar{F}$ is evaluated for each distance simulation and is then integrated according to the equation 1.22.

2. Simulations of Ca^{2+} in Aqeuous Solutions

For the study of the calcium dications' properties in aqueous solutions we chose to compute the free energy profile of Ca^{2+}-Cl^- dissociation and subsequently the water coordination number as a function of the Ca^{2+}-Cl^- interionic distance.

2.1 Force Field-based Molecular Dynamic Simulations

First, the calcium-chloride PMF was determined employing the classical force field-based molecular dynamics. Two systems were investigated. A large system containing one calcium dication and one chloride anion in a solution of 2220 SPC/E (Berendsen et al. [1983]) water molecules enclosed in a 4 nm cubic box employing periodic boundary conditions (PBC) served as a benchmark for calculations of our chosen smaller system, which will be directly comparable to ab-initio MD simulations. The small system containing one Ca^{2+}, one Cl^- and 64 SPC/E water molecules was enclosed in a cubic box with a size of 1.25 nm, see Figure 2.1. Within these systems, three different setups for three different Ca^{2+} and Cl^- parametrisations were investigated, see Table 2.1.

The number of water molecules in both unit cells was chosen to mimic the experimental salt solution density at a room temperature. Also, note that both systems have a non-zero total charge, which is taken care of by a homogenous screening charge in the Ewald summation.

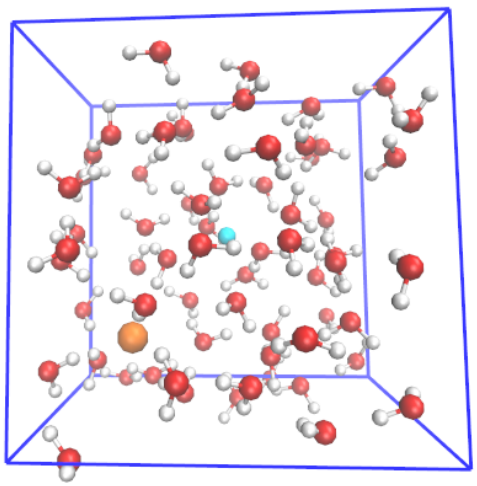


Figure 2.1: Picture of the small system. Red colour is assigned to water oxygen atoms and white to water hydrogen atoms. Chloride anion is orange and calcium dication cyan.

The first set of parameters, named full, uses the same Lennard-Jones σ and ϵ parameters and charges as defined in the Gromos 53a6 force field (Oostenbrink

force-field	ion	σ [Å]	ϵ [kJ/mol]	charge [e]
full	Ca^{2+}	2.8196	0.5072	+2.0
	Cl^-	4.4499	0.4184	−1.0
ECC	Ca^{2+}	2.8196	0.5072	+1.5
	Cl^-	4.4499	0.4184	−0.75
ECCR	Ca^{2+}	2.5376	0.5072	+1.5
	Cl^-	3.7824	0.4184	−0.75

Table 2.1: Lennard-Jones parameters used in force field based MD studies.

et al. [2004]) for Ca^{2+} and parameters from Dang and Smith [1995] for chloride anion. This force filed and many other nowadays commonly used ones do not account for the electronic polarisation (i.e. they are so-called non-polarizable force-fields). To account for electronic polarizability, Leontyev and Stuchebrukhov [2011] developed new parameters by scaling the charges of ions by the factor $1/\sqrt{\epsilon_{el}}$, where $\epsilon_{el} = 1.78$ is the electronic part of the dielectric constant of water. Therefore in the Electronic Continuum Correction (ECC) force field parameters the Ca^{2+} and Cl^- ion charges are scaled by roughly 0.75.

This charge rescaling drop, however, changes the coulombic part of the ion-ion interaction pushing the ions slightly apart at the CIP position. To better fit the neutron scattering data (Badyal et al. [2004]), another correction called Electronic Continuum Correction with Rescaling (ECCR) was made by Kohagen et al. [2014]. The σ value, corresponding to van der Waals radii, was decreased by 10 % for calcium dication and 15 % for chloride anion. Finally, the Lorentz-Berthelot combination rules were used in all the cases.

2.1.1 Computational Details

Large System

The large system containing one Ca^{2+} and one Cl^- ion was simulated for 42 ns including 2 ns as a preequilibration with a timestep of 2 fs. The canonical ensemble (NVT) at a temperature of 300 K controlled by temperature coupling using velocity rescaling with a stochastic term (v-rescale Bussi et al. [2007]) with a time constant of 0.5 ps was employed. The long-range electrostatics was treated by the Particle-Mesh Ewald (PME, Essmann et al. [1995]) method. Long-range dispersion correction for energy and pressure was applied. The initial velocities were generated according to a Maxwell distribution at the temperature of 300 K. All the bonds and angles were constrained by a LINCS algorithm (Hess et al. [1997]).

Umbrella windows were constructed for the Ca^{2+}-Cl^- distances between 2.5 Å and 10.0 Å with a spacing of 0.5 Å. Two additional windows at 3.25 Å and 3.75 Å were added in order to sample the area of interest better. The choice of the force constants used for each window was based on previous calculations, see Table 2.2.

The Gromacs 4.6.1 program package (Hess et al. [2008]) was used to run molecular dynamic simulations and we used the g_wham (Hub et al. [2010]) analysis implementation to obtain the free energy profiles.

Additionally, the role of the cutoff constant was investigated - two different umbrella sampling simulation sets were performed, with a cutoff of 1.5 nm as

Ca^{2+}-Cl^- distance [Å]	force constant [kJ/mol.nm^2]
2.5	20000
3.0	50000
3.25	30000
3.5	50000
3.75	20000
4.0	20000
4.5	10000
5.0	10000
5.5	10000 (20000*)
6.0 and more	10000

Table 2.2: Force constants used for umbrella windows. The force constant marked by asterix symbol at the 5.5 Å distance was used in the case of small system, as is discussed in the corresponding section.

a value closer to half of the box and another cutoff of 0.62 nm consistent with the small system.

Small System

The smaller system was simulated in a similar manner as the bigger one above. In this system setup there were only 9 umbrella windows (ending at 5.5 Å included) due to restrictions on cutoff constant resulting from the small size of a simulations box. For the same reasons, only the smaller cutoff constant 0.62 nm was used.

This is also the reason for assigning a slightly larger force constant in the case of 5.5 Å umbrella window. It was necessary to hold the Ca^{2+}-Cl^- atoms together a bit more strongly to avoid their distance values closely approaching half of the box size during the simulation runs. The Gromacs program does not allow for this situation, since increasing ion distance above half the box size can result in the ions seeing their counterpart's periodic image actually closer to them than half the box size. Therefore, in this case, the force constant of 20000 kJ/mol.nm^2 was employed. This change should not affect the comparison between our systems, especially with our interest lying in the position of ions' RDF CIP and TS as well as the CIP well's depth.

2.1.2 Results

Large System Investigations

Further verification is needed to confirm that our investigated aqueous calcium-chloride system can be shrank to dimensions small enough to be computable by the ab-initio molecular dynamics. Therefore, we first investigate purely by the methods of force field-based molecular dynamic simulations whether our small system is comparable with a bigger one in the investigated properties.

Figure 2.2 shows the differences in free energy profiles for all three investigated force field parameter sets (full, ECC and ECCR), when the value of the cutoff is

either 1.5 nm or 0.62 nm. These free energy profiles were calculated by umbrella sampling simulations for the big systems. The entropy correction was employed.

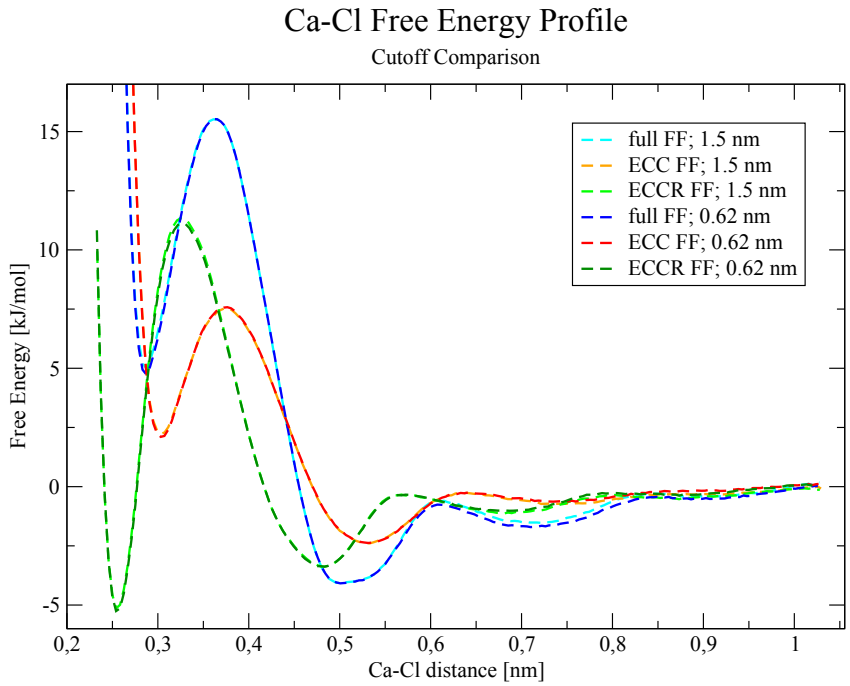


Figure 2.2: Investigations on how the cutoff value affects the free energy profiles for our three force field parameter sets in a big system.

The profiles shown in a Figure 2.2 are basically the same for the same Ca^{2+}-Cl^- force field parameters. Therefore we conclude that the value of the cutoff does not have a significant impact on the free energy profiles.

With the use of a cutoff value of 0.62 nm we investigate the free energy profile difference resulting from reducing the size of the box - Figure 2.3. The profile differences between big and a small systems for the same force field parameters are now non-negligible, although the shape of the peaks, CIP, TS and SIP positions stay almost the same. Further investigations showed this small difference is caused by the significantly different effective ion concentration in the studied systems (not shown here).

Results of a Small System Simulations

The resulting free energy profiles obtained by umbrella sampling simulations for the small system are shown in the Figure 2.4. Since the simulation box in this case is quite small and as Figure 2.3 shows the energy profiles has not yet reached the plateau at 0.55 nm, it was not possible to shift the results in Figure 2.4 so that the most distant energy value would reach the asymptotic value of zero. For the sake of easy visual comparison, the profiles are instead shifted so that the SIP of each curve lies on zero free energy value.

The CIP, TS and SIP positions of the Ca^{2+}-Cl^- ion pair are recorded in the Table 2.3 for each force field parameter set.

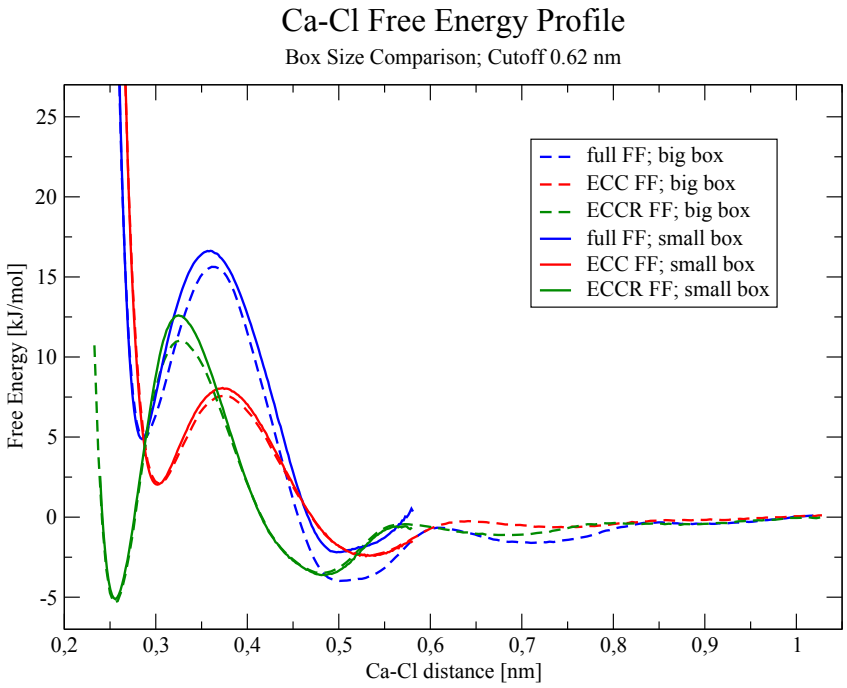


Figure 2.3: Investigations on how the simulation box size affects the free energy profiles for our three force field parameter sets.

FF	CIP [Å]	TS [Å]	SIP[Å]
full	2.85	3.60	4.98
ECC	3.02	3.73	5.30
ECCR	2.56	3.25	4.85

Table 2.3: Contact ion pair (CIP), transition state (TS) and solvent-shared ion pair (SIP) Ca^{2+}-Cl^- ion distances obtained from Umbrella sampling simulations.

The full force field profile is plotted in blue. When changing nothing else but the charges of investigated ions (ECC profile in red), only Coulomb interaction is affected. The Coulomb force between the ions is smaller in this case resulting in a shift of positions of interest (CIP, TS and SIP) towards further distances. In other words, decreasing ions' charges results in a smaller Coulomb attraction. Therefore, the contact ion pair, as well as TS and CIP Ca^{2+}-Cl^- distance value increase.

The ECCR approach (green in 2.4) scales not only the charges of the ions (in the same manner as ECC) but also their Lennard-Jones σ values corresponding to the van der Waals radius of the ions. Specifically, it reduces the van der Walls radius by 10% and 15% for calcium dication and chloride anion respectively. Therefore the ions seem "smaller", and despite their reduced charges, they are positioned at a closer contact ion pair position than in the case of two previous setups.

The free energy difference between the well minima of the CIP and the SIP ($\Delta E_{CIP,SIP}$) as well as the height of transition barrier from the CIP ($\Delta E_{TS,CIP}$) and from the SIP ($\Delta E_{TS,SIP}$) to TS are presented in the Table 2.4.

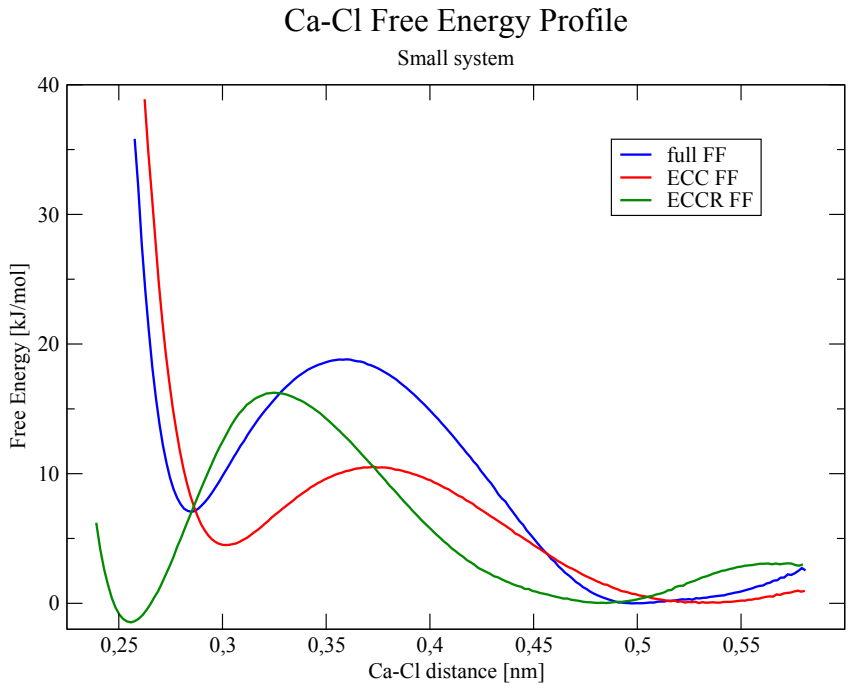


Figure 2.4: Free energy profiles for investigated force field parameters in the case of the small system.

The contact ion pair minima in the case of full and ECC FF are above the solvent-shared ion minima indicating that the latter will be more preferred. On the contrary concerning the ECCR force field, the CIP minimum is lower and therefore slightly more stable.

FF	$\Delta E_{CIP,SIP}$[kJ/mol]	$\Delta E_{TS,CIP}$[kJ/mol]	$\Delta E_{TS,SIP}$[kJ/mol]
full	7.07	11.76	18.83
ECC	4.47	6.04	10.51
ECCR	−1.49	17.72	16.23

Table 2.4: The free energy differences between CIP and SIP (second column), TS and CIP (third column) and TS and SIP (fourth column) for the systems with investigated force field parameters.

One of the very important soluble ions' properties in the intra- or extracellular environment is the number of atoms in their first solvation shell - the coordination number (CN). The CN plays a crucial role in these atoms' binding to the cell membrane or enzymes, effectively changing the curvature of the membrane or enzymes' conformation (and therefore their activity).

Table 2.5 shows the average number of oxygen atoms in the first solvation shell around calcium dication throughout the simulations. These values are calculated for all the umbrella windows in the case of three investigated force field parameter sets: full, ECC and ECCR. A boundary Ca^{2+}-O distance for adding the actual oxygen among the first solvation shell ones was chosen to be 3.25 Å, corresponding to the large zero region between the first and the second peak of Ca^{2+}-O radial

distribution function during these simulations.

Ca^{2+}-Cl^- distance [Å]	full FF	ECC FF	ECCR FF
2.5	6.0 ± 0.2	5.2 ± 0.4	5.0 ± 0.1
3.0	6.3 ± 0.4	5.4 ± 0.5	5.0 ± 0.9
3.25	6.8 ± 0.4	5.7 ± 0.5	5.3 ± 1.0
3.5	7.0 ± 0.2	6.0 ± 0.4	5.4 ± 1.0
3.75	7.1 ± 0.4	6.2 ± 0.5	6.0 ± 0.2
4.0	7.5 ± 0.5	6.4 ± 0.5	6.0 ± 0.2
4.5	7.8 ± 0.4	6.7 ± 0.5	6.0 ± 0.2
5.0	7.9 ± 0.3	6.8 ± 0.5	6.1 ± 0.2
5.5	7.9 ± 0.3	6.8 ± 0.5	6.1 ± 0.3

Table 2.5: Average Ca^{2+}-O coordination numbers for the umbrella windows.

The increase of the Ca^{2+}-O CNs is also shown in the Figure 2.5. Between Ca^{2+}-Cl^- distance of 2.5 Å and 3.0 Å, corresponding approximately to their contact ion pair, the number of oxygen atoms in the first solvation shell of calcium dication is at least smaller by one than the Ca^{2+}-O CNs around 5.0 Å, suggesting the exchange of chloride anion for one or more water molecules, as expected when increasing the Ca^{2+}-Cl^- distance.

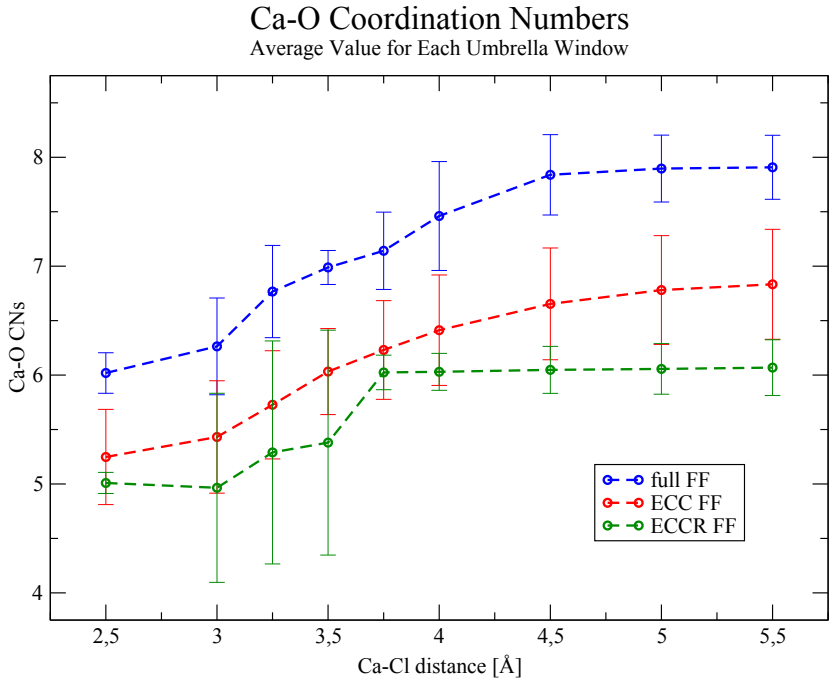


Figure 2.5: The average amount of oxygens in the first solvation shell around calcium dication for each umbrella window of investigated force field simulations.

Small System - Additional Setup Verifications

To check that the timestep of the performed simulations is not too long, which might cause non-realistic system behaviour, another set of simulations with half the used timestep ($\Delta t = 1$ fs) was performed for the small system and full FF. The resulting free energy profile (Figure 2.6) is basically the same as the profile obtained from simulations with 2 fs timestep, confirming the used timestep choice.

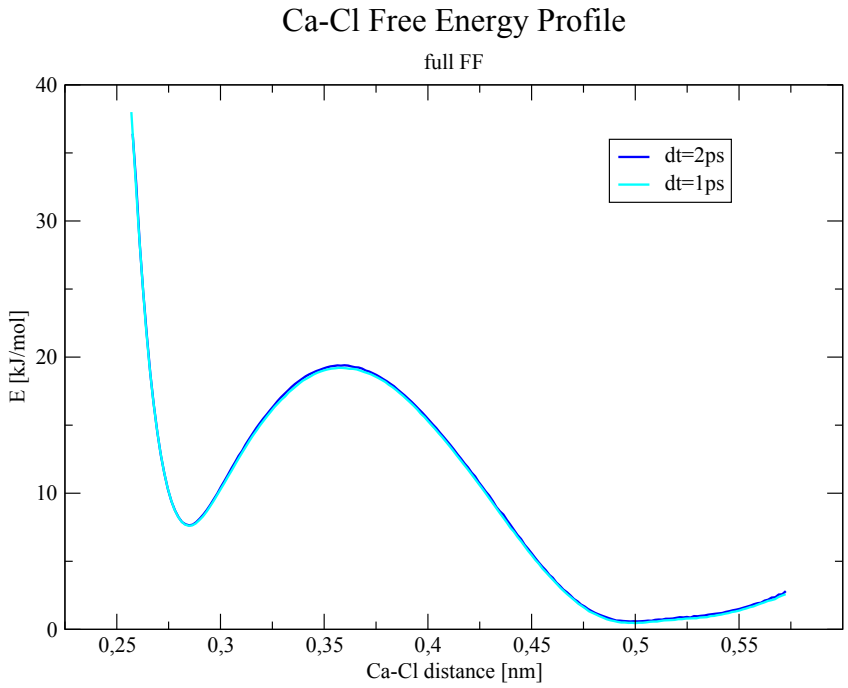


Figure 2.6: Comparison of two timestep values used in the case of force field based molecular dynamic simulations, specifically full FF Umbrella sampling MD simulations performed on a small system from Chapter 3.

The Figure 2.7 shows the effect of the simulation's length. The full FF Ca^{2+}-Cl^{-} small system free energy profile curves are calculated for four cases using differing simulation lengths. The 0.1 ns profile is not smooth and clearly contrasts to the others, implying that the simulations are not fully converged yet. The 1 ns profile is already correct. However, a bit longer simulation is needed for the profile to converge fully.

2.2 Ab-initio MD Simulations

The small system, already investigated by classical force field methods, was consequently studied by much more computationally expensive ab-initio molecular dynamic simulations.

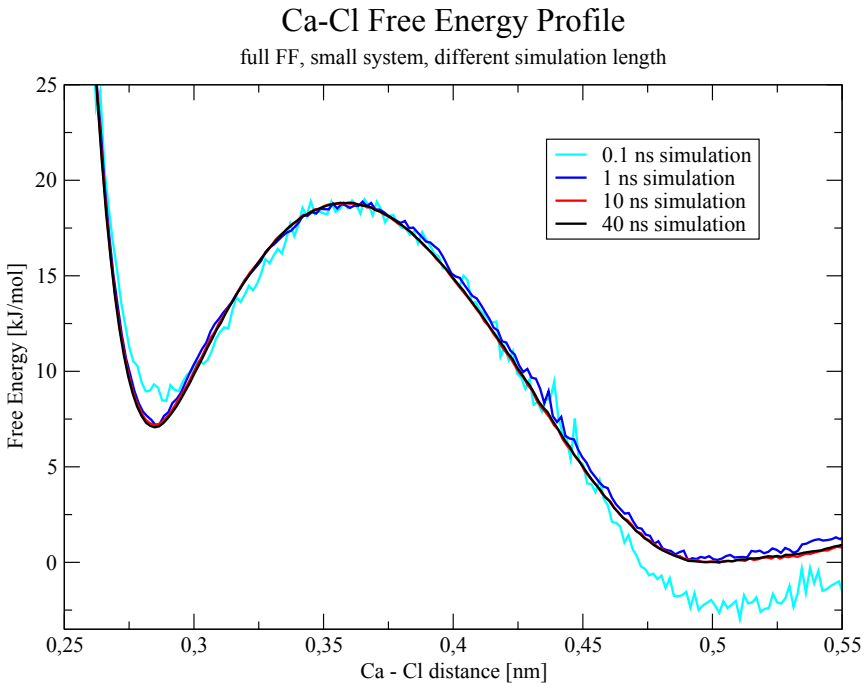


Figure 2.7: Comparison of the results accuracy connected to the simulation length.

2.2.1 Computational Details

As in the previous force field study, the system contained one Ca^{2+} ion, one Cl^- ion and 64 water molecules and was enclosed in a cubic box with a size of 1.25 nm. Periodic boundary conditions were employed with a cutoff of 0.65 nm.

Ab-initio simulations were performed using the CP2K 2.3 program package. Hybrid Gaussian plane waves method (GPW) implementation of the CP2K Quickstep module (VandeVondele et al. [2005]) was employed. The BLYP (Lee et al. [1988]; Becke [1988]) functional with the D2 Grimme correction (Grimme [2006]) to account for dispersion interactions was used along with Kohn-Sham orbitals expanded in a Gaussian basis set DZVP-MOLOPT-SR-GTH (VandeVondele and Hutter [2007]) with Goedecker et al. [1996] GTH-BLYP-q* pseudopotentials, where * denotes the number of valence-treated electrons: 1, 6, 7 and 10 for H, O, Cl^- and Ca^{2+} respectively. A cutoff of 400 Ry was used for the auxiliary plane wave basis set.

Each umbrella window was simulated for 60 ps including the first 10 ps as a equilibration, with a timestep of 0.5 fs. The canonical ensemble was established. The temperature was kept at 300 K by canonical sampling through velocity rescaling (CSVR, Bussi et al. [2007]) thermostat with a 16.68 fs time constant.

For the ab-initio umbrella sampling simulations, a total number of 10 windows were generated. Umbrella potentials were set at the $Ca^{2+}-Cl^-$ distances from 2.5 Å to 6.0 Å shifted by 0.5 Å with two additional ones at 3.25 Å and 3.75 Å. Different force constants of biasing potential were assigned to various windows to obtain better sampling, see Table 2.2.

In the case of various force fields parameters (full, ECC or ECCR), different

dominant calcium-oxygen coordination numbers (CNs) were observed for a corresponding Ca^{2+}-Cl^- ion distance (Table 2.5). We have therefore decided to simulate each umbrella window twice, starting at two different initial conditions. The input topologies were taken from previous force field based MD simulations for two chosen Ca-O coordination numbers - 5 and 6 for umbrella windows smaller or equal to 3.25 Å; 6 and 7 for those larger. The resulting free energy profile was generated by Grossfield Lab's WHAM 2.0.9 (Grossfield [2013]) implementation using all 20 windows.

2.2.2 Results

The final Ca^{2+}-Cl^- free energy profile from ab-initio molecular dynamic simulations is depicted in Figure 2.8. For comparison, there are also the three profiles from previous force field based MD study.

This profile is made from 50 ps data per window (since the first 10 ps were omitted for equilibration), which is a relatively short time due to the demanding ab-initio simulations. Therefore the error bars are non-negligible in the plot. For estimating the error, the umbrella windows were cut into ten 5 ps pieces. For this data cuts the free energy profiles were made. These profiles were then aligned into one point (TS in this case), and for each other point, the average value and standard deviation was calculated.

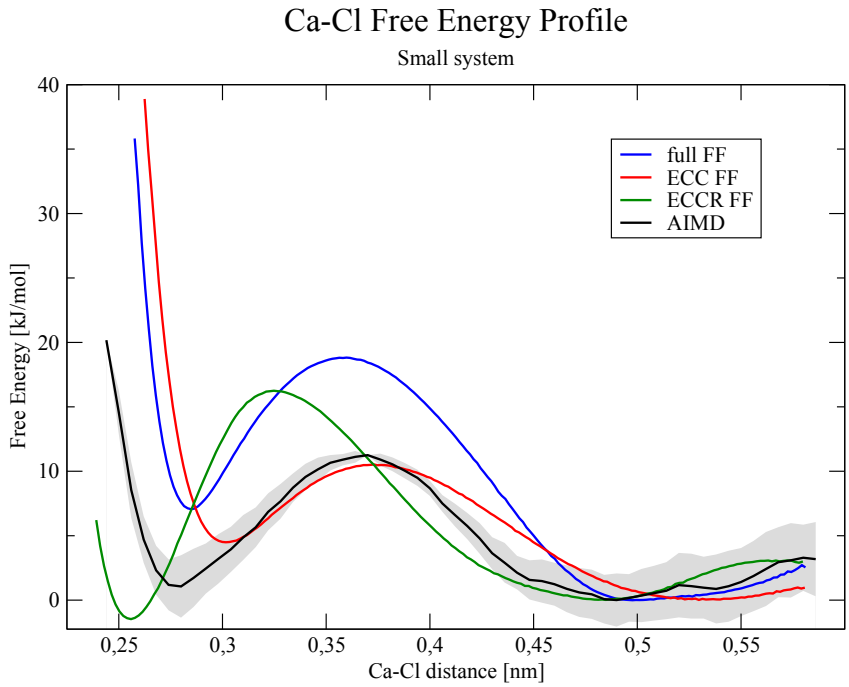


Figure 2.8: Ab-initio MD free energy profile between calcium dication and chloride anion (black).

The position of the AIMD Ca^{2+}-Cl^- contact ion pair (2.8 Å) is between the ECC and ECCR FF CIP. Moving to the transition state at 3.7 Å, the ion pair has to overcome a barrier of 10.2 kJ/mol. From the SIP (at 4.9 Å), the barrier

to TS is 11.2 kJ/mol high. This creates a free energy difference of 1.1 kJ/mol between CIP and SIP. This free energy difference is smaller than the error at CIP or SIP. We therefore conclude that the SIP and CIP lie around the same energy level and are favoured comparably. The closest CIP-SIP free energy difference to this ab-initio MD results was obtained from ECCR FF; both ECC and full FF show a strong preference to the contact ion pair.

The average calcium-oxygen coordination numbers are in the Table 2.6 and Figure 2.9. Here the full FF overestimates the Ca^{2+}-O binding by approximately one oxygen atom. In other words, in the case of the full FF, there is one more water molecule in the Ca^{2+} first solvation shell. The Ca^{2+}-O CNs results for ECC and ECCR FFs are in agreement with the ab-initio MD result within the error bars. Specifically, the ECC FF values correspond with AIMD result very well. The further Ca^{2+} and Cl^- potential properties adjustment, considering neutron scattering (the ECCR correction), moves the force field based simulation results slightly away from the ab-initio ones.

Ca^{2+}-Cl^- distance [Å]	Ca^{2+}-O CNs
2.5	5.4 ± 0.5
3.0	5.6 ± 0.5
3.25	6.0 ± 0.2
3.5	6.2 ± 0.4
3.75	6.5 ± 0.5
4.0	6.3 ± 0.5
4.5	6.8 ± 0.4
5.0	6.8 ± 0.5
5.5	7.0 ± 0.2
6.0	6.9 ± 0.3

Table 2.6: Average Ca^{2+}-O coordination numbers for the ab-initio umbrella windows.

This work was a part of a bigger project and my colleague also computed Ca^{2+}-Cl^- free energy profile in the same manner, but using the Blue Moon Sampling method (Martinek et al. [2018]). Figure 2.10 shows the comparison of these results - the profiles are aligned to the TS maximum, where the error estimation starts in both cases. The umbrella sampling SIP minimum is set to zero. These two free energy profiles are consistent with each other within the error range and also in a good agreement with the profile obtained by Baer and Mundy [2016]. The CIP and SIP minima are within the error around the same energy level. The contact ion pair is positioned to the Ca^{2+}-Cl^- distance of 2.75 ± 0.05 Å and transition state occurs at 3.75 ± 0.05 Å. The SIP location agreement is somehow smaller. Still, it lies around 5.0 Å.

The average Bader charges (Bader [1994]) for Ca^{2+} and Cl^- for the umbrella windows are depicted in the Figure 2.11. Here the blue colour and the left-hand charge scale belongs to calcium dication, red colour and right-hand scale to chloride anion. The displayed charge alterations are caused both by the inter-ionic and ion-water charge transfer. The charge values were obtained using the Bader Charge Analysis program (Arnaldsson et al. [2012]).

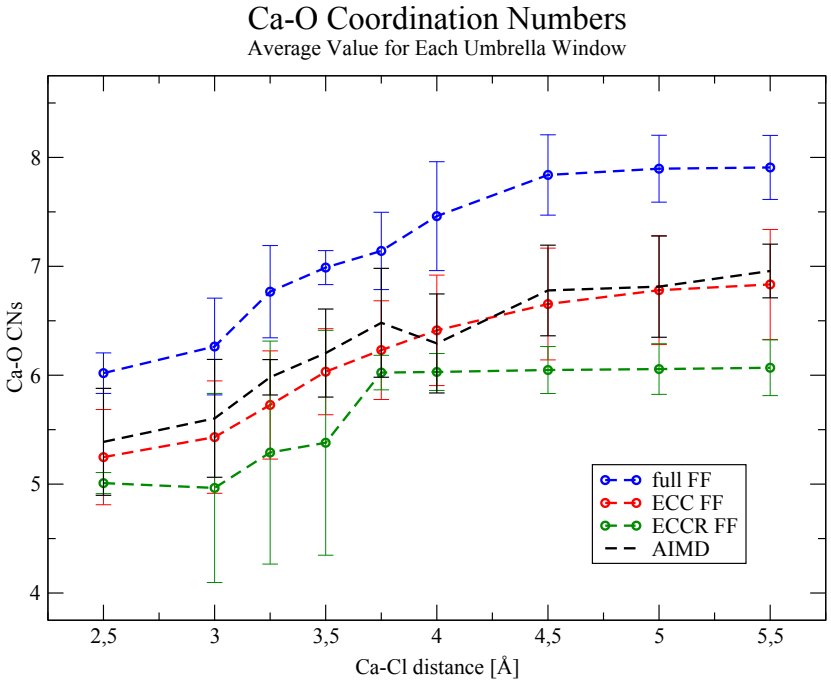


Figure 2.9: Coordination numbers of oxygen around calcium dication for the ab-initio MD (black) as well as for the three studied force field based MD setups.

2.2.3 Discussion

Within our approach of accounting for the electronic polarizability in a formally non-polarisable force field, the original parameters of calcium dication and chloride anion were adjusted to the ECC by scaling the charges by a factor of 0.75. This charge scaling resulted in lowering the free energy barrier between the CIP and SIP as well as a shift of these positions towards longer distances. The Ca^{2+}-O first solvation shell coordination numbers, being previously overestimated, are now in a reasonable agreement with both the ab-initio MD and with experimental results (Megyes et al. [2006]).

However, the ECC radial distribution functions of aqueous $CaCl_2$ solutions were not in perfect accord with the neutron scattering experiments (Badyal et al. [2004]). Thus, further adjustment was needed. Kohagen et al. [2014] reduced the Ca^{2+} and Cl^- Lennard-Jones σ value by 10 % and by 15 % respectively (ECCR correction). This resulted in the energetical preference of the CIP over SIP and, since the van der Walls radii of the ions was decreased, a shift of the CIP positions towards a closer Ca^{2+}-Cl^- distance occurred.

Three free energy profiles obtained by a force field-based molecular dynamics were compared to the profiles obtained by a computationally very demanding ab-initio MD. It turned out that the ECCR approach does not agree fully with the ab-initio profiles. In our article (Martinek et al. [2018]) we therefore proposed a slight change of Ca^{2+}-Cl^- parameters. Our newly suggested ECCR2 parametrization enlarges back the Ca^{2+} σ-parameter by 5% and rescales the Cl^- σ-parameter to 4.10 Å and the ε-parameter to 0.4928 kJ/mol as was suggested in (Pluhařová et al. [2014]). This change represents a reasonable compromise to fit well both

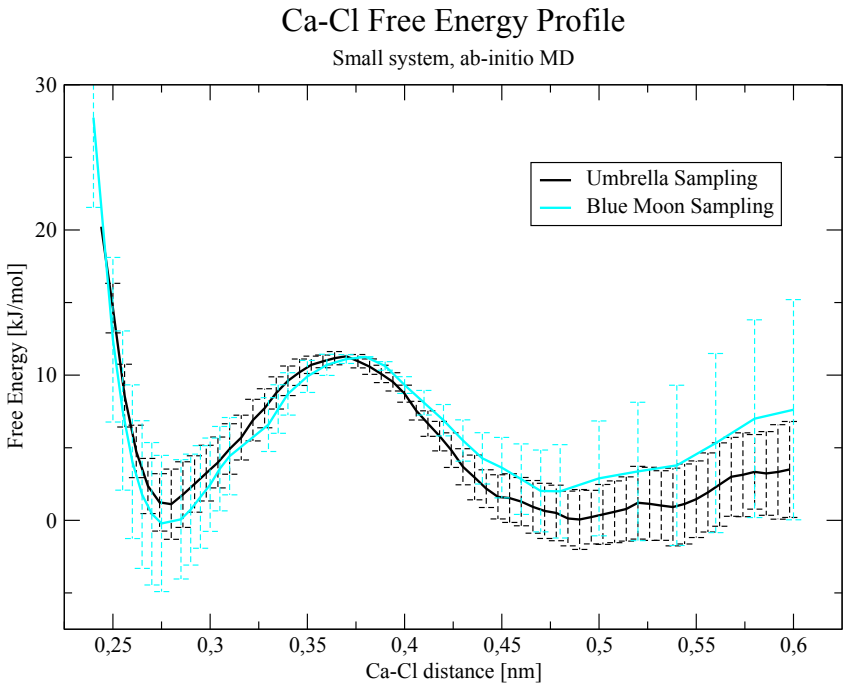


Figure 2.10: Comparison of Ca^{2+}-Cl^{-} free energy profiles obtained by ab-initio umbrella sampling method (black) and ab-initio Blue Moon Sampling method (cyan).

the ab-initio results and neutron scattering data (Badyal et al. [2004]).

An important lesson learned from this study is that in simulations of larger (biological) systems, merely scaling the charge and slightly adjusting the van der Waals parameters of the ions can lead to a significant improvement in accuracy.

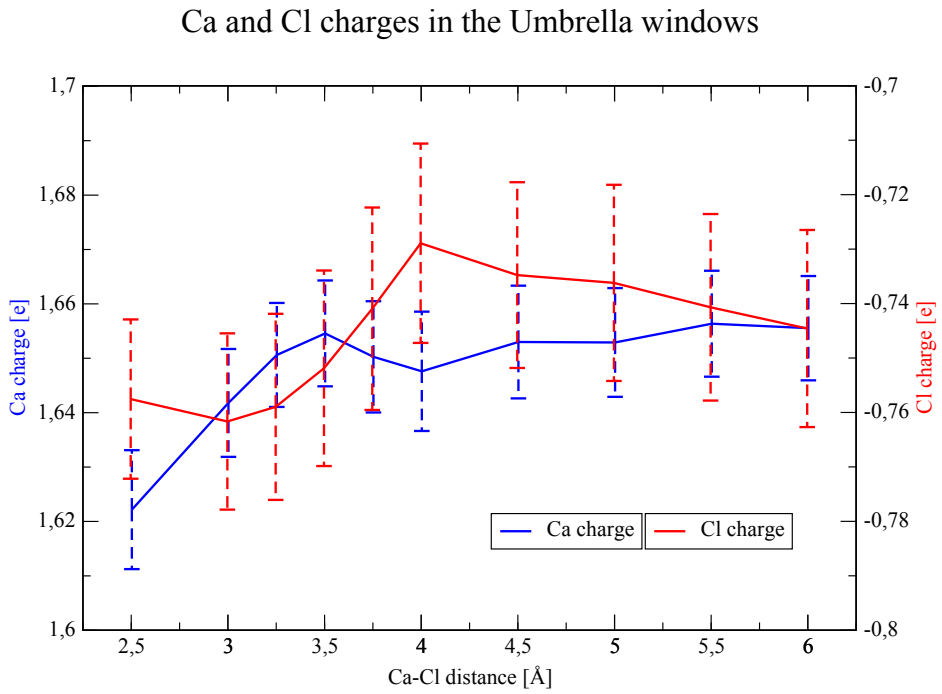


Figure 2.11: Bader charges of calcium dication (blue, left axis) and chloride anion (red, right axis) for all the Umbrella windows.

3. Guanidinium Pairing in Aqueous Solutions - Force Field MD Comparison

The existence of the previously reported guanidinium pair stacking (REF) in commonly-used force fields for water and ions was investigated by umbrella sampling simulations. The system contained two Gdm^+ molecular ions and two Cl^- counterions enclosed in a cubic box of 4x4x4 nm filled with 2171 water molecules. The periodic boundary conditions were employed with a cutoff of 1 nm. The simulations were set to run for 42 ns with 2 ns used as an equilibration. A 2 fs timestep was chosen.

The OPLS-AA (Jorgensen et al. [1996]) force field, parametrised for the liquid simulations, with TIP3P (Jorgensen et al. [1985]), TIP4P (Jorgensen and Madura [1985]) and SPC/E (Berendsen et al. [1983]) water models was employed in addition to the all-atom CHARMM22 (MacKerell et al. [1998]) force field aimed at protein simulations with TIP3P(Jorgensen et al. [1985]) and SPC/E (Berendsen et al. [1983]) water models. The force field parameters for guanidinium were derived from arginine amino acid parameters. The Electronic Continuum Correction (ECC) was not necessary in this case, since guanidinium is a monovalent ion with low charge density so the electronic polarization effect is relatively small.

Velocity rescaling with a stochastic term (v-rescale, Bussi et al. [2007]) generates a NVT canonical ensemble at the temperature of 298 K with a 0.5 ps time coupling parameter. Long-range electrostatic interactions were treated with the Particle-Mesh Ewald method (Essmann et al. [1995]). The initial velocities were generated according to Maxwell distribution at a temperature of 298 K. The LINCS (Hess et al. [1997]) method was used to constrain all the bonds.

For the guanidinium carbon-carbon distance from 3.0 Å to 15 Å, 25 umbrella windows were prepared with a step of 0.5 Å. The 10000 kJ/mol.nm^{-2} force constant was employed in the case of umbrella windows with the C-C distance smaller or equal to 7.5 Å except the CHARMM/TIP3P system, where the 5000 kJ/mol.nm^{-2} force constant was used for all umbrella potentials. For the larger C-C distances, the umbrella force constant of 5000 kJ/mol.nm^{-2} has proved to be sufficient. The Gromacs 4.6.1. (Hess et al. [2008]) program with the g_wham (Hub et al. [2010]) incorporation was used to perform these simulations and obtain free energy profiles.

3.0.1 Results and Discussion

The simulated guanidinium-guanidinium free energy profiles in an aqueous solution are shown in the Figure 3.1. After performing the entropy correction, the profiles were aligned vertically to zero at the C-C distance of 15 Å. The SPC/E water model profiles are coloured bright green, while cyan is used for the OPLS-AA and CHARMM force field respectively. The OPLS-AA and CHARMM profiles for TIP3P water model are shown in green for the former and in blue for the latter. The simulation using the TIP4P water model and OPLS-AA force

field is depicted in black.

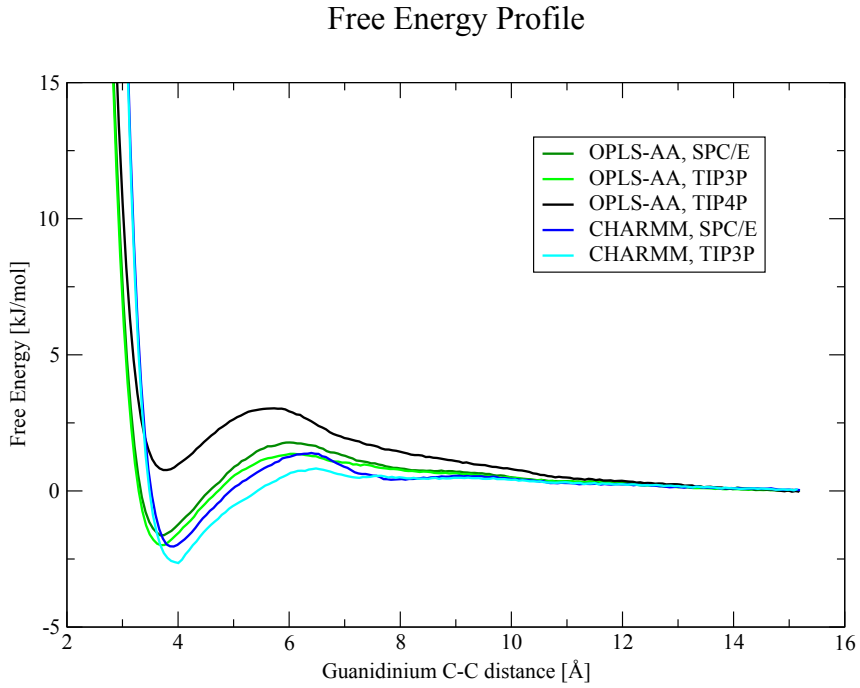


Figure 3.1: The free energy profiles of Gdm^+-Gdm^+ ion association.

The positions of contact ion pair as well as the positions of the transition state distances for dissociation are presented in Table 3.1. Table 3.2 shows the free energy differences between the contact ion pair and the solvent-separated ion pair, the transition state barrier height from the CIP and the TS height from the solvent-separated ion pair.

FF	water model	CIP [Å]	TS [Å]
OPLS-AA	SPC/E	3.7	6.0
OPLS-AA	TIP3P	3.7	6.1
OPLS-AA	TIP4P	3.7	5.7
CHARMM	SPC/E	3.9	6.4
CHARMM	TIP3P	4.0	6.5

Table 3.1: The positions of guanidiniums' dissociation free energy profile CIP and TS for the five investigated setups.

The computational evidence for the existence of a local free energy minimum representing the Gdm^+-Gdm^+ contact ion pair in aqueous solutions is provided by all the investigated force fields. Furthermore, the simulations using the SPC/E and TIP3P water models for both the OPLS-AA and CHARMM force field parametrizations indicate an absolute preference for the contact ion pair. Only the OPLS-AA force field with the TIP4P water model simulation resulted in a Gdm^+-Gdm^+ CIP energy minimum slightly above the dissociation limit.

These simulations were a part of a broader project Allolio et al. [2016]. Our published article is presented in the Attachments section.

FF	water model	$\Delta E_{CIP,SSIP}$	$\Delta E_{TS,CIP}$	$\Delta E_{TS,SSIP}$
OPLS-AA	SPC/E	−1.6	3.4	1.8
OPLS-AA	TIP3P	−2.0	3.4	1.4
OPLS-AA	TIP4P	0.8	2.3	3.0
CHARMM	SPC/E	−2.0	3.4	1.4
CHARMM	TIP3P	−2.7	3.5	0.8

Table 3.2: The free energy CIP well depth (3rd column), the energy barrier between CIP and TS (4th column) and the transition state height from the solvent separated system conformation (5th column). All the energy values are presented in kJ/mol.

Convergence with Simulation Length

Additionally, the convergence of the umbrella sampling simulations was investigated for the OPLS-AA force field with the SPC/E water model system. Figure 3.2 shows the free energy profiles comparison when 40 ns, 20 ns, 10 ns, 5 ns and 2 ns data sets per window were used. The profiles were aligned so that the energy value at 1.5 nm is zero. Clearly, using 10 ns per window or less does not result in a converged free energy profile. The simulation length of 20 ns is practically converged but is not optimal since its free energy profile is not as smooth as the 40 ns free energy profile.

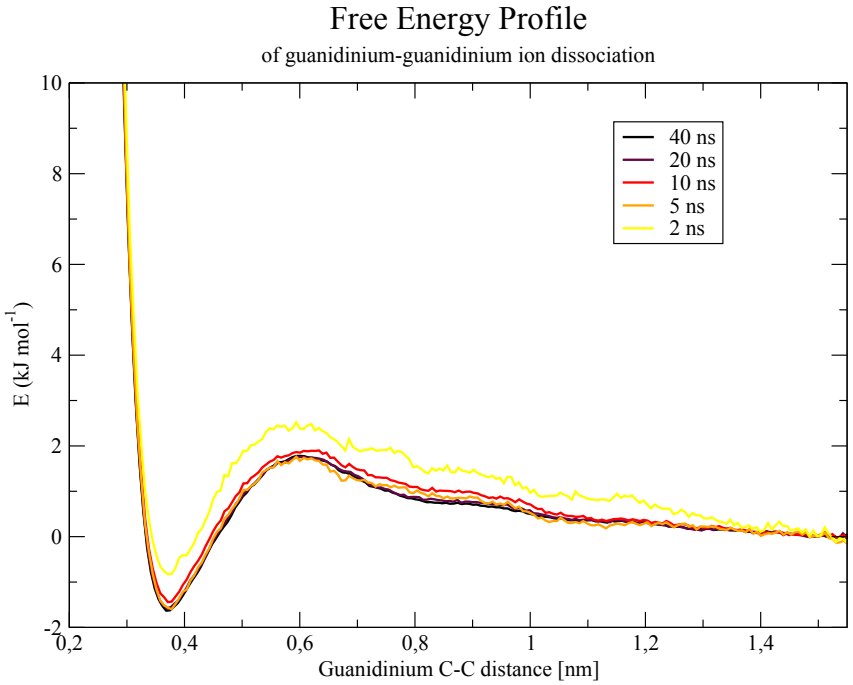


Figure 3.2: Guanidinium-guanidinium free energy profiles computed using OPLS-AA force field and SPC/E water model resulting from umbrella sampling simulations using five different data set lenghts corresponding to simulation times of 40 ns, 20 ns, 10 ns, 5 ns and 2 ns.

4. Fluorescence Spectroscopy Measurements

Cell membrane is an extremely complex system with various functions such as cell protection, regulation of ions and molecule uptake and excretion and the separation of the cell interior from the outer space. It contains many different types of proteins and lipids ordered in a bilayer structure with polar parts facing the outer aqueous environment and the non-polar parts placed inside the bilayer. At physiological conditions, the membrane-forming lipids and proteins are constantly moving in a so-called (membrane) liquid phase, determined mostly by unsaturated hydrocarbon chains of glycerophospholipids. The cell membrane is often covered by the glycocalyx, composed of glycoproteins and glycolipids, that for example plays a role in the cell-cell recognition and communication processes.

One of the simplest model systems of the cell membranes are lipid vesicles. They are ball-shaped lipid bilayer structures of varying radii. Those with a diameter of less than 100 nm are called SUVs - small unilamellar vesicles, LUVs a.k.a. large unilamellar vesicles are less than 1 micrometer in diameter and giant unilamellar vesicles (GUVs) are larger than 1 micrometer in diameter.

Within a broader investigation of cell penetrating peptides, our aim was to study the bilayer penetration by nona-arginine ($R9$), tetra-arginine ($R4$) and nona-lysine ($K9$) peptide molecules or a calcium-containing buffer. Fluorescence spectroscopy was employed to investigate the behaviour of LUVs containing the fluorescent dye calcein in an aqueous solution when introducing the peptides or calcium to the system.

4.1 Materials Used in the Fluorescence Experiments

4.1.1 Lipids

Together with proteins, nucleic acids and sacharides, lipids belong to one of the main classes of biomolecules. These are amphiphilic molecules, which means that they consist of a polar part (often called a head group) and a non-polar tail. In this study we are interested in glycerophospholipids and sterols that are the main lipid components of biological membranes.

Glycerophospholipids contain two saturated (or one saturated and one unsaturated) fatty acids attached to glycerol as esters creating the non-polar tail. The polar head group is represented by the phosphate and an additional group, which can be smaller or bigger, neutral or charged. These properties of lipid head groups as well as the fatty acid length and saturation contribute to the properties a given lipid promotes in a membrane. For example it can induce positive or negative curvature, modulate the membrane tension or contribute to the stabilization of a flat bilayer. These propertie have a further impact on the promotion or inhibition of the vesicle aggregation and fusion.

In this work we focused on glycerophospholipids with palmitoyl-oleoyl (PO) or dioleoyl (DO) tails and phosphatidylcholine (PC), phosphatidylserine (PS) or

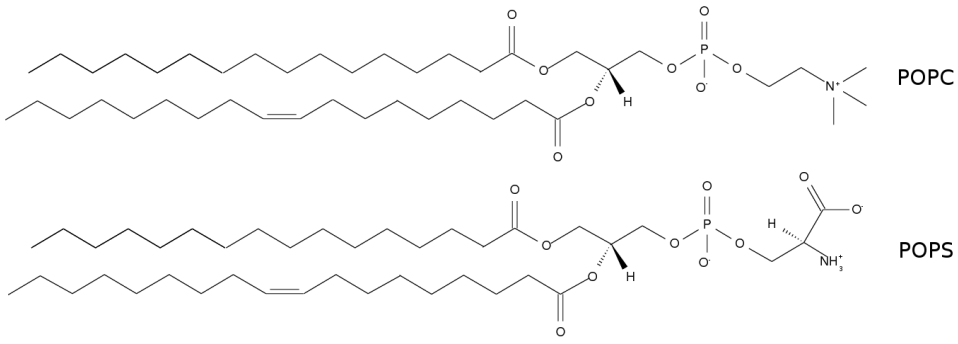


Figure 4.1: The phosphatidyloleoyl phospholipids used in this work.

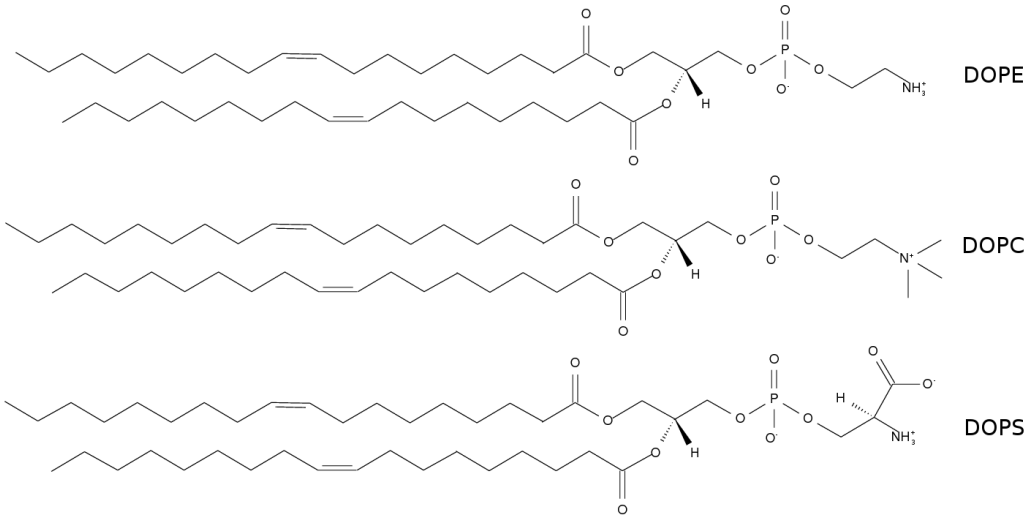


Figure 4.2: The dioleoyl phospholipids used in our experiments.

phosphatidyletanolamine (PE) head groups, as is illustrated in the Figure 4.1 and Figure 4.2.

Sterols are composed of four steroid system's rings and an alcohol group. They themselves do not form bilayer phases, however their combination with bilayer-forming lipids can result in a liquid-ordered phase. We are particularly interested in cholesterol, Figure 4.3, which is the dominant sterol of the mammalian cell membranes.

4.1.2 Fluorescent Dyes

Leakage-induced calcein fluorescence can be measured by trapping the quenched fluorescent dye calcein inside the lipid vesicles followed by measuring the leaking dye-caused fluorescence increase. Another fluorescent dye is introduced to the investigated system for the sake of the lipid amount calibration. In our case it is the DiD dye, incorporated into vesicle membranes.

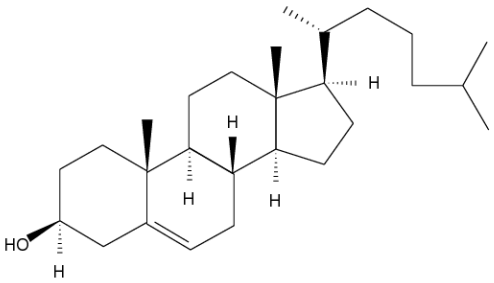


Figure 4.3: Cholesterol.

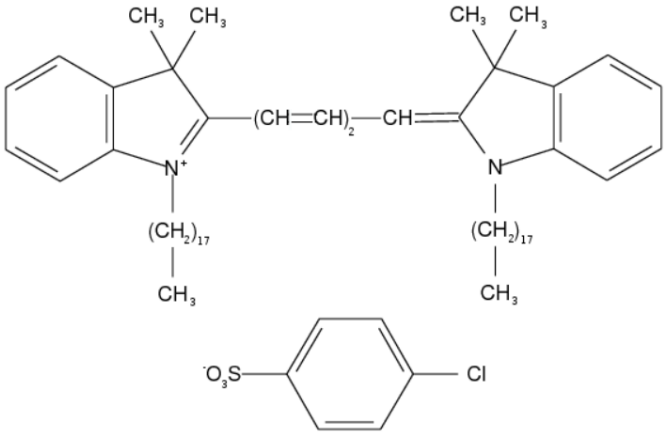


Figure 4.4: Chemical formula of DiD.

DiD

The lipophylic fluorescent dye DiD $C_{67}H_{103}ClN_2O_3S$ (1,1'-dioctadecyl-3,3,3',3'-tetramethylindodicarbocyanine,4-chlorobenzensulfonate salt), also called the $DiIC_{18}(5)$, see Figure 4.4, is an analogue to the DiI fluorescent dye with a red-shifted excitation (maximum at 644 nm) and emission spectra (maximum 665 nm). It is highly fluorescent and reasonably photostable when incorporated into the lipid membrane, while being weakly fluorescent in water.

Calcein

Calcein (see Figure 4.5), or fluorexon $C_{30}H_{26}N_2O_{13}$ is a water-soluble membrane-impermeable fluorescent dye that exhibits self-quenching while encapsulated in the lipid vesicles. When released to the environment with lower calcein concentration, a fluorescence intensity increase can be measured. It is usually used as an indicator of the lipid vesicle leakage. The employed excitation/emission wavelenghts of calcein are 495/515 nm.

4.1.3 Buffers

Normal Buffer

The here-called normal buffer consists of 10 mM HEPES, 100 mM KCl, 100 μM EDTA and 140 mM glucose. The pH was adjusted by adding KOH to $pH = 7.40$

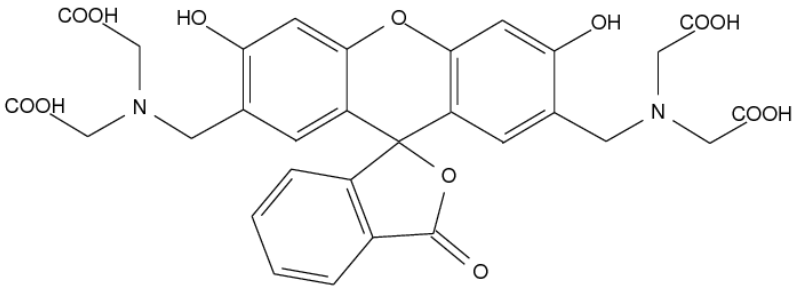


Figure 4.5: Chemical formula of calcein.

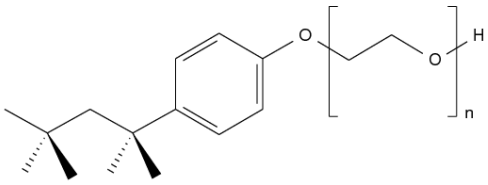


Figure 4.6: Triton - a detergent used to burst lipid vesicles.

and the osmolarity was set to be 318 mOsm/l.

Calcein Buffer

Calcein buffer contains 30 mM calcein, 10 mM HEPES, 100 mM KCl and 100 μM EDTA. Its osmolarity is 318 mOsm/l and $pH = 7.40$.

Calcium Buffer

This buffer includes calcium dication in a form of 40 mM $CaCl_2$ alongside with 100 mM KCl at the pH of 7.40 and glucose for its osmolarity to be 318 mOsm/l.

4.1.4 Detergents

Triton X

As a detergent, the 4-(1,1,3,3-tetramethylbutyl)phenyl-polyethylene glycol, or the so-called Triton X-100 was used, see 4.6. The average number n of the ethylene oxide units is 9.5 per molecule. Triton X-100 consists of hydrophilic polyethylene oxide part and a hydrophobic group containing aromatic ring. It is a non-ionic surfactant known to lyse cells (Schnaitman [1971]). 0.1 % Triton X-100 solution in water is sufficient to solubilise lipid vesicles (Inc).

4.2 Measured Systems

Table 4.1 shows five chosen lipid compositions. All the used lipid types can be found in the mammalian cell membrane. The first sample (1) consists of the widely studied POPC lipid molecules, the second lipid composition combines the POPC lipid molecules with POPS, that have a slightly bigger lipid head.

The di-oleoyl lipids, having both their fatty acid chains unsaturated, increase the membrane fluidity in comparison to the palmitoyl-oleoyl lipids. Sample (3) consists of di-oleoyl lipids with phosphatidyletanolamine and phosphatidylserine head groups. The system (4) is a combination of the PE, PS and PC lipid head groups bonded to the di-oleoyl lipid tails in the ratio of 3:1:1. Addidtionally, the sample ($5x$) was created to study the role of the cholesterol - simply by taking the (4) composition and on the top of it adding 0.5 μmol of cholesterol, corresponding to the amount of 10% of the original 5 μmol (hence the label "x").

System number	Lipid composition	Lipid percentage [%]
(1)	POPC	100
(2)	POPC/POPS	80/20
(3)	DOPE/DOPS	80/20
(4)	DOPE/DOPS/DOPC	60/20/20
($5x$)	(4)+10% CHOL	$60/20/20 + 10$

Table 4.1: The lipid compositions of five investigated systems.

4.2.1 Preparation Protocol and Measurement Procedure

Preparing the Lipid Vesicles

The lipids forming 5 μmol compositions (1) to ($5x$) (Avanti Polar Lipids, Inc. Alabaster, USA) originally dissolved in chloroform were pipetted along with the DiD dye by Hamilton syringes and dried under N_2 flux. The lipid-DiD films were further put under vacuum for at least 20 minutes. Consequently, 500 μl of the calcein-containing buffer was added, the solutions were shaken and left to hydrate in the fridge until the next day. After more shaking and in a few cases sonicating (mostly compositions containing DOPE) to dissolve the lipid solutions fully, these were extruded using an Avestin porous membrane (diameter 400 nm; Avestin, Ottawa, Canada) passing it 50 times, which formed the lipid vesicles.

To extract only the large unilamellar vesicles (LUVs), that are the lipid vesicles with a diameter range of $100 - 200$ nm, we used a chromatography column (Sephadex G25 $4x5$ ml HiTrap desalting columns, GE Healthcare, Uppsala, Sweeden) with a normal buffer as a mobile phase at a flow rate of 2 ml/min. After a lipid vesicle solution injection to the column and draining 5.5 μl, subsequently the amount of 1.2 μl of the sample was collected. The final lipid concentration was calibrated to be around 1 mM by comparing the DiD fluorescence to a calcein-free POPC standard at 1 mM concentration.

Measuring Procedure

A calibrated amount of a lipid sample was pipetted to the cuvette with a corresponding volume of the normal buffer to be of 1.0 ml or 1.5 ml in the case when 450 μl of calcium buffer was to be added. The peptide/lipid ratios were set to 1, the calcium concentration in the sample is 12 mM.

The sample was put into the Fluorolog-3 spectrofluorimeter (model FL311; JobinYvon Inc., Edison, NJ, USA) equipped with a xenon lamp. With the excitation of calcein at 495 nm, the fluorescence intensity at 515 nm was measured for

200 s. An arginine or lysine peptide in solution or a calcium-containing buffer was added. At the end of the measurement, when the fluorescence intensity reached a plato, 50 μl of Triton X-100 was added to burst all the vesicles. Thus all the calcein leaked to the outer solution causing the fluorescence reaching maximum.

Data Processing

To quantify the relative calcein leakage of the vesicle samples p in various conditions, we calculate first the average value of the counts at the beginning of the measurement, before adding anything else, i.e. at time $0 - 200$ s (denoted here as $f(t_0)$). The fluorescence at the beginning of the measurement comes from two sources. There is a small amount of calcein outside the vesicles, coming from the pipetted lipid solution, strongly diluted during the column procedure. Another part of the signal is caused by the calcein entrapped in the vesicles, because its self-quenching is not complete.

Next, we calculate an average fluorescent intensity ($f(t)$) at the time when the system stabilises after the addition of peptides or the calcium buffer. Finally, after adding Triton X-100 when all the vesicles burst and leak calcein, we obtain $f(t_\infty)$. The relative calcein leakage fluorescence from the vesicles is then evaluated as:

$$p = \frac{f(t) - f(t_0)}{f(t_\infty) - f(t_0)} \tag{4.1}$$

4.3 Results and Discussion

$R9$ and $K9$ Membrane Penetration Measurements

Figure 4.7 shows the relative vesicle leakage after adding $R9$. The added peptide amounts were pipetted to reach the final peptide concentration in the solution - 75 mM. The POPC lipid sample (1) showed no leakage, indicating that the nona-arginine was not able to get inside the vesicles. In the case of the sample (2), we can even see a so-called "negative leakage", when the fluorescent signal decreases after adding the peptide. This drop occurred in the measurements, despite stirring the sample during all the experiment. The Dynamic Light Scattering (DLS) measurements (not shown here) demonstrated vesicle aggregation in the case of adding $R9$ as well as $K9$ to lipid vesicle solutions. The $R4$-caused vesicle aggregation was orders of magnitude lower. The fluorescent intensity decrease is therefore related to the vesicle sinking without the calcein leakage. The lipid vesicles probably aggregated, but not leaked in the case of POPC/POPS peptide composition (2). In contrast, the nona-arginine penetrates the DOPE-containing sample membranes - the Figure 4.7 shows a fluorescent increase in all three cases. The lipid vesicle sedimentation does not play a significant role when leakage occurs since the leaked calcein movement is free in the solution.

The averages of measured $f(t_0)$, $f(t)$ and $f(t_\infty)$ along with the calculated relative calcein leakage are presented in Table 4.2. The error of the relative leakage p was estimated from the average values and their average deviations as an indirect measurement error. If N is the fluorescence intensity count, then the

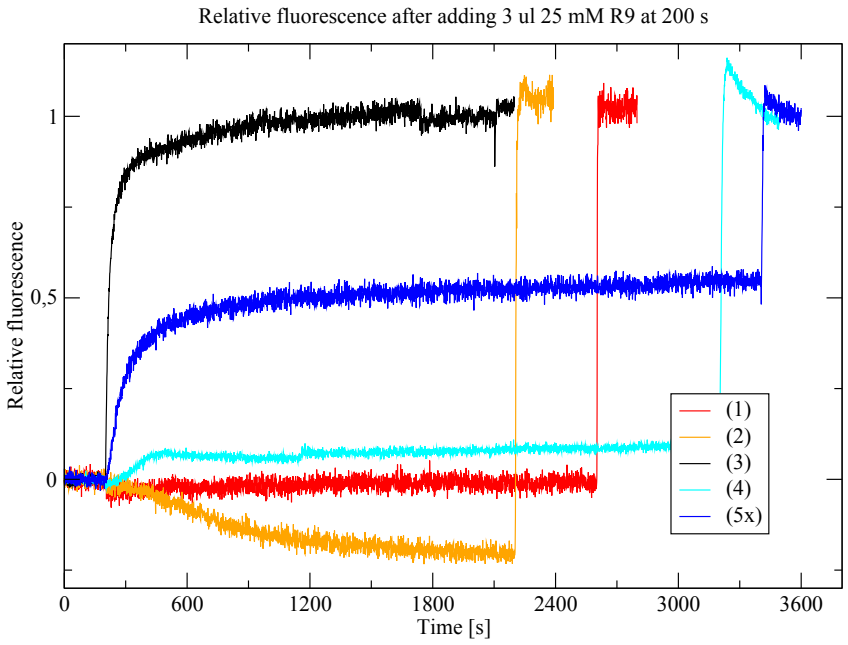


Figure 4.7: The relative fluorescent intensities during the spectroscopy measurement. The $R9$ peptide was added to the solution at 200 s.

error can be estimated as $\sqrt{N}$. For all the measurements presented, the $\sqrt{N}$-error is at least an order smaller than the calculated average deviations.

The vesicles of lipid compositions (1) and (2) do not let the nona-arginine penetrate their membrane, and therefore no calcein leakage occurs. The more fluid di-oleoyl membranes are more or less $R9$-penetrable. Lipid composition (4) leaks just a little - around 8 %. Adding a certain amount of cholesterol to the vesicle membranes causes a rapid fluorescence intensity increase - up to 54 % for lipid composition ($5x$). The highest rate of nona-arginine lipid membrane penetration occurred in the case of the DOPE-rich sample number (3). Basically all the vesicles leaked.

Lipid Composition	$f(t_\infty).10^5$	$f(t_0).10^4$	$f(t).10^4$	Leakage [%]
(1) POPC	40 ± 3	157 ± 4	153 ± 4	-2 ± 3
(2) POPC/POPS	41 ± 6	142 ± 3	89 ± 4	-20 ± -2
(3) DOPE/DOPS	80 ± 2	259 ± 5	801 ± 10	100 ± 7
(4) DOPE/DOPS/DOPC	180 ± 18	630 ± 7	848 ± 727	8 ± 9
($5x$) (4)+10% CHOL	48 ± 3	156 ± 3	331 ± 6	54 ± 9

Table 4.2: Average fluorescent intensities during the $R9$ experiment and calculated relative calcein leakage.

The experiments were repeated with the nona-lysine peptide. A decrease of calcein leakage indicated by smaller fluorescent intensities at the plateu can be seen in the Figure 4.8. Table 4.3 contains the mean intensity values as well as the calculated relative leakage p for the investigated systems.

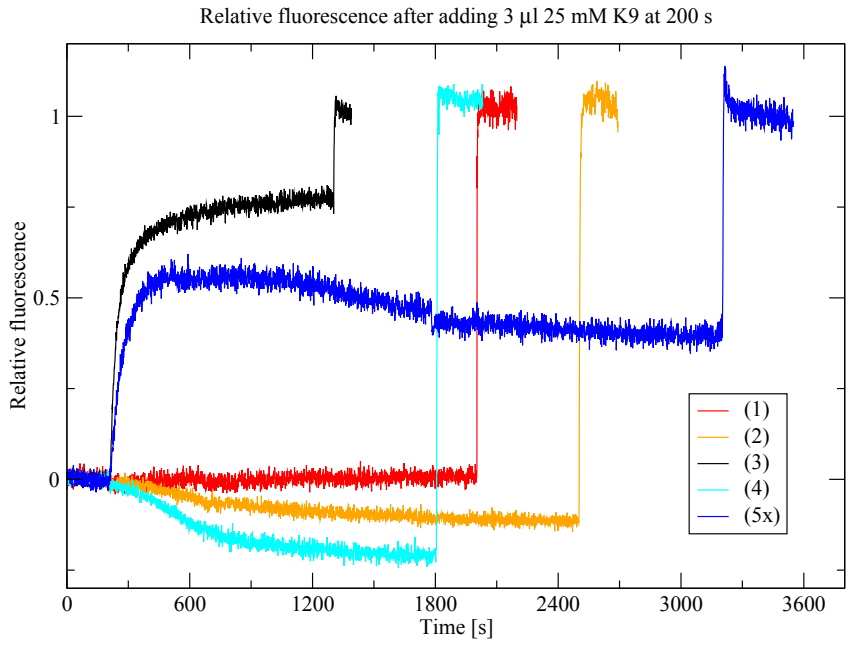


Figure 4.8: The relative fluorescent intensities with nona-lysine ($K9$) added to the lipid vesicle solution at 200 s.

In the case of the POPC sample (1), again there is no leakage or any indication of aggregation. Vesicles with the lipid composition (2) aggregate less than after adding $R9$. The fluorescent intensity decrease of the sample (4) indicates its vast aggregation, but most probably negligible leakage. The calcein leakage occurred only in the case of samples (3) and ($5x$), however being lower than in the case of the $R9$ penetration.

Lipid composition	$f(t_\infty).10^5$	$f(t_0).10^4$	$f(t).10^4$	Leakage [%]
(1) POPC	45 ± 4	159 ± 4	159 ± 4	0 ± 3
(2) POPC/POPS	49 ± 6	138 ± 4	103 ± 6	-10 ± 1
(3) DOPE/DOPS	78 ± 3	282 ± 5	660 ± 8	76 ± 8
(4) DOPE/DOPS/DOPC	98 ± 14	353 ± 6	227 ± 9	-20 ± -2
($5x$) (4) +10% CHOL	61 ± 3	249 ± 5	416 ± 23	47 ± 12

Table 4.3: Average fluorescence intensity values $f(t_\infty)$, $f(t_0)$ and $f(t)$ (column 2, 3 and 4) to compute the relative calcein leakage p (column 5).

Injecting $R4$ to the Vesicle Solutions - Fluorescence Measurements

The tetra-arginine lipid membrane penetration was investigated using double the amount of $R4$ so that the number of arginines in the lipid vesicle solution is closer to the $R9$ case for the sake of a qualitative comparison. No leakage in Figure 4.9 (and Table 4.4) indicates an absence of the $R4$ penetration through the lipid vesicle membranes. This finding is in agreement with Mitchell et al. [2000] or

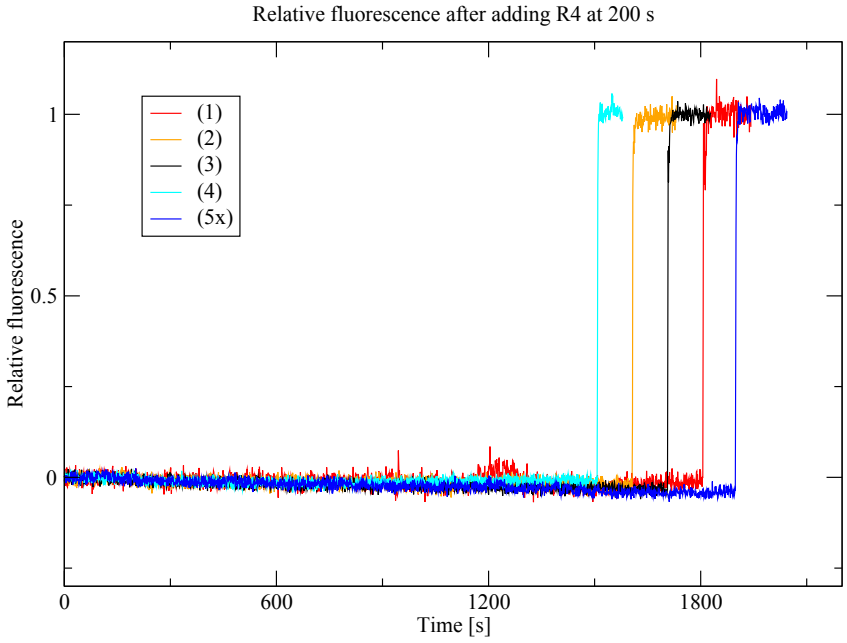


Figure 4.9: Calcein fluorescent leakage after adding tetra-arginine ($R4$) to the solution with lipid vesicle compositions (1)-(5x).

Futaki et al. [2001].

Lipid composition	$f(t_\infty).10^5$	$f(t_0).10^4$	$f(t).10^4$	Leakage [%]
(1) POPC	49 ± 7	196 ± 4	192 ± 6	-1 ± 3
(2) POPC/POPS	54 ± 9	196 ± 4	190 ± 4	-2 ± 2
(3) DOPE/DOPS	52 ± 29	353 ± 7	340 ± 7	-8 ± -6
(4) DOPE/DOPS/DOPC	111 ± 24	430 ± 6	421 ± 7	-1 ± 1
$(5x)$ (4) +10% CHOL	102 ± 2	349 ± 6	331 ± 11	-3 ± 2

Table 4.4: Average fluorescent intensities for the $R4$ calcein leakage experiment and calculated relative leakage.

Fluorescence Measurements with Calcium-containing Buffer Added to the Lipid Vesicle Systems

The calcium concentration in the 1.5 ml lipid vesicle sample was chosen to be 12 mM. Thus the 450 μl of the 40 mM calcium buffer was added to the lipid vesicle samples (1)-($5x$). The measured data with calculated relative calcein leakage values are shown in the Figure 4.10 and the Table 4.5. After adding the calcium-containing buffer to the lipid vesicle solution, the lipid vesicle concentration (as well as that of calcein) changes significantly. Therefore the fluorescence values right after adding the calcium buffer are considered to be $f(t_0)$, rather than the average of the fluorescent intensity during the initial 200 s. The measured data are also shifted accordingly in the Figure 4.10. An additional error is therefore

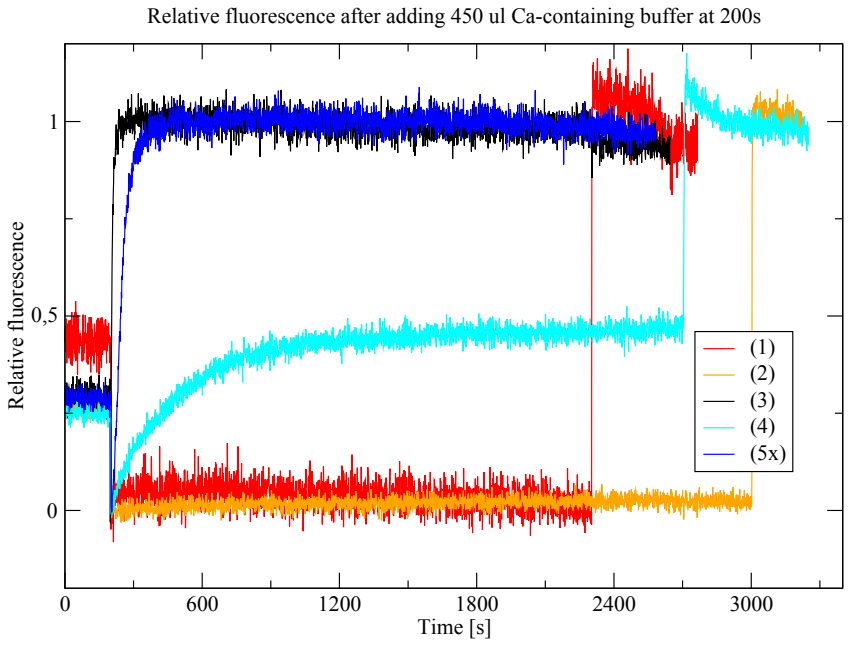


Figure 4.10: Measured fluorescent spectroscopy results for 12 mM calcium added to the LUV lipid vesicle solutions at 200 s.

introduced to the relative leakage values p, especially in the case of lipid composition (3) and ($5x$), where the fluorescent increase is so fast that the (minimum) $f(t_0)$ value taken might have already been influenced by it.

In the case of palmitoyl-oleoyl (PO) lipid compositions (1) and (2) no calcein leakage was detected. On the contrary lipid composition (4) leaked around 46 %, clearly more than after adding the $R9$ peptide. The samples (3) and ($5x$) leaked basically all the calcein. As can be seen in the Figure 4.10, the fluorescence increased much faster for lipid composition (3).

Lipid Composition	$f(t_\infty).10^5$	$f(t_0).10^4$	$f(t).10^4$	Leakage [%]
(1) POPC	19 ± 1	95	99 ± 3	4 ± 4
(2) POPC/POPS	56 ± 4	225	231 ± 6	2 ± 2
(3) DOPE/DOPS	34 ± 1	155	345 ± 5	105 ± 6
(4) DOPE/DOPS/DOPC	43 ± 2	170	292 ± 5	46 ± 5
($5x$) (4) +10% CHOL	50 ± 1	210	510 ± 7	103 ± 5

Table 4.5: Average fluorescence intensity values $f(t_\infty)$, $f(t_0)$ and $f(t)$ the relative leakage in the case of adding calcium-containing buffer to the lipid vesicle solutions.

4.3.1 Discussion

As a part of a bigger project, the penetration of oligoarginines and lysines into lipid vesicles was studied by the fluorescence spectroscopy. The fluorescence inten-

sity increase was observed only when the self-quenching fluorescent dye, calcein, originally trapped inside the vesicles, got out of them. This fluorescence increase is considered to be the proof of the lipid membrane penetration by the peptide.

The nona-arginine, reported to enter the cells via a passive mechanism (Futaki et al. [2001], Mitchell et al. [2000]), was shown to penetrate the di-oleoyl lipid vesicles. Note that vesicles with more stiff planar membranes composed of palmitoyl-oleoyl lipids do not show the fluorescence increase. Therefore, no $R9$ penetration is thought to occur. In PE-rich membranes nona-lysine peptides were shown to leak calcein less than the $R9$ peptide. This result is in agreement with the findings of Mitchell et al. [2000] only qualitatively, showing $K9$ entering the vesicles in a non-negligible amount compared to $R9$. Lipid vesicles are just a model system of the cell membranes, where $K9$ peptides might, for example, be withheld near the membrane, not penetrating it. In the case of all the investigated lipid vesicle samples, none was shown to be penetrated by the $R4$ peptide molecules, in agreement with cell experiments.

The DLS experiments showed that the membrane penetration process is linked to the vesicle aggregation (and fusion). The decrease of the fluorescence intensity shown in the Figure 4.7 and Figure 4.8 suggests vesicle aggregation, but most probably not peptide penetration. Since calcium is known to enhance membranes fusion (Papahadjopoulos et al. [1976]), another set of experiments with a calcium-containing solution was performed. A calcein leakage was observed, therefore, the process of vesicle aggregation and fusion and vesicle membrane-penetration is clearly interconnected. Further experiments within this bigger project led us to formulate a hypothesis about a novel passive mechanism of membrane penetration (the proposed article is in the Attachments).

Conclusion

The principle advantage of molecular dynamic simulations is that they can provide insights into the investigated processes at an atomistic scale. Reliable sets of parameters for the biomolecules, as well as of the surrounding water and ions, are necessary to simulate biological systems accurately. Nowadays, most force fields used for modelling of biological systems do not account for the electronic polarizability - they are the so-called non-polarisable force fields. In the case of simulating interactions between divalent ions and proteins or lipid membranes in aqueous solutions, not accounting for the electronic polarizability within a force field-based simulation can introduce significant artifacts such as for example excessive ion pairing or errors in viscosity or self-diffusion coefficients (Kohagen et al. [2014]). To correct for these inaccuracies, electronic polarisation effects can be introduced in a mean filed way by a simple but physically rigorous rescaling of ionic charges - the so-called electronic continuum correction.

In the case of the calcium dication with the chloride anion in aqueous solutions, charges were first reduced in the spirit of the electronic continuum correction (Leontyev and Stuchebrukhov [2011]). Comparing the resulting Ca^{2+}-Cl^- RDF to the neutron scattering data (Badyal et al. [2004]) led us to additionally slightly adjust the Lennard-Jones radii of Ca^{2+} and Cl^- (Kohagen et al. [2014]). Refining also the water parameters might have led to an even better agreement with the experiment. However, widely used water models, developed to reproduce the experimental properties of water, performed well and were, therefore, left untouched, with only the ionic parameters being adjusted.

In this work, the Ca^{2+}-Cl^- free energy profiles were calculated using force field-based molecular dynamics with three force field parameter sets (full, ECC and ECCR) as well as using ab-initio MD simulations. To construct the free energy profiles, the umbrella sampling method was employed. These calculations were a part of a broader project (Martinek et al. [2018]), where, among other things, we compared the simulation results of aqueous $CaCl_2$ solutions to the neutron scattering experiments. This comparison led to a final minor adjustment of Ca^{2+} and Cl^- force field parameters as a compromise to fit both the ab-initio and neutron scattering results.

Guanidinium molecular ions exhibit a rather unconventional behaviour in aqueous solutions - the experiments and simulations (Shih et al. [2013] ,Flocco and Mowbray [1994], Kubíčková et al. [2011]) show that despite bearing the same charge they form stable contact ion pairs. As a part of the larger project (Allolio et al. [2016]), the guanidinium-guanidinium dissociation free energy profiles were obtained using umbrella sampling force field-based simulations. The OPLS-AA force field with common water models (TIP3P, TIP4P, and SPCE) were employed along with the CHARMM protein force field (with TIP3P and SPC/E water models). All these simulation sets confirm the existence of the guanidinium-guanidinium contact ion pair and all-but-one (namely the OPLSAA/TIP4P system setup) indicate its global preference, as was also confirmed by the ab-initio simulation presented in the attached article. Note that we did not employ charge scaling here, since its effect is small for large monovalent ions like guanidinium, as was also confirmed by test ECC simulations.

In a human body, the guanidinium ion can be found as a part of the arginine amino acid. The aforementioned guanidinium stacking might play a role in membrane-penetration properties of oligoarginines. Arginine-rich peptides were found to be able to penetrate the cell membrane not only at the physiological temperature but also at 4° C where active ATP-driven processes shut off (Futaki et al. [2001], Vivès et al. [1997]), suggesting a passive uptake in addition to the active endocytosis. Despite an extensive attention of the scientific community, the molecular mechanism of the passive uptake process is still not fully resolved.

The final part of this work thus focuses on the study of interactions between the peptides composed of basic amino acids (arginine and lysine) or the calcium-containing solution with the lipid vesicles as model systems of a cell membrane. The reported (Wender et al. [2000], Futaki et al. [2001]) $R9$ penetration through a lipid membrane was confirmed by our fluorescence spectroscopy measurements together with the reduced $K9$ lipid membrane penetration and no penetration for $R4$.

The fluorescence measurements reported here were a part of an extensive project on cell penetrating peptides (the resulting article is in the Attachments). While performing fluorescence microscopy on giant unilamellar vesicles, we observed an aggregation and fusion of the GUVs along with the $R9$ lipid membrane penetration. To confirm that these processes are linked with each other, a DLS measurement was further performed. Since calcium is known to enhance membrane fusion, we performed additional fluorescence measurements on LUVs. These experiments showed that after adding the calcium-containing solution to the sample, the leakage of the LUV-entrapped calcein occurs similarly as in the case of $R9$. Further, a FRET analysis proved for LUVs the formation of multilamellar structures after adding the $R9$ peptide. These results along with simulations of curved membranes with $R9$, $K9$ and $R4$, led us to formulateion of a hypothesis concerning a direct link between the arginine membrane penetration and cell uptake processes. Finally, our fluorescence microscopy and the electron microscopy experiments on HeLa cells support the proposed hypothesis.

Bibliography

C. Allolio, K. Baxová, M. Vazdar, and P. Jungwirth. Guanidinium pairing facilitates membrane translocation. *The Journal of Physical Chemistry B*, 120: 143–153, 2016.

A. Arnaldsson, W. Tang, S. Chill, W. Chai, and G. Henkelman. Bader charge analysis. `http://theory.cm.utexas.edu/henkelman/code/bader/`, 2012.

R. F. W. Bader. *Atoms in Molecules: A Quantum Theory*. International Ser. of Monogr. on Chem. Clarendon Press, 1994. ISBN 9780198558651. URL `https://books.google.cz/books?id=tyVpQgAACAAJ`.

Y. S. Badyal, A. C. Barnes, G. J. Cuello, and J. M. Simonson. Understanding the effects ofconcentration on the solvation structure of ca^{2+} in aqueous solution. II: Insights into longer range order from neutron diffraction isotope substitution. *Journal of Physical Chemistry A*, 108:11819–11827, 2004.

M. D. Baer and C. J. Mundy. Local aqueous solvation structure around ca^{2+} during $ca^{2+}\cdots cl^{-}$ pair formation. *The Journal of Physical Chemistry B*, 120: 1885–1893, 2016.

A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A*, 38:3098–3100, 1988.

H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 79:6269–6271, 1983.

S. Boudon, G. Wipff, and B. Maigret. Monte carlo simulations on the like-charge guanidinium-guanidinium ion pair in water. *The Journal of Physical Chemistry*, 94:6056–6061, 1990.

G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126:014101–1–014101–7, 2007.

L. X. Dang and D. E. Smith. Comment on "mean force potential for the calcium–chloride ion pair in water" [j. chem. phys. 99, 4229 (1993)]. *The Journal of Chemical Physics*, 102:3483–3484, 1995.

U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. A smooth particle mesh ewald method. *The Journal of Chemical Physics*, 103(19):8577–8593, 1995.

M. M. Flocco and S. L. Mowbray. Planar stacking interactions of arginine and aromatic side-chains in proteins. *Journal of Molecular Biology*, 235:709–717, 1994.

A. D. Frankel and C. O. Pabo. Cellular uptake of the tat protein from human immunodeficiency virus. *Cell*, 55:1189–1193, 1988.

S. Futaki, T. Suzuki, W. Ohashi, T. Yagami, S. Tanaka, K. Ueda, and Y. Sugiura. Arginine-rich peptides. *The Journal of Biological Chemistry*, 276:5836–5840, 2001.

S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Physical Review B*, 54:1703–1710, 1996.

M. Green and P. M. Loewenstein. Autonomous functional domains of chemically synthetized human immunodeficiency virus tat trans-activator protein. *Cell*, 55:1179–1188, 1988.

S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27:1787–1799, 2006.

A. Grossfield. WHAM: the weighted histogram analysis method. `http://membrane.urmc.rochester.edu/content/wham`, 2013.

B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18:1463–1472, 1997.

B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

J. S. Hub, B. L. de Groot, and D. van der Spoel. g_wham - a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of Chemical Theory and Computation*, 6(12):3713–3720, 2010.

T. Inagaki, S. Aono, H. Nakano, and T. Yamamoto. Like-charge attraction of molecular cations in water: Subtle balance between interionic interactions and ionic solvation effect. *The Journal of Physical Chemistry B*, 118:5499–5508, 2014.

Sigma-Aldrich Inc. Triton x100. `https://www.sigmaaldrich.com/catalog/product/sial/x100?lang=en&region=CZ&gclid=EAIaIQobChMI0r-EkMnV3AIVFeAZCh2ijAI0EAAYASAAEgJEsfD_BwE`. Accessed: 2018-08-05.

W. L. Jorgensen and J. D. Madura. Temperature and size dependence for monte carlo simulations of TIP4P water. *Molecular Physics: An International Journal at the Interface Between Chemistry and Physics*, 56:1381–1392, 1985.

W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 56:926–935, 1985.

W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.

M. Kohagen, P. E. Mason, and P. Jungwirth. Accurate description of calcium solvation in concentrated aqueous solutions. *The Journal of Physical Chemistry B*, 118:7902–7909, 2014.

J. Kästner. Umbrella sampling. *WIREs Computational Molecular Science*, 1: 932–942, 2011.

A. Kubíčková, T. Křížek, P. Coufal, E. Wernersson, J. Heyda, and P. Jungwirth. Guanidinium cations pair with positively charged arginine side chains in water. *The Journal of Physical Chemistry Letters*, 2:1387–1389, 2011.

A. Laio and M. Parrinello. Escaping free-energy minima. *PNAS*, 99:12562–12566, 2002.

C. Lee, W. Yang, and R. G. Parr. Development of the colle-salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37: 785–789, 1988.

I. Leontyev and A. Stuchebrukhov. Accounting for electronic polarization in non-polarizable force fields. *Phys. Chem. Chem. Phys.*, 13:2613–2626, 2011.

Jr. A. D. MacKerell, D. Bashford, M. Bellott, Jr. R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102:3586–3616, 1998.

A. Magarkar, P. Jurkiewicz, C. Allolio, M. Hof, and P. Jungwirth. Increased binding of calcium ions at positively curved phospholipid membranes. *The Journal of Physical Chemistry Letters*, 8:518–523, 2017.

T. Martinek, E. Duboue-Dijon, Š. Timr, P. E. Mason, K. Baxová, H. E. Fischer, B. Schmidt, E. Pluhařová, and P. Jungwirth. Calcium ions in aqueous solutions: Accurate force field description aided by ab initio molecular dynamics and neutron scattering. *The Journal of Chemical Physics*, 148:222813–1–222813–9, 2018.

T. Megyes, I. Bakó, S. Bálint, T. Grósz, and T. Radnai. Ion pairing in aqueous calcium chloride solution: Molecular dynamics simulation and diffraction studies. *Journal of Molecular Liquids*, 129:63–74, 2006.

D. J. Mitchell, D. T. Kim, L. Steinman, C. G. Fathman, and J. B. Rothbard. Polyarginine enters cells more efficiently than other polycationic homopolymers. *J. Peptide Res.*, 56:318–325, 2000.

K. T. No, K-Y. Nam, and H. A. Scheraga. Stability of like and oppositely charged organic ion pairs in aqueous solution. *The Journal of American Chemical Society*, 119:12917–12922, 1997.

C. Oostenbrink, A. Villa, A. E. Mark, and W. Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53a5 and 53a6. *The Journal of Computational Chemistry*, 25:1656–1676, 2004.

D. Papahadjopoulos, W. J. Vail, W. A. Pangborn, and G. Poste. Studies on membrane fusion II. induction of fusion in pure phospholipid membranes by calcium ions and other divalent metals. *Biochemica et Biophysica Acta*, 448: 265–283, 1976.

E. Pluhařová, H. E. Fischer, P. E. Mason, and P. Jungwirth. Hydration of the chloride ion in concentrated aqueous solutions using neutron scattering and molecular dynamics. *Molecular Physics*, 112:1230–1240, 2014.

C. A. Schnaitman. Solubilization of the cytoplasmic membrane of escherichia coli by triton x-100. *Journal of Bacteriology*, 108:545–552, 1971.

O. Shih, A. H. England, G. C. Dallinger, J. W. Smith, K. C. Duffey, R. C. Cohen, D. Prendergast, and R. J. Saykally. Cation-cation contact pairing in water: Guanidinium. *The Journal of Chemical Physics*, 139:035104–1–035104–7, 2013.

D. Trzesniak, A-P. E. Kunz, and W. van Gunsteren. A comparison of methods to compute the potential of mean force. *ChemPhysChem*, 8:162–169, 2007.

J. VandeVondele and J. Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *The Journal of Chemical Physics*, 127:114105–1–114105–9, 2007.

J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter. QUICKSTEP: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Computer Physics Communications*, 167:103–128, 2005.

M. Vazdar, F. Uhlig, and P. Jungwirth. Like-charge ion pairing in water: An ab initio molecular dynamics study of aqueous guanidinium cations. *The Journal of Physical Chemistry Letters*, 3:2021–2024, 2012.

Loup Verlet. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.*, 159:98–103, 1967.

E. Vivès, P. Brodin, and B. Lebleu. A truncated hiv-1 tat protein basic domain rapidly translocates through the plasma membrane and accumulates in the cell nucleus. *The Journal of Biological Chemistry*, 272:16010–16017, 1997.

P. A. Wender, D. J. Mitchell, K. Pattabiraman, E. T. Pelkey, L. Steinman, and J. B. Rothbard. The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: Peptoid molecular transporters. *PNAS*, 97:13003–13008, 2000.

List of Figures

List of Tables

List of Abbreviations

AIMD Ab-initio Molecular Dynamics

CHARMM Chemistry at Harvard Macromolecular Mechanics

CIP Contact Ion Pair

CN Coordination Number

DFT Density Function Theory

DLS Dynamic Light Scattering

DO Dioleoyl

ECC Electronic Continuum Correction

ECCR Electronic Continuum Correction Rescaled

EDTA Ethylenediaminetetraacetic acid

FF Force Field

FFT Fast Fourier Transform

FRET Förster resonance energy transfer

Gdm^+ Guanidinium

GUV Giant Unilamellar Vesicles

HeLa Henrietta Lacks (cell line)

HEPES 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid

HIV-1 Human immunodeficiency virus type 1

K Lysine

K9 Nona-lysine

LUV Large Unilamellar Vesicles

MD Molecular Dynamics

OPLS-AA Optimized Potential for Liquid Simulations All Atom

PBC Periodic Boundary Conditions

PC Phosphatidylcholine

PE Phosphatidyletanolamine

PME Particle-Mesh Ewald method

PMF Potential of Mean Force

PO Palmitoyl-oleoyl

PS Phosphatidylserine

Q Glutamine

R Arginine

R4 Tetra-arginine

R9 Nona-arginine

RDF Radial Distribution Function

SIP Solvent-shared Ion Pair

SSIP Solvent-separated Ion Pair

SPC/E Extended Simple Point Charge model

SUV Small Unilamellar Vesicles

TIP3P Three-site Transferable Intermolecular Potential

TIP4P Four-site Transferable Intermolecular Potential

TS Transition State

WHAM Weighted Histogram Analysis Method

Attachments

Calcium ions in aqueous solutions: Accurate force field description aided by *ab initio* molecular dynamics and neutron scattering

Tomas Martinek,[1] Elise Duboué-Dijon,[1] Štěpán Timr,[1] Philip E. Mason,[1] Katarina Baxová,[1] Henry E. Fischer,[2] Burkhard Schmidt,[3] Eva Pluhařová,[4,a)] and Pavel Jungwirth[1,a)]
[1]*Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nám. 542/2, 160 00 Prague, Czech Republic*
[2]*Institut Laue-Langevin, 71 Avenue des Martyrs, CS 20156, 38042 Grenoble Cedex 9, France*
[3]*Institut für Mathematik, Freie Universität Berlin, Arnimallee 6, D-14195 Berlin, Germany*
[4]*J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, v.v.i., Dolejškova 2155/3, 182 23 Prague 8, Czech Republic*

(Received 27 September 2017; accepted 2 December 2017; published online 6 March 2018)

We present a combination of force field and *ab initio* molecular dynamics simulations together with neutron scattering experiments with isotopic substitution that aim at characterizing ion hydration and pairing in aqueous calcium chloride and formate/acetate solutions. Benchmarking against neutron scattering data on concentrated solutions together with ion pairing free energy profiles from *ab initio* molecular dynamics allows us to develop an accurate calcium force field which accounts in a mean-field way for electronic polarization effects via charge rescaling. This refined calcium parameterization is directly usable for standard molecular dynamics simulations of processes involving this key biological signaling ion. *Published by AIP Publishing.* https://doi.org/10.1063/1.5006779

I. INTRODUCTION

The calcium ion is a key signaling species involved in many biological processes including allosteric enzyme activation, neurotransmitter release, muscle contraction, and others.[1–4] Consequently, numerous recent computer modeling studies have aimed at elucidating the molecular mechanisms of the biological actions of calcium.[5–10] In general, a molecular simulation can only be as accurate as the underlying force field. With the commonly employed non-polarizable force fields, problems may arise to accurately describe aqueous ions of high charge density, such as alkali earth dications, due to important, but neglected, electronic polarization effects.[11] As a consequence, such simulations may not accurately describe without additional parameterization the ion pairing of these ions in aqueous solutions and their binding to negatively charged groups in proteins, nucleic acids, or phospholipids.[12–16]

Recently, a simple but physically well justified way of accounting in a mean-field way for electronic polarization (missing in non-polarizable force fields) via rescaling ionic charges has been suggested.[17,18] The scaling factor of ~0.75 is the inverse square root of the electronic (high frequency) part of the dielectric constant of water.[19] This scaling factor is directly applicable to atomic ions with integer charge. For atoms with partial charges, including those forming water molecules, the situation is more complicated, not least because varying extent of charge scaling has already been applied implicitly when fitting the force field against experimental observables.[18] In our previous studies, we have applied this approach to develop a charge scaled model of calcium benchmarked against structural neutron scattering data[11,20] and successfully applied it to quantify its affinity to a common calcium-binding protein, calmodulin,[13] and to phospholipids in model membranes.[14,15]

Here, we introduce an additional benchmarking approach, namely, *ab initio* molecular dynamics (AIMD), which allows us (albeit at a significant computational cost) to generate accurate free energy profiles[21] for pairing of calcium with its counter-ion in water. We apply this approach to aqueous calcium chloride and formate solutions, the latter serving as a model for the interaction of calcium with the carboxylic side chain groups of glutamate or aspartate and the protein C-terminus. Together with neutron scattering experiments on analogous systems, this allows us to refine the calcium force field, such that it is now applicable for accurate simulations of biological processes involving this crucial signaling ion. As a bonus, we obtain a faithful description of these important calcium salt solutions with quantitative molecular details concerning the hydration structure of the ions as well as their tendency to form contact ion pairs (CIPs) and solvent-shared ion pairs.

II. METHODS

A. Experimental details—Neutron scattering

Four solutions of calcium acetate were prepared with different isotopic constitutions, which are summarized in Table I. All these solutions were prepared using a common procedure as follows. Calcium oxide (anhydrous 99.99%, Sigma-Aldrich) was dissolved in the stoichiometric amount of acetic acid to yield a calcium acetate solution, which consists after

a)Authors to whom correspondence should be addressed: pavel.jungwirth@uochb.cas.cz and eva.pluharova@jh-inst.cas.cz

TABLE I. Isotopic composition of calcium acetate solutions used in neutron scattering measurements.

Sample name	Acid used for preparation	Solvent
$(H_3CCOO)_2Ca$ in H_2O	H_3CCOOH	H_2O
$(D_3CCOO)_2Ca$ in H_2O	D_3CCOOH	H_2O
$(H_3CCOO)_2Ca$ in D_2O	H_3CCOOD	D_2O
$(D_3CCOO)_2Ca$ in D_2O	D_3CCOOD	D_2O

neutralization of a ratio of water to acetate 2:56.55. For convenience, this is referred to as a 1 m calcium acetate solution in this study. The solution contains 2 m acetate and 1 m calcium concentrations.

Total neutron scattering patterns were measured for the four calcium acetate solutions (Table I) on the D4C diffractometer[22] at the ILL in Grenoble, France, with neutrons with a wavelength of λ = 0.5 Å. Data[23] were recorded for 3 h for each D_2O sample and for 6 h for each H_2O sample. The raw scattering data were then corrected for multiple scattering and absorption,[24] before being normalized versus a standard vanadium rod. This provided for each sample the total scattering pattern, $S(Q)$. First order differences, $\Delta S(Q)$, were then obtained by subtracting the total scattering patterns of CD_3COO (D_3-acetate) and CH_3COO (H_3-acetate) solutions. These first order differences were obtained both in H_2O and D_2O. Each of them can be expressed as a sum of pairwise structure factors, which contains only structural data from the substituted non-exchangeable hydrogen "*Hsub*" to any other atom in solution, "*X*," with all the other terms canceling out. The expressions for the first order differences in Q-space are provided below, "*Hex*" referring to the exchangeable hydrogen atoms on water and the prefactors (expressed in millibarns) being calculated from the concentrations and coherent scattering length of each nucleus.[25] The offset is subtracted so that the scattering patterns oscillate around zero at long Q,

$$\begin{aligned}\Delta S^{D_2O}_{HsubX}(Q) = {} & 27.66\, S_{HsubHex}(Q) + 12.03\, S_{HsubOw}(Q) \\ & + 0.85\, S_{HsubOc}(Q) + 0.97\, S_{HsubC}(Q) \\ & - 0.17\, S_{HsubCa}(Q) + 0.32\, S_{HsubHsub}(Q) - 42.02,\end{aligned} \tag{1}$$

$$\begin{aligned}\Delta S^{H_2O}_{HsubX}(Q) = {} & -15.51\, S_{HsubHex}(Q) + 12.03\, S_{HsubOw}(Q) \\ & + 0.85\, S_{HsubOc}(Q) + 0.97\, S_{HsubC}(Q) \\ & - 0.17\, S_{HsubCa}(Q) + 0.32\, S_{HsubHsub}(Q) + 1.15.\end{aligned} \tag{2}$$

The first order difference in H_2O was further subtracted from the first order difference in D_2O to yield the second order difference, ΔΔS(Q), which contains only *HsubHex* correlations,

$$\begin{aligned}\Delta\Delta S(Q) &= \Delta S^{D_2O}_{HsubX}(Q) - \Delta S^{H_2O}_{HsubX}(Q) \\ &= 43.17 S_{HsubHex}(Q) - 43.17,\end{aligned} \tag{3}$$

which can be Fourier-transformed to yield the corresponding second-order difference in r-space,

$$\begin{aligned}\Delta\Delta G(r) &= \Delta G^{D_2O}_{HsubX}(r) - \Delta G^{H_2O}_{HsubX}(r) \\ &= 43.17 g_{HsubHex}(r) - 43.17.\end{aligned} \tag{4}$$

B. Computational details

1. Force field molecular dynamics

A set of aqueous solutions was simulated to investigate the association of the calcium cation with its most common counterions—chloride anion as an example of the simplest spherical counterion—and formate or acetate (Ac) anions to mimic the association with carboxylic groups of biological molecules. Systems containing one ion pair (a calcium cation with a single chloride, formate or acetate counterion) in explicit water were simulated to obtain the free energy profiles for the corresponding ion associations and compare the behavior of different force fields with *ab initio* density functional theory (DFT) simulations. In addition, force field simulations of concentrated solutions of $CaCl_2$ and $CaAc_2$ were performed to directly compare the results computed with different force fields with experimental neutron scattering data. Table II summarizes the compositions of the different simulated systems and provides in each case additional details about the simulation setup.

All force field molecular dynamics simulations were performed with the Gromacs software[26] using the leap frog propagator with a 1 fs time step at a constant temperature of 300 K maintained by the Canonical Sampling through Velocity Rescaling (CSVR) thermostat with a time constant of 0.2 ps.[27] Water molecules were described using the extended simple point charge (SPC/E) force field,[28] their geometry being kept rigid by the Settle algorithm.[29] Periodic boundary conditions were employed and long-range electrostatic interactions were treated by the particle mesh Ewald method.[30] When the Parinello-Rahman barostat was employed, the pressure coupling constant was set to 1 ps. Each concentrated solution ($CaCl_2$ or $CaAc_2$) was equilibrated for at least 20 ns before a production run of at least 20 ns long, which was used to evaluate the radial distribution functions required to compute the neutron scattering signal.

In this study, we employed several force fields for calcium and chloride ions. First, we used a common full charge

TABLE II. Overview of simulated systems with relevant details. For simulations in the NpT ensemble, the cell size is the average value obtained from simulation runs, which slightly depends on the employed force field.

System	Composition	Ensemble	Cell size (Å)	Cutoffs (Å)
Calcium chloride ion pair	1 Ca^{2+}, 1 Cl^-, 64 H_2O	NVT	12.5	6
4 m $CaCl_2$ solution	52 Ca^{2+}, 104 Cl^-, 719 H_2O	N*p*T	~29.3	12
Calcium formate ion pair	1 Ca^{2+}, $HCOO^-$, 107 H_2O	NVT	14.73	6
1 m $CaAc_2$ solution	81 Ca^{2+}, 162 CH_3COO^-, 4581 H_2O	N*p*T	~53	12

TABLE III. Force fields for the calcium and chloride ions.

	Ca^{2+}			Cl^-		
	σ (Å)	ε (kJ/mol)	Charge (*e*)	σ (Å)	ε (kJ/mol)	Charge (*e*)
FULL	2.8196	0.5072	+2.00	4.4499	0.4184	1.00
ECCR	2.5376	0.5072	+1.50	4.1000	0.4928	0.75
ECCR2	2.6656	0.5072	+1.50	4.1000	0.4928	0.75

force field for both Ca^{2+} and Cl .[31,32] We hereafter denote this force field FULL. We compare this standard force field to the scaled charge calcium force field recently developed in our group, which employs the electronic continuum correction with ionic size refinement (ECCR).[11] Here, the ionic charges are scaled by the factor 0.75 to account for electronic polarization in a mean field way.[17] The ionic Lennard-Jones parameter σ is correspondingly reduced to recover the proper Ca^{2+}–H_2O interaction, as estimated from neutron scattering data. In this work, we propose a refined Ca^{2+} scaled charge force field, denoted as ECCR2, which provides better agreement with our *ab initio* simulation data, leading to a slightly enlarged calcium radius compared to the ECCR model. In combination with both ECCR and ECCR2 scaled charges calcium force fields, we used a scaled charge force field for the chloride anion, the parameters of which were refined compared to our previous study,[11] which was parametrized in our group using reference neutron scattering data of lithium chloride salt.[33] Details of the force fields used for Ca^{2+} and Cl are provided in Table III.

In addition, we designed a force field for the acetate and formate anions based on the Amber ff99 force field,[34] the charges being obtained from a fit of the electrostatic potential on an optimized geometry. These calculations were performed using the Gaussian 09 software[35] with the Hartree-Fock method with the 6-31G* basis set[36] employed both for the geometry optimization and the electrostatic potential calculation. The resulting Restrained ElectroStatic Potential (RESP) charges are listed in Tables SI and SII of the supplementary material. The corresponding scaled charge force field was obtained by scaling the RESP charges by the factor 0.75 without modification of the van der Waals potentials.

2. Ab initio molecular dynamics

Born-Oppenheimer *ab initio* molecular dynamics simulations were performed using the Quickstep module of the CP2K package[37] implementing the hybrid Gaussian functions and plane waves (GPW) method.[38] All simulations were performed with a time step of 0.5 fs in the NVT ensemble, using the CSVR thermostat[27] with a time constant of 50 fs. Periodic boundary conditions were used. The electronic structure of the system was treated at the DFT level of theory with the BLYP functional[39,40] with the Grimme correction scheme (DFT-D2)[41] to account for dispersion interactions. The core electrons were described by norm-conserving Goedecker-Teter-Hutter (GTH) pseudopotentials.[42] Kohn-Sham orbitals were expanded in a Gaussian basis set: TZV2P MOLOPT for O, H, and C, and DZVP-MOLOPT-SR-GTH-q10 for Ca in the aqueous Ca^{2+} HCOO system and in a DZVP-MOLOPT-SR-GTH Gaussian basis set for the aqueous Ca^{2+} Cl system.[38] Previous studies demonstrated that the DZVP basis set yields similar results to TZVP, while being significantly less computationally demanding.[43] A cutoff of 400 Ry was used for the auxiliary plane wave basis set.

Free energy profiles for the ion pairing were obtained for a single CaCl ion pair and a single calcium-formate ion pair at the *ab initio* MD level and compared to the results of force field simulations on the same systems. Converging such *ab initio* free-energy profiles is extremely challenging and computationally expensive so that only a few attempts in this direction can be found in the literature,[21,44] including studies of the Ca–Cl ion pair.[45,46] The free energy profile along the Ca–Cl distance was obtained at the *ab initio* level using two (in principle equivalent) methods: (i) integration of mean force and (ii) umbrella sampling. This allowed us to check the convergence of our computationally very demanding calculations and estimate the associated error. In method (i), the average force between the studied ions is evaluated at different fixed interionic distances and consequently integrated along the Ca–Cl distance to obtain the potential of mean force (PMF). By contrast, the umbrella sampling method uses a set of biasing harmonic potentials along the Ca–Cl distance to enhance the sampling of rare ionic configurations. The initial configurations for each window of the *ab initio* simulations were taken from classical MD simulations of the same system. The umbrella sampling windows were then combined and unbiased to obtain the free-energy profile using the Weighted Histogram Analytic Method (WHAM) algorithm.[47] The standard volume entropy correction[44] ($+2k_BT\ln(r)$) was applied to all the obtained free energy profiles.

A total of 38 simulations with fixed Ca–Cl interionic distances ranging from 2.2 Å to 6.0 Å were performed to obtain the average mean force along the Ca–Cl distance, while the umbrella sampling simulations used a total of 10 windows with distances ranging from 2 to 6 Å. Each umbrella sampling window was duplicated starting with two different initial geometries to ensure proper sampling of relevant calcium-water coordination numbers.[46] Each window was equilibrated for 10 ps before a 50 ps production run. The error bars were evaluated as the standard deviations obtained from 10 free-energy profiles generated from 5 ps cuts of the simulation production run.

Unlike the case of Ca^{2+}–Cl , *ab initio* MD simulations are too expensive to allow for satisfactorily converged sampling of all degrees of freedom of the structurally more complex Ca^{2+}–HCOO ion pair. It is thus beyond our reach to fully characterize the *ab initio* free energy landscape for the ion-ion interaction. Nevertheless, we were able to obtain the free-energy profile along the Ca–O distance in a monodentate arrangement of the ion pair using umbrella sampling simulations (13 windows) by restraining the COCa angle around 180° using a harmonic force constant of 100 kJ mol ^{1} rad ^{2}. To further investigate the interplay between the mono- and bidentate interaction mode, we generated the free-energy profile along the Ca–C distance (22 windows), restraining Ca^{2+} to the OCO plane using a harmonic restraint on the OCOCa dihedral angle with a force constant of 100 kJ mol ^{1} rad ^{2}. Due to the higher complexity of the phase space and applied restraints, the standard radial volume entropy correction cannot be employed

here. The procedure used to correct the free energy profiles for the sampled volume is described in detail in the supplementary material.

3. Comparison between ab initio and force field free energy profiles

The various *ab initio* free energy profiles were compared with the corresponding profiles obtained with different force fields, for which full convergence could be reached using longer simulation times and additional umbrella sampling windows. We further characterized with the different force fields the detailed interaction of a calcium cation with the carboxylic group of the formate anion. In particular, we computed the full 2D free energy profile in the OCOCa plane along the CaC distance and Ca–C–O_{mid} angle, where O_{mid} is the middle point between the two carboxylate oxygen atoms. The distance coordinate was binned from 2.7 to 6.1 Å and the angular coordinate was binned from 0 to 110°. The sampling was restrained to the formate plane using the same restraints as above.

III. RESULTS AND DISCUSSION

A. Refining the Ca^{2+} force field using *ab initio* simulations

First, we investigate the Ca^{2+}–Cl ion pairing using AIMD simulations of a single ion pair in water. We obtain the *ab initio* free energy profile along the Ca^{2+}–Cl distance. Figure 1 compares two independent free energy methods—umbrella sampling and integration of the mean force. The free energy profiles obtained by the two approaches are consistent with each other within the estimated error bars. Both predict a Ca^{2+}–Cl distance of 2.75 ± 0.05 Å at the contact ion pair (CIP), as well as the same position of 3.75 ± 0.05 Å of the barrier between the CIP and the Solvent Shared Ion Pair (SShIP), and an almost symmetrical barrier of 10 ± 1 kJ/mol between the two minima. The estimate of the SShIP position is less accurate because of increasing statistical uncertainties at larger distances; nevertheless it is found to be around 5.0 Å. The free energies of the CIP and SShIP are practically identical within the error bars.

The good agreement between the umbrella sampling and mean force integration methods points to a satisfactory convergence of the obtained Ca^{2+}–Cl free energy profile. This is further supported by comparison of the present results to a free energy profile obtained for an analogous system using a similar method recently.[46] With this reassurance, we use below the free energy profile obtained using the umbrella sampling technique (with the slightly smaller statistical error) as the reference for comparison with force field simulations and for further force field refinement.

We compared AIMD results with three empirical force field simulations. The first one is a standard full charge force field (denoted here as "FULL") with full integer charges on both ions. This full charge force field provides a free energy profile significantly different from the AIMD reference (Fig. 2). In particular, this force field significantly underestimates the stability of the CIP. It is found higher in free energy by about 8 kJ/mol compared to the SShIP, with the barrier between SShIP and CIP being 19 kJ/mol, i.e., 8 kJ/mol higher than using AIMD. The Ca–Cl distance at the CIP is only slightly shifted compared to AIMD, and the transition barrier is found at 3.6 Å, which is close to the AIMD value of 3.7 Å.

The second force field, ECCR, previously[11] designed by us, employs the electronic continuum correction to account for water polarization, with the size of the calcium ion (i.e., the Lennard-Jones σ-parameter) fitted to experimental neutron scattering data. While this force field reproduces quantitatively the neutron scattering data available in the literature[20] (Fig. 3), it does not compare so well with the present *ab initio* free energy profile (Fig. 2). Namely, the Ca–Cl distance at the CIP is slightly shorter than that found in the *ab initio* simulations, 2.72 vs 2.80 Å. In addition, the barrier height is found significantly larger than that for the *ab initio* reference, 20 vs 11 kJ/mol.

Analysis of the simulations shows that the relatively small discrepancy between the *ab initio* and ECCR force field profiles stems mainly from different Ca–O(water) distances. Namely, the average Ca–O distance obtained from the AIMD simulation is 2.43 Å, in contrast to 2.35 Å with the ECCR force field, which was fitted to reproduce the first peak of the neutron scattering pattern (2.38 Å) corresponding to the Ca–O and Ca–Cl correlations. To ensure that our DFT setup provides reliable Ca–O optimal distances, we optimized the geometry of a perfectly symmetric $Ca^{2+}(H_2O)_6$ cluster at different levels of

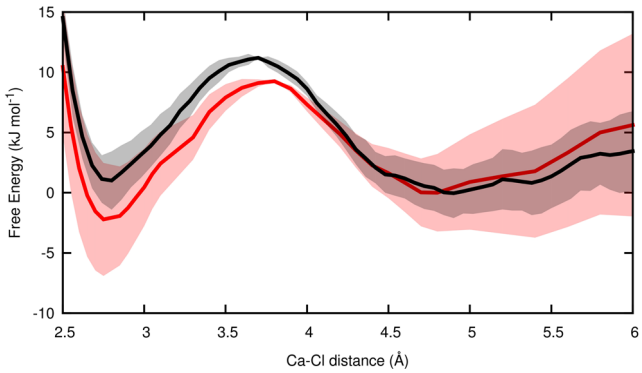


FIG. 1. Free energy profile along the Ca^{2+}–Cl^- distance obtained by mean force integration (red) and umbrella sampling (black).

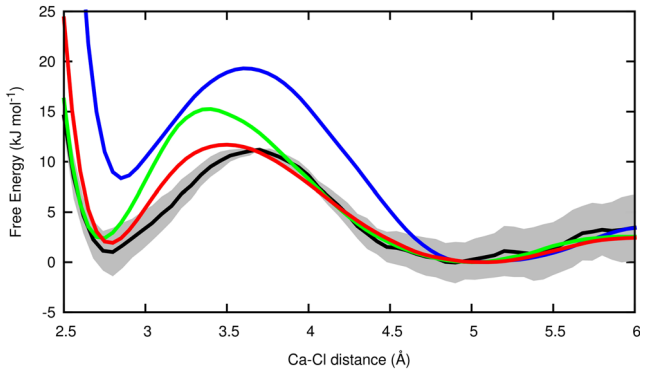


FIG. 2. Free energy profile along the Ca^{2+}–Cl^- distance obtained by umbrella sampling (black) compared with force field simulations performed using the full charges force field (blue), the ECCR (green), and the ECCR2 force field (red).

theory, employing DFT methods (BLYP and B3LYP) as well as HF, full MP2, and Coupled Cluster Singles and Doubles (CCSD) with basis sets of various sizes (6-31G*, 6-31+G*, 6-311++g**, cc-pVDZ, and cc-pVTZ). The Ca–O distances at the optimized geometries are provided in the supplementary material, showing that the Ca–O distance is practically insensitive to the level of theory, which means that the BLYP-D2 level of description used in our AIMD simulations does not suffer from any significant systematic deviation in comparison to higher levels of theory or larger basis sets. Moreover, a very similar AIMD setup was already shown to correctly reproduce the structure of pure water[48] and a very recent study[46] showed that the EXAFS spectrum of $CaCl_2$ solutions was extremely well reproduced using the same AIMD setup as in the present study.

Hence, we suggest here a refined parametrization for Ca^{2+} based on the previous ECCR one (hence we call it ECCR2), with a calcium cation σ-parameter enlarged by 5% to reach a better agreement with the AIMD free energy profile.[46] As seen in Fig. 2, the position of CIP at 2.80 Å is now fully consistent with the AIMD result. The free energies of the CIP and SShIP are found to be very similar to each other, with the CIP being only 1.8 kJ/mol higher in energy (compared to 0.9 kJ/mol with AIMD, well within the AIMD error bars). The SShIP and CIP are separated by a 11.4 kJ/mol high transition barrier, which is in much better agreement with the AIMD result of 11 kJ/mol than the original ECCR force field. Finally, we show (Fig. S4 of the supplementary material) that the free energy profile obtained with the ECCR2 scaled charge force field is in quantitative agreement with that obtained using the fully polarizable force field AMOEBA, itself being within the error bars of the AIMD calculation. The price is that the ECCR2 description does not compare as well as the original ECCR with the neutron scattering data. Nevertheless, the comparison is still good and presents a significant improvement compared to the full charges force field (Fig. 3). The new ECCR2 calcium parametrization thus appears to be the best compromise between agreement with the available neutron scattering data on the one hand and with EXAFS experiments and advanced AIMD simulations on the other hand.

B. Ca^{2+} interaction with the carboxylate group

Having in hand a refined calcium force field that provides a satisfactory agreement with both AIMD and neutron data, we now aim at testing the transferability toward a description of the interaction of calcium with negatively charged protein residues. While the calcium ion also interacts with the polar protein backbone,[49] the interaction with the negatively charged carboxylates of aspartate and glutamate is expected to be the strongest. We thus focus here on the carboxylate group using the acetate anion as a simple proxy.

The total neutron scattering patterns $S(Q)$ [Fig. 4(a)] were obtained for four 1 m calcium acetate solutions with different isotopic compositions (see Sec. II and Table I). These patterns exhibit a large background slope, due to the Placzek effect.[50] This effect is more pronounced for H_2O solutions than for D_2O, as expected from the higher incoherent scattering cross section of ^1H versus ^2H. The total $S(Q)$ is largely dominated by contributions from water-water interactions and the calcium acetate interaction is thus mostly hidden. The use of H/D isotopic substitutions, both on water and on the non-exchangeable hydrogen of the acetate, thus allows us to focus on the acetate hydration properties and indirectly on the ion pairing with calcium ion, since ion pairing displaces water molecules from the hydration shell.

The first order differences in H_2O and D_2O $\Delta S^{H_2O}_{HsubX}(Q)$ and $\Delta S^{D_2O}_{HsubX}(Q)$ are obtained by direct difference between pairs

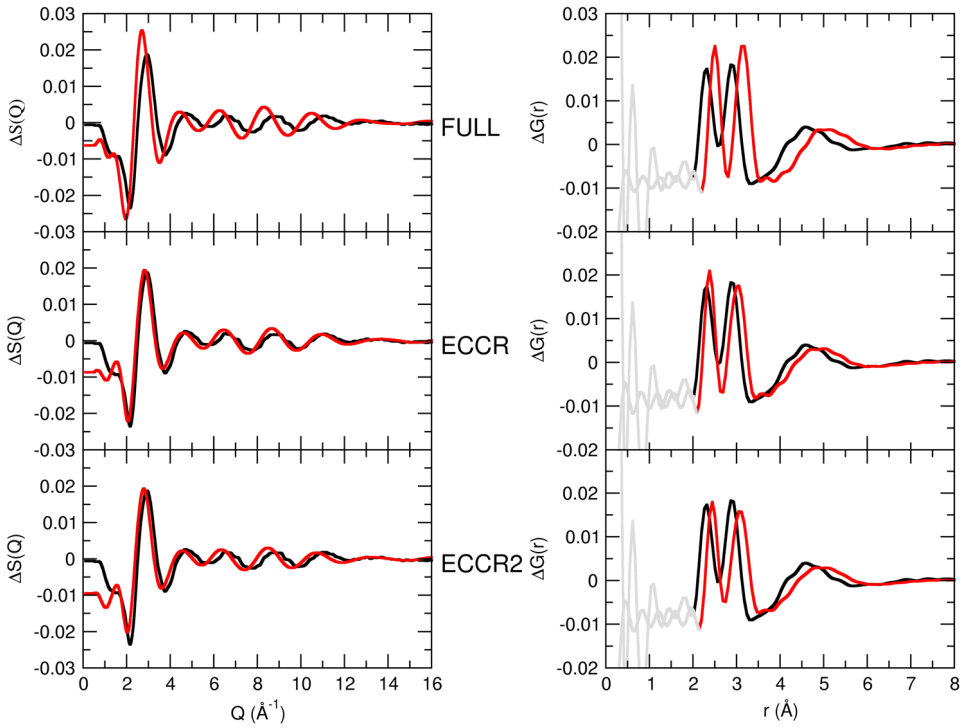


FIG. 3. First order difference (on calcium) in Q-space (left-hand side) and in r space (right-hand side) for a concentrated 4 m $CaCl_2$ solution, as obtained from the neutron scattering experiment (black) and from simulations with different force fields, full charges (top), ECCR (middle), and ECCR2 (bottom). Since the experimental signal was obtained by subtracting the neutron scattering patterns obtained with different calcium isotopes, the signal only reports on correlations involving the Ca atom. See more details in Refs. 11 and 20.

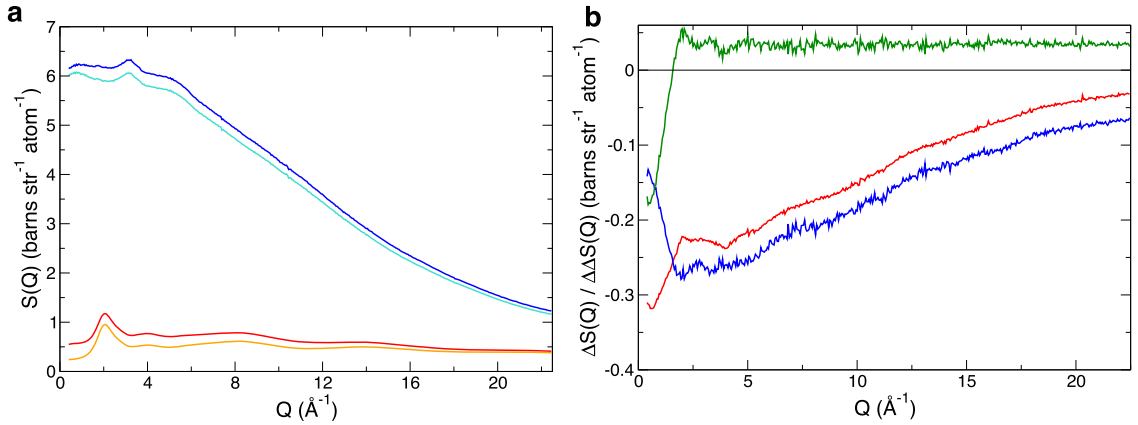


FIG. 4. (a) Total neutron scattering patterns for 1 m solutions of calcium h3-acetate (dark blue) in H_2O, calcium d3-acetate (light blue) in H_2O, calcium h3-acetate (red) in D_2O, and calcium d3-acetate (orange) in D_2O. (b) First order differences in H_2O, $\Delta S^{H_2O}_{HsubX}(Q)$ (blue), and in D_2O, (Q) (red), together with the second order difference, $\Delta\Delta S\,(Q)$ (green).

of solutions with different isotopes on the acetate. They only contain the correlation between the substituted acetate hydrogens *Hsub* and any other atom in the system, *X* [Eqs. (1) and (2)]. This subtraction thus cancels most of the background slope and removes the large water-water correlations from the signal. However, the first order differences still contain strong intramolecular correlations (*Hsub*-C and *Hsub*-O from the same acetate molecule), which dominate the signal but are of little interest in this work. This problem is solved by taking the difference between the two first order differences [Eq. (3)] yielding the second order difference $\Delta\Delta S(Q)$ (Fig. 4). This has no residual slope and reports exclusively on the part of the system we are directly interested in, i.e., the acetate hydration shell, since it contains only correlations between the substituted hydrogens *Hsub* on acetate and the exchangeable water hydrogens *Hex* [see Eq. (3)]. In all that follows, the constant offset is subtracted from the double difference so that it oscillates around zero, consistently with the way we defined $\Delta\Delta S(Q)$ in Eq. (3).

Comparison of the experimental neutron second order difference with that computed with different force fields constitutes a stringent test of the quality of a force field and of its ability to capture the hydration and ion pairing properties of acetate in concentrated calcium acetate solutions. Figure 5(a) shows that while a standard full charge force field correctly captures the structure at large values of Q, it exhibits a spurious sharp peak at low Q (below 1 Å ^{1}), which is not present in the experimental signal. Such low Q features typically point to clustering in the solution, which is confirmed by visual examination of the MD simulations [Fig. 6(a)]. Indeed, when a full charge description is used, all the ions form contact ion pairs with no free ion left in solution. This behavior results in an unrealistic depletion at short distance of the simulated r-space signal (Fig. 5), since the clustering of ion pairs effectively reduces the number of water molecules (and thus *Hex*) in the vicinity of each acetate molecule.

Scaling only the charge of the calcium ion reduces marginally this spurious peak at low Q and the structure of the solution remains very much the same as for the full charge model (see Fig. S5 of the supplementary material for comparison of the ECCR and ECCR2 parameterizations yielding essentially the same result). A good agreement with experiment is obtained only upon scaling both calcium and acetate ions. Figure 5(c) shows that in this case the experimental Q-space signal is almost perfectly reproduced with the proper low-Q limit. The r-space signal is no longer depleted at short distances, and the agreement with experiment becomes very good. A snapshot of the simulation [Fig. 6(c)] shows that ion pairing is much weaker than in the previous two simulations, now with many free ions in solution. As a result, the solution is much more homogeneous with no sign of strong ion clustering. This study thus illustrates an important point for biological simulations—when using the scaled charge description, one should scale not only the charges of the ions in the solution but also those of *charged* protein residues, as already suggested in the literature.[17]

While the neutron scattering data allows us to assess the quality of the employed force field, they represent only an indirect probe of ion pairing. Namely, the amount of ion pairing can be assessed via depletion of the number of hydrating water molecules and through the presence or absence of low Q signal due to strong ion pairing and clustering. We thus complement the neutron scattering experiment with a molecular level study of the calcium-carboxylate interaction using *ab initio* MD simulations. Since such simulations are extremely computationally expensive, we opted for the smallest carboxylate group containing species, the formate anion, which allows us to use a somewhat smaller simulation box.

An interesting property of the carboxylate group is that it can pair with Ca^{2+} either in a monodentate fashion (i.e., interacting with only one of the two carboxylate oxygens) or in a bidentate fashion (i.e., interacting simultaneously with both oxygens). We first focus on the monodentate interaction mode and obtain the free energy profile along the Ca–O(formate) distance using *ab initio* MD umbrella sampling simulations [Fig. 7(a)]. We then compare the AIMD free energy profile with those obtained with a full charge force field for both calcium and formate moieties vs. a scaled charge force field [i.e., ECCR2 for Ca^{2+} and simple charge scaling for formate (see Sec. II)]. At the *ab initio* level, the contact monodentate ion pair at a Ca–O distance of 2.3 Å is found to be 6 kJ/mol more stable than the SShIP. The height of the transition barrier from

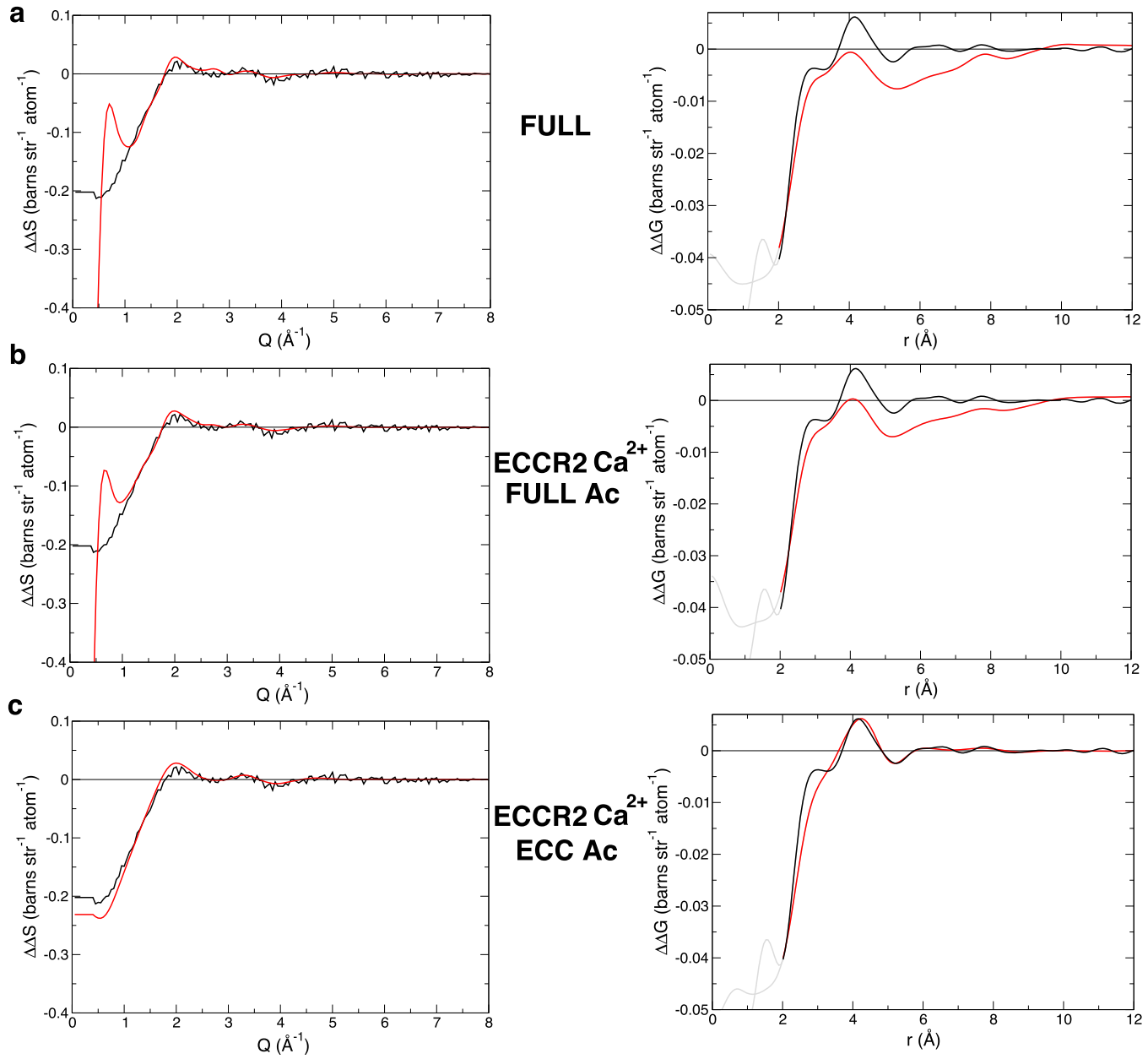


FIG. 5. Comparison between the experimental neutron double difference (black) in Q space, $\Delta\Delta S(Q)$ (left), and in r space, $\Delta\Delta G(r)$ (right), and the signal calculated from molecular dynamics simulations (red) using (a) a full charge force field, (b) the ECCR2 force field for Ca^{2+} and full charges for the acetate, and (c) the ECCR2 force field for Ca^{2+} and a ECC description for the acetate anion.

SShIP to CIP is around 11.5 kJ/mol. The stability of the contact ion pair is significantly overestimated, by 10 kJ/mol, with the full charge force field. This is in accord with the above simulations of concentrated calcium acetate, which yield a strong pairing in solutions described with a full charge force field. By contrast, the use of a scaled charge force field both

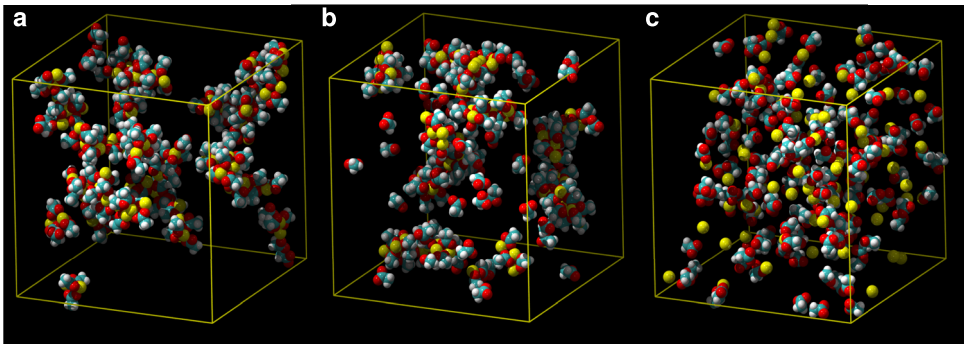


FIG. 6. Snapshots from 1 m calcium acetate MD simulations performed using (a) a full charges force field, (b) the ECCR2 force field for Ca^{2+} and full charges for the acetate, and (c) the ECCR2 force field for Ca^{2+} and a ECC description for the acetate anion.

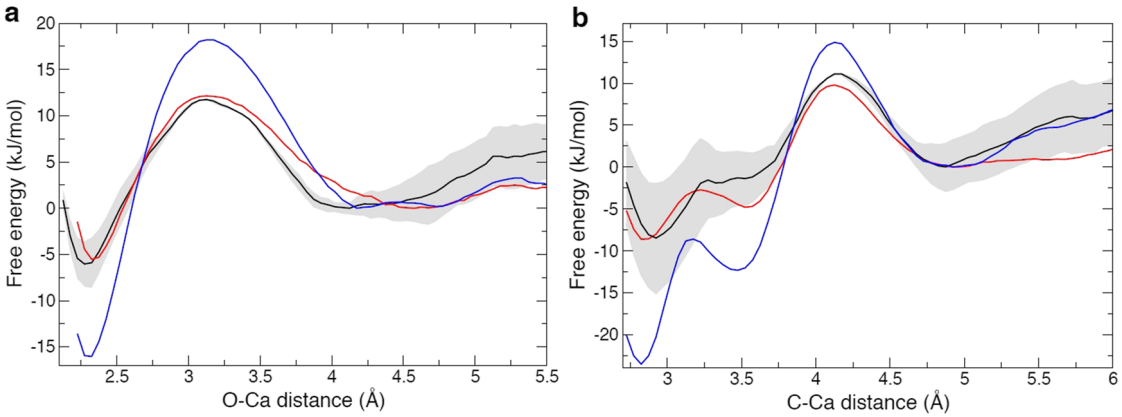


FIG. 7. (a) Free energy profile along the Ca–O(formate) distance in a linear monodentate fashion and (b) free energy profile along the C–Ca distance in the formate plane using AIMD (black), the full charges force field (blue), or a scaled charges force field both for calcium and formate (red).

for calcium and acetate provides a quantitative agreement with AIMD, both for the relative stability of the contact ion pair and the height of the barrier [Fig. 7(a)].

Since the neutron scattering measurements cannot provide full information about the mode of calcium-carboxylate interaction (monodentate vs bidentate), we explore it in more detail computing the free energy profile along the Ca-C(formate) distance in the COO plane with *ab initio* MD [Fig. 7(b)]. We see that upon prolonging this distance the system is progressively driven from a bidentate geometry (Ca–C distances of 2.9 Å) to a monodentate interaction (Ca–C distance of 3.5 Å). We find the bidentate geometry more stable than the monodentate one by about 7 kJ/mol. The full charge force field again overestimates the stability of the bidentate arrangement—the bidentate well being found 11 kJ/mol lower in free energy than the monodentate arrangement as well as the height of the barrier toward the SShIP. By contrast, the scaled charge force field is in quantitative agreement with the AIMD profile, both for the relative position of the two wells and for the barrier height.

Finally, we used this scaled charge force field, to fully characterize the calcium ion pairing in the formate plane by generating the 2D free energy landscape (Fig. 8). Note that such a 2D free energy landscape requires too much sampling to be currently obtained at the AIMD level. Figure 8 demonstrates that the global minimum is at a Ca–C distance just below 3 Å with a nearly collinear Ca–C–O_{mid} arrangement, thus corresponding to ion pairing in a bidentate fashion. Following the low energy areas (blue), we see that the dissociation from the bidentate ion pair proceeds via a rearrangement into a monodentate conformation (broad minimum with Ca–C–O_{mid} angle around 50° and a Ca–C distance of about 3.5 Å), with a very small barrier of 6 kJ/mol. From this point on, solvent shared configurations are accessible with a 14 kJ/mol barrier. The SShIP local minimum is found at around 4.8–5.4 Å CaC with a weak preference for a collinear Ca–C–O_{mid} arrangement.

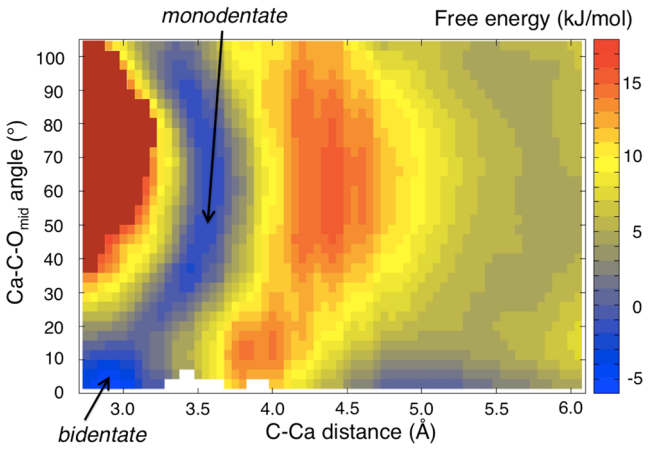


FIG. 8. 2D Free energy landscape along the C(formate)–Ca distance and the Ca–C–O_{mid} angle in the formate plane.

IV. CONCLUSIONS

We have employed *ab initio* molecular dynamics simulations together with neutron scattering experiments to quantify ion hydration and pairing in aqueous calcium chloride and formate/acetate solutions and to refine a calcium force field applicable for simulations involving this key biological ion. By inclusion of electronic polarization effects in a mean-field way via charge rescaling of both calcium and the counter-ion (be it a free anion in the solution or a negatively charged side chain group on a protein), we are able to accurately describe calcium ions in aqueous solutions. On the one hand, we quantitatively reproduced structural features of concentrated aqueous solutions, as revealed by neutron scattering with isotopic substitution. On the other hand, the refined force field recovered within the statistical error ion-pairing free energy profiles from *ab initio* molecular dynamics. We are thus making available to the scientific community a physically well justified transferable calcium parameterization, which, thanks to its standard format, can be directly used within common molecular dynamics programs and at the same time provides an accurate description including electronic polarization effects of this ion in aqueous biological environments.

SUPPLEMENTARY MATERIAL

See supplementary material for the acetate and formate force fields, the procedure used to perform volume correction on the free energy profiles, electronic structure calculations

on a $Ca(H_2O)_6^{2+}$ cluster, and comparison of our results with additional free energy profiles, including that obtained with the AMOEBA force field.

ACKNOWLEDGMENTS

P.J. acknowledges support from the Czech Science Foundation (Grant No. P208/12/G016). E.D.-D. acknowledges support from the EMBO and Marie Curie Actions (Fellowship No. ALTF 952-2015). E.P. thanks the Czech Science Foundation (Grant No. 17-01982Y).

We thank the Institute Laue Langevin for allocating us time on the neutron scattering facility (Proposal No. 8-03-807). We particularly thank the staff of D4C and of the ILL chemistry laboratory for their help with the neutron scattering experiments. Calculations of the free energy profiles were made possible through a generous allocation of computer time from the North-German Supercomputing Alliance (HLRN).

[1] A. G. Szent-Györgyi, Biophys. J. **15**, 707 (1975).
[2] T. M. Perney, L. D. Hirning, S. E. Leeman, and R. J. Miller, Proc. Natl. Acad. Sci. U. S. A. **83**, 6656 (1986).
[3] D. E. Clapham, Cell **131**, 1047 (2007).
[4] H. Tidow and P. Nissen, FEBS J. **280**, 5384 (2013).
[5] P. M. Kekenes-Huskey, V. T. Metzger, B. J. Grant, and J. Andrew McCammon, Protein Sci. **21**, 1429 (2012).
[6] A. O. Aykut, A. R. Atilgan, C. Atilgan, M. Ikura, and J. Maune, PLoS Comput. Biol. **9**, e1003366 (2013).
[7] A. Amcheslavsky, M. L. Wood, A. V. Yeromin, I. Parker, J. A. Freites, D. J. Tobias, and M. D. Cahalan, Biophys. J. **108**, 237 (2015).
[8] I. Frischauf, V. Zayats, M. Deix, A. Hochreiter, I. J. Polo, M. Muik, B. Lackner, B. Svobodová, T. Pammer, M. Litviňuková, A. A. Sridhar, I. Derler, I. Bogeski, C. Romanin, R. H. Ettrich, and R. Schindl, Sci. Signaling **8**, ra131 (2015).
[9] Y. I. Neela, A. S. Mahadevi, and G. N. Sastry, J. Phys. Chem. B **114**, 17162 (2010).
[10] J. Adiban, Y. Jamali, and H. Rafii-Tabar, Mol. BioSyst. **13**, 208 (2017).
[11] M. Kohagen, P. E. Mason, and P. Jungwirth, J. Phys. Chem. B **118**, 7902 (2014).
[12] S. Mamatkulov, M. Fyta, and R. R. Netz, J. Chem. Phys. **138**, 24505 (2013).
[13] M. Kohagen, M. Lepšík, and P. Jungwirth, J. Phys. Chem. Lett. **5**, 3964 (2014).
[14] A. Melcrová, S. Pokorna, S. Pullanchery, M. Kohagen, P. Jurkiewicz, M. Hof, P. Jungwirth, P. S. Cremer, and L. Cwiklik, Sci. Rep. **6**, 38035 (2016).
[15] A. Magarkar, P. Jurkiewicz, C. Allolio, M. Hof, and P. Jungwirth, J. Phys. Chem. Lett. **8**, 518 (2017).
[16] J. Kahlen, L. Salimi, M. Sulpizi, C. Peter, and D. Donadio, J. Phys. Chem. B **118**, 3960 (2014).
[17] I. Leontyev and A. Stuchebrukhov, Phys. Chem. Chem. Phys. **13**, 2613 (2011).
[18] I. V. Leontyev and A. A. Stuchebrukhov, J. Chem. Theory Comput. **6**, 1498 (2010).
[19] I. V. Leontyev and A. A. Stuchebrukhov, J. Chem. Theory Comput. **8**, 3207 (2012).
[20] Y. S. Badyal, A. C. Barnes, G. J. Cuello, and J. M. Simonson, J. Phys. Chem. A **108**, 11819 (2004).
[21] E. Pluhařová, O. Marsalek, B. Schmidt, and P. Jungwirth, J. Phys. Chem. Lett. **4**, 4177 (2013).
[22] H. E. Fischer, G. J. Cuello, P. Palleau, D. Feltin, A. C. Barnes, Y. S. Badyal, and J. M. Simonson, Appl. Phys. A: Mater. Sci. Process. **74**, s160 (2002).
[23] P. Mason, C. Dempsey, H. E. Fischer, J. Hladilkova, P. Jungwirth, and S. Timr, "Towards a fuller understanding of salt-bridges in proteins," Institut Laue-Langevin (ILL) Report No. 8-03-807 (2014).
[24] A. C. Barnes, S. B. Lague, P. S. Salmon, and H. E. Fischer, J. Phys.: Condens. Matter **9**, 6159 (1997).
[25] G. W. Neilson, P. E. Mason, S. Ramos, and D. Sullivan, Philos. Trans. R. Soc., A **359**, 1575 (2001).
[26] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, J. Chem. Theory Comput. **4**, 435 (2008).
[27] G. Bussi, D. Donadio, and M. Parrinello, J. Chem. Phys. **126**, 14101 (2007).
[28] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[29] S. Miyamoto and P. A. Kollman, J. Comput. Chem. **13**, 952 (1992).
[30] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[31] L. X. Dang and D. E. Smith, J. Chem. Phys. **102**, 3483 (1995).
[32] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, J. Comput. Chem. **25**, 1656 (2004).
[33] E. Pluhařová, H. E. Fischer, P. E. Mason, and P. Jungwirth, Mol. Phys. **112**, 1230 (2014).
[34] J. Wang, P. Cieplak, and P. A. Kollman, J. Comput. Chem. **21**, 1049 (2000).
[35] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, M. C. X. Li, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, GAUSSIAN 09, Gaussian, Inc., Wallingford CT, 2009.
[36] G. A. Petersson and M. A. Al-Laham, J. Chem. Phys. **94**, 6081 (1991).
[37] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Comput. Phys. Commun. **167**, 103 (2005).
[38] J. VandeVondele and J. Hutter, J. Chem. Phys. **127**, 114105 (2007).
[39] C. Lee, W. Yang, and R. G. Parr, Phys. Rev. B **37**, 785 (1988).
[40] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[41] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, J. Chem. Phys. **132**, 154104 (2010).
[42] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).
[43] M. Galib, T. T. Duignan, Y. Misteli, M. D. Baer, G. K. Schenter, J. Hutter, and C. J. Mundy, J. Chem. Phys. **146**, 244501 (2017).
[44] J. Timko, D. Bucher, and S. Kuyucak, J. Chem. Phys. **132**, 114510 (2010).
[45] J. Timko, A. De Castro, and S. Kuyucak, J. Chem. Phys. **134**, 204510 (2011).
[46] M. D. Baer and C. J. Mundy, J. Phys. Chem. B **120**, 1885 (2016).
[47] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).
[48] J. Schmidt, J. VandeVondele, I.-F. William Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy, J. Phys. Chem. B **113**, 11959 (2009).
[49] E. Pluhařová, M. D. Baer, C. J. Mundy, B. Schmidt, and P. Jungwirth, J. Phys. Chem. Lett. **5**, 2235 (2014).
[50] G. Placzek, Phys. Rev. **86**, 377 (1952).

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Guanidinium Pairing Facilitates Membrane Translocation

Christoph Allolio,[*,†,‡] Katarina Baxova,[†] Mario Vazdar,[§] and Pavel Jungwirth[†,∥]

[†]Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nám. 2, CZ-166 10 Prague 6, Czech Republic

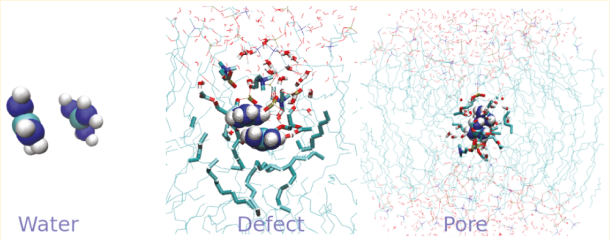
[‡]Institut für Physikalische and Theoretische Chemie, Universität Regensburg, 93040 Regensburg, Germany

[§]Institut Rudjer Bošković, Bijenička cesta 54, 10000 Zagreb, Croatia

[∥]Department of Physics, Tampere University of Technology, P.O. Box 692, FI-33101 Tampere, Finland

Supporting Information

ABSTRACT: Ab initio free energy calculations of guanidinium pairing in aqueous solution confirm the counterintuitive conjecture that the like-charge ion pair is thermodynamically stable. Transferring the guanidinium pair to the inside of a POPC lipid bilayer, like-charge ion pairing is found to occur also inside the membrane defect. It is found to contribute to the nonadditivity of ion transfer, thereby facilitating the presence of ions inside the bilayer. The effect is quantified by free energy decomposition and comparison with ammonium ions, which do not form a stable pair. The presence of two charges inside the center of the bilayer leads to the formation of a pore. Potential consequences for cell penetrating peptides and ion conduction are drawn.



INTRODUCTION

Arginine magic, i.e. the ability of many arginine-rich peptides and guanidinium-functionalized small molecules[1] to be rapidly internalized by cells, is still not a well understood phenomenon.[2] The increased usage of arginine-rich cell penetrating peptides[3] for drug delivery is in stark contrast to the lack of detailed molecular understanding of their mode of action. Part of the reason is clearly the complexity of the cells themselves and their mechanisms of active endocytosis. However, there have been numerous experimental studies, where such peptides were found to be able to cross artificial lipid bilayers or extracted cell membranes without the presence of ATP or cytoskeleton.[4−9] Therefore, a direct, passive mechanism remains an important pathway for cell penetration and, consequently, an area of active research. Several molecular dynamics (MD) studies have been performed, from coarse-grained simulations[10−12] to united or all atom MD,[13−16] in order to elucidate the details of this mechanism. So far, the results have not been fully convincing. The simulations either contained implausibly high energy barriers for penetration,[10,11,16] artifacts due to absence of counterions,[17] or the presence of counterions not available under in vivo conditions.[15] The most puzzling aspect of the direct penetration mechanism remains the question why a highly *positively* charged arginine-rich peptide should readily cross the hydrophobic membrane core. A logical place of departure for understanding how this is possible is the interaction of ions, especially charged amino acid or side-chain analogs with a model lipid bilayer.

The membrane penetration of small ions as well as amino acids has been studied extensively with MD methods.[18−24] It is a common feature of the entirety of the atomistic models and even of the better part of the less accurate coarse grained simulations that the transfer of ions toward the center of these model membranes leads to a membrane defect,[25] where water and headgroups are displaced toward the center of the bilayer. Recently it was found that there is little ion specifity involved in the formation of such a defect.[26] However, the creation of a defect for an ion pair leads to nonadditivity for ion transport. By nonadditivity we mean that once the defect is created, it is energetically less costly to introduce a second ion. This effect has notably been studied for arginine,[27] but it also was found in a coarse grained MD study of the translocation of polyarginine peptides.[11] As this nonadditivity is a consequence of defect or pore formation, the question of ion specifity in the nonadditive behavior imposes itself. Namely, like-charge ion pairing might be a powerful enhancer of nonadditivity, because it could directly contribute to the stabilization of multiple ions in a defect. While there is no definite experimental evidence for the formation of stable like-charge guanidinium pairs in water, there is a large body of literature from force-field based MD[28,29] over electronic structure calculations[30] and resolution of the identity solvation model (RISM) calculations[31] and ab initio molecular dynamics[32] to various experiments, such as X-ray[29] or neutron scattering[33] and electrophoresis experiments,[34] supporting this

Received: October 23, 2015
Revised: December 13, 2015
Published: December 16, 2015

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

notion. The aim of this study is to gather more conclusive evidence for the existence of guanidinium pairing and explore its contribution to the nonadditivity effect in membranes. Finally, we will critically examine the potential role of this nonadditivity effect for the passive translocation of cell penetrating peptides.

COMPUTATIONAL DETAILS

Force Field-Based Molecular Dynamics. *System Setup.* The large simulation boxes contained 128 POPC lipids, 1−2 GDM^+ ions, and 1−2 chloride counterions to neutralize the system. The boxes containing GDM^+ contained 6352 or 6344 H_2O, and the boxes with NH_4^+ ions contained 5716 or 5682 H_2O respectively. To construct the lipid membrane a pre-equilibrated lipid patch was used. Typical box dimensions were 6.4 × 6.4 × 8.3 nm^3. We used the SPC water[35] model, the OPLS-AA[36] force field for NH_4^+ and guanidinium (GDM^+), and the Berger united atom force field for the lipids.[37] The charges for the cations were taken from the lysine and arginine side chains. In the case of ammonium, the central atom charge was decreased by 0.02 e in order to achieve an integer charge. All bonds were constrained using the LINCS scheme.[38] The time step used in the simulations was 2 fs. Calculations were done at constant temperature and pressure (NpT ensemble) with the simulation temperature of 310 K and pressure of 1 bar. The simulations were thermostated with a CSVR thermostat[39] with a time constant of 0.5 ps. The pressure was constrained via a Parrinello Rahman barostat[40] with a time constant of 1 ps. For simulations containing the membrane we used a semi-isotropic pressure coupling, i.e., the z dimension of the box was coupled to the barostat separately from the x and y dimensions, allowing it to independently fluctuate. A short-range cutoff of 1 nm for Coulomb and Lennard−Jones interactions was employed. Long range electrostatic interactions were treated with the particle mesh Ewald method.[41] For ion pairing in aqueous solutions, smaller systems were also simulated using different force fields (for full details see Supporting Information). The simulations were performed using the GROMACS v. 4.6.3 molecular dynamics package.[42]

Free Energy Calculations. The free energy profiles were computed using the umbrella sampling method[43] along the distance between the center of mass of the ion(s) and the center of mass of the membrane, projected along the membrane normal. The membrane patch lies in the XY plane, so the normal is the z-axis. For merging of the umbrella sampling simulations we used the weighted histogram analysis method (WHAM),[44] an implementation of which[45] we modified to obtain the WHAM weighting coefficients used then in the calculation of additional properties. Umbrella potentials had a force constant of 1500 kJ mol^{-1} nm^{-2}, except for the GDM^+ pair where the force constant was relaxed to 500 kJ mol^{-1} nm^{-2} after pre-equilibration in the relevant windows in order to achieve larger overlaps between neighboring windows to sample the C−C distribution continuously. For the GDM^+ pair we also used 2000 kJ mol^{-1} nm^{-2} force constants in the very center of the bilayer, for which the geometries were generated by taking the geometry of a central membrane window and moving slowly backward after a total of more than 300 ns inside the membrane core. The guanidinium pair was restrained using a capped umbrella potential. This potential is flat until a cutoff distance is reached; then it becomes harmonic with a force constant of 1.3 × 10^5 kJ mol^{-1} nm^{-2}. The cutoff distance was chosen to be 0.6 nm to include the entire zone of stability of the GDM^+ pair in an aqueous solution. Additional simulations were performed to check for potential artifacts induced in the zone before the cutoff, but none could be found. More details are given in the Results section of the manuscript. The umbrella windows were then equilibrated for 50−100 ns each, before sampling in each case for at least 100 ns. For each free energy evaluation we used 17 or more umbrella windows and 168−390 ns of simulation time per window. Simulations inside the membrane core (z-distance < 1.2 nm from center) were equilibrated for ≥100 ns. Initial configurations for the umbrella sampling were generated by fast pulling of the ion toward the membrane center of mass, with a pulling rate of 0.5 Å ps^{-1}. Due to the use of the NpT ensemble, the free energies obtained from these simulations are Gibbs energies. We do refer to them as free energies, as the volume work is expected to be negligible. The ion pairs were also simulated without the membrane in aqueous solution, to obtain the free energy of ion pairing; for details, see the Supporting Information.

Ab Initio Molecular Dynamics. The ab initio molecular dynamics simulations are based on the Kohn−Sham density functional theory (DFT) and use the BLYP[46,47] functional with the Grimme D3[48] dispersion corrections. The GPW[49] method was used to model the Kohn−Sham orbitals. A local TZVP basis set with GTH pseudopotentials,[50] and a 280 Ry plane-wave cutoff for the density were employed. The D3 corrections were used without the three-body and C9 terms; this approximation was previously found to lead to a negligible loss of accuracy in the case of water.[51] As D3 is known to reduce known DFT problems (most prominently the over-structuring of water still present in the previous version of the Grimme corrections),[51,52] the simulation was thermostated to the physiological temperature of 310 K using a CSVR[39] thermostat with a 100 fs time constant. We used a time step of 0.5 fs and reduced the isotope masses of C, N, and O to 4 au to increase guanidinium diffusion. The mass of H was increased by 50% to prevent integration errors. In our setup, the structure only depends on the relative configurational energies, which are independent of the momenta; nevertheless, classical MD simulations were performed with both adjusted and correct nuclear masses to verify that structural properties were indeed not affected. using scaled OPLS-AA charges and SPC/E[54] water, 2 GDM^+ and 148 H_2O were pre-equilibrated in CP2K[53] for 1 ns. The GDM^+ C atoms were constrained to move along the z-axis at x,y = 7 Å. This constraint allowed us to increase the box size in one dimension toward the dissociation limit of the guanidinium pair at a reasonable computational cost. Accordingly, the box size was 14 × 14 × 23.6 $Å^3$, as determined from NpT simulation with GROMACS at 310 K. The free energy calculation was started from 15 geometries and velocities extracted from this trajectory. The geometries were placed inside regular umbrella windows at 0.6 Å intervals, beginning at 2.3 Å consisiting of 0.75 Å of a flat potential bounded by upper and lower harmonic walls of a 0.1 Ha $Bohr^{-2}$ force constant. The simulation was run inside the umbrella windows for 60 ps (up to 100 ps in the umbrella windows which explore the minima). The free energy was computed using an adapted version of the WHAM program.[55]

RESULTS AND DISCUSSION

Guanidinium Pairing in Aqueous Solution. *Force Field vs ab Initio Results.* Ion pairing in water is an active area of computational research[56,57] which can be prone to artifacts to

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

due inconsistencies in force fields. Ab initio MD is expected to give an accurate description of ion pairing, as the method is known to give quantitative results for the structure of the first solvation shell.[58] Specifically, the BLYP functional with D3 dispersion corrections has been proven to give good interaction energies of GDM^+ ions in water clusters[30] and is known to yield a rather accurate structure of liquid water in comparison to other nonhybrid DFT functionals.[51,52] Like-charge pairing of guanidinium ions was already observed in a previous ab initio MD study[32] of a 1.64 M aqueous GDM^+Cl^- solution. However, the small box size of 12.64 Å, a relatively high concentration, and short simulation time did not allow reaching a reliable answer to the question of whether the like-charge ion pair is thermodynamically stable. Our simulation setup was designed to explore the GDM^+-GDM^+ dissociation curve at low concentrations. Stabilizing counterions were omitted because their distribution at low concentration cannot possibly be equilibrated within the subnanosecond time scale accessible by ab initio MD. Moreover, they should not significantly influence the like-charge pairing at low concentrations. We computed error estimates by exactly rebuilding the ab initio MD setup, but using the force field employed for pre-equilibration to collect statistics. This allowed us to obtain enough uncorrelated data to obtain an error estimate for the amount of ab initio sampling done. Our ab initio MD study thus represents the most rigorous computational test of guanidinium like-charge ion pairing so far.

In Figure 1 we compare the ab initio free energy profile of the guanidinium pair to force-field-based MD. Thermodynamically stable guanidinium pairing is present in both models, and the depths of the minima are comparable. There are, however, some subtle differences in their location and in the shape of the profiles. The ab initio curve shows a stable zone beginning at 0.35 nm, close to the mininum found in most force field descriptions (see Supporting Information). This minimum corresponds to the stacked conformation typically associated with guanidinium ion pairing. At around 0.46 nm we find a relatively deep mininum (Figure 1). This minimum is due to a T-shaped conformation of the guanidinium pair. The T-shaped conformation was also observed in the previous ab initio study during the pre-equilibration phase.[32] The transition from the T-shaped to the stacked conformation requires a change in orientation of a guanidinium ion. Even though we run the simulation for 100 ps per window in this region, it is not possible to weigh the T-shaped conformation fully quantitatively against the stacked conformation, because the exchange events are very slow and both conformations are possible inside the binding region of the GDM^+-GDM^+ contact ion pair. Also, there is a minimum at 0.7 nm representing the formation of the solvent shared ion pair; see Figure 1.

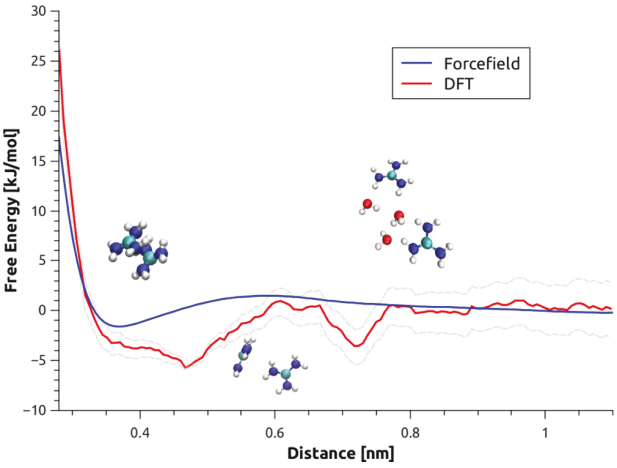


Figure 1. Free energy of the guanidinium pair as a function of central atom distance, in force field-based (OPLS-AA) and ab initio (BLYP-D3) simulation. The given free energies are volume-entropy corrected. Pictures are snapshots from the trajectories and represent the three interchanging forms of ion pairing. The gray lines represent the standard deviations expected for a trajectory of this length, as estimated from an equivalent force field setup. The measurement of the error is referenced to the energy minimum at 0.46 nm.

These details on the free energy curve are absent in the OPLS-AA description. Other force fields, such as CHARMM, exhibit better agreement with the ab initio data in this respect (see Supporting Information); however, the differences are not large and the force field methods generally give reasonable results. The ab initio data point toward a somewhat deeper minimum for the guanidinium ion pairing, in agreement with values given for example from CHARMM[29] simulation or from RISM.[31] They also show the surprising stability of the T-shaped conformation. Nevertheless, the onset of contact pair stabilization and the position of the dissociation transition state of the OPLS-AA force field are in good agreement with the ab initio method. We, therefore, believe our OPLS-AA free energy can serve as a reasonable (lower end) estimate of guanidinium like-charge ion pairing. Using the OPLS-AA force field furthermore allows comparability with our previous study of polyarginine agglomeration at a POPC bilayer.[13]

Guanidinium Pairing vs Ammonium Pairing. In the previous section, we have established guanidinium pairing in dilute aqueous solutions. In order to quantify the ion specificity arising from this like-charge ion pair a comparison must be established with analoguous pairing for more "regular" cations. For this we chose the ammonium ion, which is derived from the lysine side chain in the same way as guanidinium from arginine. In contrast to guanidinium, ammonium it is not expected to form any stable ion pairs. In order to test this, we computed the ion pairing free energy profile for the ammonium pair using the OPLS-AA force field. Comparison to the GDM^+-GDM^+ pair is given in Figure 2.

The stable minimum for the GDM^+-GDM^+ pairing in aqueous solution is shallow, with a depth of about 2 kJ/mol. In contrast to this, the NH_4^+ pairing free energy increases nearly monotonously with shortening of the distance; see Figure 2. In the region of GDM^+ pairing around 4 Å it is already about +5 kJ/mol, so that in aqueous solution the energy difference between an "ammonium pair" at this distance vs the guanidinium pair at its minimum is about 7 kJ/mol. There is an extremely shallow local minimum for the ammonium contact ion pair around 0.47 nm, but it is not thermodynamically stable. The position of this local minimum is in good agreement with results from our previous ab initio study.[32] From minimum to minimum the difference between ammonium and guanidinium pairs is about 4 kJ/mol.

Free Energy Cycle Including a Lipid Bilayer. To fully characterize the free energy of ion pairing inside a lipid bilayer, it is necessary to know the pairing free energy at each depth inside the membrane. The ion pairing free energy for two guanidinium ions far away from the membrane is already known from the aqueous solution simulations presented above.

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

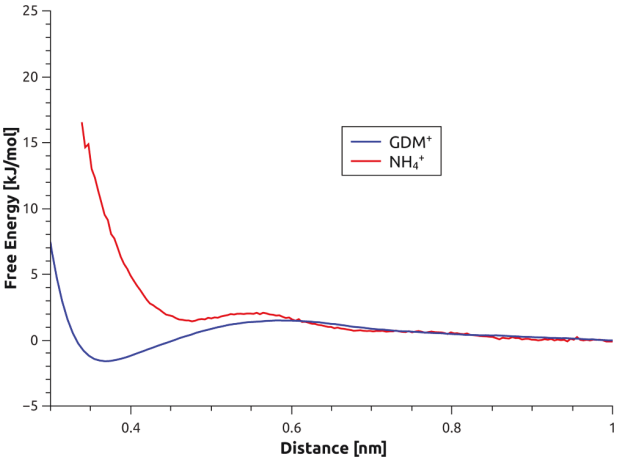


Figure 2. Like-charge ion paring: Guanidinium vs ammonium free energy as a function of central atom distance. The presented free energies are volume-entropy corrected and referenced to zero at 1 nm.

However, as the ions proceed to enter the membrane, the ion-pairing free energy becomes dependent on their depth inside the membrane. Instead of computing a full 2D free energy profile for the dissociation along the membrane normal, we construct a free energy cycle to determine the stabilization of pairs of ions inside the membrane; see Figure 3. Using this cycle we compute the ion pair stabilization by comparison with infinitely separated ions, which are simulated separately.

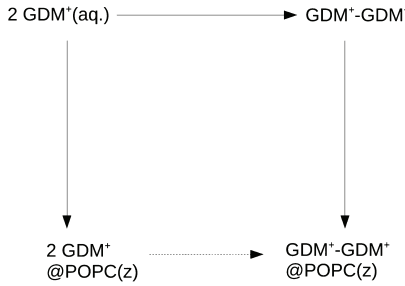


Figure 3. Schematic of the free energy cycle for guanidinium pairs inside the lipid bilayer. The free energy cycle can be constructed for any POPC membrane depth z.

In the case of NH_4^+, no stable ion pair is expected; therefore, the ions had to be constrained to a certain distance (0.4 nm). This is an arbitrary definition of an ammonium ion pair, at a distance close to the minimum of the guanidinium pair. However, in the case of guanidinium ions, it is desirable to know if the genuine minimum for ion pairing in water persists in the bilayer and, if so, if it is shifted to a different equilibrium distance as the solvation structure and dielectric permittivity in the bilayer change. As the guanidinium pairing free energy is very low and the corresponding free energy coordinate is rather flat, we performed a limited two-dimensional umbrella sampling with a biasing potential.

$$V(q) = 0;\ q < q_{\mathrm{cut}} \tag{1}$$

$$V(q) = c(q - q_{\mathrm{cut}})^2;\ q > q_{\mathrm{cut}} \tag{2}$$

The generalized coordinate q is chosen to be the distance between the two GDM^+ carbon atoms, and the value of q_{cut} is 0.6 nm. The biased probability in a one-dimensional umbrella sampling simulation is defined as

$$\tilde{P}(q, s_k) = \exp[\beta A_k] \int d^N\mathbf{r} \exp[-\beta U(\mathbf{r}) - \beta W(f_1(\mathbf{r}), s_k)]\delta(f_1(\mathbf{r}) - q) \tag{3}$$

with the generalized coordinate q, the umbrella window position s_k, and the umbrella potential W. The probability is normalized with the configurational partition function of the biased system $\exp[-\beta A_k]$. The function f_1 maps the Cartesian coordinates on the generalized coordinate q, i.e., the interionic distance. In the case of a flat potential, in a zone of $q < q_{\mathrm{cut}}$ the bias potential $W(f_1(\mathbf{r}),s_k) = 0$. The unbiased probability, therefore, reads as

$$P_k(q) = \exp[\beta A_0 - \beta A_k]\tilde{P}(q, s_k) = \alpha\tilde{P}(q, s_k) \tag{4}$$

This means that inside the cutoff zone the unbiased probability is simply a multiple of the biased probability. This translates to an offset in the free energy.

$$A_k(q) = -\beta^{-1} \log [\alpha\tilde{P}(q, s_k)] = A_k - A_0 - \beta^{-1} \log \tilde{P}(q, s_k) \tag{5}$$

This makes the evaluation highly convenient, as no further unbiasing in this coordinate is necessary and only an energy offset has to be added to the restrained potential free energy profile. Furthermore, it allows computation of the free energy profile along the normal to the bilayer with a standard 1-d WHAM. All other structural observables can then easily be obtained from the WHAM along the membrane normal, by weighing with the histogram coefficients.[44]

Single vs Multiple Ions: The Nonadditivity Effect. *Guanidinium vs Ammonium.* The free energy profiles of single ions and ion pairs along the z-coordinate perpendicular to the bilayer are given in Figure 4, top. The paired guanidinium and ammonium profiles are shifted from the baseline by the ion pair formation free energy of the ions in the aqueous solution. This leads to a total shift of the pairs of $\Delta\Delta E_{\mathrm{pair}} = -7$ kJ/mol in favor of guanidinium over ammonium pair formation. We observe a lower transition energy of the guanidinium pair from water into the bilayer compared to the ammonium pair. The difference in ion pairing free energies constitutes a major part of this effect. Consequently, the transition free energies of the unpaired ions do not differ as much. The ions and ion pairs exhibit a different adsorption to the membrane headgroup region. Due to the delocalized positive charge and the possibility to form more hydrogen bonds to phosphate moieties, the stronger adsorption of the guanidinium ions to the phospholipid headgroup is not surprising, and the differences to ammonium correspond quite well to values found for various side-chain analogues in the literature.[19−21] To exemplify the effect of adsorption, we referenced the free energy curves to their respective minima, as shown in Figure 4, bottom. It becomes clear that the only statistically significant differences between the guanidinium and ammonium pairs are in the adsorption and in the pairing energy, in the aqueous phase. We do not see any ion specific effect in the energy profile when moving from the headgroups to the center of the bilayer. For the single guanidinium or ammonium ions the situation is identical to that of the pairs with the main difference in the headgroup adsorption free energy.

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

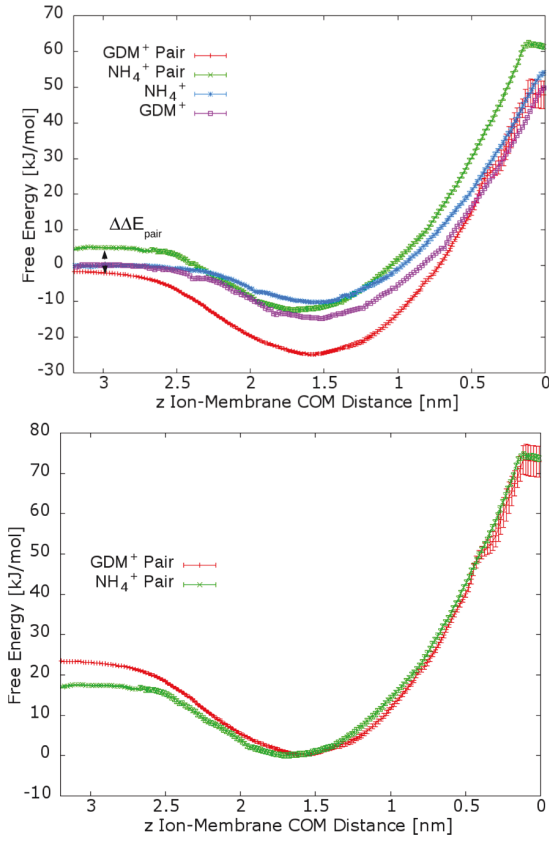


Figure 4. Upper panel: Free energy profiles for the ions and ion pairs as a function of the distance to the POPC bilayer COM, referenced to the free ions in solution, and error bars are standard deviations obtained from bootstrapping of complete histograms[45] across the GDM^+−GDM^+ profile. Lower panel: The same profiles for the ion pairs referenced to their respective minima.

Nonadditivity in Crossing the Membrane. The ion pairs exhibit very pronounced nonadditivity; i.e., the free energy of pulling an ion pair inside the membrane core is drastically lower than twice the free energy of pulling a single ion to the same position. Nonadditivity of the behavior of ions inside the bilayer is a known phenomenon,[27] the reason for it being visible in Figure 5. Namely, the ions forming a pair occupy a common defect inside the bilayer. This defect has to be formed only once and can host the two ions with little additional energy cost compared to the energy for single ion. A major contribution to the formation of the defect is most likely the high free energy cost of desolvating an electric charge. However, it is not a kinetic deformation effect caused by pulling the ion inside the bilayer. We confimed this by placing the ion pair directly inside the core of a pre-equilibrated membrane, which quickly lead to the formation of a hydrophilic defect as seen in the picture. When an ion pair is transported, the free energy becomes flat in the very center of the bilayer. This behavior is caused by the formation of a pore. In order to observe the pore formation, it was necessary to constrain the ion pair to the very center of the membrane and relax the system for up to 300 ns with an elevated force constant (2000 kJ mol^{-1} nm^{-2} in the case of the GDM^+ pair). In the presence of a pore, the energy to transfer one GDM^+ ion becomes practically indistinguishable from the energy to transfer two of them. We have tabulated the transfer energies of ions to the center of the lipid bilayer both from the center of the bilayer and from their preferred mininum in Table 1.

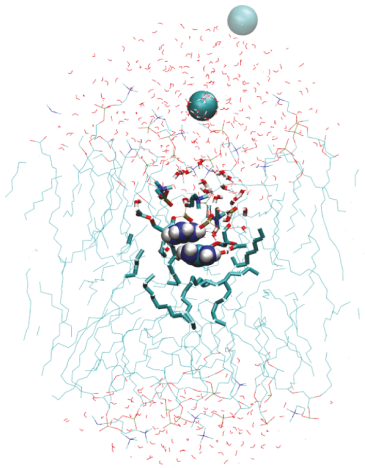


Figure 5. A guanidinium pair inside the POPC bilayer between headgroups and the membrane core. Chloride counterions are depicted as cyan balls.

Table 1. Ion Transport Energies from the Aqueous Bulk to the Membrane Center in kcal/mol

ine ion	ΔG_{center}
NH_4^+	13
GDM^+	12
GDM^+−GDM^+	12
NH_4^+−NH_4^+	15

Nonadditivity and Conductance. The experimental free energy of ion transport across a DPhPC (Ph stands for the branched fatty acid phytanoyl) bilayer has been estimated to be 18 kcal/mol[26] for GDM^+, as well as K^+, determined from conductance experiments at 0.1 M concentration. The authors compared the experimental results to the free energy of single-ion transport across a DPPC bilayer using the CHARMM force field. They found good agreement of the computed free energy of about 22 kcal/mol with a diffusion barrier-corrected experimental value of 20 kcal/mol. (See ref 26 for details.) Hence, ion transfer energies were not found to be ion specific. The energies reported in Table 1 are lower than the reported experimental value due to the use of a POPC bilayer rather than a DPhPC membrane. The difference in lipid composition lowers the (Lysine) transfer energy by about 4 kcal/mol.[20] We acknowledge possible issues with our force field (OPLS-AA/Berger) description of the system, namely the likely overestimation of the ion binding to the membrane. Nevertheless, ion adsorption at the headgroup region was found for both side chain analogues in a recent benchmark study of force fields[19] and the obtained values of free energy differences agree well with the results presented here. Moreover, numerical agreement with simulations of side chains/amino acids crossing POPC is satisfactory.[19,20,27] Since the transferred ion will most likely be initially bound at the bilayer, it may also be appropriate to reference the transition free energy to the global minimum of the free energy profile. The need to desorb the ions increases the transfer barrier to 15−16 kcal/mol, and differences between GDM^+ and NH_4^+ ions then become even smaller. The available conductivity data for GDM^+ do not show

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

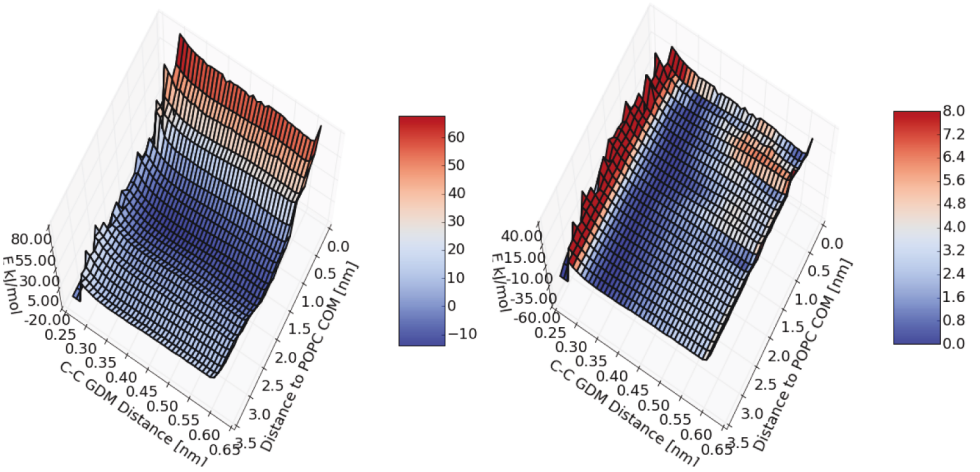


Figure 6. Left: 2D free energy profile of the guanidinium pair. Right: Entropy corrected GDM^+−GDM^+ 2D free energy profile, referenced to the minimum in the C−C distance at a given z.

any increase over other, nonpairing ions even at 1 M concentration.[26] It is, therefore, helpful to clarify whether this is in contradiction with our theoretical result of higher nonadditivity for the guanidinium pair. We will now show that differences in the transition energies of the GDM^+ ion pair to the NH_4^+ pair have only a small influence on the conductance, even though the GDM^+ pair has a comparable transition barrier to the single GDM^+ ion. Nonadditivity in a transition barrier ($\Delta G^{\ddagger} > 0$) has different implications from cooperativity in an exergonic ($\Delta G < 0$) process. Cooperativity in ion adsorption, for example, means that the adsorption free energy of several ions at an interface has a larger negative value than the sum of the single ion adsorption free energies. The expected consequence is agglomeration of ions at the interface. In the case of a positive energy barrier, nonadditivity means that the multi-ion transition barrier $\Delta G^{\ddagger\prime}$ has a lower free energy than the sum of n single ion transition barriers $n\Delta G^{\ddagger}$. The transition rates correspond to the experimentally measured conductance. A rough estimate of the ion transition rate ν can be given by the Arrhenius equation:

$$\nu = Bn \exp[-\beta \Delta G^{\ddagger}] \tag{6}$$

where the pre-exponential factor B is an unknown proportionality constant, n is the number of ions in the transition state, and $\Delta G^{\ddagger}$ is the transition barrier. We now look for a condition for the multi-ion rate ν' to be larger than the single ion rate ν by simply substituting the Arrhenius relation for the rates. Assuming an equal proportionality constant B, the rate for a barrier with n' ions over a barrier with n ions is then only increased if the following inequality holds for the difference in transition free energies of single and multi-ion states $\Delta\Delta G^{\ddagger}$ ($= \Delta G^{\ddagger\prime} - \Delta G^{\ddagger}$):

$$\beta^{-1}(\log n' - \log n) > \Delta\Delta G^{\ddagger} \tag{7}$$

To show that nonadditivity is not relevant to rates even for equal transition barriers for an ion and ion pair, it suffices to show the left-hand side of this equation to be small. In our case the left-hand side is smaller than 2 kJ/mol and the right-hand side is ∼0, thus still indicating a small positive effect on the rate due to ion pairing. However, the value is very much within the error bar for $\Delta\Delta G^{\ddagger}$ in the simulation and smaller than the error of the method. In fact the left-hand side is even smaller than the volume entropy penalty of forming the pair at low concentrations. Increasing the number of ions in the transition state will further reduce the entropy of the transition state and most likely also lower the prefactor B (for further elaboration see the Supporting Information). This indicates that at low to intermediate concentrations nonadditivity has a negligible impact on transition rates, unless $\Delta\Delta G^{\ddagger}$ is negative. This would be the case if the multi-ion barrier is smaller than the single ion barrier corresponding to true cooperativity in the sense of the previous example. Note that these considerations are very general; they are valid when inserting ions into the pore in sequence and also apply for any multipeptide translocation mechanism from the aqueous phase.

When an ion crosses a membrane, this is likely due to a temporary pore, as has been pointed out by several authors.[22,59,60] Our simulations can be interpreted as showing the ease of ion transport through such a temporary pore. The accompanying nonadditivity is present for both examined ions and is fairly unspecific. The ions sharing the temporary pore do not necessarily need to be paired; several unpaired ions may actually pass before the temporary pore closes. These reflections suggest that a conductivity study will not show a strong effect due to ion pairing, at least as long as the concentrations are low and the ion pairing free energy is in the few kJ/mol range as is true in this case. Summarizing, our free energy profiles show lower transition barriers for guanidinium pairs compared to ammonium pairs and very strong nonadditivity when paired ions are transported across a bilayer.

Ion Pairing through the Bilayer. From the observation of nonadditivity we know that an ion pair is more stable inside a defect in a membrane than two separate ions each in their own defect. However, we do not know yet if inside the membrane defect there still exists a minimum for the ion pair, even though this is of critical importance for ion pairing in guanidinium-rich molecules within a direct translocation mechanism. An absence of a minimum for the ion pair would not be surprising, when considering the putatively lower dielectric permittivity inside the pore. The central role of the permittivity effect for like-charge ion pairing can easily be estimated from calculations involving a polarizable continuum model (PCM), where a continuum dielectric with $\epsilon_r = 80$ is sufficient to create an energy minimum.[28] As mentioned in the previous sections, the

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

probability distribution from the WHAM procedure can be binned as a function of the GDM^+-GDM^+ C−C distance q. The resulting 2D free energy plot is given in Figure 6. On the left-hand side of the figure the free energy energy profile along both coordinates is given as referenced to the bulk aqeuoeus solution. As expected, the energy distribution is dominated by the membrane normal coordinate and all except for very short GDM^+-GDM^+ distances remain easily energetically accessible. This not only confirms the validity of our choice of a flat umbrella potential but also makes it more difficult to monitor guanidinium pairing throughout the membrane. In order to better observe the effect of the bilayer on guanidinium pairing, we show only the C−C dependent part of the free energy along the z-axis. (see Figure 6, right). This was achieved by setting the energy of the local minimum of the GDM^+-GDM^+ free energy in each z-coordinate to zero.

We observe the presence of a free energy minimum for the GDM^+-GDM^+ pair throughthout the membrane. More interestingly, when proceeding to the membrane core, the minimum becomes actually deeper until a pore forms. The deepening of the minimum for the guanidinium pair starts in the headgroup region and becomes rather strong just before the pore formation threshold. To better understand the nature of ion−membrane interactions when leaving the headgroup region, we explored the $NH_4^+-NH_4^+$ N−N pairing free energy surface around this zone in the same way as we did it for the guanidinium pair. For both ions, we see an offset of about 2 kJ/mol for the barrier of pair dissociation compared with the bulk water (see Figure 7). This means that the ion pairing seems to

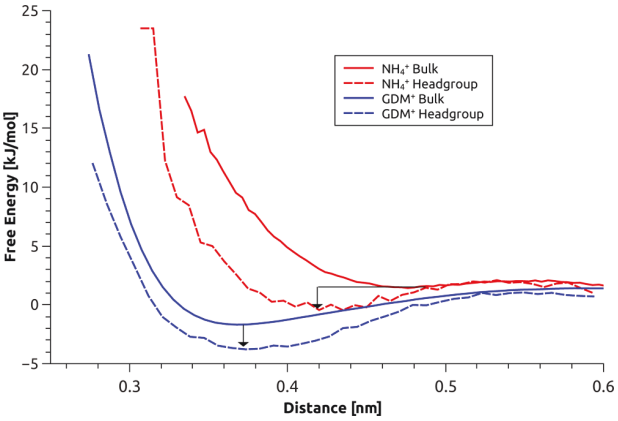


Figure 7. Comparison of GDM^+ and NH_4^+ ion pairing minima in the aqeuoeus bulk vs in the membrane (at ca. 1.5 nm COM depth).

be enhanced. We also observe a shift to a lower N−N distance minimum for the NH_4^+ pairing. This minimum is around 0.4 nm and justifies our choice of this offset when comparing transfer free energies. We have verified that this is also the location of the $NH_4^+-NH_4^+$ free energy minimum at the center of the bilayer. Because of the similar offsets for the two pairs, the effect does not seem to be ion specific. Anything else would be surprising, as the total free energy surface in the z-coordinate perpendicular to the bilayer does not show ion specificity. It is, therefore, highly unlikely that there is an increase in ion pairing in the sense of an enhanced interactions of the ions with each other. Instead, the effect must be caused by the confinement in the membrane defect.

Interpreting the two-dimensional free energy profiles of the ion pairs with respect to the 1-D free energy profiles of the unpaired ions is not as straightforward as it first may appear: When comparing the membrane normal position of the ion pair to the single ions, it is not clear in which position each ion of the pair is on the membrane normal. This missing information can be supplied through the angle θ between ion pair connector and the membrane normal in addition to the center of mass distance. We have plotted the free energy as a function of this angle and the center of mass z-coordinate in Figure 8. Inside the defect, the C−C axis has a strong energetic preference to be aligned with the membrane normal. The preference for the membrane normal is absent before entering the headgroup region and seemingly reversed inside the pore. The reason for this is shown in Figure 8, on the left-hand side: The membrane defect (before pore creation) envelops the ions from all sides, but one. Under the condition of a constant center of mass, increasing the ion−ion distance, therefore, also pushes one of the ions deeper into the defect, effectively enlarging it. Rotating the ion pair has the same effect, creating an impression of reinforced ion pairing. The apparent reinforcement of ion pairing is, therefore, due to the choice of this coordinate and it should be interpreted as resistance to defect enlargement by the membrane. However, the systematic nature of the effect and the presence of the free energy minimum means that ion pairs exhibit stability in the membrane which goes beyond the effect of nonadditivity. In the headgroup region, the free energy profile along the membrane normal shows that the ions at the adsorption minimum in the headgroup region are more stable when being separated ions. This global minimum position on the membrane normal does not coincide with a free energy minimum in the θ coordinate at high angles. Therefore, optimum adsorption at the membrane requires larger interionic distances.

When the membrane defect covers the entire depth of the membrane and allows for a free flow of water, it turns into a pore. As the pore is formed, part of the confinement effect disappears; increasing the interionic distance along the z-axis now simply means increasing the ion distance from the defect center. Therefore, the ion pairing free energy minimum is expected to have lower stability, as is indeed observed. More surprisingly, the alignment of the ions inside the bilayer pore shows no preference for the z-axis. Instead, the GDM^+ ions seemingly prefer to bind to the pore wall, thus weakening the pair. Once a pore is created, it is large enough to allow the ion pair to rotate inside it. In contrast, the ammonium ions still have a slight reinforcement of pairing.

Membrane−Ion Interactions. Until now, we have discussed the membrane defect due to the presence of ions, as if it was simply a cavity inside a hydrophobic phase, such as, e.g., in decane. This ignores the specific interactions of the ions with the bilayer, such as hydrogen bonding to phosphates and carbonyls. In Figure 9, the average contacts of the guanidinium ions with their environment are plotted across the membrane bilayer. The hydrogen bonds to water are gradually replaced by hydrogen bonds to carbonyls and phosphates. The total amount of contacts is the highest in the headgroup region, coinciding with the total free energy minimum. The ions are never fully stripped of water, and the number of water contacts is comparable for single ions and ion pairs. The composition of the contacts inside the membrane core is different when transporting two ions, compared to having only one ion inside the membrane defect. In the case of a single ion the contacts

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

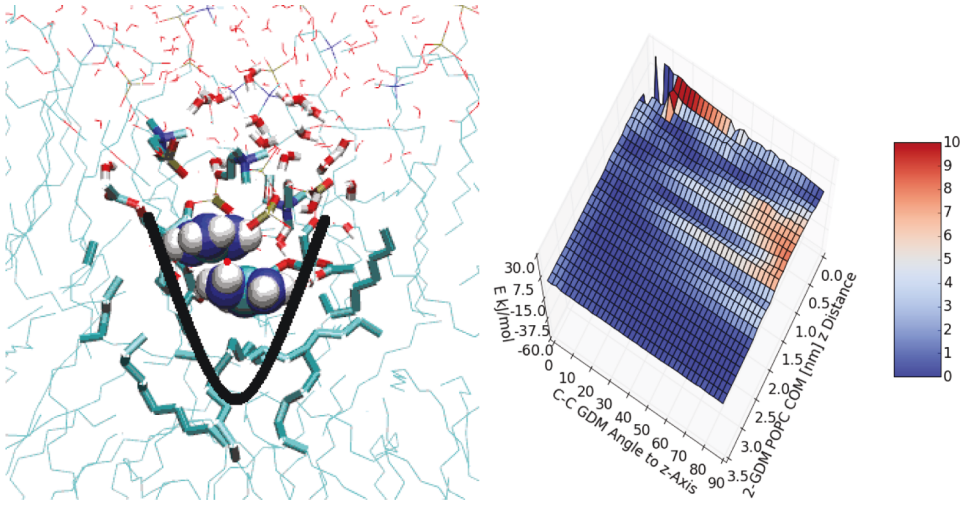


Figure 8. Left: Membrane defect enveloping the GDM^+ pair. Right: Free energy of the membrane/GDM^+ pair angle θ; the free energy is cone corrected and referenced to the energy minimum along the z-coordinate.

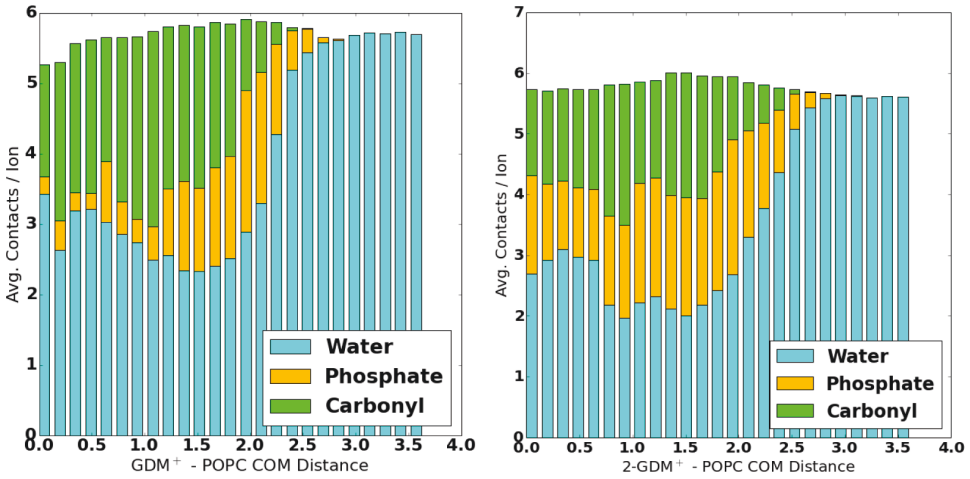


Figure 9. Left: Decomposition of solvation contacts for a single GDM^+ ion across the POPC bilayer. Right: Same decomposition of contacts for a GDM^+ ion pair.

near the center of the defect are dominated by carbonyl groups, whereas for the ion pair there is a much larger share of phosphate binding, especially near the bilayer center.

The bending down of the phosphate headgroups is usually sufficient to open a pore.[25] In fact the energy to bring two ions from their adsorption minimum to the membrane center (75 kJ/mol) is comparable with the transition barrier of transporting a lipid from one leaflet to the other. This process is known as the lipid flip-flop, and its barrier has been reported to be 80 kJ/mol[61] for POPC in a similar setup. In contrast to the carbonyl group, the phosphate bears a full negative charge. The plot shows that the single ions hydrogen bonding can be saturated using only the carbonyl groups. The reason for triggering phosphate flip-flop is, therefore, likely to lie in the higher total positive charge of the ion pair. While it is difficult to create a total energy decomposition of the process, judging from the first-shell contacts, the ions appear to remain well solvated. This also partially explains why the minimum for ion pairing persists throughthout the bilayer despite the lower permittivity. In order to achieve the solvation of ions the membrane has to undergo defect formation. The major energetic contributions to defect formation are likely the desolvation of water molecules and the mechanical stress on the lipid chains, in short the surface tension of the membrane–water interface.

Impact on the Direct Poration Mechanism. The results presented here show the clear effect of GDM^+ like-charge ion pairing on the free energy of transporting ions into the membrane, especially when compared to an "ordinary" cation such as NH_4^+. While discussing the relation of the simulation results to conductance experiments, we have pointed out the caveats of nonadditivity and the concentration of ions. However, for a cell penetrating peptide such as R_9 these concerns do not apply as the nine side chain arginine GDM^+ groups cannot diffuse far away from each other. Following our observation of the coupling of ion pairing and defect size, a cationic, hydrophilic peptide crossing the membrane should have a lower free energy of penetration if the size of the related hydrophilic defect is reduced. In other words, ion pairing allows the charged peptide to compress inside the pore, thus minimizing the deformation penalty. Guanidinium pairing can assist in this task, and extrapolating from our results for a

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143–153

noncyclic R_9 peptide the free energy could be reduced by up to 49 kJ/mol (7 pairs at 7 kJ/mol per pair; arginine cannot form a GDM^+−GDM^+ pair with the next amino acid in the sequence, but has to skip one neighbor[30]) with respect to K_9 by pairing alone. The effect should be at least as important during the prepore stage of the peptide in the membrane, as a higher charge concentration in the defect will lead to quicker pore formation. In addition to this ion pairing effect there are important diffences in the adsorption of the peptides to the bilayer.[62] Guanidinium pairing, as well as adsorption at the bilayer, might, therefore, well be at the core of *arginine magic*. However, it has to be mentioned that so far no satisfyingly low energy barrier for entry has been found in a reliable simulation of polyarginines. There is a simple condition for such a barrier: it has to be lower than that for lipid flip-flop. The flip-flop is a slow process, and as it creates a temporary pore, it should allow for the passage of materials. If the barrier of entry is higher than this value, which is typically around 80−100 kJ/mol, the natural leakage, e.g., of vesicles will be faster than peptide translocation, rendering the notion of a cell-penetrating peptide obsolete. The contribution of GDM^+ pairing inside the defect might be decisive, but the viability of R_9 passive translocation through a pore as well as its exact mechanism remain to be fully confirmed.

CONCLUSION

We have quantified the effect of like-charge ion pairing on the transport across a lipid bilayer via a free energy molecular dynamics study. The transport of ion pairs is nonadditive and leads to the creation of a temporary membrane pore. The differences between ions arise from the energy necessary to create a like-charge ion pair and the differences in ion adsorption. Using ab initio molecular dynamics we have confirmed the existence of stable guanidinium ion pairs. Free energy calculations confirmed that a free energy minimum for guanidinium ion pairing exists throughtout the bilayer. We find that ion pairing in water is a good predictor for ion pairing inside the lipid bilayer, as in the presence of an ion pair there is little ion specificity in the formation of the defect and the pore. This is a possible important factor in the more effective translocation of arginine-rich cell-penetrating peptides as compared to lysine-based peptides.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b10404.

More computational details, a free energy comparison of guanidinium force fields, an estimate of systematical biases in the AIMD setup, and some further technical discussion (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: christoph.allolio@marge.uochb.cas.cz. Phone: (+420) 220 410 316.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

P.J. thanks the Czech Science Foundation for support via Grant No. GA13-06181S. He also acknowledges the Academy of Sciences for the Praemium Academie award and the Academy of Finland for the FiDiPro award. C.A. thanks the German Academic Exchange Service (DAAD) for support via a P.R.I.M.E. fellowship and Dominik Horinek for computer time and valuable discussions.

REFERENCES

(1) Wexselblatt, E.; Esko, J. D.; Tor, Y. On Guanidinium and Cellular Uptake. *J. Org. Chem.* **2014**, *79*, 6766−6774.
(2) Bechara, C.; Sagan, S. Cell-Penetrating Peptides: 20 Years Later, Where do we Stand? *FEBS Lett.* **2013**, *587*, 1693−1702.
(3) Milletti, F. Cell-Penetrating Peptides: Classes, Origin, and Current Landscape. *Drug Discovery Today* **2012**, *17*, 850−860.
(4) Mishra, A.; Lai, G. H.; Schmidt, N. W.; Sun, V. Z.; Rodriguez, A. R.; Tong, R.; Tang, L.; Cheng, J.; Deming, T. J.; Kamei, D. T.; et al. Translocation of HIV TAT Peptide and Analogues Induced by Multiplexed Membrane and Cytoskeletal Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16883−16888.
(5) Lamazière, A.; Burlina, F.; Wolf, C.; Chassaing, G.; Trugnan, G.; Ayala-Sanmartin, J. Non-Metabolic Membrane Tubulation and Permeability Induced by Bioactive Peptides. *PLoS One* **2007**, *2*, e201.
(6) Pae, J.; Säälik, P.; Liivamägi, L.; Lubenets, D.; Arukuusk, P.; Langel, U; Pooga, M. Translocation of Cell-Penetrating Peptides Across the Plasma Membrane is Controlled by Cholesterol and Microenvironment Created by Membranous Proteins. *J. Controlled Release* **2014**, *192*, 103−113.
(7) Katayama, S.; Nakase, I.; Yano, Y.; Murayama, T.; Nakata, Y.; Matsuzaki, K.; Futaki, S. Effects of Pyrenebutyrate on the Translocation of Arginine-rich Cell-Penetrating Peptides through Artificial Membranes: Recruiting Peptides to the Membranes, Dissipating Liquid-Ordered Phases, and Inducing Curvature. *Biochim. Biophys. Acta, Biomembr.* **2013**, *1828*, 2134−2142.
(8) Rydberg, H. A.; Matson, M.; Åmand, H. L.; Esbjörner, E. K.; Nordén, B. Effects of Tryptophan Content and Backbone Spacing on the Uptake Efficiency of Cell-Penetrating Peptides. *Biochemistry* **2012**, *51*, 5531−5539.
(9) Walrant, A.; Matheron, L.; Cribier, S.; Chaignepain, S.; Jobin, M.-L.; Sagan, S.; Alves, I. D. Direct Translocation of Cell-Penetrating Peptides in Liposomes: A Combined Mass Spectrometry Quantification and Fluorescence Detection Study. *Anal. Biochem.* **2013**, *438*, 1−10.
(10) Hu, Y.; Sinha, S. K.; Patel, S. Investigating Hydrophilic Pores in Model Lipid Bilayers Using Molecular Simulations: Correlating Bilayer Properties with Pore-Formation Thermodynamics. *Langmuir* **2015**, *31*, 6615−6631.
(11) Hu, Y.; Liu, X.; Sinha, S. K.; Patel, S. 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* **2014**, *118*, 2670−2682.
(12) Wu, Z.; Cui, Q.; Yethiraj, A. Why Do Arginine and Lysine Organize Lipids Differently? Insights from Coarse-Grained and Atomistic Simulations. *J. Phys. Chem. B* **2013**, *117*, 12145−12156.
(13) Vazdar, M.; Wernersson, E.; Khabiri, M.; Cwiklik, L.; Jurkiewicz, P.; Hof, M.; Mann, E.; Kolusheva, S.; Jelinek, R.; Jungwirth, P. Aggregation of Oligoarginines at Phospholipid Membranes: Molecular Dynamics Simulations, Time-Dependent Fluorescence Shift, and Biomimetic Colorimetric Assays. *J. Phys. Chem. B* **2013**, *117*, 11530−11540.
(14) Huang, K.; García, A. Free Energy of Translocating an Arginine-Rich Cell-Penetrating Peptide across a Lipid Bilayer Suggests Pore Formation. *Biophys. J.* **2013**, *104*, 412−420.
(15) Herce, H. D.; Garcia, A. E.; Cardoso, M. C. Fundamental Molecular Mechanism for the Cellular Uptake of Guanidinium-Rich Molecules. *J. Am. Chem. Soc.* **2014**, *136*, 17459−17467.
(16) Sun, D.; Forsman, J.; Lund, M.; Woodward, C. E. Effect of Arginine-rich Cell Penetrating Peptides on Membrane Pore Formation

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

and Life-Times: A Molecular Simulation Study. *Phys. Chem. Chem. Phys.* **2014**, *16*, 20785−20795.
(17) Herce, H. D.; Garcia, A. E. Molecular Dynamics Simulations Suggest a Mechanism for Translocation of the HIV-1 TAT Peptide Across Lipid Membranes. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 20805−20810.
(18) MacCallum, J. L.; Bennett, W. F. D.; Tieleman, D. P. Distribution of Amino Acids in a Lipid Bilayer from Computer Simulations. *Biophys. J.* **2008**, *94*, 3393−3404.
(19) Sun, D.; Forsman, J.; Woodward, C. E. Evaluating Force Fields for the Computational Prediction of Ionized Arginine and Lysine Side-Chains Partitioning into Lipid Bilayers and Octanol. *J. Chem. Theory Comput.* **2015**, *11*, 1775−1791.
(20) Bonhenry, D.; Tarek, M.; Dehez, F. Effects of Phospholipid Composition on the Transfer of a Small Cationic Peptide Across a Model Biological Membrane. *J. Chem. Theory Comput.* **2013**, *9*, 5675−5684.
(21) Ou, S.; Lucas, T. R.; Zhong, Y.; Bauer, B. A.; Hu, Y.; Patel, S. Free Energetics and the Role of Water in the Permeation of Methyl Guanidinium across the Bilayer-Water Interface: Insights from Molecular Dynamics Simulations Using Charge Equilibration Potentials. *J. Phys. Chem. B* **2013**, *117*, 3578−3592.
(22) Wang, Y.; Hu, D.; Wei, D. Transmembrane Permeation Mechanism of Charged Methyl Guanidine. *J. Chem. Theory Comput.* **2014**, *10*, 1717−1726.
(23) Li, L.; Vorobyov, I.; Allen, T. W. Potential of Mean Force and pKa Profile Calculation for a Lipid Membrane-Exposed Arginine Side Chain. *J. Phys. Chem. B* **2008**, *112*, 9574−9587.
(24) Cardenas, A. E.; Shrestha, R.; Webb, L. J.; Elber, R. Membrane Permeation of a Peptide: It Is Better to be Positive. *J. Phys. Chem. B* **2015**, *119*, 6412−6420.
(25) Bennett, W. F. D.; Tieleman, D. P. The Importance of Membrane Defects-Lessons from Simulations. *Acc. Chem. Res.* **2014**, *47*, 2244−2251.
(26) Vorobyov, I.; Olson, T.; Kim, J.; Koeppe, R.; Andersen, O.; Allen, T. Ion-Induced Defect Permeation of Lipid Membranes. *Biophys. J.* **2014**, *106*, 586−597.
(27) MacCallum, J. L.; Bennett, W. D.; Tieleman, D. P. Transfer of Arginine into Lipid Bilayers Is Nonadditive. *Biophys. J.* **2011**, *101*, 110−117.
(28) Vondrášek, J.; Mason, P. E.; Heyda, J.; Collins, K. D.; Jungwirth, P. The Molecular Origin of Like-Charge Arginine-Arginine Pairing in Water. *J. Phys. Chem. B* **2009**, *113*, 9041−9045.
(29) Shih, O.; England, A. H.; Dallinger, G. C.; Smith, J. W.; Duffey, K. C.; Cohen, R. C.; Prendergast, D.; Saykally, R. J. Cation-Cation Contact Pairing in Water: Guanidinium. *J. Chem. Phys.* **2013**, *139*, 035104.
(30) Vazdar, M.; Vymětal, J.; Heyda, J.; Vondrášek, J.; Jungwirth, P. Like-Charge Guanidinium Pairing from Molecular Dynamics and Ab Initio Calculations. *J. Phys. Chem. A* **2011**, *115*, 11193−11201.
(31) Inagaki, T.; Aono, S.; Nakano, H.; Yamamoto, T. Like-Charge Attraction of Molecular Cations in Water: Subtle Balance between Interionic Interactions and Ionic Solvation Effect. *J. Phys. Chem. B* **2014**, *118*, 5499−5508.
(32) Vazdar, M.; Uhlig, F.; Jungwirth, P. Like-Charge Ion Pairing in Water: An Ab Initio Molecular Dynamics Study of Aqueous Guanidinium Cations. *J. Phys. Chem. Lett.* **2012**, *3*, 2021−2024.
(33) Mason, P. E.; Neilson, G. W.; Enderby, J. E.; Saboungi, M.-L.; Dempsey, C. E.; MacKerell, A. D.; Brady, J. W. The Structure of Aqueous Guanidinium Chloride Solutions. *J. Am. Chem. Soc.* **2004**, *126*, 11462−11470.
(34) Kubíčková, A.; Křížek, T.; Coufal, P.; Wernersson, E.; Heyda, J.; Jungwirth, P. Guanidinium Cations Pair with Positively Charged Arginine Side Chains in Water. *J. Phys. Chem. Lett.* **2011**, *2*, 1387−1389.
(35) Toukan, K.; Rahman, A. Molecular-Dynamics Study of Atomic Motions in Water. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1985**, *31*, 2643−2648.
(36) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(37) Berger, O.; Edholm, O.; Jähnig, F. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophys. J.* **1997**, *72*, 2002−2013.
(38) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(39) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(40) Parrinello, M.; Rahman, A. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Phys. Rev. Lett.* **1980**, *45*, 1196.
(41) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(42) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(43) Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187−199.
(44) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. The Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(45) Hub, J. S.; de Groot, B. L.; van der Spoel, D. gwham A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713−3720.
(46) Becke, A. D. Density-Functional Exchange-Energy Approximation With Correct Asymptotic Behavior. *Phys. Rev. A: At., Mol., Opt. Phys.* **1988**, *38*, 3098.
(47) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron-Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(48) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(49) Lippert, G.; Hutter, J.; Parrinello, M. A Hybrid Gaussian and Plane Wave Density Functional Scheme. *Mol. Phys.* **1997**, *92*, 477−487.
(50) Goedecker, S.; Teter, M.; Hutter, J. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 1703−1710.
(51) Jonchiere, R.; Seitsonen, A. P.; Ferlat, G.; Saitta, A. M.; Vuilleumier, R. Van der Waals Effects in Ab Initio Water at Ambient and Supercritical Conditions. *J. Chem. Phys.* **2011**, *135*, 154503.
(52) Bankura, A.; Karmakar, A.; Carnevale, V.; Chandra, A.; Klein, M. L. Structure, Dynamics, and Spectral Diffusion of Water from First-Principles Molecular Dynamics. *J. Phys. Chem. C* **2014**, *118*, 29401−29411.
(53) Mundy, C. J.; Mohamed, F.; Schiffman, F.; Tabacchi, G.; Forbert, H.; Kuo, W.; Hutter, J.; Krack, M.; Iannuzzi, M.; McGrath, M. et al. CP2K, version 2.5.1. http://www.cp2k.org [accessed Jan 2015].
(54) Berendsen, H.; Grigera, J.; Straatsma, T. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269−6271.
(55) Grossfield, A. WHAM: the Weighted Histogram Analysis Method, version 2.0.9. http://membrane.urmc.rochester.edu/content/wham [accessed March 2015].
(56) Jungwirth, P. Ion Pairing: From Water Clusters to the Aqueous Bulk. *J. Phys. Chem. B* **2014**, *118*, 10333−10334.
(57) Allolio, C.; Salas-Illanes, N.; Desmukh, Y. S.; Hansen, M. R.; Sebastiani, D. H-Bonding Competition and Clustering in Aqueous LiI. *J. Phys. Chem. B* **2013**, *117*, 9939−9946.

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

(58) Baer, M. D.; Mundy, C. J. An Ab Initio Approach to Understanding the Specific Ion Effect. *Faraday Discuss.* **2013**, *160*, 89−101.
(59) Wang, S.; Larson, R. G. Water Channel Formation and Ion Transport in Linear and Branched Lipid Bilayers. *Phys. Chem. Chem. Phys.* **2014**, *16*, 7251−7262.
(60) Glaser, R.; Leikin, S.; Chernomordik, L.; Pastushenko, V.; Sokirko, A. Reversible Electrical Breakdown of Lipid Bilayers: Formation and Evolution of Pores. *Biochim. Biophys. Acta, Biomembr.* **1988**, *940*, 275−287.
(61) Sapay, N.; Bennett, W. F. D.; Tieleman, D. P. Thermodynamics of Flip-Flop and Desorption for a Systematic Series of Phosphatidylcholine Lipids. *Soft Matter* **2009**, *5*, 3295−3302.
(62) Robison, A. D.; Sun, S.; Poyton, M. F.; Johnson, G. A.; Pellois, J.-P.; Jungwirth, P.; Vazdar, M.; Cremer, P. S. Polyarginine Interacts more Strongly and Cooperatively than Polylysine With Phospholipid Bilayers. Submitted to *J. Am. Chem. Soc.*, **2015**.

DOI: 10.1021/acs.jpcb.5b10404
J. Phys. Chem. B 2016, 120, 143−153

Arginine-rich cell-penetrating peptides induce membrane multilamellarity and subsequently enter via formation of a fusion pore

Christoph Allolio[1a,b,c], Aniket Magarkar[1a,d], Piotr Jurkiewicz[e], Katarina Baxova[a], Philip E. Mason[a], Radek Šachl[e], Marek Cebecauer[e], Martin Hof[e], Dominik Horinek[b], Veronika Heinz[f], Reinhard Rachel[g], Christine Ziegler[f,h], Adam Schröfel[i], and Pavel Jungwirth[2a]

[a]Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Flemingovo nám. 2, CZ-166 10 Prague 6, Czech Republic; [b]Institut für Physikalische und Theoretische Chemie Universität Regensburg, 93040 Regensburg,Germany; [c]Fritz Haber Institute for Molecular Dynamics and Department of Chemistry, The Hebrew University of Jerusalem Edmond J. Safra Campus, Giv'at Ram, Jerusalem 9190401, Israel; [d]Faculty of Pharmacy, University of Helsinki, Viikinkaari 5E, Helsinki 00014, Finland ; [e]J. Heyrovský Institute of Physical Chemistry, Czech Academy of Sciences, v.v.i., Dolejškova 2155/3, 182 23 Prague 8, Czech Republic; [f]Institute of Biophysics and Biophysical Chemistry, University of Regensburg, Regensburg, Germany; [g]Lehrstuhl für Mikrobiologie und Archaeenzentrum, Universität Regensburg, Universitätsstrasse 31, D-93053 Regensburg, Germany; [h]Institute of Biophysics and Biophysical Chemistry, University of Regensburg, Regensburg, Germany; [i]Max Planck Institute of Molecular Cell Biology and Genetics, Pfotenhauerstr. 108, D-01307 Dresden, Germany

This manuscript was compiled on July 4, 2018

We show that arginine-rich peptides do not enter cells in a passive mode by directly passing through a lipid membrane; instead, they passively enter vesicles and live cells by inducing membrane multilamellarity and fusion. The molecular picture of this penetration mode, which differs qualitatively from the previously proposed direct mechanism, is provided by molecular dynamics simulations. The kinetics of vesicle agglomeration and fusion by an iconic cell penetrating peptide - nonaarginine - is documented via real time fluorescence techniques, while the induction of multilamellar phases in vesicles and live cells is demonstrated by a combination of electron and fluorescence microscopies. This concert of experiments and simulations reveals that the newly identified passive cell penetration mechanism bears analogy to vesicle fusion induced by calcium ions, indicating that the two processes may share a common mechanistic origin.

Cell penetrating peptides | Vesicle fusion | Ion membrane interactions | Molecular dynamics | Fluorescence microscopy | Electron microscopy

Cell penetrating peptides have a unique potential for targeted drug delivery, therefore, mechanistic understanding of their membrane action has been sought since their discovery over 20 years ago.(1) While ATP-driven endocytosis is known to play a major role in their internalization(2) , there has been also ample evidence for the importance of passive translocation(3–5) for which the direct mechanism, where the peptide is thought to directly pass through the membrane via a temporary pore, has been widely advocated. (4, 6–9) Here, we question this view and show that arginine-rich cell penetrating peptides instead passively enter vesicles and live cells by inducing membrane multilamellarity and fusion.

Ions do not dissolve in oil. From this point of view the direct passive mechanism of cell penetration is intuitively problematic, as cationic peptides such as polyarginines or the trans-activating transcriptional activator (TAT) are too highly charged to be able to pass through the "oily" interior of a lipid membrane. The concept of direct penetration was seen plausible due to the action of the related antimicrobial peptides, which are also charged, but in addition contain a large fraction of hydrophobic residues(10): These peptides are known to stabilize pores in membranes(11). At a close inspection, however, it becomes clear that their charged sidechains do not interact directly with the aliphatic chains in the low dielectric interior of the phospholipid bilayer, but rather stabilize transient water channels or act as terminal residues anchoring the transmembrane helix.(10, 12) Taken together, the passive action of cell penetrating peptides (CPP) seems to be very different from direct translocation across an otherwise unperturbed cell membrane.

To make matters even more confusing, experimental facts and suggested mechanisms often seem contradictory to each other. For example, there are conflicting reports whether or not nonaarginine (R_9) is able to penetrate vesicles composed purely of 1-palmitoyl-2-oleoyl-phosphatidylcholine (POPC). (5, 13, 14) Additionally, fluorescence microscopy suggests that R_9 is able to deform membranes(5, 15) and small angle X-ray scattering (SAXS) experiments reveal phase transitions induced in lipid systems by polyarginines.(4) An important factor in these observations appears to be the membrane composition with negatively charged lipids facilitating membrane translocation of cationic peptides.(7, 16) Indeed, there is some evidence that a direct mechanism may be enabled by hydrophobic counterions, such as pyrene butyrate(14, 17) or presence of an unphysiological concentration of phosphatidic acids.(7) The relevance to of these phenomena to actual cellular uptake

Significance Statement

The passive translocation mechanism of arginine-rich cell penetrating peptides has puzzled the scientific community for more than twenty years. In this study we propose a new mechanism of passive cell entry involving fusion of multilamellar structures generated by the cell penetrating peptides. The geometry of entry for this mechanism is completely different from previously suggested direct translocation mechanisms leading to a new paradigm for designing molecular carriers for drug delivery to the cell.

A.M. and C.A. performed the simulations, supervised by P.Jun and D. H., vesicle fluorescence related experiments and DLS were performed by P.Jur., C.A., K.B. and supervised by M. H. fluorescence data were interpreted by C.A. and R.S. Cryoelectron microscopy on LUVs was performed by V.H. and R.R., supervised by C.Z.; fluorescence and electron microscopy on HeLa cells was done by M.C., A.S. and P.E.M.; simulation visualizations were done by A.M. and C.A.; C.A. and P.Jun wrote the paper.

The authors declare no competing financial interests.

[1]C.A. and A.M. contributed equally to this work

[2]To whom correspondence should be addressed. E-mail: pavel.jungwirth@marge.uochb.cas.cz.

is not clear and current discussions present direct mechanisms side by side with endocytosis-like membrane deformations induced by the CPP.(18)

Another fundamental cellular process involving membranes and charged species is fusion of vesicles with the cell membrane during calcium-triggered exocytosis. In cells, vesicle-membrane fusion is mediated by the SNARE protein complex(19, 20) with synaptotagmins(21), nevertheless, it can also be induced in *in vitro* lipid vesicles without the need for the presence of the protein machinery(22, 23). It is experimentally well established that Ca^{2+} is a key player capable of promoting vesicle fusion(24) and there is general consensus about the fusion mechanism, which proceeds via a stalk intermediate, followed by formation of a hemifused structure and opening of a fusion pore.(25, 26) In this context, it is worth mentioning that cationic cell penetrating peptides, especially TAT and its derivatives, are known to aggregate at phospholipid membranes and occasionally fuse vesicles.(2, 5, 22, 27) This brings an idea, which is examined further in this study, that the processes of passive cell penetration and membrane fusion may be mechanistically more intimately connected than thought so far.(27)

Results and Discussion

Exploring vesicle penetration by a fluorescence leakage essay. In order to explore the potential connection between cell penetration and membrane fusion, we start by investigating the abilities of R_9 as an archetypal CPP, in contrast to non-CPPs like tetraarginine (R_4) or nonalysine (K_9) (28), to penetrate and cause leakage of large unilamellar vesicles (LUVs) of varying lipid compositions using a fluorescence leakage assay (for details see Methods and the SI Appendix). While, contrary to some published data(14), pure POPC LUVs do not show leakage upon exposure to R_9 even at high peptide concentrations, LUVs composed of mixtures of 1,2-dioleoyl-phosphatidylethanolamine (DOPE) and 1,2-dioleoyl-phosphatidylserine (DOPS) exhibit leakage as long as the content DOPE is sufficiently high (see Table S3 in the SI Appendix). Despite different lipid composition of the present vesicles compared to live cells, we did recover the arginine 'magic', i.e., the high activity of CPPs with more than about seven aminoacids and high arginine content (29). In these vesicles, similarly to experiments in cells, R_9 was always found to be an efficient leakage agent, in contrast to K_9 or R_4 (see Figure 1, top left). Using dynamic light scattering (DLS) measurements we showed that leakage is accompanied with vesicle aggregation, as the leakage kinetics mimics the increase in time of the mean diamater of the aggregates (see Figure 1, top right and the SI Appendix, Table S3 for the DOPE-rich lipid composition).

Membrane fusion induced by calcium as well as by cationic cell penetrating peptides. The range of lipid compositions of vesicles capable of being leaked by R_9 is at odds with simulations of direct translocation, where a far higher translocation free energy has been predicted for DOPE-rich bilayers than for those rich in POPC.(30) However, it seems to match compositions known to enhance vesicle fusion by calcium.(22, 23, 31) In order to verify this correlation, we repeated the experiments with Ca^{2+} instead of R_9. Indeed, we were able to observe calcium-induced leakage for lipid compositions that were most susceptible to leakage by the CPPs as well (see Table S3 in the SI Appendix). In order to obtain a comparable effect, however, the concentration of Ca^{2+} had to be significantly higher than that of R_9, even when taking into account the significantly larger charge carried by the latter species.

At high peptide content, the LUV leakage kinetics is described *quantitatively* by a second order rate law in the vesicle concentration (for details see Figure S8 and the kinetic model in the SI Appendix). This indicates that aggregation of vesicles and the double bilayer formed during this process is essential for vesicle leakage. It is indicative that for all lipid compositions at which significant leakage occurs the vesicles also exhibit R_9-induced fusion, as detected by a near complete lipid mixing in a Förster resonance energy transfer (FRET) assay (see the SI Appendix, esp. Figure S8). Moreover, the leaky lipid compositions are those known to be susceptible to Ca^{2+} mediated fusion. Note that leaky fusion has been observed previously in conjunction with cell penetrating peptides and has even been used to classify them.(27) In addition to the above circumstantial evidence for a direct connection between cell penetration and membrane fusion we add further experimental support using confocal microscopy finding agglomeration of giant unilamellar vesicles (GUVs) together with leakage (see Figure 1). We were able to observe fusion of GUVs directly (see the Supporting Video) and, using Oregon Green 488 (OG)-labelled peptides, we confirmed a preferential adsorption of fluorescently labeled R_9 to the vesicles (see Figure S5 in the SI Appendix). Finally, adding Ca^{2+} instead of R_9 to the GUVs we found a functionally analogous behavior (see the SI Appendix).

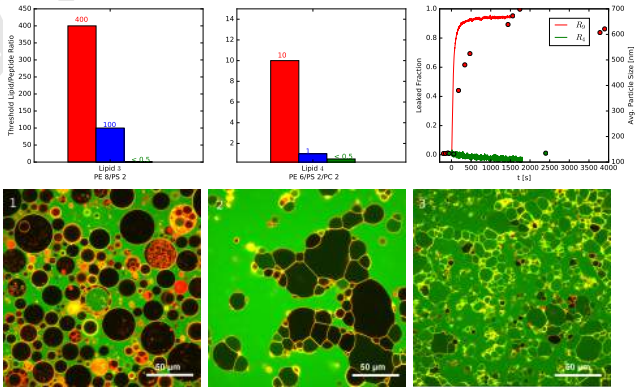


Fig. 1. Fluorescence spectroscopy results. Top Left: Threshold concentrations for leakage induced by R_9, R_9 and R_4 given as inverse of the peptide/lipid ratios for two lipid compositions: DOPE/DOPS 80/20 (3) and DOPE/DOPC/DOPS 60/20/20 (4) (the higher the threshold value, the more efficient the peptide is in leaking the vesicles). Top Right: DLS measurements showing particle growth overlayed with leakage kinetics for R_9 for the composition (3) and absence of particle growth and leakage for R_4. Bottom panel: Fluorescence microscopy images showing the effect of R_9 on GUV with composition (4). From left to right: **1**. No peptide added, **2**. Shortly after addition of R_9 **3**. Final state after 1h.

Ideal fusion topologically precludes cell penetration. The similarities in aggregation/fusion caused by R_9 and Ca^{2+} are illustrated in Figure 2, **A**-**J**. In the context of the present study, it is important to note that there is no topological way for peptides to enter the vesicles from the outside (or do the reverse) by an ideal fusion process, within which two unilamellar vesicular structures coalesce, as it merely connects the

interiors of the two vesicles. A previous electron microscopy study of Ca^{2+}-mediated fusion reported content loss along the fusion diaphragm, attributing it to the strain induced by the deformation during agglomeration.(32). A close observation of the GUVs in Figure 1 indeed reveals that they are significantly deformed as the agglomeration creates nearly planar surfaces at the regions of contact, indicating high tension. However, leaky fusion can hardly be the mechanism of penetration of peptides into vesicles since it would either rapidly close the adjacent pores or lead to vesicle deflation via membrane rupture. Neither can it explain continous leakage into intact vesicles. We show below that a solution lies in locally bifurcating the membrane, leading to multilamellar structures. Such a pathway allows the peptides to enter by fusion, as shown schematically in Figure 2, **K-V**.

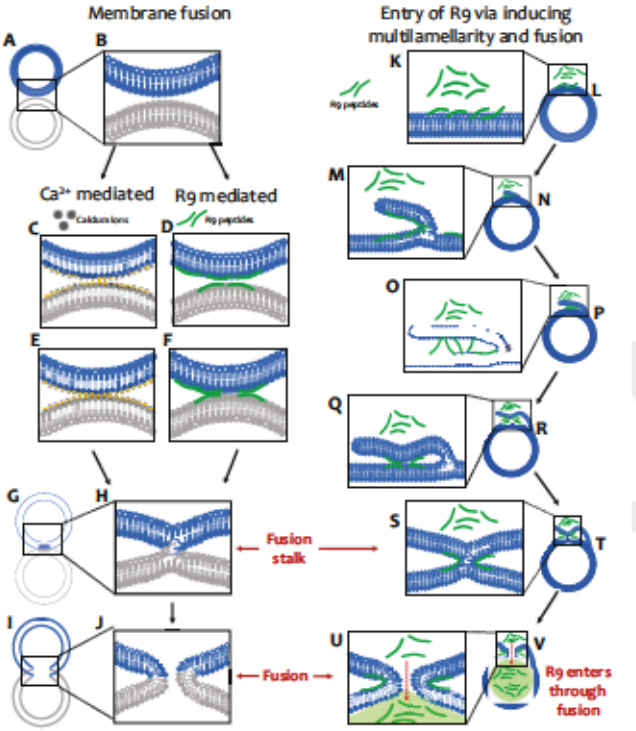


Fig. 2. The schematic mechanisms of R_9 and Ca^{2+} mediated vesicle fusion: **A** Fusion of different vesicles (in red and blue), by **B** interface contact. **C, D** Adsorption of the charged particles (R_9 in green and Ca^{2+} in yellow). **E, F** Agglomeration of the bilayers induced by cross-linking. **G** Stalk formation. **H** Opening of the fusion pore. **I,J** R_9 translocation via self-fusion of a single vesicle, **K-V** starting from a flat vesicle surface bilayer. **K** Strong adsorption of R_9. **K,L** Membrane bifurcation through adhesion and curvature.**M,N** Extension of the bifurcated bilayer through R_9 cross-linking **O,P**. Agglomeration of the bilayers induced by cross-linking of two bilayers on the same vesicle **Q,R**. Stalk formation**S,T** and opening of the fusion pore through which additional R_9 peptides enter **U,V**.

Induced multilamellarity as a solution to the topological conundrum.

Cryo-electron microscopy: Seeing is believing. A tendency of GUVs(15) or cells(14) to become multilamellar upon addition of CPPs has been suggested recently. To further explore this idea, we first conducted cryo-EM experiments on LUVs. The obtained cryo-EM images indeed reveal formation of multilamellar domains and lipid bilayer bifurcations after the addition of R_9, see Figure 3. Additional time-resolved FRET

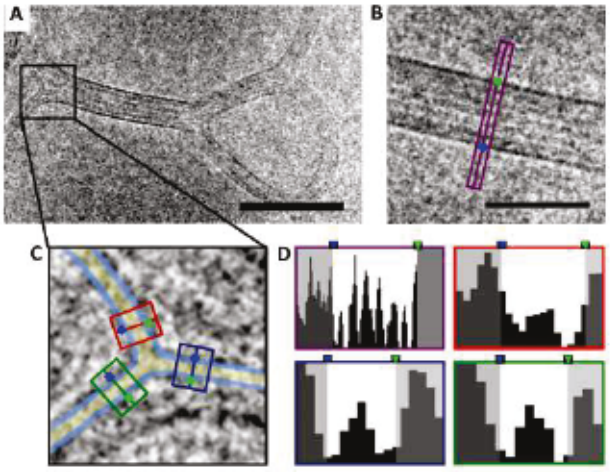


Fig. 3. Electron micrographs of LUVs in the presence of R_9. (**A**): vesicles treated with R_9 (>60 sec.) fuse with each other and exhibit bifurcated, multilamellar membranes; Bar = 100 nm. (**B**) Example of a multilamellar membrane. Bar = 50 nm.; The violet box is analyzed in (**D**). (**C**): Example of a membrane bifurcation. The membranes before and after the bifurcation site are analyzed by line-scans. The line-scan areas are marked with colored boxes. (**D**): The histograms are boxed in the same color as the respective line- scan areas in (**B**) and (**C**). The histogram (upper left panel) corresponding to the multilamellar membrane (shown in (**B**)) exhibits seven distinct minima attributed to individual membranes.

experiments on fluorescently labeled LUVs reveal presence of inter-bilayer energy transfer, which provides independent confirmation for the induction of multilamellar lipid structures by R_9 (see the SI Appendix and Figure S8 therein). Importantly, calcium ions are also able to fuse and collapse vesicles to multilamellar phases due to the ability of Ca^{2+} to bridge phosphates from different bilayers.(32) The cryo-EM structures also provide some additional evidence for fusion, as the LUVs in Figure 3, A are many times larger than those found in the initial state (again see Figure S8, A).

Multilamellar structures can be formed via folding of a membrane or by stacking of deflated vesicles. Any process based on direct membrane stacking would, however, add an even number of bilayers in between the vesicles and, therefore, would not lead to leakage via fusion. It is thus a key finding that by counting the lipid bilayers we frequently find odd numbers (see Figure 3, D). Moreover, a close inspection of the EM micrographs provides direct evidence for bilayer bifurcation at multiple positions (see Figure 3, C for an example). We conclude that R_9 is indeed capable of inducing multilamellarity by membrane adsorption and bifurcation, rendering a cell penetration mechanism via fusion feasible.

Fluorescence and electron microscopy on HeLa cells. In order to directly explore the mechanism behind cellular uptake of CPPs in the absence of endocytosis, we first observed penetration of CPPs into living human HeLa cells by fluorescence confocal microscopy (Figures S6, S7 in the SI Appendix). HeLa cells incubated with 15 μM Oregon-Green (OG) labeled-R_9 peptide for 3 minutes at 4 °C already exhibited surface fluorescence and, in particular, the presence of highly fluorescent foci (SI Appendix, Figure S6, arrows). Longer incubation of cells with the peptide increased the number of foci detected on HeLa cells. Cytosolic presence of the peptide was observed

as early as 5 min after addition of the peptide to the cells and the number of penetrated cells increased over prolonged incubation time (even when cells were washed after 3 minutes to remove free peptide from incubation medium). All cells with cytosolic peptide exhibited at least 1-2 fluorescent foci on its surface. Moreover, these foci were found only in cells with exocytosis blocked by low temperature.

These results suggest that R_9 rapidly accumulates in a very few places on a cell at low temperatures. This phenomenon cannot be explained by the peptide forming pores on the surface of exposed cells. Rather, some sort of membrane reorganisation with accumulated peptide facilitates entrance of the peptide into cells. Previously, electron microscopy (EM) on cells revealed that addition of R_9 leads to the formation of layered membrane domains.(33) Motivated by these observations, we performed EM experiments on HeLa cells with added fluorescently labeled R_9. In accord with our observations on vesicles and with previous EM on cells(33) we see regions of protrusions, bifurcations, and multilamellarity by incubating cells with R_9 at 4 °C. At these conditions active endocytosis is switched off and only passive cell penetration is operational (see Figure 4, B). The branched structures (see zoomed-in images in Figure 4, C) are topologically identical to those suggested in the schematic drawing above (Figure 2, M) and to the structures observed in the LUVs (Figure 3). The observed protrusions indicate that a strong curvature-generating interaction is at work, as can be seen in Figure 4, D. Specifically, we interpret the budding spherical protrusions as being due to negative (Gaussian) curvature generation, similar as occuring in a fusion stalk. By overlaying fluorescence images with EM we clearly see that the changes in the bilayer structure correlate with the location of the fluorescently labeled OG-R_9, as can be seen in Figure 4, A. Our EM images on HeLa cells are thus consistent with the above results on vesicles, as well as with previous observations showing that giant plasma membrane vesicles (GPMVs) are susceptible to R_9 penetration, but only in the presence of intact membrane proteins.(3). Note that in contrast to our leakage experiments on LUVs the local, encapsulated foci will allow only very limited diffusion, thereby limiting potential cell toxicity of the CPP. Their locality at the cell membrane also points to a yet unknown specific interaction which is not present in the synthetic vesicle systems.

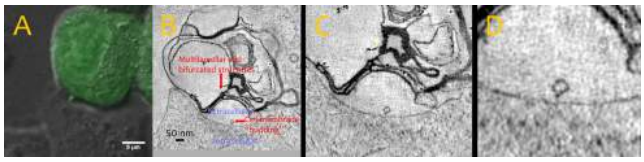


Fig. 4. Electron microcopic and fluorescence microscopy images of the same spot on a fixated HeLa cell in the presence of OG-R_9. (**A**): A fluorescence microscopy image of the multilamellar spot showing the presence of the labeled peptide. (**B-D**): An electron microscopy image at three zooms exhibiting bifurcated, multilamellar membranes and vesicle budding. (**C**) An example of a multilamellar membrane structure.; (**D**): Focus on a budding protrusion.

Molecular dynamics simulations: Atomistic insights. In order to gain atomistic insight into the fusion process and its connection to cell penetration we performed molecular dynamics simulations. Previous studies, based on continuum and coarse-grained models agree that fusion proceeds via a stalk intermediate. (25, 34) The stalk is strongly concave, which explains the observed lipid selectivity toward phosphatidylethanolamine (PE) headgroups which spontaneously induce negative curvature. Indeed, to form the stalk a significant deformation of the membrane has to take place,(25) implying a high activation energy of the bilayer-bilayer fusion ($\approx$ 70 kJ/mol) (35, 36).

Our simulation setups involve strongly curved bilayer geometries, with their shapes being designed such as to lower the barrier for stalk formation via a positive initial curvature. This is consistent with the fact that the surface of cells is also densely decorated with membrane extrusions and intrusions which expose highly curved membranes to the environment. (37) The stress hereby induced in the PE-rich bilayers allows us to forego the use of explicit biasing toward the stalk geometry, which in our simulations occurs spontaneously (see Methods and the SI Appendix for full details). Snapshots from the R_9 or Ca^{2+} mediated fusion processes are presented in Figure 5. We find both Ca^{2+} ions and the charged R_9 side chains to bind to lipid headgroups, primarily at the negatively charged phosphates. Subsequently, we observe mechanistic similarities in the membrane fusion mediated by Ca^{2+} and R_9. The first step in the fusion process is crosslinking, i.e., simultaneous binding of lipids from two membrane bilayers by either Ca^{2+} or R_9, with the latter being found to be a particularly effective crosslinker. As the ions keep the bilayers in close contact, the lipid tails eventually crosslink, too, in what appears to be the rate-determining step of the whole fusion process. This lipid tail cross-linking occurs within about half a microsecond. After the lipid tail flip-flop takes place, a stalk starts forming within a few nanoseconds (Figure 5, B and C).

In Figure 5, D we examine the action of R_9 on the membrane in close detail. First, we note the long-range cross-linking capability of R_9, which is likely to be responsible for its strongly agglomerating effect on vesicles and for stabilizing the multilamellar structures we find in electron microscopy. R_9 tends to be only partially adsorbed at the membrane and can thus easily reach across the interface and attach to the second bilayer. It is clear that for such a cross-linking a minimum chain length is necessary, which explains the ineffeciency of R_4 in this respect (see also the control simulation data, in Figure S3 in the SI Appendix). In addition, R_9 also forms agglomerates when cross-linking the interfaces. It is known that R_9 binding to membranes is more cooperative than that of K_9(38), making the former a much more efficient agent of membrane fusion and cell penetration than the latter. The attraction of R_9 to negatively charged lipids also leads to a lipid sorting effect, with phosphatidylserine lipids (Figure 5, D, orange) being accumulated next to the R_9 agglomerate. Electrostatics-based lipid sorting (involving, e.g., gangliosides) may contribute to the action in cells, membranes of which would otherwise not be sufficiently active. In the bottom right panel of Figure 5, we show that R_9 is preferentially adsorbed to regions of negative curvature(39) (marked by the color-coded surface), as this allows it to efficiently bind its sidechains to lipid headgroups (marked in gold). Aggregation of headgroups increases membrane tension in the headgroup region and exposes hydrophobic patches (visible in all panels of Figure 5, D). The exposure of hydrophobic patches in turn lowers hydration repulsion. Once lipid tails from opposing bilayers are in contact, lipid flip-flop can readily occur, starting thus the fusion cascade. At the surface of the fusion stalk and

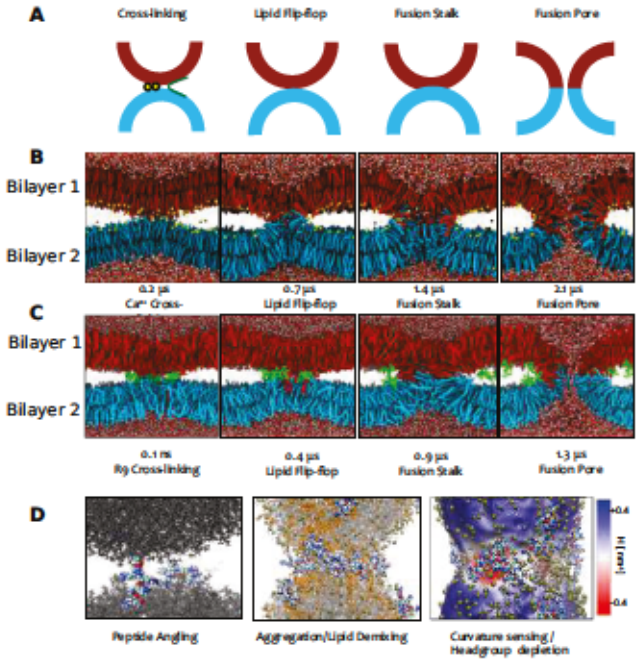


Fig. 5. A: Schematic drawing of vesicle fusion - Lipid crosslinking, stalk initialization and subsequent onset of stalk formation through lipid flip-flop. **B**: Time evolution of the Ca^{2+} fusing bilayer system. **C**: Time evolution of the R_9 vesicle-self fusion system. Cross section of systems undergoing fusion: crosslinking, flip-flop, fusion stalk , fusion pore. **D**: Driving forces of the mechanism. Peptide "angling" crosslinks vesicles and aggregates membranes, peptide agglomeration and lipid demixing create fusable interface, generation of negative curvature through strong binding to headgroups.

the pore, a very strongly negative Gaussian curvature and a negative mean curvature occur simultaneously. This stabilizes the actual geometry and, therefore, lowers the barrier for fusion. The induction of negative Gaussian curvature has previously been proposed as a characteristics of cell penetrating peptides.(4). Importantly for the present mechanism, we do not observe membrane rupture or water permeation during the fusion process.

Conclusions

In summary, we unraveled here a novel passive entry mechanism of cell penetrating peptides via branching and layering of membranes followed by fusion of the agglomerated systems. The layering is induced by a cooperative bridging of bilayers via adsorbed R_9. The peptides also induce membrane bifurcations that allow to connect the vesicle exterior and interior via fusion and thus to translocate the cell penetrating peptides (which would not be topologically possible within an ideal vesicle fusion). The actual R_9-induced fusion process then mirrors that of Ca^{2+} assisted vesicle fusion. The viability of this mechanism is supported by experimental results from electron microscopy, fluorescence microscopy, and light scattering, together with with molecular dynamics simulations. The atomistic simulation data shed further light on the molecular mechanism of formation of the fusion stalk and pore. While the new mechanism has been unraveled for lipid vesicle systems, which allow for investigations with unprecedented molecular detail, induction of membrane branching and multilamellarity by polyarginines has been observed by electron microscopy also in cellular membranes of live cells at low temperatures, indicating that the passive cell penetration process analogously involves membrane fusion.

Future work will be directed toward unraveling further molecular details of the cell penetration mechanism suggested in this discovery study. In the next step, we need to understand the interaction of cell penetrating peptides with biological cell surfaces. Increased experimental understanding of the specific binding will allow us to develop more realistic models and vice versa. This will not only allow to firmly establish all the details of this hithertho unrecognized mechanism of passive cell penetration, but will also have a direct impact on development of smart cell delivery strategies for therapeutic molecules employing cell penetrating peptides. Should we get this passive cell penetration mechanism under full control, we may eventually be able to expoit it to directly deliver cargo into the cell without the need for releasing it from the transport vesicles, as is the case in active endocytosis.

Materials and Methods

The SI Appendix contains HeLa cell and GUV fluorescence microscopy images, further details on LUV leakage, the kinetic model for leakage as well as the FRET fits and the details of the FRET model. It also contains further cryo-EM images, a cryo-EM-based membrane thickness determination as well as further simulation details. We also provide a Supporting Video with direct evidence of fusion induced by R_9.

Experimental Details.

Leakage assay for LUVs. Calcein containing vesicles were stirred at room temperature with LUV buffer in a quartz cuvette to obtain 1.5 ml of solution. The Calcein fluorescence was monitored at 520 nm, with excitation at 495 nm. After an initial stirring phase of no less than 200 s, 3-6 μl peptide in buffer solution was added. For Ca^{2+} mediated leakage, we added a Ca^{2+} buffer (40 mM $CaCl_2$, 100 mM KCl, pH 7.40, Glucose, 318 mOsm/l). Fluorescence intensity measurements were performed on a Fluorolog-3 spectrofluorimeter (model FL3–11; JobinYvon Inc., Edison, NJ, USA) equipped with a xenon-arc lamp. After the fluorescence intensity reached a plateau 50 μl of TRITON-X were added after which the signal was again allowed to stabilize.

FRET assay for LUVs. The fluorescent lipid probes used as a FRET pair were DOPE-Atto 488 and DiD. We used the LUV preparation protocol described in the SI. The fluorescent probe to lipid molar ratios were 1:200 for the donor, and 1:100 for the acceptor. We measured at 0.2 mM lipid concentration. Time-resolved measurements were performed on a 5000U Single Photon Counting setup using a cooled Hamamatsu R3809U-50 microchannel plate photomultiplier (IBH, UK). DOPE-Atto 488 was excited at 470 nm with LDH-P-C-470 laser (75 ps pulse width, 5 MHz repetition rate, PicoQuant, Berlin, Germany). Flourescence decays were measured immediately after mixing the donor labelled vesicles with the acceptor labelled ones and after incubation of the vesicles with R_9 peptide. FRET was measured at the P/L ratios given in the SI Appendix in Table S3, results are shown in Figure S8). The obtained signals were interpreted in terms of multilamellar fused structures (see the SI Appendix).

Confocal Microscopy for GUVs. GUVs labeled with DiD were prepared using electroformation(40) in a 300 mOsm/l sucrose solution. Sinusoidal altering voltage was used. First, the peak to peak voltage was increased stepwise from 0.02 V to 1.1 V at 10 Hz for 45 min, then kept constant for further 90 min, and finally the frequency was lowered to 4 Hz at 1.3 V and maintained for 30 min. Prepared GUVs were diluted with a glucose buffer (9 mM HEPES, pH 7.40 (KOH), 90 mM KCl, 90 mM EDTA, 120 mM glucose, 300 mOsm/l, filtrated) with 20 μl 50 nM Atto-488 to a total volume of 300 μl. Images were recorded using Olympus IX81 laser scanning confocal microscope(Olympus, Hamburg, Germany). Atto-488, OG-R_9 and

DiD were excited using 488 or 632 nm lasers, respectively. Depending on the measurement the final lipid concentration was between 3 and 15 μM. After the begin of the measurement, 1.2 μl R_9 solution (3 nmol) were added to the GUV chamber. In the case of the labelled peptides, we omitted the Atto-488 and added 3 nmol of labelled peptide OG-R_9. For the Ca^{2+} experiment, we added 3μl of 0.5 M $CaCl_2$ solution to achieve a concentration of 5 mM.

Cryoelectron Microscopy for LUVs. For cryo-EM sample preparation, 4 μl of the sample were applied to plasma-cleaned EM grids (400 mesh copper grids, covered with Quantifoil film (R1.2/1.3)). Samples were plunge-frozen on the grids in liquid Ethane in a Grid Plunger (Leica EM GP, Leica Microsystems GmbH) with the following parameters: pre-blotting exposure 5 sec., blotting time 1.7 sec., no post-blotting exposure. Chamber humidity was set to 95% at 22 °C. Untreated samples, containing LUVs of Lipid composition *3* (10 mM) were used as a control, and the LUV solution was treated with R_9 (c = 25 mM) in a ratio of 10:1 for t >60 sec. immediately before plunge freezing. Cryo-electron micrographs were collected on a JEM-2100F (JEOL Germany GmbH) operated at 200 kV. Images were recorded in a semi-automated procedure using EM-SPC from the EM-Tools package (TVIPS GmbH) with a TemCam-F416 (TVIPS GmbH) at a nominal magnification of 20.000x under low-dose conditions, with defocus values ranging between 1.5 and 3.5 μm. Bilayer thickness was measured (see the SI Appendix).

Dynamic Light Scattering for LUVs. Samples were transferred to UV grade poly(methyl methacrylate) cuvettes (Kartell, Noviglio, Italy), and equilibrated at 298 K for 3 min before each measurement. The light scattering setup of Zetasizer Nano ZS (Malvern Instruments Ltd., Worcestershire, UK) consisted of a He–Ne laser (532 nm) and an avalanche photodiode detector (APD). The scattering intensity was collected at the angle of 173 °. Mean particle sizes were obtained using contin fitting implemented in Zetasizer Software 6.2 (Malvern Instruments Ltd., Worcestershire, UK). Datapoints were obtained by averaging over 20 scans, so that every datapoint represent averaging over ca. 2 min of kinetics.

Details on fluorescence spectroscopy and electron microscopy on HeLa cells. HeLa cells (ATCC CCL-2) were maintained in 37 °C incubator with 5% CO_2 in DMEM tmedium containing 10% fetal bovine serum. Forty thousand cells were seeded to a well of 8-well μ-slide (ibiTreat; ibidi, Germany) 16-20 hours before the experiment. Cells were first washed with DMEM medium lacking serum and pH colour indicator (SF-DMEM) and kept at 4 °C for 15 min to inhibit endocytic processes. For treatment, pre-cooled (4 °C) 15 μM solution of a peptide in SF-DMEM was added to cells via media exchange and incubated for indicated periods of time at 4 °C. In selected cases, cells were treated for 3 min with a peptide at 4 °C, washed with pre-cooled SF-DMEM and further incubated for indicated period of time at 4 °C in fresh SF-DMEM.

To verify cell viability, a test experiment was performed where after incubation with the R_9 peptide for 2h at 4 °C cells were returned to the incubator and cultured for 1-2 h at 37 °C. First, morphology of cells remained normal. Second, dramatic endocytosis was observed in cells previously incubated with R_9 peptide upon warming to 37 °C, indicating fully functional vital functions. This indates that incubation of cells in buffer at 4 °C for time periods required for passive entrance of the peptide to cells (i.e., up to 1 h) did not affect cellular viability in any significant way.

Cells were imaged using scanning confocal microscope (FluoView 1000, Olympus) equipped with 60x water immersion, NA 1.2 objective (UPlanSApo, Olympus). 488 nm steady-state semiconductor laser (Coherent) was used as an excitation source. Light power at the specimen did not exceed 3 μW. Triple dichroic mirror 488/543/633 and 490-525 nm bandpass filter were used. 512x512 pixel images were collected in monodirectional scanning mode (scanning speed of 4 μs/pixel) using Kalman filtering of 8 consecutive lines. Samples were kept on ice except for 1-2 minutes of image acquisition at room temperature. The images were processed for publishing using Fiji/ImageJ.(41)

Confocal fluorescence and DIC images for CLEM analysis were acquired on inverted Leica TCS SP8 SMD, using 63x NA1.2 water objective, 5x5 tile scan, total size 0.68x0.68 mm with 90 nm pixel size, bidirectional scan 600 Hz, line accumulation 3, photon counting mode, excitation at 488 nm with 10 μW at sample, emission detection window 496-587 nm. Cells were grown on glass coverslips with grid (MatTek Corporation, USA). Cells were fixed with 2% glutaraldehyde and 1% formaldehyde, postfixed with 1% OsO_4, dehydrated in ethanol series (10%, 30%, 50%, 70%, 90%, 95% and 2x 100%) and embedded in EMBed 812 resin (Electron Microscopy Sciences, USA). The cells of interest were chosen based on the fluorescence data and the selected area was sectioned to 300 nm thick serial sections. Sections were post-contrasted with uranyl acetate and lead citrate and decorated on both sides by 15 nm colloidal gold. The tomograms we acquired on Titan Halo transmission electron microscope, operating on 300 kV, using K2 direct detector (Gatan, USA) in counting mode and post-column energy filter (Gatan, USA). The acquired tomograms were reconstructed using IMOD software package(42).

Computational Details. In a first setup, we created two curved membranes via lipid population imbalances at the two leaflets of each bilayer. In the second setup, we put a very small vesicle composed of DOPE (80 %) and DOPS (20 %) in the unit cell and let it fuse with its periodic image. Both of these approaches facilitate formation of the stalk without enforcing its shape. The simulations used the Amber99sb force field for peptides(43) and ions and the SLIPIDS lipid force field(44, 45) and TIP3P water(46). In addition we used the ILDN dihedral modification(47) in Setup 1 and Dang K^+(48) ions throughtout. All bonds were constrained using the LINCS algorithm(49), a timestep of 2 fs was employed. The box pressure was held to 1 bar, using a Parrinello Rahman barostat(50) with semi-isotropic pressure coupling and a time constant of 10 ps. The temperature was maintained using a Nosé-Hoover thermostat(51, 52) with a timeconstant of 500 fs and separate temperature groups for the lipids and the rest of the system. As in the first simulation setup, long-range electrostatics was accounted for using the particle-mesh Ewald (PME)(53) method. The GROMACS 5(54) software package was used for all simulations.

ACKNOWLEDGMENTS. P.Jun. acknowledges support from the Czech Science Foundation (grant no. 16-01074S). C.A. thanks the German Academic Exchange Service (DAAD) for support via a P.R.I.M.E. fellowship and the Minerva foundation for a postdoctoral fellowship. A.M. acknowledges Magnus Ehrnrooth foundation, Finland for funding. R.S. and M.H. acknowledge the Czech Science Foundation (grant no. 17-03160S). Allocation of computer time from the CSC Finland is appreciated. VH, RR and CZ acknowledge the use of the cryoTEM in the Department of Molecular Cell Anatomy, Univ. Regensburg, headed by Ralph Witzgall. We acknowledge the Imaging Methods Core Facility at BIOCEV, Faculty of Sciences, Charles University, institution supported by the Czech-BioImaging large RI project (LM2015062 and CZ.02.1.01/0.0/0.0/16_013/0001775 funded by MEYS CR) for their support with obtaining imaging data presented in this paper. We thank Aleš Bendar and Markéta Dalecká, Mario Vazdar, Daniel Harries, Šarka Pokorná, Uri Raviv, and Lea Fink for discussions and technical assistance.

References

1. Bechara C, Sagan S (2013) Cell-penetrating peptides: 20 years later, where do we stand? *FEBS Lett.* 587(12):1693 – 1702.
2. Wadia JS, Stan RV, Dowdy SF (2004) Transducible tat-ha fusogenic peptide enhances escape of tat-fusion proteins after lipid raft macropinocytosis. *Nat. Med.* 10(3):310–315.
3. Pae J, et al. (2014) Translocation of cell-penetrating peptides across the plasma membrane is controlled by cholesterol and microenvironment created by membranous proteins. *J. Control. Release* 192(0):103 – 113.
4. Mishra A, et al. (2011) Translocation of hiv tat peptide and analogues induced by multiplexed membrane and cytoskeletal interactions. *Proc. Natl. Acad. Sci. USA* 108(41):16883–16888.
5. Lamazière A, et al. (2007) Non-metabolic membrane tubulation and permeability induced by bioactive peptides. *PLoS ONE* 2(2):e201.
6. Sun D, Forsman J, Lund M, Woodward CE (2014) Effect of arginine-rich cell penetrating peptides on membrane pore formation and life-times: A molecular simulation study. *Phys. Chem. Chem. Phys.* 16(38):20785–20795.
7. Herce HD, Garcia AE, Cardoso MC (2014) Fundamental molecular mechanism for the cellular uptake of guanidinium-rich molecules. *J. Am. Chem. Soc.* 136(50):17459–17467.

8. Herce HD, Garcia AE (2007) Molecular dynamics simulations suggest a mechanism for translocation of the hiv-1 tat peptide across lipid membranes. *Proc. Natl. Acad. Sci. USA* 104(52):20805–20810.
9. Ciobanasu C, Siebrasse JP, Kubitscheck U (2010) Cell-Penetrating HIV1 TAT Peptides Can Generate Pores in Model Membranes. *Biophys. J.* 99(1):153–162.
10. Zasloff M (2002) Antimicrobial peptides of multicellular organisms. *Nature* 415:389–395.
11. Lee MT, Sun TL, Hung WC, Huang HW (2013) Process of inducing pores in membranes by melittin. *Proc. Natl. Acad. Sci. USA* 110(35):14243–14248.
12. Schow EV, et al. (2011) Arginine in membranes: The connection between molecular dynamics simulations and translocon-mediated insertion experiments. *J. Membr. Biol.* 239(1):35–48.
13. Marks JR, Placone J, Hristova K, Wimley WC (2011) Spontaneous membrane-translocating peptides by orthogonal high-throughput screening. *J. Am. Chem. Soc.* 133(23):8995–9004. PMID: 21545169.
14. Guterstam P, et al. (2009) Elucidating cell-penetrating peptide mechanisms of action for membrane interaction, cellular uptake, and translocation utilizing the hydrophobic counter-anion pyrenebutyrate. *BBA - Biomembranes* 1788(12):2509 – 2517.
15. Maniti O, Piao HR, Ayala-Sanmartin J (2014) Basic cell penetrating peptides induce plasma membrane positive curvature, lipid domain separation and protein redistribution. *Int. J. Biochem. Cell Biol.* 50(0):73 – 81.
16. Swiecicki JM, et al. (2015) Accumulation of cell-penetrating peptides in large unilamellar vesicles: A straightforward screening assay for investigating the internalization mechanism. *Peptide Science* 104(5):533–543.
17. Katayama S, et al. (2013) Effects of pyrenebutyrate on the translocation of arginine-rich cell-penetrating peptides through artificial membranes: Recruiting peptides to the membranes, dissipating liquid-ordered phases, and inducing curvature. *BBA - Biomembranes* 1828(9):2134 – 2142.
18. Gestin M, Dowaidar M, Langel Ü (2017) *Uptake Mechanism of Cell-Penetrating Peptides*, eds. Sunna A, Care A, Bergquist PL. (Springer International Publishing, Cham), pp. 255–264.
19. Chen YA, Scheller RH (2001) Snare-mediated membrane fusion. *Nat. Rev. Mol. Cell Biol.* 2(2):98–106.
20. Chen YA, Scales SJ, Patel SM, Doung YC, Scheller RH (1999) {SNARE} complex formation is triggered by ca2+ and drives membrane fusion. *Cell* 97(2):165 – 174.
21. Babai N, Kochubey O, Keller D, Schneggenburger R (2014) An alien divalent ion reveals a major role for ca2+ buffering in controlling slow transmitter release. *J. Neurosci.* 34(38):12622–12635.
22. Papahadjopoulos D, Vail W, Pangborn W, Poste G (1976) Studies on membrane fusion. ii. induction of fusion in pure phospholipid membranes by calcium ions and other divalent metals. *Biochim. Biophys. Acta* 448(2):265 – 283.
23. Brock TG, Nagaprakash K, Margolis DI, Smolen JE (1994) Modeling degranulation with liposomes: Effect of lipid composition on membrane fusion. *J. Memb. Biol.* 141(2):139–148.
24. Schneggenburger R, Rosenmund C (2015) Molecular mechanisms governing Ca2+ regulation of evoked and spontaneous release. *Nat. Neurosci.* 18:935.
25. Kozlovsky Y, Kozlov MM (2002) Stalk model of membrane fusion: Solution of energy crisis. *Biophys. J.* 82(2):882 – 895.
26. Zhao WD, et al. (2016) Hemi-fused structure mediates and controls fusion and fission in live cells. *Nature* 534(7608):548–552.
27. Yang ST, Zaitseva E, Chernomordik LV, Melikov K (2010) Cell-penetrating peptide induces leaky fusion of liposomes containing late endosome-specific anionic lipid. *Biophys. J.* 99(8):2525–2533.
28. Tünnemann G, et al. (year?) Livecell analysis of cell penetration ability and toxicity of oligoarginines. *J. Peptide Sci.* 14(4):469–476.
29. Mitchell D, Steinman L, Kim D, Fathman C, Rothbard J (2000) Polyarginine enters cells more efficiently than other polycationic homopolymers. *J. Peptide Res.* 56(5):318–325.
30. Hu Y, Sinha SK, Patel S (2015) Investigating hydrophilic pores in model lipid bilayers using molecular simulations: Correlating bilayer properties with pore-formation thermodynamics. *Langmuir* 31(24):6615–6631.
31. Churchward MA, et al. (2008) Specific lipids supply critical negative spontaneous curvature—an essential component of native ca2+-triggered membrane fusion. *Biophys. J.* 94(10):3976 – 3986.
32. Kachar B, Fuller N, Rand R (1986) Morphological responses to calcium-induced interaction of phosphatidylserine-containing vesicles. *Biophys. J.* 50(5):779 – 788.
33. Hirose H, et al. (2012) Transient focal membrane deformation induced by arginine-rich peptides leads to their direct penetration into cells. *Mol. Ther.* 20(5):984–993.
34. Marrink SJ, , Mark AE (2003) The mechanism of vesicle fusion as revealed by molecular dynamics simulations. *J. Am. Chem. Soc.* 125(37):11144–11145.
35. Martens S, Kozlov MM, McMahon HT (2007) How synaptotagmin promotes membrane fusion. *Science* 316(5828):1205–1208.
36. Martens S, McMahon HT (2008) Mechanisms of membrane fusion: disparate players and common principles. *Nat. Rev. Mol. Cell Biol.* 9(7):543–556.
37. Parmryd I, Önfelt B (2013) Consequences of membrane topography. *FEBS J.* 280(12):2775–2784.
38. Robison AD, et al. (2016) Polyarginine interacts more strongly and cooperatively than polylysine with phospholipid bilayers. *J. Phys. Chem. B* 120(35):9287–9296. PMID: 27571288.
39. Allolio C, Haluts A, Harries D (2018) A local instantaneous surface method for extracting membrane elastic moduli from simulation: Comparison with other strategies. *Chem. Phys.* in press:–.
40. Angelova MI, Soléau S, Méléard P, Faucon F, Bothorel P (1992) *Preparation of giant vesicles by external AC electric fields. Kinetics and applications*, eds. Helm C, Lösche M, Möhwald H. (Steinkopff, Darmstadt), pp. 127–131.
41. Schindelin Johannes, et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods* 9:676–682.
42. Kremer JR, Mastronarde DN, McIntosh J (1996) Computer visualization of three-dimensional image data using imod. *J. Struct. Biol.* 116(1):71 – 76.
43. Hornak V, et al. (2006) Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* 65(3):712–725.
44. Jämbeck JPM, Lyubartsev AP (2012) Derivation and systematic validation of a refined all-atom force field for phosphatidylcholine lipids. *J. Phys. Chem. B* 116(10):3164–3179.
45. Jämbeck JPM, Lyubartsev AP (2012) An extension and further validation of an all-atomistic force field for biological membranes. *J. Chem. Theory Comput.* 8(8):2938–2948. PMID: 26592132.
46. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
47. Lindorff-Larsen K, et al. (2010) Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins* 78(8):1950–1958.
48. Dang LX, Schenter GK, Glezakou VA, Fulton JL (2006) Molecular simulation analysis and x-ray absorption measurement of ca2+, k+ and cl- ions in solution. *J. Phys. Chem. B* 110(47):23644–23654. PMID: 17125322.
49. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) Lincs: A linear constraint solver for molecular simulations. *J. Comp. Chem.* 18(12):1463–1472.
50. Parrinello M, Rahman A (1980) Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.* 45:1196.
51. Nosé S (1984) A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* 5:255–268.
52. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* 31(3):1695–1697.
53. Essman U, et al. (1995) A smooth particle mesh ewald method. *J. Chem. Phys.* 103:8577–8593.
54. Abraham MJ, et al. (2015) Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2:19 – 25.