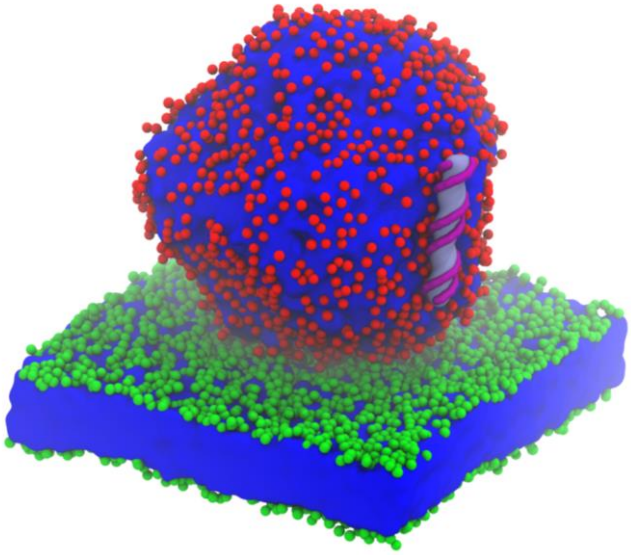
Practical applications for Martini coarse-grained DNA

Coarse-grained double stranded DNA base flipping, and cationic lipid-mediated gene transfer

Bart Bruininks s1826417
BartBruininks@gmail.com
18th of October 2015



Supervisors: dr. Ignacio Faustino, dr. Helgi I. Ingólfsson, Jaakko J. Uusitalo, prof. dr. Siewert J. Marrink

Molecular Dynamics Research Project 2015 (master)

Contents

Abstract

In this project the applicability of the newly developed coarse-grained (CG) Martini double stranded DNA (dsDNA) model was explored in two ways. One being the investigation of the applicability of the usage of the CG dsDNA model for base flipping research. The other being the usage of the model to build and analyze lipoplexes (dsDNA encapsulated cationic).

Base flipping is involved in both expression and maintenance of the genome. It plays an important role for the recognition of damaged DNA and is often needed for modification of the nucleotides – e.g. methylation of cytosine. The exact mechanisms by which these protein-DNA interactions occur is unclear and CG molecular dynamics simulations could provide new insights. By locally uncoupling the elastic network – which is present in the Martini dsDNA model – base(s)/nucleotide(s) should have enough freedom to allow base flipping to occur. Different degrees of uncoupling were tested and visually analyzed.

Uncoupling of the elastic network did allow base flips to occur with the uncoupling of a single nucleotide or nucleotide-pair giving the best results. Uncoupling a single base did not lead to base flips and uncoupling more than one (pair) caused loss of the double helix character. However, uncoupled bases interacted too strong with their complementary bases independent of their directionality. Including directionality in the non-boned potential could improve the behavior of the flipped out bases and the behavior of the model as a whole, though this would increase computational load. For now, one should be extremely careful when using the current CG dsDNA model for base flipping related research.

Complexes of DNA and cationic lipids (lipoplexs) can be used as an alternative for viral-vector mediated gene transfer. They do not cause an immune response and are easier to make. The downside of non-viral vectors is that they often have a low transfection efficiency due to the low escape of genetic material from the endosome. A CG model could allow the study of lipoplex-membrane interactions at an otherwise inaccessible resolution – both in time and size. The development of the lipoplex model was performed in three steps: first a periodic lipoplex was built containing lipids in the inverted hexagonal phase (H_{II}) with the strands of DNA in the hydrated channels. Second, the complex was solvated in water with additional lipids to create the coat around the H_{II} phase inner core of the lipoplex. Third, the fusion of such lipoplexes with an endosomal resembling membrane was explored. Three levels of saturation for the cationic lipid (DOTAP) were tested to study their effect on the fusion speed of the lipoplex with the membrane.

All three steps of building the lipoplex succeeded. The internal DNA distances were close to the values previously observed by small-angle-X-ray diffraction and CG periodic lipoplex models. The solvated lipoplex appeared stable in the microsecond time range. Remarkably, if the channels of the lipoplex were closed from the bulk, exchange of water between channels and the bulk occurred quite readily. No difference in fusion speed between saturated and single unsaturated tails of cationic lipids was found. Double-unsaturated tails in the cationic lipid did seem to have an effect on fusion speed though more simulations are needed to make the trend significant. The solvated lipoplex model has been built and appears to be a new asset for cationic lipid mediated gene transfer research but much more can be analyzed, like the effects of anionic lipids in the membrane on lipoplex-membrane fusion.

Overall the Martini dsDNA force field showed to be accurate if the internal structure was rigid – as was the case for the lipoplex simulations. If internal movement of the dsDNA is mandatory – e.g. base flipping – the model showed to be of less value, though this could be overcome by future improvements.

Introduction

"Everything is made of atoms ... Everything that animals do, atoms do. ... There is nothing that living things do that cannot be understood from the point of view that they are made of atoms acting according to the laws of physics.".[1] Though sub-atomic particles – protons, neutrons, electrons, etc. – have been found which eventually led to the development of the standard model in the mid-1970s, the movement of atoms is still the lion's share for understanding biology. Molecular dynamics is a method which predicts the behavior of such particles over time by solving the Newtonian equations of motion for each particle. The collective of the forces acting on the particles define a force field, which is obtained by carefully parameterizing the bonded and non-bonded interactions of the particles. All forces acting on the particles in the simulation are calculated and propagated to the next frame. With time steps in the femtosecond range frame by frame is calculated – to create a stepwise approximation of atomistic and/or molecular movement. One of the first biological macromolecules for which molecular dynamics simulations were used was the bovine pancreatic trypsin inhibitor (BPTI). The simulation took place in vacuum and lasted for 9.2 picoseconds using a crude molecular mechanics potential. It showed that the interior of a protein was fluid-like instead of a rather rigid body.[2] Demonstrating that molecular dynamics can be used to gain insights in otherwise inaccessible magnitudes and timescales of biological systems and processes.[3]

38 years later computing power has been multiplied by roughly a billion creating the opportunity to study larger systems over longer periods of time.[4] However, for the study of biological systems this increase in computational power was necessity, not luxury. Their generally large sizes and relatively slow dynamics – timescales in the range of microseconds or more are huge for all atoms (AA) simulations – make them expensive with respect to computational load. Thus it is of major importance to design the simulations in an efficient manner with respect to computational load e.g. by the use of periodic boundary conditions. Another approach to reduce system complexity – and thus computational load – is to reduce the resolution of the model by grouping atoms together. A side effect of the grouping of atoms is smoothening of the energy landscape. This loss of resolution of the energy landscape is considered to be a downside, though it speeds up the effective simulation time and helps to study other regions of conformational space. The time steps of integration can also be increased due to the lower frequencies of the bonded interactions – and in lesser amount the non-bonded interactions – which further increases the effective simulation time. This so called coarse-graining (CG) allows the study of many larger systems such as the thylakoid membrane and protein-mediated vesicle fusion.[5,6]

One such CG force field is Martini – named after the Martini tower and also the nickname of the city where it was mainly developed: Groningen.[7,8] Martini was used in all simulations performed in this report. The Martini force field distinguishes itself from other CG force fields by focusing on reproducing the experimental thermodynamic properties such as oil/water partitioning coefficients of its CG beads, allowing for a more general and less parameterization state-dependent modelling. Focusing on oil/water partitioning coefficients for a biological model makes sense, because the main driving forces in many biological processes – such as bilayer formation – are based on the interaction between polar and non-polar groups. The Martini force field roughly maps 4 heavy atoms onto 1 CG bead. It makes a distinction between charged (Q), polar (P), nonpolar (N) and apolar (C) beads allowing the CG bead to generally resemble the chemical group it represents. For the charged and nonpolar beads a distinction is made between particles that can act as: a hydrogen bond acceptor ($_a$), donor ($_d$), acceptor and donor ($_{ad}$), or those which have no hydrogen bonding capabilities at all ($_0$). The polar and apolar beads have an index indicating their polar affinity (1-5). For ring particles smaller (s) bead types have been developed which have a comparable but slightly different mapping philosophy, focusing on the preservation of the ring symmetry. Though the Martini force field was initially developed for modelling of lipids, it has been extended to other biomolecules such as carbohydrates, proteins and DNA.[9–12] Tiny beads – even smaller than the s beads used for the mapping of ring systems – were added to allow the tight packing of the bases inside the double helix of the DNA. The backbone of the CG-DNA is represented by one normal

sized bead for the phosphate and two small beads for the sugar moiety. The pyrimidines and purines are represented by three and four tiny beads respectively.

CG-DNA base flipping

This report will address two processes involving CG double-stranded (ds) DNA. The first topic is CG-dsDNA base flipping. Base flipping is thought to be of importance for many protein-DNA interacting processes and is generally described as a single base leaving the internal structure of the double helix, flipping out roughly 180° rotating over the axes of the backbone which is perpendicular to the base pair plane. It is unknown if the base flips out first and then interacts with the protein, or if the protein assist in the base flipping – or a combination of both. Previous AA studies showed the tendency of certain bases to base flip and the pathways possible to do so. Though these studies gave a high resolution on the actual base flip event itself more macroscopic properties like the life time of a flipped out base and the influence proteins can have are still largely unknown.[13] Given that many properties of base flipping falls into the microsecond or millisecond time scale, the AA simulation of this process is currently unreachable.[14] Studying base flipping at the CG level could offer the possibility to simulate milliseconds, even in the presence of protein and bulk solvent. Although the CG-dsDNA force field was designed with an elastic network to keep the two DNA strands together in their B-DNA helical conformation, local uncoupling of this elastic network will allow the base flip to occur without melting the dsDNA. I have studied the DNA base flipping process using the CG-dsDNA Martini model.

Cationic lipid-mediated gene transfer

The second topic which will be addressed is cationic lipid-mediated gene transfer (CLMGT). This project also made use of the new Martini CG-dsDNA force field, but the imposed elastic network was left untempered. Gene transfer is the infection of a host cell with non-native genetic material. In medical sciences it is used for gene therapy. Gene therapy uses genetic material as a drug to treat disease. Gene therapy is based on the introduction of a short oligonucleotide molecule in a cell to alter the targeted protein expression, the transcription of the corresponding mRNA or to mutate genes.

Since the genetic information is stored within each cell the alien genetic material has to cross the cell membrane. This is problematic due to the charged nature of DNA and RNA – one negative charge per nucleotide. Thus a vector must be used to cloak the DNA, shielding its negative charges from the hydrophobic core of the lipid bilayer. A distinction is made between viral and non-viral vectors. Viral vectors are altered viruses, which have been made non-pathogenic. The alien material is cloned into the virus at the site which is normally for the insertion of viral genes. Though the use of viral vectors is common in gene therapy due to their transfection efficiency they have a risk of interacting with the immune system – triggered by the viral coat/envelope or the non-native proteins expressed by the infected cell.[15] Other disadvantages of using viral vectors are: carcinogenesis, their usual broad tropism, production of the viral vector, and limited capacity for genetic material.[16]

Non-viral vectors do not make use of a modified virus to transfect the target cell with the therapeutic genetic material. Examples of non-viral gene methods for transfection are the injection with naked DNA, electroporation, sonoporation and magnetofection. However, these techniques are limited for systemic gene delivery in humans. Thus a wide range of synthetic vectors have been developed. These vectors are often polymers or lipids with a positive charge, inorganic nanoparticles and membrane-penetrating peptides.[16] They have a low immunogenicity due to the fact that patients do not have a native immunity against them. The complex of genetic material with lipids is called a lipoplex (Figure 1). Another advantage is that they are – in general – easy to synthesize and can carry larger loads of genetic material.[17,18] Drawbacks of non-viral vectors are their often low gene-transfer efficiency and their possible toxicity.[18,19] In this research coarse-grained modelling was used to investigate CLMGT of oligonucleotides at the molecular level.

Though the challenges for *in vivo* delivery reside at many levels – systemic clearance, toxicity, targeting, etc. – it is thought that there is a more general barrier for CLMGT. It is has been shown that particles smaller than 200 nm in diameter are mainly internalized via the clathrin-coated pits.[20] After internalization the particles are captured in an endosome. The endosome is targeted towards the lysosome where its content will be broken down – in this case the lipoplex and its bound genetic material. For successful transfection the genetic material inside the lipoplex has to escape the early/late endosome before it is fused with the lysosome. It is this breaking free out of the endosome which is thought to be the bottleneck for CLMGT.[21] It appears that bilayer destabilizing lipids favoring the inverted hexagonal phase (H_{II}) tend to increase fusion of the lipoplex and the endosomal membrane, promoting the escape of genetic material from the endosome and increase the nuclear delivery of the genetic material.[22,23]

Optimization of this escape has been mainly empirical, for the exact molecular mechanism by which the genetic material leaves the endosome is unknown. Insight in the molecular mechanism of lipoplex endosomal-like bilayer fusion might allow the intelligent design of new cationic lipids focusing on their endosome escaping capabilities. Previous molecular dynamics models investigated the lipoplex under periodic conditions. Their results showed that lipoplex modelling was well in line with experimental results.[24] I have investigated lipoplex-bilayer fusion using coarse-grained molecular dynamics simulations.

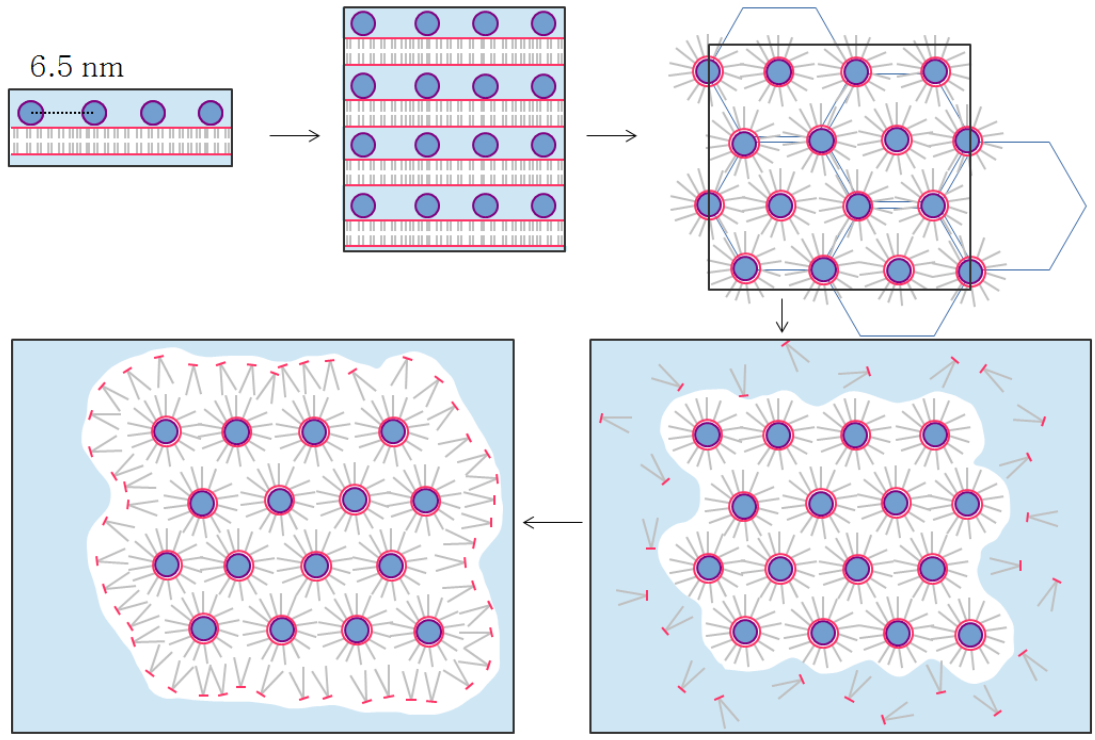


Figure 1 Building the lipoplex. A bilayer with DNA on top, was stacked onto itself. Due to under-hydration the stacked bilayers quickly form the H_{II} phase. With the channels of the H_{II} being perpendicular to the image. Solvation of the period lipoplex and the addition of lipids around the naked lipoplex result in the final solvated state. Red depicts lipid headgroups. Grey is used for lipid tails and purple is DNA. Blue shows where the solvent/ions can be present.

Methods

The following workflow was used for all molecular dynamics simulations: First a coarse-grain model was obtained/built for each individual component. Then the components were combined to form the initial structure which was subsequently energy minimized, equilibrated and ran. All simulations were ran by GROMACS 4.6.7[25] and made use of the Martini force field[12]. For energy minimization the steepest descent algorithm was used. Equilibration and the final run used a leap-frog algorithm for the integration of Newton's equations of motion. Unless specifically specified otherwise the temperature coupling made used was V-rescale with a reference temperature of 320 K and an interval of 1.0 ps. The standard values for the Berendsen pressure coupling were a reference pressure of 1 bar, with 3e-4 bar^{-1} compressibility at and interval of 5 ps. [26] The Lennard-Jones interactions were shifted between 0.9-1.2. The Coulomb interactions were shifted from 0-1.2, which were uniformly screened by a relative electric permittivity (ε_r = 15). A dodecahedron shaped box was used for the periodic boundary conditions.

DNA base flipping simulations

The initial AA structure of a 12 base pair (bp) piece of dsDNA – 1BNA d(CpGpCpGpApApTpTpCpGpCpG)[27] – was coarse-grained with a stiff elastic network (RUBBER_F=500) using *martinize-dna*[28].[11] Specific elastic bonds were severed to allow base flipping (Table 1). Each base flipping experiment contained solvated DNA with counter-ions and roughly 50 mM of NaCl (1 piece of dsDNA, 1222 waters, 25 Na^+ and 3 Cl^-). The time step of integration was set to 5 and 10 fs for the final run of 1 µs. Parrinello-Rahman coupling[29,30] was used for isotropic pressure coupling with the standard settings used for the Berendsen pressure coupling. In total each specific type of uncoupling of the base(s)/nucleotide(s) was simulated 25 times – 5 different initial configurations before energy minimization of which each was initiated with 5 different random seeds used for the initial velocities during equilibration. The cut-off for bonded interactions was set to 1.2 nm. Crashed simulations were not reinitiated.

Building the lipoplex

The creation of the periodic lipoplex was based on an article published by Corsi *et al.*[24] First a symmetrical membrane of lipids was made with DOTAP and DOPE in a 1:4 ration using the *insane* program (130*111 Å).[31] Parallel to the membrane two strands of DNA were placed with a minimum distance to the membrane slightly below 5 Å and an interspacing of 65 Å with regard to each other (see Figure 1). The AA-dsDNA contained 24 bp – two times the 1BNA strand used for the base flipping simulations – and was coarse grained with martinized-dna.[11,28] Water was added to the box at a ratio of 2 AA waters per lipid. Since every CG-W represents 4 AA waters the CG-W to lipid ratio was 2 lipids per water. Two of these 1membrane-2dna boxes were stacked along their z-axes on top of each other using GROMACS *genconf* tool.[25] The system was charge neutralized using the GROMACS *genion* tool.[25] The final simulation contained dsDNA, cationic and helper lipids, water and counter ions (3 12bp-dsDNA, 760 DOPE; 188 DOTAP; 4 dsDNA; 474 W and 4 CL^-). A time step of 3, 5 and 10 fs was used for pre-equilibration, equilibration – 10 ns – and the production run – for 1 µs. During pre-equilibration isotropic pressure coupling was used. For equilibration and the production run isotropic pressure coupling was used – x, y, and z were set to the standard. Orthorhombic periodic boundaries were used.

The periodic lipoplex was used as a starting configuration to create the solvated lipoplex. Waters and ions were removed and the lipoplex was made whole using trjconv – using the -pcb whole flag.[25] After the individual periodic movement of some lipids which flipped to the wrong side of the periodic boundary – not oriented with their headgroups towards the channels – the complex was either directly solvated – naked lipoplex –, or coated with an extra layer of lipids which were placed closely around the lipoplex using *Packmol* – a total of 226 DOTAP and 907 DOPE was added.[32] Then it was solvated using genbox.[25] The system was minimized, pre-equilibrated at high pressure (1000 bar reference pressure; 3 fs time step; 300 ps total length), followed by the same pre-equilibration at normal pressure (1 bar; 3fs; 300 ps), equilibration (1 bar; 5 fs; 10 ns) and a production run (1 bar; 10 fs; 1 µs). Around 68000 W beads were

added for both the naked and coated lipoplex. Pressure coupling was the same as for the periodic lipoplex.

Lipoplex-membrane fusion

The membrane to be fused with the lipoplex was made using insane (DPPC:POPC:DAPC:DPPS:POPS:DAPS in a 4:8:4:1:2:1 ratio; 22x25x9 nm).[31] After initial relaxation of the membrane the lipoplex was placed above the membrane with at least one layer of water in between. The periodic boundaries were orthorhombic and pressure coupling was set to semi-isotropic in the plane of the membrane – standard settings for both xy and z. The cationic lipids used for the lipoplexes were DPTAP DOTAP and DITAP – all 18 C carbon tails but subsequently with none, one or two unsaturated carbon-carbon bonds in each tail where TAP refers to the tri-methylamine headgroup (Figure 2). The bonded parameters were conform the Martini standard.[31] The ratio of lipids, starting configuration and orientation was the same for each fusion simulation. Each fusion experiment was performed 10 times per saturation level.

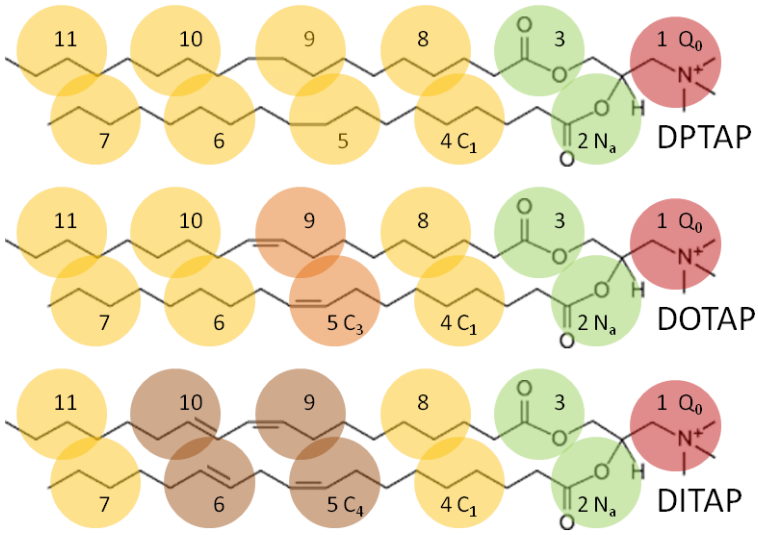


Figure 2 Cationic lipid mapping. DPTAP, DOTAP and DITAP. Bead type is shared between all beads with the same color and the numbering is that of the itp file which was conform the Martini lipid standard.[31]

Analysis

Visual analysis was performed using VMD.[33] Python was used as the main programming language for the scripts and MDAnalysis was used for the helical parameter calculations together with a script kindly provided by dr. I. Faustino.[34,35] The GROMACS toolkit was used for the manipulation of the associated GRO and XTC files.[25] Wolfram Mathematica was used for the statistical analysis of the survival data (*LogRankTest* and exponential fitting).[36] D-spacing data was acquired using VMD, the distance between the helical extremes was measured after 1 µs for 100 ns. This was done for each saturation level of the lipoplex.

Results

DNA base flipping

Stability of the systems

For the study of CG DNA base flipping the new Martini DNA force field was used.[11] Due to the elastic network (EN) which is present in the current dsDNA force field base flipping cannot take place. By uncoupling the EN locally, base flips should be possible.

Table 1 The names and corresponding positions of the uncoupled residues from the 12 bp DNA. Paired indicates if the opposing base was uncoupled too.
Chain sequence: d(CpGpCpGpApApTpTpCpGpCpG)

Residue(s)	position(s)	paired (Y/N)
a	18	N
att	6-8	N
c	21	N
cga	3-5	N
g	4	N
gaa	4-6, 19-21	Y
gc	4, 21	Y
gct	20-22	N
t	7	N
ta	6, 19	Y
taa	17-19	N

For each uncoupling described in Table 1 twenty-five simulations were run for 1 µs to study the stability of the simulation and the behavior of the DNA. This was done for different time steps and led to a total amount of 1150 simulations. For the 10 µs simulations more than 60 % did not reach 1 µs. Some special cases *never* reached 1 µs (rgaapb, rgctb and rtaab), all of which have more than one nucleotide uncoupled from the EN.

For the 5 fs time step simulations (Figure A- 1) the stability landscape looks completely different. Almost no crashes occurred – less than 20 % – and most uncoupling setups showed no crashes at all. The only simulations which crashed were ra18b, rgaapb, rgctb, rtaab and rtapb. Again there seems to be a correlation between uncoupling the nucleotide as a whole and the stability of the simulation.

In both the 5, 10 fs time step simulations the initial configuration and random velocity did not seem to have an effect on the ratio of simulations that reached 1 µs.

Base flipping

To study the possible base flipping events in the 1150 simulations with a maximum length of 1 µs would have been a tedious job if some form of automated detection would not have been possible. However, using the helical parameters a script was made which flagged each frame in which two opposing bases would have a stretch of more than 2 Å – the helical parameters are covered in depth in a paper by Lu and Olson.[37] Using this form of automated detection 31 events out of the 571 initiated simulations were found for the 10 fs time step simulations – 4 never passed the pre-equilibration. For the 5 fs time step simulations 67 events were found out of the 571 initiated simulations. An overview of the detected events per uncoupling type can be found in the appendix (Figure A- 2).

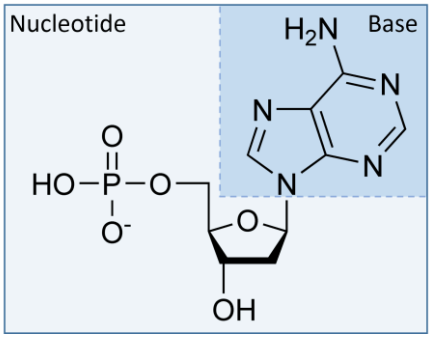


Figure 3 Adenine. The nucleotide is shown in light blue, dark blue was used to mark just the base.

They can be categorized into three groups. The first being a flipped out base in solution – a clean flip. Second is a flip which starts to interact with the grooves – both the major and minor groove can be interacted with. Third is the local loss of the double helix character (Figure 4). In general it can be said that more uncoupling causes bigger distortions in the double helix character. Uncoupling of a single nucleotide did not result in loss of the double helix character, though it does give enough freedom for a base flip to occur. Detected base flips sorted by their uncoupling and integration time step are depicted in Figure A- 3.

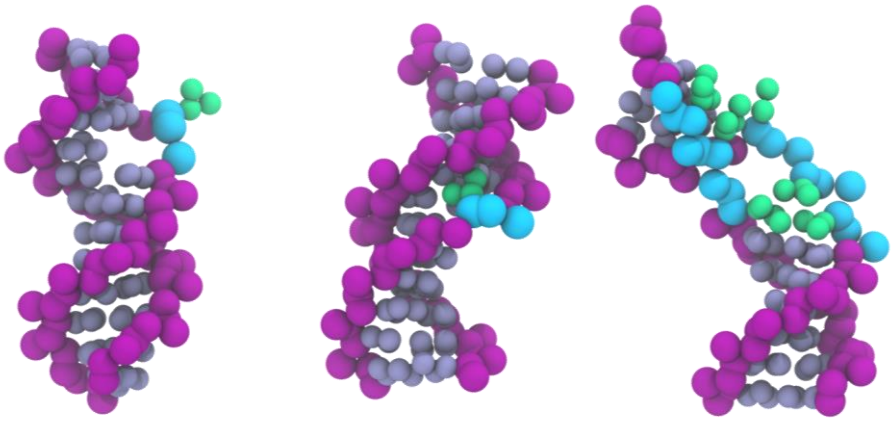


Figure 4 DNA base flip conformations. From left to right: a clean flip, groove-base interaction, local distortion of the double helix. Purple show the backbone beads with an intact EN. Ice-blue depicts the bases with an intact EN. Turquoise was used for the uncoupled backbone beads, and seafoam shows bases which were uncoupled from the EN.

Cationic lipid-mediated gene transfer

Building the lipoplex

Escaping from the endosome – after uptake of the lipoplex by the cell – is thought to be one of the major hurdles for CLMGT.[21] New insights in the exact molecular manner of lipoplex-endosome interaction could help with the rational design of lipids which escape more readily from the endosome. The use of a molecular dynamics model allows the study of this escape at a resolution and timescale inaccessible to empirical studies. Previous work has been done by Corsi *et al.* to describe the behavior of a periodic lipoplex in its inverted hexagonal phase (Figure 5) describing the channels with and without the presence of DNA.[24,38] By using the lipid ratio and level of hydration described in their work it was relatively easy to attain a periodic lipoplex. DOTAP was the cationic lipid and DOPE the helper lipid to promote the inverted hexagonal phase. The periodic lipoplex that formed had a clear hexagonal unit cell and showed a d-spacing – the inter-DNA distance – of roughly 6.7 Å (Table 2).

Solvation of the lipoplex was accomplished by cutting out a piece of the periodic lipoplex which was solvated in a bigger box. Extra lipids were added in close proximity of the periodic fragment to coat the naked lipoplex. This created an outer leaflet which resembles the outer leaflet of a bilayer. The inner leaflet however is very different from the one found in a bilayer (Figure 5). If more lipids than needed were used to coat the lipoplex extra channels/pockets would form inside the lipoplex without loss of the hexagonal periodicity of the inner core. Upon solvation no difference in d-spacing of the DNA was observed (Table 2). After the addition of the extra lipids and an equilibration of 1 μs, all channels of the inverted hexagonal phase of the inner core of the lipoplex were no longer continuous with the bulk solvent – closed. Remarkably, water seems quite capable to enter and leave the channels through the hydrophobic region of the lipoplex – visual analysis with VMD – causing a slight swelling of the channels after solvation. Ions seem to be trapped in the channels.

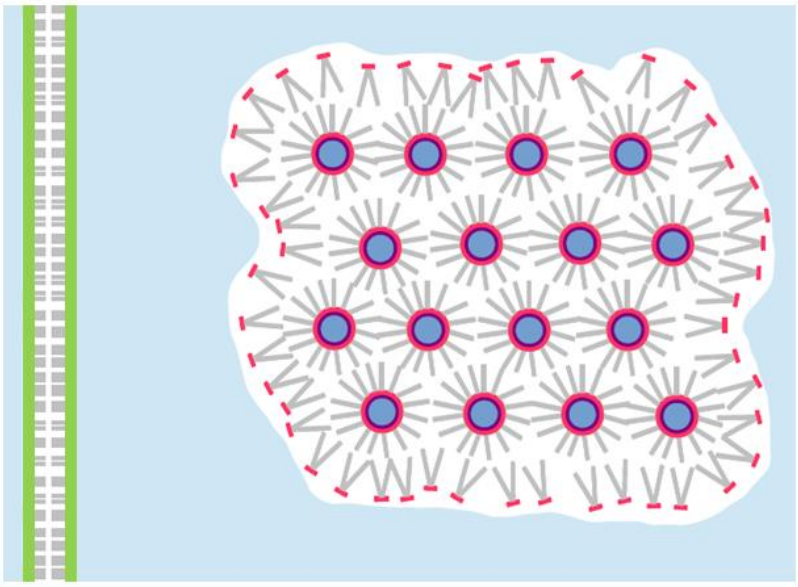


Figure 5 Bilayer and lipoplex (H_{II}) conformation. Green depicts the headgroups of lipids in the bilayer. Red are the headgroups of lipids which form the lipoplex. Grey represents the lipid tails and purple is used for DNA. Both the bilayer and lipoplex are solvated.

Lipoplex-membrane fusion

The last step of the lipoplex simulations was the actual fusion of such a particle with a membrane. Several setups were tested including long range electrostatics and assymetrical bilayers. Due to the increased computational load of long range electrostatics and an already quite large system size – roughly 135,000 beads – these interesting setups were saved for future research. In the end a symmetrical bilayer – 1:4 XXPS/XXPC (1:2:1 DAXX/DPXX/POXX) – was used with normal Martini water. Three levels of saturation were tested for the XXTAP in the lipoplex – saturated, single unsaturated and double unsaturated tails. The same initial lipoplex was used for all different saturation levels and after changing the saturation the complex was equilibrated briefly – 100 ns. To validate the usage of the same initial structure to construct the lipoplexes with varying amounts of saturation, the overall structure of the lipoplexes was analysed over 1 μs. No major reorganization of the lipoplex took place over this period of time (Figure 6). All fusion experiments started from the same position and orientation of the lipoplex with regard to the membrane.

The different levels of saturation of the XXTAP did not have an effect on the d-spacing of the DNA inside the lipoplex (Table 2). However, it did change the average positions of the lipid tails, causing them to be more hydrated (Figure 6).

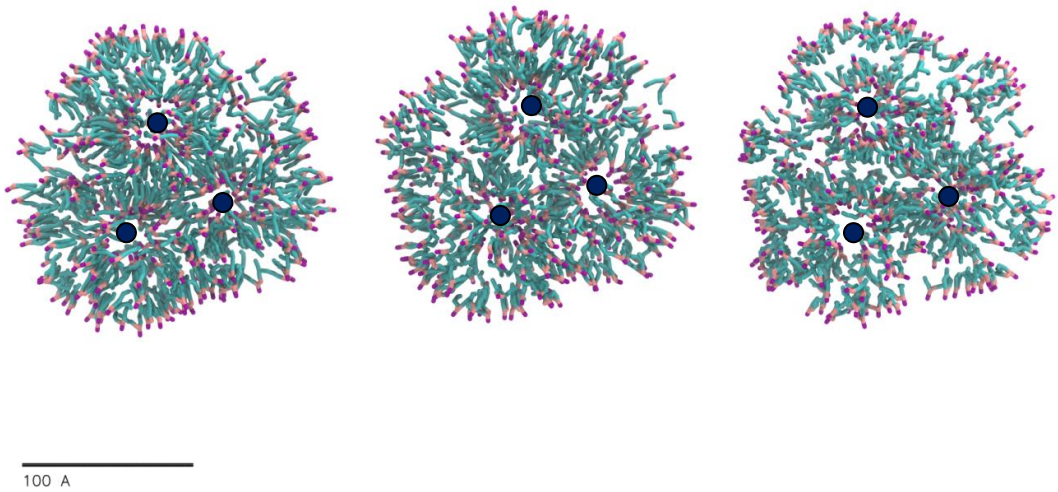


Figure 6 Lipoplex XXTAP at varying saturation levels. From left to right: DPTAP, DOTAP, DITAP. Lipid positions have been averaged over 10 ns after alignment of the complete lipoplex with respect to its internal DNA – channels are depicted as dots.

For each saturation level ten simulations of 4 μs were initiated to study the time and manner at which fusion took place. The survival of each saturation level is graphed in (Figure 7). From the survival data there seems to be no difference between the fusion of a lipoplex with the membrane containing saturated or single unsaturated tails using only ten fusion events (P-value < 0.001 for them being the same). Merger of the lipoplex with the membrane *seems* to be faster for DITAP – which has double unsaturated tails – in comparison to DPTAP and DOTAP (P-value 0.190 and 0.205 for them being different). P-values were obtained by comparing the unbiased raw data using the *LogRankTest*.

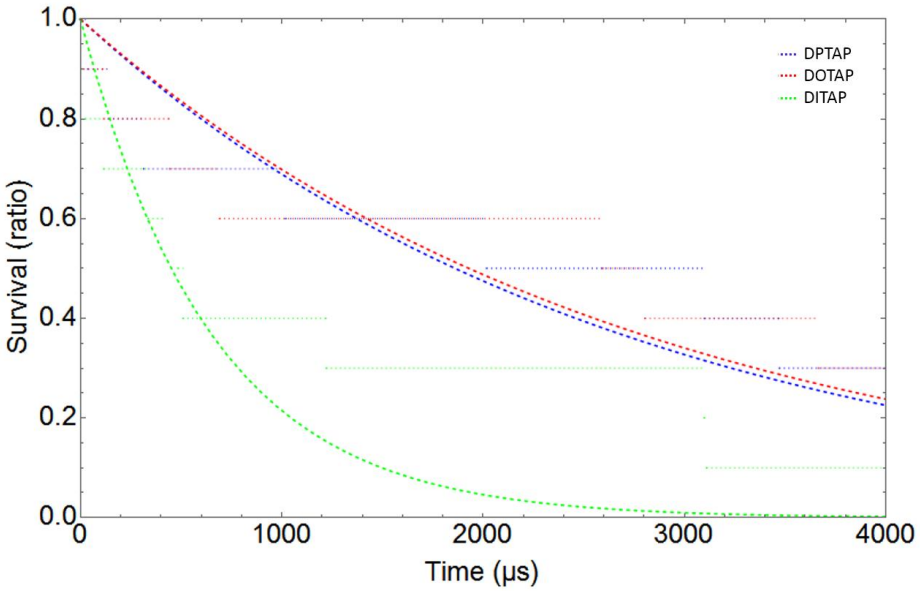


Figure 7 Survival of DPTAP, DOTAP and DITAP containing lipoplexes upon fusion with a bilayer. Dotted lines represent the raw survival data, the dashed line was used to depict the single exponential fit. N = 10; Lambda = [-3.7E-4±6.4E-05 (DPTAP), -3.6E-4±7.0E-5 (DOTAP), -1.5E-3±2.9E-4 (DITAP)]

Table 2 D-spacing DNA in the H_{II} phase of the lipoplex.

System	AVG(nm)	STD(nm)
Periodic	6.7	0.9
DPTAP	6.8	0.9
DOTAP	6.8	0.9
DITAP	6.8	0.9
PERIODIC[24]	6.4	-
SAXS[24]	6.1	-

Conclusion/Discussion

DNA base flipping

DNA base flipping is a biological important phenomenon which is involved in gene regulation and preservation. Knowing the mechanism by which proteins can induce or recognize flipped out bases is a major contribution to our understanding of genetic material. In this research the possibility of using the Martini DNA model to study base flipping was explored. The EN was locally released to allow base flipping to occur. Since the EN was put in place with a specific reason – to maintain the double helix conformation – uncoupling too much was expected to distort the helix, where uncoupling too little would prevent the base/nucleotide to ever flip out. Next to multiple uncoupling schemes two time steps of integration were tested – 5 and 10 fs – to maximize sampling time.

The uncoupling of a single base was not found to be enough to allow a base/nucleotide to flip out of the helical core. Uncoupling of a single nucleotide did allow the base to flip out of the helical core without causing deformations in the double helix (Figure 3). The downside is the rather static behavior of the flanking bases. Atomistic simulations have shown that base flipping is a combined change of local and more distant dihedrals in the strand of the flipped base.[13] This complex motion would be impossible if only one nucleotide is uncoupled from the EN. Since we do see base flipping – in this model – if only one nucleotide was uncoupled, the change in the distant dihedral is not mandatory for a CG base flip. Uncoupling a pair of nucleotides also led to base flipping. A flipped out base was rarely observed to restore, though the expected relaxation time lies in the millisecond range and thus far outside our current scope – 1 µs. A flipped out base could interact with both the minor and major groove, preferring the interaction with its complementary base in either the same or the opposing strand. This interaction is probably too strong since the strength of base-base interaction was parameterized for the Watson-Crick base-pairing. To improve the specificity of (complementary) base-base interactions, directionality could be included in the potential. In the case of base flipping this potential could be used only for the interactions with the flip-able base.

The difference of crashed simulations between the 5 and 10 fs integration steps was huge. More than 60 % of the simulations with the larger time step did not reach 1 µs of simulation time. For the 5 fs time step less than 20 % of the simulations crashed. However, it was a select group which crashed almost all the time causing the 20 % – crashes were not evenly distributed. Remarkably the uncouplings causing the crashes were also those which showed base flipping. This can be explained by the fact that I used an rdd of 1.2, while the DNA model is known run best at an rdd of 2 or even higher. At an integration step size of 10 fs beads can move more and get further removed from the bonded partner than the rdd. The fact that there are still well defined crashes for the 5 fs time step simulations probably has something to do with the distortion of the normal base stacking – i.e. base flipping, or the loss of the double helix conformation.

Thus using a larger rdd – 2 or higher – could allow the usage of a bigger time step without affecting overall stability.

To conclude, it seems coarse-grain simulations of base flip events could be studied, but one should be very careful with the level of uncoupling of the EN. Probably one of the biggest hurdles will be to define the base-base interactions in a "better way" – *e.g.* directionality. Doing so would not only improve the flipped out base behavior, but would improve almost all properties of the DNA model and could possible make the EN superfluous. The downside of increasing the accuracy of the interaction potential is the associated increase in computational load. Finding the right balance between the two will always be key in any coarse-grain modelling.

Cationic lipid-mediated gene transfer

Cationic lipid-mediated gene transfer makes use of cationic lipids to transfect cells with non-native genetic material. Cationic lipids can be used a replacement for viral vectors, but struggle with relatively low transfection efficiency. Increasing the transfection efficiency of CLMGT could allow gene therapy to reach its full potency. For cationic lipid-mediated gene delivery the membrane that has to be crossed to enter the cytoplasm is that of the endosome. After endocytosis of the lipoplex the complex is trapped in the endosome which is targeted for destruction in the lysosome. For successful transfection the lipoplex – or at least the genetic material – has to escape the endosome, before it fuses with the lysosome. It has been generally accepted that using lipids which prefer the inverted hexagonal phase improves the transfection efficiency.[23]

In this study we succeeded in building a lipoplex consisting out of DOTAP and DOPE with a 1:4 ratio of roughly 13 nm in diameter and fused it with a lamellar membrane containing 20 % anionic lipids. The d-spacing of the DNA inside the lipoplex was close to that measured by SAXS experiments for real lipoplexes. The internal inverted hexagonal phase of the lipoplex formed at a hydration level of two waters per lipid appeared to be stable after solvation, though the H_{II} channels of the lipoplex did slightly swell, indicating that adding two waters per lipid during the formation of the inverted hexagonal phase results in a slightly underhydrated lipoplex core.

Since this was the first time a model was made which simulated a completely solvated lipoplex it was tempting to see how lipoplex fusion with a membrane takes place. After investigation of multiple fusion events – nanosecond per nanosecond – it became clear that the initial fusion of the outer layer of the lipoplex with the lipoplex facing monolayer of the bilayer follows the same initial steps as vesicle-vesicle fusion.[39] Initial lipid handshaking between the lipids of the outer shell of the lipoplex and the monolayer facing the lipoplex is followed by stalk formation which smoothly transitions in the formation of a diaphragm. After the formation of the diaphragm two things can happen. The first pathway is the linear expansion of the diaphragm which will eventually close on itself. This leads to the formation of the *hemi-fused* state. This state has a water containing bubble between the two layers of the bilayer. In hemi-fusion the outer layer of the lipoplex and the lipoplex facing monolayer become continuous, but the outer layer of the bilayer is still intact. This complex appeared to be quite stable even if a strand of the DNA was caught in the water compartment between the two layers of the membrane. From this state no DNA was ever observed to cross the membrane, though this is probably due to the relatively short length of the simulations – 4 µs. The second pathway which was observed in which the lipoplex can fuse with the membrane, resembles that of the *direct* fusion of two vesicles. Upon formation of the diaphragm the diaphragm expands evenly in all directions in the plane of the bilayer. Lipids from the monolayer facing away from the lipoplex are donating their tails at the center of the diaphragm. This direct fusion of the inverted hexagonal phase with the bilayer then differentiates from vesicle fusion by forming a funnel which is metastable. This funnel is followed by the rotation of the inverted hexagonal phase until the DNA is perpendicular to the membrane interface. The funnel becomes continuous with one of the DNA containing channels resulting in the translocation of one strand of DNA from one side of the membrane to

the other. It is not clear from the current simulations if eventually all DNA will be released from the complex (videos of fusion events can be provided by email).

Since the escape from the endosome is thought to be the rate limiting step in transfection, the speed at which the lipoplex fuses with the membrane – partial or complete – might be a robust indicator of total transfection efficiency. This has been tested by increasing the amount of unsaturation in the lipid tails of the XXTAP in the lipoplex from completely saturated to double unsaturated. This showed that an increase from completely saturated (DP) to single unsaturated (DO) – in both tails – did not have a measurable effect on the fusion speed. However, the double unsaturated DITAP did show an increase in the fusion speed. It is at the moment not clear why a single unsaturation does not cause the fusion to occur more readily compared to the fusion with saturated cationic lipids. Perhaps the difference in fusion speed is not large enough to be accurately measured with only ten fusion events. Due to the fact that increasing the unsaturation of a lipid also increases its preference for the inverted hexagonal phase – which is thought to increase the total transfection efficiency – these results indicate that the model might qualitatively predict the effects of lipid saturation on transfection efficiency.

From this study is seems that coarse-grained modelling can be used to investigate properties of lipoplexes at a resolution and time unavailable previous to this study. Though some general characterization of the model has been done many question remain unanswered, like the importance of the anionic lipids in the membrane and the effect of lipid domains in the bilayer and/or lipoplex. This model could provide an answer to these questions by including more accurate electrostatic interactions and a more complex representation of the endosomal bilayer. I believe that some steps have been made – solvated lipoplex structure, lipoplex membrane interaction and the effects of saturation on these interactions – to a better understanding of the molecular interactions involved in CLMGT which play a major role in the future development of cationic lipids used for gene transfer both *in vitro* and *in vivo*.

Bibliography

1. Feynman RP. *Six Easy Pieces: Essentials of Physics, Explained by Its Most Brilliant Teacher*. (Leighton RB, Matthew L. Sandes, eds.). Helix Books; 1995.

2. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. *Nature*. 1977;267(5612):585-590. doi:10.1038/267585a0.

3. Berendsen HJC, Gunsteren WF. Molecular Dynamics Simulations: Techniques and Approaches. In: *Molecular Liquids*. Dordrecht: Springer Netherlands; 1984:475-500. doi:10.1007/978-94-009-6463-1_16.

4. Mims C. Moore's Law Over, Supercomputing "In Triage," Says Expert. 2012. http://www.technologyreview.com/view/427891/moores-law-over-supercomputing-in-triage-says-expert/.

5. van Eerden FJ, de Jong DH, de Vries AH, Wassenaar TA, Marrink SJ. Characterization of thylakoid lipid membranes from cyanobacteria and higher plants by molecular dynamics simulations. *Biochim Biophys Acta - Biomembr*. 2015;1848(6):1319-1330. doi:10.1016/j.bbamem.2015.02.025.

6. Baoukina S, Tieleman DP. Direct simulation of protein-mediated vesicle fusion: lung surfactant protein B. *Biophys J*. 2010;99(7):2134-2142. doi:10.1016/j.bpj.2010.07.049.

7. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, De Vries AH. The MARTINI force field: Coarse grained model for biomolecular simulations. *J Phys Chem B*. 2007;111(27):7812-7824. doi:10.1021/jp071097f.

8. Periole X, Marrink SJ. The martini coarse-grained force field. *Methods Mol Biol*. 2013;924:533-565.

9. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. The MARTINI coarse-grained force field: Extension to proteins. *J Chem Theory Comput*. 2008;4(5):819-834.

10. López CA, Rzepiela AJ, de Vries AH, Dijkhuizen L, Hünenberger PH, Marrink SJ. Martini coarse-grained force field: Extension to carbohydrates. *J Chem Theory Comput*. 2009;5(12):3195-3210. doi:10.1021/ct900313w.

11. Uusitalo JJ, Ingólfsson HI, Akhshi P, Tieleman DP, Marrink SJ. Martini Coarse-Grained Force Field: Extension to DNA. *J Chem Theory Comput*. 2015;11(8):3932-3945. doi:10.1021/acs.jctc.5b00286.

12. Marrink SJ, Tieleman DP. Perspective on the Martini model. *Chem Soc Rev*. 2013;42(16):6801-6822. doi:10.1039/c3cs60093a.

13. Bouvier B, Grubmüller H. A molecular dynamics study of slow base flipping in DNA using conformational flooding. *Biophys J*. 2007;93(3):770-786.

14. Yin Y, Yang L, Zheng G, et al. Dynamics of spontaneous flipping of a mismatched base in DNA duplex.

Proc Natl Acad Sci U S A. 2014;111:8043-8048. doi:10.1073/pnas.1400667111.

15. Thomas CE, Ehrhardt A, Kay MA. Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet*. 2003;4(5):346-358. doi:10.1038/nrg1066.

16. Yin H, Kanasty RL, Eltoukhy A a, Vegas AJ, Dorkin JR, Anderson DG. Non-viral vectors for gene-based therapy. *Nat Rev Genet*. 2014;15(8):541-555. doi:10.1038/nrg3763.

17. Mintzer MA, Simanek EE. Nonviral vectors for gene delivery. *Chem Rev*. 2009;109(2):259-302. doi:10.1021/cr800409e.

18. Pack DW, Hoffman AS, Pun S, Stayton PS. Design and development of polymers for gene delivery. *Nat Rev Drug Discov*. 2005;4(7):581-593. doi:10.1038/nrd1775.

19. Lv H, Zhang S, Wang B, Cui S, Yan J. Toxicity of cationic lipids and cationic polymers in gene delivery. *J Control Release*. 2006;114(1):100-109.

20. Rejman J, Oberle V, Zuhorn IS, Hoekstra D. Size-dependent internalization of particles via the pathways of clathrin- and caveolae-mediated endocytosis. *Biochem J*. 2004;377(Pt 1):159-169. doi:10.1042/BJ20031253.

21. Hoekstra D, Rejman J, Wasungu L, Shi F, Zuhorn I. Gene delivery by cationic lipids: in and out of an endosome. *Biochem Soc Trans*. 2007;35(Pt 1):68-71. doi:10.1042/BST0350068.

22. Koltover I, Salditt T, Rädler JO, Safinya CR. An inverted hexagonal phase of cationic liposome-DNA complexes related to DNA release and delivery. *Science*. 1998;281(5373):78-81. doi:10.1126/science.281.5373.78.

23. Zuhorn IS, Bakowsky U, Polushkin E, et al. Nonbilayer phase of lipoplex-membrane mixture determines endosomal escape of genetic cargo and transfection efficiency. *Mol Ther*. 2005;11(5):801-810. doi:10.1016/j.ymthe.2004.12.018.

24. Corsi J, Hawtin RW, Ces O, Attard GS, Khalid S. DNA lipoplexes: Formation of the inverse hexagonal phase observed by coarse-grained molecular dynamics simulation. *Langmuir*. 2010;26(14):12119-12125. doi:10.1021/la101448m.

25. 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(3):435-447. doi:10.1021/ct700301q.

26. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola a, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys*. 1984;81(8):3684-3690. doi:10.1063/1.448118.

27. Drew HR. WRM. TT. BC. TS. IK. DRE. Structure of a B-DNA dodecamer: conformation and dynamics. *ProcNatlAcadSciUSA*. 1981;78:2179-2183. doi:6941276.

28. De Jong DH, Singh G, Bennett WFD, et al. Improved parameters for the martini coarse-grained protein force field. *J Chem Theory Comput*. 2013;9(1):687-697. doi:10.1021/ct300646g.

29. Nosé S, Klein ML. Constant pressure molecular dynamics for molecular systems. *Mol Phys*. 1983;50(5):1055-1076. doi:10.1080/00268978300102851.

30. Parrinello M. Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys*. 1981;52:7182-7190. doi:10.1063/1.328693.

31. Wassenaar TA, Ingólfsson HI, Böckmann RA, Tieleman DP, Marrink SJ. Computational Lipidomics with insane : A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J Chem Theory Comput*. 2015;11(5):2144-2155. doi:10.1021/acs.jctc.5b00209.

32. Martinez L, Andrade R, Birgin EG, Martínez JM. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J Comput Chem*. 2009;30(13):2157-2164. doi:10.1002/jcc.21224.

33. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *J Mol Graph*. 1996;14(1):33-38. doi:10.1016/0263-7855(96)00018-5.

34. Python Software Foundation. Python Language Reference, version 2.7. *Python Softw Found*. 2013. http://www.python.org.

35. Michaud-Agrawal N, Denning EJ, Woolf TB, Beckstein O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J Comput Chem*. 2011;32(10):2319-2327. doi:10.1002/jcc.21787.

36. Wolfram Research I. *Mathematica*. Version 10. Champaign, Illinois: Wolfram Research, Inc.; 2015.

37. Lu XJ, Olson WK. 3DNA: A software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res*. 2003;31:5108-5121. doi:10.1093/nar/gkg680.

38. Khalid S, Bond PJ, Holyoake J, Hawtin RW, Sansom MSP. DNA and lipid bilayers: self-assembly and insertion. *J R Soc Interface*. 2008;5 Suppl 3:S241-S250. doi:10.1098/rsif.2008.0239.focus.

39. Markvoort AJ, Marrink SJ. Lipid acrobatics in the membrane fusion arena. *Curr Top Membr*. 2011;68:259-294. doi:10.1016/B978-0-12-385891-7.00011-8.

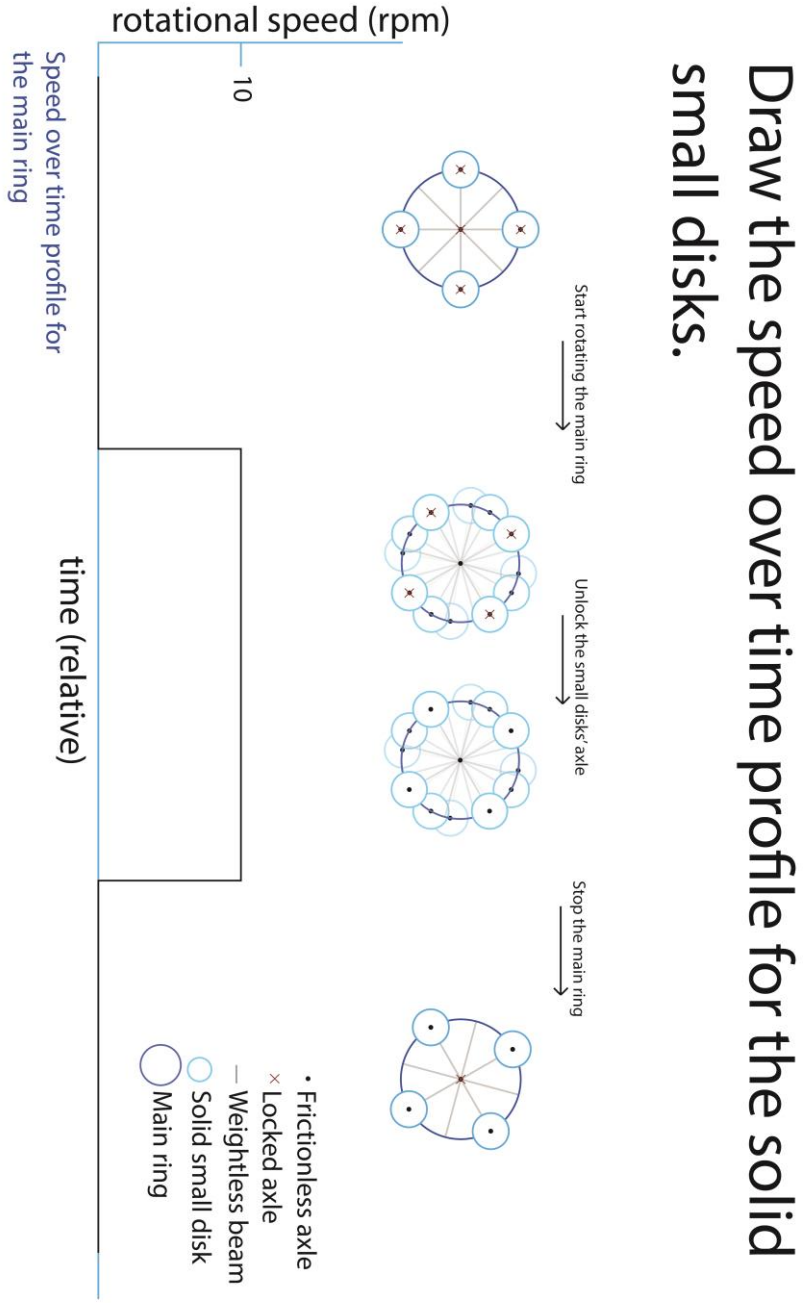
Aknowledgements

This work was made possible by the superb supervision the whole MD group at the RUG has provided. That being said, I would like to explicitly thank the following people for the amount of time and effort they have invested in me and my research in the past months:

Alex, Dick, Helgi, Ignacio, Jaakko, Manuel, Paulo, Tsjerk and Siewert Jan

The Big Eye's problem

During my stay at the Molecular Dynamics lab in Groningen I thought of a "simple" physics problem which caused many people around me to disagree with each other. The final proof of the problem was given by J.J. Uusitalo and can be supplied if requested. If your brain is not tired yet after reading this report, or if you only wanted this report to gain excess to this problem, I would like to invite you to grab a glass of wine – or any other alcoholic beverage you prefer – sit down and give it a shot. As with any problem, draw while you think and think before you write.



Appendix

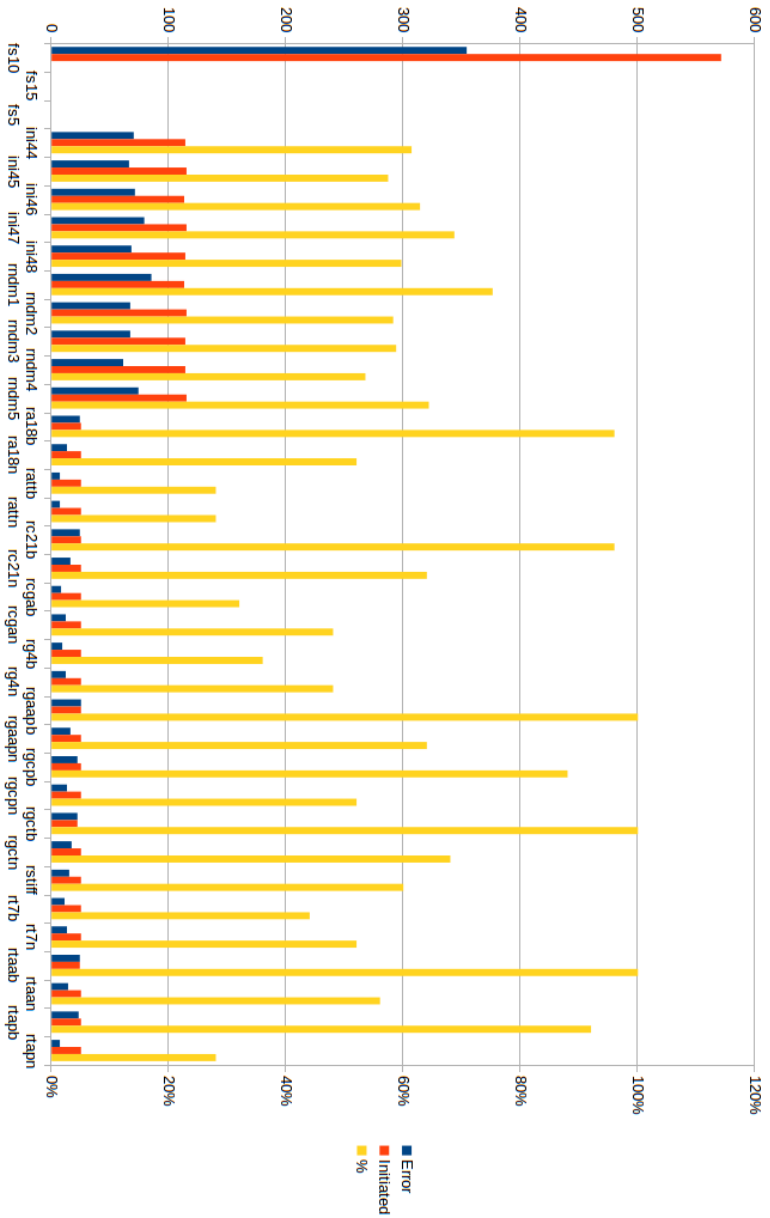


Figure A- 1 Crashed/Initiated percentage for the 10 fs base flip simulations. Red depicts the absolute amount of initiated simulations for each group. Blue depicts the absolute number of simulations that stopped before reaching 1 µs. Yellow shows the percentage of crashed simulations compared to the amount of initiated ones. Ini[xx] stands for the initial configuration used. Rndm[x] is the random seed used to generate the initial velocities. The uncoupling code starts with an [r] which stands for the stiff elastic network that was used. The end of the uncoupling code is either a [n] which stands for the uncoupling of only the base, or a [b] which stands for uncoupling of the full nucleotide. If the second last character of the uncoupling code is a [p] the stated residue(s) was/were uncoupled as pairs.

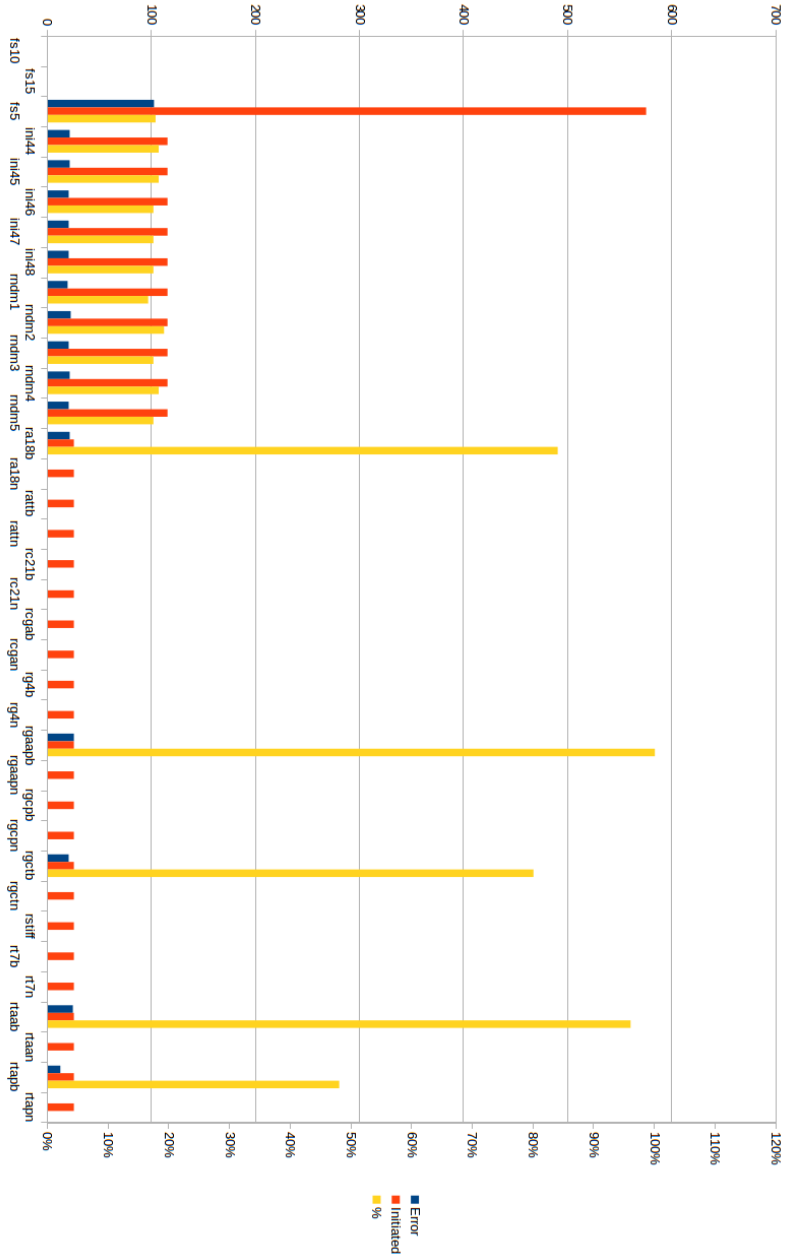


Figure A- 2 Crashed/Initiated percentage for the 5 fs base flip simulations. Red depicts the absolute amount of initiated simulations for each group. Blue depicts the absolute number of simulations that stopped before reaching 1 µs. Yellow shows the percentage of crashed simulations compared to the amount of initiated ones. Ini[xx] stands for the initial configuration used. Rndm[x] is the random seed used to generate the initial velocities. The uncoupling code starts with an [r] which stands for the stiff elastic network that was used. The end of the uncoupling code is either a [n] which stands for the uncoupling of only the base, or a [b] which stands for uncoupling of the full nucleotide. If the second last character of the uncoupling code is a [p] the stated residue(s) was/were uncoupled as pairs.

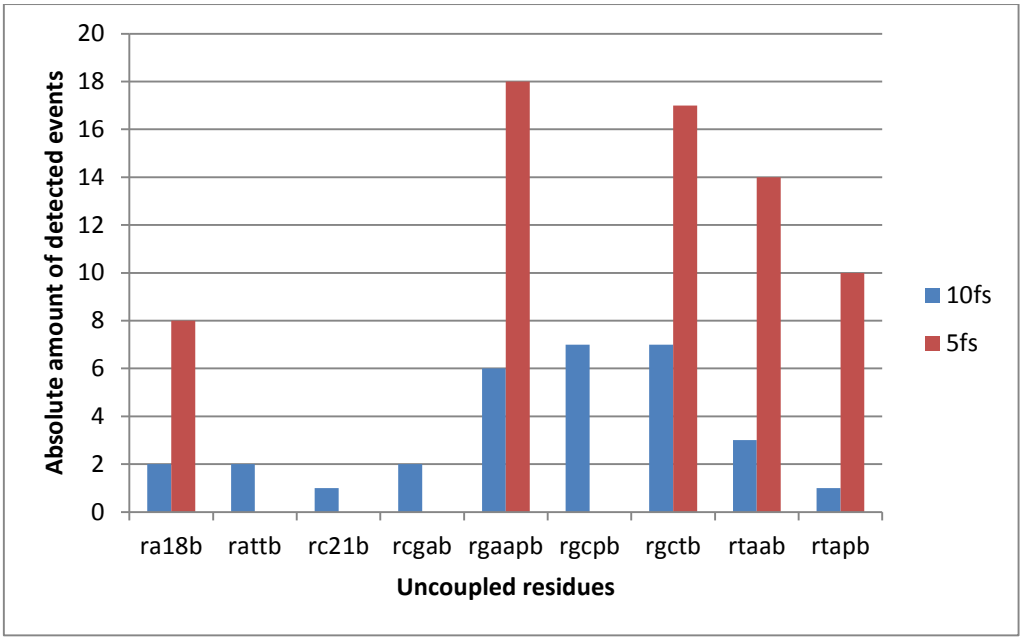


Figure A- 3 The detection of base flipping for the 5 and 10 fs simulations. The 5 fs time step simulations show more events in total, but within less different types of uncoupling with regard to the 10 fs time step simulations.