Molecular Dynamics Simulations of DNA Translocation through a biological Nanopore

Simen Eidsmo Barder

Abstract

Experimental and simulation studies of nucleic acid transport through nanosized channels, both biological and synthetic, has become a rapidly growing research area over the last decade. While the utilization of the α-hemolysin channel as a sequencing device is soon to be realized, other biological nanochannels may hold advantages that are yet unknown. Motivated by this, the first reported molecular dynamics simulations of DNA translocation through a connexon 26 channel were accomplished, for single-strandeed DNA with a length of 24 nucleotides and with a sequence containing only adenine, cytosine, guanine or thymine bases. Transmembrane voltages between 40 mV and 8.4 V were applied for up to ~2 ns, and the minimum voltage needed for translocation was found to be at 2.4 V. Higher voltages led to shorter translocation times in most cases. Non-translocation or slow translocation events were normally the result of a high degree of foldedness at the entrance of the "funnel" region, the narrowest part of the connexon channel. Distinct differences were seen between the bases, in particular through slower translocations for the purines than for the pyrimidines. Comparison with published literature of α-hemolysin translocation found that some of the results were on the same order of magnitude for translocation through connexon channels subject to constraints. It was concluded that to characterize the translocation mechanisms, further investigations should be carried out; both by the use of experiments as well as more simulation studies.

1

Preface

This master thesis was accomplished in the spring term 2012, at the Department of Structural Engineering, under the Faculty of Engineering Science and Technology (IVT) at NTNU. The work was carried out in close collaboration with the Nanomechanics group directed by Professor Zhiliang Zhang. Acknowledgements go to the High Performance Computing Group at the University of Tromsø, for the use of their supercomputer.

I would like to thank Professor Zhiliang Zhang for giving me the opportunity to write a master thesis at his research group. The project has been very inspiring, and I appreciate the encouragement that was given to come up with new ideas. Many thanks also go to Associate Professor Jianying He and Post-doctoral research associate Shijo Nagao for many valuable contributions during group meetings, and for assistance during the writing of the thesis. Finally, I would like to thank my supervisor, Research fellow Jianyang Wu, for teaching me LAMMPS, helping me with a countless number of simulations and for giving valuable feedback on the thesis. I am deeply grateful for all the hours you spent helping me finish my thesis.

Contents

Part I

Introduction

One of the main focus areas within biomedical research over the past century has been to determine the connection between specific gene sequences (genotypes) and the resulting attributes (phenotypes) that they code for in humans[1]. The aim of so-called association studies is to gain insight into the correlation between genetic variations among individuals, and the resulting differences in certain disease susceptibilities[2]. For so-called monogenic disorders, the presence of "disease-coding" genotypes may completely predict the occurence of a certain disease in an individual, whereas for genetically complex diseases, the presence of such genotypes may only increase the probability of disease by a small amount. Common diseases like diabetes, heart diseases and cancer, arise due to a combination of both genetic and environmental factors[3]. This leads to the acknowledgement that surveying of genetic information will be central in the development of an improved type of healthcare. This improved healthcare will involve determination of the potential risk of acquiring specific diseases in an individual, thereby enabling preventive efforts like observation and screening to minimize the disease probability[4]. Specific treatment based on genomic profiling will also be a part of this new type of healthcare. Through analysis of their genomes, individuals may be divided into subpopulations and receive a more effective treatment, since some therapeutic treatments may have undesirable side effects or even prove ineffective for specific patient populations.

While mapping of the entire human genome (about $3.2 \cdot 10^9$ base pairs[5]) was completed by the Human Genome Project in 2003[6], there is still an increasing need for techniques that facilitate fast and cheap sequencing of individuals. Automated Sanger sequencing[7] has been a standard method for about two decades, and was the method utilized in the Human Genome Project[8]. The method includes amplification of DNA, modification of nucleotides, fluorescent labeling and electrophoretic separation, which makes it both time consuming, expensive and labour-intensive[9]. The cost of the Human Genome Project was estimated to about US$300 million [8], and even though its cost has decreased dramatically since, methods that are both faster and cheaper have gradually replaced the Sanger method[8]. These newer methods, termed next-generation sequencing (NGS) techniques, have decreased the genome sequencing cost to below US$10'000 in 2012[10].

To further emphasize the comprehensive focus on the development of faster, cheaper and more accurate sequencing techniques, the Archon Genomics X PRIZE Competition is rewarding US$10 million to the first research group able to sequence 100 human genomes within 30 days, to an accuracy of 1 error per 1'000'000 bases and with 98% completeness, to a maximum cost of US$1'000 per sequenced genome[11].

While NGS techniques may be superior to Sanger sequencing, the emerge of so-called third-generation techniques that involve single molecule sequencing,

are now under development[12]. Single molecule sequencing is beneficial because it may be carried out in real-time, without the need for DNA amplification steps or large amounts of DNA. Among the different single molecule sequencing techniques, sequencing by enforcing DNA through a nanosized channel also holds the advantage of being reagent free and rapid.

By electrophoretically pulling a single molecule of DNA through a nanochannel, a reduction in the ionic current can be detected due to DNA-blockage of the channel[13]. By analyzing the characteristics of the current drop during DNA translocation, a potential read-out of the base sequence may thereby be accomplished. Different types of channels have been investigated for this utilization, which can broadly be separated into biological and solid-state nanochannels[12]. The first commercial machine that is claimed to be capable of successfull sequencing will be launched by Oxford Nanopore Technologies during 2012[14, 15]. Their technology is based on a modified α-hemolysin channel, and they have reported reading a continuous DNA molecule about 48'000 bases long, longer than postulated by anyone before[14].

In this project work, a connexon 26 gap junction protein is investigated regarding its translocation properties of single-stranded DNA, with potential future applications as a sequencing device. Analysis of DNA-translocation through the protein channel is accomplished by the use of molecular dynamics (MD) simulations.

Part II
Theory

In this part, the basic theory of nucleic acid translocation through nanochannels is first presented. Different types of nanochannels (biological and solid-state) are then discussed, ending in a short description of the commonly usedα-hemolysin channel. Theory about connexon channels in general, as well as the specific one used in this study, is thereafter presented, followed by some theory about the DNA molecule. Finally, a rather comprehensive section about the theory behind molecular dynamics and force fields is given.

1 Nucleic acid translocation through nanochannels

1.1 Main concept: Coulter-counter technique

The concept of using a nanopore for detecting and analyzing nucleic acids is based on the Coulter-counter technique[16]. The method involves filling two adjacent chambers with an electrolyte solution and separating the chambers, only allowing movements through a narrow channel. By adding charged particles to one of the chambers (the "cis" chamber) and applying a voltage over the
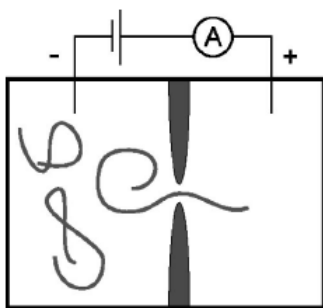
Figure 1: Schematic illustration of translocation of a polymer by applying a voltage across a nanosized pore. [17]
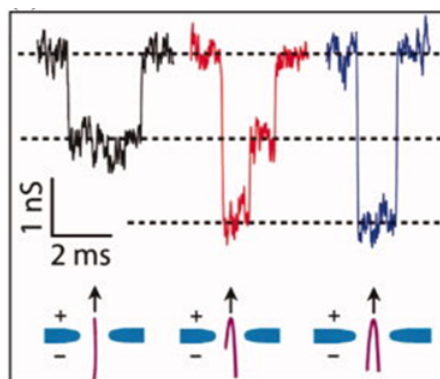
Figure 2: Illustration of the different current drops arising due to a straight, partially folded and folded polymer, upon translocation through a nanosized region. [20]

channel, the particles will be electrophoretically driven through the channel, and over to the other chamber (the "trans" chamber) (Fig. 1).

Since the presence of a particle inside the pore results in a temporary (and partial) current blockage, the particle can be detected through a corresponding drop in the ionic current, which is constantly monitored. The magnitude of the current drop is proportional to the size of the particle blocking the channel, because a larger particle will displace a larger amount of ions[17]. For long polymers like nucleic acids, the duration of the current drop will be proportional to the length of the polymer. In addition, different folding conformations may be read out of the current drops, as shown in figure 2. A folded polymer give rise to a deeper current drop than a linear one, and by analyzing how the signal changes in time, the conformation of the polymer during translocation can be revealed.

At low ion concentrations, the phenomenon of current "spikes" instead of

7

current drops has been observed during DNA-translocation[18, 19]. This is assumed to be caused by introduction of the counterions that surround DNA into the channel.

1.2 Biological and solid-state nanochannels

For detection of nucleic acids by using a Coulter-counter approach, the channels connecting the two electrolyte chambers need to be on the nanoscale; they are therefore often termed nanopores or nanochannels. Two main approaches for polynucleotide translocation through nanochannels have been investigated, involving translocation through either biological or solid-state nanochannels. Common biological nanochannels include α-hemolysin[13] and the phi29 motor protein[21]. In addition, synthetic silicon nitride[22], silicon oxide[17] and graphene-based[23] nanochannels as well as nanotubes[24, 19] have been developed for nanochannel experiments. Biological and solid-state nanochannels have their own advantages and drawbacks. Biological channels have the advantage of being genetically controlled, thereby providing channel diameters and topologies that are more predictable than the ones resulting from solid-state nanochannel synthesis[25]. On the other hand, solid-state nanochannels are more mechanically robust, and do not depend on displacement within fragile lipid bilayers. They can also tolerate a broader range of temperature, pH and chemical conditions.

The idea of sequencing single molecules of DNA or RNA by using methods similar to the Coulter-counter technique, is suggested through various approaches; by electronic or optical interactions, or through force measurements[20]. The electronic detection mechanisms normally make use of additional probe electrodes positioned in the nanochannel region, enabling them to perform local single-molecule spectroscopy by applying a transverse electric field (Fig. 3).

Examples of this include measuring the variations in tunneling currents traversing through the different base types[26]. Recently, the identification of a single base within a nanochannel has been achieved through this method[27]. It has also been suggested that the translocation of different base types can be directly read out from the current blockage signal[13].

1.3 Major challenges

With the possible exception of the upcoming sequencing device from Oxford Nanopore Technologies (as mentioned in the introduction), nanochannel sequencing of a polynucleotide with single-base resolution has not yet been achieved due to several different challenges. One of the major ones suggested by Kasianowicz et al.[13] is the ability to control the motion of the DNA molecule during translocation. The translocation speed must be slow compared to the time resolution of the ionic current measurement system, to be able to detect separate signals from each base. Change of external factors such as viscosity, applied voltage, ion concentration and temperature has been shown to decrease the translocation time[28]. Modification of the solid-state nanochannel geometry[25] and

8
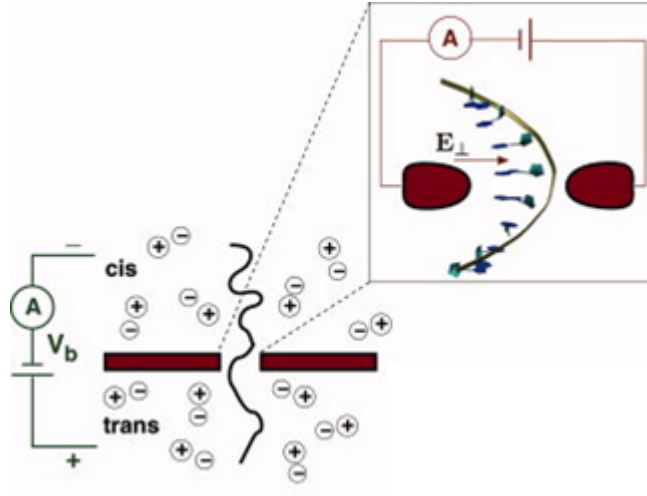
Figure 3: To discriminate between the bases of a DNA molecule, a transverse electric field is applied as DNA translocates through a nanopore region, from the "cis" to the "trans" chamber. Each base may give rise to a distinct signal, enabling sequencing of the DNA molecule. Adapted from [20].

the use of optical tweezers[29] are other speed-reducing techniques that have been demonstrated. Among other challenges are thermal fluctuations and the secondary structure of the DNA molecule, which both contribute to disturbances in the detected signals and must be minimized to be used in a potential sequencing device[9].

As stated in Ref. [20], a first approach towards successful sequencing involves differentiating between DNA molecules of different lengths or sequences - without necessarily distinguishing the separate bases.

1.4 Structure of the α-hemolysin channel

The α-hemolysin protein is secreted by the Staphylococcus aureus bacteria, and is believed to induce cell death in different cell types [30]. It is also been used as a biological channel for DNA-translocation experiments for more than a decade [13]. Figure 4 shows a schematic α-hemolysin channel, with the main regions labeled. The length of the whole α-hemolysin is about 100 Å, whereas the length of the transmembrane channel is about 50 Å[30, 31]. The extramembranal section of the protein is termed the "cap", while the transmembrane channel is a β-barrel known as the "stem". The cap entrance is 26 Å in diameter, and widens to 36 Å within the "vestibule", the widest part of the whole channel. The transmembrane channel possesses a restricting minimum diameter of 15 Å at its entrance, which is the narrowest part of the whole channel. The transmembrane channel is also thought to be hydrophilic.
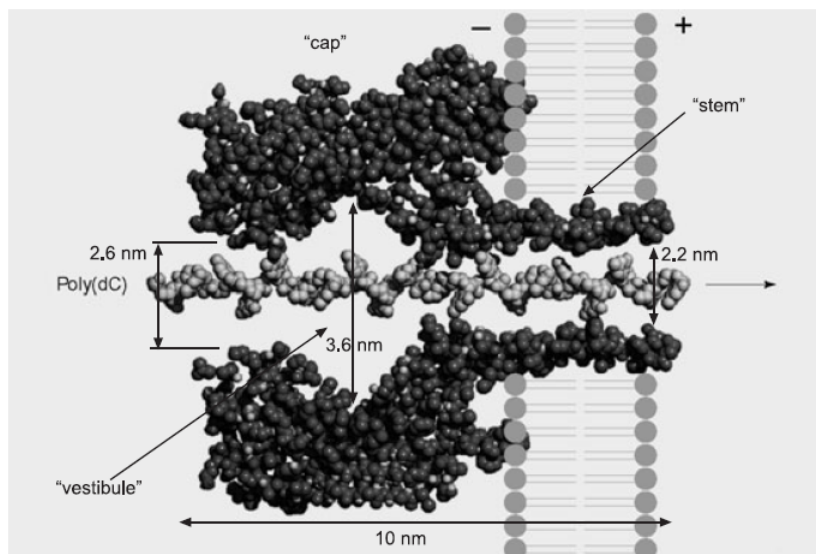
Figure 4: Cross-section of a α-hemolysin channel displaced in a cellular membrane. [30]

2 Connexons

2.1 Cell-cell junctions

To be able to form tissues and organs, multicellular organisms need to carry out mechanisms of attachment between single cells[32]. These mechanisms typically involve modifications of the cell membranes of the connecting cells, and the resulting structures are called cell-cell junctions. The three most common cell junction types in the animal kingdom are known as adhesive junctions, tight junctions and gap junctions. Adhesive junctions link cells together, enabling them to form tissues and serve as a unit. Tight junctions, on the other hand, seal the two neighboring cells tightly together, preventing the passage of molecules. Gap junctions are regions where the cell membranes of two cells are brought together at a fixed distance of 2-3 nm (Fig. 5). The gap junction regions contain aqueous channels, which allow the intercellular exchange of ions and small molecules under ~1.0 kDa[33]. This way, the gap junctions enable direct electrical and chemical intercellular communication [32].

2.2 Gap junctions and connexons

A gap junction consists of clusters of up to thousand intercellular channels[35]. Each of the gap junction channels are composed of two sub-channels, one from each of the connecting cells, termed hemichannnels or connexons. Connexons are multiprotein complexes formed by six protomers called connexins, which
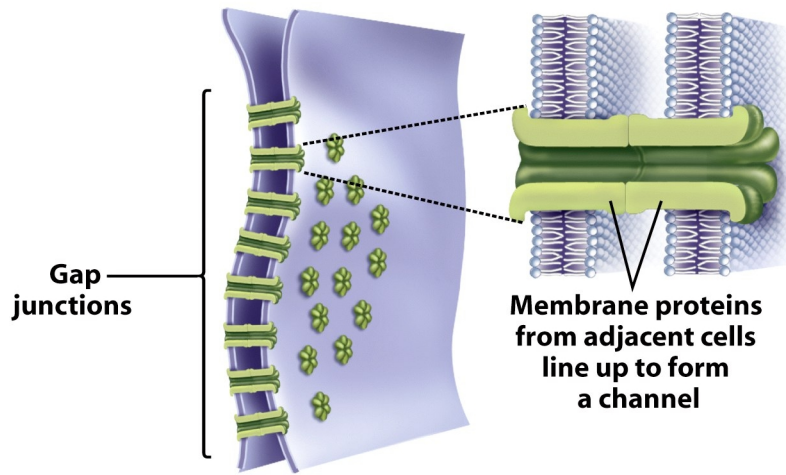
Figure 5: Illustration of gap junction channels, which connect the intercellular environments of two neighboring cells. The two membrane proteins (connexons) that form the gap junction channel can be seen in the magnified section on the right. [34]

surround the central channel (Fig. 6).

While there exist different types of membrane proteins involved in intercellular communication (e.g. ion channels and G-protein coupled receptors), the connexons are the only ones that mediate intercellular communication by connecting the cytoplasms of the adjacent cells[35]. In a multicellular organism, intercellular communication is crucial to carry out complex biological functions like neural transmission, immune reazction and reproductive function. The proper functioning of gap junction channels in humans is crucial, and mutations in connexin genes are associated with a wide variety of diseases like neuropathies, deafness, epidermal diseases, cataracts and oculodentodigital dysplasia[37]. It is also shown that connexons have various biological functions even when they are not paired up with a connexon from a neighboring cell[38].

2.3 Connexin genes and proteins: Nomenclature and subgroups

The human genome contains 21 different genes known till date, that code for 21 corresponding connexin proteins[38, 37]. The standard nomenclature for proteins in the connexin family is "Cx" followed by a suffix that indicates the predicted molecular mass of its mouse isoform in kilodaltons. Examples include Cx26 and Cx31.1, which refer to connexins with masses of 26 and 31.1 kDa, respectively. It has also been common to divide connexins into an α-group and a β-group, partially based on similarities in the amino acid sequence. The official
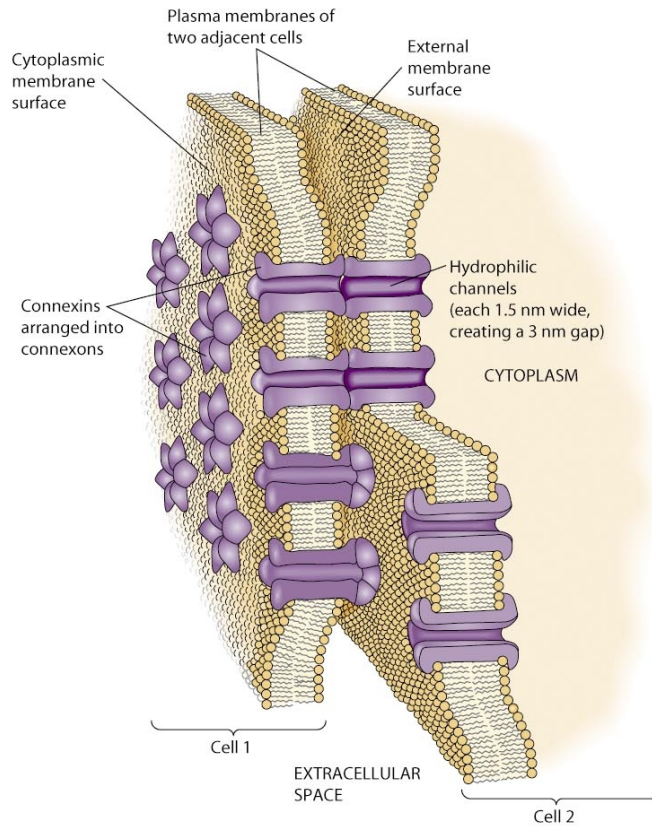
Figure 6: Detailed illustration of the hydrophilic gap junction channels, with the hexameric structure of the connexon channels visible. [36]

nomenclature for the connexin genes starts with the letters "GJ" (abbreviation for "gap junction") followed by a letter indicating the Greek-letter subgroup, and ends with a number denoting the order of discovery in that subgroup. The gene for Cx43, GJA1, is for example the first identified connexin gene within the α-group. In 2007, the Greek letter subdivision system was extended because of the requirement for additional subgroups[37].

2.4 Connexin proteins: General structure and functionality

The main constituents of all connexin proteins include four transmembrane segments (TM1-4), two extracellular loops (E1 and E2), an N-terminal helix (NTH), a C-terminal segment (CT) and a cytoplasmic loop (CL) (Fig. 7)[37].

The extracellular loops play an important role in docking two connexons
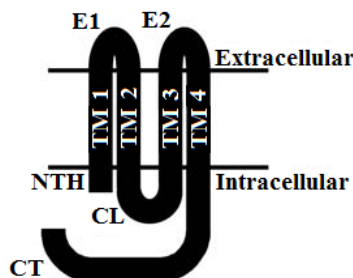
Figure 7: Schematic illustration of a general connexin protein, showing four transmembrane segments, two extracellular loops, an N-terminal helix, a C-terminal segment and a cytoplasmic loop. Adapted from [37]

together to form a gap junction channel, and are also also some of the most conserved structures within the connexin family: For all connexins (except Cx31 and Cx23), the extracellular loops contain three cysteines each, and are separated by an equal number of other amino acids. The lengths and sequences of the transmembrane and N-terminal helices are also relatively similar when comparing different connexins. In contrast, the amino acid sequences of the C-terminal segment and the cytoplasmic loop are highly divergent.

In addition to the heterogeneity within the connexin family, there exists an even larger variety of connexons and gap junction channels[33]. This is due to the fact that different connexins may interact to form various heterogeneous structures (Fig. 8): Homomeric connexons are formed by six identical types of connexins, while heteromeric connexons are formed by six connexins where at least two of them are of different types. Similarly, in homotypic gap junction channels the opposing connexins are of the same type, while in heterotypic gap junction channels they are of different types.

Even though there are variations within connexins, and different ways in how they may form connexons and gap junction channels, the fundamental structure of six protomers surrounding a central channel is conserved among all connexons[33]. The connexon channel permeability, however, holds large variations, probably because different connexons normally reside in different tissue types, where they have different functions[38].

Unpaired connexons are closed by default, and respond to specific physiological signals or stress factors by opening their channels[38]. By pairing up with an opposing connexon to form gap junction channels, the opposite occurs: To enable communication between the cells, the "ground state" is an open gap junction channel, which only closes as a response to specific physiological conditions. Factors that affect the opening and closing of connexin-containing structures include $Ca2^+$ concentration[39], pH[40], mechanical stimulation [38] and voltages[41].

Specific values for the channel diameters are experimentally challenging to
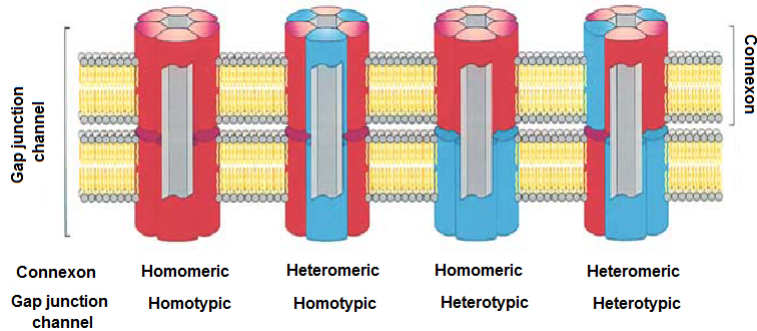
Figure 8: Illustration of the different possible combinations of connexins and connexons, to form gap junction channels of varying degree of homogenity. Adapted from [33].

obtain, but the diameters of the Cx26 have been measured down to a resolution of 3.5 Å, to vary between 14-25 Å[42]. In addition, experimental data suggest the following two rankings of channel diameters for different (homotypic) gap junction channels, in decreasing order[43]:

-Cx43 > Cx32 > Cx26 > Cx37

-Cx43 ~ Cx46 > Cx40.

2.5 Connexon 26 channel structure and gating mechanism

X-ray analysis has determined the structure of the human connexon 26 gap junction channel in the open state, at a resolution of 3.5 Å[42], and the details of this structure will be presented in this section. Reference [42] is used for the whole section, unless stated otherwise.

The connexons formed by Cx26 protomers occur among others in cells located in the retina and inner ear, functioning in cone regulation and Ca2+-mediated ATP, respectively [38]. Connexon 26 gap junction channels have been experimentally shown to close at a pH<6.5 and open when the pH is increased to 7.6[44].

The connexin 26 protomer is composed of the general connexin components: four transmembrane helices (TM1-4), two extracellular loops (E1 and E2), an N-terminal helix (NTH), a C-terminal segment (CT) and a cytoplasmic loop (CL). The 226 amino acid residues of the Cx26 protomer can be seen in figure 9.

Residues 110-124 and residues 218-226 are not resolved, which means that the CL and the CT are excluded from the model. This is also the case for the side chains of residues K15, S17 and S19, as well as the M1 residue.

Figure 10a shows the symmetric relationship between the six protomers in a connexon (when looking through the channel), as well as the entrance and innermost diameter of the channel. The figure also indicates the relative dis-

14
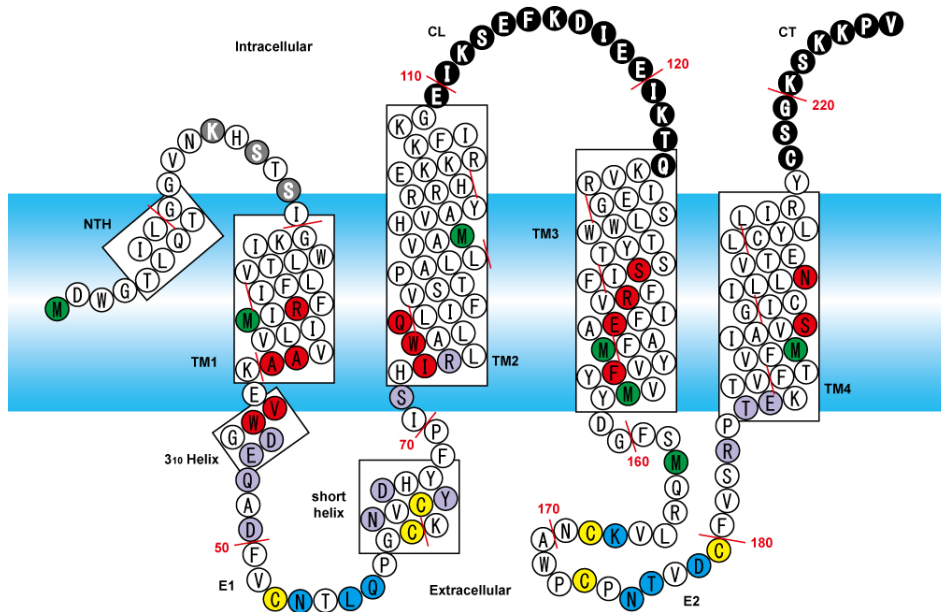
Figure 9: Illustration of the 226 amino acid residues of the Cx26 protomer. The black, filled circles with white letters show residues that were not resolved. Supp. info from [42].
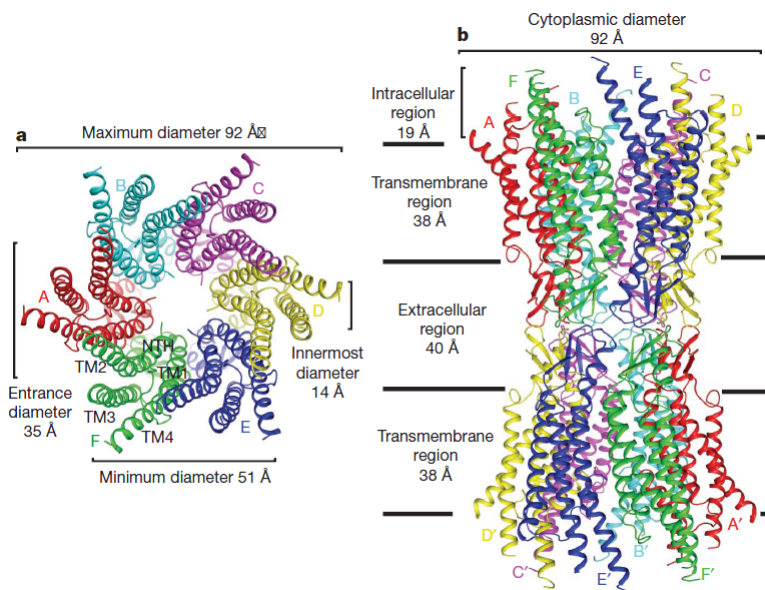
Figure 10: Illustration of a connexon 26 gap junction channel, as seen through the channel (a), and from the side (b). Adapted from [42].
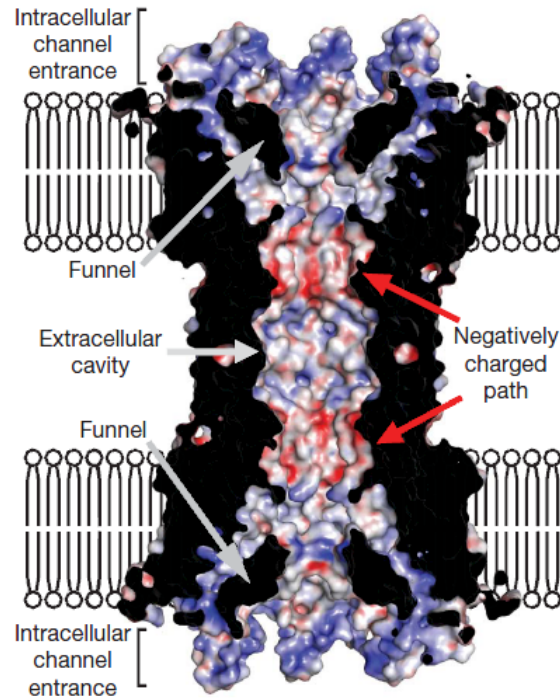
Figure 11: Cross-section of the connexon 26 gap junction channel, with the electrostatic potential shown. Red areas are negatively charged, and the blue areas are positively charged. Notice the positively charged intracellular area and the negatively charged path below the funnel region. Adapted from [42].

placement of TM1-4 in the four-helix bundle: TM1 and TM2 face the interior, while TM3 and TM4 face the membrane environment. Figure 10b shows the dimensions of two connexons forming a gap junction channel. The whole length of the gap junction channel (without the CT and CL) is estimated to be around 155 Å, while the length of one single connexon is about 77 Å.

2.5.1 Structure of the connexon and gap junction channel

The connexon channel consists of a cytoplasmic entrance, a main channel with a narrow "funnel", and an extracellular exit. When the connexon forms a gap junction channel with an opposing connexon, the extracellular exit area instead becomes an extracellular cavity (Fig. 11). This extracellular cavity is encased by extracellular loops E1 and E2 from both connexons, which interact through hydrogen bonds and salt bridges, contributing to a tight wall that shields the interior from the extracellular environment.

Figure 11 also shows the variation in surface potentials. There is a positively charged environment at the intercellular channel entrance, caused by positively
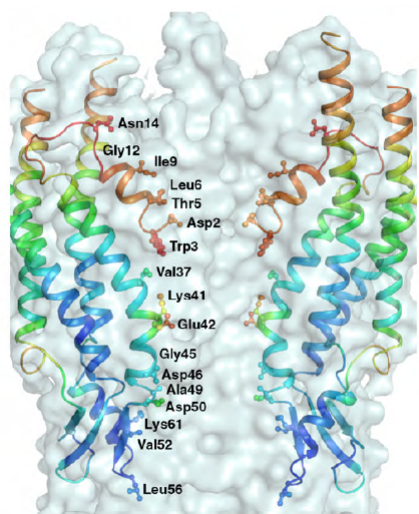
Figure 12: Highlight of the amino acid residues lining the connexon 26 channel. The narrowest part of the channel is located at the Asp 2-residue. (For clarity, only two of the six subunits are shown.) [45]

charged residues in TM2 and TM3. Further inside the channel, close to the extracellular cavity, amino acid residues Asp 46 and Asp 50 create a negatively 9 Å long charged path (See also fig. 12). There are three regions in the channel that are thought to contribute to size restriction and charge selectivity, and this negatively charged path is one of them. One of the other regions are the Lys 41 residues at the TM1/EM1 boundary, which form a narrowing of about 17 Å.

The last size restricting and charge-selective area is a funnel formed by the six NTHs of each protomer. The funnel has a diameter of 14 Å in an open state, which makes it the narrowest part of the whole channel.

Figure 13 shows the funnel when observed from the cytoplasmic side. The minimum channel diameter of 14 Å is localized at the bottom of the funnel, where Asp 2 from one NTH interacts through hydrogen bonds with Thr 5 from the neighboring NTH. In addition, Trp 3 situated next to the Asp 2, interacts hydrophobically with Met 34 at the TM1 of the neighboring protomer. The Trp 3–Met 34-interactions draw the NTH to the inner wall of the channel, and keep the funnel in an open state. It should be remarked that Kwon et al. determined a minimum channel diameter of 10 Å at the Asp 2 residues (instead of 14 Å) during MD simulations[46], presumably because they took larger atom diameters into consideration. Kwon et al. also pointed out that inclusion of the Met 1 of the NTH would have resulted in a further narrowing of the channel.
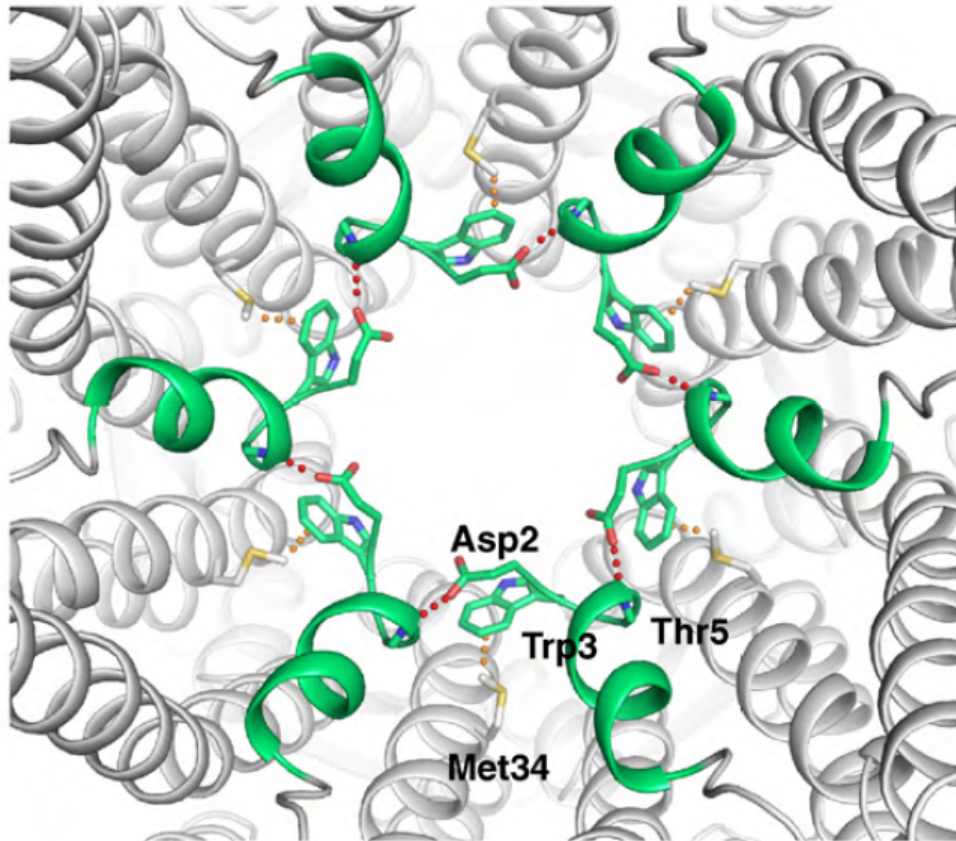
Figure 13: Illustration of the funnel in an open conformation, as seen through the connexon 26 channel. The NTHs from all six subunits are colored green. Red dots indicate hydrogen bonds, whereas orange dots indicate hydrophobic interactions. [45]
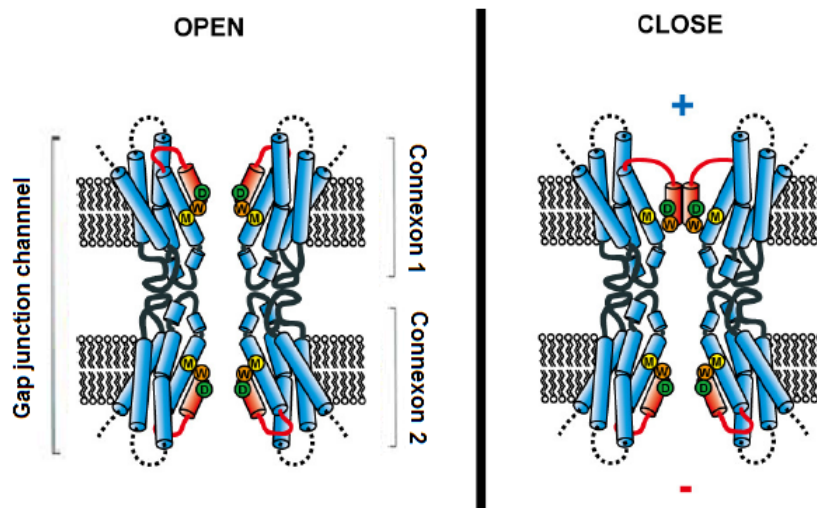
Figure 14: Voltage gating mechanism of the connexon 26 gap junction channel, consisting of two connexons. For clarity, only two connexin protomers are shown for each connexon. Adapted from [42].

2.5.2 Voltage gating mechanism

The voltage dependent gating mechanism is believed to involve displacement of the negatively charged Asp 2 residue. By an inside positive potential the Asp 2 residue is thought to move inward (in cytoplasmic direction), preventing both Asp 2-Trp 5 and Trp 3-Met 34-interactions (Fig. 14). This may then lead to a "release" of the six NTHs that make up the funnel, causing them to form a plug that physically blocks the channel in the connexon experiencing an inside positive potential. As can be seen in figure 14, the opposing connexon will in the same situation be subject to an inside negative potential, keeping its channel in an open state. This is thought to occur because the Trp 3 – Met 34-interactions prevent any drastic movement towards the extracellular direction. It should be noted that the voltage difference is measured between the intercellular environments of the connected cells, and not over the membrane. According to[41], each of the two channels in a homotypic connexon 26 gap junction channel closes at transjunctional voltages above 80-120 mV, and the gating mechanism is on the order of several seconds.

3   DNA

Reference [5] is used for the whole section, unless stated otherwise.
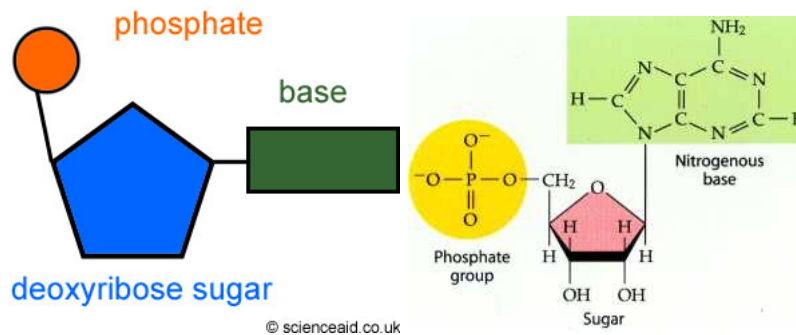
Figure 15: Left: Schematic illustration of a single nucleotide, the monomer of the DNA molecule. Right: Detailed illustration of a nucleotide, with the adenine base shown as an example. [47, 48]

3.1 DNA structure

The DNA (deoxyribonucleic acid) molecule is a polynucleotide chain, where each monomer is a nucleotide composed of a purine or pyrimidine base, a 2'-deoxyribose monosaccharide and a negatively charged phosphate (Fig. 15).

There are four different bases, the two purines adenine (A) and guanine (G), and the two pyrimidines cytosine (C) and thymine (T) (Fig. 16).

Alternating monosaccharides and phosphates form a sugar-phosphate backbone, which links the bases together into a single-stranded DNA molecule (Fig. 17). The phosphate group is linked to the neighboring monosaccharides through phosphodiester bonds, a 3'-hydroxyl and 5'-hydroxyl, respectively. This makes each DNA-strand directionally dependent, and the sequence of consecutive bases is normally read from the 5' to the 3'-end.

DNA usually occurs in a double-stranded configuration, which is formed by interaction of two complementary and antiparallel single strands. The complementarity is found between the different bases, which bind to each other through hydrogen bonds. The adenine bases are always complementary to the thymine bases, whereas the guanine bases are always complementary to the cytosine bases. This can be understood by observing the hydrogen bond interactions between the bases in figure 16: Different base shapes and different possibilities of interacting through hydrogen bonds results in the specific complementarities. When two complementary single-stranded chains of DNA interact to form a double-stranded chain of DNA, the secondary structure of the famous double helix is formed (Fig. 18).

The double helix stability is mainly caused by stacking interactions between adjacent bases located on the same DNA-strand[52]. Stacking interactions involve shearing of π-electrons between the adjacent "stacked" bases, and is accomplished because of the flat surfaces of the bases.

The complete genetic information in humans, the genome, consists of around $3.2 \cdot 10^9$.

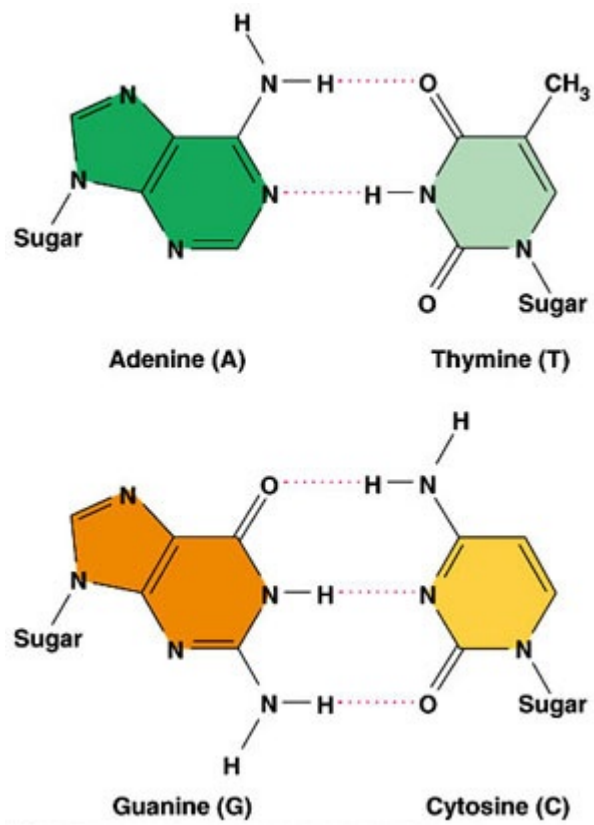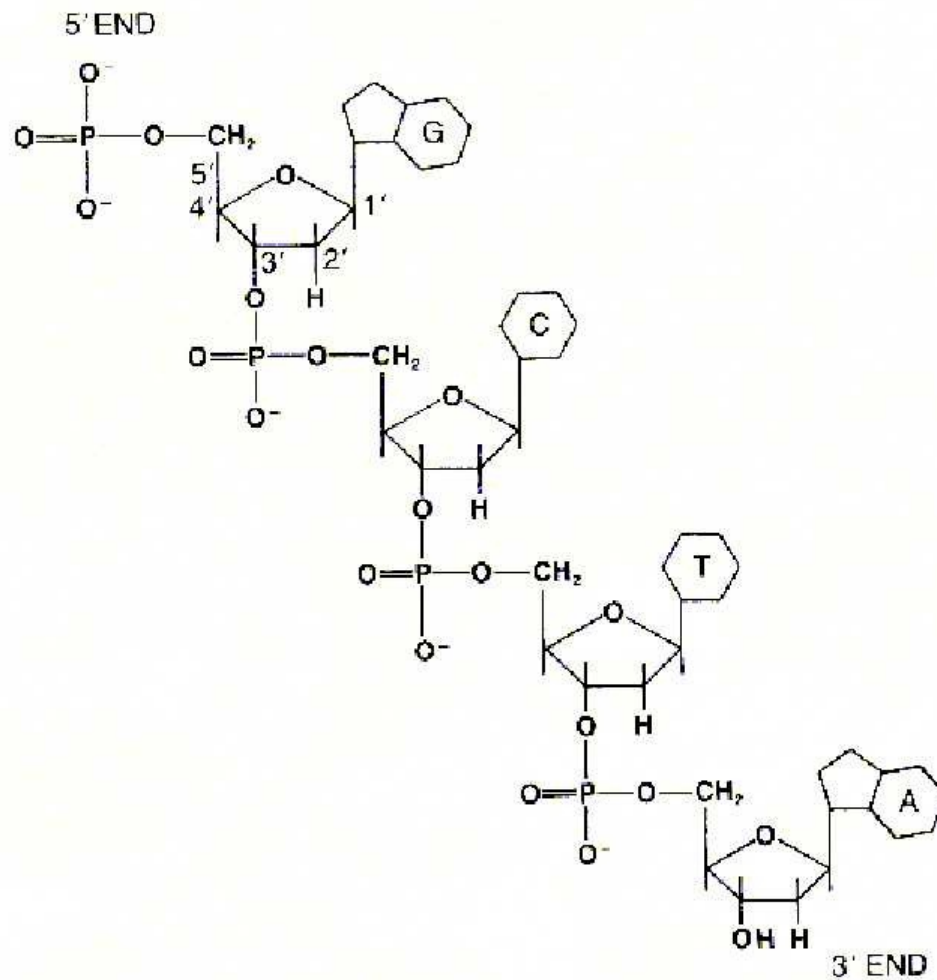Figure 16: The structure of the four bases of DNA, with A-T and G-C complementarity highlighted.[49]

Figure 17: Structure of the DNA molecule shown for four consecutive bases; G, C, T and A. The 5' and 3' end is indicated.[50]

22
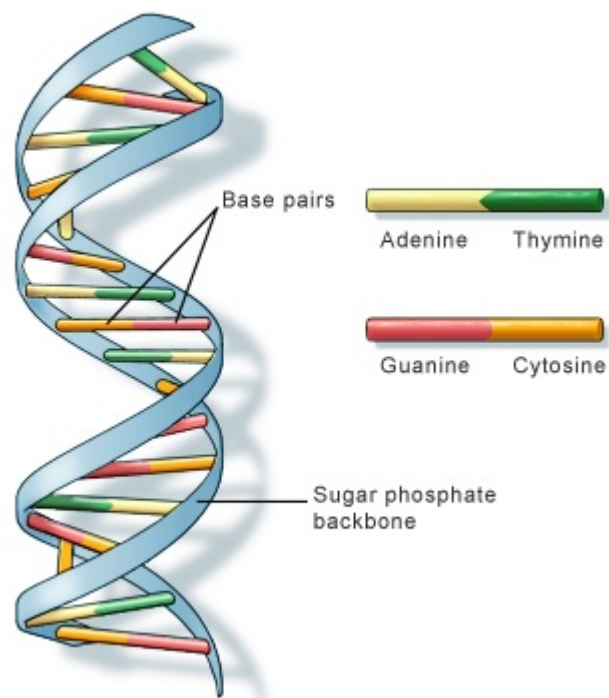
Figure 18: Illustration of the double helix, formed by complementarity between the bases of two single strands of DNA.[51]

3.2 DNA dimensions and conformations

The double-stranded DNA helix can exist in three different conformations, called B-, A- and Z-DNA (Fig. 19, upper). The most common conformation is the B-DNA structure, which is adopted by a great majority of the DNA in cells. The B-DNA structure involves 10 base pairs per helix turn, and a base pair distance of 0.34 nm.

B-DNA forms a wide major groove, and a narrow minor groove (Fig. 19, lower), which are the areas of the double helix where the bases are accessible. The major groove has a width of 1.2 nm, while the minor groove has got a width of 2.2 nm. The diameter of the double-stranded helix in B-DNA is 23.7 Å.

It should be noted that real DNA in cells is not perfectly regular − because of the base sequence variation; the B-DNA structure is therefore an idealized structure that most closely corresponds to the average structure of DNA found in cells. As an example, the experimental value for the number of base pairs per helix turn is found to be approximately 10.4 instead of 10.0[55].

A-DNA is more compact than the B-DNA structure, and can be observed in certain DNA-protein complexes. As can be seen in figure 19 (upper), the bases in A-DNA are tilted in relation to the helical axis, in a much larger degree than the B-DNA bases. Whereas the A- and B-DNA are right-handed, DNA can also in rare occasions form a left-handed DNA helix, which is denoted Z-DNA (Fig. 19, upper).

3.3 Persistence length of DNA

The rigidity of the DNA molecule and other polymers can be defined through the so-called persistence length[56], which is proportional to the stiffness of the polymer [57]. The persistence length can be defined as the length scale where the polymer maintains its tangent orientation[58]. For polymer segments longer than the persistence length, the segment will be flexible for thermal fluctuations, while for shorter segments it will act as a stiff rod[59]. The persistence length for double-stranded DNA in solutions containing high $Mg2^+$ or Na^+ concentrations, has been found to be around 350-550 Å, depending on experimental and modeling methods[60]. For single-stranded DNA, the persistence length has been found to be between 8 and 13 Å, for high salt concentrations[61].

B DNA A DNA Z DNA

0.34 nm

3.4 nm

Minor

Major
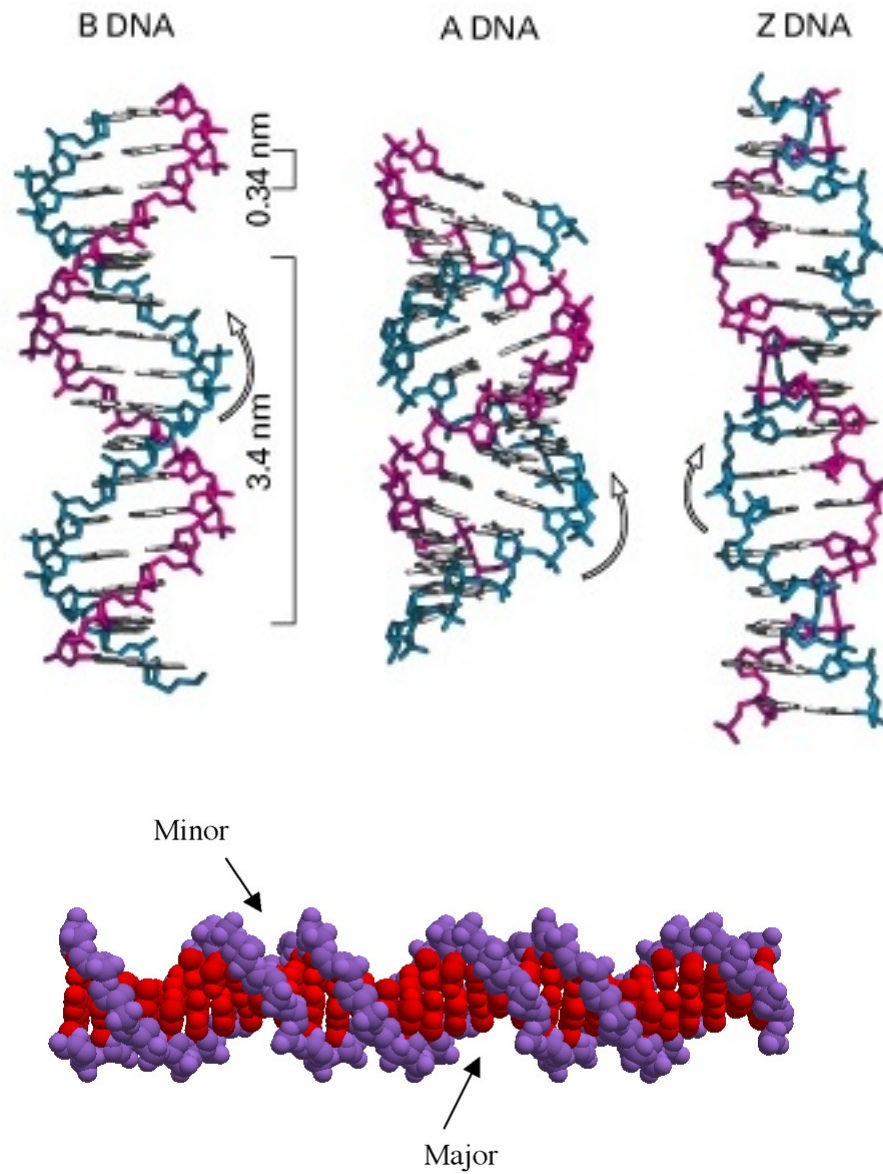
Figure 19: Upper: Illustration of the three conformations of DNA, with B-DNA being the most common one. Lower: Illustration of a double helix with the minor and major groove highlighted.[53, 54]
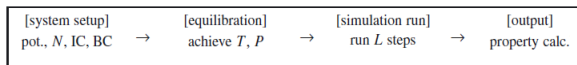
[system setup]   [equilibration]   [simulation run]   [output]
pot., N, IC, BC $\rightarrow$ achieve T, P $\rightarrow$ run L steps $\rightarrow$ property calc.

Figure 20: The progress of a typical MD simulation illustrated as a flowchart. Adapted from [62]

4 Molecular dynamics

References [62]-[65] are used for the whole section, unless stated otherwise.

4.1 General

A molecular dynamics (MD) simulation can be defined as a method that generates the atomic trajectories of a system consisting of N particles. The atomic trajectories are created through numerical integration of Newton's equation of motion, in combination with defined initial and boundary conditions (IC and BC). By knowledge of IC and BC, all atomic trajectories may be predicted, which makes MD-simulations deterministic. In a MD simulation, molecules are modeled by a collection of atom masses that are centered at their nuclei and bound together with springs. This mechanical consideration of the molecules corresponds to a "classical" approach to the system, and is based on the Born-Oppenheimer approximation. This approximation involves separating the wave function for a molecule into one for the nuclei and one for the electrons. The wave function for the electrons can be considered constant, and the electrons are therefore only implicitly modeled. In the mechanical approach of a typical MD-simulation, bond formation and breakage is seldom considered. Although quantum mechanical properties are not explicitly modeled in MD simulations, the information about structures and force fields may often originate from quantum mechanical energy calculations (as well as experimental data obtained from different kinds of experiments).

MD simulations are so-called atomistic simulations because all atoms are explicitly represented (sometimes with the exception of hydrogen). An empirically obtained potential energy function, or force field, is applied so that the molecules may stretch, bend and rotate about their bonds. In addition, van der Waals and Coulombic interactions are usually evaluated between non-bonded atom pairs. During a MD simulation, the molecules will change their displacements and conformations to minimize the total internal energy of the system, in accordance with the force field. It is also common to add external driving forces which will influence molecule displacements and conformations together with the force field. A typical flowchart of a MD simulation can be seen in figure 20.

The system setup involves defining the force field as well as initial and boundary conditions. Before starting the simulation itself, it is common to equilibrate the system until it reaches a desired condition – for example a specific value of the system temperature and pressure. The simulation is thereafter run for a cer-

tain number of discrete timesteps, before data ouput of chosen system properties can be analyzed.

The features of a MD simulation can be divided into six main sections, which will be described below: initial conditions, boundary conditions, force calculation, integrators, ensembles and property calculations.

For a system consisting of N atoms, the internal energy can be defined as $E \equiv K + U$, where the kinetic energy K is given by

$$K \equiv \sum_{i=1}^{N} \frac{1}{2} m_i |\dot{\mathbf{x}}_i|^2, \tag{1}$$

and the potential energy U is given by

$$U = U(\mathbf{x}^{3N}(t)). \tag{2}$$

$\mathbf{x}^{3N}(t)$ describes the 3D coordinates of the N particles; $\mathbf{x}_1(t), \mathbf{x}_2(t), ..., \mathbf{x}_N(t)$. The initial conditions of such a system will normally be given as $\mathbf{x}^3 N(t = 0)$ and $\dot{\mathbf{x}}^3 N(t = 0)$, corresponding to initial particle positions and velocities, respectively. To define the state of the system of N particles with known masses,$6N$ values are required; $3N$ coordinates for atom positions and $3N$ for velocity. Each of these $6N$ values defines a point in the so-called phase space, which consists of all possible states for the system. A MD simulation thus generates a sequence of points in phase space that are connected in time.

The boundary conditions of a system can be either isolated ("non-periodic") or periodic. For isolated boundary conditions, the system is surrounded by vacuum, and the N particles only interact with themselves; not with anything on the outside. For periodic boundary conditions, the particle dynamics are monitored in one "supercell" that is surrounded by an infinite amount of identical, similar supercells (Fig. 21).

Figure 21 shows a system of periodic boundary conditions, with the cutoff radius for one atom highlighted (See section 4.2.3 for further information about cutoff radius).

For isolated boundary conditions, a particle that exits the simulation volume will be lost, and its movements will be omitted during subsequent trajectory creations. In periodic boundary conditions however, a particle that exits the simulation volume will instantaneously re-enter on the opposite side of the volume. All the particles in the supercell may interact both with each other as well as with particles in adjacent supercells. The most common shape of a supercell is a parallelepiped. It is also possible to combine isolated and periodic boundary conditions, thereby creating a system that can be periodic in one direction and non-periodic in other directions.

Force calculations involve evaluating the right-hand side of the classical equation of motion;

$$m_i \frac{d^2 \mathbf{x}_i(t)}{dt^2} = \mathbf{f}_i \equiv -\frac{\partial U}{\partial \mathbf{x}_i}, i = 1, ..., N, \tag{3}$$
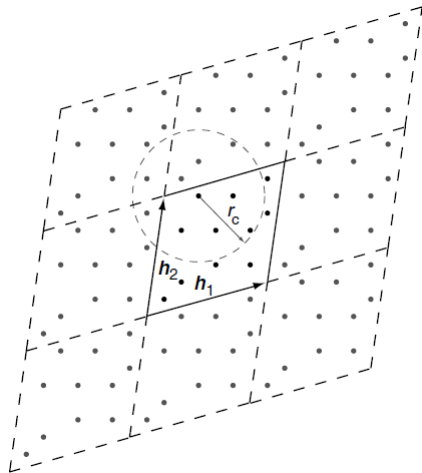
Figure 21: Illustration of periodic boundary conditions (PBC). Vectors h1 and h2 define a supercell which is repeated an infinite amount of times. The cutoff radius for one atom is also highlighted. [62]

which is often the most computationally demanding part of a MD simulation. Two simplifications of the potential energy are therefore commonly applied. The first simplification (Eq. 4) can be carried out by considering the potential in terms of the sum of pairwise potentials, while the second approximation (Eq. 5) can be applied by considering the interactions as interactions between spherical atoms. Equation 2 then becomes:

$$U = \frac{1}{2} \sum_{i \neq j}^{N} U(\mathbf{x}_i, \mathbf{x}_j) \tag{4}$$

$$= \frac{1}{2} \sum_{i \neq j}^{N} U(r_{ij}), \tag{5}$$

where r_{ij} is the interatomic distance between atoms i and j.

In a molecular model, the force field is the origin of the potential energies, which will be further described in section 4.2.

There exist several algorithms that can be utilized in the generation of atomic trajectories during a MD simulation. These algorithms are called integrators, and they all have in common that they advance the trajectories over small, discrete time increments, or time steps Δt:

$$\mathbf{x}^{3N}(t_0) \rightarrow \mathbf{x}^{3N}(t_0 + \Delta t) \rightarrow \mathbf{x}^{3N}(t_0 + 2\Delta t) \rightarrow ... \rightarrow \mathbf{x}^{3N}(t_0 + L\Delta t), \tag{6}$$

where L normally is between 10^4 and 10^7.

28

For biomolecular dynamics, one of the simplest and best families of integrators is the Verlet group, which is known to be exceptionally stable over long simulation times. A popular integrator within the Verlet group is the velocity-Verlet algorithm, where the initial information involves values of $\mathbf{x}^{3N}(t_0)$ and $\mathbf{v}^3 N(t0)$. The following formula is first used to find particle positions after one time step Δt:

$$\mathbf{x}_i(t_0 + \Delta t) = \mathbf{x}_i(t_0) + \mathbf{v}_i(t_0)\Delta t + \frac{1}{2}\left(\frac{\mathbf{f}_i(t_0)}{m_i}\right)(\Delta t)^2. \tag{7}$$

The forces $\mathbf{f}^{3N}(t_0 + \Delta t)$ are thereby calculated for the new particle positions, which enables finding the particle velocities after one time step:

$$\mathbf{v}_i(t_0 + \Delta t) = \mathbf{v}_i(t_0) + \frac{1}{2}\left[\frac{\mathbf{f}_i(t_0)}{m_i} + \frac{\mathbf{f}_i(t_0 + \Delta t)}{m_i}\right]\Delta t \tag{8}$$

Thus, by continuing for successive time steps, the trajectories for particle positions and velocities are created.

The maintenance of different ensembles often rely on Verlet algorithms. The traditional ensemble to operate under in MD simulations is the microcanonical ensemble (constant NVE). While this is often used for benchmarking of the system or during equilibration, it might be desirable to apply a canonical (constant NVT) or isothermal-isobaric (constant NPT) ensemble during the main simulation, to keep a constant temperature or pressure, respectively.

All properties within classical and statistical mechanics can be calculated for the atoms in the system, which is the central advantage of a MD simulation. Major challenges however, arise from inaccuracies originating from the force field, and the amount of computational power needed for efficient simulations.

As mentioned previously, the limitations of MD simulations include the inability to calculate properties at the quantum mechanical level. In addition, simulation durations are normally limited to nanoseconds, and the system sizes are of many orders of magnitude smaller than the corresponding "natural" systems. While computational quantum mechanics can be utilized to implement simulations at the quantum mechanical level, so-called mesoscale simulations can be utilized to enable direct observations of phenomena on the order of microseconds.

Despite its temporal limitations, observations made on the scale of MD simulations may have implications on the macroscale, which can be tested experimentally. Results from MD simulations may also be extrapolated to give predictions about real systems.

4.2 CHARMM Force field

4.2.1 Potential energy function

As mentioned above, the interactions between atoms and molecules in the system are governed by potential energies, which are described through a force field and applied to pairwise interactions of atoms. The force field is a function of the

atomic positions, and is usually separated into local terms (bond length strain, bond angle strain and dihedral strain) and nonlocal terms (Lennard-Jones potential, Coulombic potential):

$$U = U_{bonds} + U_{angle} + U_{dihedrals} + U_{LJ} + U_{Coul}. \tag{9}$$

The nonlocal (or nonbonded) terms are applied for all atoms separated by three or more covalent bonds [66]. The CHARMM force fields are force fields designed for biomolecular systems[70], and several different versions exist. The CHARMM22[66] force field is designed for simulation of proteins, while the CHARMM27[68, 69] force field is designed for simulation of DNA, RNA and lipids. Their potential energy function also includes a Urey-Bradley term and an improper dihedral term:

$$U = U_{bonds} + U_{UB} + U_{angle} + U_{dihedrals} + U_{impropers} + U_{LJ} + U_{Coul}. \tag{10}$$

The potential energy terms for these force fields are further given as follows:

$$U_{bonds} = \sum_{bonds} K_b(b - b_0)^2 \tag{11}$$

$$U_{UB} = \sum_{UB} K_{UB}(S - S_0)^2 \tag{12}$$

$$U_{angle} = \sum_{angle} K_\theta(\theta - \theta_0)^2 \tag{13}$$

$$U_{impropers} = \sum_{impropers} K_{imp}(\varphi - \varphi_0)^2 \tag{14}$$

$$U_{dihedrals} = \sum_{dihedrals} K_\chi(1 + cos(n\chi - \delta)) \tag{15}$$

$$U_{LJ} = \sum_{nonbond} \epsilon_{ij} \left[\left(\frac{R_{min_{ij}}}{r_{ij}} \right)^{12} - \left(\frac{R_{min_{ij}}}{r_{ij}} \right)^6 \right] \tag{16}$$

$$U_{Coul} = \sum_{nonbond} \frac{q_i q_j}{e r_{ij}}. \tag{17}$$

4.2.2 Local terms of the potential energy function

In the first four equations, K_b, K_{UB}, K_θ and K_{imp} are the bond, Urey−Bradley, angle and improper dihedral angle force constants, while b, S, θ and φ are the bond length, Urey-Bradley 1,3-distance, bond angle and improper torsion angle, respectively. The variables with a subscripted zero refer to equilibrium values, and the different strains arise because of deviations from these "ideal" reference values. All variables with a subscript zero, as well as all force constants, are given by the empirically obtained parameters contained in the force field, for
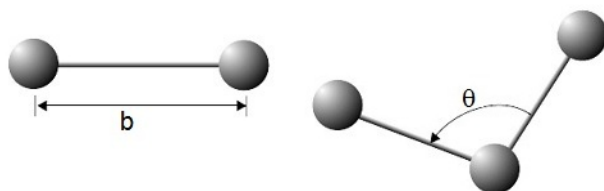
30

Figure 22: Illustrations of bond length and bond angle between two atoms. Adapted from [71].
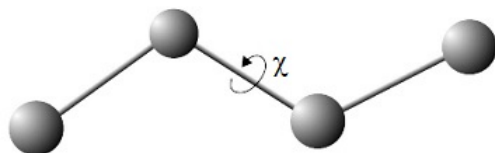

Figure 23: Dihedral angle describing the rotation about a central bond in a molecule. Adapted from [72].

different chemical compounds[70]. The Urey-Bradley term accounts for so-called 1,3-interactions, that is, interactions between atoms that are bonded to one common neighbour in between. The improper dihedral angle term is employed to compensate for "out-of-plane"-movements.

In figure 22, illustrations of the bond length and bond angle can be seen.

For the dihedral term, K_χ is the dihedral angle force constant, χ is the dihedral angle, n is the multiplicity of the function and δ is the phase shift. For four successively linked atoms, the dihedral angle χ accounts for torsional rotations about the central bond (Fig. 23).

4.2.3 Nonlocal terms of the potential energy function

The Lennard-Jones (LJ) potential (Eq. 16) gives the potential between pairs of permanent or induced dipoles as a function of distance , and is constructed by the sum of relatively long-ranged van der Waals attraction ($\propto -1/r^6$), and short-ranged repulsion ($\propto 1/r^{12}$) (Fig. 24) [73].

While the inverse 6-power dependence of the attractive term has a physical foundation, the inverse 12-power dependence of the repulsive term is chosen out of mathematical convenience[73]. In the CHARMM notation, $R_{min_{ij}}$ gives the interatomic distance where the LJ-potential is at minimum, and ϵ_{ij} is the LJ well depth[68]. The LJ-potential, can only be considered up to a certain cutoff radius, so a method of truncation should be applied. So-called switching functions are commonly used, which smoothly steers the potential to zero between an inner
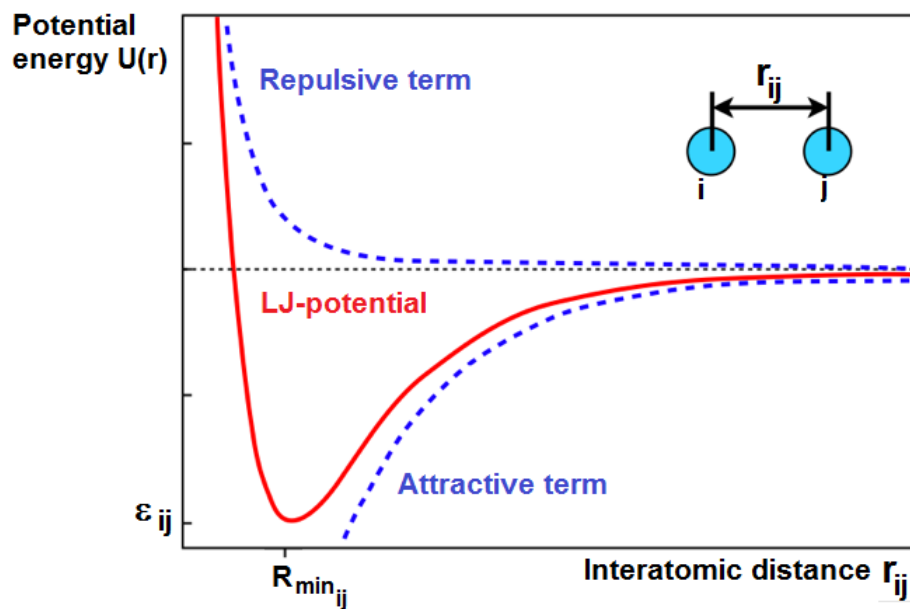
Figure 24: Illustration of the Lennard-Jones potential, which changes from being attractive to repulsive at interatomic distances shorter than $R_{min_{ij}}$. The Lennard-Jones potential describes the relatively long-ranged van der Waals interactions. Adapted from [74].

and outer cutoff radius[66].

The Coulomb potential (Eq. 17) accounts for electrostatic interactions between fully or partially charged atoms. e is the dielectric constant, and is always set to 1 for the CHARMM22/27 force fields [66, 68]. The Coulomb potential decays slowly with distance, unlike the Lennard-Jones potential. To account for these long-range interactions, so-called "fast-electrostatics algorithms" can be applied, to increase the time efficiency of the simulation[70]. The Particle-Particle Particle-Mesh (PPPM) method is a commonly applied method, which involes separating the electrostatic interactions into short- and long-ranged forces[75]. Up to a certain cutoff radius, the pairwise interactions are computed in real-space (Particle-Particle interactions). Interactions beyond the cutoff radius are approximated by mapping all atomic charges onto a three-dimensional mesh, using Fast Fourier Transformations, and interpolating the electrostatic fields from the mesh points back to the real-space atomic positions (Particle-Mesh). The PPPM method is considered highly accurate, and not too computationally demanding[76].

4.2.4 PDB/PSF/Parameter and Topology files

The Protein Data Bank (PDB) is an archive containing structural data of biological macromolecules, which is often acquired through X-ray analysis or NMR[77]. So-called PDB-files, which contain information about these structures, may be freely downloaded from the PDB website[78]. The information include, but is not limited to, amino acid sequences, secondary structure locations, atom coordinates (except hydrogen atoms), as well as the methods that were utilized to obtain the data. A protein structure file (PSF) contains information needed to apply a specific force field to the different molecules in a PDB-file[79]. To create a PSF-file for the CHARMM force field, a so-called topology file is needed, which defines the PDB-file information in accordance to the force field. In addition, to apply the CHARMM force field in a simulation, a parameter file is needed, which contains quantitative information about all local and nonlocal interactions between different types of atoms[79].

4.2.5 Additional information

The CHARMM 22/27 force fields are optimized for use with the TIP3P water model , which decreases the amount of computations needed[70].

CMAP correction, which corrects certain small errors concerning the protein backbone, has been applied to newer versions of the CHARMM 22 force field [67].

Part III

Materials and methods

In this part, the methods that were used for building the model are first presented. Afterwards, the general simulation procedure is presented, followed by a justification of the choice of simulation parameters.

5 Building the model

Visual Molecular Dynamics (VMD) for LINUXAMD64, version 1.9.1[80], was used for visualization of the various MD simulations, generation of PSF files, lipid membrane building, as well as editing and merging of the PDB structures. Much of the model editing was carried out by using Tool Command Language (TCL) scripts, in the VMD embedded "Tk console". The TCL scripts used for PSF generation and DNA separation were taken from the associated files of Comer et al.[81], while the others were found at various website forums. Some of the scripts required manual editing to suite this particular model.

The model basically consisted of three parts; a single-stranded DNA molecule, a single connexon (Cx26) protein (not paired up with another one) and a POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) lipid membrane, which were all merged together and solvated by addition of water and ions. The PDB file for the connexon was acquired from the Protein Data Bank[42]. The protein was aligned to the principal axes, and a PSF file was generated by using the VMD Automatic PSF Builder tool in combination with the topology file for the combined CHARMM22/CHARMM27 force field [66, 68, 69] (Fig. 25). During this process, atoms that are missing from the PDB file are also created, in particular hydrogen atoms[81].

By using the VMD Sequence Viewer tool, all the residues of the protein was manually inspected and compared with figure 9, to ensure accordance. The VMD Membrane Builder tool was thereafter used, to build a square POPC lipid bilayer with dimensions of x=120 Å and y=120 Å (Fig. 26).

The PDB/PSF files of the connexon and the membrane were then combined with the VMD Merge Structures tool. Any water molecules created during the membrane building were manually removed. The coordinates of the center of mass of the protein was then found by using a TCL script. This was used to move the membrane in such a way that the protein was located exactly in the middle of it. (Fig. 27).

Afterwards, all the water molecules were removed, and a TCL script was used to remove all lipid molecules overlapping the connexon, as well as lipids closer than 2 Å to the connexon. (Fig. 28).

The PDB file for the double-stranded DNA molecule of a specific base sequence and length was generated by using the 3D-DART DNA generator[82]. By using various TCL scripts, the two strands were then separated, and new PDB/PSF files of the single strand were created. During this process, the topol-
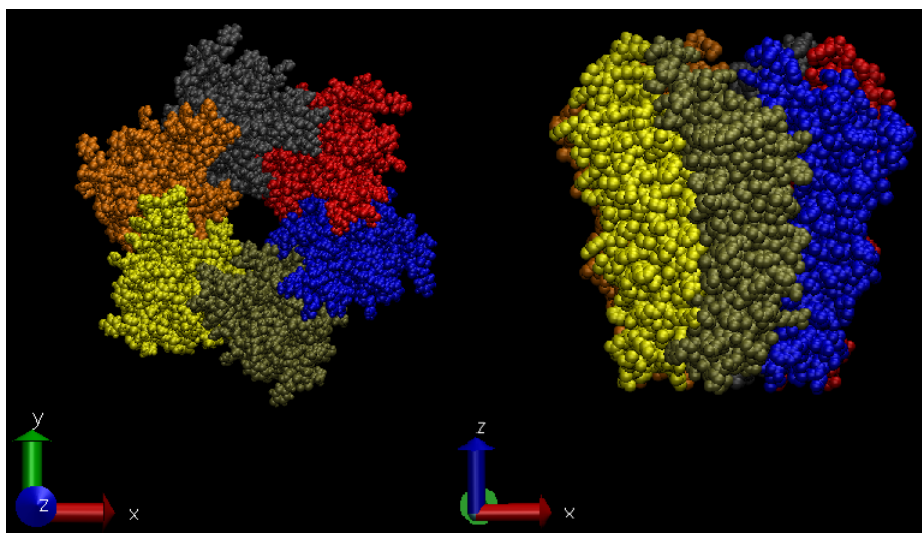
Figure 25: A connexon as seen through the channel (left) and from the side (right). Each color depicts one of the six connexin subunits.
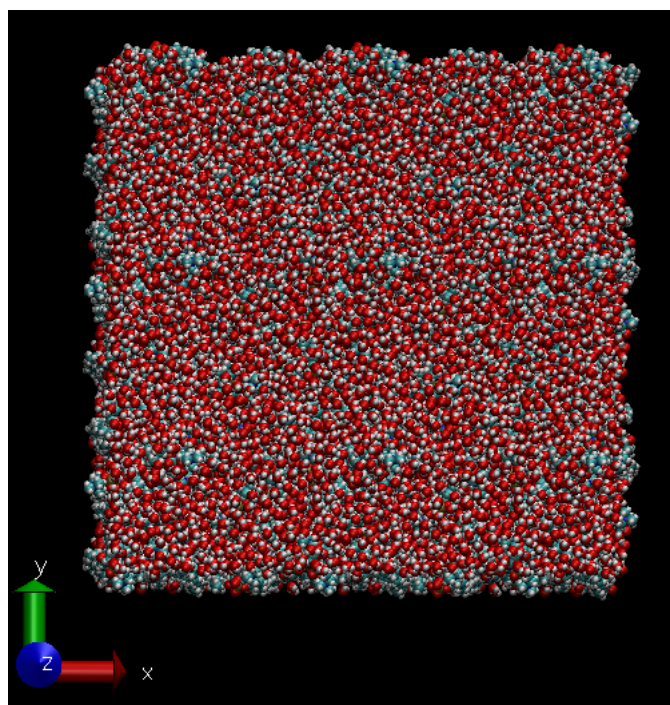
Figure 26: A POPC lipid bilayer with dimensions of $X = 120\mathring{A}$ and $Y = 120\mathring{A}$, created with the VMD Membrane Builder tool.
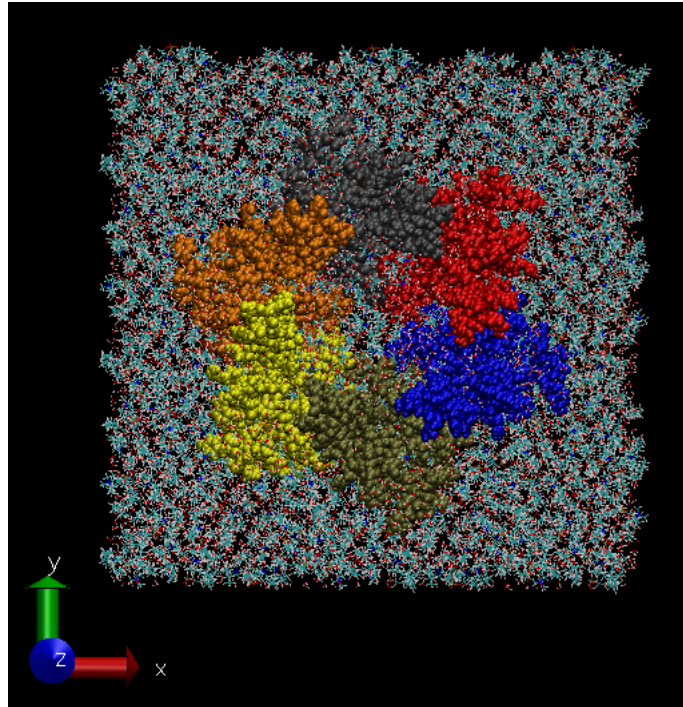
Figure 27: The connexon structure displaced in the middle of the lipid bilayer. Notice the overlap of molecules.
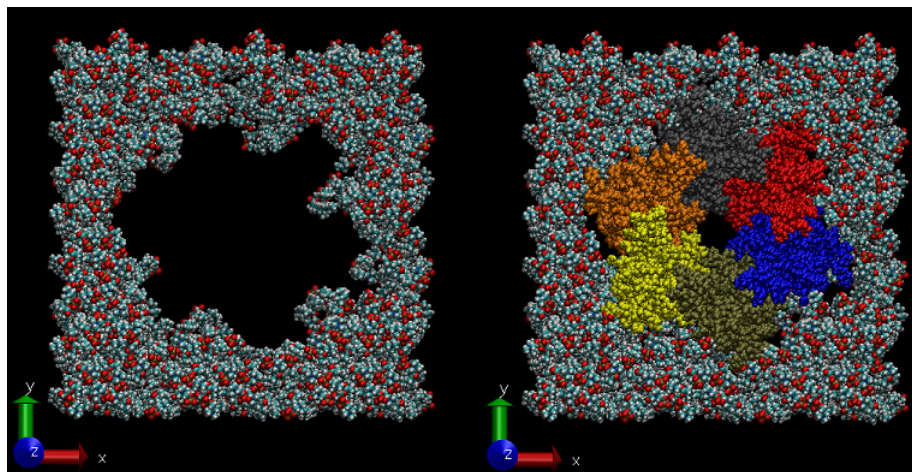
Figure 28: The lipid bilayer after removal of overlapping molecules, depicted without (left) and with (right) the connexon.

ogy file for the CHARMM27 force field[68, 69] was also applied. The DNA strand was finally combined with the connexon/membrane-system by again using the "Merge structures" tool. The DNA molecule was then moved along the positive z-axis to be placed on the cytoplasmic side of the protein channel. A 24 bases long DNA molecule was used in all simulations, which is a bit longer than the connexon, that is $24 \cdot 0.34^{nm}/_{bases} = 82\mathring{A}$, compared with a connexon length of about $77\mathring{A}$. For the main simulations, the DNA molecule was displaced so that the lowermost part of it was at approximately the same z-level as the uppermost part of the protein residues (Fig. 29).

After placing the DNA molecule on the cytoplasmic side of the connexon, the molecule also had to be rotated 180 degrees, to obtain an orientation with the 3' end facing the channel (See section 7.4). The rotation was carried out by using a TCL script that applied a rotation matrix to the nucleic acids.

The PDB/PSF files containing the DNA, connexon and membrane was converted to a LAMMPS data file by using the integrated LAMMPS tool "ch2lmp" (charmm to lammps) in combination with the CHARMM force field files [66, 68, 69]. During this conversion, a solvation box consisting of explicit TIP3P water molecules and Na^+/Cl^- ions of a specified concentration were also added to the system. The ratio of the Na^+ and Cl^- ions was automatically set to ensure charge neutrality of the system. Because of the use of periodic boundary conditions during the simulations, it was important to add a box large enough to avoid interactions between the supercells, especially regarding the DNA along the z-axis. Because of this, the solvation box always included a "buffer" of water between the DNA end facing the simulation border, and the border itself. This involved defining the following rectangular system sizes, : $X = 133\mathring{A}$, $Y = 133\mathring{A}$, $Z = 207\mathring{A}$, which resulted in a total atom number of $\sim 350'000$.
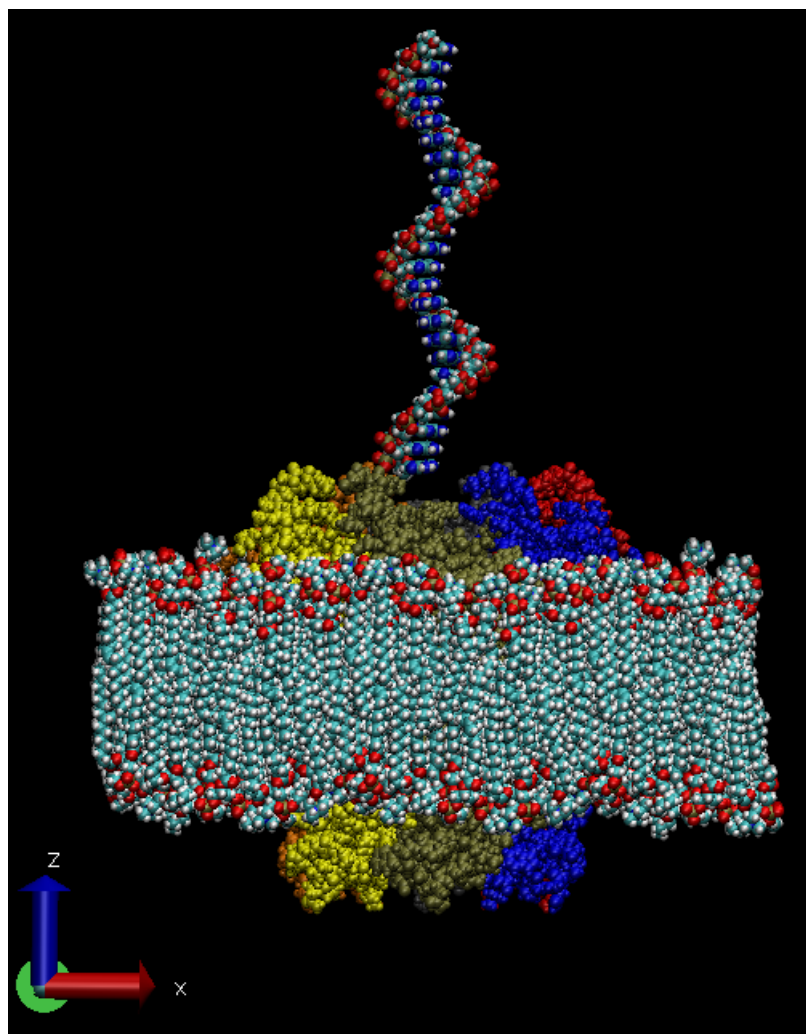
Figure 29: The full model seen from the side, showing a DNA molecule displaced directly above the connexon displaced in the lipid bilayer. The solvent is omitted from this illustration.

6   Simulation procedures

For the simulations, Large-scale Atomic-Molecular Massively Parallel Simulation (LAMMPS)[83] version *29Apr12* was used.

6.1   Initial system setup

For all simulations, periodic boundaries was used. The inner cutoff radius for LJ interactions was set to 8.0 Å, and the outer cutoff radius was set to 10.0 Å. A switching function was set to smooth the potential to zero between the inner and outer cutoff. The cutoff radius for Coulombic interactions was also set to 10.0 Å; within this radius electrostatic interactions were calculated from equation 17, while a PPPM solver was defined to calculate long-range electrostatic interactions beyond the cutoff value. A random number generator was used to create an ensemble of initial atom velocities, with 310 K as the defined initial system temperature. To decrease simulation time, bond and angle constraints were applied to all water molecules during the whole simulation. This was accomplished by using the SHAKE algorithm[84].

6.2   Energy minimization

To obtain a local potential energy minimum of the system before starting the molecular dynamics, a version of the conjugate gradient algorithm[85] was applied at the beginning of all simulations. This involved minimizing the total potential energy of the system by iteratively adjusting the atom coordinates. The minimization was set to run a maximum of 200 iterations.

6.3   Equilibration and molecular dynamics

After energy minimization, the system was equilibrated under NVE ensemble for 8.65 ps. During this period, a Berendsen[86] thermostat was also applied to the system, to keep the temperature fluctuations around 310 K. The first 0.05 ps of this equilibration was carried out without any constraints.

6.3.1   Equilibration with constraints

After the first 0.05 ps of the NVE equilibration, constraints were applied on the membrane, protein and DNA during the last 8.60 ps. Constraints were applied to the DNA molecule to inhibit DNA-protein attachment, because of the positively charged residues at the entrance. Membrane constraints were applied to avoid membrane displacement during application of the voltage, while protein constraints were applied to inhibit extensive protein deformations. The constraints were analogous to attaching one end of a spring to the initial atom position and the other end to the atom itself. Thus, each atom would experience a spring force proportional to the displacement away from the initial position. A spring working in the z-direction was applied for the membrane, while two

springs; one working in the z-direction and one working in the xy-plane, were applied for the protein and DNA-atoms. The spring constant of the spring working in the z-direction on the membrane atoms was set to $K = 0.1\,Kcal/mole\mathring{A}^2$. The spring constant of the spring working in the z-direction on the protein atoms was also set to $K = 0.1\,Kcal/mole\mathring{A}^2$, while the one working in the xy-plane was set to $K = 0.05\,Kcal/mole\mathring{A}^2$. For the DNA atoms, both the spring constant in the z-direction and the one in the xy-plane were set to $K = 10.0\,Kcal/mole\mathring{A}^2$. Several test simulations were carried out to find appropriate values of the harmonic constraints.

After the 8.65 ps of NVE equilibration, the system was equilibrated for 100 ps under a NPT ensemble of zero external pressure and a temperature of 310 K, under the same constraints as defined above.

6.3.2 Translocation dynamics with constraints

To simulate the dynamics of DNA-translocation, a uniform electric field was applied to all the atoms in the system, analogous to an experimentally applied transmembrane voltage. When applying electric fields, the equation for uniform electric fields, $E = \Delta V/d$, where ΔV is the transmembrane voltage and d is the membrane thickness, was considered. Voltages between 40 mV and 8.4 V were applied, and with a membrane thickness of 38 Å[42], this corresponded to electric fields ranging from 0.001 V/Å to 0.221 V/Å. During all simulations, the constraints set on the DNA-atoms were removed before applying the voltage. The translocation dynamics were also carried out under a NVT ensemble at 310 K, and lasted up to a couple of nanoseconds. The simulations were manually aborted, either when translocation was observed or when the visualizations gave strong indications that translocation was unlikely to occur (within a reasonable simulation time range).

6.3.3 Equilibration and translocation dynamics with DNA constraints only

For some simulations, constraints were only applied on DNA during equilibration; the same as described in the previous section. During translocation dynamics, constraints were applied on the membrane only. These constraints were analogous to a spring working in the z-direction, with a spring constant of $K = 0.1\,Kcal/mole\mathring{A}^2$. Except from this, the simulation protocol was identical with the one described in the previous sections.

7 Simulation parameters

7.1 Transmembrane voltage magnitudes and connexon constraints

It is desirable to use as small voltage magnitudes as possible in the simulations, but they still need to be large enough to induce translocation over a timespan

of maximum a few nanoseconds. Voltages of several volts therefore had to be applied. Applied voltages of several volts are way too high to be used in real-life experiments, since rupture of the membrane is observed at about ± 300 mV[87]. However, since bond-breakage is not accessible in these simulations, both the membrane and the connexon are instead greatly deformed and gain various velocities. Translocation through deformed connexon channels in membranes that would actually rupture in real-life experiments is clearly not too interesting to simulate. For most simulations, it was therefore decided to apply constraints on the connexon channel, to inhibit radical deformations. The drawback by doing this, is that higher voltages are needed to accomplish translocation, which again is further away from real-life experiments. A trade-off had to be made, and it was decided that non-deformation of connexon channels (as a result of constraints) was more important than keeping the voltage low. The connexon PDB-file also depicts the connexon channel in an open conformation, which is a desirable condition for real translocation experiments. Section 9.0.6 discusses more sophisticated methods to overcome challenges concerning voltage magnitudes and simulation durations.

It could be argued that the gating mechanism of the connexon channel would disturb DNA-translocation, but since the gating mechanism takes place on the order of several seconds, it could safely be neglected (See section 2.5.2).

7.2 Membrane and DNA constraints

During the preliminary simulations, it was observed that the application of a voltage resulted in displacements of the whole system, in particular the membrane. For some simulations, this resulted in a "rotating" behavior of the membrane. During equilibration of the system, the negatively charged phosphate groups of the DNA molecule and the positively charged residues at the cytoplasmic connexon entrance were attracted to each other, resulting in DNA attachment and no translocation – even at high voltages. Constraints were therefore applied on the membrane during the whole simulation, and on the DNA molecule during equilibration.

7.3 Ion concentrations

Simulations were carried out in a solution with ion concentrations of 0.1 M and 1.0 M. In accordance with the theory (Section 3.3), the persistence length (and thereby also the stiffness) of the DNA molecule seemed to be larger at 0.1 M than at 1.0 M. This could be seen because of the increased capability of entering the channel in a straight conformation, at concentrations of 0.1 M. For subsequent simulations, ion concentrations of 0.1 M was therefore used. 1.0 M KCl is however the typically used ion concentration[13, 88].
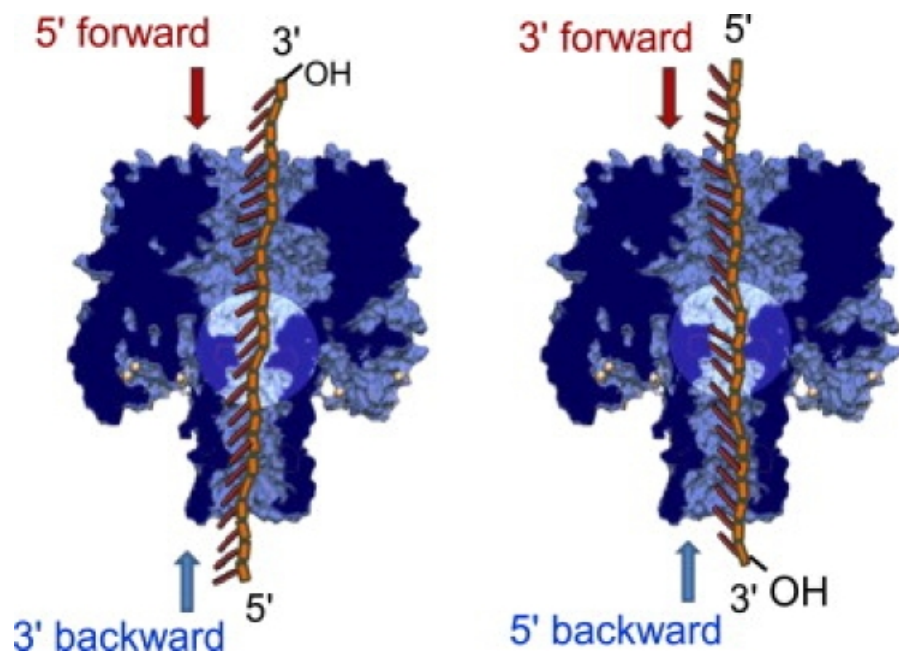
Figure 30: Illustration of the orientational dependence on DNA conformation during translocation. DNA is found to translocate easier with the 3'end pointing in the forward direction.[89]

7.4 DNA orientation

Experimental and theoretical studies have shown the importance of the orientation of the DNA molecule during translocation through α-hemolysin[88, 89], where larger frictional forces occur for translocations where the 5' end enters the channel first. This is proposed to be due to an effect compared with forcing a tree through a door, where the branches of the tree tilt in the same or opposite direction as the translocation direction, analogous to a comparable tilt of the DNA bases (Fig. 30).

To increase the probability of translocation, all simulations were carried out with the 3' end facing the opening of the connexon channel.

Part IV
Results and discussion

The first section of this part presents main simulation results as well as corresponding discussions. The next section includes discussions when considering published literature.

8 Simulation cases

In this part, the simulation results are presented together with general discussions. Some of the discussions are separated into parts numbered with greek letters (I, II, etc.), mainly to ease the reading. Comparison with published literature will later be discussed in section 9. When describing the different simulations, DNA observed in a particular simulation may for example be referred to as "8.4 V-DNA", if it was used in a simulation with an applied voltage of 8.4 V. The simulation were 8.4 V was applied will also occasionally be referred to as the "8.4 V-simulation". Similarly, a 24 bases long DNA molecule consisting of only adenine bases may be referred to as "24A-DNA", and "24C-DNA" if it only contains cytosine, etc.

There are three main result sections, which all focus on translocation times: In section 8.1, translocation of 24A-DNA without connexon constraints are presented. Section 8.2 describes similar results, but with connexon constraints applied. Finally, in section 8.3, the differences in translocation times between 24A-, 24C-, 24G- and 24T-DNA with connexon constraints are presented.

8.1 Translocation times - membrane constraints only

Simulations were carried out to find the translocation time dependence on the voltage magnitude, with constraints applied on the membrane only. These simulations were carried out on a 24 bases long DNA molecule consisting of only adenine bases.

Table 1 shows that the minimum "threshold" voltage where translocation was observed, was found to be 2.4 V. When applying high transmembrane voltages with constraints on the membrane only, the connexon showed a deforming stretching behavior, especially for the positively charged residues at the cytoplasmic entrance region (Fig. 31).

The stretching was directly proportional to the applied voltage magnitude, and it could clearly be seen that it also affected the residues lining the interior of the channel (Fig. 32).

Figure 32 shows that the main channel diameter was comparably similar for voltages above and below the "threshold" voltage of 2.4 V. The protruding residues of the channel interior however, rapidly became lined up parallel with the inner channel surface for voltages above the threshold voltage, resulting in a "smoother" channel, with an effectively larger inner channel diameter.

24A - Translocation times, without constraints		
Voltage (V)	Translocation (Yes/No)	Cancelation time/Translocation time (ns)
0.04	No	1.49
0.12	No	1.63
0.6	No	1.06
1.2	No	2.27
2.4	Yes	1.81
3.6	Yes	0.64
4.8	Yes	0.57
6	Yes	0.19
7.2	Yes	0.13

Table 1: Translocation times of 24A-DNA, at different transmembrane voltages and without connexon constraints. Translocation was observed at voltages of 2.4 V and above.
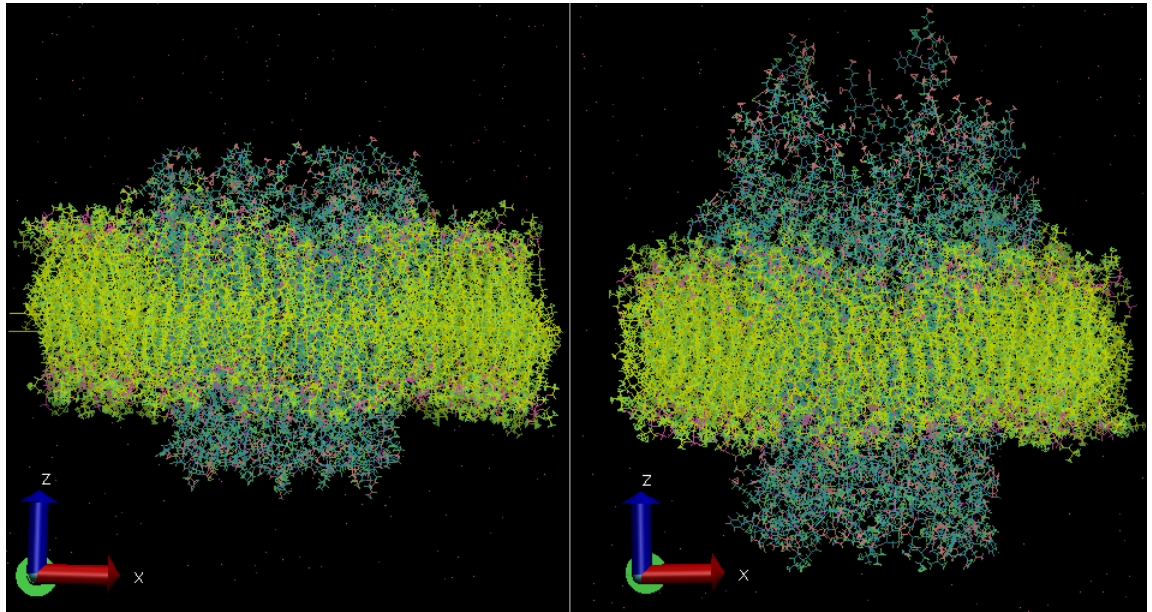

Figure 31: Snapshots of connexon (blue) embedded in a membrane (yellow), before (left) and after (right) the application of a high transmembrane voltage. At high transmembrane voltages, the connexon shows a stretching behavior.
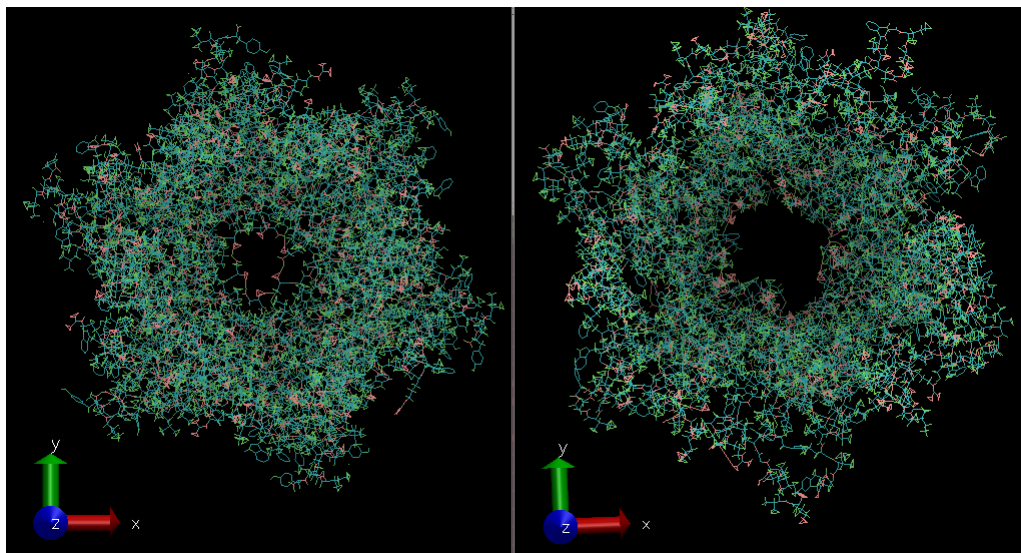
Figure 32: Connexon as seen through the channel, without connexon constraints. Left: At low transmembrane voltages, the residues lining the interior of the channel contribute to a smaller channel diameter. Right: At high voltages, the channel-lining residues become lined up with the channel, parallel with the z-direction.
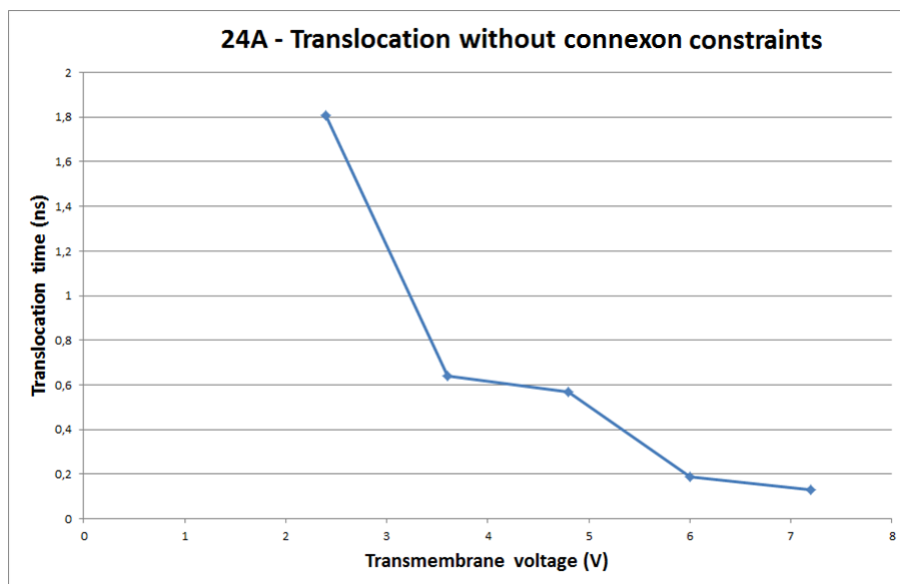
Figure 33: Plot of translocation times for 24A-DNA at different voltages, without connexon constraints.

For transmembrane voltages above the threshold of 2.4 V, the translocation time was strictly decreasing with increasing voltage magnitudes (Fig. 33).

Figure 33 shows that for voltages between 3.6 V and 7.4 V, a linear relationship may be seen between the data points. The translocation time for 2.4 V does however deviate from this linear relationship to a large extent.

8.1.1 Discussion of case 8.1

Table 1 shows that no translocations are observed for voltages of 1.2 V and lower. Figure 34 shows excerpts from the simulation at 1.2 V, and it can be seen that although the DNA molecule enters the channel in a straight manner, the voltage is not high enough for the molecule to make any further progress. This may be due to a possible energy barrier that the interior channel residues need to overcome, to be able to form a "smooth" channel (Sec. 8.1). The front of the DNA is stuck inside the channel, while the "tail" keeps getting pushed downwards. This results in a "clumping" behavior combined with a fluctuating tail, which can be seen around 800 ps. At the end of the simulation, around 2 ns, the DNA tail also gets fixated by attachment to the positive residues located around the channel entrance. A similar behavior could be seen for voltages lower than 1.2 V.

When considering figure 33, the translocation time for an applied voltage of 2.4 V clearly deviates from the linear relationship between the other voltage magnitudes. Figure 35 and 36 shows frames from 2.4V- and 3.6V-simulations
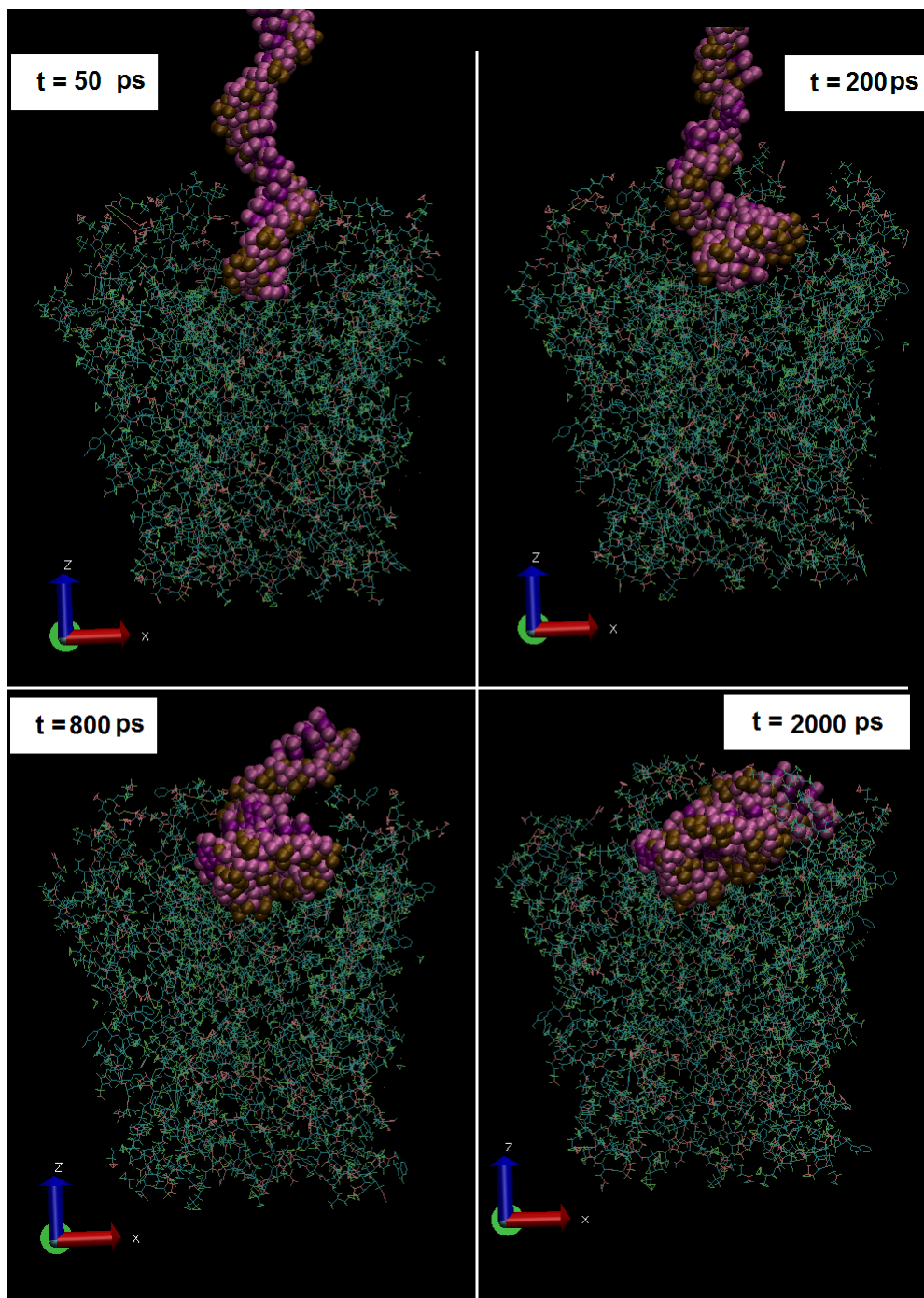
46

Figure 34: Snapshots depicting the non-translocation of 24A at various time points, under an applied voltage of 1.2 V and without connexon constraints.

for comparison. In the 3.6V-simulation, the positive residues at the entrance stretch up while the DNA molecule is forced through the middle of the channel. These opposite movements might reduce the interaction time between the DNA tail and the positive residues, thereby avoiding DNA-attachment. A similar translocation character was also seen for the voltages larger than 3.6 V. On the other hand, in the 2.4V-simulation the DNA molecule is subject to a smaller force and is exposed to the positive residues over a larger timespan. As can be seen at 350 ps, this causes the tail of the DNA to attach to the positive residues, resulting in a greatly increased translocation time. In addition, 2.4V might not be enough to completely "line up" the inner channel residues, as mentioned in section 8.1, which may be further contributing to a reduction of translocation speed.
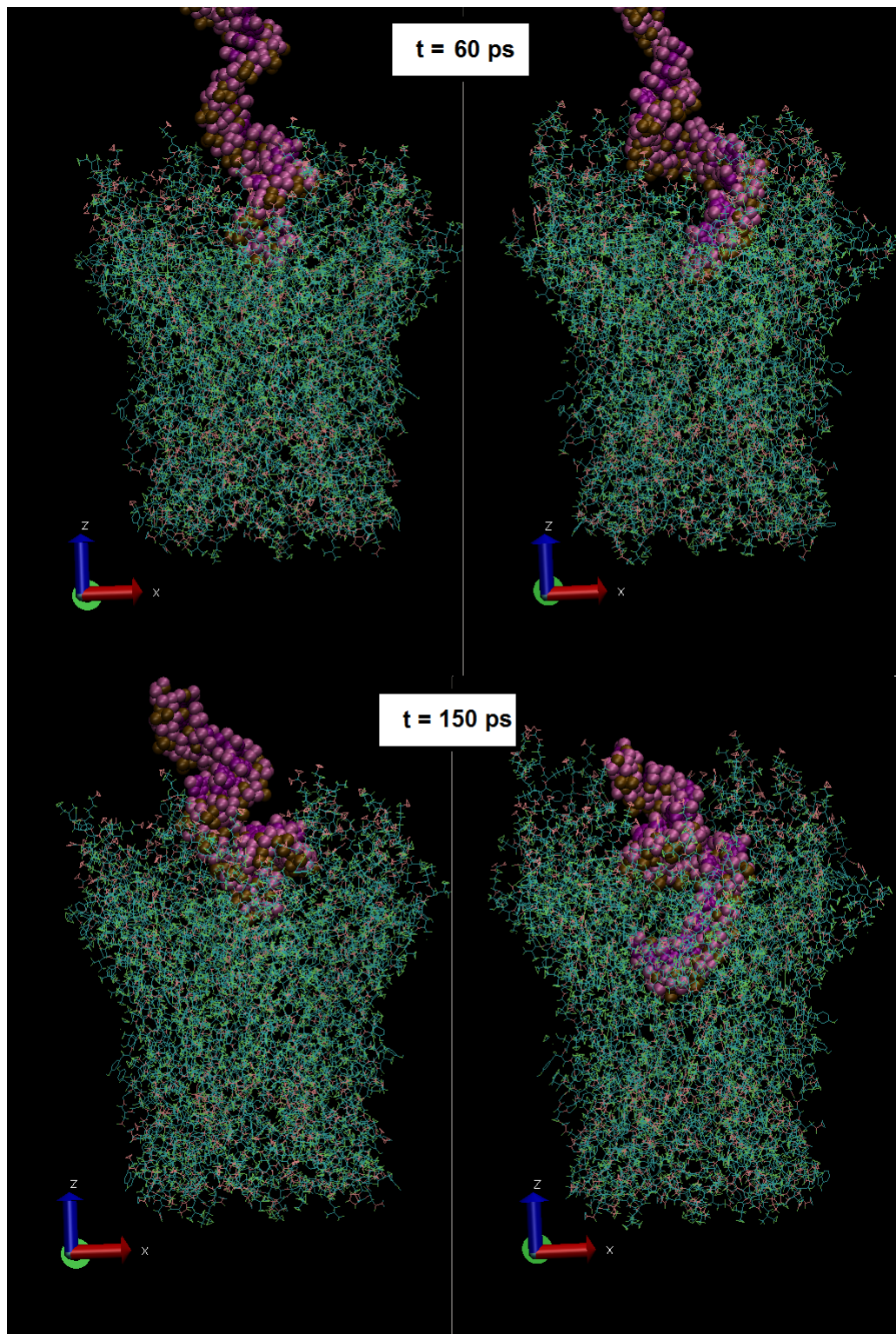
Figure 35: Snapshots showing the translocation of 24A-DNA at an applied voltage of 2.4V (left) and 3.6V (right), without connexon constraints.
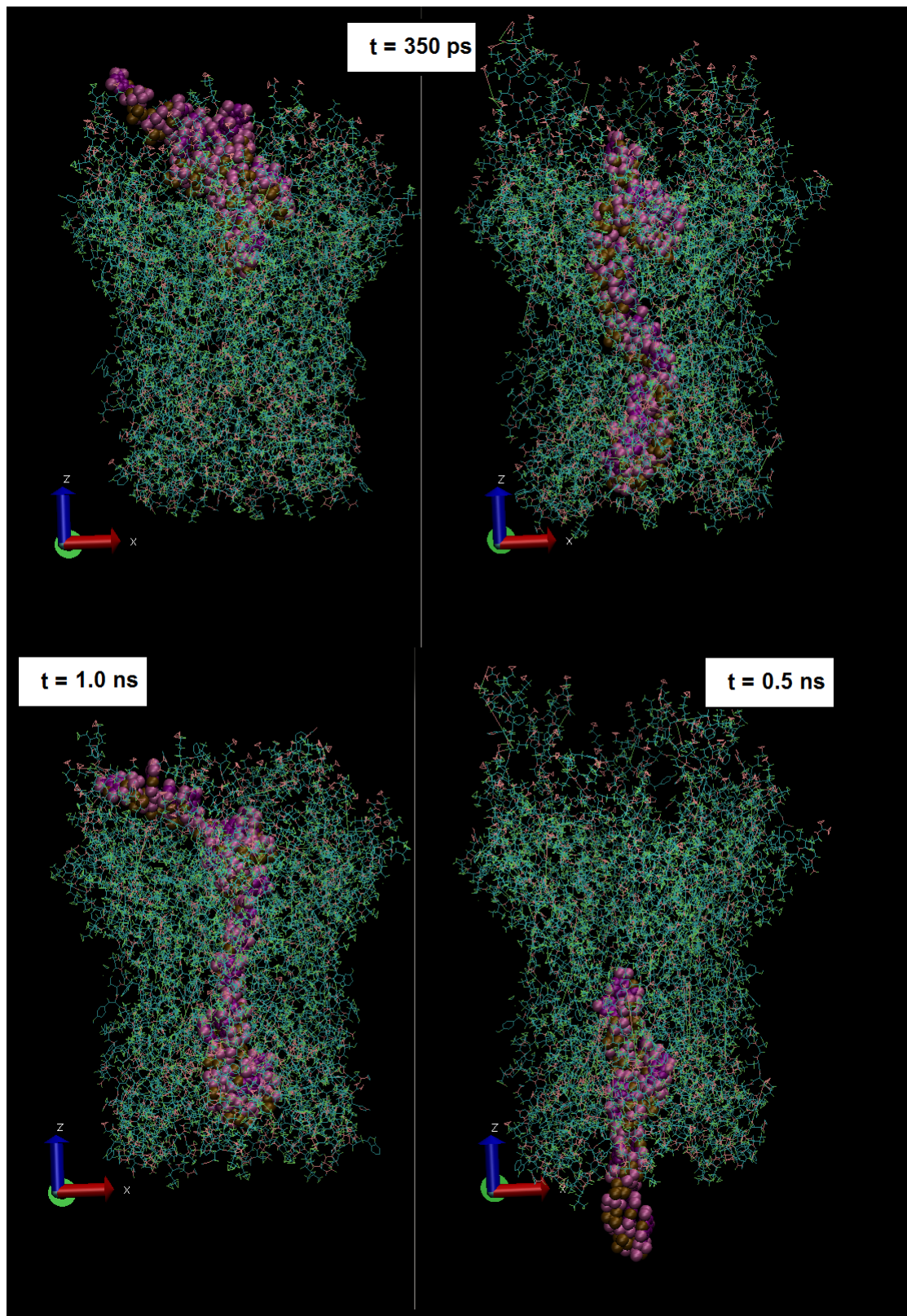
Figure 36: Snapshots showing the translocation of 24A-DNA at an applied voltage of 2.4V (left) and 3.6V (right), without connexon constraints.

24A - Translocation time, with connexon constraints		
Voltage (V)	Translocation (Yes/No)	Cancelation time/Translocation time (ns)
1.2	No	1.65
2..4	No	1.42
3.6	No	1.23
4.8	Yes	1.76
6	Yes	1.18
7.2	Yes	0.4
8.4	Yes	0.33

Table 2: Translocation times of 24A-DNA, at different transmembrane voltages and with applied connexon constraints. Translocation was observed at voltages of 4.8 V and above.

8.2 Translocation times – membrane and connexon constraints

Simulations were carried out to find the translocation time dependence on the transmembrane voltage magnitudes, with constraints applied on both the membrane and the connexon. These simulations were carried out on a 24A-DNA.

Table 2 shows that the threshold voltage for translocation during these conditions was found to be 4.8 V, which is twice as large as the voltage that was needed for translocation in the simulations without connexon constraints (Sec. 8.1).

For voltages above the threshold of 4.8 V, the translocation time was strictly decreasing with increasing voltage magnitudes. Figure 37 shows that a linear curve may be fitted to the data points, with the exception of the translocation time for 8.4 V.

8.2.1 Discussion of case 8.2 - I

As seen in table 2, translocations are only observed for voltages of 4.8 V and higher, when constraints are applied on the connexon. The constraints force the residues of the channel interior to stay protruded, which results in an effectively smaller inner channel diameter. This probably increases the minimum voltage needed for translocation, at least during the limited time range of the simulations.

Figure 38 shows three frames from the simulation under 3.6 V, which resulted in no translocation. The behavior is similar to the one observed for voltages of 1.2 V and below, without connexon constraints (Fig. 34). That is, even though the DNA enters the channel in a straight conformation, the energy barrier is too high for any further downward movements, so the DNA molecule ends in a clumped conformation. Note that in figure 38, towards the end of the simulation, the DNA tail does not attach to the positive residues at the channel entrance, probably due to the greatly reduced mobility of the positive residues.
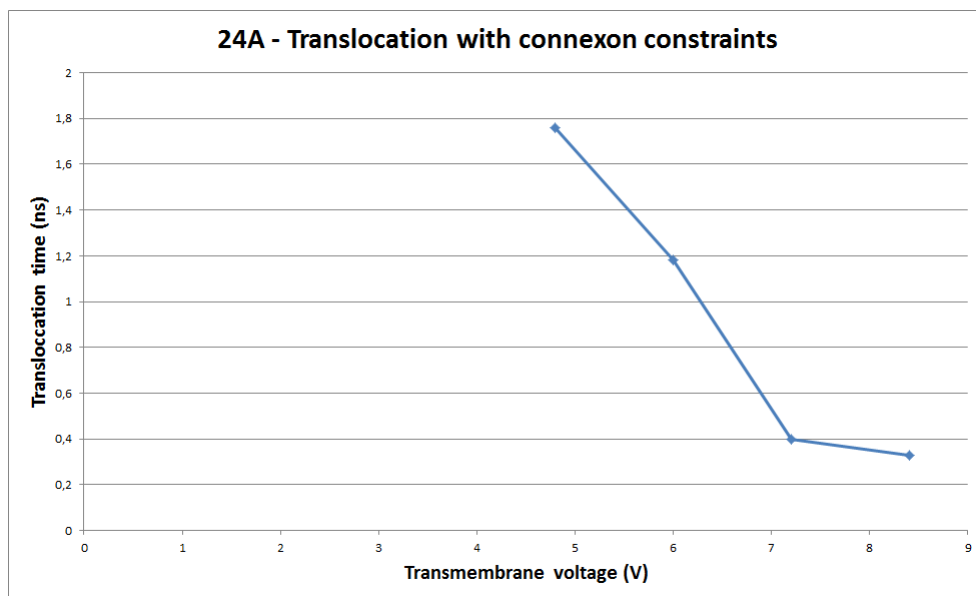
Figure 37: Plot of translocation times for 24A-DNA at different voltages, with applied connexon constraints.
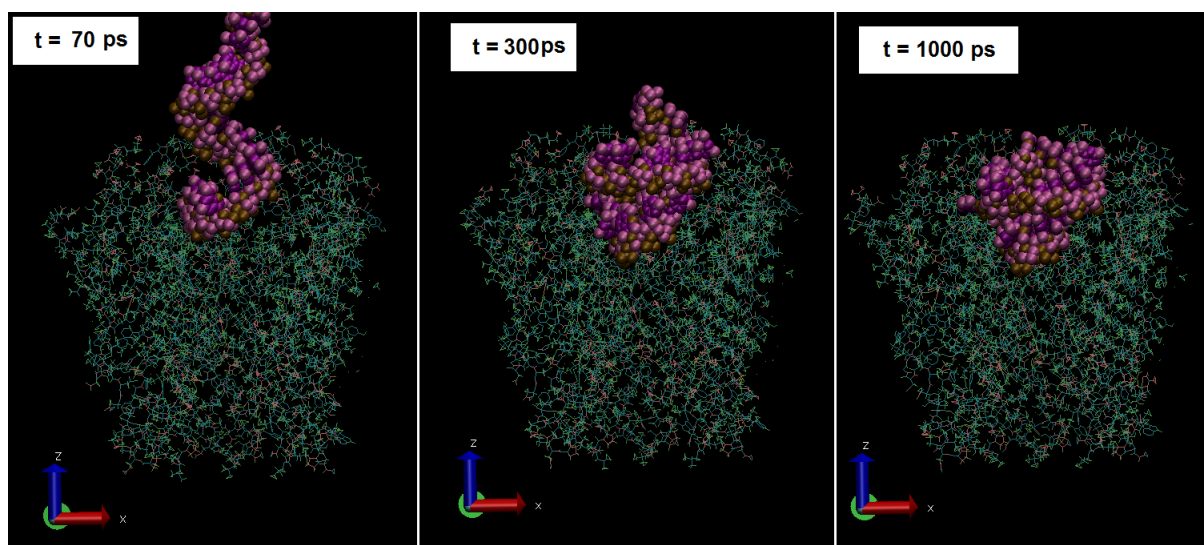

Figure 38: Snapshots showing the non-translocation of 24A-DNA at an applied voltage of 3.6 V, with connexon constraints.

8.2.2 Discussion of case 8.2 - II

Figure 37 shows a linear decrease in translocation time as the transmembrane voltage is increased, with the exception of the 8.4V-simulation. Figure 39 and 40 shows the sequence of events during 4.8 V- and 6 V-simulations. Both simulations started with the DNA entering the channel in a straight manner at 40 ps, followed by a clumping at about 175 ps. In contrast to the behavior seen in figure 34 and 38, the force exerted on the "lump" of DNA is large enough to maintain a very slow downward movement. At around 850 ps for the 4.8 V-simulation, and 950 ps for the 6 V-simulation, the DNA molecule experiences a sudden downward "release", which is followed by further stepwise progressive movements. This stepwise release was observed in many of the simulations carried out. Just before the sudden release of the DNA molecule, it seems to be at the funnel, which inhibits the downward motion for a long time period. After overcoming this barrier, the speed of the subsequent downward motion drastically increases. The increase in translocation velocity could be caused by the sudden absence of an energy barrier in combination with the fact that the region below the funnel is negatively charged (See fig. 11). It should be noted in figure 40 (at t=850/950 ps), that the 6 V-simulation needed 100 more ps to cover the same distance as the 4.8 V-simulation, even though a higher voltage was applied. This could be due to the fact that the phosphate groups of the whole DNA are pulled down more rapidly than for the 4.8 V-simulation, which results in a more compressed DNA at the beginning of the translocation (Fig. 39, t=40 ps). However, after overcoming the funnel region, the higher voltage of the 6 V-simulation accelerates the DNA in such a way that it still finishes its translocation before the 4.8 V-simulation.

8.2.3 Discussion of case 8.2 - III

When inspecting figure 37, the 8.4 V-simulation is not included in the linear relationship between the other results. That is, the difference between the 7.2 V- and the 8.4 V-simulation is very small, despite their voltage differences. Figure 41 shows a slight difference in the initial conformations of the lower part of the DNA molecules. The DNA molecule in the 8.4 V-simulation is not pointing in a straight-down direction, probably because some posterior phosphate groups have "catched up with" the leading ones, due to the large voltage. This small difference in initial conformations between the 7.2 V- and 8.4 V-simulations evolve into major differences as the DNA molecules reaches the funnel. Due to the initial slope of the DNA in the 8.4 V-simulation, its leading tip is folded back onto itself in a much larger degree than for the 7.2 V-simulation. Figure 42 shows the differences in how much the DNA molecules are folded, as they traverse the funnel. Being in a folded conformation while passing the funnel is clearly time-consuming , and figure 42 shows that a much larger amount of the DNA in the 8.4 V-simulation is folded. However, as seen in figure 43, the high voltage causes the 8.4 V-DNA to "catch up" with the 7.2 V-DNA after overcoming the funnel barrier.

Figure 39: Snapshots showing the translocation of 24A-DNA at an applied voltage of 4.8 V (left), and 6 V (right), with connexon constraints.
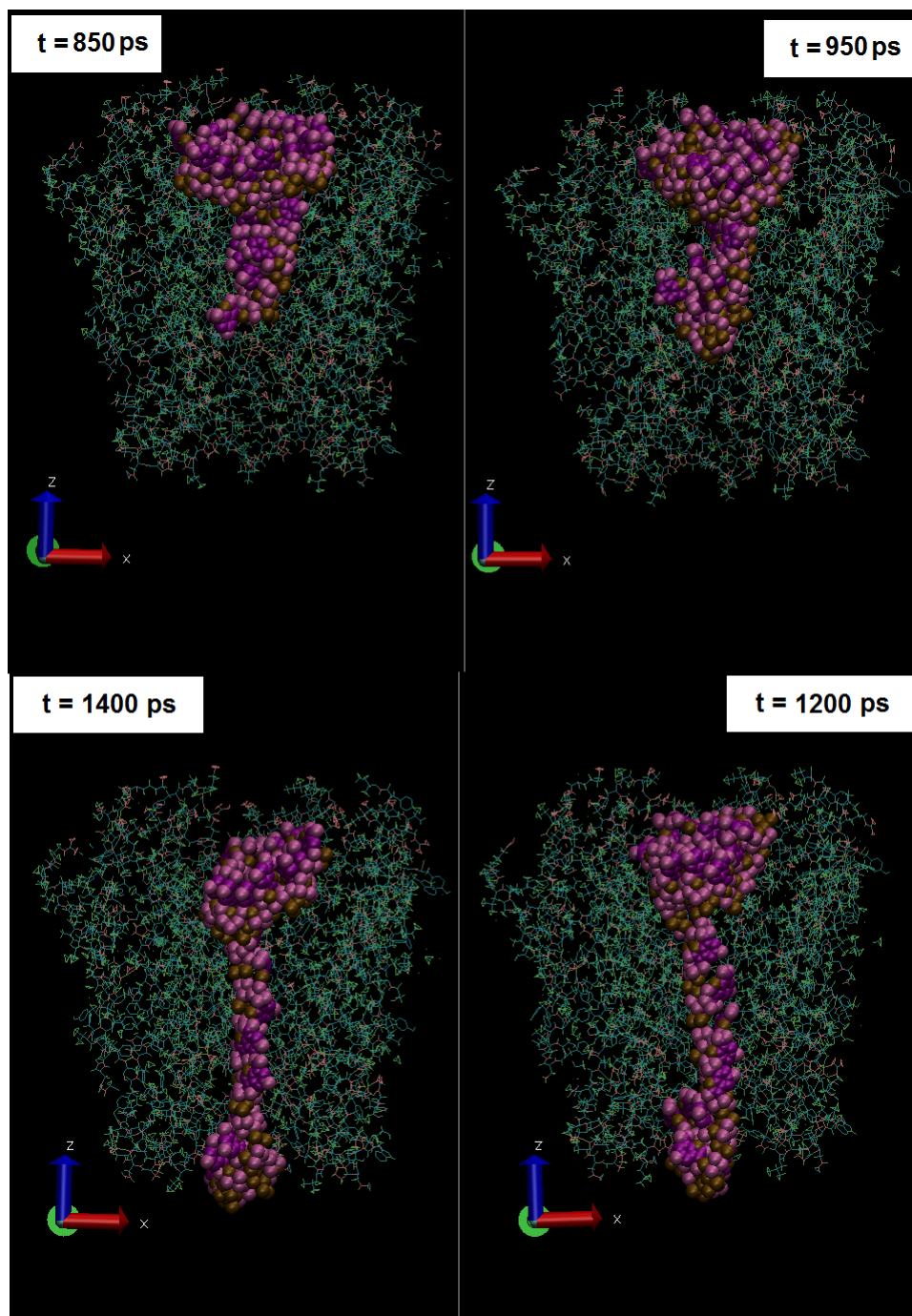
Figure 40: Snapshots showing the translocation of 24A-DNA at an applied voltage of 4.8 V (left), and 6 V (right), with connexon constraints.
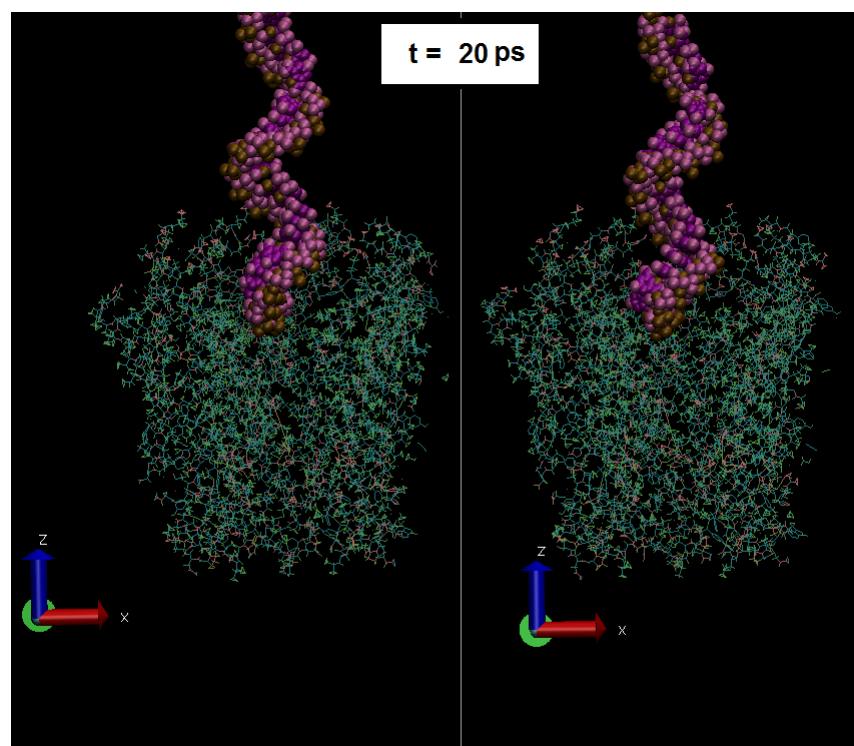
Figure 41: Snapshots showing the translocation of 24A-DNA at an applied voltage of 7.2 V (left), and 8.4 V (right), with connexon constraints.
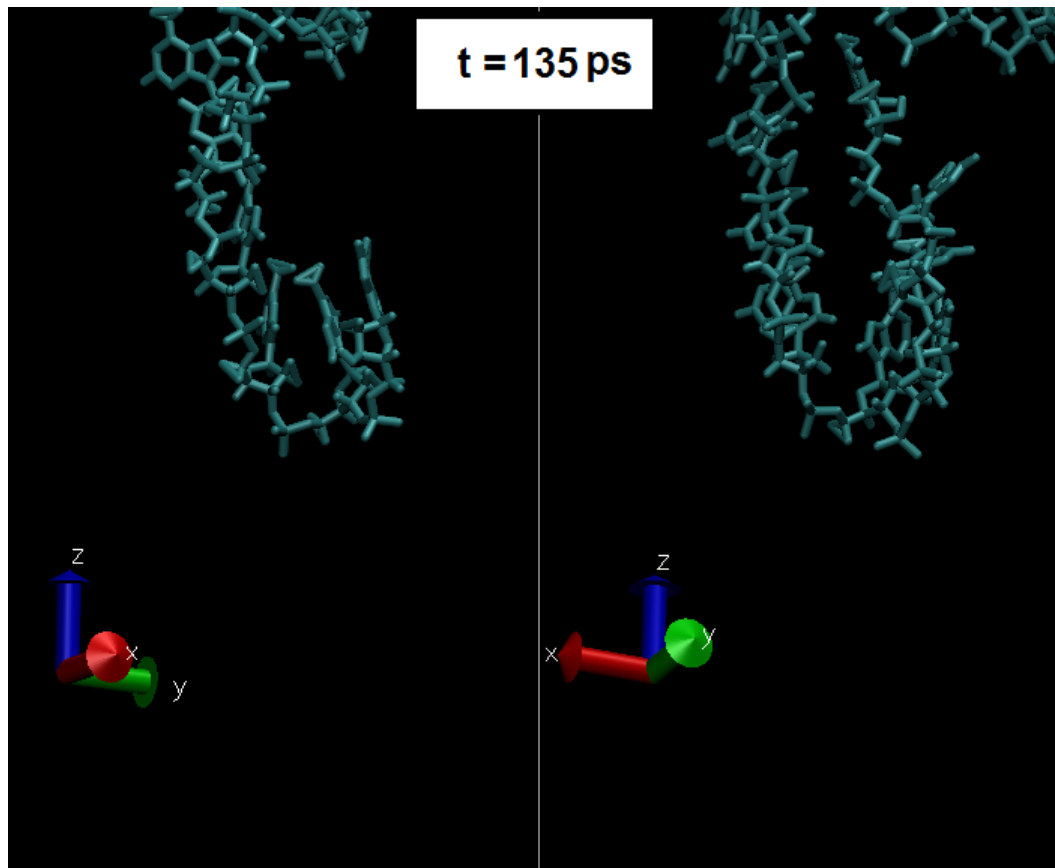
Figure 42: Detail from the translocation of 24A-DNA at an applied voltage of 7.2 V (left), and 8.4 V (right), with connexon constraints. The snapshot is taken during translocation through the narrow funnel region. Note the higher amount of folding for the 8.4 V-simulation. For clarity, only the DNA molecules are depicted.
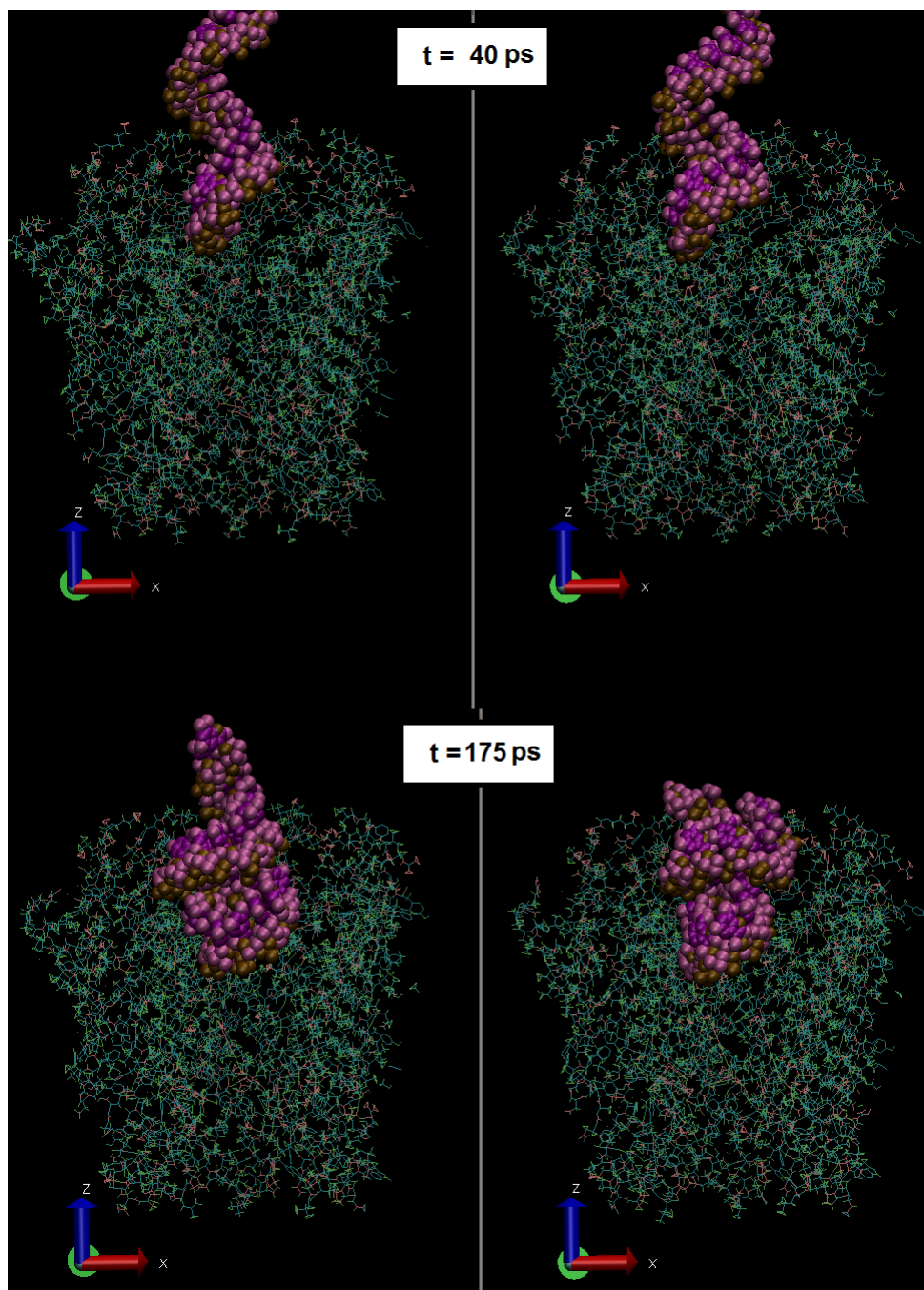
Figure 43: Snapshots showing the translocation of 24A-DNA at an applied voltage of 7.2 V (left), and 8.4 V (right), with connexon constraints.
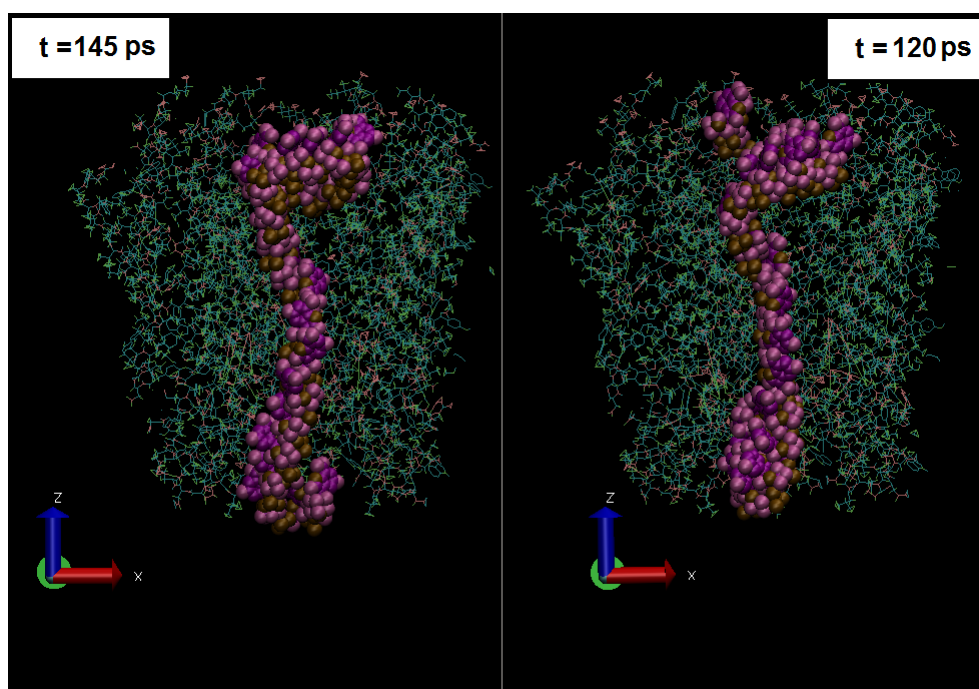
24A/C/G/T Translocation time comparison, with connexon constraints				
Transmembrane voltage (V)	24A (ns)	24C (ns)	24G (ns)	24T (ns)
6	1.18	0.21		0.36
7.2	0.4	0.16	0.34	0.36
8.4	0.33	0.12	0.57	0.19

Table 3: Translocation times of 24A-, 24C-, 24G- and 24T-DNA, at different transmembrane voltages and with applied connexon constraints. (Translocation was not observed for 24G-DNA at 6 V, and the simulation was canceled after 1.18 ns).
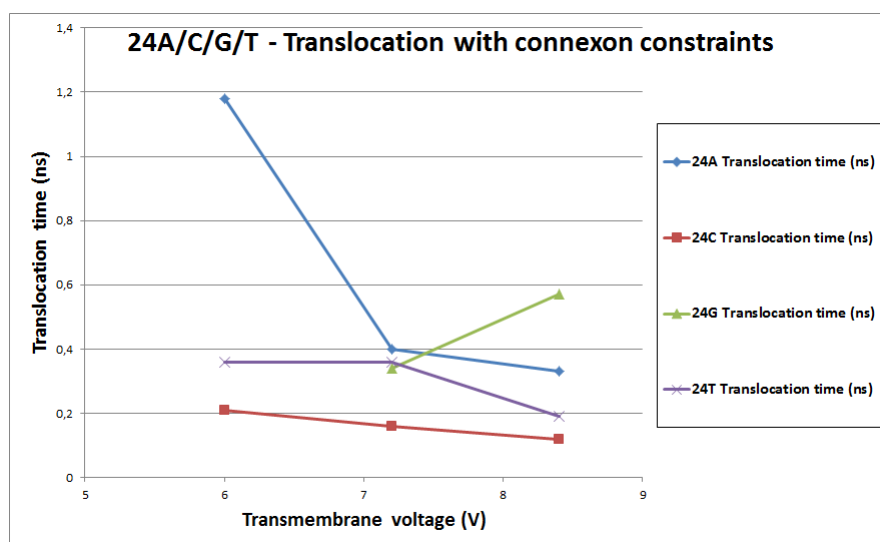
Figure 44: Plot of translocation times for 24A-, 24C-, 24G- and 24T-DNA at different voltages, with applied connexon constraints.

8.3 Translocation times – comparison between A, C, G, T-bases

Simulations were carried out to find the translocation time differences between the four different bases, adenine (A), guanine (G), cytosine (C) and thymine (T). Constraints were applied on both the membrane and the connexon during all simulations. Table 3 and figure 44 shows the results for the four bases.

It can be seen that the translocation times are decreasing (albeit not strictly) with increasing voltage magnitude, with the exception of 24G. At an applied voltage of 6 V, no indications of imminent translocation was observed, so the simulation was canceled after 1.18 nanoseconds.

8.3.1 Discussion of case 8.3 - I

Table 3 and figure 44 implies a difference in translocation opposition between the bases, where G opposes translocation the most, followed by A, T and C. It is interesting to note that this relationship is directly transferable to the differences in molecular sizes of the bases (Fig. 16). When the bases are attached to the sugar-phosphate backbone, the effective diameter of the single-stranded DNA should be larger for the purines (A and G) than for the pyrimidines (C and T). (This difference would probably not be observed by double-stranded DNA, because the complementary between large and small bases would even out the diameter differences of the DNA molecule.)

8.3.2 Discussion of case 8.3 - II

From figure 3, it should be questioned why the translocation time is the same, 0.36 ns, for 24T at 6 V and 7.2 V. By inspecting the snapshots from the two simulations (Fig. 45 and 46), it can be seen that the initial "compression" of the DNA molecule in the 7.2 V-simulation results in a non-beneficial "clumped" translocation. As seen previously, after passing the narrow funnel, the simulation with a larger applied voltage catches up with the other one, which causes the molecules to exit the connexon at the same time point.

The high degree of foldedness for 24T at the 7.2 V-simulation could also explain why the translocation time is comparable with the 7.2V-simulations of 24A and 24G (Fig. 44). According to the size differences, as discussed in section 8.3.1, 24T should slip more easily through the channel. The simulations of 24A and 24G at 7.2 V both showed straighter conformations, leading to the assumption that a beneficial conformation with a low degree of foldedness could be equally important to the size of the bases. In this context, it should be noted that the 24C-translocations were significantly faster than the other ones, and all of these simulations included "straight-down" translocations without any major foldedness. An example of this type of translocation can be seen in figure 47. It could be discussed whether 24C-DNA possesses a smaller diameter than 24T, but it may be said to be smaller than 24A and 24G. It is therefore plausible to assume that the probability of adopting a beneficial straight conformation is correlated to small base dimensions.

This hypothesis is further supported by the simulations with the large 24G DNA. Of the three simulations carried out, both the 6 V- and the 8.4 V-simulation involved a high degree of foldedness. At 40 ps in the 6 V-simulation, figure 48 shows a "tilted" conformation of the DNA tip, with respect to the z-axis. This results in a high degree of foldedness, and the "lump" of DNA is stuck in the channel from around 220 ps till the end of the simulation at 1.18 ns, resulting in no translocation.

8.3.3 Discussion of case 8.3 - III

When comparing the 8.4 V-simulation with the 7.2 V-simulation (Fig. 49 and 50), the initial conformation of the 7.2 V-DNA is much straighter than the 8.4

Figure 45: Snapshots showing the translocation of 24T-DNA at an applied voltage of 6 V (left), and 7.2 V (right), with connexon constraints.

Figure 46: Snapshots showing the translocation of 24T-DNA at an applied voltage of 6 V (left), and 7.2 V (right), with connexon constraints.

Figure 47: Snapshots showing the translocation of 24C-DNA at an applied voltage of 6 V, with connexon constraints. Note the absence of substantial DNA-folding.

Figure 48: Snapshots showing 24G-DNA at an applied voltage of 6 V, with connexon constraints. Note the high degree of foldedness, resulting in no translocation.

V-DNA, at the beginning of the simulation (40 ps). Despite the higher voltage in the 8.4 V-simulation, the DNA has covered a much smaller distance than in the 7.2 V-simulation after 220 ps. The 7.2 V-DNA translocates in a fairly straight conformation, and starts exiting the channel at 275 ps. On the other hand, the 8.4 V-DNA traverses the whole channel in a folded manner, which results in a much longer translocation time. It starts exiting the channel after 500 ps (Fig. 51), approximately twice as slow as the 7.2 V-DNA. It can be seen that although the 8.4 V-DNA possesses a less linear conformation at the channel entrance (Fig. 49, 40 ps) than the 6 V-DNA (Fig. 48, 40 ps), the large voltage forces it through anyhow. When considering that a voltage of 8.4 V was needed for translocation of a completely folded DNA, it should be emphasized that such a translocation is unlikely to occur in real-life experiments, where voltages several orders of magnitudes smaller are applied (See also section 7.1). Anyhow, all the results indicate that the translocation time should be considered as an interplay between base type (size), initial conformation and voltage magnitude.

Figure 49: Snapshots showing the translocation of 24G-DNA at an applied voltage of 7.2 V (left), and 8.4 V (right), with connexon constraints
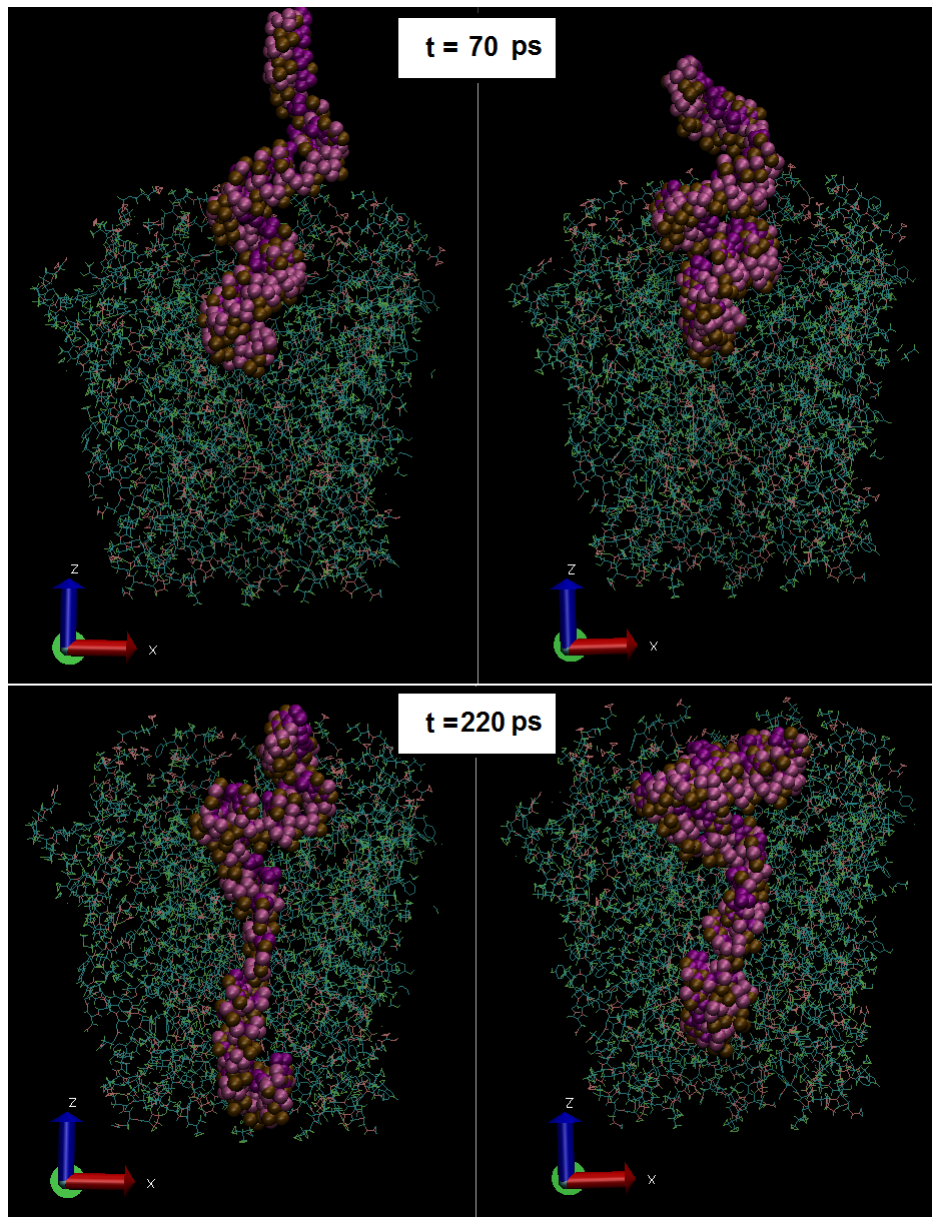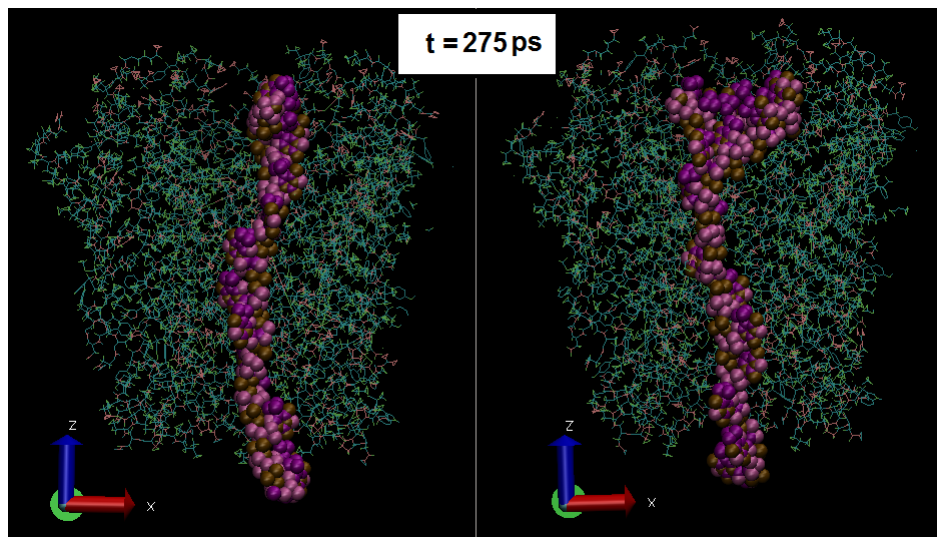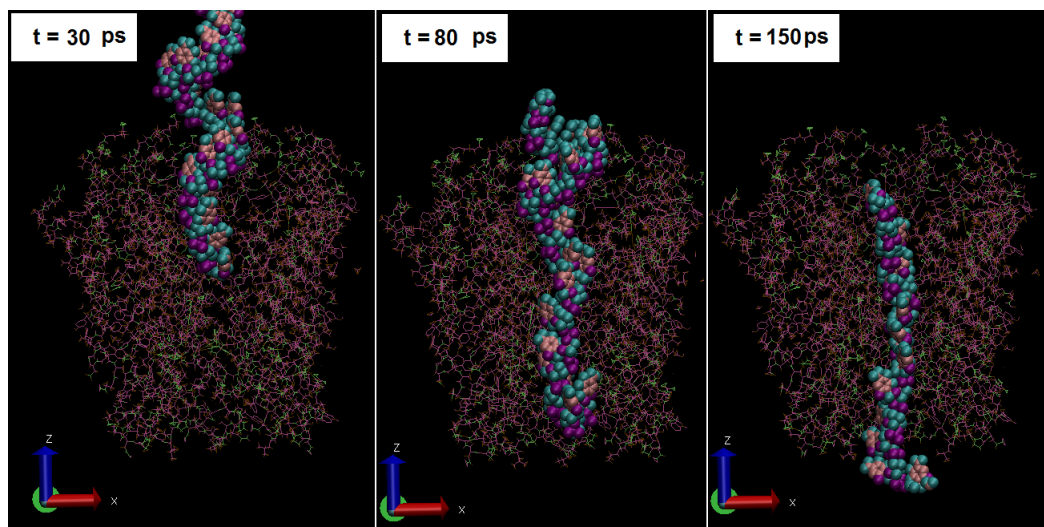
Figure 50: Snapshots showing the translocation of 24G-DNA at an applied voltage of 7.2 V (left), and 8.4 V (right), with connexon constraints.

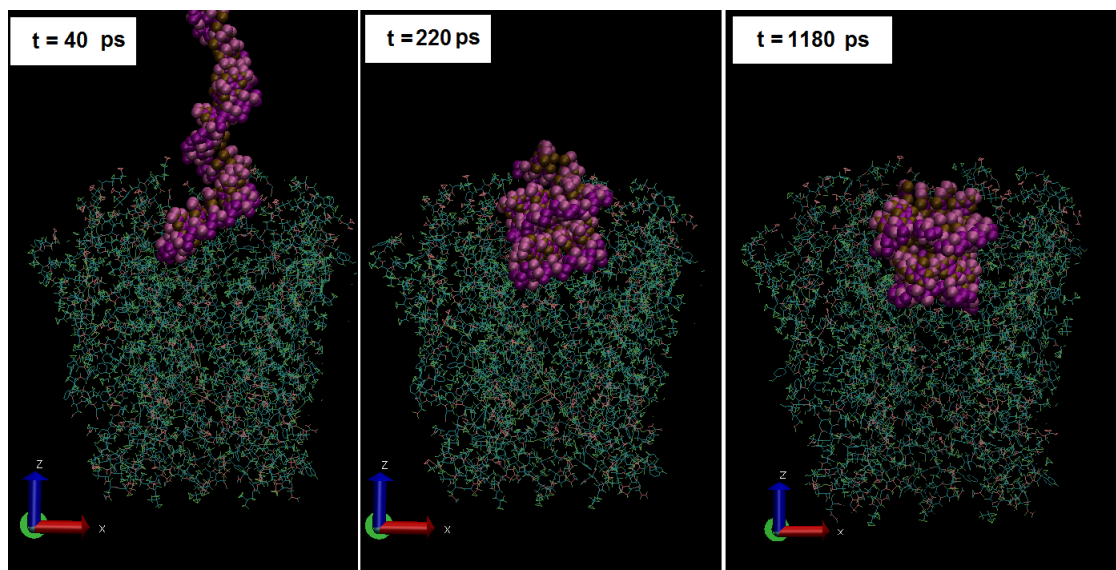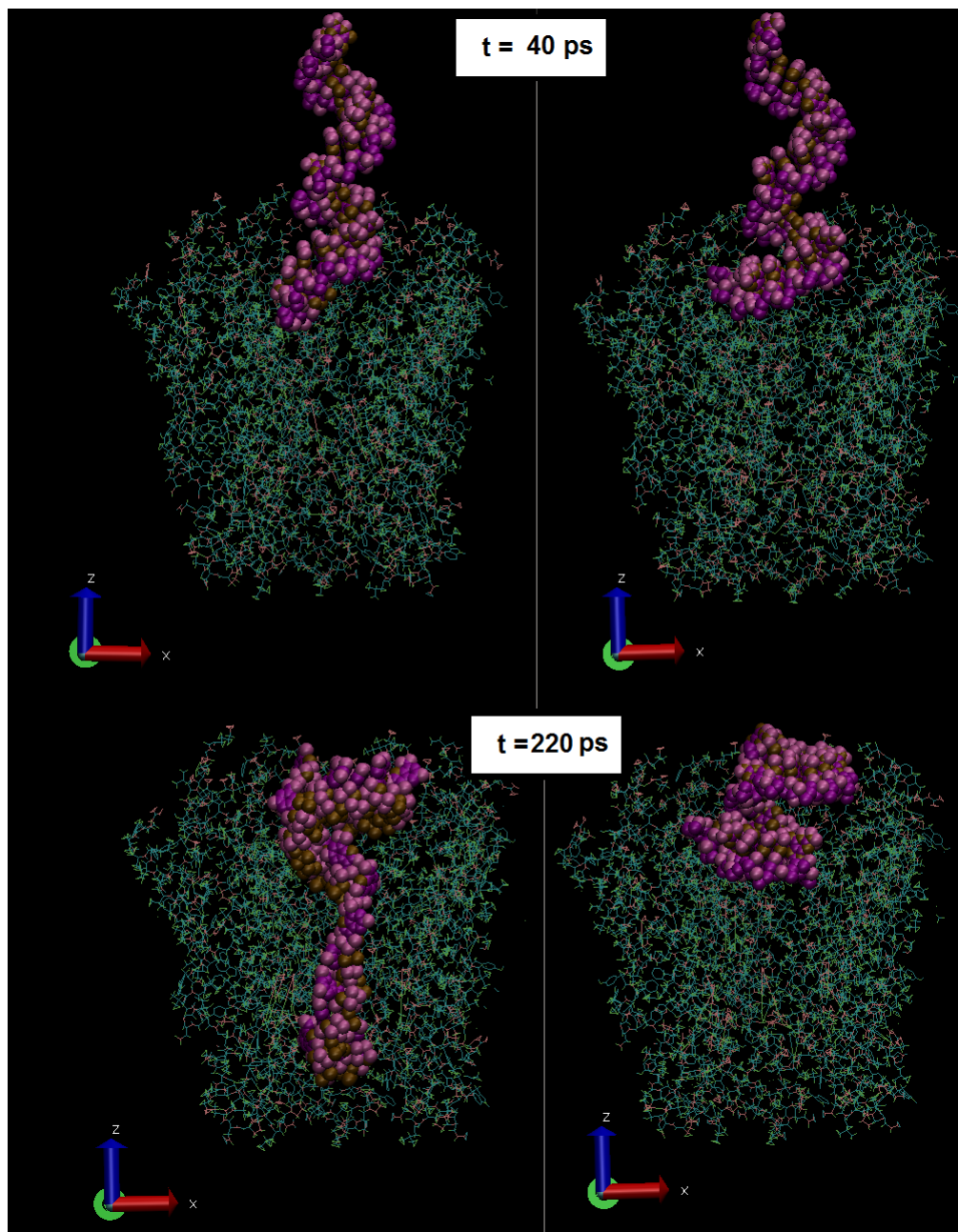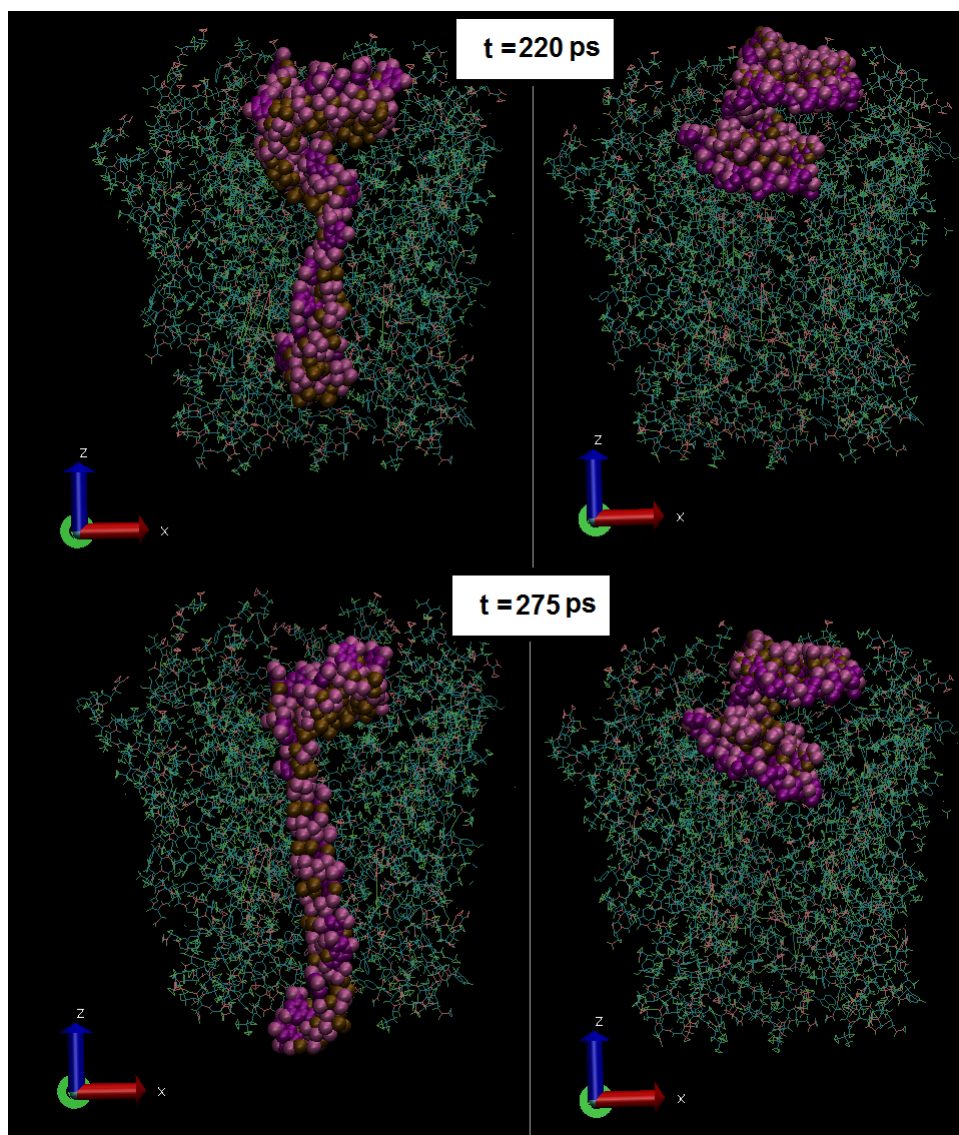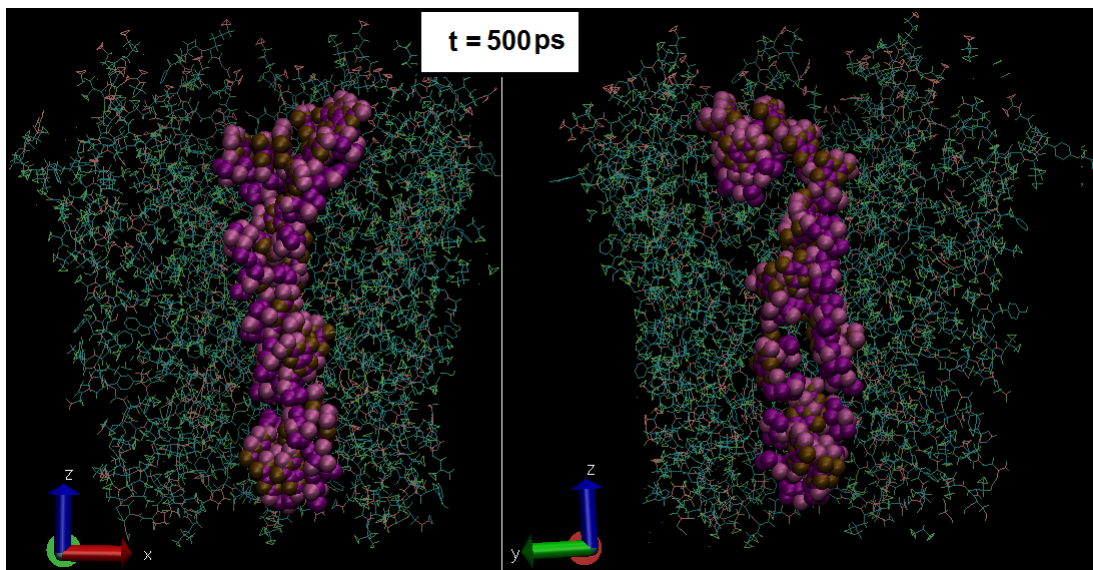Figure 51: Snapshots showing the translocation of 24G-DNA at an applied voltage of 8.4 V, with connexon constraints. The same DNA-molecule is shown from two different directions to point out the high degree of foldedness.

9 Comparison with published literature

Direct comparison with the recently discussed results with experimental results from the published literature is challenging, due to several reasons: To be able to observe DNA translocations within a temporal limit range of a couple of nanoseconds, the applied voltages in this study had to be at least a couple of volts. This is several orders of magnitude larger than the ones used in experimental studies, which are typically 120 mV for experiments with the α-hemolysin channel[13, 88, 90]. In addition, the translocation times observed in this study range from around 200 ps up to 1.8 ns, whereas the experimental translocation times reported are typically between microseconds and milliseconds[13, 88, 91]. This is a much larger timescale, which is not easily accessed by traditional MD simulations. Even though experimental studies are numerically different from the results of this study, some general trends and characteristics of the translocation events may still be compared. It is reasonable to compare the results with experiments carried out with the α-hemolysin channel, since its limiting inner diameter is 15 Å, close to the connexon limiting diameter of 14 Å.

9.0.4 Translocation time dependence on voltage magnitude

In a set of experimental studies by Meller et al.[92, 93, 94], the relationship between translocation velocity and voltage magnitude for DNA-translocation through an α-hemolysin channel is found to scale by $v \propto V^2$, for a DNA strand

Figure 52: Fitting of a power-law function to the curve from figure 33.

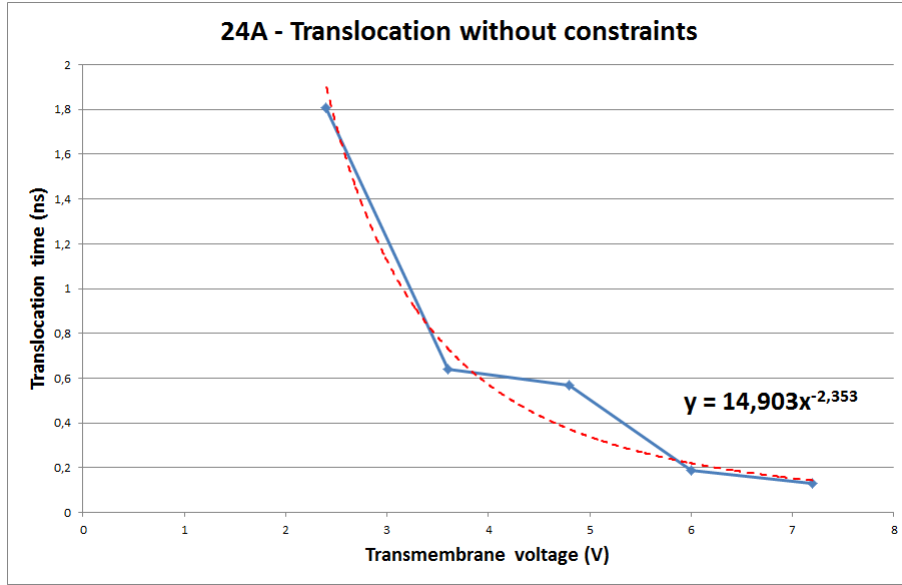containing only A-bases, and about the same length as the narrow channel stem. This relationship can also be expressed as $t \propto V^{-2}$, where t is the translocation time. In the discussion of translocation times of 24A-DNA with and without connexon constraints (Fig. 33 and 37), it was assumed that there might be a linear relationship between applied voltage and translocation time, and that conformational differences resulted in some data points deviating from the linearity. There is however also a possibility that there is no linear relationship, and that a power-law function may instead be fitted to the data. Despite the relatively few data points, power-law functions were fitted to the data, as can be seen in figure 52 and 53.

Although there is not a clear time-voltage relationship scaling as $t \propto V^{-2}$, a trend of the translocation time decaying as a power-law function may be seen in both figures. By extrapolating the fitted curve for translocations without connexon constraints to the typically used voltage of 120 mV for (Fig. 52), the following translocation time is found: $y = 14.903 \cdot (0.120)^{-2.353} \approx 2.19 \mu s$. Since the length of the connexon and the α-hemolysin channel are different, a direct comparison of translocation times should not be carried out. Instead, a comparison of translocation velocities can be made. The DNA travels about the whole length of the connexon, which is 77 Å, and the resulting translocation velocity therefore becomes $\overline{v} = \frac{77}{2.19} \mathring{A}/\mu s \approx 35.16 \mathring{A}/\mu s$.

By doing the same for the fitted curve from translocations with connexon constraints (Fig. 53), the following translocation velocity can be found: $y = 323.95 \cdot (0.120)^{-3.272} ns \approx 333 \mu s \Longrightarrow \overline{v} = \frac{77}{333} \mathring{A}/\mu s \approx 0.231 \mathring{A}/\mu s$. For translocation of 100A-DNA at temperatures about 40 degrees Celsius (which is comparable
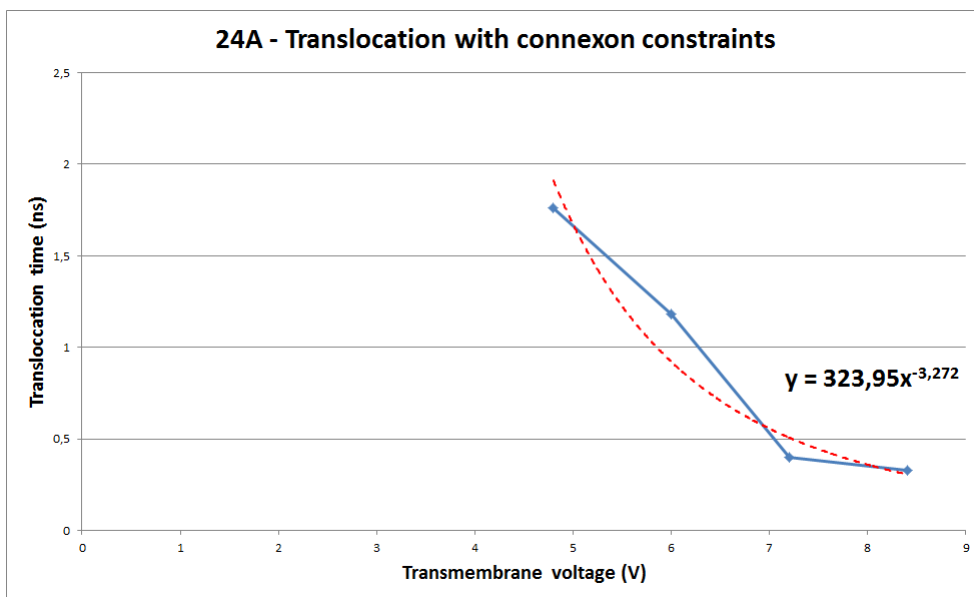
Figure 53: Fitting of a power-law function to the curve from figure 37.

with $310K \approx 37$ degrees Celsius, the temperature used in this report), Meller et al. showed a translocation time of about $\sim 60\mu s$. When considering an α-hemolysin stem length of 50 Å, this results in a translocation velocity of $\overline{v} = \frac{50}{60} \mathring{A}/\mu s \approx 0.83 \mathring{A}/\mu s$. This is on the same order of magnitude as $0.231 \mathring{A}/\mu s$, but in disagreement with $35.16 \mathring{A}/\mu s$. It thereby supports the assumption that placing constraints on the connexon channel results in more realistic simulations.

9.0.5 Translocation time dependence on base type

Meller et al. experimentally pointed out the translocation time differences of DNA containing distinct base types. They also found that the translocation time decreased with increasing temperature for DNA-sequences that were 100 bases long, for transmembrane voltages of 120 mV. The ratio between the translocation time of 100A-DNA and the translocation time of 100C-DNA was seen to to be inversely proportional with temperature.. This ratio decreased from ~3.2 at 15 degrees Celsius to ~2.1 at 40 degrees Celsius. The ratio at 40 degrees Celsius is fairly comparable with the results seen in table 3, when considering 24A and 24C at voltages of 7.2 V and 8.4 V. The ratios between the translocation times for 24A and 24C at these voltages are 2.5 and 2.75, respectively, and all simulations were carried out at about $310K \approx 37$ degrees Celsius. The ratio between translocation times of 24A and 24C at 6 V however, differs greatly from this relationship. The deviation arises because no power-law relationship can be seen for 24C (Fig. 44) as was indicated for 24A (Fig. 53). Further simulations for 24C should be carried out, to check whether the seemingly linear relationship

also holds for lower voltages.

In section 8.3.1, it is stated that G seems to be the base opposing translocation the most, followed by A, T and C. In the study by Meller et al., translocations of DNA containing only T or G is not reported. However, the similarity of the differences between A- and C-translocations can clearly be seen, between Meller et al. and the results of this study. In section 8.3.1, it was hypothesized whether the different sizes of the bases (mainly between purines and pyrimidines) could be the source of translocation time differences. In Meller et al., the difference was mainly explained by a stronger attractive interaction between A-DNA and the α-hemolysin channel. MD-simulations of a simplificated pore system[95] support that the differences in translocation times for A-DNA and C-DNA is a result of base A having a stronger attractive interaction with the channel . Although this characteristic of base A versus base C cannot be directly transferred to the DNA-connexon channel, it still shows that in addition to base sizes, the differences in chemical properties should also be considered when trying to explain the different translocation times seen in table 3.

9.0.6 Simulation artifacts, limitations and challenges

In the present study, conformational differences were assumed to play an important role when considering differences in translocation times. The very large applied voltages were shown to induce "compressed" conformations of the DNA molecule (e.g. Fig. 50), which was not seen for more realistic voltages (around 120 mV). It is therefore valid to assume these conformations to be artifacts arising from the unrealistically high voltages. A more sophisticated method that has been used to simulate DNA translocations through α-hemolysin is the method of grid-steered molecular dynamics (G-SMD)[87, 96]. In G-SMD, the translocation of solutes can be scaled up by a factor N to increase translocation velocity, without having to strain the membrane or the protein channel. This is accomplished by mapping the electrostatic potential of the system under low voltages, and amplifying them for the translocating solutes only.

In addition, the "phantom pore method" may be applied to shape the DNA into a conformation that better fits the channel of translocation[96]. This involves compressing the DNA inside a gradually shrinking mathematical surface (the phantom pore), resulting in a straight DNA molecule that is better suited for translocation through the specific channel. To avoid the dominating effects of conformation differences seen in the simulations of this report, the phantom pore method should be considered for further studies.

Qualitatively, the results from Ref. [96] includes a stepwise translocation, as was also seen in several of the results in this report (Sec. 8.2.2). It is suggested that the permeation through the narrow constriction of the α-hemolysin channel takes place in pulses of 2-4 bases. By inspection of the simulations of this report, it can be verified that the stepwise motion is not as "fine-tuned" as it would be on a single-base level. In other words, single-base resolution is doubtfully accomplishable through exploitation of this mechanism alone.

Part V
Conclusion

Molecular dynamics simulations of DNA-translocation through a connexon 26 channel were accomplished, for single-stranded DNA with a length of 24 nucleotides and with a sequence containing only adenine, cytosine, guanine or thymine. The simulations were carried out by using the LAMMPS simulation tool and the CHARMM force field. The applied transmembrane voltages ranged between 40 mV and 8.4 V, although translocations were not observed for voltages below 2.4 V. The two main barriers found to inhibit translocation were the positively charged residues at the channel entrance, and the narrow funnel region inside the channel. The translocation time decreased with increasing voltages for most simulations, with the exception of DNA molecules that adopted non-beneficial conformations upon channel entrance. It was seen that a small difference in initial conformation could result in major differences later in the translocation process. Non-beneficial conformations during translocation were more frequently observed at higher voltages and for DNA containing only purines. Although purines are larger than pyrimidines, results from published literature suggests that different channel affinities should also be taken into consideration. When extrapolating the translocation times of 24A with and without connexon constraints, comparison with published literature of α-hemolysin translocation found that the results when connexon constraints were applied agreed to the same order of magnitude. Without connexon constraints, the results deviated to a much larger extent, so connexon constraints are probably a more realistic simulation approach. Further simulations at different voltages are needed to decide if the relationship between applied voltage and translocation time is decreasing as a linear or power-law function. A rapid stepwise translocation was observed after passage of the funnel region, but it was unlikely at the single-base level, according to visual inspection.

Since this is the first reported study of DNA-translocation through a connexon channel, further investigations are needed to determine the complex mechanisms involved in translocation. In addition to experiments, simulations without substantial constraints should be pursued, e.g. by the use of steered molecular dynamics. Anyhow, the connexon channel undoubtedly interacts differently with the four different DNA bases, and thereby holds the potential of becoming a sequencing device on equal terms with the α-hemolysin channel. This certainly qualifies for further investigations.

Part VI
Future work

Section 9.0.6 described the possibility of using G-SMD in combination with the phantom pore method as an improvement to the traditinoal MD-simulation used in this study. Apart from this, the traditional MD-simulation used here could also be repeated with several variations applied, including:

-Different voltages.

-Different DNA lengths.

-Translocation of double-stranded DNA (if possible)

-Ion concentrations of 1.0 M instead of 0.1 M, to obtain a stronger ionic current signal, regarding potential sequencing

-Different base type combinations, including heterogeneous DNA sequences.

-Different DNA orientations: Simulations both with the 5' and with the 3' end facing the channel entrance.

-Longer simulation durations, up to tens of nanoseconds[88]

-Reverse translocation: Translocation with an initial DNA displacement on the extracellular side of the connexon. Could be interesting in terms of avoiding DNA-attachment at the positive residues on the cytoplasmic side.

-Translocation of RNA.

Finally, experimental studies could be carried out, to determine if translocation events give rise to current drops in the ionic current, and additionally compare the results with obtained simulation results. It could also be questioned whether translocation is at all possible, since the PDB-structure of the connexon 26 structure excludes a Met 1-residue inside the channel (See section 2.5.1). Experiments could however also be performed with one of the other connexon types, e.g. the connexon43, which possesses a larger inner diameter than connexon 26.

References

[1] Kathiresan, S. and Srivastava, D. Genetics of human cardiovascular disease. *Cell,* 148(6):1242-57, 2012

[2] Burgner, D. and Hull, J. In defence of genetic association studies. *Lancet,* 356(9229):599, 2000

[3] Hirschhorn, J.N. et al. A comprehensive review of genetic association studies. *Genet. Med.,* 4(2):45-61, 2002

[4] Brunicardi, F.C. Overview of the Development of Personalized Genomic Medicine and Surgery. *World J. Surg.,* 35(8):1693–1699, 2011

[5] Watson et al. *Molecular biology of the gene, 6th ed.,* Pearson Education, Inc., 2008

[6] International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature,* 431:931-345, 2004

[7] Hartwell, L.H. et al. *Genetics: From genes to genomes, 4th ed.,* McGraw-Hill Companies, Inc., 2008

[8] Metzker, M.L. Sequencing technologies - the next generation. *Nat. Rev. Genet.,* 11(1):31-46, 2010

[9] Timp et al. Third generation DNA sequencing with a nanopore. *In:* Iqbal, S.M. and Bashir, R. *Nanopores: Sensing and fundamental biological interactions.* Springer Science+Business Media, LLC 2011

[10] Wetterstrand, K.A. DNA Sequencing Costs: Data from the NHGRI Large-Scale Genome Sequencing Program available at: www.genome.gov/sequencingcosts. Accessed June 2012

[11] Kedes, L. and Campany, G. The new date, new format, new goals and new sponsor of the Archon Genomics X PRIZE Competition. *Nature Genetics,* 43(11):1055-58, 2011

[12] Bayley, H. Sequencing single molecules of DNA. *Curr. Opin. Chem. Biol.,* 10(6):628-37, 2006

[13] Kasianowicz, J.J. et al. Characterization of individual polynucleotide molecules using a membrane channel. *PNAS,* 93(24):13770-73, 1996

[14] Schaffer, A. Nanopore sequencing. *Technology Review,* 115(3):44, 2012

[15] Oxford Nanopore Technologies. Press Releases. http://www.nanoporetech.com/news/press-releases/view/39. Accessed June 2012

[16] DeBlois, R.W. and Bean, C.P. Counting and sizing of submicron particles by the resistive pulse technique. *Rev. Sci. Instrum* 41:909, 1970

[17] Storm, A.J. et al. Translocation of double-strand DNA through a silicon oxide nanopore. *Phys. Rev. E.,* 71(5):051903, 2005

[18] Chang, H. et al. DNA-mediated fluctuations in ionic current through silicon oxide nanopore channels. *Nano Letters,* 4(8):1551-56, 2004

[19] Fan, R. et al. DNA translocation in inorganic nanotubes. *Nano Letters,* 5(9):1633-37, 2005

[20] Fyta, M., Melchionna, S. and Succi, S. Translocation of biomolecules through solid-state nanopores: Theory meets experiments. *Journal of Polymer Science Part B: Polymer Physics,* 49(14):985-1011, 2011

[21] Wendell, D. et al. Translocation of double-stranded DNA through membrane-adapted phi29 motor protein nanopores. *Nature Nanotechnology,* 4:765-72, 2009

[22] Li, J. et al. Ion-beam sculpting at nanometre length scales. *Nature,* 412:166-69, 2001

[23] Garaj, S. et al. Graphene as a subnanometre trans-electrode membrane. *Nature,* 467:190-93, 2010

[24] Haitao, L. et al. Translocation of single-stranded DNA through single-walled carbon nanotubes. *Science,* 327:64, 2010

[25] Chen, Z. et al. DNA translocation through an array of kinked nanopores. *Nature Materials,* 9:667-75, 2010

[26] Krems, M. et al. Effect of noise on DNA sequencing via transverse electronic transport. *Biophysical Journal,* 97(7):1990-96, 2009

[27] Huang, S. Identifying single bases in a DNA oligomer with electron tunnelling. *Nature Nanotechnology,* 5:868-73, 2010

[28] Peng, H., Luan, B. and Stolovitzky, G. Nanopore-based DNA sequencing and DNA motion control. *In:* Iqbal, S.M. and Bashir, R. *Nanopores: Sensing and fundamental biological interactions.* Springer Science+Business Media, LLC 2011

[29] Dekker, C. Solid-state nanopores. *Nature Nanotechnology,* 2:209-15, 2007

[30] Meller, A. Dynamics of polynucleotide transport through nanometre-scale pores. *Journal of Physics: Condensed Matter,* 15:581-607. 2003

[31] Winters-Hilt, S. The α-hemolysin nanopore transduction detector - single-molecule binding studies and immunological screening of antibodies and aptamers. *BMC Bioinformatics,* 8:S9, 2007

[32] Becker, W.M. et al. *The world of the cell, 7th ed.,* Pearson Education, Inc., 2009

[33] Yeager, M. Gap junction channel structure. *In:* Harris, A. and Locke, D., *Connexins: A guide.* Springer Science+Business Media, LLC 2009

[34] http://www.uic.edu/classes/bios/bios100/f06pm/gap-junctions.jpg. Accessed June 2012

[35] Maeda, S., Tsukihara, T. Structure of the gap junction channel and its implications for its biological functions. 68(7):1115-29, 2011

[36] http://www.mun.ca/biology/desmid/brian/BIOL2060/BIOL2060-17/1711.jpg. Accessed June 2012

[37] Beyer, E.C. and Berthoud, V.M. The family of connexin genes. *In:* Harris, A. and Locke, D., *Connexins: A guide.* Springer Science+Business Media, LLC 2009

[38] Chandrasekhar, A. and Bera, A.K. Hemichannels: permeants and their effect on development, physiology and death. *Cell. Biochem. Funct.,* 30(2):89-100, 2012

[39] Thimm, J. et al. Calcium-dependent open/closed conformations and interfacial energy maps of reconstituted hemichannels. *J. Biol. Chem.,* 18(11):10646-54, 2005

[40] Wang, X. et al. pH-dependent channel gating in connexin26 hemichannels involves conformational changes in N-terminus. *Biochim. Biophys. Acta.,* 1818(5):1148-57, 2012

[41] Bargiello, T. and Brink, P. Voltage-gating mechanisms of connexin channels. *In:* Harris, A. and Locke, D., *Connexins: A guide.* Springer Science+Business Media, LLC 2009

[42] Maeda, S. et al. Structure of the connexin 26 gap junction channel at 3.5 Å resolution. *Nature,* 458:597-602, 2009

[43] Harris, A.L. and Locke, D. Permeability of connexin channels. *In:* Harris, A. and Locke, D., *Connexins: A guide.* Springer Science+Business Media, LLC 2009

[44] Yu, J. Aminosulfonate modulated pH-induced conformational changes in connexin26 hemichannels. *J. Biol. Chem.,* 282(12):8895-904, 2007

[45] Nakagawa, S., Maeda, S., Tsukihara, T. Structural and functional studies of gap junction channels. *Curr. Opin. Struct. Biol.,* 20(4):423-30, 2010

[46] Kwon, T. et al. Molecular dynamics simulations of the Cx26 hemichannel: Evaluation of structural models with Brownian dynamics. *J. Gen. Physiol.,* 138(5):475-93, 2011

[47] http://asingaporeanbiologist.files.wordpress.com/2011/10/nucleotide.jpg. Accessed June 2012

[48] http://www.dnareplication.info/images/dnadoublehelix.jpg. Accessed June 2012

[49] http://weloveteaching.com/2011/organics/base-pairs.jpg. Accessed June 2012

[50] http://www.uic.edu/classes/phys/phys461/phys450/ANJUM04/DNA_bbone.jpg. Accessed June 2012

[51] http://ghr.nlm.nih.gov/handbook/illustrations/dnastructure.jpg. Accessed June 2012

[52] Yakovchuk, P., Protozanova, E. and Frank-Kamenetskii, M.D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.,* 34(2): 564–574, 2006

[53] http://bioweb.wku.edu/courses/biol495g/DNA1/Images/F04-06.JPG. Accessed June 2012

[54] http://courses.biology.utah.edu/horvath/biol.3525/1_DNA/Fig2/marty_1.jpg. Accessed June 2012

[55] Wang, J.C. Helical repeat of DNA in solution. *PNAS,* 76(1): 200–203, 1979

[56] Benham, C.J. and Mielke, S.P. DNA Mechanics. *Annu. Rev. Biomed. Eng.,* 7:21–53, 2003

[57] Wen, Q. and Janmey, P.A. Polymer physics of the cytoskeleton. *Current Opinion in Solid State and Materials Science,* 15(5):177-182, 2011

[58] Wiggins, P.A. and Nelson, P.C. Generalized theory of semiflexible polymers. *Physical Review E,* 73:031906, 2006

[59] Adamcik, J. et al. Understanding amyloid aggregation by statistical analysis of atomic force microscopy images, *Nature Nanotechnology,* 5:423-28, 2010

[60] Lu, Y., Weers, B., Stellwagen, N.C. DNA Persistence length revisited., *Biopolymers,* 61:261-275, 2002

[61] Tinland, B. Persistence length of single-stranded DNA. *Macromolecules,* 30(19):5763-5765, 1997

[62] Li, J. Basic molecular dynamics. *In:* Yip, S., *Handbook of materials modeling.* Springer, 2005

[63] Leach, A.R. *Molecular modelling: Principles and applications, 2nd ed.,* Pearson Education, Inc., 2001

[64] Rapaport, D.C. *The art of molecular dynamics simulation, 2nd ed.,* Cambridge University Press, 2004

[65] Schlick, T. *Molecular modeling and simulation: An interdisciplinary guide, 2nd ed.,* Springer Science+Business Media, LLC 2010

[66] MacKerell Jr., A.D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B,* 102(18):3586-3616, 1998

[67] MacKerell Jr., A.D, Feig, M. and Brooks III, C.L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.,* 25(11):1400-1415, 2004

[68] Foloppe, N. and MacKerell Jr., A.D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.,* 21(2):86-104, 2000

[69] MacKerell Jr., A.D and Banavali, N.K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.,* 21(2):105-120, 2000

[70] MacKerell Jr., A.D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.,* 25:1584-1604, 2004

[71] http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node23.html. Accessed June 2012

[72] http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node23.html. Accessed June 2012

[73] Hiemenz, P.C. and Rajagopalan, R. *Principles of colloid and surface chemistry, 3rd ed.,* CRC Press, 1997

[74] http://homepage.mac.com/swain/CMC/DDResources/mol_interactions/molecular_interactions_files/pa ljp.png. Accessed June 2012

[75] Deserno, M. and Holm, C. How to mesh up Ewald sums. I. A theoretical and numerical comparison of various particle mesh routines. *J. Chem. Phys.,* 109(18):7678, 1998

[76] Deserno, M. and Holm, C. How to mesh up Ewald sums. II. An accurate error estimate for the particle-particle-particle-mesh algorithm. *J. Chem. Phys.,* 109(18):7694, 1998

[77] Berman, H.M. et al. The Protein Data Bank. *Biological Crystallography,* 58(1):899-907, 2002

[78] The Protein Data Bank. http://www.rcsb.org. Accessed June 2012

[79] http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-unix-html/node21.html. Accessed June 2012

[80] Humphrey, W., Dalke, A. and Schulten, K. VMD - Visual Molecular Dynamics (http://www.ks.uiuc.edu/Research/vmd/). *J. Molec. Graphics*, 14:33-38, 1996

[81] Comer, J.R., Wells, D.B. and Aksimentiev, A. Modeling nanopores for sequencing DNA. *Mol. Biol.*, 749:317-58, 2011

[82] Dijk, M.v. and Bonvin, A.M.J.J. 3D-DART: A DNA structure modelling server. *Nucl. Acids Res.*, 37 (Web Server Issue):W235-W239

[83] Plimpton, S. Fast parallel algorithms for short-range molecular dynamics (http://lammps.sandia.gov). *J Comp Phys*, 117:1-19, 1995

[84] Ryckaert, J.P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Comp Phys*, 23(3):327-341, 1995

[85] Hestenes, M.R. and Stiefel, E. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409-436, 1952

[86] Berendsen, H.J.C et al. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684,1984

[87] Aksimentiev, A. Deciphering ionic current signatures of DNA transport through a nanopore. *Nanoscale,* 2:468-483, 2010

[88] Mathé, J. et al. Orientation discrimination of single-stranded DNA inside the α-hemolysin membrane channel. *PNAS,* 102(35):12377-82, 2005

[89] Muzard, J. DNA translocation and unzipping through a nanopore: Some geometrical effects. *Biophysical Journal,* 98:2170-78, 2010

[90] Purnell, R.F., Mehta, K.K. and Schmidt, J.J. Nucleotide identification and orientation discrimination of DNA homopolymers immobilized in a protein nanopore. *Nano letters,* 8(9):3029-34, 2008

[91] Hall, A.R. et al. Hybrid pore formation by directed insertion of α-hemolysin into solid-state nanopores. *Nature Nanotechnology,* 5:874-77, 2010

[92] Meller, A., Nivon, L. and Branton, D. Voltage-driven DNA translocations through a nanopore. *Physical Review Letters,* 86(15):3435-38, 2001

[93] Meller, A. and Branton, D. Single molecule measurements of DNA transport through a nanopore. *Electrophoresis,* 23:2583-91, 2002

[94] Meller, A. et al. Rapid nanopore discrimination between single polynucleotide molecules. *PNAS,* 97(3):1079-84, 2000

[95] Luo, K. Sequence dependence of DNA translocation through a nanopore. *Physical Review Letters,* 100(5):058101, 2008

[96] Wells, D.B., Abramkina, V. and Aksimentiev, A. Exploring transmembrane transport through α-hemolysin with grid-steered molecular dynamics. *The Journal of Chemical Physics,* 127(12):125101, 2007