Diploma Thesis

Simulation von Nucleoporin-Netzwerken

Andreas Russek*

March 7, 2011

Max Planck Institute for Biophysical Chemistry†
Department of Theoretical and Computational Biophysics
Supervisor: Helmut Grubmüller
-
Georg-August-University Göttingen
Faculty of Physics‡
Supervisor: Annette Zippelius

*arussek@gwdg.de
†Am Fassberg 11, 37077 Göttingen
‡Freidrich-Hund-Platz 1, 37077 Göttingen

Contents

1 Introduction

Every eucaryotic cell contains a nucleus. It is enclosed by a double membrane, the nuclear envelope which protects the genetic information stored in the DNA inside the nucleus. The DNA itself contains the construction plans of all components needed by a cell, such as proteins or RNA. The production of proteins takes place in ribosomes, which are not located in the nucleus. Therefore, the RNA has to be transported through the nuclear envelope. This is done through the *nuclear pore complex* (NPC), which is the only type of pore in the nuclear envelope. It is responsible for transport from the nucleus to the cytoplasm and vice versa. To protect the DNA from harm, this gating has to be highly selective. Therefore, every molecule that needs to be transported into the nuclues carries a nuclear localization signal (NLS), whereas molecules that need to be exported have a nuclear export signal (NES), as can be seen in Figure 1.1. These signals are special protein sequences that can be recognized by transport receptors. These receptors can then distinguish between inert proteins and proteins that are intended to pass through the nuclear pore.
To transport a molecule with a transport signal through the pore, a transport receptor binds to that molecule and carries it to the opposite side of the NPC. There, the dissociation of the cargo-transporter complex is triggered by binding of RanGTP on the nucleoplasm side and by the hydrolysis of RanGTP to RanGDP on the cytoplasmic side. RanGDP (Ran is a small GTPase protein) is then again trasported from the cytoplasm to the nucleoplasm by the RanGDP importer NTF2, where it is converted to RanGTP (Miao and Schulten, 2009).

The NPC has a diameter of approximately 120 nm and is composed of about 30 different proteins, the so called nucleoporins (nups). It can be seen in Figure 1.2. The structure of its cytoplasmic part differs from the one reaching into the nucleus. Where on one side there are eight long filaments reaching into the cytoplasm, on the other side there is a so-called basket. The pore has an hourglass shape with a diameter between 45 and 70 nm.

The central pore contains long filaments of nups, which consist of repeating phenylanaline glycine sequences connected by polar linkers of varying amino acid composition (called FG-repeats). The most abundant FG-motifs are FG, FxFG (where x stands for any amino acid) and GLFG. How these FG-repeats are arranged in the nuclear pore is not yet clear as they are very flexible and largely disordered and, therefore, their configuration is difficult to determine experimentally. So, at present, there are several main

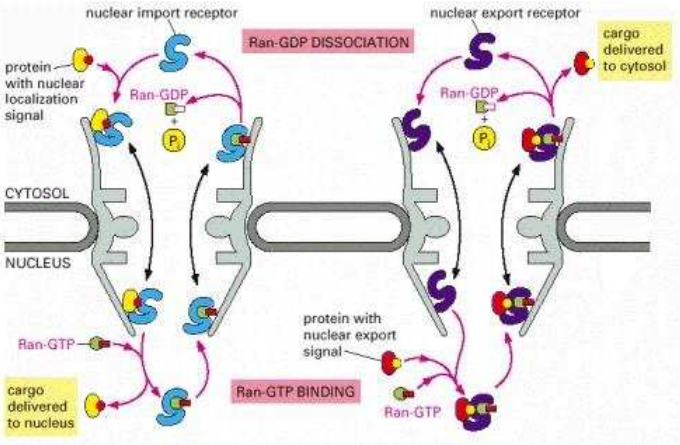

Figure 1.1: Diagram of the nuclear transport taken from Molecular Biology of the Cell, Alberts

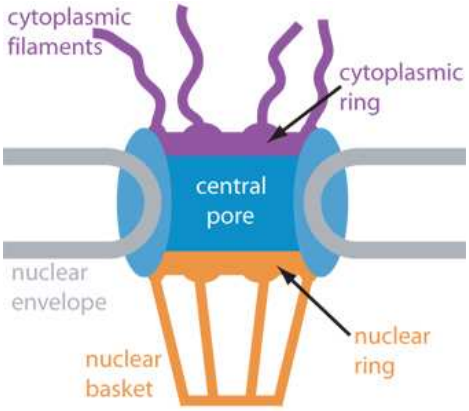

Figure 1.2: Diagram of the nuclear pore complex adapted from (Peters, 2005)

models of the structure inside the NPC.
Frey and Görlich (Frey et al., 2006) propose that the FG-repeats can form a saturated hydrogel in vitro. The meshwork is formed due to interactions between the hydrophobic FG-motifs, which tend to aggregate and form the joint connections of the FG-FG network. Newly gained results lead them to the assumption that even the linkers might be involved in the formation of the network (Ader et al., 2010). They generated a gel in which nuclear transport receptors diffuse faster than a control protein. A model cargo that was bound to special nuclear transport receptors (NTR), was even transported 20,000 times faster than the free one. The diffusion of smaller particles is regulated by the size of the FG mesh. They assume that the transporter protein can dissolve the FG-FG network of the sieve-like meshwork in the gel. This would allow larger cargoes to pass through the pore when they are bound to the transporter. The authors propose that such a gel will also be formed in vivo, in the NPC.
However, the FG concentration in such a gel has to be high and not every FG-hydrogel showed such a selective behaviour. Additionally, the formation of this gel requires ex-

treme pH and salt concentrations. It is, therefore, not clear to what extent it reflects the situation in the nuclear pore.
In contrast to the "selective phase model" mentioned above, the "virtual gating" model of M. Rout et al. (Rout et al., 2003) regards the NPC as a catalyst which lowers the activation energy for passage. The entropic barrier for large molecules - those with a mass larger than 30 to 40 kDa - to pass through the narrow pore is usually very high, as the centre is already crowded with FG-repeats. In contrast to the model of Frey and Görlich, the FG-repeats are regarded as mostly non-interacting. This is proposed by R. Lim and U. Aebi, who performed atomic force microscopy (AFM) experiments on nucleoporins (Lim et al., 2006). For transporter proteins the entropic barrier is lower, due to the fact that they are able to bind to the FG-repeats, thus pulling their bound cargo through the pore.
The third idea was published by R. Peters (Peters, 2005) and is based on a reduction of dimensionality for diffusion through the pore. The model suggests that the FG-repeats cover the filaments and the central channel with a coherent FG surface. The FG surface on the cytoplasmic filaments serves as an antenna for transporter proteins to attract them and bring them closer to the entrance of the pore channel.
Inside the channel, the FG surface is used to guide transporters through the pore, as they search the surface by a two-dimensional random walk. To create a selective filter, long hydrophilic unfolded peptide chains reach into the pore centre. These chains can be pushed aside by large transport complexes. The selectivity filter will reject neutral molecules because they are not bound to the FG-surface.
Last but not least, the "forest model", which was introduced by J. Yamada et al. (Yamada et al., 2010) combines the "selective phase" with the "entropic barrier" model. They observed that the FG-repeats are mainly in two different conformations. One is a globular, collapsed coil conformation, which they called "shrubs" and the other is an extended-coil conformation, which is refered to as "trees". The "trees" consist of an uncoiled part reaching from the scaffold of the pore into the central channel, where it ends up in a coiled conformation that may be able to aggregate, forming a meshwork with hydrogel-like properties (Wälde and Kehlenbach, 2010). The "shrubs", however, are already coiled close to the scaffold and therefore do not reach far into the central channel.
Both conformations can be found through the entire pore, generating two zones for transportation. The first is created by the "trees" as they reach far into the pore, leaving an open pathway in the middle. Transport through the first zone functions as in the "selective phase" model described above and is mostly used by large cargo.
The second zone contains the "shrubs" and the uncoiled parts of the "trees". Empty NTRs or NTRs with just small cargoes may travel through this zone as it works like an entropic barrier to larger molecules. Regarding the models above, it is not clear which parameters influence the structure of the selectivity barrier most. It is only known that the compostion of the FG-repeats plays a role, as well as their length and their density inside the channel. Even the exact interaction with the transporter receptors remains unclear. The problem for researchers is that the FG-repeats are highly flexible and even the most structured model by Görlich and Frey is still disorderd, so that the config-

uration of the selectivity barrier inside the central channel still cannot be resolved by microscoping and scattering experiments.
For atomistic Molecular Dynamic (MD) simulations the NPC is too large to be simulated as a whole. Additionally, the protein folding lies within timescales of miliseconds, which is much too long for usual MD simulations.
Therefore, we aim at developing a simple coarse grained (CG) model to carry out Monte Carlo simulations and study the effect of the different factors mentioned above. With this model we want to examine the dependency of the selectivity barrier from the cohesion of different parts of the FG-nups. This may help to solve the question of whether the nups tend to collapse and coil individually or whether they prefer to build a meshwork. We further want to observe whether the length of the FG-nups influences the assembly of the barrier.
It is also interesting to know whether there is a difference between the behaviour of freely moving FG-repeats, as is usually the case in the in vitro experiments of Görlich and Frey, and FG-repeats that are anchored in the scaffold of the pore. Fixing one end of the FG-repeats to the scaffold of the pore will lead to a decrease in the degrees of freedom of the nups. Therefore, a possible gel formation could be hindered.
The experiments of Görlich and Frey lead to another interesting point, as for them it was not possible to generate a gel-phase with all the FG-repeats they used. Hence, the structure of the selectivity barrier could be influenced by the amino acid sequence of the FG-nups. With our Monte Carlo method, we will try to determine the extend to which the distance between two FG sequences will influence the aggregation of the FG-repeats. In a first step, the model is parameterized to reproduce results of all-atom (aa) Molecular Dynamic simulations of smaller FG-repeat model systems. The results of such a comparison are used to define a set of parameter, which will later describe the properties of longer FG-repeats. Therefore, we focussed on persistence lengths of single FG-repeats and clustering properties of small ones.
In a second step, when all the parameters are defined, we will examine which interactions will have which effect on the network by changing single parameters.

2 Methods

2.1 Monte Carlo Simulation

The term "Monte Carlo method" is very general, and the different methods only have in common that they use stochastic techniques to solve a problem.
The way a Monte Carlo simulation is performed depends strongly on the system of interest. It can be used to evaluate definite integrals, especially multi-dimensional ones with complicated boundaries. It is also suitable for studying systems with a large number of coupled degrees of freedom and many more.
In biophysics highdimensional integrals, where the calculation cannot be done analytically and even the numerical solution exceeds computational feasibility, have to be solved. The partition sum of a system, for example, is such an integral. It is given by

$$Z = c \int \mathrm{d}p^N \mathrm{d}q^N \exp\left[-\beta H(q^N, p^N)\right], \tag{2.1}$$

where q^N stand for the coordinates of the N particles and p^N for the momenta. $H(q^N, p^N)$ is the Hamiltonian of the system. It is usually needed to calculate the average of an observable O

$$\langle O \rangle = \frac{\int \mathrm{d}q^N \exp\left[-\beta U(q^N)\right] O(q^N)}{\int \mathrm{d}q^N \exp\left[-\beta U(q^N)\right]}, \tag{2.2}$$

where $U(q^N)$ is the potential energy. These high dimensional integrals can only be solved analytically for a few systems. In general they have to be treated numerically.

2.1.1 Metropolis Algorithm

The Metropolis algorithm provides a way of calculating the average of an observable without the need to calculate the partition sum. Therefore it is neccessary to generate random points in configurational space according to the probability distribution given by

$$\rho = \frac{\exp\left[-\beta U(q^N)\right]}{Z}, \tag{2.3}$$

where $Z = \int \mathrm{d}q^N \exp\left[-\beta U(q^N)\right]$. It can easily be seen that ρ is nonnegative. The mean value of the observable can then be written as

$$\langle O \rangle = \frac{1}{L} \sum_{i=1}^{L} n_i O(r_i^N). \tag{2.4}$$

L is the total number of generated points n_i. To sample the region of interest in our phase space a starting conformation r^N, further declared as o, with a non-vanishing Boltzmann factor $\exp\left[-\beta U(o)\right]$ is generated. Then a new configuration n is generated by adding a small random displacement Δ to o. Clearly the Boltzmann factor of the new state is $\exp\left[-\beta U(n)\right]$. To decide whether the new state is accepted or rejected, the Metropolis algorithm is used, as it is simple and generally applicable.

Detailed Balance and Acceptance Criteria

The acceptance criteria of a metropolis algorithm are directly connected to a principle called "detailed balance". In equilibrium, the probability of being in state o and going to state n must be the same as the probability of being in state n and going to state o

$$\rho(o)\pi(o \to n) = \rho(n)\pi(n \to o) \tag{2.5}$$

ρ refers to the equilibrium probability to be in state o, j and π describes the transition probability to go from one state into another. In a Monte Carlo simulation the transition probability is split into two parts. The first part describes the probability to perform the trial move from o to n. This probability is denoted as $\alpha(o \to n)$ below. The second part is the probability $\mathrm{acc}(o \to n)$ to accept the trail move from state o to state n. Now the transition probability π can be rewritten as follows:

$$\pi(o \to n) = \alpha(o \to n) \times \mathrm{acc}(o \to n). \tag{2.6}$$

If we now choose α to be a symmetrical probability matrix, equation 2.5 can be rewritten as

$$\rho(o) \times \mathrm{acc}(o \to n) \;=\; \rho(n) \times \mathrm{acc}(n \to o), \tag{2.7}$$

which will lead to

$$\frac{\mathrm{acc}(o \to n)}{\mathrm{acc}(n \to o)} \;=\; \frac{\rho(n)}{\rho(o)} \;=\; \exp\left\{-\beta\left[U(n) - U(o)\right]\right\} \tag{2.8}$$

Still, there are many choices left for the acceptance probability that will satify this condition and, of course, the one that it must not be larger than 1. The choice that Metropolis et al. used and which I will also use as acceptance criterion is

$$\begin{aligned} \mathrm{acc}(o \to n) = \;& \frac{\rho(n)}{\rho(o)} && \text{for } \rho(n) < \rho(o) \\ =\; & 1 && \text{for } \rho(n) \geq \rho(o), \end{aligned} \tag{2.9}$$

because the acceptance probability turns out to be rather simple

$$\mathrm{acc}(o \to n) \;=\; \exp\left\{-\beta\left[U(n) - U(o)\right]\right\} \qquad \text{for } U(n) > U(o) \tag{2.10}$$

and the partition sum is also no longer needed. If the potential energy of the new state is equal to or even lower than that of the old state, it will be accepted right away.

2.1.2 MC Move Methods

In order to reduce the size of the conformational space that needs to be sampled, we decided to build a simplified coarse grained model of the nucleoporins. Therefore, we replaced the single atoms of an amino acid by three beads on a straight line. The first and the last bead deal as junctions to the previous and the following segments, respectively. In nature these beads would be best described by the peptide bond between the carbon C of the carboxyl groub and the nitrogen N of the following amino acid. The distance between them is kept fixed at length l_{seg}. The bondangle, however, is flexible, as it is only restricted by harmonic potential.

The second bead is a property bead used for the calculation of the energies, as it contains the information about the Lennard-Jones and the Coulomb interaction. In our model there are 4 different types of property beads: neutral, polar, charged and hydrophobic. These four are chosen because amino acids of these types can mainly be

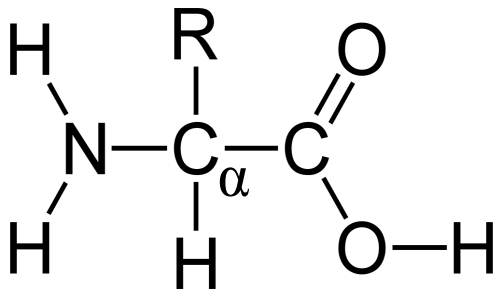

Figure 2.1: A general composition of an amino acid. The R stands for a specific side chain. In the MC simulation a the backbone N-C_α-C is represnted by the line and the information of the side chain is written on the middle bead.

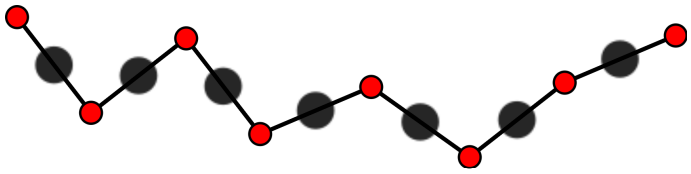

Figure 2.2: Simple model of an amino acid chain used in the MC simulation. The dark grey circles represent the property beads containing the information of the underlying AA. The red circles represent the junctions.

found in FG-repeats. In nature the property of an amino acid is given by its specific side chain, which is bonded to the C_α. Thus, the property bead can be understood as the coarse grained sidechain. A diagram of what a nup would look like in the MC simulation can be seen in Figure 2.2.

For MC simulations of nonbonded systems, the MC move can easily be performed by just randomly displacing particles. However, in bonded systems, the displacement has to be more restrained, as the bondlength has to stay fixed. Therefore, the problem arises of how to obtain meaningful conformations with reasonable acceptance ratios, while still sampling enough of the conformational space.
To deal with this problem, we had two main ideas for performing a Monte Carlo move. The first idea was to rebuild the chain, starting from a randomly chosen number, and the second was a so-called fixed end method.

Chain Rebuilding

This method is based on rebuilding the chain starting from a randomly chosen startpoint. From here on, all the following segments are deleted and a set of k random vectors b_i is generated. Between all these vectors and the previous segment the anglepotential is calculated. To calculate the acceptance probability of a new segment, the potential energy has to be split into an orientantional part U_{or} and into a part that only depends

on the position of the atoms U_{pos}. A new segment is then accepted with the probability of

$$p(b_n) = \frac{\exp\left[-\beta U_{or}(b_n)\right]}{\sum_{i=1}^{k} \exp\left[-\beta U_{or}(b_i)\right]} \tag{2.11}$$

where U_{or} is the energy that depends on orientations and the sum $W(n) = \sum_{i=1}^{k} \exp\left[-\beta U_{or}(b_i)\right]$ denotes the so-called Rosenbluth factor.
This is done until the whole chain is rebuilt. The acceptance criteria has changed to

$$acc(o \to n) = \min\left(1, \frac{W(o)}{W(n)} \exp\left\{-\beta\left[U_{pos}(o) - U_{pos}(n)\right]\right\}\right) \tag{2.12}$$

with

$$W(o) = \exp\left[-\beta U_{or}(b_o)\right] + \sum_{j=2}^{k} \exp\left[-\beta U_{or}(b_j)\right]. \tag{2.13}$$

We found that this method works well for single chains, but leads to high rejection rates when multiple chains are involved. This arises because the step size for a move cannot be defined. Additionally, as the whole chain is rebuilt, even small changes can lead to very large changes in the overall conformation, which again can cause clashes with other chains.
However, a positive aspect of this method is that it has low correlation times, as the conformational changes from the old state to the new one are usually very large.

Fixed End Method

The Monte Carlo steps in the fixed end method are performed as follows. First, a random number n_{seg} between 1 and N_{seg}, which is the total amount of segments of the chain, is generated. n_{seg} refers to the first bead of the segment which will be changed. There are now three possible cases.

- The first possibility is that the randomly chosen segment lies between $1 \leq n_{seg} \leq N_{seg} - 2$ as shown in Figure 2.5. In this situation, three segments will be displaced. The endpoint j_2 of segment s_1 is moved according to a gaussian distribution. After displacement, it is projected back on a sphere around its first junction j_1. The sphere has the radius of a segment length. After that, a vector t is generated from the endpoint of the newly created segment $s_{1,new}$ to junction j_4. The length of $\vec{t}$

may not exceed the length of two segments. Otherwise, the displacement of j_2 has to be repeated until a connection $\vec{t}$ between the two mentioned junctions can be made. After that, the new segments $s_{2,new}$ and $s_{3,new}$ are generated so that they span an isosceles triangle. The junction $j_{3,new}$ is put into a plane spanned by $\vec{t}$ and the old junction j_3. Even by not moving the junction out of the mentioned plane, the simulation is still ergodic. So it is not neccessary to introduce yet another random number that would move the junction out of the plane. The displacement described here can be seen in Figure 2.5 and 2.6.

- When the chosen segment is the penultimate segment, the method above cannot be applied. In this case, we move the endpoint of the chain. Again this point may not be moved further away from the first junction than twice the length of a segment. This is guaranteed by the connecting vector $\vec{t}$. The displacement is now carried out by randomly choosing the length of $\vec{t}$ between 0 and its maximal length (Figure 2.3). All lengths are equally distributed. After this, the new endpoint of $\vec{t_{new}}$ is displaced on a sphere with the radius of $|\vec{t_{new}}|$.
  Again, the newly generated segments $s_{1,new}$ and $s_{2,new}$ form an isosceles triangle and the junction between them is put into the plane spanned by the old junction and the connecting vector $\vec{t_{new}}$ (figure 2.4).

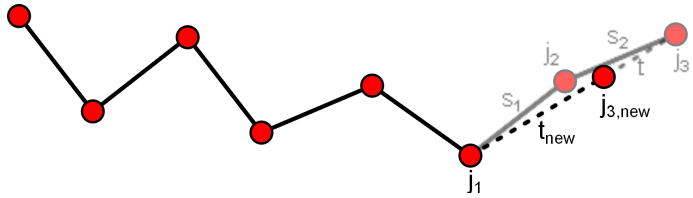

Figure 2.3: First part of the second possible fixed end move. The vector t is shortend to t_{new}.

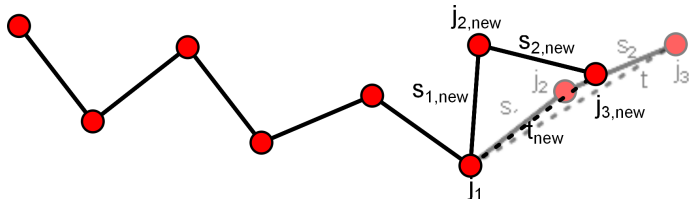

Figure 2.4: Second part of the second possible fixed end move. The the new endpoint $j_{3,new}$ is diplaced and the segments $s_{1,new}$ and $s_{2,new}$ are generated, so that they form an isosceles triangle.

- The third and also the last possibility is to move only the last segment. Here, the endpoint of the segment is moved on a sphere, as already described in the first option.

To determine the parameters, we introduced another random number. This number is either one or zero and dictates whether the MC move is done forwards or backwards.

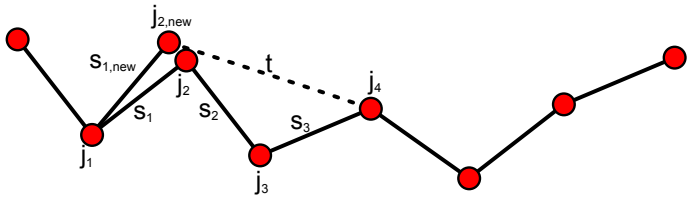

Figure 2.5: The displacement of the first segment s_1 is based on a gaussian distributed random number that is added to junction j_2. The length of the new segment $s_{1,new}$ is then reduced to the fixed segment length. The connecting vector t may not exceed a length of twice a segment length.

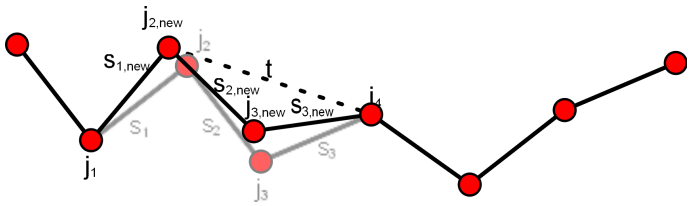

Figure 2.6: The second part of the segment displacement of the fixed end method.

Therefore, the enumeration of all junctions and segments is mirrored and both ends will be moved equally.
For a better description of the situation inside the nuclear pore, this random number is switched off, so that the first bead of each chain stays fixed. It can be understood as the junction between the nups and the nuclear pore itself.
If more than a single chain is simulated, yet another random number is generated right at the beginning of an MC move. This number n_{chain} then defines on which chain the MC move is carried out and it is essential that $1 \leq n_{chain} \leq N_{chain}$, where N_{chain} is the total number of chains in the simulation.

To finally perform an MC step I chose the fixed-end method, as the stepsize of a single displacement can easily be determined by width of the gaussian distributions. Additionally, as only three segments are displaced within a single MC step, the newly generated conformations are close to the old one. Therefore, in systems with more than one chain, clashes can be avoided. However, as the conformational changes are rather small, the correlation times are high.

After a move is completed, the energy E_{new} of the new state is calculated. The energy function includes the bondangle potential between all neighbouring segments and the Lennard-Jones potential for all beads that contain the information of the segment. For the Lennard-Jones potential we weakened the interaction between neighbouring particles by a factor of 4. Such a scaling is commonly applied in MD and MC calculations to avoid artificially high energies. Then the Boltzmann weight referring to the old state is calculated. The new state is accepted if an equally distributed random number between

0 and 1 is smaller than the Boltzmann factor. If the new state is accepted, it is stored and replaces the old state. Otherwise, a new MC move is carried out, starting from the old confirmation. The functions describing the interactions are explained in detail in section 2.2.3. To account for the effect of the solvent, the Coulomb interaction is extended by the Debye Hückel equation, which describes the weakening of the Coulomb interaction due to the surrounding material. The extended equation is given by

$$U_{DH,coul} = \exp(-\kappa r)\frac{q_1 q_2}{4\pi\epsilon_0\epsilon_r r}, \tag{2.14}$$

where q_i are the charges, ϵ_0 is the dielectric constant and ϵ_r is the relative permittivity, which is set to $\epsilon_r = 80$. In the exponential function, κ stands for the inverse Debye Hückel length. It is calculated via

$$\kappa = \sqrt{\frac{N_A e^2 \sum c_i z_i 2}{\epsilon_0\epsilon_r k_B T}}. \tag{2.15}$$

In this equation, the concentration of the salt ions is denoted as c_i with their charge number z_i.
The results of a Monte Carlo simulation strongly depend on random numbers. Therefore, the quality of the algorithm of the random number generator needs to be high. In our simulaiton, we used the Mersenne twister mt 19937 algorithm which has a long period. It was especially designed to meet the needs of Monte Carlo simulations.

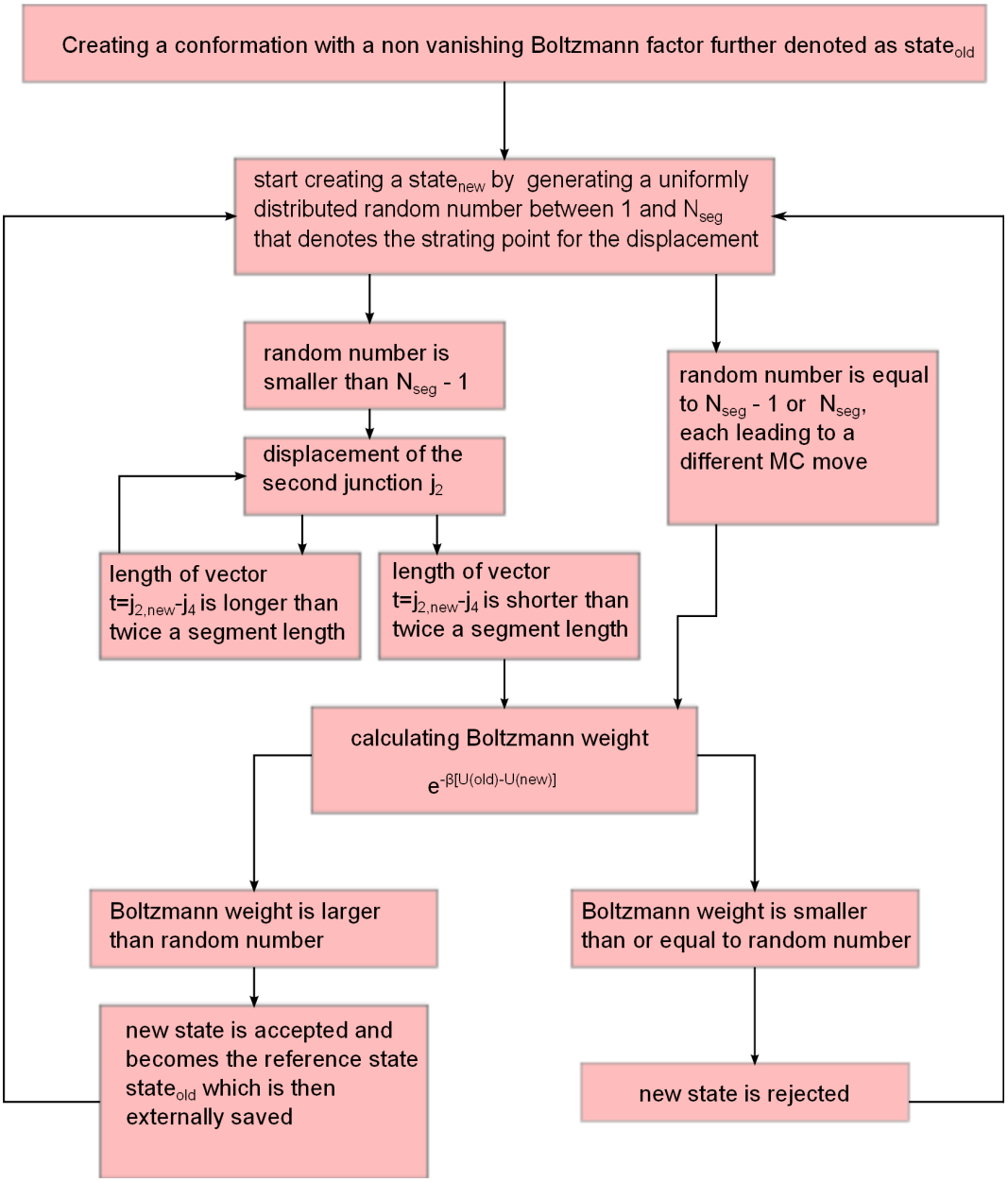

Figure 2.7: Flowchart of the Monte Carlo simulation. The second and the third case of an MC move are combined in a single box, as no displacement is possible, which will not lead to the calculation of the Boltzmann weight.

2.2 Molecular Dynamic Simulation

Molecular dynamics (MD) simulations describe molecular systems as they develop in time. Due to the large number of atoms in these systems, Newton's equations of motion are calculated numerically. Therefore, it is neccessary to apply three basic approximations. The first is called the Born Oppenheimer Aproximation, which allows a decoupling of nuclear and electronic motion. Secondly, it is assumed that the nuclei behave in a classical way, and thirdly, that all interactions between particles can be described in empirical force fields. The three approximations are detailed below.

2.2.1 Born-Oppenheimer Approximation

On an atomistic level, the evolution of a system in time is described by the time dependent Schrödinger equation.

$$\hat{H}\psi = i\hbar\frac{\partial\psi}{\partial t} \tag{2.16}$$

Here, $\hat{H}$ is the Hamiltonian operator, i.e the sum of the potential and the kinetic energy operators, ψ is the wavefunction and $\hbar$ is Planck's constant divided by 2π. The wavefunction ψ depends on the coordinates and momenta of the nuclei and electrons. The slowly moving nuclei are much heavier than the fast moving electrons. For such a situation, the Born Oppenheimer approximation allows us to separate the two systems. The general idea behind this is that the electrons will adjust instantaneously to the nuclei. Consequently, the electronic wavefunction depends only parametrically on the coordinates of the nuclei, so that the total wavefunction can be written as

$$\psi_{tot} = \psi_n(\mathbf{R})\psi_e(\mathbf{R};\mathbf{r}), \tag{2.17}$$

where $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, ..., \mathbf{R}_N)$ denotes the coordinates and momenta of N nuclei and $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_M)$ stands for the coordinates and momenta of M electrons. The electronic coordinates can now be approximated by their average values, due to their fast motion. The nuclei move in the potential which is created by the average field of the electrons. Using the Born Oppenheimer approximation, the so-called potential energy surface is obtained by solving the time independent Schrödinger equation for the electronic system

$$\hat{H}_e\psi_e = E_e\psi_e \tag{2.18}$$

where $\hat{H}_e$ is the electronic Hamiltonian and E_e is the energy eigenvalue, which depends on the nuclear postions $\mathbf{R}$.

2.2.2 Classical Approximation

As already mentioned above, we assume that the nuclei can be treated as classical particles, i.e. they obey Newton's equations of motion.

$$-\nabla_i V(\mathbf{R}) = m_i \frac{\mathrm{d}^2 \mathbf{R}_i(t)}{\mathrm{d}t^2}, \tag{2.19}$$

$$\mathbf{F}_i = m_i \mathbf{a}_i, \tag{2.20}$$

where $V(\mathbf{R})$ equals the potential energy and $\mathbf{R}_i$ and m_i are the coordinates and mass of the atom i. A change in the potential energy results in a force $\mathbf{F}_i$ acting on atom i, which is then accelerated by $\mathbf{a}_i$. The acceleration again leads to a change in the atom's velocity and position within a discrete time step Δt. During a simulation, this time step has now to be small enough to include even the fastest motions of the system. Usually, these are the bond and angle vibrations of the lightest atoms, in biological systems mostly hydrogens. This restriction causes time steps of about 1 fs.

2.2.3 Force Fields

In MD simulations, force fields contain all information about particle interactions. These can be split into two main groups, bonded and non-bonded interaction. Although there are already numerous quantum mechanical methods available, they are still too expensive to calculate them on a whole system. This is why the bonds are dealt with in a classical approximation.

Bonded Interactions

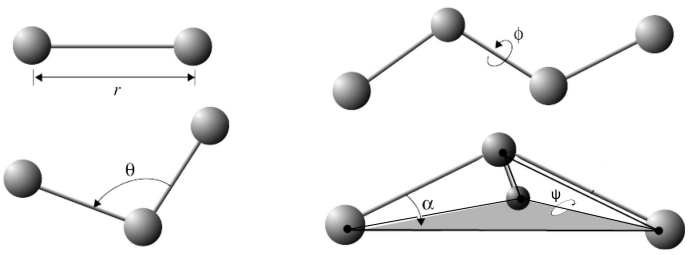

Figure 2.8: Graphical description of the potentials used in MD simulations

Instead of solving Schrödinger's equation to get the potential energy of each bond, they are approximated by different classical potentials.

- *Pair bonds*
  Covalent bonds are treated as springs with a certain spring constant $k_{b,ij}$ and a relaxed bond length $b_{0,ij}$.

$$V_b = \sum_{\text{bonds}} \frac{1}{2} k_{b,ij}(b_{ij} - b_{0,ij}) \tag{2.21}$$

  The spring constant defines how strong a bond between two atoms i,j is. Usually, this parameter is quite high, so that only an oscillation around its equilibrium length is tolerated. The amplitudes of such an oscillation are small at room temperature, where most simulations are carried out. Bonds that are defined cannot be broken during a simulation.

- *Bond angle potential*
  The potential energy of the bond angle is also approximated by a harmonic potential.

$$V_a = \sum_{\text{angles}} \frac{k_{\theta,ijk}}{2}(\theta_{ijk} - \theta_{0,ijk}) \tag{2.22}$$

  Here, the spring constant $k_{\theta,ijk}$ characterizes the stiffness of the bond angle θ_{ijk} between atoms i, j and k, as shown in Figure 2.8.

- *Dihedral potential*
  Another potential energy dealing with angles is the dihedral potential. Dihedral angles are shown in Figure 2.8. They describe the potential energy, which depends on the angle between plane A that is spanned by the atoms i,j and k and plane B spanned by atoms j,k and l. In contrast to the bond angle potential, the dihedral potential does not behave like a harmonic potential, but is commonly described by a cosine function.

$$V_{dih} = \sum_{\text{dihedrals}} \frac{V_n}{2}(1 + \cos[n\omega - \gamma]) \tag{2.23}$$

- *Improper Dihedrals*
  With the potentials described above, it is not possible to stabilise planar rings. Therefore, a harmonic potential for the angle between the planes spanned by three out of four consecutively bound atoms is defined. The potential for the so-called improper dihedrals is given by

$$V_{id} = \sum_{\text{improper dih.}} \frac{k_\phi}{2}(\phi_{ijkl} - \phi_0)^2 \tag{2.24}$$

Non-bonded Interactions

Two different types of non-bonded interactions are commonly used in MD simulations, i.e. the Lennard-Jones potential and the Coulomb interaction.

- *Lennard-Jones potential*
  The Lennard-Jones potential is defined in the following way

$$V_{LJ} = 4\epsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right) \tag{2.25}$$

  It is used to describe all neutral interaction between two particles. The $\frac{1}{r^{12}}$ term is a repulsive term that comes from Pauli's principle, which characterizes the repulsion of electrons surrounding a nucleus. $\frac{1}{r^{12}}$ is chosen for computational efficiency. The attractive $\frac{1}{r^6}$ term arises from the Van der Waals interaction, which is a classical description of induced dipole-dipole interactions. The values for the parameter ϵ and σ are specified for each atom type individually.
  As can be seen, the Lennard-Jones potential is very short ranged, due to its high exponents. Therefore, a cut-off point for the calculation is usually set at large distances, as it hardly contributes to the potential energy.

- *Coulomb interaction*
  The potential that describes the interaction of two point charges $q_{i,j}$ is called Coulomb interaction

$$V_{Coul} = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r}\frac{1}{r_{ij}}, \tag{2.26}$$

  where $r_{i,j}$ is the distance between the two charges and ϵ_0 is the electric constant. ϵ is called the dielectric constant, which represents the weakening of the interaction due to the surrounding material. For calculations in a vacuum, the dielectric constant is given as $\epsilon_r = 1$.
  As can be seen, the Coulomb interaction is a long-range interaction. This means that its contribution to the total energy of the system cannot be neglected even for large distances. To reduce the computational costs for the calculation, new techniques were developed. The **Ewald Summation** reduces the scaling for computational costs from an order of N^2 to $\mathcal{O}(N^{3/2})$. The Ewald summation can only be used in periodic systems. MD simulations often use periodic boundaries, so that an artificial periodic system is generated to which this method can be applied. It splits the electric potential into two parts. One part, dealing with short-range

interactions, stays in real space and the second part will be converted into Fourier space. The Fourier calculation can be further improved by introducing a mesh on which the charges are positioned. An algorithm using this method would be the Particle Mesh Ewald algorithm which only scales with $\mathcal{O}(N \ln N)$.

Parameters

The parameters that are neccessary for the force field can be obtained in two different ways. The experimental methods calculate them via a multi-dimensional fitting to experimental data. The other method is to derive them from quantum mechanical ab initio calculations. Once a set of parameters is gained, they only make sense as a whole, i.e. one single parameter does not need to have a real physical meaning, but when all parameters are taken into account, they can reproduce experimental observations.

2.2.4 Virtual Interaction Sites

Virtual interaction sites, short vsites, are dummy atoms in MD simulations. They can carry charges or be involved in any other interaction. In our case we replaced hydrogens by vsites to remove the fastest degrees of freedom, allowing us to use larger integration steps during the simulation. The dummy atoms are constructed in different ways for different types of hydrogens that are to be replaced.

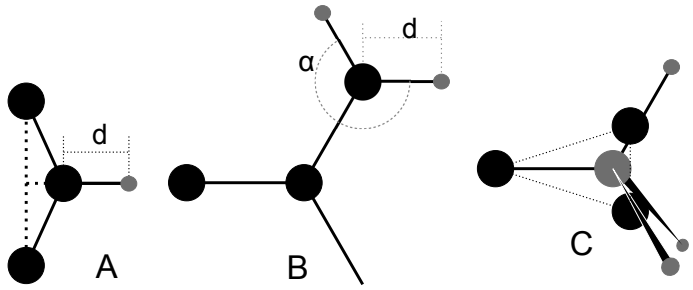

Figure 2.9: The three different types of virtual site construction for hydrogens. The black circles represent the atoms which are used to construct the virtual sites (grey circles). The larger circles stand for heavy atoms.

- *single amine or amide hydrogens:* The hydrogen is placed on a line going through the heavy atom and a point on the line through both second bonded atoms. The distance between the hydrogen and the heavy atom is fixed (Figure 2.9A).

- *planar amine hydrogens:* In this case, the nitrogen is only bonded to a single other heavy atom. So the method described above cannot by applied because of the missing second heavy atom. Hence, the hydrogens are again placed in a certain fixed distance to the nitrogen but this time with a fixed angle to the carbon atom

and in the plane spanned by the nitrogen, carbon and one of the other heavy atoms (Figure 2.9B).

- *amine group hydrogens:* To keep the rotational degree of freedom, the three hyrogens are replaced by two virtual sites to keep the rotational freedom, while removing the bond angle degrees of freedom. The two dummy atoms contain the same masses, moment of inertia and centre of mass as the three hydrogens and the nitrogen, but they have no other interaction than a connection to each other and to the carbon. The replaced atoms can be reconstructed by a linear combination of the two carbon mass vectors and their outer product (Figure 2.9C).

3 Results

After implementing the MC scheme (Fig. 2.7) for our CG-model described in section 2.1.2, the bond-angle, LJ, charge etc. parameters had to be determined to describe the behaviour of FG nups. To this and, we carried out aa-MD simulations of small model peptides and adjusted the parameters of the CG model to reproduce the aa calculations. We focussed on the aggregation of small peptides and on the persistence length of single large molecules as both characteristics are from a special point of interest to describe the seletivity barrier of the NPC.

3.1 Aggregation of Small Peptides

To observe the aggregation of small peptides, we set up MD simulation of six heptamers. Each heptamer was built the same way. To cover all the parameters we wanted to imply in the Monte Carlo simulation, we used four different heptamers in the MD simulations. The first heptamer was completely built of alanine. For the remaining 3 heptamers we exchanged the middle amino acid by either lysine, asparagine or aspartic acid. These 4 amino acids represent the following characteristics: hydrophobic, positively charged, polar and negatively charged. On both ends of a heptamer, we added caping groups. The simulations were done using Gromacs with the amber99 force field. They were performed in "SPC" water at a physiological salt concentration of 150 mM NaCl. The temperature was kept constant using V-rescale coupling at $T = 300\ K$ with a coupling time of $\tau_T = 0.1\ ps$. The pressure was coupled to a Parrinelly Rahman barostat with a coupling time of $\tau_p = 4\ ps$ and an isotropic compressibility of 4.5×10^{-5} bar^{-1}. Long range interactions were calculated by particle-mesh Ewald summation with a grid spacing of 0.12 nm and fourth order B-spline interpolation. The structures were written out every 10 ps. In all simulations we replaced the hydrogens by virtual sites to remove the fastest degrees of freedom. Because of this, an intergration time step of 4 fs could be used. The construction of the virtual sites used in the simulations is explained in section 2.2.4.

Each heptamer was equilibrated separately in a simulation of 20 ns before they where put together in one dodecahedral simulation box with periodic boundaries. The dodecahedral shape has the advantage that it fits the shape of the proteins better than a cubic box. Therefore, less water has to be simulated, which leads to a decrease of computational costs.
The 6 heptamers were placed randomly inside the simulation box. The cluster sizes were

determined by using the *g_mindist* program, which computes the minimum distance between two chosen groups. If the distance is shorter than 4 $\mathring{A}$, the two groups are assumed to form a cluster. This calculation is done for each possible pair of peptide chains in each frame. If more than one cluster is found in a single frame, the two clusters are analyzed to determine whether they are two idependent clusters or whether they belong to a single, but larger cluster. To give an example, if the program finds chain A and chain D to be in a cluster, but also finds chain D and F to form a cluster in the same frame, then the new cluster will be ADF, having a cluster size of 3. In contrast, if the program finds A and D to form a cluster and B and E, then it will treat them as two separate clusters, each having a cluster size of 2.
In Figure 3.1 the clustering behaviour of the observed heptamers is shown. It stands out that heptamers where the middle alanine was replaced by aspartic acid aggregate less than the others. In fact it seems that this set up prefers a cluster size of two heptamers. The largest clusters are formed by the pure alanine heptamers, followed by aspartic acid containing heptamers.

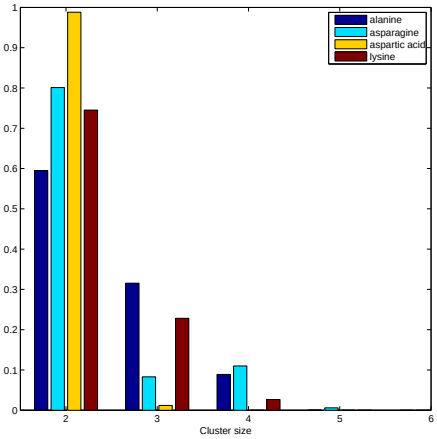

Figure 3.1: Clustering of heptamers observed with MD

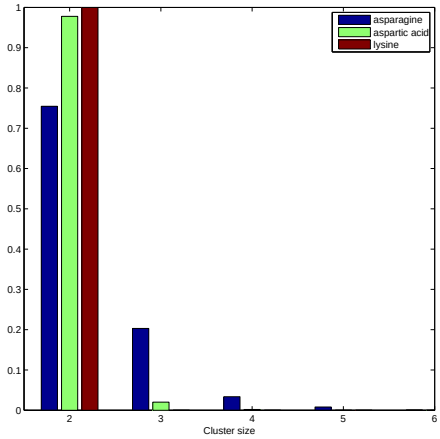

Figure 3.2: Clustering of nonamers observed with MC

In the MC simulation nonamers were used instead of heptamers. The two additional amino acids represent the capping groups in the MD simulation. Therefore, the peptide chains in the MC simulation consist of 9 alanines, where the middle one is again replaced by the amino acid of interest. Each simulation was performed at $T = 300\ K$, and a step size of 0.15, leading to a gaussian distribution with a mean value of 0 and a variance of 0.15^2, was used to perform 10,000,000 Monte Carlo steps. The nonamers were placed randomly, but not further away than $\sqrt{30}\ \mathring{A}$. Before they were put together in a single simulation, the nonamers were equilibrated in an additional MC simulation.

The first set of parameters for MC simulation was taken from Bellesia and Shea (Bellesia and Shea, 2007), who examined the structure and stability of chiral β-tapes

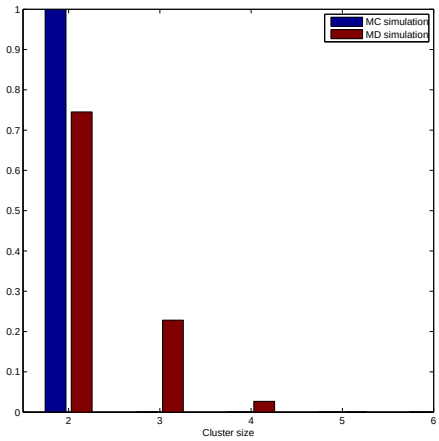

Figure 3.3: Aggregation of the peptide chain containing lysine in MD and MC simulation

using a coarse-grained model.
Figure 3.2 shows the cluster sizes obtained for the nonamers, except the ones made of pure alanine. Only for the negatively charged amino acid aspartic acid are the results nearly identical to the MD simulations. The behaviour of the set up where an asparagine replaced the middle alanine is close to the result found with MD simulations. The aggregation of pure alanine nonamers and of the nonamers where the middle amino acid is a lysine (Fig. 3.3) are completely different. Therefore, we focussed on generating a parameter set-up that will describe the aggregation as predicted by the MD simulations predict.
To this end, we performed MC simulations in which the interaction of two hydrophobic parts are made stronger. The results can be seen in Figure 3.4. With rising interaction strength, the clustering of the nonamers got closer to that observed in MD simulations. However, the difference in the size of the clusters is still too large. Therefore, the strength of the hydrophobic interaction has yet to be increased.

3.2 Persistence Length

The persistence length l_p describes the stiffness of a peptide or chain. It is given by

$$\langle \cos \beta \rangle \quad = \quad \mathrm{e}^{\frac{l}{l_p}} \tag{3.1}$$

where β is the bondangle and l is the length of a segment. Now, the persistence length is the length where the correlation between the bondangles has dropped to $\frac{1}{\mathrm{e}}$. For computation of the persistence length, we only used the C_αs of the peptide chains.

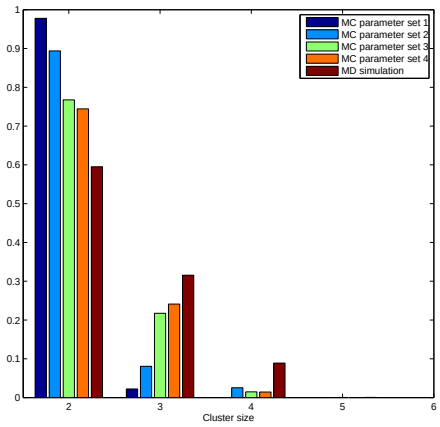

Figure 3.4: clustering of pure alanine peptide chains in MD and MC simulations. The strength of the hydrophic interaction used in the MC simulation rises from parameter 1 to 4

3.2.1 Vacuum Coiled Peptide

To begin with, we computed the the persistence length of a protein with 263 amino acids. Therefore, the protein was generated in *pymol* and first simulated in a vacuum to speed up the folding. The vacuum simulation was done at a temperature of 500 K using a temperature coupling time $\tau_T = 0,1\ ps$ for the Berendsen thermostat. The integration time step was chosen to be 2 fs. No periodic boundary conditions were applied. We let the simulation run until the linear starting structure was completely coiled. After this, a new simluation was started using the folded structure in a dodecahedral box. This time the simulation was done in "SPC" water at 300 K and a salt concentration of 150 mmol/L. The pressure was controlled by using Parrinello-Rahman isotropic pressure coupling with a coupling constant of $\tau_p = 5\ ps$ at a compressibility of 4.5×10^{-5} bar^{-1}. For V-rescale temperature coupling, we used a coupling time of $\tau_T = 2.5\ ps$. The integration time step was again chosen to be 2 fs. For computation of the long range interactions we used fourth order particle-mesh Ewald summation, with a grid spacing of 0.12 nm and cut off of 10 $\mathring{A}$. The structure was written out every 2 ps.
The persistence length was then calculated using only every tenth frame of the trajectory and we obtained a persistence length of $l_p = 1.23 \pm 0.02$ amino acids, which will lead to a persistence length of about 4 $\mathring{A}$. A similar persistence length of $l_p = 3.9 \pm 1.4\ \mathring{A}$, which is that of the length of an amino acid, was also found by Roderick Y. H. Lim et al (Lim et al., 2006), who determined the persistence length by using atomic force microscopy.

4 Discussion and Outlook

We showed that it is generally possible to reproduce the clustering behaviour of short peptide chains. The persistence lengths of longer molecules are of the same order of magnitude as found by experimentalists. According to the clustering behaviour of the peptide chains, where the middle alanine was replaced by a different amino acid, it seems that the ratio between the non-hydrophobic hydrophobic interaction and the hydrophobic hydrophobic interaction was already described well by the first parameter set. However, some parameters still need to be determined before simulations of a large network can be carried out. At present, we are continuuing zo perform MC simulations on pure alanine chains, until the clustering of this system, obtained by MD simulations, can be reproduced. After that, we we will go on with the parametrisation of lysine, which will be more difficult, as it also contains a coulomb interaction. In alanine the coulomb interaction was set to zero as it does not contain charged groups. Therefore, only a single parameter has to be changed in order to obtain a different aggregation. However, even in a set up of pure lysine chains, two different parameters have to be taken into account.
After the parameters have been set, we will simulate long peptide chains to determine the persistence length, as we did with MD simulations. We have already tried to compare peptides with a length of 20 amino acids regarding their persistence length. However, the obtained persistence lengths were much shorter than a segment length and in most cases it was not even possible to fit the persistence length to datapoints. Therefore, to calculate reliable values for the persistence length, the peptide chains have to be much larger than 20 amino acids.
As already mentioned in 2.1.2, we have the possibility to let the first part of a chain stay fixed. Later, we want to use this option to see whether the clustering of longer chains behaves differently when one end of them is fixed, compared to a situation in which the chains may move around freely. Furthermore, we want to observe how a change in the parameters will influence the properties of a meshwork. Last but not least, it has to be noted that the simulations of even short peptide chains are very time consuming. Therefore, the source code of the MC simulation will be optimized to reduce the computational time needed for a single MC step. This will allow simulations of larger systems, in which we are interested, in an appropriate amount of computational time.

Bibliography

C. Ader, S. Frey, W. Maas, H.B. Schmidt, D. G
"orlich, and M. Baldus. Amyloid-like interactions within nucleoporin FG hydrogels. *Proceedings of the National Academy of Sciences*, 107(14):6281, 2010.

G. Bellesia and J.E. Shea. Self-assembly of β-sheet forming peptides into chiral fibrillar aggregates. *The Journal of chemical physics*, 126:245104, 2007.

G. Bellesia, M.V. Fedorov, Y.A. Kuznetsov, and E.G. Timoshenko. Structure and stability of chiral β-tapes: A computational coarse-grained approach. *The Journal of chemical physics*, 122:134901, 2005.

M.R. Betancourt. Efficient Monte Carlo trial moves for polypeptide simulations. *The Journal of chemical physics*, 123:174905, 2005.

D.P. Denning, S.S. Patel, V. Uversky, A.L. Fink, and M. Rexach. Disorder in the nuclear pore complex: the FG repeat regions of nucleoporins are natively unfolded. *Proceedings of the National Academy of Sciences of the United States of America*, 100 (5):2450, 2003.

P.M. Diesinger and D.W. Heermann. Hydrophobicity as a possible reason for gelation of FG-rich nucleoporins. *European Biophysics Journal*, 39(2):299–306, 2010. ISSN 0175-7571.

N. Dölker, U. Zachariae, and H. Grubmüller. Molecular Dynamics Studies on the Permeability Barrier of Nuclear Pore Complexes. *Statistical Mechanics of Molecular Biophysics*, page 37.

N. Dölker, U. Zachariae, and H. Grubmüller. Hydrophilic Linkers and Polar Contacts Affect Aggregation of FG Repeat Peptides. *Biophysical journal*, 98(11):2653–2661, 2010. ISSN 0006-3495.

Daan Frenkel and Berend Smit, editors. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, Inc., Orlando, FL, USA, 1st edition, 1996. ISBN 0122673700.

S. Frey, R.P. Richter, and D. Gorlich. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science*, 314(5800):815, 2006.

R.Y.H. Lim, N.P. Huang, J. Köser, J. Deng, KH Lau, K. Schwarz-Herion, B. Fahrenkrog, and U. Aebi. Flexible phenylalanine-glycine nucleoporins as entropic barriers to nucleocytoplasmic transport. *Proceedings of the National Academy of Sciences*, 103(25): 9512, 2006.

E. Lindahl, B. Hess, and D. van der Spoel. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7(8):306–317, 2001. ISSN 1610-2940.

M. Mezei. Efficient Monte Carlo sampling for long molecular chains using local moves, tested on a solvated lipid bilayer. *Journal of Chemical Physics*, 118(8):3874–3879, 2003.

L. Miao and K. Schulten. Transport-Related Structures and Processes of the Nuclear Pore Complex Studied through Molecular Dynamics. *Structure*, 17(3):449–459, 2009.

R. Peters. Translocation Through the Nuclear Pore Complex: Selectivity and Speed by Reduction-of-Dimensionality. *Traffic*, 6(5):421–427, 2005. ISSN 1600-0854.

M.P. Rout, J.D. Aitchison, M.O. Magnasco, and B.T. Chait. Virtual gating and nuclear transport: the hole picture. *Trends in cell biology*, 13(12):622–628, 2003. ISSN 0962-8924.

Wilfred F. van Gunsteren, Anthony J. Wilkinson, and Paul K. Weiner. *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications Volume 2 (Computer Simulations of Biomolecular Systems) (v. 2)*. Springer, 1994. ISBN 9072199154.

K. Weis. The Nuclear Pore Complex: Oily Spaghetti or Gummy Bear? *Cell*, 130(3): 405–407, 2007.

S. Wälde and R.H. Kehlenbach. The Part and the Whole: functions of nucleoporins in nucleocytoplasmic transport. *Trends in cell biology*, 2010. ISSN 0962-8924.

J. Yamada, J.L. Phillips, S. Patel, G. Goldfien, A. Calestagne-Morelli, H. Huang, R. Reza, J. Acheson, V.V. Krishnan, S. Newsam, et al. A Bimodal Distribution of Two Distinct Categories of Intrinsically Disordered Structures with Separate Functions in FG Nucleoporins. *Molecular & Cellular Proteomics*, 9(10):2205, 2010. ISSN 1535-9476.

Acknowledgement

I would like to thank Helmut Grubmüller and Nicole Dölker for all their support and many interesting ideas, and for supervising my work. Whenever I ran into larger problems, they helped me out by showing me several potential solutions. I would also like to thank all members of the Department of Theoretical and Computational Biophysics for very helpful discussions and providing answers to all kinds of scientific and technical questions.
Last but not least, I would like to thank Annette Zippelius from the Institute for Theoretical Physics at the Georg-August-University Göttingen for being my official supervisor at the Faculty of Physics. Without her agreement I would not have been able to write an external diploma thesis.