1 Introduction

1.1 Focal adhesion kinase

Focal adhesions (FA) are macromolecular protein complexes which act as a connection hub between the cell, i.e. the cytoskeleton, and the extracellular matrix (ECM). They enable the cell to exert tension forces, but can also transduce mechanical stimuli from the ECM. One important protein associated to FA is the focal adhesion kinase (FAK). FAK occurs in several signalling pathways and is a key player in integrating extracellular stimuli. It is of large interest not least because in cancer cells often an overexpression of FAK can be found and understanding the activation processes and dynamics of FAK could give rise to new cancer treatments.

1.1.1 Structure

FAK consists of four domains: (i) a FERM domain at the N-terminus, (ii) a tyrosine kinase, (iii) a proline-rich region and (iv) a focal adhesion targeting (FAT) domain at the C-terminus (see Figure 1.1).
FERM (4.1 protein, ezrin, radixin and moesin) is a common protein domain which targets proteins to membranes [10] and consists of three subdomains: F1, F2 and F3. In the F2 subdomain a basic patch (^{216}KAKTLRK^{222}) can be found, which is a prominent binding site for phosphatidylinositol-4,5-bisphosphate (PI(4,5)P_2, see below).
The kinase consists of the C-lobe, the activation loop and the N-lobe. Catalytic activity of kinase is mainly regulated by the phosphorylation of Y^{576} and Y^{577}, which are located in the activation loop [8]. The kinase also provides binding sites for PI(4,5)P_2. One important one is located next to the basic patch of the FERM domain, but others (namely R^{508}, R^{514}, K^{515}, K^{621} and K^{627}) can be found on the larger surface of the kinase [13, 17].
The FERM domain and the kinase are connected by a linker region. In contrast to other kinase domains, the main autophosphorylation site of FAK, Y^{397}, can be found in this region and not in the kinase itself [16].
The FAT domain is linked to the kinase by a flexible proline-rich region. FAT targets FA by interacting with talin and paxillin, which are proteins associated with FA [2].

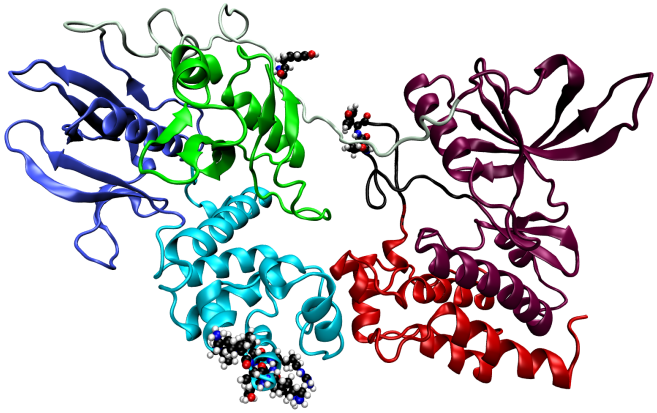


Figure 1.1: **Structure of FAK** FERM domain consiting of F1 (green), F2 (iceblue) and F3 (blue) connected via the linker to the kinase consisting of N-lobe (violet), the activation loop (black) and the C-lobe (red). The basic patch (bottom), Y^{397} (top) and Y^{567}, Y^{577} (activation loop) are highlighted.

1.1.2 PI(4,5)P_2 binding and activation

It is known that FAK triggers several stimuli. In this thesis, however, the focus is on the link of FAK activation and PI(4,5)P_2 binding. Therefore, the different binding sites of FAK for PI(4,5)P_2 and their impacts on FAK activation are discussed below.

In the autoinhibited conformation described above, the FERM domain shields the active site of the kinase. An activation of FAK therefore requires the dissociation of these domains [25].
PI(4,5)P_2 is a phospholipid, which is locally generated in FA due to integrin signaling [**CIT34**]. It has a net charge of -4, but in presence of K the deprotonated state gets promoted resulting in a net charge of -5 [29]. The electrostatic binding of PI(4,5)P_2 to the basic patch in the F2 subdomain results in long ranged configurational changes, which also influence the interface between the F1 subdomain and the N-lobe. Also the linker region gets less strongly bound, so that an autophosphorylation of Y^{397} is promoted. However, the PI(4,5)P_2 binding alone has no effect on the catalytic activity, which suggests that the FERM domain is still associated to the kinase [16, 37]. For activation, an additional stimulus, either biochemical or mechanical, is needed.

If Y^{397} is phosphorylated, it becomes a suitable binding site for SH2 or SH3, which are subdomains of proteins from the Src kinase family. Due to the conformational changes induced by PI(4,5)P_2, this kinase has access to Y^{576} and Y^{577}. As said, the phosphorylation of these residues makes FAK fully active through a dissociation of the FERM domain and the kinase [16].

Another stimulus leading to the dissociation of the FERM domain from the kinase is me-

chanical force. Forces onto the FAT domain are transduced to the interface of the FERM domain and the kinase, because the linker is connected to the kinase, while the FERM domain is anchored onto a PI(4,5)P_2 containing membrane. These forces can lead to activation of FAK. In that way it is acting as an mechanical sensor [38].

The binding sites on the larger surface of the kinase were hypothesized by computer simulations [13] and have been confirmed recently in experiments [17]. Their findings show that these residues are required for catalytic activity of the kinase, and that they bind to phospholipids *in vivo*. However, since the catalytic activity is not regulated by PI(4,5)P_2, this binding is supposed to act as a stabilisation of the active state on membranes only [17].

1.1.3 Dimerization, clustering and autophosphorylation

As said above, autophosphorylation of Y^{397} is an important event in FAK activation. It has been shown that this happens in intact cell in *in trans*, for which a self-association of FAK is required.

The FERM domain induces a dimerization of FAK, as it is doing in other proteins containing a FERM domain as well. The interaction emerges around W^{266} in the connected FERM domains and is stabilised by an interaction of the FAT domain with the basic patch of the other FERM domain respectively. The presence of W^{266} is also required for autophosphorylation of FAK.
For the dimerization of the FERM domains PI(4,5)P_2 is not needed. However, an enriched concentration of FAK is needed to observe FAK dimers in cells, which is the case at FA. It is still unclear how the dimer is stabilized at membranes, where the basic patch is also required for ligand binding [7].
It has been shown that FAK is autophosphorylated by other FAK molecules (intermolecular) and that this process is promoted by dimerization of FAK (not only due to FERM-FERM domains) [22].

Although PI(4,5)P_2 is not required for dimerization, it induces clustering of several FAK *in vitro* on membranes [16]. Because dimers support autophosphorylation of FAK, it is not surprising that the same effect is observed in clusters. However, these clusters can trigger additional biochemical stimuli [22] and may play an important role in the scaffolding function of FAK for FA [16].

2 Motivation

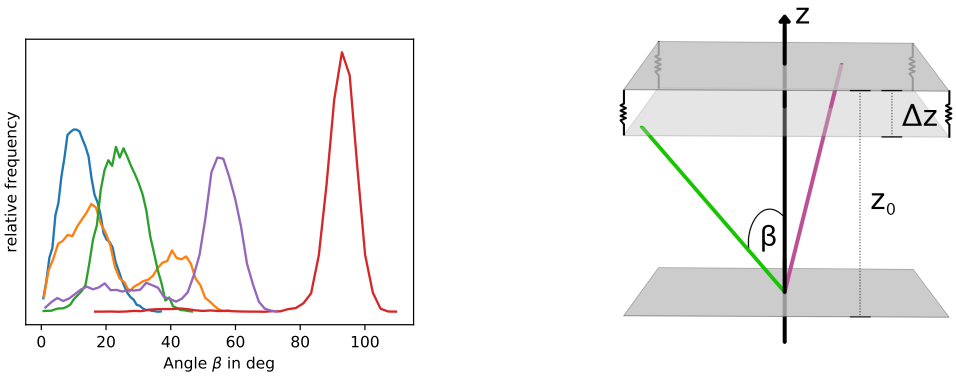
As described in subsection 1.1.3 the process of FAK clustering and its effect upon activation, especially on an atomistic scale, are still not understood and part of current research. In this thesis the results of MD simulations with the Martini force field (see section 3.1.2) are presented, which address the clustering process of FAK molecules. In this context Martini is a necessary simplification due to the large number of particles in systems containing sever FAK molecules.

However previous work in the group [3] obtained difficulties in the use of Martini for simulations of FAK on a PI(4,5)P_2 containing membrane. In some simulations equivalent to setup 3 (see section 4.4) except for an external force the protein rapidly changed its inclination to the membrane. In the following part this shall be characterised by the angle β between the z-axis and the vector connecting F1 and F2, $\vec{d}_F$.

$$\cos(\beta) = \frac{\vec{d}_{F,z}}{d_F}, \quad d_F = \left|\vec{d}_F\right| \tag{2.1}$$

Becker et al. [3] simulated five different copies for 10 μs each. The resulting distributions of the angle can be seen in Figure 2.1a. Exemplary the red line shows a mean value of 90 deg, which is what meant with a falling of FAK in the following part. The angle changed in less than XXX and stayed constant for the remaining simulation time.
There are several reasons why this is rather an artefact of the Martini force field than a possible binding pose of FAK to the membrane as suggested by Feng and Mertz [13]. The first one is, that FAK falls to both extensive sites, which means that the key residues for an interaction of the extensive site of the kinase with the membrane proposed by Feng and Mertz [13] are located on top of the FAK and not on the side of the membrane. Indeed contact analysis showed, that virtually all residues on the surface (in both, FERM domain and kinase) were interacting with the membrane. Another one is, that this behaviour was not observed in equivalent all atom simulations in C36 (1.5 μs in total). Here only two maxima were observed around 8 deg and around 20 deg, the largest observed angle was 40 deg.

In the course of this project several experiments were made to understand the cause of this falling, e. g. to review the binding of the basic patch in Martini, which is presented in subsection 5.2.1. However the reason could not have proven beyond doubt. In order



(a) (b)

Figure 2.1: **Inclination angle of FAK** (a): Distributions of β obtained by Becker et al. [3]. (b): Illustration of the applied force.

to still perform reasonable simulations of multiple FAK an external force was applied to each FAK molecule. This is called stabilizing force in the following parts.
The force is acting onto F1 and F2 parallel to the z-axis and is proportional to the deviation of their z-distance Δz from a reference distance z_0. An illustration of the force can be found in Figure 2.1b. For the determination of z_0 only the green and the blue distribution from Figure 2.1a were considered, because the large angles observed in the other distributions have not been observed in C36 simulations. The mean value of $\vec{d}_{F,z}$ for these two distributions is 2.228 nm, which was therefore set as z_0. *@Csaba: sorry, but how did we get this force constant of 1000 out of the distributions?*

In the following two chapters the used methods, i.e. MD simulation, are explained and the used setups introduced. Afterwards the obtained results are presented. For this purpose FAK in solution and FAK bound to PI(4,5)P_2 are analysed and compared to known information from experiments or other simulations. Also the impacts of the stabilizing force onto the simulations are commented. At last the focus is set on interactions between multiple FAK molecules.

3 Methods

3.1 Molecular dynamic simulation

Because the measurable time- and length scale of biological experiments are usually larger than those of the system, the dynamics in the system can only be measured indirectly. To get a more precise insight into the concrete processes, Molecular dynamics simulation (MD) are a suitable tool.
In the following section the main concepts of MD and characteristics of the used models are outlined, with a special focus on the MARTINI force field. Nevertheless, this can only be a small survey in the scope of this thesis. Because GROMACS (*GROningen MAchine for Chemical Simulations*) [1, 4] was used as MD engine in this thesis, the following explanations refer to GROMACS conventions and features.

3.1.1 The physics of MD

In MD atoms are modelled as solid beads, which comprise the nucleus as well as the electron shell and follow classical equations of motion. Due to this implicit treatments of electrons Quantum mechanical (QM) effects, such as excitation or electron transfer processes, are not accessible in MD. However, the beads are parametrized by effective parameters motivated by QM or experiments [30, p. 127f].
Newtons equation of motion can be turned into two first order differential equation

$$\frac{\mathrm{d}\vec{r}_i}{\mathrm{d}t} = \vec{v}_i \tag{3.1}$$

$$m_i \frac{\mathrm{d}\vec{v}_i}{\mathrm{d}t} = \vec{F}_i \tag{3.2}$$

which can be integrated numerically with e.g. Leapfrog- or Verlet integration scheme. Both are time reversible and symplectic, which ensures a small long term error in energy conservation [15, p. 72ff]. The force acting on particle i is given by the potential V at its position. In MD bonded and non-bonded interactions contribute to this potential, but also external forces can be applied.

$$\vec{F}_i = -\frac{\partial V}{\partial \vec{r}_i} = -\frac{\partial}{\partial \vec{r}_i}\left(V_{\text{bonded}} + V_{\text{non-bonded}} + V_{\text{external}}\right) \tag{3.3}$$

Bonded interactions

Bonded interactions act intramolecularly and describe chemical bonds. They can occur between two, three or four particles and refer to bond stretching, bending and torsion respectively. An illustration can be found in Figure 3.1a

Deviations of the bond length r from an equilibrium distance r_0 results in potential energy, which is usually described by a harmonic oscillator with the force constant k_{dist}.

$$V_{\text{dist, bond } i} = \frac{k_{\text{dist}}}{2} (r - r_0)^2 \tag{3.4}$$

For larger deviations a Morse potential, which assumes an exponential decay of the potential energy, is more precise, but has a much higher computational cost.

Bending of a chemical bond refers to deviations of the angle between three bonded θ partners from an equilibrium angle θ_0. The resulting harmonic potential is usually described by a harmonic oscillator as well.

$$V_{\text{angle}} = \frac{k_{\text{angle}}}{2} (\theta - \theta_0)^2 \tag{3.5}$$

The dihedral angle describes the angle between two planes, which go through 3 beads respectively and have 2 beads in common. Therefore it can be understood as a torsion angle, but can also be used to preserve plane rings and the chirality of four particle groups. The resulting potential energy is usually approximated with a periodic approach

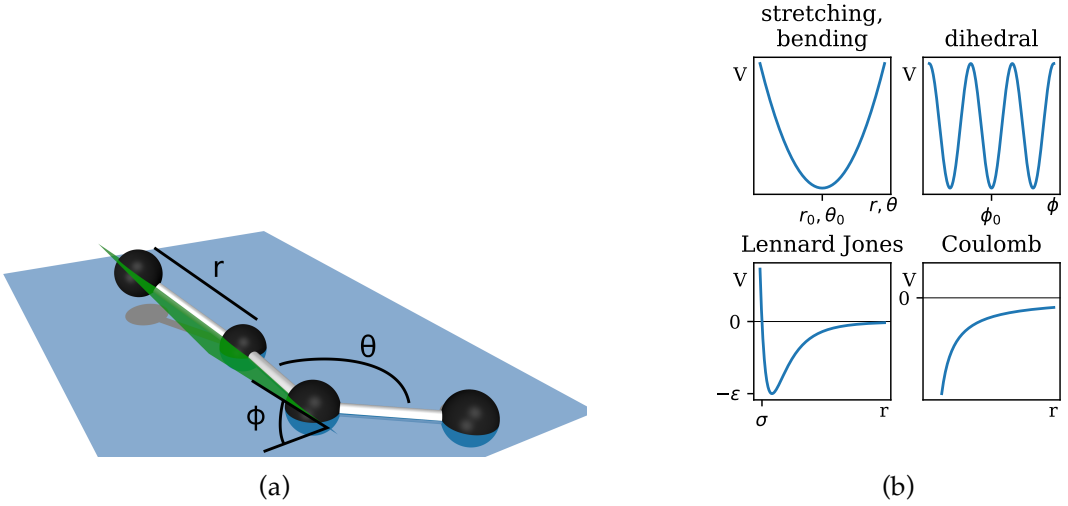
$$V_{\text{dihedral, periodic}} = \frac{k_{\text{dihedral}}}{2} (1 + \cos(n\phi - \phi_0)) \tag{3.6}$$

where k_{dihedral} describes the energy barrier for turning the dihedral angle, n the number of minima in the energy function (multiplicity) and ϕ_0 a phase factor [1, p. 71-83].

Non-bonded interactions

Non-bonded interactions are present between all atoms in the system and act pairwise. In MD Pauli repulsion, van der Waals (vdW) forces and electrostatic forces are taken into account. The Lennard-Jones potential combines Pauli repulsion (r^{-12} term) and the vdW force (r^{-6} term).

$$V_{\text{Lennard-Jones}} = \sum_{\text{non-bonded pairs i,j}} 4\epsilon \left(\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right) \tag{3.7}$$



(a) (b)

Figure 3.1: **Interactions in MD** (a): illustration of bonded interactions. (b): schematic course of contributing potentials

ϵ is related to the potential depth and σ to the potential range (Figure 3.1b).

The Coulomb potential is given by

$$V_{\text{Coulomb}} = \frac{q_1 q_2}{4\pi\epsilon_0\epsilon_r r} \tag{3.8}$$

where q_1, q_2 are the (partial) charges of the interacting particles, r their distance and ϵ_r the relative dielectric constant [1, p. 65-71].

In general non-bonded interactions act between all atoms in the system, which come along with a very large computational cost.
The easiest solution for this problem is to use a cut-off radius r_c. Particles beyond this radius are not taken into account. This can be implemented very efficiently with a Verlet neighbour lists. For each particle a neighbour list is created, which contains all particles inside a second radius r_v with $r_v > r_c$. This reduces the number of distance calculations a lot. The lists are updated, if the maximal displacement in the system is larger than $r_v - r_c$. The application of a cut-off radius is suitable especially for Lennard-Jones potential as it decays very rapidly and r_c can be chosen very small [30, p. 144].
Because the electrostatic potential is proportional to $1/r$ the use of a cut-off radius would lead to large jumps in the potential. That is why long range interactions have to be considered, which can be effectively done with Particle Mesh Ewald (PME) summation [11]. Particle mesh methods in general split the electrostatic potential up into a short range and a long range part via a switching function. The short range part can be calculated with a small cut-off radius in real space. The long range part however is calculated by solving the Poisson equation of the actual charge distribution, for which a discrete grid (mesh) is used. In contrast to other particle mesh methods, this grid is transformed to Fourier

space in PME with FFT techniques. Here the solution of the Poisson equation is a sum over the gridpoints. Afterwards the potential can be back transformed into real space. The use of PME requires of course periodic boundary conditions, which are described in section 3.1.3 [30, p. 246-251].

External forces

With GROMACS it is possible to perform pulling forces onto groups of atoms in the system. In this thesis pulling was used to bias distances between groups. For this GROMACS provides an option to apply an umbrella potential to two groups which yields in a force proportional to the deviation of the distance between the groups from a reference distance. The force can be applied on one or two spatial dimensions only or along a predefined vector and the reference distance can change during in time [1, p. 154-159].

3.1.2 Forcefields

The parameters for the potentials described above are provided by force fields. There is a wide range of force fields, which are optimized for different application fields. In this thesis tow different ones, CHARMM36 and MARTINI, are introduced.

Force fields define specific atom types, which allow a mapping of atoms (particles in physical system) onto beads (particles used in the simulation). This mapping can take the environment of the particle (binding partners, solvents, nearby charges a.o.) into account, but can also neglect details by f.e. mapping several atoms to one bead (coarse-graining). Force fields do not only define the atom types and their properties, but also all parameters for the calculation of the potential, namely force constants, equilibrium distances a.o. Therefore forcefields define the physics of the system.

All-atom and the CHARMM36 force field

CHARMM36 [6, 23] (C36) is part of the CHARMM (*Chemistry at HARvard Macromolecular Mechanics*) MD engine and was published in 2010. In C36 all atoms are considered (all-atom force field). Parameters were mainly optimized to structural experimental data, such as nuclear magnetic resonance (NMR) or X-ray data, but also QM and semi empirical QM calculations were used (e.g. for dihedral angles of the sidechains of proteins [6] and partial charges of lipids [23]).
All simulations for optimizing have been done with a time step of 2fs. Therefore very detailed dynamics are still included in the simulations.

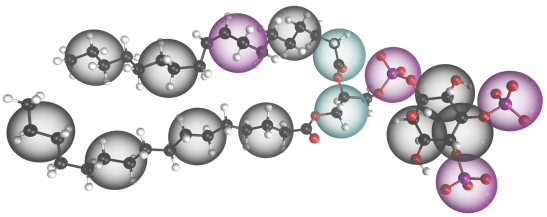


Figure 3.2: **MARTINI bead mapping** MARTINI structure of PI(4,5)P_2. Four heavy atoms are mapped to one MARTINI bead.

For all simulations the TIP3P water model [21] was used, which was also used in the parametrisation of C36. Here each water molecule is modelled with three polar beads.

Coarse graining and the MARTINI force field

The MARTINI force field [28, 20, 26] is one of the most famous coarse graining force fields. It maps usually four heavy atoms onto a single bead (Figure 3.2), which implicates of course a loss of chemical information, but also an enormous reduction of computational cost. With this approach much larger time- and spatial scales are accessible for MD simulations.

The parametrisation of MARTINI is mainly based on reproducing free energies, i.e. partitioning free energies and dimerization free energies of amino acid side chain analogues. For lipids also thermodynamic properties, such as area per lipid, have been considered. The parameters were optimized with a time step of 20fs up to 30fs [20, 26].

There are several side effects of coarse graining. Due to the mapping of four atoms to one bead, different atomistic structures can end up in the same Martini structures, which can be problematic e.g. for lipid tails. Also the secondary structure of proteins is affected as it becomes less stable in coarse grained models. Therefore it has to be to be constrained with elastic networks (additional bonds between backbone beads) and can not change during the simulation [27, p. 6812].

The decrease of degrees of freedom in the system also smooths the energy surface in coarse grained models. Because smaller local irregularities in the energy surface, which would slow down the evolving of the system, are smoothed out, coarse graining speeds up the dynamics in the system. The speed up factor is not constant, but can be relatively good approximated by factor 4 (obtained in most diffusion simulations) [27, p. 6810, 28, p. 7815].

Coarse graining also has an effect on the entropy and the temperature dependency of the system. In NpT ensembles the Gibbs free energy G is given by

$$G = H - TS \tag{3.9}$$

where S is the entropy and H the enthalpy. Due to a lower number of degrees of freedom in coarse grained systems the configurational entropy is reduced. Because MARTINI is tuned to free energy calculations, this implicates also a reduction of the enthalpy. The temperature dependency is therefore systematically biased [27].

Consistently with the four to one mapping, solvent beads in MARTINI represent four water molecules. The MARTINI water has a freezing temperature of $\approx 300\,\mathrm{K}$ and the freezing process in irreversible. Therefore antifreeze beads were introduced, which have a larger σ_{LJ} regarding interactions to normal solvent beads. By inserting these antifreeze beads the lattice conformation is disturbed and the freezing temperature reduced [28].
In contrast to water the solvent beads are not polarizable. For this reason electrostatic screening has to be treated implicitly, which is done by an relative dielectric constant $\epsilon_r = 15$. However, this approximation fails near to polar beads resulting in an underestimation of the interaction strength between polar beads [27]. To overcome implicit screening a polarizable water model was introduced to MARTINI, which is denoted as PW in the following. PW consists of three beads, two of them carrying charges. This model enables orientational polarizability and decreases the freezing point as well. Nevertheless it comes with a higher computational cost [36].

As reviewed by Marrink and Tieleman [27] MARTINI have been already used to study protein-membrane interactions as well as protein oligomerization on membranes. Also changes in tertiary structure have been addressed using MARTINI.

3.1.3 External constrains

Periodic boundary conditions

Because the simulation of an open system is not possible, boundary conditions have to be considered. Closed boundaries often lead to surface interaction artefacts and are therefore in most cases not suitable for MD simulation. That is why usually periodic boundary conditions (PBC) are used.
To use PBC the shape of the simulation box has to have a space filling geometry (e.g. rectangular or rhombic dodecahedron). With periodic boundary conditions images of the simulation box are repeated in every direction. If a particle leaves the simulation box, its periodic image is coming in from the opposite. In that way the number of particles is kept constant while surface interactions are avoided.
A particle interacts only with the nearest image of another particle, which means that particles near boundaries can interact with periodic images of other particles instead of the real particle. Nevertheless molecules can have long range interaction with their own

periodic images, which leads to artefacts and has to be considered when choosing the size of the simulation box [30, p. 141f].

Thermostats

Integration schemes in MD are designed to conserve the energy of a system. However in real biological systems not the energy but rather temperature is kept constant. This can be achieved with thermostats. For the simulations in this thesis the Berendsen thermostat [5] (BT), the Parrinello-Bussi thermostat [12] (PBT) and the Nosè-Hoover (NHT) thermostat [31, 19] were used. BT couples the system to a heat bath by rescaling the velocities of the particles, which results in an exponential decay of temperature deviations. The coupling strength is given by the time scale, over which the energy is conducted. The rescaling involves a transfer of kinetic energy from internal degrees of freedom to translational and rotational kinetic energy of the systems COM [18]. In order to prevent this transfer PBT extends BT by a stochastic term. Both BT and PBT are usefull for equilibration runs or non equilibrium MD simulations as they are stable upon large deviations [5, 1, p. 31].
In comparism NHT extends the Hamiltonian of the system by a friction term representing a heat bath. The friction parameter follow a differential equation depending on the temperature deviation. This ansatz samples the phase space more accurately making NHT suitable for production runs. However, for large deviations NHT gets unstable [1, p. 32f].

Often groups are coupled to independent thermostats. This is helpful, because the heat exchange between f.e. proteins and solvents is often not correct. Therefore proteins would cool down and the solvent would heat up [1, p. 34].

Barostats

In biological systems often not the volume but the pressure is constant, which can be achieved in MD with different barostats. For equilibration runs the Berendsen barostat [5] was used, which couples the system, analogously to the BT, to a external pressure by rescaling the positions of the particles [1, p. 36]. For production runs the Parinello-Rahman barostat [33, 32] was used, because it samples the phase space more accurately. This barostat allows also a rotation of the position vectors. The appropriate matrix follows a differential equation depending upon the current deviation of the pressure from the external pressure. However, large deviations lead to oscillations in the box [1, p. 36].

GROMACS provides the possibility to couple the z-direction independently from the x- and y-direction, which is called semiisotropic pressure coupling. This feature is useful for membrane and pulling simulations, because the dynamics differ a lot between these axes.

3.2 Free energy calculations

3.3 Free energy

In order to understand state transitions in physical systems the free energy is a key quantity. It is directly linked to the probability distribution for different states and other quantities can be derived easily.
The free energy depends on the partition function and therefore on the considered thermodynamic ensemble. In this section only the Helmholtz free energy A, which refers to a canonical ensemble, is examined, but the concepts can be applied to other ensembles as well.

Partition function

In the microcanonical ensemble (N, V and E constant) the probability that a system enters a microstate with energy $E' = E(\mathbf{q}, \mathbf{p})$ ($\mathbf{q}, \mathbf{p}$ are the positions and momenta of the particles respectively), is equal for all $|E' - E| < \mathrm{d}E$ and 0 else. Therefore the partition function Ω is given as

$$\Omega(N, V, E) = C_0 \int \delta\left(\mathcal{H}(\mathbf{q}, \mathbf{p}) - E\right) \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p} \tag{3.10}$$

where $\mathcal{H}(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + K(\mathbf{p})$ is the Hamiltonian and C_0 a proportional constant, in which the smallest phase space volume and the indistinguishability of particles have to be taken into account [9, p. 16].

In the canonical ensemble however the temperature T is kept constant instead of the energy. Therefore the partition function has to include all possible energies weighted with their probability given by the Boltzmann factor. Below $\frac{1}{k_B T}$ is shortened with β.

$$Q(N, V, T) = \int \exp\left(-\beta E\right) \Omega(N, V, E) \mathrm{d}E \tag{3.11}$$

$$= C_0 \int \exp\left(-\beta \mathcal{H}(\mathbf{q}, \mathbf{p})\right) \mathrm{d}\mathbf{q}\mathrm{d}\mathbf{p} \tag{3.12}$$

The configurational integral is defined as

$$Z(N, V, T) = \int \exp\left(-\beta U(\mathbf{q})\right) \mathrm{d}\mathbf{q} \tag{3.13}$$

It is important to see, that $\mathcal{H}$ only depends on the quadrature of $\mathbf{p}$, so the integral over the momenta can always be solved analytical by turning it into a Gauss integral. This implies, that for two related systems, in which the particle masses are the same, the integral over

p does not change and therefore

$$\frac{Q_2}{Q_1} = \frac{Z_2}{Z_1} \tag{3.14}$$

holds [9, p. 17].

The Helmholtz free energy A is definied as

$$A = -\beta \ln(Q) \tag{3.15}$$

Usually the exact value of the free energy is unknown, but the main interest lies in free energy differences between two states of a system. The difference ΔA can be written as

$$\Delta A = A_2 - A_1 = -\beta \ln(Q_2) + \beta \ln(Q_1) = -\beta \ln\left(\frac{Q2}{Q1}\right) = -\beta \ln\left(\frac{Z_2}{Z_1}\right) \tag{3.16}$$

Therefore the normalisation constant C_0 in Equation 3.10 cancels out.

Free energy in MD and Umbrella Sampling

The free energy of a system as a function of a parameter ξ is given as

$$A(\xi) = -\beta \ln(\rho(\xi)) \tag{3.17}$$

$\rho(\xi)$ is the probability distribution of ξ and can be easily obtained in MD simulation by counting the number of the appropriate states. Therefore it is also referred as histogram. ξ is called reaction coordinate and could be e.g. the distance between two molecules. However in reality $A(\xi)$ can have big barriers and because the potential energy U is sharply distributed around its mean in NVT simulations, $\rho\xi$ can hardly be sampled during a finite simulation.

One possibility to overcome this sampling problem is umbrella sampling [34]. In this approach the path ξ is split up into distinct windows $[\xi_0, \xi_n]$. To each window i a biasing potential $\hat{U}_i(\xi)$ can be applied to ensure a well sampling of ξ around ξ_i. This changes the potential energy to

$$U_{B,i}(\mathbf{q}) = U(\mathbf{q}) + \hat{U}_i(\xi(\mathbf{q}))) \tag{3.18}$$

After a simulation with the biased potential $U_{B,i}$ the unbiased measured probability distribution $\tilde{\rho}_i(\xi)$ has to be reconstructed from the observed biased one $\tilde{\rho}_{B,i}(\xi)$. At this point only the result is given, the full derivation can be found in Frenkel and Smit [15, p. 179ff].

$$\tilde{\rho}_i(\xi) = \exp\left(-\beta\left(\Delta A_i - \hat{U}_i(\xi)\right)\right)\tilde{\rho}_{B,i}(\xi) \tag{3.19}$$

Of course ΔA_i is not known, but assuming that $\rho(\xi)$ is a continuous function, the results from the single windows can be combined and afterwards normalized.

With this method however only two histograms can be considered in the overlapping region. Another problem is, that the sampling in the tails is usually poor and statistical errors, which propagate through all overlapping regions, can become very large [9, 236ff]. Therefore umbrella sampling is usually combined with the Weighted histogram analysis method (WHAM) [14, 24]. This algorithm is able to combine several histograms in one overlapping region and it is designed to keep the statistical errors small. The main idea is to combine probability densities linearly with an additional weighting factor $\omega_i(\xi)$ to the total probability density $\tilde{\rho}$. The weighting factors ω_i are chosen iteratively in a way, that the overall statistical error is minimized.

3.4 Contact Analysis

3.4.1 Intromolecular

The interface of the FERM domain and the kinase is of special interest since the domains have to dissociate for a full activation. Therefore contacts and contact areas at the interface are investigated using the following methods.

ConAn

404 Section not found. Please come back later...

Contact area

The contact area of the interface is determined with the solvent accessible surface area (SASA) and is defined as

$$\mathrm{CA} = \frac{1}{2}\left(\mathrm{SASA}_{\mathrm{FERM}} + \mathrm{SASA}_{\mathrm{kinase}} - \mathrm{SASA}_{\mathrm{FERM\text{-}kinase}}\right) \tag{3.20}$$

The calculation of the SASA values was done with GROMACS sasa tool.

3.4.2 Intermolecular

In order to determine the cluster behaviour of FAK intermolecular contacts of FAK molecules have to be considered. For this purpose, the following terms are defined.

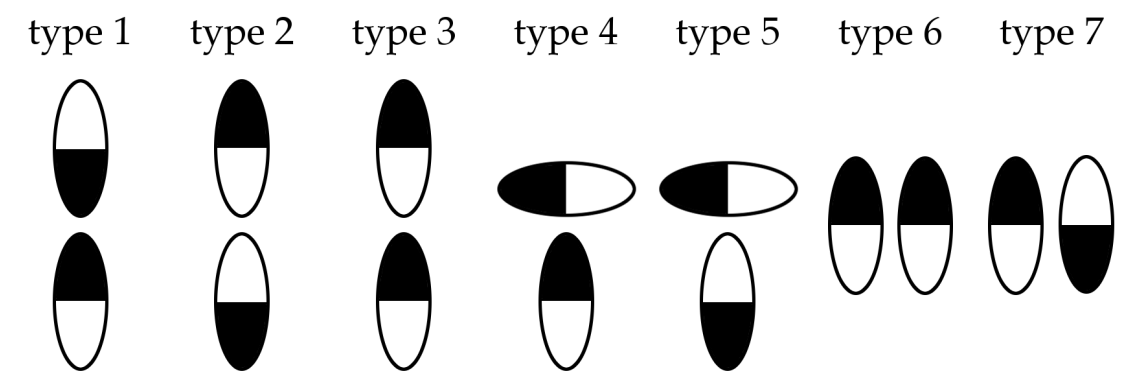


Figure 3.3: **Different two-molecule interaction types** The black part refers to the FERM domain, the white to the kinase.

Interaction Proteins or parts of proteins interact, if their minimal distance is smaller than a cut-off distance (here 1.5 nm).

Neighbour Protein A is a neighbour of Protein B, if they are interacting. One protein can have several neighbours. For a more detailed characterisation the following two-molecule interaction types are defined (see also Figure 3.3):

type 1: only the FERM domain interacts with only the FERM domain of the other protein

type 2: only the kinase interacts with only the kinase of the other protein

type 3: only the FERM domain interacts with only the kinase of the other protein

type 4: the FERM domain is interacting with both, the FERM and kinase of the other protein

type 5: the kinase is interacting with both, the FERM and kinase of the other protein

type 6: the FERM domain is interacting with the FERM domain of the other protein and the kinase is interacting with the kinase of the other protein

type 7: the FERM domain is interacting with the kinase of the other protein and the kinase is interacting with the FERM domain of the other protein

Cluster Protein A belongs to a cluster, if it has at least one neighbour inside the cluster. Two neighbouring proteins form a cluster of size 2. One protein can only belong to one cluster.

4 Setup

4.1 Protein structure

All simulations have been done with starting configurations adapted from previous work in the group (C36 forcefield: Zhou et al. [37], Martini forcefield: Becker et al. [3]). These configurations contain only a FERM-kinase fragment without the FAT domain and its linker (residues 35 to 686, PDB 2J0J [25]).
As explained in section 3.1.2 the secondary structure of proteins have to be stabilized in Martini using elastic networks. This was set up by Becker et al. [3]. It acts only between residues of the same domain with a force constant of 830 $kJmol^{-1}nm^{-2}$. Therefore the interface between FERM domain and kinase is not affected and the linker is still flexible.

4.2 Setup 1 - FAK in solution

This setup contains a single FAK molecule in Martini structure, which was placed in a waterbox with ions to eliminate net charges. After a short equilibration the system was simulated for $20\,\mu s$ at a temperature of 300K. The standard parameters of the Martini forcefield were used as simulation parameters.

4.3 Setup 2 - Free energy of basic patch

For this setup only a part of F2 (residues 107 to 219, referred as F2 lobe in the following), which contains the basic patch, was used. The lobe was placed above a single $PI(4,5)P_2$ embedded into a phosphatidylethanolamine (POPE) membrane (see **??**). After a short equilibration the F2 lobe was pulled slowly away from the membrane using a distance pull between the COM of the F2 lobe and the COM of $PI(4,5)P_2$. This simulation was used to retrieve starting conformations for the umbrella windows.
The starting configuration was set up in C36 and transferred to a Martini (with PW) structure with provided transformation tools [35]. In the Martini structures the mentioned elastic network was applied.
The number of umbrella windows was chosen accordingly to the sampling (between 90

and 120 windows). Each window was shortly equilibrated and then simulated for 6 ns in C36 and 10 ns in Martini (with PW). From the trajectories of the umbrella windows the free energy profile was calculated using GROMACS WHAM implementation. For each forcefield five different profiles were obtained to estimate the statistical error.
The presented results are based on a total simulation time of $6.33\,\mu$s for Martini, $5.64\,\mu$s for Martini with PW and $3.88\,\mu$s for C36. The temperature in the simulation was 300K.

4.4 Setup 3 - FAK on a PI(4,5)P_2 membrane

Setup 3 was adopted from Becker et al. [3] and contains a single FAK molecule in Martini structure, which was placed on a phosphatidylcholine (POPC) and PI(4,5)P_2 (15%) membrane. Ions were added to eliminate net charges. In contrast to Becker et al. [3] the stabilizing force explained in chapter 2 was applied to the protein.
Five different copies were simulated for $10\,\mu$s each with a temperature of 300K.

4.5 Setup 4 - FAK cluster

In order to set up a cluster of multiple FAK molecules 25 different frames were chosen from setup 3 equidistant in time. The frames were arranged on a 5x5 grid. The stabilizing force was applied on each FAK molecule independently. After a short equilibration the system was simulated for $9\,\mu$s. Five different copies (regarding to the arrangement of the frames) were set up, resulting in a total simulation time of $45\,\mu$s. The temperature was set to 300K.

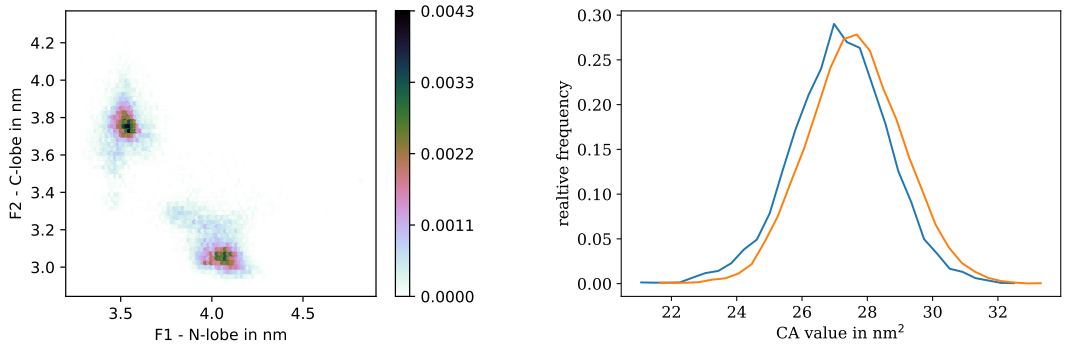
5 Results

5.1 FAK in solution

In this section the conformation of soluted FAK (FAK-SOL) for the MARTINIforce field is investigated. Since the secondary structure of the two domains is fixed with the elastic network, the focus is on the FERM kinase interface.

First the COM distances of F1 to the N-lobe ($d_{\text{F1-N}}$) and F2 to the C-lobe ($d_{\text{F2-C}}$) are considered. The two dimensional histogram of the distances reveals two different states (Figure 5.1a). Spot 1 refers to conformations, which are partially opened at the F2 - C-lobe interface, but close at the F1 - N-lobe interface. In contrast to this the conformations of spot 2 refers to states, in which F2 and the C-lobe gets closer while the distance between F1 and the N-lobe is increased. During the simulation several transitions between the spots were obtained, which indicates no convergence to a preferred spot. However, there is only a minor effect upon the contact area (Figure 5.1b). Spot 2 show a slightly larger mean of 27.6 nm^2 in comparison to spot 1 (27.1 nm^2).

A contact map of the interface between the FERM domain and the kinase for frames of spot 2 can be found in Figure 5.2. Two contact areas can be identified at the FERM-kinase interface. The first one (area 1) is located between F1 and the N-lobe/activation



(a) (b)

Figure 5.1: **Domain distances and contact area of FAK-SOL** (a): two dimensional, normed histogram of $d_{\text{F1-N}}$ and $d_{\text{F2-C}}$. (b): distribution of CA

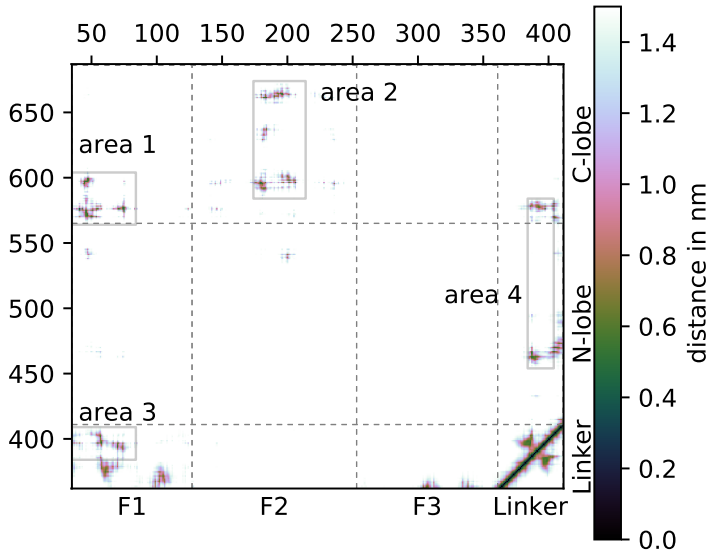


Figure 5.2: Contactmap of interface between FERM domain and kinase

loop. It shows i.e. contacts between Y^{576} and Y^{577} and residues of the FERM domain. The minimal distance in this area, occurring between residue H^{41} and Y^{576}, is 0.45 nm with an RMSF value of 0.03 nm. This area reflects the burying of the activity regulating residues in closed state.
The second contact area (area 2) is located between F2 and the C-lobe. The spots occur around the residues Y^{180} and D^{200} of F2 as well as F^{596} and R^{665} of the C-lobe. The minimal distance in this area occurs between Y^{180} and F^{596} with 0.45 nm and an RMSF value of 0.02 nm. Mutation experiments showed, that these two residues have an important effect upon the interface [25], which fits to the obtained contact.
The linker show contacts with both domains. The minimal distances in the marked areas occur between the autophosphorylation site Y^{397} and H^{58} of F1 (0.45 nm, RMSF 0.03 nm) as well as Y^{397} and Y^{576} of the kinase (0.50 nm, RMSF 0.10 nm). These contacts support the thesis, that autophosphorylation is prevented in closed conformation by a binding of the linker to the FERM domain [37].

5.2 FAK binding to PI(4,5)P_2

In this section the impacts on the conformation due to a binding to a PI(4,5)P_2 containing membrane are investigated and compared to results from section 5.1.
First of all measurements of the free energy profile of the basic patch bound to a PI(4,5)P_2

are presented. Afterwards simulations of setup 4 are considered. Because a stabilizing force was applied to the FAK molecules in these simulations, the impacts of the force are estimated before conformational changes are investigated.

5.2.1 Free energy profile of basic patch

In order to understand the falling of FAK onto the membrane the free energy profile of the basic patch PI(4,5)P_2 was investigated. Since this binding is the main contact between FAK and the membrane in our model, a short report on the results is given at this point. The profiles are obtained from setup 2. The reaction coordinate is the z component of the COM distance between PI(4,5)P_2 and the protein fragment. For each forcefield, C36, MARTINI and MARTINI with PW, the average profile out of the five copies together with the standard deviation can be found in Figure 5.3. The range $6\,\text{nm} \leq z \leq 7\,\text{nm}$ was set as zero point. Both, C36 and MARTINI, show a similar energy depth of $\approx 17\text{kJ/mol}$ and a similar slope between $3\,\text{nm} \leq z \leq 4\,\text{nm}$. Certainly, MARTINI shows systematically larger values than the all atom simulation. This can be attributed to the proposed underestimation of electrostatic forces due to the unpolar water beads (see section 3.1.2). The difference at $z = 4.2\,\text{nm}$ originates from the different treatment of long range electrostatics: MARTINI uses a cut-off radius, C36 uses PME.
Also MARTINI with PW uses PME for long range electrostatic interactions and indeed it fits much better to the C36 profile for larger distances. However the binding strength is crucially underestimated in Martini with PW.
The extent to which MARTINI reproduces the results from all-atom simulations is remarkable, even though the parameters for MARTINI were obtained a.o. from free energy calculations (see section 3.1.2)

In the used starting configurations, the proteins are already bound to PI(4,5)P_2. Therefore a correct binding strength and the shape near the minimum is of larger interest than a correct sampling of farther distances. In addition MARTINI with PW required a much higher computational effort. That is why in the following simulations only MARTINI with the standard water model is considered.

5.2.2 Impacts of the stabilizing force

In setup 3 a and setup 4 each protein was stabilized with an external force acting on the FERM domain (see chapter 2). Therefore the dependency of the used observables is examined below.

The force has a mean value of $1.47\,\text{pN}$ and a standard deviation of $13.13\,\text{pN}$. It is skewed to positive values. This is expected since positive values of the force require negative

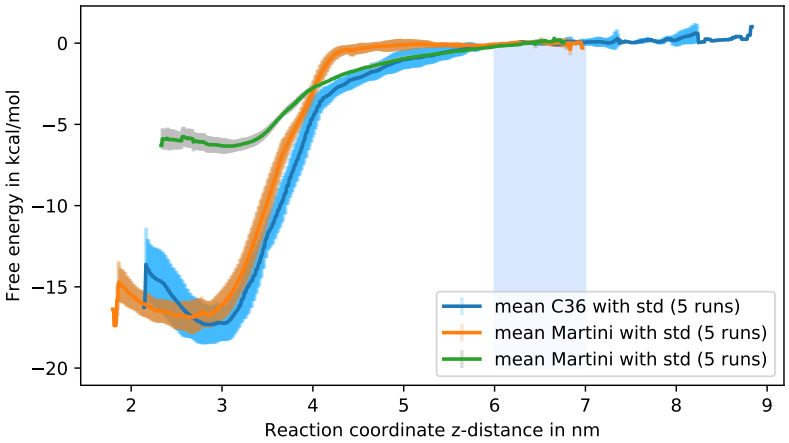


Figure 5.3: **Free energy profile of basic patch** For each forcefield the average and standard deviation of the five copies is presented.

elongations Δz. If the connecting vector of F1 and F2 $\vec{d}_F$ is parallel to the z-axis, larger negative deviations require a stretching of d_F, which suppress these forces.

Linear regressions show that none of the quantities d_F, $d_{\text{F1-N}}$, $d_{\text{F2-C}}$ and CA have a remarkable correlation to the applied force. Here remarkable means that either the regression result was not significant or that the obtained slope was so small, that a change of two standard deviations in force would not change the quantity noticeably.

Also the distances between residue pairs are tested for correlation with the applied force. To this end, 10 different proteins without neighbours were picked out of the trajectories of setup 4, each for $1\,\mu$s. On this dataset a linear regression was performed for each residue pair. Figure 5.4 shows the calculated Pearson correlation coefficient (only significant correlations with Pearson $|r| > 0.3$). The mean value of the slope for the positive correlated pair distances is $20.3\,\text{pN/nm}$ and $-19.7\,\text{pN/nm}$ for the negative correlated pair distances. Thus, the force can have large influences on these pairs. However, not all observed contacts at the interface (compare to Figure 5.2) are effected.
It is important to note that in this section only the observed quantities were tested for correlation with the force. It is however possible and probable, that due to this restriction, configurations or states are completely missing while others are over expressed. This effect can not be estimated, but should be kept in mind.

5.2.3 Conformational changes

In this section the conformation of FAK bound to PI(4,5)P_2 containing membrane (FAK-MEM) is compared to the observations from section 5.1 (FAK-SOL). For this purpose the simulation data of setup 4 was used with the condition, that other FAK molecules are more

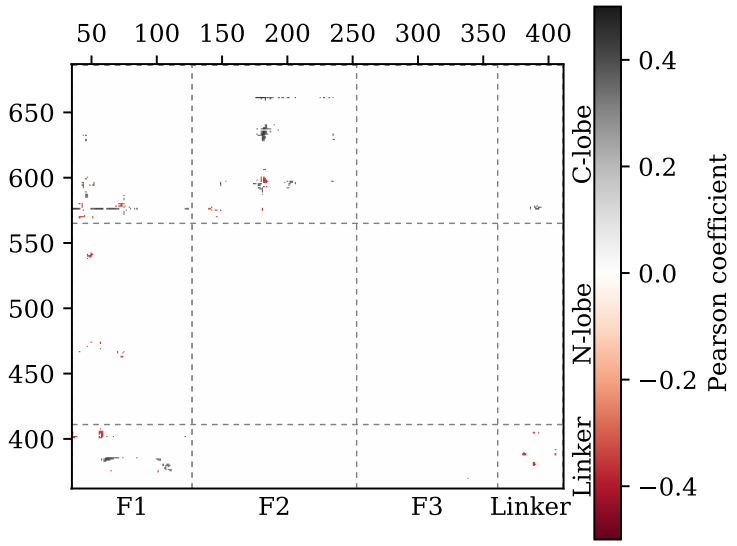


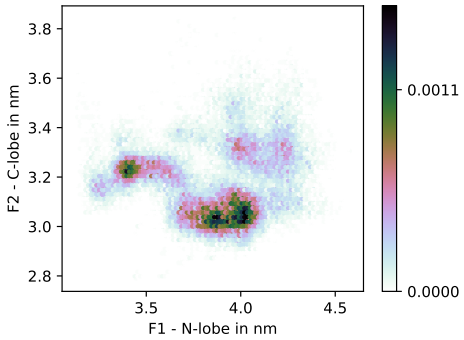
Figure 5.4: **Correlations in contact map** The contact map shows residue pairs, whose distance correlates with the applied force. The obtained slope is about 20 pN/nm.

than 2 nm away (0 neighbours). The contact map is based on the same dataset, which was used in **??**.

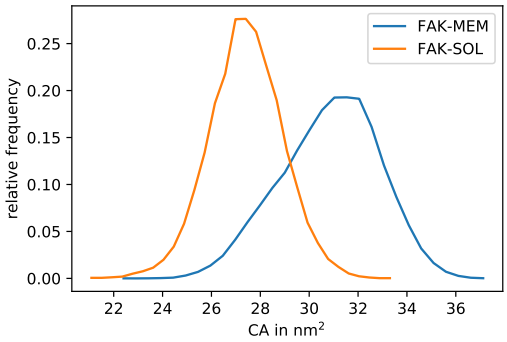
Analogously to section 5.1 the distribution of the COM distances is presented in Figure 5.5a as an hexagonal binning plot. Again two spots can be identified, which show a large number of encounters. However these spots are much more smeared out than obtained in FAK-SOL. In addition larger values of $d_{\text{F2-C}}$ disappear. This is expected, as the kinase is also bound to the membrane, resulting in less deviations in the lower part of the protein. Also the contact area of the interface, which can be found in Figure 5.5b, indicates a closure of the protein.

In contrast the linker region containing the autophosphorylation site Y^{397} seems to be released from the two domains. On the one hand the mean distance to residues of the domain increases by up to 0.67 nm (S^{74} and T^{394}), on the other hand the RMSF values are increasing as well, which indicates a more flexible region.

The obtained results are consistent with previous observations. PI(4,5)P_2does not induce the dissociation of the FERM domain and the kinase, but it enhances autophosphorylation. Later can not be simulated in MD simulation, but the changes in the linker region support this assumption.



(a) Two dimensional histogram of $d_{\text{F1-N}}$ and $d_{\text{F2-C}}$



(b) Contactarea of FERM-kinase interface

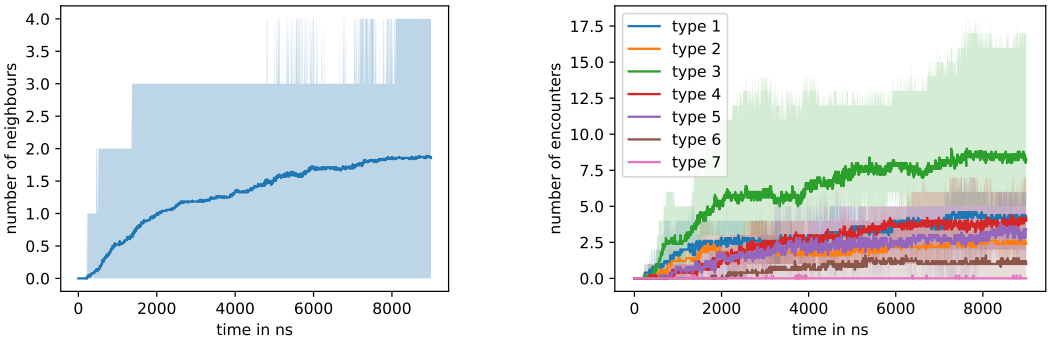
5.3 Multiple FAK interactions

In this section the interactions occuring between multiple FAK molecules are analysed, for which the data from setup 4 is used. At this point the reader shall be reminded, that the used protein structure lacks the FAT domain, which is in full length FAK connected to the kinase via a linker region. This might make an important difference for clustering processes.

5.3.1 Structure of FAK oligomers

The characterisation of the emerged FAK clusters is very difficult as they differ a lot in size and shape. The largest cluster observed in setup 4 had a size of 21 proteins, while there are other proteins, which did not join any cluster at all. Present shapes of the clusters include long chains as well as ring like conformations or just an agglomeration (see **??**).

First of all the mean number of neighbours is examined. One can see in Figure 5.6a a fast rising in the number of neighbours in the beginning and a flattening after 6 ns. The average of the five copies is at the end of the simulation 1.86. This indicates a tendency to chains of FAK.
In Figure 5.6b the average number of encounters of the different interaction types is plotted against the time. It shows, that FERM-kinase interactions (type 3) occur the most, while the others occur equally often. Only type 6 and type 7 (interactions, in which all four domains are involved) occur much less than the others.
From these observations one could draw the conclusion, that the preferred arrangement of FAK molecules is a chain, in which the FERM domain interacts with the kinase of the next molecule (FK-Chain). A second possibility would be a chain, in which the FERM domain interacts with the FERM domain of the next molecule, while the kinase interacts



(a) Mean number of neighbours against the time. Filled area is observed minimum and maximum.

(b) Mean number of interactions against the time. Filled area is observed minimum and maximum.

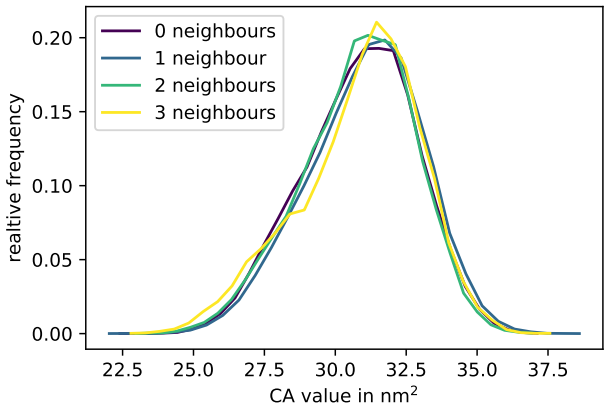


Figure 5.7: CA for different number of neighbours

with the kinase of the previous molecule (FFKK-Chain). Assuming a FAK chain of length n, FK-Chains would show n encounters of type 3 interactions, FFKK-Chains only $n/2$, but for both type 1 and type 2. This would also be consistent with the observed data.

5.3.2 Activation due to clustering

At last the impacts of clustering on FAK activation are addressed. Activation means here the dissociation of the FERM domain from the kinase, therefore the obtained trajectories are analysed with respect to the contact area (CA) of the FERM-kinase interface as a key quantity. Unfortunately in no FAK molecule a full dissociation took place at any time, therefore only trends can be considered at this point.

At first glance the CA seems to be independent of the number of neighbours a protein has as well of the interactions it is in (see explanatory for the number of neighbours Figure 5.7). Therefore the data has to be filtered more.

Motivated from subsection 5.3.1, only FAK molecules inside chains are taken into account.

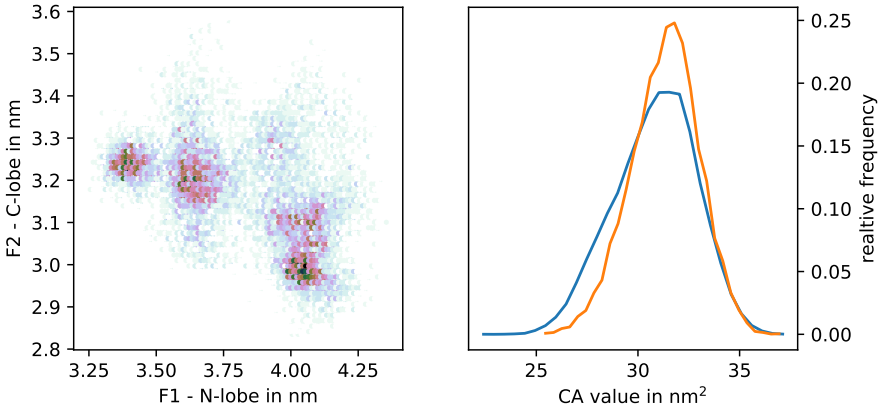


Figure 5.8: Analysis of the FERM-kinase interface in FK-Chains. The blue line was obtained from FAK-MEM.

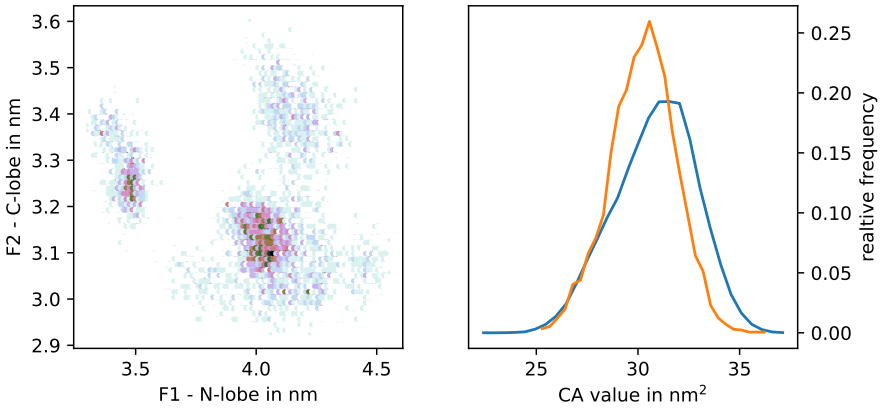


Figure 5.9: Analysis of the FERM-kinase interface in FFKK-Chains. The blue line was obtained from FAK-MEM.

A FAK molecule can be seen as a chain member, if it has exactly two neighbours and if these neighbours are not neighbours of one another. For FK-Chain only type 3 interactions were allowed, for FFKK-Chains both, type 1 and type 2. The resulting distribution of CA as well as the COM distances $d_{\text{F1-N}}$ and $d_{\text{F2-C}}$ can be found in Figure 5.8 for FK-Chain and in Figure 5.9 for FFKK-Chain.

As one can see in the plots FK-Chains do not have an influence upon the CA. Also the distribution of $d_{\text{F1-N}}$ and $d_{\text{F2-C}}$ is very similar to the one obtained in Figure 5.5a, except for less sampling. However in FFKK-Chains the mean CA value is 2 nm smaller than the one for FAK-MEM. Also the $d_{\text{F2-C}}$ seems to be populated more at larger values. Nevertheless all these changes are very small.