

# 1 Introduction

Focal adhesions (FA) are macromolecular protein complexes which act as a connection hub between the cell, especially the cytoskeleton, and the extracellular matrix (ECM). They enable the cell to exert tension forces, but can also transduce mechanical stimuli from the ECM to the inside of the cell and integrate this information. One important protein associated to FAs 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 often an overexpression of FAK can be found in cancer cells and understanding the activation processes and dynamics of FAK could give rise to new cancer treatments [39].

## 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 (Figure 1.1).

FERM (4.1 protein, ezrin, radixin and moesin) is a common protein domain which links proteins to membranes [10] and consists of three subdomains: F1, F2 and F3. In the F2 subdomain a basic patch ( $^{216}\text{KAKTLRK}^{222}$ ) can be found, which is a prominent binding site for phosphatidylinositol-4,5-bisphosphate ( $\text{PI}(4,5)\text{P}_2$ ).

The kinase consists of a C-lobe, an activation loop and a N-lobe. Catalytic activity of the kinase is mainly regulated by the phosphorylation of  $\text{Y}^{576}$  and  $\text{Y}^{577}$ , which are located in the activation loop [8]. The kinase also provides binding sites for  $\text{PI}(4,5)\text{P}_2$ . One is located next to the basic patch of the FERM domain, but others (namely  $\text{R}^{508}$ ,  $\text{R}^{514}$ ,  $\text{K}^{515}$ ,  $\text{K}^{621}$  and  $\text{K}^{627}$ ) can be found on the side 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,  $\text{Y}^{397}$ , 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 links to FAs by interacting with talin and paxillin [2].

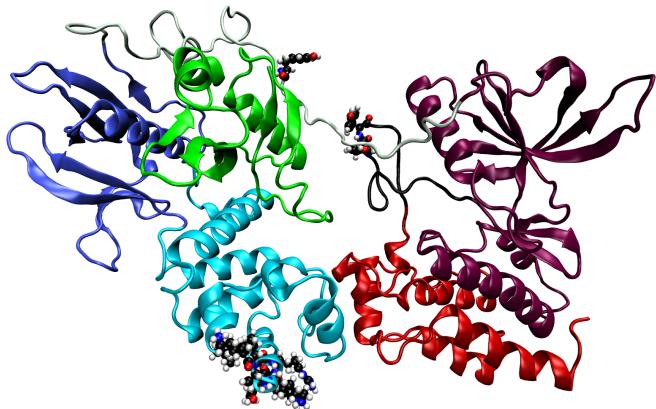


Figure 1.1: **Structure of FAK**

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

## 1.2 PI(4,5)P<sub>2</sub> binding and activation

It is known that FAK triggers several stimuli. Since in this thesis the focus is on the interactions with PI(4,5)P<sub>2</sub>, only the different binding sites of FAK for PI(4,5)P<sub>2</sub> and their impacts on FAK activation are discussed in the following.

In the autoinhibited conformation the FERM domain shields the active site of the kinase. An activation of FAK therefore requires the dissociation of these domains [26].

PI(4,5)P<sub>2</sub> is a phospholipid, which is locally generated in FAs due to integrin signaling [25]. Its charge depends strongly on the pH, but at physiological conditions a net charge of -4 is the preferred state. However, in presence of the amino acid K the deprotonated state gets promoted resulting in a net charge of -5 [30]. The electrostatic binding of PI(4,5)P<sub>2</sub> 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. Moreover the linker region gets less strongly bound, which promotes autophosphorylation of  $Y^{397}$ . However, the PI(4,5)P<sub>2</sub> binding alone has no effect on the catalytic activity, which suggests that the FERM domain is still associated to the kinase [16, 40]. 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<sub>2</sub>, this kinase has access to  $Y^{576}$  and  $Y^{577}$ . As said, the phosphorylation of these residues makes FAK fully active resulting in dissociation of the FERM domain and the kinase [16].

Mechanical forces can lead to the dissociation of the FERM domain from the kinase as well. Forces acting on 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<sub>2</sub> containing membrane. These forces can lead to activation of FAK. In that way, it is acting as a mechanical sensor [41].

The binding sites on the side of the kinase were identified by computer simulations [13] and have been confirmed recently in experiments [17]. The findings from Hall and Schaller [17] 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<sub>2</sub>, this binding was hypothesized to act as a stabilisation of the active state only [17].

### 1.3 Dimerization, clustering and autophosphorylation

Autophosphorylation of Y<sup>397</sup> is an important event in FAK activation. It has been shown that this happens in intact cells *in trans*, for which a self-association of FAK is required [36].

The FERM domain induces a dimerization of FAK, as it does in other proteins containing a FERM domain as well. The interaction emerges around W<sup>266</sup> 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<sup>266</sup> is also required for autophosphorylation of FAK.

PI(4,5)P<sub>2</sub> is not needed for the dimerization of the FERM domains. However, an enriched FAK concentration is needed to observe FAK dimers in cells, which is the case at FAs. 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 *in trans* autophosphorylation of FAK is promoted by dimerization [22].

Albeit PI(4,5)P<sub>2</sub> is not required for dimerization, it induces clustering of several FAK molecules on the membrane *in vitro* [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 FAs [16].

## 2 Motivation

As described in section 1.3 intermolecular interactions of FAK molecules effect their activation. However, the process is still not understood, especially on an atomistic scale.

In this thesis we used MD simulations with the MARTINI force field to investigate these interactions. MARTINI lacks chemical details, but it is a necessary simplification since systems with several FAK molecules involve a large number of particles. Unfortunately, previous work in the group [3] revealed problems in the use of MARTINI regarding simulations of FAK, which are summarized in the following.

Becker et al. [3] obtained in simulations of a single FAK molecule on a PI(4,5)P<sub>2</sub> containing membrane rapid changes in the inclination of the protein with respect to the membrane. In the following part, this inclination is characterized by the angle  $\beta$  between the z-axis and the vector connecting F1 and F2,  $\vec{d}_F$ .  $\beta$  is given as

$$\cos(\beta) = \frac{\vec{d}_{Fz}}{d_F}, \quad d_F = |\vec{d}_F| \quad (2.1)$$

The distributions of  $\beta$  for different simulations (10  $\mu$ s) are presented in Figure 2.1a. The red line, for example, shows a mean value of 90 deg. The angle changed in less than 50 ns and stayed constant for the remaining simulation time. We refer to this behaviour as a fall of FAK in the following.

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]. First, FAK fell to both sides, which means that the interaction sites for PI(4,5)P<sub>2</sub> proposed by Feng and Mertz [13] were also located on top of the protein instead of at the protein membrane interface. Indeed, contact analysis between the protein and PI(4,5)P<sub>2</sub> lipids showed that virtually all residues on the surface (in both, FERM domain and kinase) were interacting with the membrane. A second reason is, that this behaviour was not observed in equivalent all atom simulations in CHARMM36 (1.5  $\mu$ s 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 we carried out several experiments to understand the cause of this falling. However, we were not able to prove the reason beyond doubt. In order to still perform reasonable simulations of multiple FAK molecules, we decided to apply a stabilizing force on each FAK molecule.

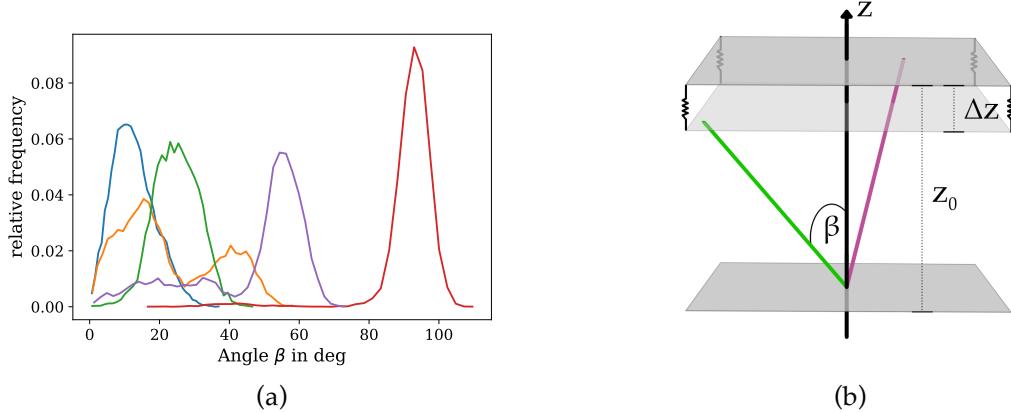


Figure 2.1: Inclination angle of FAK

(a): Distributions of  $\beta$  obtained by Becker et al. [3]. The red curve shows the distribution for a fallen FAK. (b): Illustration of the applied force. Pink and green line represents possible orientations of  $\vec{d}_{F,z}$ . The force is proportional to  $\Delta z$ .

The force is acting onto F1 and F2 parallel to the z-axis and is proportional to the deviation of their z-distance  $\Delta z$  from a reference distance  $z_0$ . This is illustrated in Figure 2.1b. For the determination of  $z_0$  we took only the green and the blue distribution into account, because the large angles observed in the other distributions have not been observed in the CHARMM36 simulations. The mean value of  $\vec{d}_{F,z}$  for these two distributions is 2.228 nm, which we therefore set as  $z_0$ . From the standard deviation of  $\vec{d}_{F,z}$ ,  $\Delta z = 0.11$  nm we estimated the maximum force constant  $k_m$ , which does not sharpen the distribution.

$$k_B T = \frac{1}{2} k_m \Delta z^2 k_m = \frac{2k_B T}{\Delta z^2} = 412 \frac{\text{kJ}}{\text{molnanom}^2} \quad (2.2)$$

We choose  $k = 100 \frac{\text{kJ}}{\text{molnanom}^2}$ . In the following chapter we give a brief overview on the used methods, especially Molecular dynamics and umbrella simulations. Afterwards the different simulation setups are explained.

# 3 Methods

## 3.1 Molecular dynamics simulations

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 specific 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 [31, p. 127f].

Newton's equation of motion can be turned into two first order differential equation

$$\frac{d\vec{r}_i}{dt} = \vec{v}_i \quad (3.1)$$

$$m_i \frac{d\vec{v}_i}{dt} = \vec{F}_i \quad (3.2)$$

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.

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

## 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  $k_{\text{dist}}$ .

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

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  $\theta$  partners from an equilibrium angle  $\theta_0$ . The resulting harmonic potential is usually described by a harmonic oscillator as well.

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

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

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

where  $k_{\text{dihedral}}$  describes the energy barrier for turning the dihedral angle,  $n$  the number of minima in the energy function (multiplicity) and  $\phi_0$  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).

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

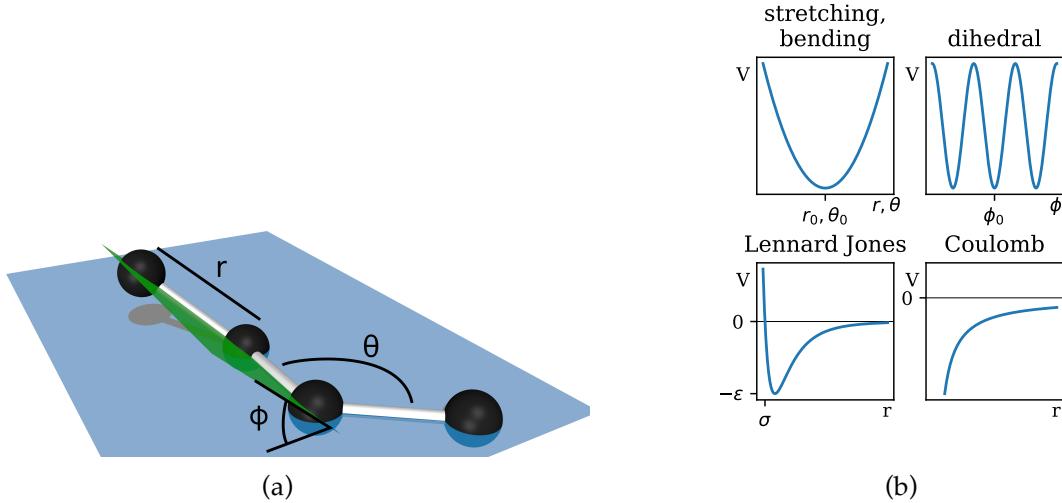


Figure 3.1: **Interactions in MD**

(a): Illustration of stretching  $r$ , bending  $\theta$  and the dihedral angle between blue and green plane  $\Phi$ . (b): schematic course of contributing potentials

$\epsilon$  is related to the potential depth and  $\sigma$  to the potential range (Figure 3.1b).

The Coulomb potential is given by

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

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

In general non-bonded interactions act between all atoms in the system, which imply 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 [31, 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 [31, 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 in time [1, p. 154-159].

### 3.1.2 Force fields

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 two 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 can be distinguished in non-polarizable and polarizable force fields. The latter offer the possibility to model electronic polarization, f.e. with additional shell beads. Orientational and geometric polarisation are also accessible with non-polarizable force fields since they only rely on partial charges [31, p. 215ff]. Both, MARTINI and CHARMM36, are non-polarizable force fields.

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 force fields 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,

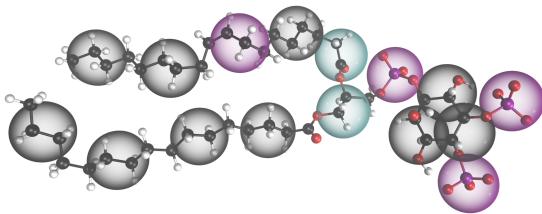


Figure 3.2: **MARTINI bead mapping**

MARTINI structure of PI(4,5)P<sub>2</sub> with underlying atomistic structure. Four heavy atoms are mapped to one MARTINI bead. The colour indicates the bead type.

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.

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 partially charged beads.

### Coarse graining and the MARTINI force field

The MARTINI force field [29, 20, 27] 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, 27].

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 constrained with elastic networks (additional bonds between backbone beads) and can not change during the simulation [28].

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) [28, 29].

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

$$G = H - TS \quad (3.9)$$

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 [28].

Consistently with the four to one mapping, solvent beads in MARTINI represent four water molecules. The MARTINI water has a freezing temperature of  $\approx 300$  K and the freezing process is irreversible. Therefore antifreeze beads were introduced, which have a larger  $\sigma_{\text{LJ}}$  regarding interactions to normal solvent beads. By inserting these antifreeze beads the lattice conformation is disturbed and the freezing temperature reduced [29].

In contrast to water the solvent beads are not (orientational) polarizable. For this reason electrostatic screening has to be treated implicitly, which is done by a 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 [28]. 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 [38].

As reviewed by Marrink and Tieleman [28] 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 constraints

#### 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 [31, 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 [32, 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 useful for equilibration runs or non equilibrium MD simulations as they are stable upon large deviations [5, 1, p. 31].

In comparison NHT extends the Hamiltonian of the system by a friction term representing a heat bath. The friction parameter follows 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 [34, 33] 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 on 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

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  $|E' - E| < dE$  and 0 else. Therefore the partition function  $\Omega$  is given as

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

where  $\mathcal{H}(\mathbf{q}, \mathbf{p}) = U(\mathbf{q}) + K(\mathbf{p})$  is the Hamiltonian and  $C_0$  a proportionality 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  $\beta$ .

$$Q(N, V, T) = \int \exp(-\beta E) \Omega(N, V, E) dE \quad (3.11)$$

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

The configurational integral is defined as

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

It is important to see, that  $\mathcal{H}$  only depends  $\mathbf{p}^2$ , so the integral over the momenta can always be solved analytically 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  $\mathbf{p}$  does not change and therefore

$$\frac{Q_2}{Q_1} = \frac{Z_2}{Z_1} \quad (3.14)$$

holds [9, p. 17].

The Helmholtz free energy  $A$  is defined as

$$A = -\beta \ln(Q) \quad (3.15)$$

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  $\Delta A$  can be written as

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

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  $\xi$  is given as

$$A(\xi) = -\beta \ln(\rho(\xi)) \quad (3.17)$$

$\rho(\xi)$  is the probability distribution of  $\xi$  and can be easily obtained in MD simulation by counting the number of the appropriate states. Therefore it is also referred to as histogram.  $\xi$  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 with a realistic computational cost.

One way to overcome this sampling problem is umbrella sampling [35]. In this approach the path  $\xi$  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  $\xi$  around  $\xi_i$ . This changes the potential energy to

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

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].

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

Of course  $\Delta A_i$  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  $\omega_i$  are chosen iteratively in a way, that the overall statistical error is minimized.

### 3.3 Contact Analysis

In the following part the quantities and tools used for contact analysis are briefly introduced.

Intramolecular distances have been analysed with ConAn. This analysis tool measures inter-residue distances and performs statistical analysis on them. ConAn is still under development and not published yet.

The contact area of the FERM kinase interface can be determined with the solvent accessible surface area (SASA) of the involving domains

$$CA = \frac{1}{2} (SASA_{\text{FERM}} + SASA_{\text{kinase}} - SASA_{\text{FERM-kinase}}) \quad (3.20)$$

For the calculation of SASA values GROMACS sasa tool was used. The v.d.W radii were adapted to MARTINI beads.

For an appropriate description of intermolecular interactions between FAK molecules we defined the following terms.

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

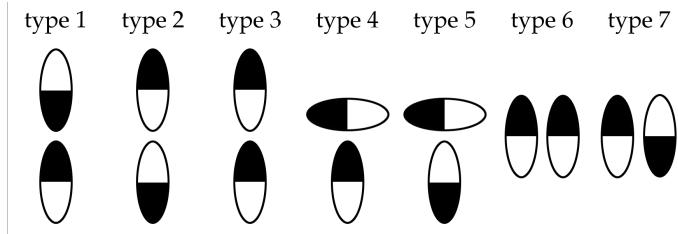


Figure 3.3: **Different two-molecule interaction types**

The black part refers to the FERM domain, the white to the kinase.

**Neighbour** Two proteins are neighbours, if they are interacting with each other. One protein can have several neighbours. For a more detailed characterisation we defined the following neighbour types:

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

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

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

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

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

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

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

**Cluster** Neighbouring proteins form a cluster. A protein belongs to a cluster, if it has at least one neighbour inside the cluster. One protein can only belong to one cluster. The clustersize is the number of proteins belonging to the 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. [40], 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 [26]).

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  $830 \text{ kJmol}^{-1}\text{nm}^{-2}$ . 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  $20\mu\text{s}$  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<sub>2</sub> 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<sub>2</sub>. 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 [37]. 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<sub>2</sub> 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<sub>2</sub> (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

In this chapter we present our results from the simulations. First of all FAK in solution is considered. Afterwards the focus is on the effect of PI(4,5)P<sub>2</sub>. We will start this section with the free energy profile of the basic patch. Since the simulations containing a membrane were carried out with a stabilizing force, we will examine the impact of the force on the observables, before we investigate the conformational changes of FAK induced by binding to the membrane. At the end the results obtained for multiple FAK molecules are presented.

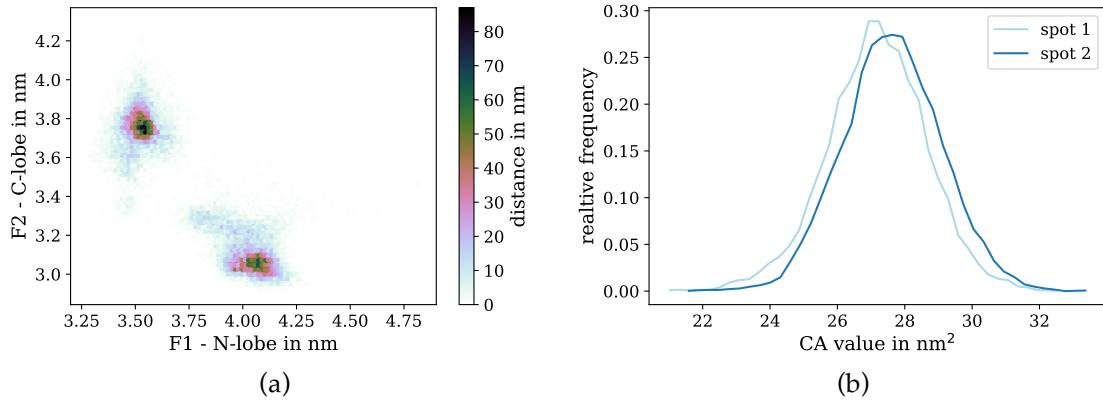
## 5.1 FAK in solution

In this section we investigate the conformation of soluted FAK (FAK-SOL) for the MARTINI force field. Since the secondary structure of the two domains is fixed due to the elastic network, the focus is on the FERM kinase interface.

First the COM distances of F1 to the N-lobe ( $d_{F1-N}$ ) and F2 to the C-lobe ( $d_{F2-C}$ ) 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. The corresponding 3D structures are shown in ???. They induce that spot 1 refers to a configuration, in which the kinase is slightly tilted against the FERM domain while it is in line for configurations associated with spot 2. We obtained several transitions between the spots during the simulation, which indicates a sufficiently long simulation time. At the end 47.4% of the obtained distances were located in spot 1 and 52.6% in spot 2.

However, there is only a minor effect upon the contact area (Figure 5.1b). Spot 2 show a slightly larger mean value of 27.6 nm<sup>2</sup> in comparison to spot 1 (27.1 nm<sup>2</sup>).

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



**Figure 5.1: Domain distances and contact area of FAK-SOL**

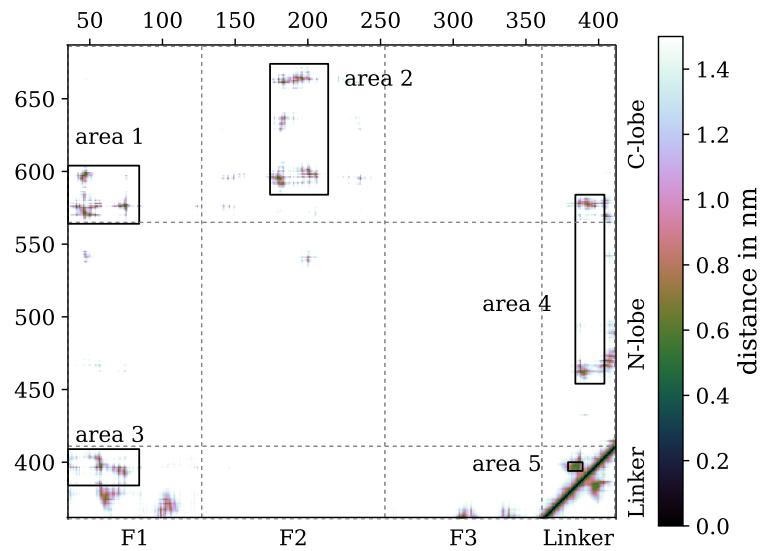
(a): two dimensional hexbinning histogram of  $d_{F1-N}$  and  $d_{F2-C}$ , which shows the two spots. (b): distribution of CA for the two spots.

$Y^{576}$  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.

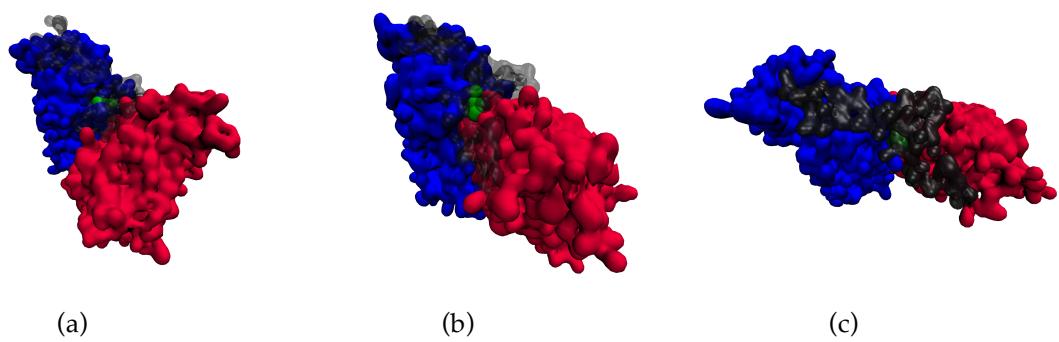
The second contact area (area 2) is located between F2 and the C-lobe. The spots occur around the residues  $Y^{180}$  and  $D^{200}$  of F2 as well as  $F^{596}$  and  $R^{665}$  of the C-lobe. The minimal distance in this area occurs between  $Y^{180}$  and  $F^{596}$  with 0.45 nm and an RMSF value of 0.02 nm. Mutation experiments showed, that these two residues have an important effect upon the interface [26], which fits to the obtained contact.

The linker shows 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). Furthermore several interactions occur between the residues in the linker itself, namely between  $S^{379}$  to  $V^{389}$  and  $T^{394}$  to  $I^{400}$  (area 5). Also in this spot the minimal distance occurs between  $Y^{397}$  and  $T^{386}$  (0.52 nm, RMSF 0.04 nm). The density in this area induces that the region forms a ball, which is slightly plunged into the interface of the FERM domain and the kinase (regarding area 3 and area 4). These contacts support the thesis that autophosphorylation is prevented in closed conformation by a binding of the linker to the FERM domain [40].

The contact map for configurations of spot 1 shows similar features as obtained for spot 2, except for area 2, in which less contacts occur.



**Figure 5.2: Contactmap of FAK-SOL**  
Contactmap of the FERM-kinase interface and the linker region.



**Figure 5.3: 3D structure of FAK in the different states**  
The 3D structures show the FERM domain (blue), the kinase (red) and the linker (black, transparent). The autophosphorylation site Y<sup>397</sup> (green) is plunged into interface. (a): configuration of spot 1. The kinase is tilted against the domain. (b and c): configuration of spot 2. Both domains are in line.

## 5.2 Free energy profile of basic patch

### 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 binding of the basic patch to PI(4,5)P<sub>2</sub> 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<sub>2</sub> 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 5.4](#). 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 more shallow wells 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.

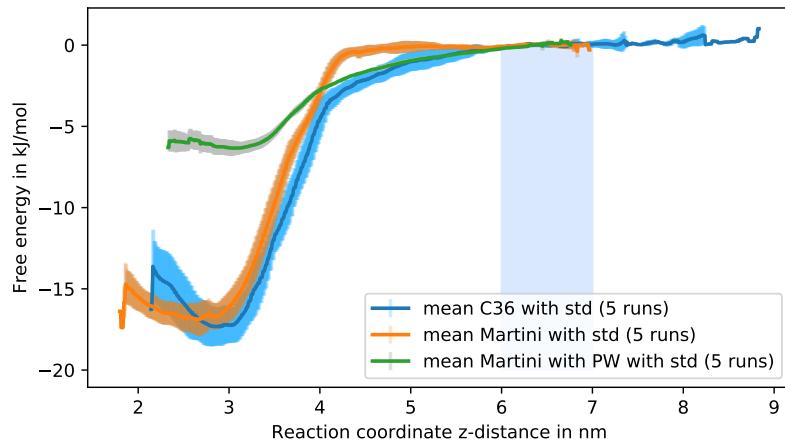
The extent to which MARTINI reproduces the results from all-atom simulations is remarkable, even though the parameters for MARTINI were obtained from free energy calculations (see [section 3.1.2](#))

In the following starting configurations, the proteins are already bound to PI(4,5)P<sub>2</sub>. 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 we consider only the standard water model in the following simulations.

## 5.3 Impact of stabilizing force

### 5.3.1 Impacts of the stabilizing force

In setup 3 and setup 4 we stabilized each protein with an external force acting on the FERM domain (see [chapter 2](#)). To ensure that this restraining does not cause major artefacts the dependency of the used observables is examined below.



**Figure 5.4: Free energy profile of basic patch**

For each forcefield the average and standard deviation of the five copies is presented.

The force has a mean value of  $1.47 \text{ kJ/mol}^{-1}\text{nm}^{-2}$  and a standard deviation of  $13.13 \text{ kJ/mol}^{-1}\text{nm}^{-2}$ . It is skewed to positive values. This is expected since positive values of the force require negative elongations  $\Delta z$ . If the connecting vector of F1 and F2  $\vec{d}_F$  is parallel to the z-axis the maximal negative elongation is limited by the length of  $\vec{d}_F$ . Therefore the force would have to stretch the distance between F1 and F2.

Linear regressions show that all of the quantities  $d_F$ ,  $d_{\text{F1-N}}$ ,  $d_{\text{F2-C}}$  and CA have a negligible correlation to the applied force. Here negligible means that either the regression result was not significant or that the obtained slope was so small, that a change of one standard deviations in force would not change the quantity noticeably.

We tested also the inter-residue distances for correlation with the applied force. To this end, 10 different proteins without neighbours were picked, each for  $1 \mu\text{s}$ . For each residue pair in this dataset we performed a linear regression. Figure 5.5 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 influence residue pairs contributing to the interface. However, we do not expect major changes in the contact regions, since the majority of residue pairs show only weak correlations.

Summarizing we do not expect large perturbations due to the force. However, there is the possibility to miss configurational states, f.e. of binding poses which require a large inclination of FAK. Therefore investigations of the cause for the falling and alternatives to our approach should be also addressed in further studies.

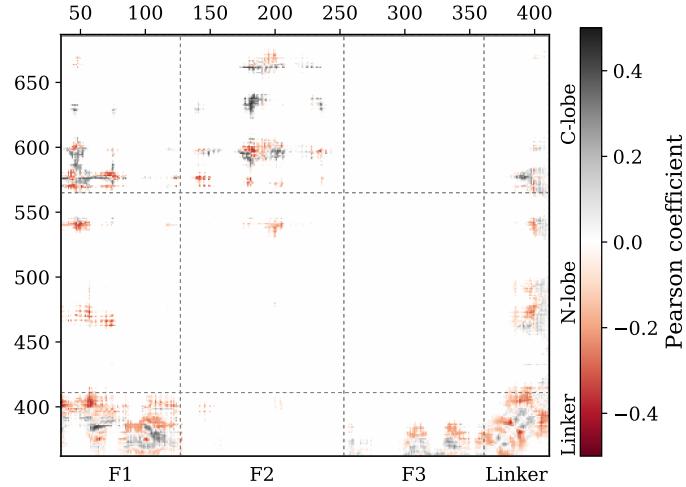


Figure 5.5: **Correlations in contact map**

The contact map shows residue pairs, whose distance correlates with the applied force. The obtained slope is about 20 pN/nm.

## 5.4 Conformational changes on a membrane

### 5.4.1 Conformational changes

In this section the conformation of FAK bound to a PI(4,5)P<sub>2</sub> containing membrane (FAK-MEM) is compared to the observations of FAK-SOL. To this end FAK structures of setup 4 were used with the condition that the FAK molecules have no neighbours. The contact map is based on the same dataset, which was used for [Figure 5.5](#).

Regarding  $d_{F2-C}$  a closure of F2 and the C-lobe can be obtained. Large values up to 3.8 nm, such as in spot 1 of FAK-SOL, disappear ([Figure 5.6a](#)). Since both, FERM domain and kinase, have a docking site for PI(4,5)P<sub>2</sub> which binds them to the membrane, this stabilisation is reasonable. In contrast the distribution of  $d_{F2-C}$  induces that a partial opening of the interface between F1 and the N-lobe gets more preferred. As explained in [section 5.1](#) the linker is located in this region for FAK-SOL.

The closure of F2 and the C-lobe influences the contact area as well. In comparison to FAK-SOL, it is increased to 30.8 nm<sup>2</sup> ([Figure 5.6b](#)).

A more detailed insight can be obtained from the contact map. For an easier comparison, the difference between the contact map of FAK-SOL and FAK-MEM is shown in [Figure 5.7](#). At the interface of F2 and the C-lobe (area 1) the residue distances get closer, which is consistent with the decreasing  $d_{F2-C}$ . In area 2 both, increasing and decreasing distances can be found. Regarding the 3D structure one can see that the residues of the FERM domain,

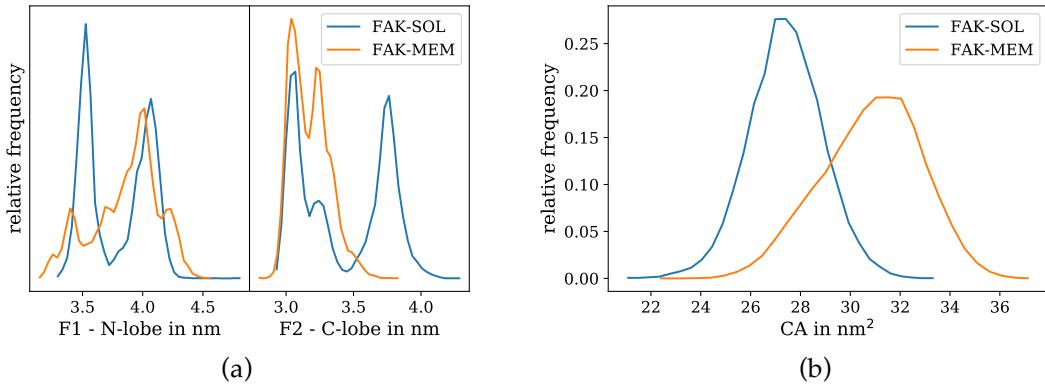


Figure 5.6: **Domain distances and contact area of FAK-MEM**

(a): distribution of  $d_{F1-N}$  and  $d_{F2-C}$  in comparison to FAK-SOL. (b): contact area in comparison to FAK-SOL

which get closer to the kinase, are located nearer to F2 than those getting farther away. In addition all residues of the FERM domain getting farther away from residues S<sup>574</sup> to A<sup>579</sup> of the kinase, which also include the activity regulating residues Y<sup>576</sup> and Y<sup>577</sup>.

Also in the linker region conformational changes can be observed. Area 3 indicates that the ball structure of the Y<sup>397</sup> containing ball structure observed in FAK-SOL disappears. As shown in ?? the residues unravel and cling to the FERM domain. By this the autophosphorylation site Y<sup>397</sup> gets more exposed.

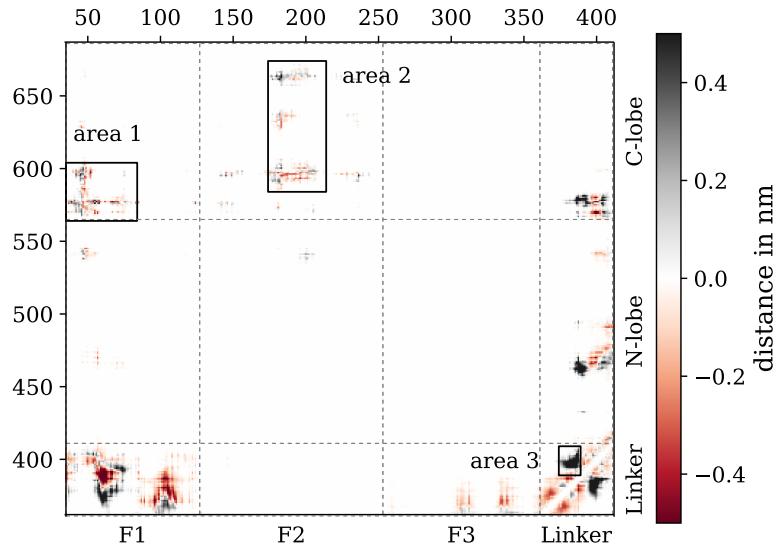
As suggested by experimental results the FERM domain does not dissociate from the kinase because of binding to PI(4,5)P<sub>2</sub>. However this binding leads to conformational changes resulting in a partial opening of the FERM kinase interface and the exposure of the autophosphorylation site Y<sup>397</sup> [16, 40].

## 5.5 Multiple FAK interactions

The focus in this section is on interactions occurring between multiple FAK molecules, which are obtained from setup 4. At this point we want to remind the reader 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 have a significant effect on clustering processes.

### 5.5.1 Dimerization of FAK

First we want to investigate interactions between two FAK molecules. For this purpose we filtered the dataset for interactions, in which both partners have exactly one neighbour. It turns out that FERM-kinase dimers occur the most followed by FERM-FERM dimers.



**Figure 5.7: Difference contact map of FAK-MEM**

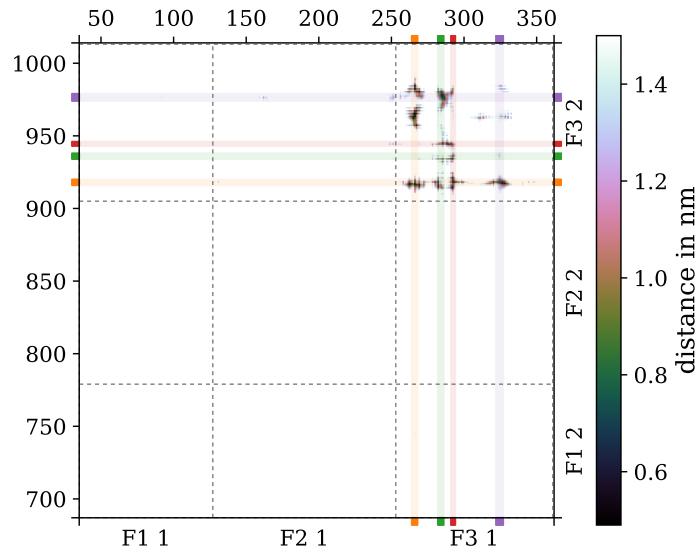
The contact map shows the difference FAK-MEM - FAK-SOL in the distances.

Because the latter have been experimentally investigated by Brami-Cherrier et al. [7] a short comparison between the obtained results and the experimental findings is given. For this purpose we chose the dimer with the longest lasting time as a representative example.

The contact map of the interaction surface is shown in Figure 5.8. From this the importance of residue W<sup>266</sup> can be confirmed (orange region). Beside this also the residues Y<sup>282</sup> to K<sup>286</sup> (green region), T<sup>291</sup> to A<sup>394</sup> (red region) and G<sup>322</sup> to L<sup>327</sup> (violet region) appear as important actors at the interaction surface since they interact with W<sup>266</sup> and among each other. However, the RMSF value of these contacts is larger than for the contacts involving W<sup>266</sup>.

At this point it has to be said that other FERM-FERM dimers in the simulation show also contacts between F1 and F3. Since they are not visible in other samples (e.g. the chosen one), they are not presented here.

FERM-FERM dimerization affects the FERM kinase interface of the involved proteins. In comparison to FAK-MEM almost all distances of the interacting residues gets closer for both protein instances (about 0.2 nm). A reason for this could be that the connection of the FERM domains require an uplift of F3 since the interaction interface points towards the membrane in FAK-MEM. Because the FERM domain is anchored on the membrane via the basic patch, this would push the kinase partly into the membrane. Therefore a force is acting on the kinase, which could be the reason of the closure.



**Figure 5.8: Contactmap of FERM-FERM dimers**

The contact map shows the interaction site. The most important residues are highlighted.

### 5.5.2 FAK clusters

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 agglomerations (see ??).

First of all we focus on the occurring neighbouring types, which were introduced in section 3.3. In Figure 5.9a the time evolution of the average number of encounters of the different interaction types is determined. It shows, that FERM-kinase interactions (type 3) occur the most, while type 1, 2, 4 and 5 occur equally often. Type 6 and type 7 are interactions, in which all four domains are involved. They occur less often, especially type 7, which is the asymmetric one.

We also analysed the altering of neighbour types, which is presented in Figure 5.9b. This indicates that the first contact of two FAK molecules appear at the end of their long axis. Some of these in line neighbours alter than to type 4 and 5, but also backwards. Also type 6 neighbours are only formed out of already interacting proteins.

Lastly the mean number of neighbours is considered. It turns out, that there is a fast rising in the beginning and a flattening after 6 ns. The average over the five copies is at the end of the simulation 1.86 neighbours.

From these observations one could draw the conclusion, that the preferred arrangement of FAK molecules are chains along their long axis. The FAK molecules inside the chain

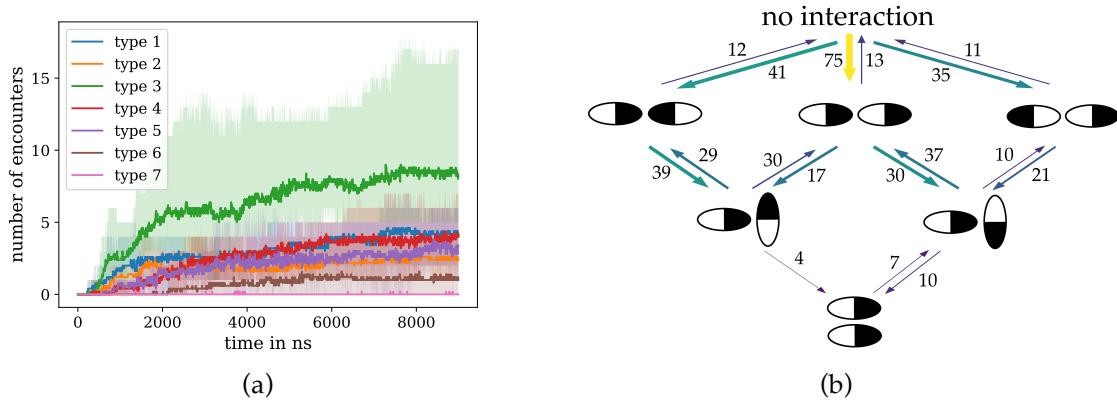


Figure 5.9: Evolution of neighbouring types

(a): Mean number of encounters over the five copies for the different neighbouring types. The filled area shows the range from minimum to maximum. (b): Transition diagram of the neighbouring types. Width and colour of the arrows correspond to the total number of observed transitions, which is also given beside the arrows.

interact either as type 3 (FK chains) neighbours or as type 1 and type 2 neighbours (FFKK chains). Of course also combinations of these two are possible.

### 5.5.3 Activation due to clustering

Lastly, the impacts of clustering on FAK activation are addressed. Activation means here the dissociation of the FERM domain from the kinase, therefore we analysed our data with respect to the contact area (CA) of the FERM-kinase interface.

Unfortunately in no FAK molecule did a full dissociation take place at any time. Because only a small number of FAK molecules were in the system and because the clustering process is not finished at the end of the simulations, this does not mean, that activation due to clustering would not be possible at all. In the hope of seeing trends regarding to activation a more detailed analysis of the CA for the different proteins was performed.

At first glance CA seems to be independent of the number of neighbours (??), the interaction type and the clustersize.

Motivated from subsection 5.5.2 we focus on FAK molecules inside chains in the following. 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 chains only type 3 interactions were allowed, for FFKK chains both, type 1 and type 2. We compare the results to FAK-MEM in Figure 5.10 for FK chains and in Figure 5.11 for FFKK chains.

Figure 5.10 indicates that an arrangement of FAK molecules to FK chains does not influence the mean value of CA. The distribution gets slightly sharper compared to FAK-MEM, which could imply a stabilisation of the interface. Not even the distributions of  $d_{F1-N}$  and  $d_{F2-C}$  differs from those obtained in FAK-MEM.

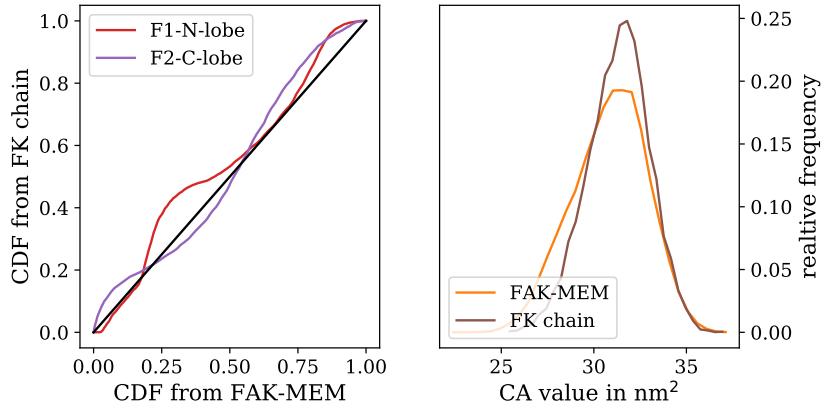


Figure 5.10: **Domain distances and contact area of FK chains**

(left): Q-Q plot of  $d_{F1-N}$  and  $d_{F2-C}$  distributions in comparison to FAK-MEM. The distributions are quite similar. (right): CA value in comparison to FAK-MEM.

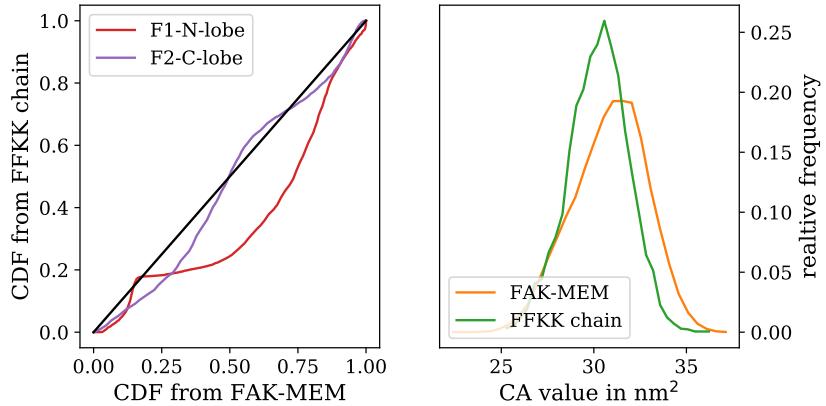


Figure 5.11: **Domain distances and contact area of FFKK chains**

(left): Q-Q plot of  $d_{F1-N}$  and  $d_{F2-C}$  distributions in comparison to FAK-MEM. FFKK chain members tend to larger values of  $d_{F1-N}$ . (right): CA value in comparison to FAK-MEM.

Also an arrangement to FFKK chains affects the value of CA hardly (Figure 5.11). The distribution gets slightly sharper and shifted to smaller values by 0.6 nm. However, the distribution of  $d_{F1-N}$  indicates a trend to larger values in comparison to FAK-MEM, which refers to an more open state of FAK.

The filtering leads of course to a smaller sampling of the distributions of  $d_{F1-N}$ ,  $d_{F2-C}$  and CA . E.g. for FFKK chains only 10 different FAK instances were taken into account (21.7  $\mu$ s simulation time in total). However, the obtained changes are visible in almost all of the 10 instances.

## 6 Conclusion

The aim of this work was to gain insights into multiple FAK interactions and their impacts on conformational changes using MD. Because these systems are very large the coarse graining force field MARTINI was used.

In order to avoid falling of FAK on the membrane we introduced a stabilizing force, which acts on the FERM domain of the FAK molecules. In subsection 5.3.1 we showed that this approach doesn't influence the observables used in the remaining part. However, it has important limitations. The force does not only prevent tiltings around the long axis of FAK (falling to side), but also around the short axis, which happens f.e. in FERM-FERM dimerization. Another problem is the absolute reference to the z axis. A reference to the membrane would be better to allow tiltings arising from membrane curvature.

From the configurations of FAK in solution we identified important residues contributing to the FERM kinase interface. The observations fit well with experimental studies of Lietha et al. [26]. Also the burying of the active site of the kinase as well as the hiding of the autophosphorylation site was observed in these simulations.

We recognized configurational changes of FAK when it binds to a PI(4,5)P<sub>2</sub> containing membrane, namely a partial opening, the promotion of the autophosphorylation site and that the domains stays associated. These changes are consistent with previous studies on the influence of PI(4,5)P<sub>2</sub> binding to FAK by Goñi et al. [16] and Zhou et al. [40]. Also the free energy profile of the PI(4,5)P<sub>2</sub> binding site in the FERM domain we obtained from MARTINI simulations samples the equivalent profile from CHARMM36 simulations in the range of the statistical error.

Albeit MARTINI is a coarse graining force field, even atomistic details were reproduced, which also show the power of MARTINI and confirms it as a suitable tool regarding FAK simulations.

In section 5.5 we investigated the interactions between multiple FAK molecules on a membrane. We interpreted the observations as a tendency to a chain like arrangement of the FAK molecules along their long axis and tried to estimate the impacts of such arrangements on the configurations of the involved FAK molecules. Our results show only a small impact on the quantities associated with FAK activation. Therefore we can not draw the conclusion, that these arrangements could promote activation of FAK in larger clusters.

Currently we try to cluster the obtained configurations with more general approaches in order to reveal the parameters inducing configurational changes in multiple FAK interactions. However, arrangement of large molecules such as FAK is a time consuming process, which has not come to an end in our simulations. Longer simulation times could therefore give new insights into the consequences of FAK clustering.