# 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.
FERM (4.1 protein, ezrin, radixin and moesin) is a common protein domain which targets proteins to membranes [4] 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 [3]. 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 [5, 7].

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 [6].
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 [1].

### 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 [9].
PI(4,5)$P_2$ is a phospholipid, which is locally generated in FA due to integrin signaling [**CITATION34**]. It has a net charge of -4, but in presence of K the deprotonated state gets promoted resulting in a net charge of -5 [10]. 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 [6, 12]. 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 [6].

Another stimulus leading to the dissociation of the FERM domain from the kinase is mechanical 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 [13].

The binding sites on the larger surface of the kinase were hypothesized by computer simulations [5] and have been confirmed recently in experiments [7]. 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 [7].

### 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 [2].
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) [8].

Although PI(4,5)$P_2$ is not required for dimerization, it induces clustering of several FAK *in vitro* on membranes [6]. 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 [8] and may play an important role in the scaffolding function of FAK for FA [6].

# 2 Motivation

As described in **??** 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 **??**) 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 [**sara**] 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 $\beta$ between the z-axis and the vector connecting F1 and F2, $\vec{d}_F$.

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

**sara** simulated five different copies for $10\,\mu$s each. The resulting distributions of the angle can be seen in **??**. 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 [5]. 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 [5] 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 ($1.5\,\mu$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 several efforts 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 **??**. However the reason could not have proven beyond doubt. In order 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 $\Delta z$ from a reference distance $z_0$. An illustration of the force can be found in **??**. For the determination of $z_0$ only the green and the blue distribution from **??** 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 $2.228\,\text{nm}$, 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 MD

## 3.2 Free energy calculations

# 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. [12], Martini forcefield: **sara**). These configurations contain only a FERM-kinase fragment without the FAT domain and its linker (residues 35 to 686, PDB 2J0J [9]).
As explained in **??** the secondary structure of proteins have to be stabilized in Martini using elastic networks. This was set up by **sara**. 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 [**PW**, 11]. 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 [**gromacsWHAMpaper**]. 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 **sara** 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 **sara** 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 FAK in absence of PI(4,5)$P_2$ and other FAK molecules is analyzed. For this purpose the simulation data of setup 1 is used.

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. In Figure 5.1 a hexagonal binning plot of both values can be found, which indicates, that there are two different states. Spot 1 shows large values of $3.8\,\text{nm}$ for $d_{\text{F2-C}}$, while $d_{\text{F1-N}} \approx 3.5\,\text{nm}$. In contrast to this the values of $d_{\text{F2-C}}$ are much smaller in spot 2 and $d_{\text{F1-N}}$ gets larger. Certainly these spots can not be seen in the contact area of the interface (see Figure 5.6b), which is symmetrically distributed around $27.5\,\text{nm}^2$. 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 interface. The first one (area 1) is located between F1 and the N-lobe/activation loop. It shows i.e. contacts between $Y^{576}$ and $Y^{577}$ and residues of the FERM domain. The minimal distance in this area, between residue $H^{41}$ and $Y^{576}$, is $0.45\,\text{nm}$ with an RMSF value of $0.03\,\text{nm}$. This 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 $R^{665}$. The minimal distance in this area occurs between $Y^{180}$ and $F^{596}$ with $0.45\,\text{nm}$ and an RMSF value of $0.02\,\text{nm}$. These two residues have been reported as important actors in the interface, because a mutation disturbs the interface and enhances the activation of FAK.
The linker show contacts with both domains. Interestingly the minimal distance in the marked areas (area 3 and area 4) occur between the autophosphorylation site $Y^{397}$ and $H^{58}$ ($0.45\,\text{nm}$, RMSF $0.03\,\text{nm}$) in F1 as well as $Y^{576}$ ($0.50\,\text{nm}$, RMSF $0.10\,\text{nm}$) in the kinase. This is consistent with the concept, that autophosphorylation is prevented in closed conformation by a binding of the linker to the FERM

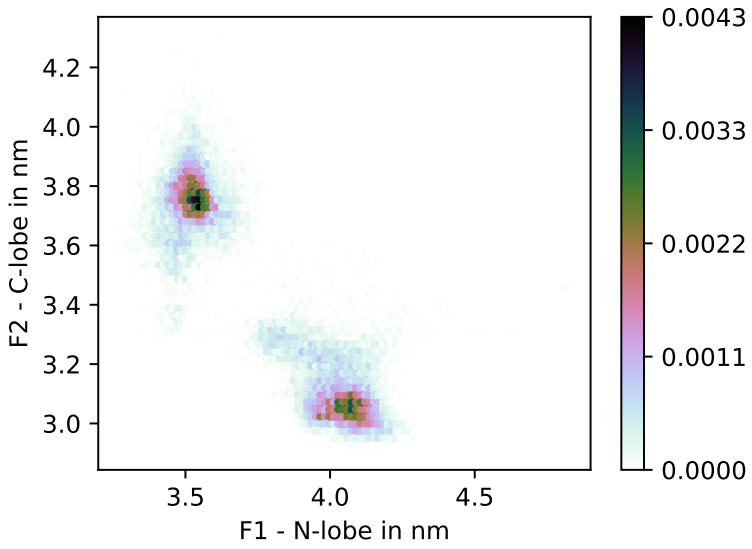



Figure 5.1: Two dimensional histogram of $d_{\text{F1-N}}$ and $d_{\text{F2-C}}$. The histogram is normed and the color indicates the relative frequency.

domain. In contrast to Figure 5.2, the contact map for frames of spot 1 show less contacts between F2 and the C-lobe, i.e. around the mentioned residues $Y^{180}$ to $M^{183}$. A few additional contacts appear between F1 and the N-lobe, but these are only minor spots.

## 5.2 FAK binding to PI(4,5)$P_2$

### 5.2.1 Free energy profile of basic patch

In the context of the investigation of FAK falling the free energy profile of the basic patch in F2 was investigated. Because this might be of interest for other applications as well, a short report on the obtained results is given at this point. The used setup is setup 2 described in section 4.3.
The average profile of the five copies together with the standard deviation for the forcefields C36, Martini and Martini with PW can be found in Figure 5.3. Here the range between $z = 6\,\text{nm}$ and $z = 7\,\text{nm}$ was defined as zero point.
C36 and Martini show both a similar energy depth of $A \approx 17\text{kJ/mol}$ and the same slope between $z = 3\,\text{nm}$ and $z = 4\,\text{nm}$. Certainly Martini results in larger values for the free energy, which could originate from an underestimation of electrostatic forces due to the unpolar water beads (see **??**). For values above $z = 4\,\text{nm}$ the free

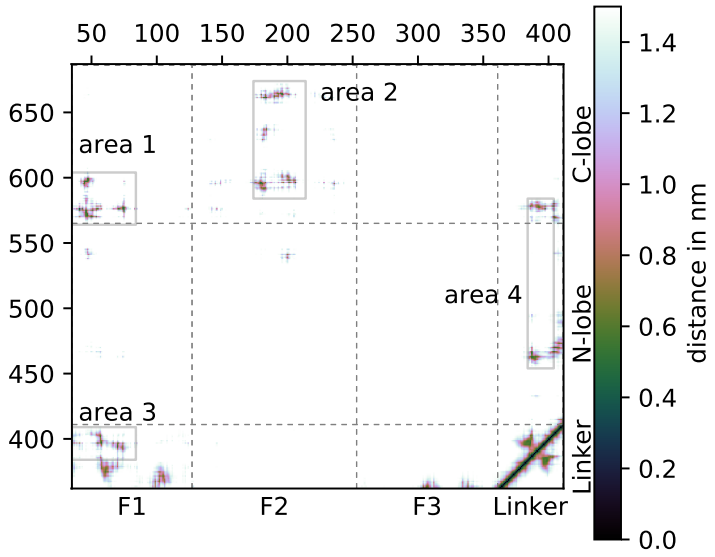



Figure 5.2: Contactmap of interface between FERM domain and kinase

energy profile flattens out in C36 while a kink can be observed in Martini. This is due to 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.
It is remarkable, in which extent the Martini force field redraws the results from all-atom simulations. A contribution to this might be, that the parameters for Martini were obtained a.o. from free energy calculations and measurements (see **??**).

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 further distances. In addition Martini with PW showed a much higher computational cost. That is why in the following simulations only Martini has been considered.

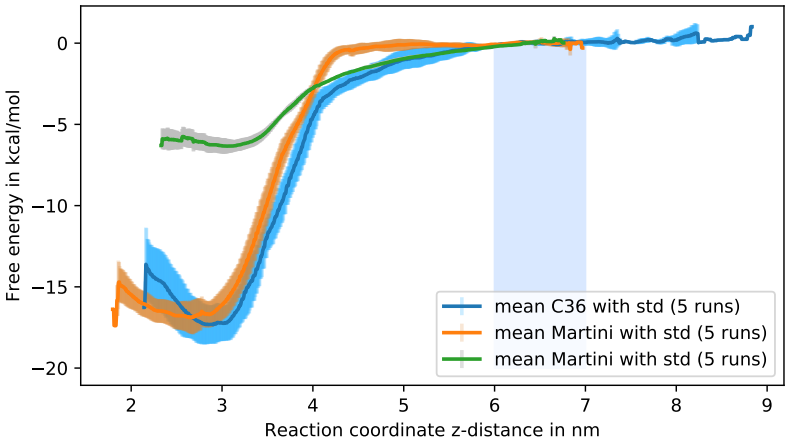



Figure 5.3: Free energy profiles of basic patch in different force fields

### 5.2.2 Impacts of the applied force

In setup 3 and setup 4 each protein was stabilized with an external force acting on the FERM domain (see chapter 2). These forces have of course impacts on the FAK molecules, which are analysed below.
The force itself has a mean value of $1.47\,\mathrm{kJ/mol/nm}$ and a standard deviation of $13.13\,\mathrm{kJ/mol/nm}$. It is skewed to positive values, which was expected as deviations with a negative sign are restricted due to $z_0 + \Delta z \leq d_F$ (see below). Neither additional maxima nor minima are observable in its distribution.

#### Distance between F1 and F2

First the COM distance of F1 and F2, namely $d_F$, is considered. In Figure 5.4 a hexbinning plot of $d_F$ against the force can be found. The plot shows, that some combinations (below the orange line) are never visited. For a given force, $d_F$ seems to have a lower boundary. This boundary is defined by $d_{F,\mathrm{min}} = z_0 - \Delta z$, where $z_0$ is the reference distance of the force and $\Delta z$ the elongation (proportional to the force), and is reached, if the vector between F1 and F2, $\vec{d_F}$, is parallel to the z-axis. The plot indicates, that $d_F$ and the force are hardly correlated, which is tested with a linear regression. In order to not bias the regression by $d_{F,\mathrm{min}}$ only points with a force lower than $0\,\mathrm{kJ/mol/nm}$ are taken into account. With a p-Value of $23\%$ for the H0 hypothesis slope $m = 0$, the distance $d_F$ and the applied force can indeed be seen as uncorrelated.

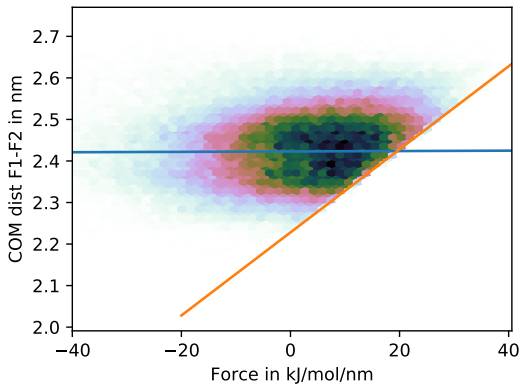



Figure 5.4: $d_F$ against the force. The orange line defines the minimum value of $d_F$. The blue line shows the result of the linear regression.

### Intramolecular distances

First all distances between interacting residues are tested on correlation with the applied force. For this 10 different proteins without neighbours were picked out of the trajectories from setup 4, each for $1\,\mu$s. In Figure 5.5 the calculated Pearson correlation coefficients for significant correlations are shown. There are 270 residue pairs contributing to the interface, which show a large correlation (Pearson $|r| > 0.3$). The calculated slope of these pairs is $\approx \pm 20$force unit/nm. A change of the force of one standard deviation would therefore lead to a change of roughly $0.6$ nm. Furthermore spots exists, which show the same sign in the correlation. Thus the force can have large influences on the interface.
Also the COM distances between F1 and the N-lobe as well as F2 and the C-lobe were tested upon correlation with the external force using a linear regression. Both reveal a significant correlation, but the maximum obtained slope is $0.0008$ nm/force unit. Changes in the distance upon the external force can therefore be neglected.
It is important to see, that in this section only the observed quantities were tested for correlation with the force. It is however possible and probable, that due to the restriction, configurations or states are completely missing while others become over expressed. This effect can not be estimated, but should be kept in mind.

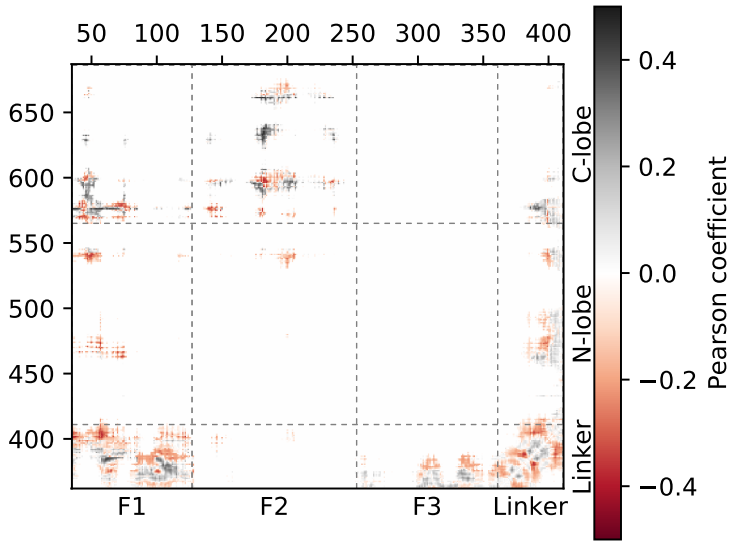


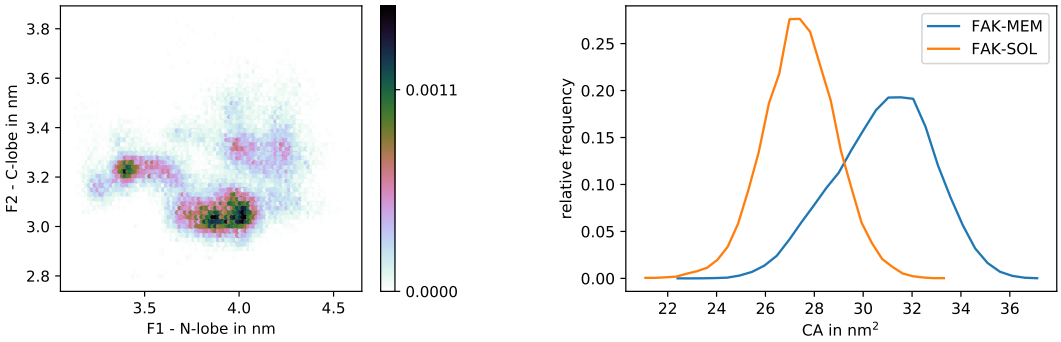

Figure 5.5: Correlation of residue-residue distances upon external force

### 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 than 2 nm away (0 neighbours). The contact map is based on the same dataset, which was used in section 5.2.2.

Analogously to section 5.1 the distribution of the COM distances is presented in Figure 5.6a 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 5.6b, 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.




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

(b) Contactarea of FERM-kinase interface

The obtained results are consistent with previous observations. PI(4,5)$P_2$does 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.

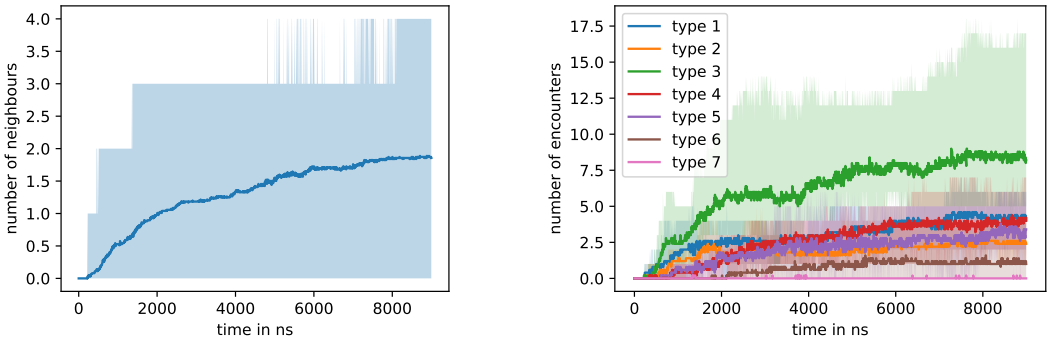
## 5.3 Multiple FAK interactions

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 5.7a a fast rising in the number of neighbours in the beginning and a flattening after




(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.

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 5.7b 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 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

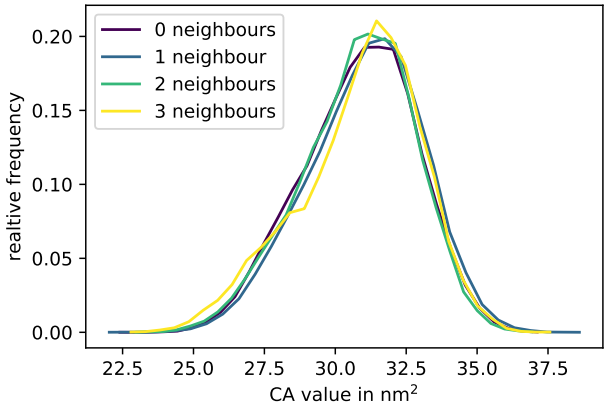



Figure 5.8: CA for different number of neighbours

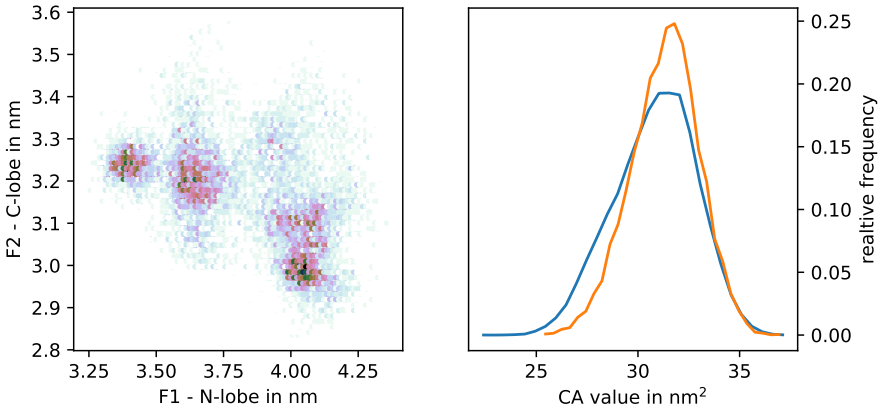



Figure 5.9: Analysis of the FERM-kinase interface in FK-Chains. The blue line was obtained from FAK-SOL.

neighbours Figure 5.8). Therefore the data has to be filtered more.
Motivated from subsection 5.3.1, only FAK molecules inside chains are taken into account. 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 5.9 for FK-Chain and in Figure 5.10 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 5.6a, 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.

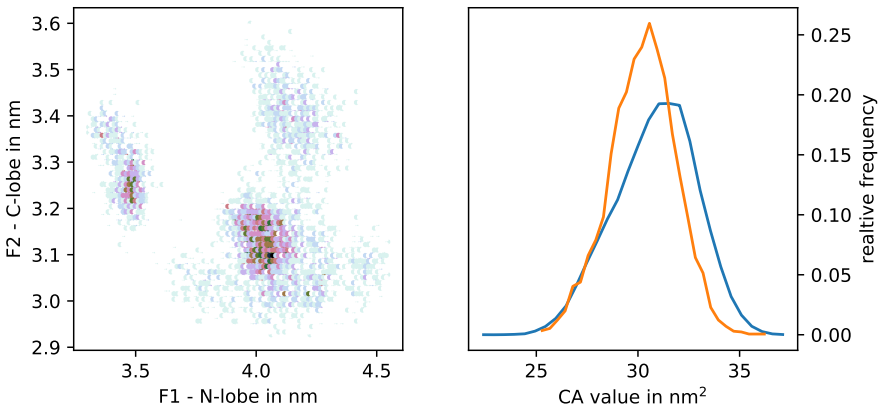



Figure 5.10: Analysis of the FERM-kinase interface in FFKK-Chains. The blue line was obtained from FAK-SOL.