

Rate limiting step of the allosteric activation of the bacterial adhesin FimH investigated by molecular dynamics simulations

Gianluca Interlandi¹

¹University of Washington Department of Bioengineering

January 27, 2023

Abstract

The bacterial adhesin FimH is a model for the study of protein allostery because its structure has been resolved in multiple configurations, including the active and the inactive state. FimH consists of a pilin domain (PD) that anchors it to the rest of the fimbria and an allosterically regulated lectin domain (LD) that binds mannose on the surface of infected cells. Under normal conditions, the two domains are docked to each other and LD binds mannose weakly. However, in the presence of tensile force generated by shear the domains separate and conformational changes propagate across LD resulting in a stronger bond to mannose. Recently, the crystallographic structure of a variant of FimH has been resolved, called FimH^{FocH}, where PD contains 10 mutations near the inter-domain interface. Although the X-ray structures of FimH and FimH^{FocH} are almost identical, experimental evidence shows that FimH^{FocH} is activated even in the absence of shear. Here, molecular dynamics simulations combined with the Jarzynski equality were used to investigate the discrepancy between the crystallographic structures and the functional assays. The results indicate that the free energy barrier of the unbinding process between LD and PD is drastically reduced in FimH^{FocH}. Rupture of an inter-domain hydrogen bond involving R166 constitutes a rate limiting step of the domains separation process and occurs more readily in FimH^{FocH} than FimH. In conclusion, the mutations in FimH^{FocH} shift the equilibrium towards an equal occupancy of bound and unbound states for LD and PD by reducing a rate limiting step.

**Rate limiting step of the allosteric activation
of the bacterial adhesin FimH investigated by
molecular dynamics simulations**

Gianluca Interlandi*

Department of Bioengineering, University of Washington, Seattle, WA
98195, USA

Running title: *Rate limiting step*

*Correspondence to: Gianluca Interlandi, Department of Bioengineering,
University of Washington Box 355061, 3720 15th Ave NE,
Seattle, WA 98195-5061, USA, Tel.: (206) 685-4435
Fax: (206) 685-3300, E-mail: gianluca@u.washington.edu

Abstract

The bacterial adhesin FimH is a model for the study of protein allostery because its structure has been resolved in multiple configurations, including the active and the inactive state. FimH consists of a pilin domain (PD) that anchors it to the rest of the fimbria and an allosterically regulated lectin domain (LD) that binds mannose on the surface of infected cells. Under normal conditions, the two domains are docked to each other and LD binds mannose weakly. However, in the presence of tensile force generated by shear the domains separate and conformational changes propagate across LD resulting in a stronger bond to mannose. Recently, the crystallographic structure of a variant of FimH has been resolved, called FimH^{FocH}, where PD contains 10 mutations near the inter-domain interface. Although the X-ray structures of FimH and FimH^{FocH} are almost identical, experimental evidence shows that FimH^{FocH} is activated even in the absence of shear. Here, molecular dynamics simulations combined with the Jarzinski equality were used to investigate the discrepancy between the crystallographic structures and the functional assays. The results indicate that the free energy barrier of the unbinding process between LD and PD is drastically reduced in FimH^{FocH}. Rupture of an inter-domain hydrogen bond involving R166 constitutes a rate limiting step of the domains separation process and occurs more readily in FimH^{FocH} than FimH. In conclusion, the mutations in FimH^{FocH} shift the equilibrium towards an equal occupancy of bound and unbound states for LD and PD by reducing a rate limiting step.

Keywords: Allosteric regulation, molecular dynamics, pulling simulations, Jarzinski equality, FimH adhesin, bacterial adhesion, biophysics.

Introduction

In allosterically regulated proteins, generally an effector molecule binds to a site of the protein triggering or stabilizing a conformational transition that propagates to a distally located functional site.^{1,2} The bacterial adhesin FimH is a well-studied model for protein allostery. It consists of an allosterically regulated mannose-binding lectin domain (LD) and a pilin domain (PD) that anchors FimH to the rest of the bacterial fimbria.³ When the two domains are docked to each other, LD binds mannose weakly.^{4,5} However, upon separation of the two domains from each other, for example under tensile force generated by shear flow, LD undergoes conformational changes that strengthen its bond to mannose.^{4,5} Thus, LD is predominantly found in either a low affinity state (LAS) or a high affinity state (HAS) for binding mannose. Such an allosteric auto-inhibitory mechanism confers to FimH the property of a so called "catch bond", i.e., tensile force increases the life time of the bond with mannose.⁶ Previous studies by us have indicated a weak correlation between the conformational changes in the inter-domain region of LD and those occurring at the distally located mannose binding site during the conversion between LAS and HAS.^{7,8} Furthermore, both PD and mannose act as effector molecules of the allosteric transition of LD according to a so called "population shift" model.⁸ While docking of PD shifts the conformational equilibrium of LD towards LAS, the presence of mannose increases the likelihood of LD to be found in HAS⁹ although the switch from LAS to HAS is strongly enhanced by the application of tensile force.⁶

Hence, it is plausible that an equilibrium exists even in the absence of shear between a state where LD and PD are bound to each other with LD in LAS and a state where the two domains are separated with LD in HAS. While in the absence of mannose the LD-PD bound state is more predom-

inant, when mannose is present the equilibrium slightly shifts towards the state where LD and PD are separated while tensile force provides sustained activation. Evidence for such an equilibrium shift in the absence of shear has been provided by a recent study by us where wild-type FimH was compared to variants that are more easily activated.¹⁰ In particular, in one of the variants LD of FimH was fused in a recombinant manner to a structurally altered PD where a segment of 16 residues was replaced with a related sequence from FocH, a naturally occurring homologue of FimH with different lectin specificity. Overall, this fusion results in 10 mutations in PD (Figure 1). I refer here to the combination of LD from FimH and PD with the amino-acid sequence replacement from FocH as FimH^{FocH} (in a previous study,¹⁰ it has been termed LD-PD^{FocH}). The X-ray structure of FimH^{FocH} crystallized in the absence of mannose reveals that LD and PD are bound to each other with LD in LAS in a very similar fashion as in wild-type FimH¹⁰ (Figure 1a,b). However, in functional assays FimH^{FocH} was found in the presence of mannose to be activated and displayed HAS-like phenotype even in the absence of shear while wild-type FimH required shear-induced tensile force for sustained activation.¹⁰ The observation that FimH^{FocH} is activated under static conditions while wild-type FimH requires shear despite their crystallographic structures having no significant differences suggests a differential activation mechanism between FimH^{FocH} and FimH. This motivates studying the dynamics of the activation mechanism at atomic level of detail.

A previous study by us found that the mutation A188D in FimH^{FocH}, which is known to be activating,^{11,12} weakens the LD-PD interface by disrupting hydrophobic inter-domain side-chain interactions.¹⁰ However, the mutations in FimH^{FocH} may also affect inter-domain electrostatic interactions, which are thought to be stronger than hydrophobic contacts. Hence,

to thoroughly understand how the FimH^{FocH} mutations alter the LD-PD interface it is necessary to investigate, besides hydrophobic-driven side chain contacts, also electrostatic interactions including intra-PD as the mutations may alter networks of hydrogen bonds and salt bridges.

In this study, I used molecular dynamics (MD) simulations to study how the mutations in FimH^{FocH} alter networks of interactions at the interface between LD and PD. Furthermore, a method combining pulling simulations and the Jarzynski equality¹³ as elaborated in a previous study¹⁴ was used here to investigate free energy barriers in the unbinding process of LD from PD and compare how they differ between FimH and FimH^{FocH}. The overall goal was to provide a mechanistic explanation for the discrepancy between X-ray crystallography where FimH^{FocH} is found in LAS and functional experiments where this variant is activated even in the absence of shear in contrast to wild-type FimH, which requires tensile force for sustained activation. The results improve the understanding about the rate limiting steps of the allosteric transition in a two-domain protein where one of the domains acts as the effector.

Materials and Methods

Initial conformations

The initial coordinates of LAS of FimH and FimH^{FocH} were obtained from the crystallographic structures with PDB codes 3JWN⁵ and 7SZO,¹⁰ respectively. In the X-ray structures, the proteins are incorporated into the entire fimbrial tip of which two identical copies are present in the crystallographic unit. While LD and PD are parts of the same subunit, the preceding domains (such as FimG, FimF, etc.⁴) are connected through a so called β -strand

swapping mechanism. For example, PD is donated a β strand from FimG (Figure 1). For the simulations, chain H (which forms LD and most of PD) and part of chain G (the donated FimG strand) were used for both FimH and FimH^{FocH}. The donated FimG strand was truncated after residue 13 and the resulting C-terminus was capped with N-methylamide in order to neutralize its charge. The initial conformations were minimized with 100 steps of steepest descent in vacuo and 500 steps of conjugate gradient in a dielectric continuum using the program CHARMM.¹⁵

General setup of the simulations

The MD simulations were performed with the program NAMD¹⁶ using the CHARMM all-hydrogen force field (PARAM22)¹⁷ with the CMAP extension^{18,19} and the TIP3P model of water. The different simulation systems are summarized in Table 1. The proteins were inserted into a cubic water box with side length of 115 Å, resulting in a system with in total ca. 146,000 atoms. Chloride and sodium ions were added to neutralize the system and approximate a salt concentration of 150 mM. The water molecules overlapping with the protein or the ions were removed if the distance between the water oxygen and any atom of the protein or any ion was smaller than 3.1 Å. To avoid finite size effects, periodic boundary conditions were applied. After solvation, the system underwent 500 steps of minimization while the coordinates of the heavy atoms of the protein were held fixed and subsequent 500 steps with no restraints. Each simulation was started with different initial random velocities to ensure that different trajectories were sampled whenever starting with the same initial state. Electrostatic interactions were calculated within a cutoff of 10 Å, while long-range electrostatic effects were taken into account by the Particle Mesh Ewald summation method.²⁰ Van der Waals

interactions were treated with the use of a switch function starting at 8 Å and turning off at 10 Å. The dynamics were integrated with a time step of 2 fs. The covalent bonds involving hydrogens were rigidly constrained by means of the SHAKE algorithm with a tolerance of 10^{-8} . Snapshots were saved every 10 ps for trajectory analysis.

Before production runs, harmonic constraints were applied to the positions of all heavy atoms of the protein to equilibrate the system at 300 K during a time length of 0.2 ns. After this equilibration phase, the harmonic constraints were released. The systems were simulated for in total 50 ns, and the first 10 ns of unconstrained simulation time were also considered part of the equilibration and were thus not used for the analysis. During both the equilibration and production phases, the temperature was kept constant at 300 K by using the Langevin thermostat²¹ with a damping coefficient of 1 ps^{-1} , while the pressure was held constant at 1 atm by applying a pressure piston.²²

Determination of persistent contacts

The simulation trajectories obtained in the absence of tensile force were screened for persistent inter-domain side chain contacts, hydrogen bonds and salt bridges. Contacts in the vicinity of R166 were also analysed. A side chain contact was defined to be formed if the distance between the centers of geometry of two side chains was less than or equal 6 Å. To define a hydrogen bond, a H...A distance cutoff of 2.7 Å and a D-H...A angle cutoff of 120° was used, where a donor D could either be an oxygen or a nitrogen, and an acceptor A could be either an oxygen or a nitrogen as long as it is not part of an amino group. An interaction was defined as a salt bridge if the atoms N_ζ of Lys or C_ζ of Arg were closer than 4 Å or 5 Å, respectively, from either

the C_γ of Asp or C_δ of Glu. All histidines were assumed neutral. A contact was considered persistent if it was present in at least 66% of the frames of a particular simulation. Some of the contacts found to be persistent in the simulations with no tensile force were also used to monitor separation of the domains in the pulling runs.

Studying the free energy of unbinding with a method based on the Jarzynski equality

In order to determine the free energy profile using simulations one would have to sample multiple binding and unbinding events between LD and PD. Then, from such an ensemble where FimH or FimH^{FocH} are at equilibrium between the LD-PD bound and LD-PD unbound states it would be theoretically possible to determine the free energy profile along the binding and unbinding pathways. However, sampling such events in atomistic MD simulations would be prohibitive since they would occur in a time scale that is currently not accessible to computer simulations. A viable alternative is to use external forces to accelerate the separation of PD from LD and then calculate the work performed during the stretching. Normally, the work from such a non-equilibrium process would only provide an upper estimate for the free energy of the equilibrium process occurring in the absence of force. However, a method has been developed that combines pulling simulations at constant velocity with the Jarzynski equality¹³ and the second-order cumulant expansion to approximate the free energy profile corresponding to the process occurring at equilibrium with no external forces applied.¹⁴ The free energy profile is then expressed as a function of the internal coordinate that describes the extension of the entire protein. Such a function is called a potential of mean force (PMF) and the coordinate is thought of as a reaction

coordinate.¹⁴

The method is explained in detail in reference.¹⁴ A brief description is presented here with emphasis on the equation used for the calculation of PMF. Jarzynski¹³ discovered the following equality that describes the relationship between an equilibrium free energy difference ΔF and work W done through a non-equilibrium process:

$$e^{-\beta\Delta F} = \langle e^{-\beta W} \rangle, \quad (1)$$

where β is $1/kT$ and $\langle \cdot \rangle$ denotes an average. However, this equality can only be applied to slow processes where the fluctuation of work is comparable to the temperature,¹⁴ but as shown in the results this is not the case for this system. It has been shown that a better estimate of the free energy is achieved when using the second-order cumulant expansion to approximate the logarithm of the right side of equation 1:

$$\log \langle e^{-\beta W} \rangle \approx -\beta \langle W \rangle + \frac{\beta^2}{2} (\langle W^2 \rangle - \langle W \rangle^2). \quad (2)$$

Since it has been shown that this introduces a bias, one way to correct it is to use the unbiased estimator for the variance, yielding to the following approximation for the PMF (corresponding to equation 19 in reference¹⁴):

$$\Psi_M \equiv \frac{1}{M} \sum_{i=1}^M W_i - \frac{\beta}{2} \frac{M}{M-1} \left[\frac{1}{M} \sum_{i=1}^M W_i^2 - \left(\frac{1}{M} \sum_{i=1}^M W_i \right)^2 \right], \quad (3)$$

where M is the number of trajectories used in the calculation (here $M = 10$ since the in total 20 simulations performed were grouped into two sets). Details about the pulling simulations and how equation 3 is applied are given next.

Constant-velocity pulling simulations

Setup of the simulations. The constant-velocity pulling simulations were started from snapshots sampled during the runs with no tensile force (Table 1). The protein and a bulk layer of 6 Å were removed from the original cubic system and placed into a rectangular water box with side lengths of 140 Å in the direction of pull and 65 Å in the other two directions. The system was then equilibrated at 300 K during 0.2 ns with harmonic constraints applied to the heavy atoms of the protein similarly as for the runs with no tensile force. Positional restraints were then applied to the coordinates of the C_α atom of the N-terminus of LD. The C_α atom of the C-terminus of the donated FimG strand in PD (Figure 1) was attached through a virtual spring with a stiffness constant of 2 kcal/mol/Å² to a dummy atom that was pulled at a constant velocity of 1 Å/ns. This mimics the situation where LD is bound to mannose on a host cell while the rest of the fimbria is pulled away from it as the bacterial cell experiences drag forces due to shear stress. The initial direction of pull was parallel to the axis through the fixed N-terminal C_α atom and the pulled atom. As the dummy atom is pulled, the spring extends. Using Hook’s law, the resulting applied tensile force is defined as $F = k\Delta x$, where Δx is the extension and k the stiffness constant of the spring. It needs to be noted that from a physical point of view the direction of force is not relevant, because the protein would rotate as a rigid body until the axis through the pulled atoms is aligned parallel to the direction of pull, if this was not the case at the start of the simulation. In total, 20 pulling simulations were performed with each FimH and FimH^{FocH} and each run lasted 25 ns (Table 1). This number was estimated from a previous study where the Jarzynski equality was applied to simulations with deca-alanine.¹⁴

Calculation of PMF. When applying the Jarzynski equality in conjunction with pulling simulations, the PMF needs to be calculated as a function of a parameter λ that is varied along the simulation.¹⁴ In this case, the parameter λ is the distance between the fixed N-terminal C_α atom and the position of the dummy atom. Because of the stiff-spring approximation the reaction coordinate generally follows the parameter λ ,¹⁴ and the spring constant of 2 kcal/mol/Å² used here is a bit softer but still of comparable magnitude than what was used in the benchmark study with deca-alanine.¹⁴ The work performed during pulling is calculated as:

$$W = F_{\parallel} \Delta\lambda, \quad (4)$$

where F_{\parallel} is the component of the force parallel to the direction of pull and $\Delta\lambda$ is the increase in λ between saved steps. The resulting value for W is then used in equation 3 to calculate the PMF. The 20 pulling simulations were grouped into two sets of 10 and the PMF was calculated for each set individually. Averages and standard errors of the mean (SEM) were then calculated ($n = 2$) and plotted as a function of λ . The fluctuation of work was estimated by calculating the standard deviation of W over all 20 runs, and its comparison to kT yields an estimate how accurate the free energy calculation is.

Results

Comparison of backbone flexibility and inter-domain contacts

In order to investigate differences in backbone flexibility and inter-domain contacts, and to generate snapshots for pulling simulations, 50-ns long runs at 300 K were performed with wild-type FimH and FimH^{FocH} (Table 1, from here on "FimH" will always refer to "wild-type FimH"). Although the resulting trajectories were already used in a previous published study to provide atomistic explanations for experimental observations,¹⁰ here a more detailed analysis is provided to investigate how the mutations alter the dynamics of residues at the inter-domain interface.

Time series of the C_α root mean square deviation (RMSD) indicate that the proteins are stable in the 50-ns long simulations and no significant differences are observed between FimH and FimH^{FocH} for the total and individual domains' RMSD (Figure 2). This is consistent with experimental data showing that in the absence of mannose both FimH and FimH^{FocH} are likely to be found in LAS with LD and PD docked to each other.¹⁰

Despite no differences in overall flexibility, previous studies by the author on protein complexes have provided evidence that mutations near a binding interface can often induce changes in the network of inter-protein (in this case, inter-domain) contacts.^{23,24} For this reason, I analysed the persistence of inter-domain contacts between LD and PD, and compared FimH to FimH^{FocH}.

Consistent with a prior study,¹⁰ the activating mutation A188D causes the loss of a side chain contact in FimH^{FocH} involving residue 188 (Figure 3), although the total number of persistent inter-domain side chain contacts was

the same in both runs with both variants (Table 2 and Figure 3). Besides side chain contacts, which are mostly stabilized by hydrophobic interactions, it is necessary to compare also electrostatic contacts at the inter-domain interface to better understand how the mutations in FimH^{FocH} alter its activation mechanism. Generally, electrostatic interactions are thought to be stronger than hydrophobic ones and to confer specificity in both protein folding and protein-protein binding.^{25,26}

Analysis of inter-domain hydrogen bonds between LD and PD shows a slightly stronger network in FimH than FimH^{FocH}. Two hydrogen bonds were persistent in both runs with FimH, and both are present also in the X-ray structure (Table 2 and Figures 3 and 4). In contrast, only one hydrogen bond was persistent in both simulations with FimH^{FocH} and it is not observed in the X-ray structure (Table 2 and Figures 3 and 4). In particular, the hydrogen bond involving the side chain of R166 is broken in most frames of one run with FimH^{FocH} while persistent in both simulations with FimH (Table 2 and Figures 3 and 4). This hydrogen bond is likely to be important in stabilizing the hook-shaped structure of FimH because it involves a large charged side chain buried at the interface and the large entropic cost of burying an arginine needs to be compensated by the hydrogen bond formation. Buried hydrogen bonds are known to significantly contribute to the stability of protein structure. No persistent inter-domain salt bridges were observed in any of the simulations although a salt bridge was observed in a previous study between LD and the neighboring FimG domain.⁴ Interestingly, a hydrogen bond between the side chain of N33 and the backbone carboxyl group of G159 in the linker was observed to be persistent in all four simulations (Figure 3). Although this is not an inter-domain interaction, it may contribute to the stability of the LAS conformation of LD.

In order to understand why there is a difference in hydrogen bond formation between FimH and FimH^{FocH}, I analyzed the mobility of the R166 side chain and its interaction with other charged side chains along the simulations. A persistent salt bridge was observed between R166 and D162 in both simulations with FimH (Figure 5a,c). In contrast, in the simulations with FimH^{FocH} the D162-R166 salt bridge was rarely formed (Figure 5b,c). Instead, the D162 side chain formed a salt bridge with R186, which was persistent in one of the two simulations with FimH^{FocH} and still present in 37% of the time in the other run (Figure 5b,c). The observed switch in salt bridge formation is most likely due to the Y186R mutation, which introduces an arginine competing with R166 for salt bridge formation with D162 (Figure 5b). Interestingly, the formation of the D162-R186 salt bridge mostly coincides with the rupture of the A115-R166 hydrogen bond (Figure 5c). Also, visual analysis reveals that when the D162-R166 salt bridge is not formed R166 moves away from the inter-domain interface towards the solvent (Figure 5b). Taken together, this analysis reveals that the D162-R166 salt bridge stabilizes the position of R166 facilitating its hydrogen bond formation with A115. However, the mutation Y186R in FimH^{FocH} causes D162 to move away from R166 leading to the destabilization of the inter-domain hydrogen bond between the R166 side chain and the backbone of A115. Taken together, although residue 186 is not directly involved in inter-domain contacts, the Y186R mutation still alters the interface by disrupting hydrogen bond formation through a domino effect on nearby residues.

Comparing free energies between FimH and FimH^{FocH} during domain separation using the Jarzynski equality

Analysis of simulation trajectories that sample around the native state can provide clues as to how the mutations in FimH^{FocH} may weaken the LD-PD inter-domain interface compared to wild-type FimH. Such destabilization of the LD-PD complex is thought to lead to the activation of FimH^{FocH} in the presence of mannose but absence of shear while both, mannose and shear, are necessary to activate FimH.¹⁰ However, it is *a priori* not clear whether and how destabilization of specific inter-domain contacts alters the strength of the interaction between LD and PD. One effective way of measuring how strongly LD and PD bind to each other and compare FimH to FimH^{FocH} is to estimate the free energy of the binding and unbinding process between the two domains. Calculation of free energy profiles in principle requires the unbiased sampling of binding and unbinding events between LD and PD. However, such simulations would not be computationally feasible as this type of events occur in a time scale not accessible through current computational capabilities. A method has been developed to estimate the equilibrium free energy profile from a non-equilibrium process such as the application of tensile force using the Jarzynski equality¹³ (equation 1) and its second-order cumulant expansion (equations 2-3).¹⁴ This method was used here and I performed 20 pulling simulations at constant velocity with FimH and 20 similar runs with FimH^{FocH}. Then, I calculated for each protein the PMF (equation 3) from the work performed during pulling (equation 4) as detailed in Materials and Methods. This allowed me to compare the free energy profile of LD unbinding from PD between FimH and FimH^{FocH} and analyse differences in the sequence of events during the unbinding process. In order to estimate the amount of variability in the simulations, I grouped the runs into blocks

of 10 similarly as in a previous study,¹⁴ computed the PMF for each block individually (equation 3) and calculated the average and SEM for $n = 2$.

Comparison of the PMF profiles indicates a significantly lower free energy barrier for the unbinding process of LD from PD in the case of FimH^{FocH} than FimH (Figure 6a). Also, while the PMF profile increases rapidly for FimH, it remains roughly constant in the case of FimH^{FocH} in the early stage of stretching. Interestingly, the PMF for FimH^{FocH} displays a small dip at around $\lambda = 86$ Å, which coincides with the rupture of the inter-domain hydrogen bond between the R166 side chain and the backbone of A115. In the case of FimH, rupture of such hydrogen bond occurs later and at a significantly larger PMF value than for FimH^{FocH} (Figure 6a). These observations suggest that rupture of the inter-domain hydrogen bond involving the R166 side chain occurs much more readily in FimH^{FocH} than FimH. It is important to note that for $\lambda > 90$ Å the SEM for both proteins becomes very large indicating that as the domains separate more configurational space becomes accessible limiting the accuracy of the calculations. Another measure for the reliability of the free energy estimate is the fluctuation of work, W , in comparison to kT . For this reason, the standard deviation of W (δW) was calculated over all 20 simulations, divided by kT and plotted for both FimH and FimH^{FocH} (Figure 6b). In a previous benchmark study with deca-alanine,¹⁴ where application of the Jarzynski equality was shown to provide a good approximation for the free energy, the standard deviation for the total work was between $3.1 kT$ and $7.1 kT$ depending on pulling speed. Here, large error bars for the PMF coincide with δW values larger than $5 kT$ indicating that PMF values for $\lambda > 90$ Å are probably not reliable. Nonetheless, application of the Jarzynski equality to the separation process of LD from PD reveals a lower free energy barrier and an easier rupture of the inter-domain hydrogen bond in-

volving the R166 side chain in FimH^{FocH} compared to FimH. Rupture of the inter-domain hydrogen bond involving R166 is likely to lead to hydration of the inter-domain interface causing further separation of the domains. Thus, the lower energy barrier observed in the simulations in the early stage of stretching may explain the easier activation of FimH^{FocH} compared to FimH.

Discussion

The bacterial adhesin FimH consists of two domains: LD, which binds mannose on host cells; and PD, which connects to the rest of the fimbria. Such two-domain architecture confers to this protein an auto-inhibitory property whereby PD acts as an effector molecule regulating the mannose-binding affinity of LD in an allosteric manner.^{5,6} There is compelling evidence that LD and PD exist at an equilibrium between a state where they are bound to each other and a state where they are separated.¹⁰ In the situation where both mannose and tensile force are absent, the two domains are rather found to be bound to each other and LD is in LAS, which is characterized by weak binding to mannose. When mannose is bound and upon the application of tensile force, which is for example generated by shear stress in the urinary tract, the equilibrium is shifted towards the state where the domains are separated and LD binds strongly to mannose. Normally, mannose alone does not fully shift the equilibrium towards the active state. Indeed, FimH has been crystallized in an intermediate state where LD and PD are bound to each other but LD is in a hybrid conformation with mannose in the binding pocket.²⁷ However, in the case of a recombinant variant referred to here as FimH^{FocH}, which contains 10 mutations in PD near the inter-domain interface, mannose alone is enough to be activating even in the absence of tensile

force.¹⁰ This study aimed at understanding how the mutations in FimH^{FocH} alter contacts at the inter-domain interface and how this affects the strength of the interaction between LD and PD. The latter was accomplished by estimating the free energy profile of the separation process between LD and PD and monitoring changes in the sequence of events during the separation.

Analysis of persistent inter-domain interactions along simulations performed in the absence of external forces revealed that the hydrogen bond between the side chain of R166 and the backbone carboxyl oxygen of A115 is significantly weakened in FimH^{FocH} with respect to FimH (Table 2 and Figures 3 and 4). Rupture of the hydrogen bond involving R166 was found to coincide with a switch in salt bridge formation where the side chain of D162, which formed a persistent salt bridge with R166 in FimH, switched to form a salt bridge with R186 in FimH^{FocH} (Figure 5b,c). Such a salt bridge between residues 162 and 186 is not possible in FimH where a tyrosine is present at position 186. Hence, the intra-PD D162-R166 salt bridge may have a stabilizing function on the inter-domain hydrogen bond involving the R166 side chain, and the Y186R mutation in FimH^{FocH} disrupts such interaction. These results highlight how a mutation, such as Y186R in this case, may destabilize a region of interest, such as the interface between LD and PD, even though it is located distally and, in this case, residue 186 is not directly involved in any inter-domain contacts. Yet through a cascade of side chain rearrangements, Y186R leads to the loss of an inter-domain hydrogen bond.

In order to measure the strength of the interaction between LD and PD and compare FimH to FimH^{FocH}, pulling simulations at constant velocity were combined with the Jarzynski equality. The results indicated that in FimH^{FocH} there is likely to be a relatively lower free energy barrier for the

separation of LD from PD compared to FimH at least in the initial stage of the stretching (Figure 6). Furthermore, rupture of the inter-domain hydrogen bond involving R166 occurs significantly earlier in FimH^{FocH}. This is in agreement with the fact that this hydrogen bond was observed to be weaker in FimH^{FocH} than in FimH in the simulations with no applied force.

The stabilizing function of the inter-domain R166-A115 hydrogen bond is in agreement with the observation in *in vitro* studies that the mutation R166H has an activating effect. In fact, variants of FimH from the multidrug resistant *E. coli* sequence type 131 (ST131) lineage that carried the R166H mutation were shown to have increased mannose-binding ability.²⁸ Furthermore, a recent study showed that *E. coli* cells expressed with the specific variant FimH30, which is also from the ST131 lineage and contains the R166H mutation, increased adhesion to bladder cells, invasion of intestinal cells and biofilm formation compared to other variants of the same lineage.²⁹ In the same study,²⁹ MD analysis of the R166H mutation revealed that the electrostatic interaction between residue 166 and A115 was lost due to the replacement of arginine with histidine. Hence, there is a common mechanism between the activation observed in FimH^{FocH} and in the variants containing R166H, i.e., the disruption of inter-domain hydrogen bonding between R166 and A115 facilitates separation of the FimH domains from each other. Therefore, similarly as the PMF profile is reduced in FimH^{FocH} relative to FimH it is plausible that a similar reduction in free energy would be observed if the same method used here was applied to R166H-bearing variants.

The current work presents an application of the Jarzynski equality to the separation process of two relatively large protein domains. Furthermore, it provides a way to assess how reliable the calculation of the PMF is along the extension. In the case of the comparison between FimH and FimH^{FocH}, the

rupture of key inter-domain contacts occurs in the early stage of the stretching where the estimation of the free energy can still be considered reliable according to the fluctuation of work, δW , relative to thermal energy. Applications of the Jarzynski equality to biomolecules reported in the literature up to date are limited to rather small systems such as the unfolding of small proteins,^{14,30} the unbinding of ligands from receptors,³⁰ or, the pulling of an ion through a potassium channel.³¹ In contrast, here the Jarzynski equality was applied to the estimation of a free energy profile involving the complex between two relatively large protein domains. The ability of this analysis to reproduce an experimental difference between the two variants suggests that the method described in this study based on the Jarzynski equality can provide relevant insight to larger systems.

The two-domain architecture observed in FimH is typical of glycan-binding lectin-like bacterial adhesins, which are common especially among gram-negative bacteria. Hence, it can be postulated that LDs of such adhesins are regulated by an allosteric mechanism as in FimH. In fact, a similar allosteric activation mechanism has been shown also for the galactose-specific adhesin FmlH³² located at the tip of F9 fimbriae. Therefore, it is thinkable that the method described here based on the Jarzynski equality can be applied to other two-domain adhesins in order to, for example, study at atomic level of detail the effect of activating mutations near the inter-domain interface, or even screen for mutations that may facilitate activation by lowering the free energy barrier of the separation between LD and PD.

In conclusion, the present study provides an explanation for the relatively easier activation of FimH^{FocH} based on an analysis of persistent contacts near the inter-domain interface and the estimation of the free energy profile of the domains separation process using a method based on the Jarzynski

equality. The study also employed a criterion to assess the reliability of the PMF calculation along the projected axis. Given how ubiquitous two-domain bacterial adhesins are the method described here can be used to study the allosteric activation mechanism of a vast family of proteins.

Acknowledgments

I would like to thank Wendy Thomas for helpful discussions about the Jarzynski equality and for help editing the manuscript. I would also like to thank Evgeni Sokurenko for providing helpful comments about the manuscript. The simulations were performed on the Comet supercomputer at the San Diego Supercomputing Center thanks to a XSEDE allocation³³ with grant number TG-MCB140143, which is made available through support from the National Science Foundation. This research was financially supported by NIH grants R01AI50940, R01AI171570 and R01HL153253.

Conflict of interest

The author declares no potential conflict of interest.

References

1. Changeux, J. Feedback control mechanism of biosynthetic L-threonine deaminase by L-isoleucine. *Cold Spring Harb. Symp. Quant. Biol.* 26:313–&, 1961.
2. Monod, J., Wyman, J., and Changeux, J. On nature of allosteric transitions - a plausible model. *J. Mol. Biol.* 12(1):88–&, 1965.
3. Hahn, E., Wild, P., Hermanns, U., Sebbel, P., Glockshuber, R., Haner, M., Taschner, N., Burkhard, P., Aebi, U., and Muller, S. Exploring the 3D molecular architecture of Escherichia coli type 1 pili. *J. Mol. Biol.* 323(5):845–857, 2002.
4. Aprikian, P., Interlandi, G., Kidd, B., Le Trong, I., Tchesnokova, V., Yakovenko, O., Whitfield, M., Bullitt, E., Stenkamp, R., Thomas, W., and Sokurenko, E. The Bacterial Fimbrial Tip Acts as a Mechanical Force Sensor. *PLoS Biology* 9(5), 2011.
5. Le Trong, I., Aprikian, P., Kidd, B. A., Forero-Shelton, M., Tchesnokova, V., Rajagopal, P., Rodriguez, V., Interlandi, G., Klevit, R., Vogel, V., Stenkamp, R. E., Sokurenko, E. V., and Thomas, W. E. Structural Basis for Mechanical Force Regulation of the Adhesin FimH via Finger Trap-like β Sheet Twisting. *Cell* 141(4):645–655, 2010.
6. Yakovenko, O., Sharma, S., Forero, M., Tchesnokova, V., Aprikian, P., Kidd, B., Mach, A., Vogel, V., Sokurenko, E., and Thomas, W. E. FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. *J. Biol. Chem.* 283(17):11596–11605, 2008.

7. Rodriguez, V. B., Kidd, B. A., Interlandi, G., Tchesnokova, V., Sokurenko, E. V., and Thomas, W. E. Allosteric coupling in the bacterial adhesive protein fimh. *Journal of Biological Chemistry* 288(33):24128–24139, 2013.
8. Interlandi, G. and Thomas, W. E. Mechanism of allosteric propagation across a beta-sheet structure investigated by molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* 84(7):990–1008, Jul, 2016.
9. Thomas, W. E., Vogel, V., and Sokurenko, E. Biophysics of catch bonds. *Annu. Rev. Biophys.* 37:399–416, 2008.
10. Thomas, W. E., Carlucci, L., Yakovenko, O., Interlandi, G., Le Trong, I., Aprikian, P., Magala, P., Larson, L., Sledneva, Y., Tchesnokova, V., Stenkamp, R. E., and Sokurenko, E. V. Recombinant FimH Adhesin Demonstrates How the Allosteric Catch Bond Mechanism Can Support Fast and Strong Bacterial Attachment in the Absence of Shear. *J. Mol. Biol* 167681, Jun, 2022.
11. Aprikian, P., Tchesnokova, V., Kidd, B., Yakovenko, O., Yarov-Yarovoy, V., Trinchina, E., Vogel, V., Thomas, W., and Sokurenko, E. Interdomain interaction in the FimH adhesin of *Escherichia coli* regulates the affinity to mannose. *J. Biol. Chem.* 282(32):23437–23446, 2007.
12. Tchesnokova, V., Aprikian, P., Yakovenko, O., Larock, C., Kidd, B., Vogel, V., Thomas, W., and Sokurenko, E. Integrin-like allosteric properties of the catch bond-forming FimH adhesin of *Escherichia coli*. *J. Biol. Chem.* 283(12):7823–7833, 2008.

13. Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* 78:2690–2693, Apr, 1997.
14. Park, S., Khalili-Araghi, F., Tajkhorshid, E., and Schulten, K. Free energy calculation from steered molecular dynamics simulations using Jarzynski’s equality. *J. Chem. Phys.* 119(6):3559–3566, 2003.
15. Brooks, B. R., Brooks, III, C. L., Mackerell, Jr., A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caffisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* 30(10, SI):1545–1614, 2009.
16. Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. NAMD2: Greater scalability for parallel molecular dynamics. *Journ. Comp. Physics* 151:283–312, 1999.
17. MacKerell, A., Bashford, D., Bellott, M., Dunbrack, R., Evanseck, J., Field, M., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F., Mattos, C., Michnick, S., Ngo, T., Nguyen, D., Prodhom, B., Reiher, W., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102:3586–3616, 1998.

18. MacKerell, A., Feig, M., and Brooks, C. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* 126(3):698–699, Jan 28, 2004.
19. Mackerell, A., Feig, M., and Brooks, C. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25(11):1400–1415, 2004.
20. Darden, T., York, D., and Pedersen, L. Particle mesh ewald - an $N \cdot \log(N)$ method for ewald sums in large systems. *J. Chem. Phys* 98:10089–10092, 1993.
21. Schneider, T. and Stoll, E. Molecular-Dynamics study of a 3-dimensional one-component model for distortive phase-transitions. *Physical Review B* 17(3):1302–1322, 1978.
22. Feller, S. E., Zhang, Y. H., Pastor, R. W., and Brooks, B. R. Constant-pressure molecular-dynamics simulation - the Langevin piston method. *Journal of chemical physics* 103(11):4613–4621, Sep 15, 1995.
23. Interlandi, G. and Thomas, W. The catch bond mechanism between von Willebrand Factor and platelets investigated by molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics* 78(11):2506–2522, 2010.
24. Interlandi, G., Yakovenko, O., Tu, A.-Y., Harris, J., Le, J., Chen, J., Lopez, J. A., and Thomas, W. E. Specific electrostatic interactions between charged amino acid residues regulate binding of von Willebrand factor to blood platelets. *J. Biol. Chem.* 292(45):18608–18617, Nov 10, 2017.

25. Zhou, H. X. and Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* 118(4):1691–1741, 02, 2018.
26. Hubbard, R. E. and Haider, M. K. Hydrogen bonds in proteins: Role and strength, , Feb, 2010.
27. Sauer, M. M., Jakob, R. P., Eras, J., Baday, S., Eris, D., Navarra, G., Berneche, S., Ernst, B., Maier, T., and Glockshuber, R. Catch-bond mechanism of the bacterial adhesin FimH. *Nat. Commun.* 7:10738, 2016.
28. Paul, S., Linardopoulou, E. V., Billig, M., Tchesnokova, V., Price, L. B., Johnson, J. R., Chattopadhyay, S., and Sokurenko, E. V. Role of homologous recombination in adaptive diversification of extraintestinal *Escherichia coli*. *J. Bacteriol.* 195(2):231–242, Jan, 2013.
29. Qin, J., Wilson, K. A., Sarkar, S., Heras, B., O’Mara, M. L., and Totsika, M. Conserved FimH mutations in the global *Escherichia coli* ST131 multi-drug resistant lineage weaken interdomain interactions and alter adhesin function. *Comput. Struct. Biotechnol. J.* 20:4532–4541, 2022.
30. Xiong, H., Crespo, A., Marti, M., Estrin, D., and Roitberg, A. E. Free energy calculations with non-equilibrium methods: applications of the Jarzynski relationship. *Theor. Chem. Acc.* 116(1-3):338–346, 2006.
31. Bastug, T. and Kuyucak, S. Application of jarzynskis equality in simple versus complex systems. *Chemical Physics Letters* 436(4):383–387, 2007.
32. Kisiela, D., Magala, P., Interlandi, G., Carlucci, L., Ramos, A., Tchesnokova, V., Basanta, B., Yarov-Yarovoy, V., Avagyan, H., Hovhannisyan, A., Thomas, W., Stenkamp, R., Klevit, R., and Sokurenko, E. Toggle

switch residues control allosteric transitions in bacterial adhesins by participating in a concerted repacking of the protein core. *PLoS Pathog.* 17(4), Apr, 2021.

33. Towns, J., Cockerill, T., Dahan, M., Foster, I., Gaither, K., Grimshaw, A., Hazlewood, V., Lathrop, S., Lifka, D., Peterson, G. D., Roskies, R., Scott, J. R., and Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* 16(5):62–74, Sep-Oct, 2014.

Figure 1: **Comparison of the crystallographic structures of (a) wild-type FimH and (b) FimH^{FocH} highlighting the location of the mutations.** In both structures, the proteins are in LAS with LD and PD docked to each other in a hook-shaped conformation. The PDB codes for FimH and FimH^{FocH} are 3JWN and 7SZO, respectively (chain H was used for each structure). The location of the mannose binding site is indicated. Side chains that differ between FimH and FimH^{FocH} are shown in the stick and ball representation and labeled. The backbone of LD and PD is colored in cyan and magenta, respectively, while the backbone of the donated β strand of FimG (truncated after residue 13) is colored in orange.

Figure 2: **Time series of the C_α RMSD along the trajectories with no external forces applied.** The C_α RMSD was calculated from the initial conformation in the two simulations with wild-type FimH (FimH_300K_1,2) and the two runs with FimH^{FocH} (FimH^{FocH}_300K_1,2, Table 1). N- and C-terminal residues were excluded from the calculations as they are generally found to fluctuate. Thus, the values for LD were determined using residues 2 to 158, for PD using residues 161 to 278 and 2 to 12 of the donated strand of FimG, and for the entire protein (LD-PD) the residues 2 to 278 and 2 to 12 of FimG were used.

Figure 3: **Time series of the formation of inter-domain side chain contacts and hydrogen bonds along the trajectories with no external force applied.** Side chain contacts are in turquoise while hydrogen bonds are colored in blue. The fraction of frames in a run where a specific contact is formed is indicated on the right side of each graph. A value in boldface highlights that a contact is persistent (i.e., it is formed in at least 66% of the simulation frames). If a contact is persistent in both runs the corresponding axis label on the left side of the graphs is in boldface. If it is persistent in only one run it is indicated in the regular font while if it is not persistent in either run the label is italicized. Vertical cyan dashed lines indicate that the first 10 ns of each simulation were not used to calculate how persistent a contact was. The hydrogen bond between the side chain of N33 and the backbone of G159 is also shown although it is not an inter-domain contact, but it may help stabilize the LAS conformation of LD.

Figure 4: **Persistent inter-domain hydrogen bonds.** (a) Simulations with FimH. (b) Simulations with FimH^{FocH}. Hydrogen bonds persistent in both simulations performed with the same structure are highlighted with blue dashed lines while hydrogen bonds persistent in only one simulation with a given structure are colored in green. Side chains or backbone functional groups involved in the described hydrogen bonds are shown in the stick and ball representation. For simplicity, only donated hydrogen atoms are shown. The displayed conformations were sampled after 10 ns in the runs FimH_300K_1 for (a) and FimH^{FocH}_300K_1 for (b), respectively.

Figure 5: **Electrostatic interactions network at the R166 site.** (a) Snapshot sampled after 20 ns in run FimH_300K_2. (b) Snapshot sampled after 20 ns in run FimH^{FocH}_300K_2. (c) Time series of the formation of the inter-domain hydrogen bond between A115 and R166 and intra-domain salt bridges involving D162. Hydrogen bonds are highlighted with blue dashed lines while salt bridges are colored in red. Side chains or backbone functional groups involved in the described electrostatic interactions are shown in the stick and ball representation. The Y186 side chain in FimH is also displayed for illustration purposes. The fraction of frames in a run where a specific contact is formed is indicated on the right side of the time series plot in (c). A value in boldface highlights that a contact is persistent (i.e., it is formed in at least 66% of the simulation frames). If a contact is persistent in both runs the corresponding axis label on the left side of the time series is in boldface. If it is persistent in only one run it is indicated in the regular font while if it is not persistent in either run the label is italicized. Vertical cyan dashed lines in the plots in (c) indicate that the first 10 ns of each simulation were not used to calculate how persistent a contact was.

Figure 6: **Approximation of the free energy profile along the separation process of LD and PD from each other comparing FimH and FimH^{FocH}.** (a) Projection of the free energy onto the direction of pulling (referred to as potential of mean force (PMF)) approximated through equation 3, which is based on the Jarzynski equality (see Materials and Methods). The PMF was calculated by applying equation 3 individually to two blocks of 10 pulling simulations each and then averaging the two values. The "envelope" along the averages represents standard errors of the mean ($n = 2$). Averages and SEM were calculated every 2 ps and thus the error bars are very close to each other forming an envelope. Indicated are the time points where two inter-domain hydrogen bonds break with averages and standard deviations calculated from all 20 pulling simulations. Besides the R166-A115, the C161-S114 hydrogen bond is also shown as a previous study by us showed that this contact is closest to the hinge axis⁴ and thus its rupture indicates full separation of the domains. The time point of hydrogen bond rupture was defined as the last snapshot in a pulling simulation where the contact was observed to be formed. (b) Estimate of the accuracy of the PMF approximation by comparing the fluctuation of Work δW to kT . The horizontal cyan dashed line highlights the threshold where we define that δW is significantly larger than kT .

Table 1: **Simulation Systems**

Name ^a	Starting structure	Type	Duration [ns]
FimH_300K_1,2	wild-type FimH	300 K ^b	2 x 50
FimH ^{FocH} _300K_1,2	FimH ^{FocH}	300 K ^b	2 x 50
FimH_pull_1-20	FimH_300K_1 10-29 ns	pull ^c	20 x 25
FimH ^{FocH} _pull_1-20	FimH ^{FocH} _300K_1 10-29 ns	pull ^c	20 x 25

^aIndexes are used to denote different replicas. ^bNo forces were applied in these runs to sample the native state. ^cThe pulling simulations were started from the snapshots sampled after 10 ns, 11 ns, ..., 29 ns of the respective 300-K simulation where no forces were applied.

Table 2: **Number of Inter-Domain Contacts**
FimH

	X-ray	FimH_300K_1	FimH_300K_2	Both runs
Side chain contacts	14 (14)	10 (10)	9 (9)	9 (9)
Hydrogen bonds	3 (3)	2 (2)	2 (2)	2 (2)

FimH^{FocH}

	X-ray	FimH ^{FocH} _300K_1	FimH ^{FocH} _300K_2	Both runs
Side chain contacts	15 (14)	11 (10)	10 (8)	9 (8)
Hydrogen bonds	3 (3)	3 (2)	1 (0)	1 (0)

The number of contacts in the "X-ray" column was calculated for both, chain H (reported outside the parenthesis) and chain N (reported inside the parenthesis). The numbers in parenthesis in the simulation columns are the subset of contacts present in the respective X-ray structure (chain H) that was used to start the simulations. "Both runs" refers to contacts persistent in both simulations started with the same structure. The simulations are listed in Table 1.

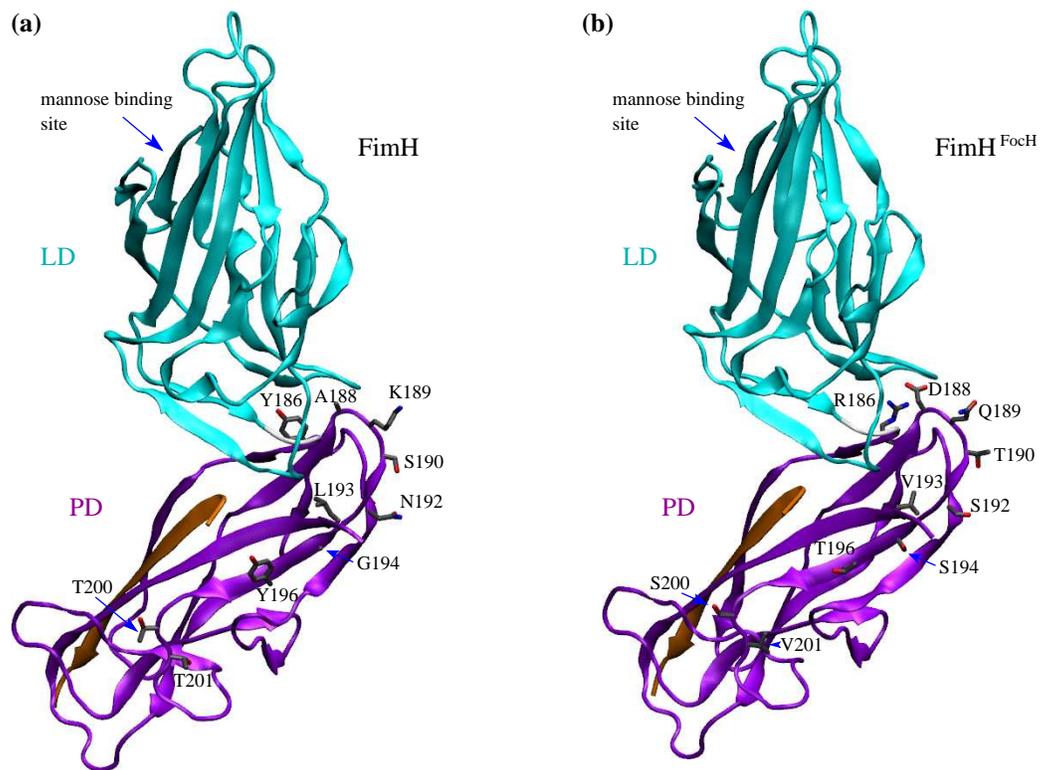


Figure 1:

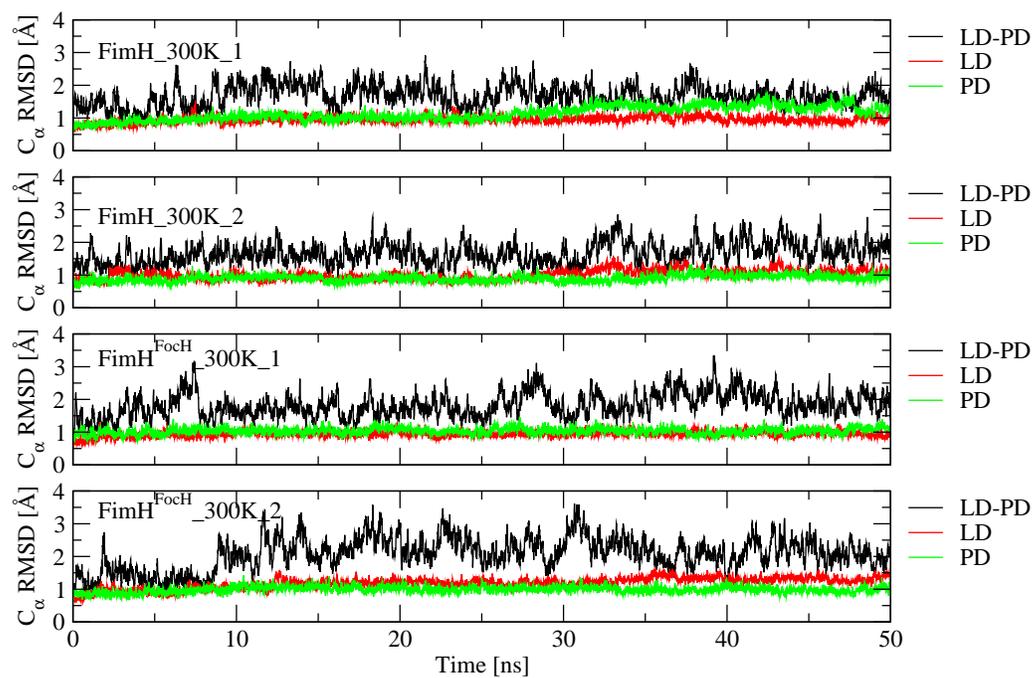


Figure 2:

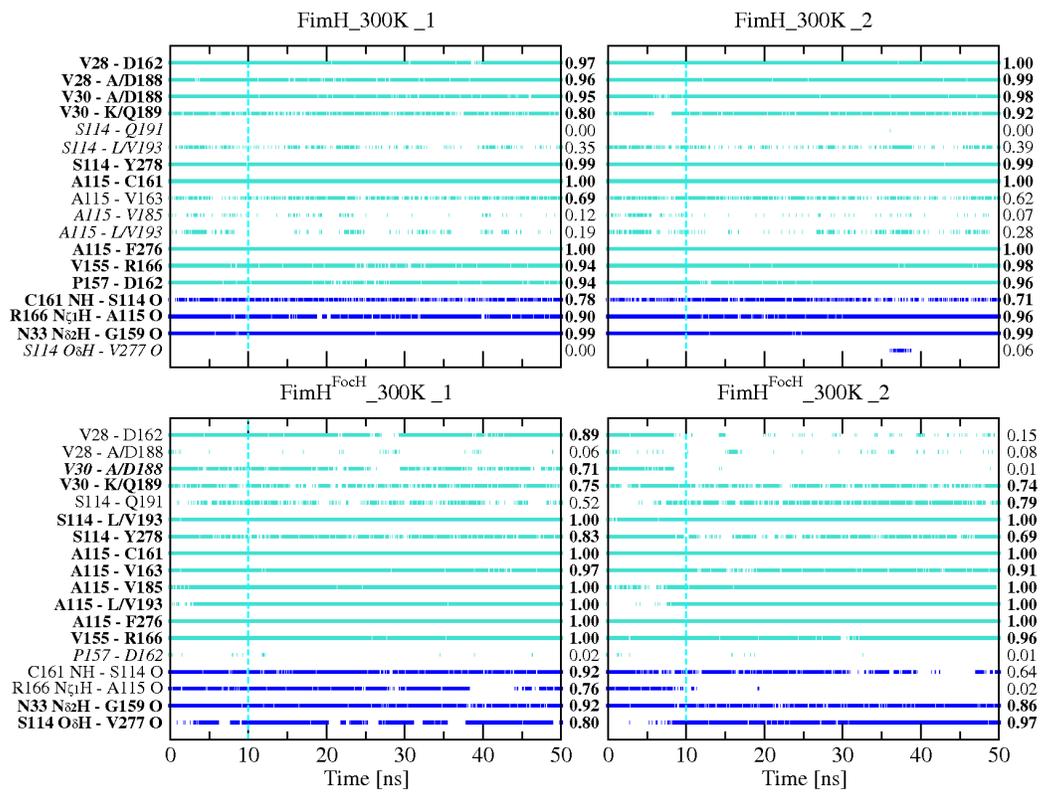


Figure 3:

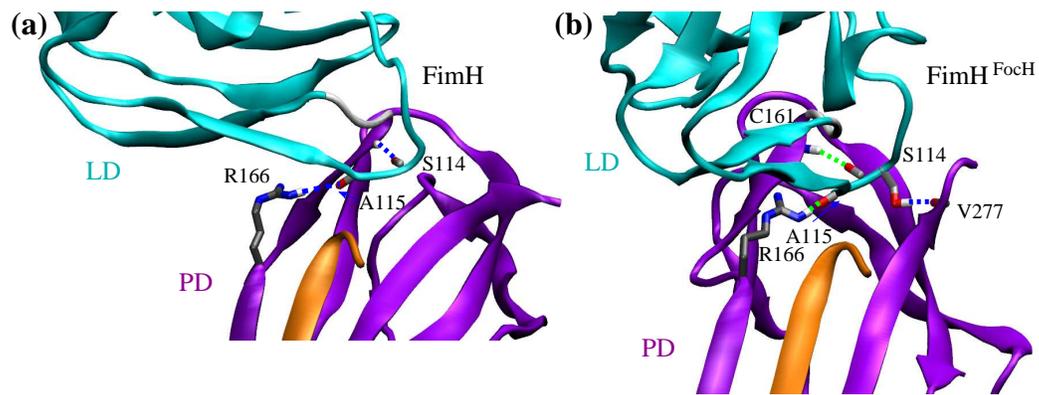


Figure 4:

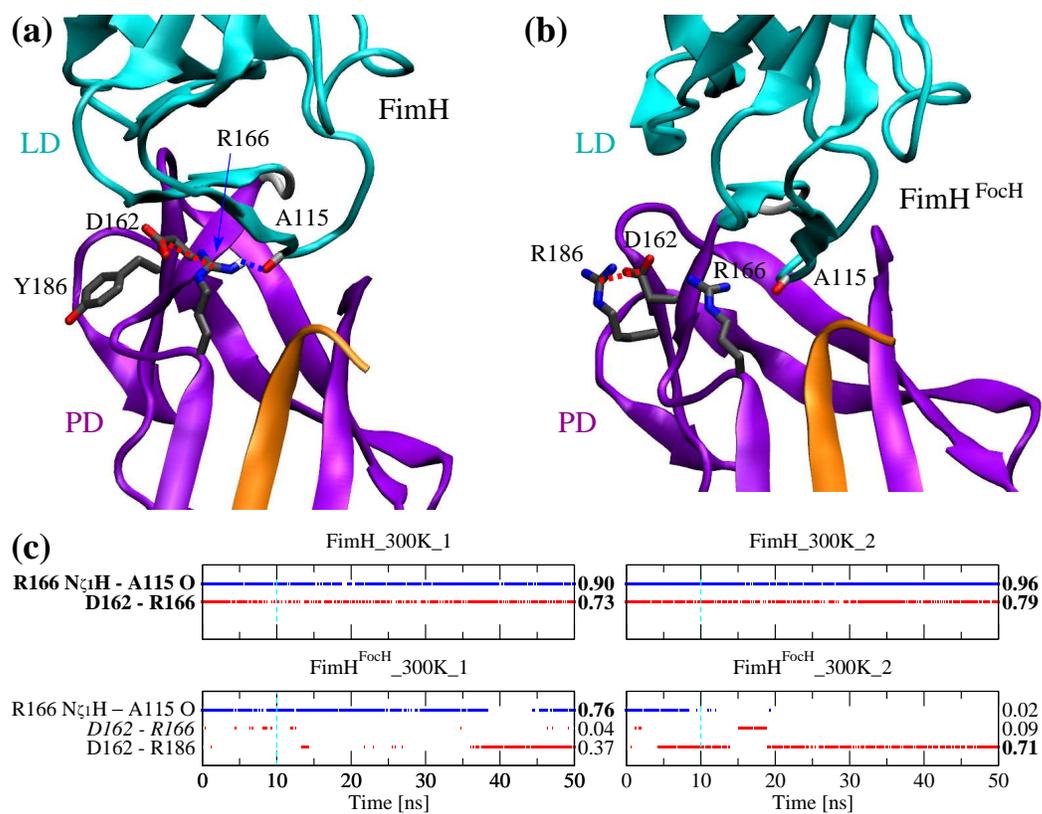


Figure 5:

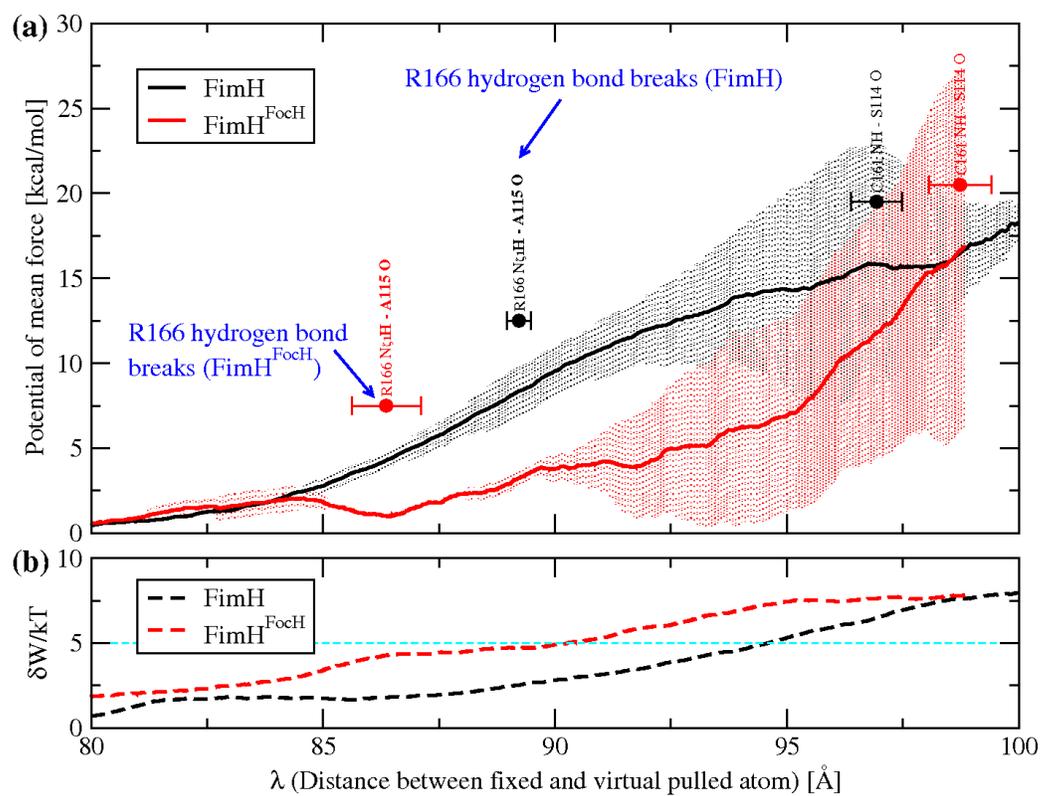


Figure 6: