PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Mechanism of allosteric propagation across a β-sheet structure investigated by molecular dynamics simulations

Gianluca Interlandi* and Wendy E. Thomas

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

ABSTRACT

The bacterial adhesin FimH consists of an allosterically regulated mannose-binding lectin domain and a covalently linked inhibitory pilin domain. Under normal conditions, the two domains are bound to each other, and FimH interacts weakly with mannose. However, under tensile force, the domains separate and the lectin domain undergoes conformational changes that strengthen its bond with mannose. Comparison of the crystallographic structures of the low and the high affinity state of the lectin domain reveals conformational changes mainly in the regulatory inter-domain region, the mannose binding site and a large β sheet that connects the two distally located regions. Here, molecular dynamics simulations investigated how conformational changes are propagated within and between different regions of the lectin domain. It was found that the inter-domain region moves towards the high affinity conformation as it becomes more compact and buries exposed hydrophobic surface after separation of the pilin domain. The mannose binding site was more rigid in the high affinity state, which prevented water penetration into the pocket. The large central β sheet demonstrated a soft spring-like twisting. Its twisting motion was moderately correlated to fluctuations in both the regulatory and the binding region, whereas a weak correlation was seen in a direct comparison of these two distal sites. The results suggest a so called "population shift" model whereby binding of the lectin domain to either the pilin domain or mannose locks the β sheet in a rather twisted or flat conformation, stabilizing the low or the high affinity state, respectively.

Proteins 2016; 84:990–1008.

Key words: allosteric regulation; molecular dynamics; FimH adhesin; bacterial adhesion; biophysics; protein structure; thermodynamics; population shifts; cooperativity; correlations.

INTRODUCTION

Protein allostery is generally described as the regulation of the active site of a protein through an effector binding to a distal location and triggering or stabilizing a conformational transition across the protein.[1,2] As a result, allosterically regulated proteins can be found natively in two distinct states depending whether the effector is bound or not. Characterization of the allosteric propagation necessitates knowledge of both states. However, determining the structure of even only one functional state of a protein is challenging and relatively few proteins have been structurally characterized in both allosteric states.[3] Furthermore, most of the reported allosteric changes involve propagation through loop regions, or across regions of high helical content,[4] which are likely to confer a certain degree of flexibility.[5] For example, a α helix is observed to shift in the hinging region of myosin heads[6] and integrins.[7,8] To date, allosteric propagation through β sheets has only been characterized for the lectin domain of the protein FimH, an adhesin located at the fimbrial tip of uropathogenic *E. Coli*.[9,10] However, β sheets are also known to contain a certain degree of flexibility,[11] and 23% of structurally characterized protein domains contain β sheets but no α helices, while 52% contain both β sheets and α helices (in

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: NIH; Grant numbers: K25HL118137, 1R01AI119675; Grant sponsor: NSF; Grant number: CBET-1132860.

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

Received 3 November 2015; Revised 24 March 2016; Accepted 8 April 2016
Published online 18 April 2016 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.25050

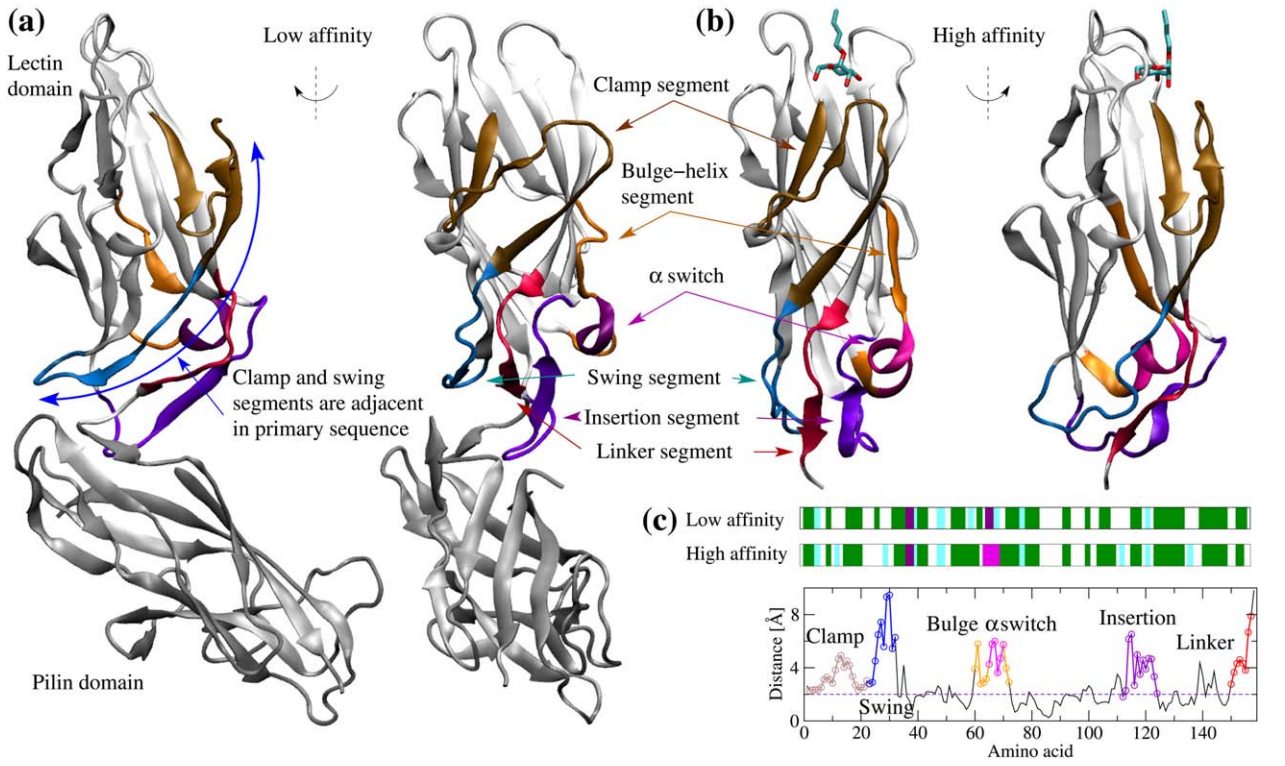

Figure 1

Tertiary and secondary structure of the lectin domain in two distinct crystallographic states. (**a**) X-ray structure of FimH in low affinity (PDB code 3JWN), where the domains are docked onto each other. The blue arch with arrows indicates that the clamp and swing segments are adjacent in primary sequence. (**b**) X-ray structure of isolated lectin domain in high affinity and with butyl α-D-mannose in the binding pocket (PDB code 1UWF). The simulations in this study were performed with the butyl group removed. The structures are displayed from two different angles, which differ by a rotation of ca. 90° around the axis through the longest dimension of the lectin domain. Regions that were identified to undergo significant conformational changes are colored as detailed in (c). (**c**) Secondary structure elements and displacements of C_α atoms between the crystallized low and high affinity conformations. Secondary structure is indicated by green for β strands, cyan for hydrogen bonded turns, magenta for α helices and purple for 3_{10} helices. The coordinates of the structures were first minimized (see "Materials and Methods") and then residues 2 to 157 were aligned before calculating the displacements shown in the plot. Segments consisting of at least seven consecutive amino acids with displacements larger than 2 Å were considered to undergo significant conformational changes between low and high affinity state and colored differently. If a segment started or ended within a loop (defined as a sequence not containing β strands or helices), the entire loop was included. Brown indicates the clamp, blue the swing, orange the bulge-helix, violet the insertion and red the linker segment. Residues 65 − 70 are highlighted in magenta in all structures and in the plot to indicate that this region switches from a 3_{10} to an α helix from low to high affinity. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

contrast only 16% contain exclusively α-helical structure). The protein FimH may thus provide a model system for understanding what may prove to be a common phenomenon.

The protein FimH consists of a pilin domain that anchors it to the rest of the fimbria and a lectin domain that binds mannose on the surface of infected cells[12] (Fig. 1). Both domains consist mainly of β sheets forming so called β-sandwich folds. Under static conditions, lectin and pilin domains are docked onto each other and FimH forms a hook-shaped structure[13] [Fig. 1(a)] with a low binding affinity to mannose.[10] When tensile force is present across FimH, for example under shear flow, the two domains separate from each other[13,14]and the lectin domain is found in a rather elongated conformation [Fig. 1(b)]. This is characterized by a large β sheet in the lectin domain adopting a less twisted configuration than in the docked state and the closing of a loop around mannose (colored in brown in Fig. 1), conferring a higher binding affinity to the ligand.[9] Therefore, the lectin domain is allosterically regulated because docking of the pilin domain affects the distally located active site, in this case resulting in lower binding affinity to mannose. The pilin domain can be said to have the role of an allosteric effector, which binds to the regulatory interdomain region of the lectin domain. This auto-inhibitory mechanism confers to FimH the property of a so called "catch bond", that is, tensile force increases the life time of the bond.[9] A recent mutagenesis study that used a conformation-sensitive monoclonal antibody, which recognizes only the high affinity state, determined that the conformation of the binding pocket and of the inter-

domain region of the lectin domain are weakly coupled to each other.[15] This suggests the presence of intermediate states that are significantly populated during the allosteric transition.

Molecular dynamics simulation studies have revealed that the allosteric pathway often consists of a combination of concerted motions (that is, coupled thermodynamically) and events that can only occur sequentially[4] (like in the case of the monomeric single-domain chemotaxis protein Y, CheY[16]). Furthermore, allostery is often described in terms of "population shifts".[17,18] According to this, the state in which the allosteric protein is found in the presence of the ligand can be visited even in its absence but is less populated. Binding of the ligand alters the energy landscape such that the conformational equilibrium shifts toward the state in which the protein is found after the allosteric transition. In order to understand how conformational changes are propagated in the FimH lectin domain across the central β sheet, it is necessary to characterize the coupling between the following processes: interaction with the pilin domain, re-arrangements in the inter-domain region, twisting of the β sheet, conformational changes in the binding pocket, and binding to mannose.

This manuscript investigated with molecular dynamics (MD) simulations how the backbone flexibility of the lectin domain is related to its allosteric regulation and mannose-binding function. Simulations performed at room temperature analyzed how structural differences between the two states in the inter-domain region and the mannose binding site might be related to the allosteric regulation of the lectin domain. Simulations of the low affinity state with mannose computationally docked were run to elucidate why the closed conformation of the binding site provides a stronger interaction with mannose. Finally, μs-long simulations performed at 330 K on a specialized supercomputer provided details how the twisting motion of the large central β sheet is correlated with fluctuations in the mannose binding site and conformational re-arrangements in the inter-domain region that contacts the pilin domain. This study emphasizes that studying protein dynamics at atomic level of detail is necessary to understand how allosteric conformational changes are propagated. Finally, the knowledge gained from studying the FimH lectin domain can have implications for the development of methods to predict conformational changes in other β-sheets containing proteins suspected of being allosterically regulated.

MATERIALS AND METHODS

Initial conformations

The structure of the low affinity state of the lectin domain used in the simulations was obtained from the crystallographic structure containing FimH incorporated into the fimbrial tip (PDB code 3JWN).[10] The rest of the fimbrial tip except the lectin domain was truncated after residue at position 160. The resulting C-terminus was capped with N-methylamide in order to neutralize its charge. The simulations with the high affinity state were started from the crystallographic structure of the isolated lectin domain bound to butyl α-D-mannose (PDB code 1UWF)[19] after truncating the butyl chain for simplicity. For the purposes of this manuscript, the ligand α-D-mannose is simply referred to as "mannose". All 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.[20]

Docking of mannose to the low affinity state

To dock mannose to the binding site of the low affinity state of the lectin domain, the X-ray structure with PDB code 3JWN was aligned onto the high affinity structure (PDB code 1UWF) in order to minimize the RMSD of residues 1–6 and residues 44–48. The coordinates of mannose, which is present in the high affinity structure, were then used to create a complex between the protein in low affinity and the ligand. The entire system was then subjected to 100 steps of steepest descent minimization *in vacuo* and 500 steps of conjugate gradient minimization in a dielectric continuum using the program CHARMM.[20] The model constructed here of mannose bound to the low affinity state of the lectin domain has been used in a previous work,[21] which has been published during the preparation of this manuscript.

Simulations with NAMD

The MD simulations were performed with the program NAMD[22] using the CHARMM all-hydrogen force field (PARAM22)[23] and the TIP3P model of water. The different simulation systems are summarized in Table I. The proteins were inserted into a cubic water box with side length of 86 Å, resulting in a system with in total ca. 60,000 atoms. Chloride and sodium ions were added to neutralize the system and approximate a salt concentration of 150 m*M*. 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

Table I
Simulation Systems

Name[a]	Start structure (PDB code)	Temp[b] [K]	Duration [ns]	Software
Ld_low_1[c]	low affinity (3JWN[d])	300	50	NAMD
Ld_low_2[c]	low affinity (3JWN[d])	300	50	NAMD
Ld_high_1[c]	high affinity (1UWF)	300	50	NAMD
Ld_high_2[c]	high affinity (1UWF)	300	50	NAMD
Ld_low_mann_1[c]	low affinity + mannose (3JWN[d])	300	50	NAMD
Ld_low_mann_2[c]	low affinity + mannose (3JWN[d])	300	50	NAMD
Anton_low_330K_1[e]	10 ns Ld_low_1[f]	330	5801	Anton
Anton_low_330K_2[e]	10 ns Ld_low_2[f]	330	2280	Anton
Anton_high_330K_1	10 ns Ld_high_1[f]	330	1028	Anton
Anton_high_330K_2[e]	10 ns Ld_high_1[f]	330	912	Anton

[a]Names used in the text to denote a specific simulation.
[b]Temperature.
[c]Ld stands for lectin domain.
[d]The lectin domain (residues 1 − 160) of chain H was used.
[e]In these runs, constraints were applied onto oxygen atoms of mannose and atoms of the protein that form hydrogen bonds with the ligand in order to prevent mannose from escaping the binding pocket.
[f]The snapshot sampled after 10 ns in the indicated simulation was used as starting conformation.

taken into account by the Particle Mesh Ewald summation method.[24] 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 (or 330 K in the equilibration runs for Anton) 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, or 330 K, respectively by using the Langevin thermostat[25] with a damping coefficient of 1 ps^{-1}, while the pressure was held constant at 1 atm by applying a pressure piston.[26] The program VMD[27] was used to visualize the trajectories and to render the structures used in the figures of this manuscript. Analysis of the simulations was performed using the programs VMD[27] and CHARMM.[20] Plots were made with the program xmgrace.

Simulations on Anton

Anton is a specialized supercomputer for MD simulations.[28] All systems simulated on Anton were first equilibrated for 10 ns using NAMD on a conventional supercomputer in order to avoid crashes due to the fixed point arithmetic used on Anton. The simulations were run at constant temperature and pressure with a 2 fs time step using Anton software version 2.7. The non-bonded interactions were evaluated analytically within a distance of 9 Å. Beyond this cutoff, electrostatic interactions were calculated by means of the *k*-space Gaussian split Ewald (GSE) method[29] and evaluated every 6 fs via a multi-stepping algorithm. The temperature was held constant by means of the Nosè Hoover thermostat[30,31] while the pressure was kept constant through the Martyna-Tobias-Klein (MTK) barostat.[32] In simulations where mannose was constrained in the binding pocket, a flat bottom harmonic potential was applied to the distance between the following two groups of atoms: Phe^{1} N, Asp^{47} N, Asp^{54} $O_{\delta 1}$, Asp^{54} $O_{\delta 2}$, and Gln^{133} N_{ϵ}, and, O_2, O_3, O_6, and O_4 of mannose. The flat bottom harmonic potential was zero if the distance between the centers of mass between these two groups was smaller than 2.45 Å and increased with a spring constant of 100 kcal/mol/$Å^2$ beyond this cutoff.

Determination of hydrogen bonds

A hydrogen bond analysis was performed along the trajectories to determine state-dependent backbone hydrogen bonds and to study the interaction between mannose and the lectin domain. 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 or an acceptor A could both be either an oxygen or a nitrogen. A hydrogen bond formed in at least 66% of the frames of a trajectory was said to be persistent in that particular simulation. A backbone hydrogen bond was said to be formed weakly if the H...O distance was between 2.7 Å and 3.0 Å as long as the angle cutoff was satisfied. If a hydrogen bond was formed in at least 66% of the simulation frames using the larger cutoff (but in less than 66% of the frames using the stricter cutoff), then it was called a persistent weak hydrogen bond. If a backbone hydrogen bond was observed to be persistent in at least one of the two simulations with one state but was not persistent in any simulation with the other state even when using the larger cutoff, it was called

Table II
Number of State-Dependent Backbone Hydrogen Bonds

	X-ray[a]		Simulation[b]		Simulation weak[c]	
Segment (residues)	Low	High	Low	High	Low	High
Total$^{1–158}$	21	25	19 (18)	22 (22)	2	1
Clamp$^{2–22}$	0	3	0 (0)	2 (2)	0	1
Swing$^{23–32}$ (total)	4	3	3 (3)	2 (2)	0	0
Swing$^{23–32}$-Linker$^{150–158}$	0	2	0 (0)	2 (2)	0	0
Bulge-helix$^{60–72}$	7	12	7 (7)	12 (12)	1	0
α-switch$^{65–70}$ (total)	3	6	3 (3)	6 (6)	1	0
α-switch$^{65–70}$-Insertion$^{112–124}$	0	1	0 (0)	1 (1)	0	0
Insertion$^{112–124}$ (total)	5	6	4 (4)	5 (5)	1	0
Insertion$^{112–124}$-Linker$^{150–158}$	5	2	4 (4)	2 (2)	1	0
Linker$^{150–158}$	5	5	4 (4)	5 (5)	1	0

A hydrogen bond observed only in the low or the high affinity state, respectively, is said to be state-dependent (see "Materials and Methods" for the exact definition). The values indicate total numbers, number of hydrogen bonds involving amino acids in a particular segment and number of hydrogen bonds between segments. A complete list of hydrogen bonds along with their frequency is presented in Supporting Information Tables S1 and S2.
[a]Number of state-dependent hydrogen bonds based on the comparison between the two static X-ray structures (see "Materials and Methods").
[b]Number of state-dependent hydrogen bonds based on the analysis of the trajectories. The number in parenthesis indicates the subset of contacts observed to be state-dependent according to a comparison of the X-ray structures.
[c]Number of state state-dependent weak hydrogen bonds (see "Materials and Methods").

a state-dependent backbone hydrogen bond. Similarly, a weak hydrogen bond observed to be persistent in at least one simulation with only one state was called a state-dependent weak hydrogen bond. When comparing two crystallographic structures, a hydrogen bond was considered state-dependent when formed in one structure but not in the other even when using the larger distance cutoff. A complete list of state-dependent backbone hydrogen bonds along with their frequencies is presented in Supporting Information Tables S1 and S2 and summarized in Table II. In the case of the hydrogen bond analysis between mannose and the lectin domain, no distinction was made which hydrogen atom was donated whenever a donor had multiple protons (for example, the N-terminal amide group of Phe^{1}). Similarly, with hydrogen bonds involving a side-chain with two oxygen atoms (for example, Asp^{54}) no distinction was made which of the two was the acceptor.

Determination of side-chain contacts

Side chain contacts were determined between the protein and mannose, and between key residues in the binding pocket. A side-chain contact was defined to be formed if the distance between the centers of geometry of two side chains, or of a side-chain and mannose was less or equal 6 Å. A side-chain contact was considered persistent during a trajectory if observed in at least 66% of the simulation frames.

RESULTS

Comparison of backbone flexibility

Backbone flexibility has often been linked to biomolecular function, in particular when binding a target.[33,34] In the case of the lectin domain of FimH there are two regions with key functions: the inter-domain regulatory region that binds the pilin domain and the distal mannose binding site. Comparing the flexibility of these regions between the two states can give clues about how the docking or separation of a ligand can initiate conformational changes that propagate across the protein. For this purpose, two 300-K MD simulations were performed with the low and two with the high affinity state of the lectin domain (Table I). In both cases, for computational efficiency, the lectin domain was simulated in isolation, that is, no other domains of the fimbrial tip were present. This might introduce artefacts in the simulations of the low affinity state where the two domains of FimH contact each other and removal of the pilin domain will result in the solvent exposure of otherwise buried residues of the lectin domain. However, the relatively short simulation time of 50 ns is not expected to cause significant conformational changes as indicated by the relatively low total C_α root mean square deviation (RMSD) from the initial structure in all four simulations [Fig. 2(a)].

Identification of segments that change conformation

In order to analyze structural differences between the two states, it was necessary to first identify the regions of the lectin domain that undergo significant conformational changes when switching between one state and another. For this purpose, the low and the high affinity structure were aligned and the following segments were identified to have significant conformational differences (see caption of Figure 1(c) for a detailed description of the method): clamp segment (residues 2–22), swing segment (residues 23–32), bulge-helix segment (residues 60–72), insertion segment (residues 112–124) and linker

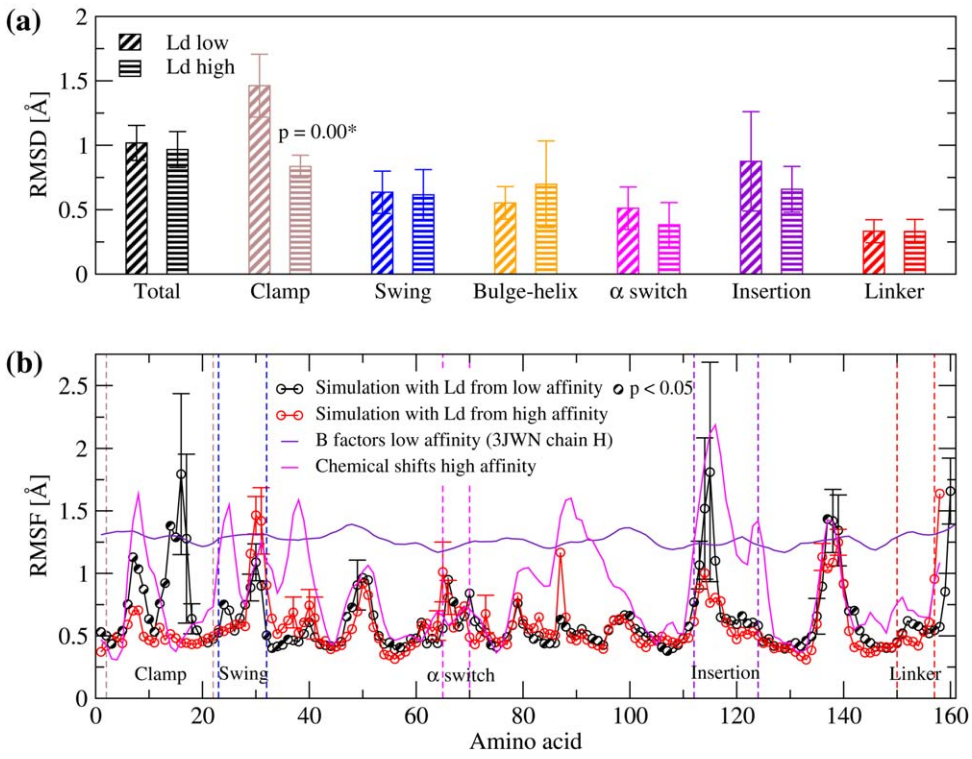

Figure 2

Comparison of backbone flexibilities at room temperature. (**a**) The C_α RMSD from the initial conformation averaged over two 300-K simulations of each the low and the high affinity state (only frames from the last 40 ns of in total 50-ns trajectories were used). The RMSD is calculated for the total length of the lectin domain except for the terminal amino acids (residues 2 − 157) and for the segments identified to present significant conformational differences between low and high affinity state [Fig. 1(c)]. Error bars indicate the standard deviation of the RMSD along the trajectories. *For the clamp segment, the difference in RMSD was statistically significant (the *p*-value from a one-tailed Student's *t* test is indicated), while the difference in standard deviation was marginally statistically significant with a *p*-value of 0.08. (**b**) Average C_α RMSF calculated from two simulations with the low and two simulations with the high affinity state. From each simulation, only the last 40 ns of in total 50-ns simulation time were used. The RMSF was calculated for each simulation independently and then the values of two runs started from the same state were averaged. The error bars indicate the standard error of the mean. Segments that switch conformation are indicated with vertical dashed lines. Semi-filled circles indicate that the difference is statistically significant according to a one-tailed Student's *t* test (for simplicity, only values of the low affinity state are labeled). Values of RMSF derived from crystallographic B factors of the low affinity state structure were calculated using the formula $\mathrm{RMSF}_i^{\exp} = \sqrt{\frac{3}{8\pi^2} B_i}$, where B_i is the B factor of C_α of residue *i*. Chemical shifts measured for the high affinity state[35] and deposited in the BMRB database[60] were used to derive RMSF values by utilizing the Random Coil Index server.[61,62] [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

segment (residues 150–157). Contained within the bulge-helix segment is the so called α switch (residues 65–70) which is also monitored separately because it switches from a 3_{10} helix in low affinity to a α helix in high affinity [Fig. 1(c)] and for consistency with previous publications.[10,15] The clamp segment is located in part within the mannose binding pocket, whereas swing, insertion and linker segments are part of the inter-domain regulatory region and contain residues that contact the pilin domain in the low affinity state of FimH [Fig. 1(a)]. According to analysis of the C_α root mean square fluctuations (RMSF), major differences in flexibility were observed in the clamp segment (mannose binding site) and in the inter-domain region [Fig. 2(b)]. For this reason, the discussion below is limited to these areas.

Flexibility of the mannose binding site

The clamp segment, which is part of the mannose binding pocket, presented a relatively higher mobility in the low affinity state and sampled a vaster conformational space as indicated by the larger RMSD from the initial structure and a larger standard deviation [Fig. 2(a)]. This was consistent with generally larger C_α RMSF values in the low affinity conformation of the clamp segment [Fig. 2(b)]. It is worth noting that fluctuations derived from the crystallographic B values show less mobility for the clamp segment in low affinity than in the simulations [indigo line in Fig. 2(b)]. This discrepancy is probably due to intra-unit crystallographic contacts between two copies of the fimbrial tip leading to lower mobility in the crystalline state than in solution. The rigidity of the clamp segment in the high affinity

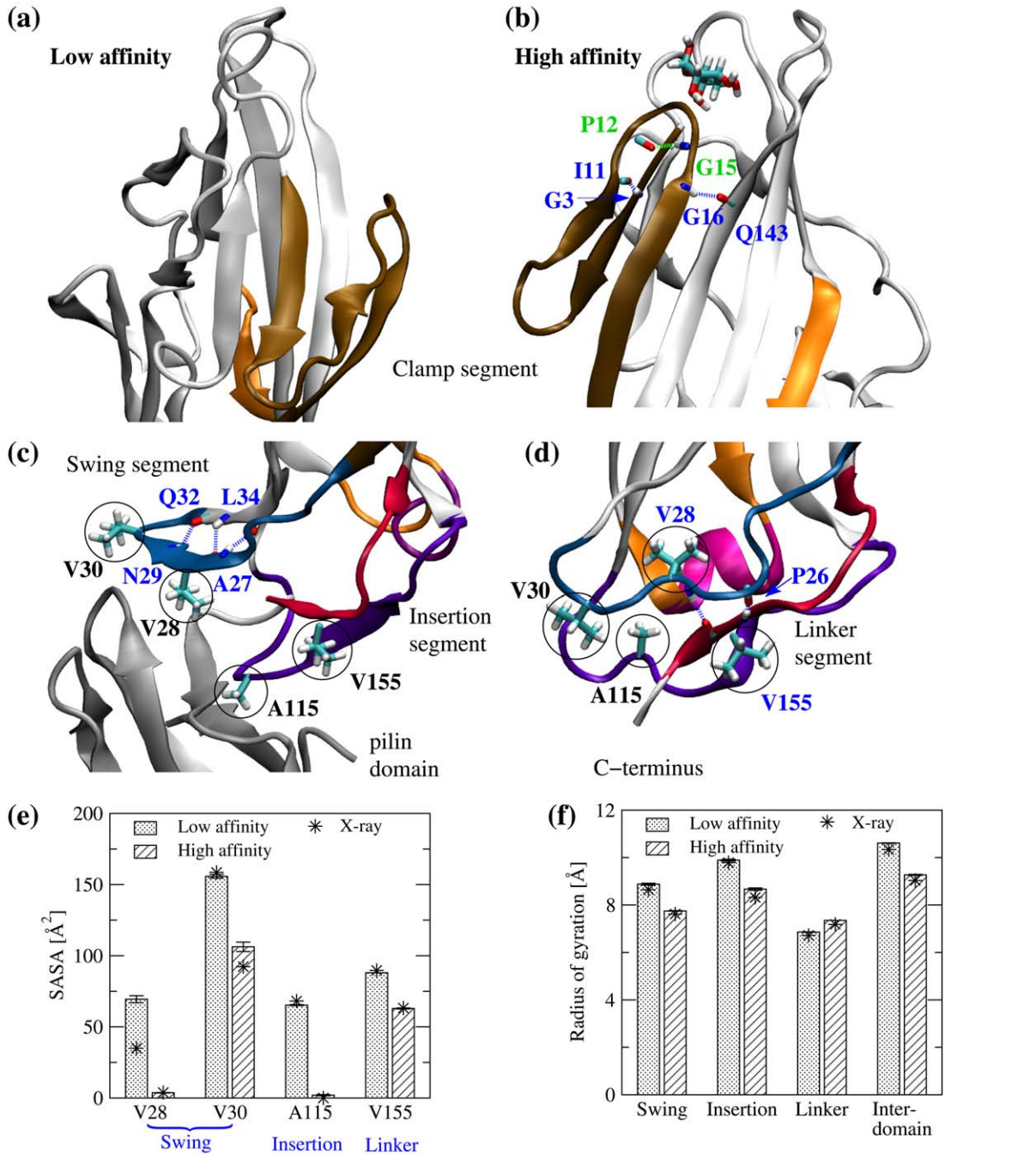

Figure 3

Structural differences in the mannose binding site and the inter-domain region between low and high affinity state. (**a**,**b**) The mannose binding site in the low and the high affinity state, respectively, showing hydrogen bonds in the clamp segment formed only in high affinity. Hydrogen bonds are indicated by blue dashed lines, except for one weak hydrogen bond which is colored in green. Residue names are also colored accordingly. (**c**, **d**) Conformation of the inter-domain segments in the low and high affinity state, respectively, highlighting state dependent backbone hydrogen bonds involving the swing segment and hydrophobic side chains (surrounded by circles) contacting the pilin domain when docked. Names of residues involved in hydrogen bonds are colored in blue. (**e**) SASA of hydrophobic residues that contact the pilin domain. The labels below the x axis indicate in which segments the amino acids are located. (**f**) Radius of gyration of inter-domain segments. The values are averages over two simulations and the error bars are standard errors of the mean. The differences are statistically significant with p-values smaller than 0.01 according to one-tailed Student's t-tests. Values calculated from the crystallographic structures are also included and indicated with stars. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

state is consistent with the presence of three state dependent (of which one is weak) backbone hydrogen bonds [Fig. 3(a,b), Table II and Supporting Information Tables S1 and S2; see "Materials and methods" for the definition of (weak) state dependent hydrogen bonds]. The flexible and open configuration of the clamp segment in the low affinity state likely facilitates docking of the ligand before tensile force activates FimH and the pocket is tightened.

Flexibility of the inter-domain region

Interestingly, the clamp segment is adjacent in primary sequence to the swing segment, which undergoes the largest displacement between low and high affinity state [Fig. 1(c)]. In low affinity, the swing segment forms a β hairpin stabilized by backbone hydrogen bonds [Fig. 3(c), Table II and Supporting Information Table S1] that swings away from the linker segment allowing the pilin domain to wedge itself between the two segments [Fig. 3(c)]. On the other hand, in high affinity, the swing segment takes a disordered conformation that is located closer to the linker segment with which it forms two hydrogen bonds [Fig. 3(d), Table II and Supporting Information Table S2]. This is consistent with its slightly higher mobility than in low affinity [Fig. 2(b); out of in total 10 residues in the swing segment, five had higher mobility with $p < 0.10$ and two with $p < 0.05$, according to a one-tailed Student's t-test]. The mobility of the swing segment and, in general, the inter-domain region is evident also from RMSF values calculated from chemical shifts of the high affinity state[35] [Fig. 2(b)]. In fact, the backbone flexibility observed in the simulations is consistent with the mobility derived from the experiments for most of the protein, except for segments in the inter-domain region and a region adjacent to the binding pocket (residues 90–100), which present a much larger mobility in the experiments [Fig. 2(b)]. This discrepancy might be due to limited conformational sampling in the simulations. The flexibility of the inter-domain region in high affinity is probably crucial to allow docking of the pilin domain and the conversion to low affinity.

Conformational changes in the inter-domain region

Conformational differences in the inter-domain region between the low (pilin domain docked) and high (pilin domain separated) affinity state have been experimentally shown to be associated with the allosteric pathway that regulates the function of the lectin domain.[10,15] In order to understand what drives the conversion of the inter-domain region from one conformation to another, it is necessary to physically characterize the conformational differences between the two states. A previous analysis of inter-domain contacts found that in total four hydrophobic residues in the inter-domain region contact the pilin domain [Fig. 3(c,d)].[13] Analysis of the solvent accessible surface area (SASA) along the trajectories revealed that all four amino acids presented a lower SASA in the high affinity than in the low affinity structure of isolated lectin domain [Fig. 3(e)]. Another important observation is that in the high affinity state all inter-domain segments presented a relatively more compact structure as indicated by the radius of gyration [Fig. 3(f)]. Notably, the observed differences were consistent between averages from the simulations and values measured from the crystallographic structures [Fig. 3(e,f)]. Taken together, these measurements suggest that after separation from the pilin domain, structural rearrangements occur aimed at achieving a more compact structure of the inter-domain region. At the same time, hydrophobic residues that are partially or completely buried in high affinity serve as hotspots for binding to the pilin domain.

Affinity for mannose: low versus high affinity state

Comparison of the crystallographic structures shows that the clamp segment in the binding pocket is in an open and rather unstructured conformation in the unbound low affinity state [Fig. 1(a)], and in a closed and structured conformation when the lectin domain is in the high affinity state and bound to mannose [Fig. 1(b)]. However, it is not clear how the clamp segment modulates the binding affinity to mannose. In order to understand this, two simulations were performed with the lectin domain in the low affinity state after mannose was docked in the binding site. The location of mannose in the binding site was determined through superimposition to the high affinity structure where mannose is present in the crystal (see "Materials and Methods" for a detailed description). In both 50-ns simulations performed with the obtained complex, mannose remained docked to the protein, allowing a comparison of persistent contacts formed between the protein and the ligand with those observed in the runs with the high affinity structure (Ld_high_1,2).

Hydrogen bond analysis

In total six hydrogen bonds between mannose and the lectin domain were observed to be persistent (see "Materials and Methods" for the definition of a persistent hydrogen bond) in both simulations with the high affinity state [see Fig. 4(a,b) and Table III and a previous publication by us[15]]. In contrast, the simulations of the low affinity state in complex with mannose presented only five persistent hydrogen bonds, which are a subset of the six observed in high affinity. The hydrogen bond Gln^{133} $N_{\epsilon}H\ldots$ mannose O3, which is persistent in high affinity, is formed in less than 20% of the frames in each of the two simulations with low affinity [Fig. 4(c,d)]. The fact that five out of six hydrogen bonds remained formed in almost every frame throughout the

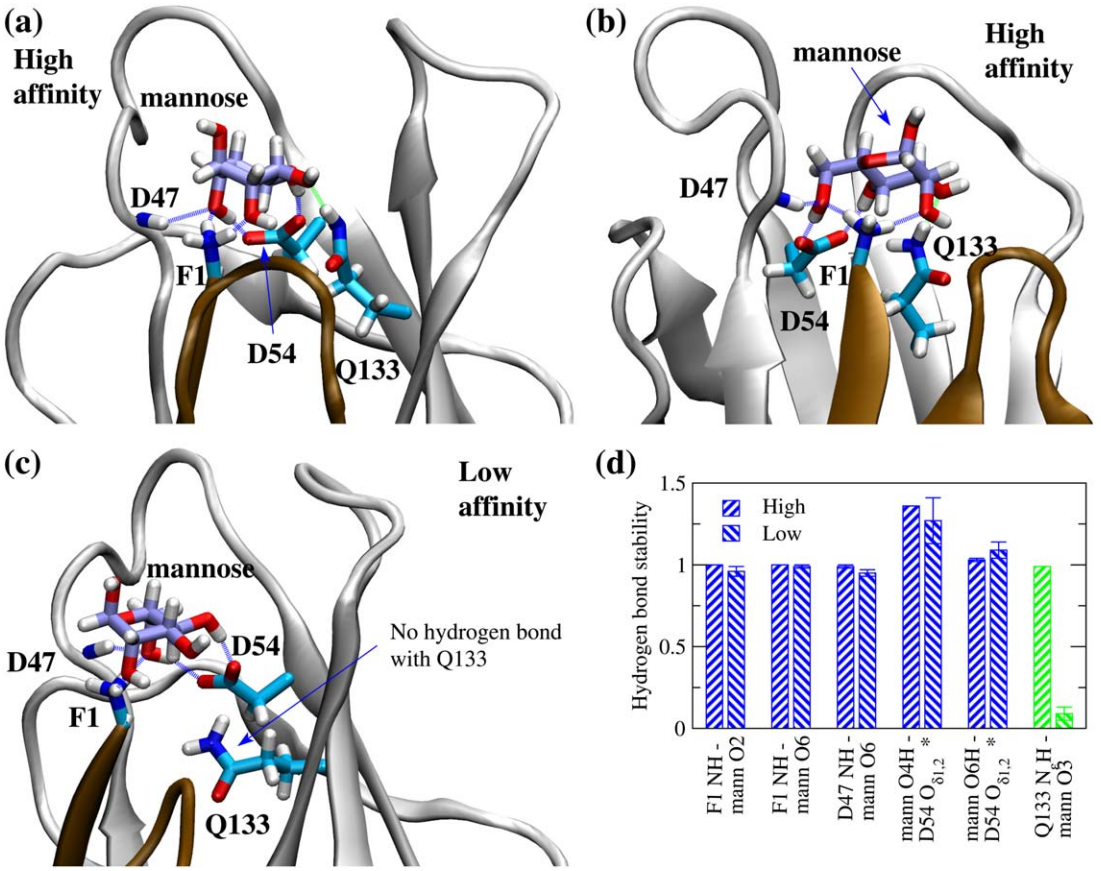

Figure 4

Hydrogen bonds between the lectin domain and mannose. (**a, b**) Two different angles of the mannose binding pocket in high affinity (snapshot after 14 ns of the run Ld_high_1). (**c**) Mannose binding pocket in low affinity highlighting that the hydrogen bond with the Gln^{133} side-chain is not formed (snapshot after 20 ns of the run Ld_low_mann_1). (**d**) Ratio of frames where the respective hydrogen bond is seen to be formed (the values are averages over two simulations and the error bars denote standard errors of the mean). *Values are larger than 1 because both oxygen atoms might form a hydrogen bond with the same donor at the same time. All hydrogen bonds persistent in the simulations of both, the high and the low affinity state, are highlighted by blue dashed lines and their stability is reported with blue bars in the plot. The hydrogen bond with the Gln^{133} side-chain, which is stable only in high affinity is colored in green in both, the graphical representations and the bar plot (the difference is statistically significant). Mannose and the functional groups involved in hydrogen bonds are shown in the stick and ball representation (the carbon atoms of mannose are colored in purple to facilitate distinction). The backbone of the protein is colored in white except the clamp segment which is shown in brown. The definition of a persistent hydrogen bond is given in "Materials and Methods". [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

simulations with docked mannose suggests that the ligand was probably positioned correctly. Also, during the docking protocol, the salt bridge between the positively charged N-terminus and the side-chain of Asp^{54} ruptured. A previous computational study suggested that rupture of the salt bridge between the N-terminus and Asp^{54} is necessary in order for mannose to bind correctly.[36] In fact, this salt bridge was observed to be persistent only in the simulations with the low affinity state without mannose (Table IV).

Hydration of the binding pocket

It is necessary to understand how the open conformation of the binding pocket in low affinity causes the lack of a stable hydrogen bond between mannose and the side-chain of Gln^{133}. For this reason, the solvent exposure of mannose and of contacting residues was analyzed in both, the low and the high affinity state [see Fig. 5(a,b) for a side to side comparison of the binding site in both states with mannose]. Mannose presented a remarkably larger SASA in the low than in the high affinity state [Fig. 5(c)]. This is probably mainly due to the loss of a side-chain contact between mannose and Ile^{13} located in a turn of the clamp segment that swings outward in low affinity [see Fig. 5(a,b) and Table III]. The side chains of Asp^{54} and Gln^{133}, which both form hydrogen bonds with mannose in high affinity, are completely buried in both states according to their SASA values [Fig. 5(c)]. Nevertheless, they still engaged in hydrogen bonds with water molecules in the low affinity state simulations [Fig. 5(d)]. This indicates that in low affinity a

Table III
Contacts between Mannose and the Lectin Domain and Side Chain Contacts between Q133 and the Rest of the Protein

	Ld_high_1	Ld_high_2	Ld_low_mann_1	Ld_low_mann_2
Hydrogen bond protein-mannose				
F1 NH - mannose O2	1.00	1.00	0.93	0.98
F1 NH - mannose O6	1.00	1.00	0.98	1.00
D47 NH - mannose O6	0.98	0.99	0.97	0.93
mannose O4H - D54 $O_{\delta 1,2}$[a]	1.36	1.36	1.13	1.40
mannose O6H - D54 $O_{\delta 1,2}$[a]	1.02	1.04	1.13	1.04
Q133 $N_{\epsilon}H$ - mannose O3	0.99	0.99	0.13	0.05
Side chain contact protein-mannose[b]				
I13 - mannose	0.71	0.81	0.00	0.00
D47 - mannose	1.00	1.00	1.00	0.99
Y48 - mannose	0.72	0.81	0.73	0.68
I52 - mannose	1.00	1.00	1.00	1.00
D54 - mannose	1.00	1.00	1.00	1.00
Q133 - mannose	1.00	1.00	0.17	0.10
N135 - mannose	0.98	0.98	0.96	0.71
Side chain contact Q133 - protein[b]				
Q133 - F1	1.00	1.00	1.00	1.00
Q133 - I13	0.00	0.00	0.82	0.37
Q133 - D54	1.00	1.00	0.96	1.00
Q133 - V56	0.42	0.38	0.88	0.96
Q133 - N135	1.00	1.00	0.32	0.33
Q133 - F142	0.99	0.88	0.99	0.86
Q133 - F144	0.92	0.90	0.93	0.83

The values indicate the ratio of simulation frames where a contact is formed.
[a]Since the side-chain of Asp^{54} can flip, no distinction was made which oxygen atom was making a hydrogen bond with mannose, leading to values larger than 1 for the frequency of hydrogen bond formation (see "Materials and methods").
[b]See "Materials and methods" for the definition of a side-chain contact.

channel might be present in the binding pocket through which water molecules can infiltrate and interfere with hydrogen bonds between protein and ligand.

In order to understand how water can penetrate into the binding pocket, side-chain interactions between Gln^{133} and vicinal residues were determined. Persistent contacts were observed between the side-chain of Gln^{133} and the phenyl rings of Phe^{1}, Phe^{142}, and Phe^{144} in both simulations with the high and both simulations with the low affinity state [Fig. 5(a,b) and Table III]. However, in the low affinity runs, Phe^{142} presented a relatively larger SASA [Fig. 5(c)] and the side chains of Phe^{142} and Phe^{144} were observed to be more flexible than in the runs with the high affinity state. The increased mobility of Phe^{142} and Phe^{144} is likely related to two major topological differences between the high and the low affinity state. One difference is due to the open configuration of the clamp segment in low affinity, which causes the phenyl ring of Phe^{142} and of Phe^{144} to lose contacts that are formed in high affinity with the backbone CH2 groups of Gly^{14} and Gly^{15}, respectively [visual analysis and Fig. 5(a,b)]. The other topological difference is that the C_{α} atoms of Phe^{142} and Phe^{144} are located at a greater distance from the C_{α} atom of Phe^{1} in low affinity [Fig. 5(e)], thus indicating an overall wider binding pocket than in high affinity. Furthermore, in low affinity, the side-chain of Gln^{133} loses a contact with Asn^{135} that is observed in high affinity [Fig. 5(a)] and forms instead a contact with Val^{56} located distally from the clamp segment [Fig. 5(b) and Table III]. These observations, that is, a wider binding pocket, an increased mobility of the phenylalanine side chains and a different placement of

Table IV
Contacts in the Binding Pocket that Differ between Bound and Unbound States

	Ld_high_1	Ld_high_2	Ld_low_mann_1	Ld_low_mann_2	Ld_low_1	Ld_low_2
Electrostatic contacts						
N-term - D54*	0.00	0.00	0.02	0.16	0.89	1.00
Side chain contacts						
Y48 - I52	0.81	0.86	0.97	0.88	0.37	0.15
I52 - Y137	0.98	1.00	0.97	0.99	0.68	0.28

Ratio of simulation frames where a contact is formed. A contact is considered persistent if observed in at least 66% of the simulation frames. *The salt bridge between the positively charged N-terminus and the side-chain of Asp^{54} is considered formed if the distance between the backbone nitrogen of Phe^{1} and the C_{γ} atom of Asp^{54} is not larger than 4 Å.

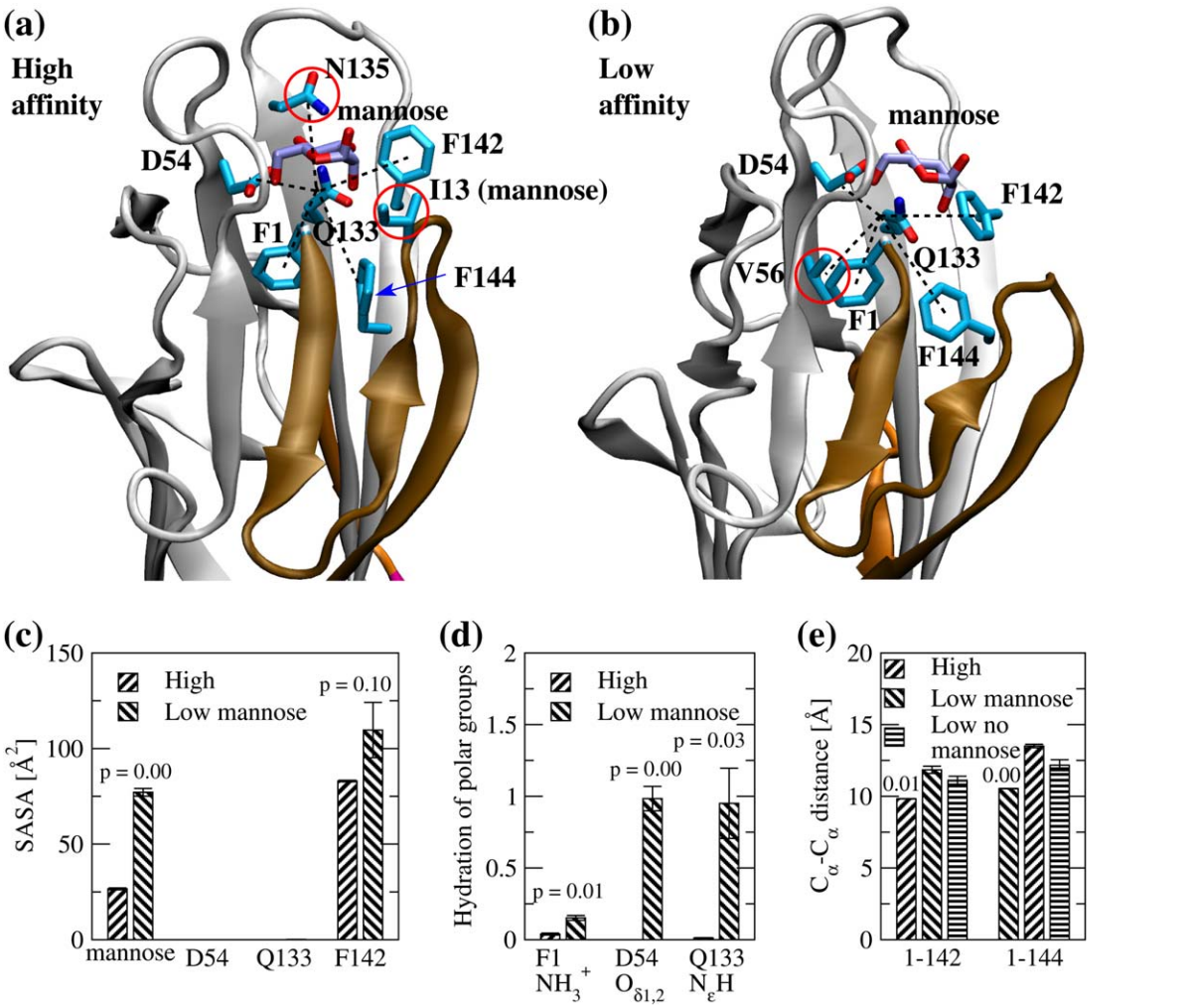

Figure 5

Side chain interactions and hydration of the mannose binding pocket. (**a**) Side chains undergoing persistent contacts with Gln^{133} in both high affinity simulations (dashed lines). Asn^{135} is marked by a red circle to highlight that it forms a persistent contact only in the high affinity state. Although it does not interact with Gln^{133}, the side-chain of Ile^{13} is also indicated because it undergoes persistent contacts with mannose only in the high affinity state (Table III). (**b**) Side chains undergoing persistent contacts with Gln^{133} in both low affinity simulations (dashed lines). Val^{56} is marked by a red circle to indicate that it forms a persistent contact only in the low affinity state (Table III). (**c**) SASA of mannose and of side chains that form hydrogen bonds with it. The SASA of Phe^{142} is also indicated because it contacts Gln^{133} (the SASA of other side chains contacting Gln^{133} did not change more than 1 Å^2 between the two states and thus it is omitted). (**d**) Ratio of frames where a specific functional group that is involved in hydrogen bonds with mannose forms hydrogen bonds with water molecules. (**e**) Average distance of C_α atoms of hydrophobic residues contacting Gln^{133} from the N-terminal C_α atom. In all plots, the reported values are averages over two simulations with each state while the error bars indicate standard errors of the mean. Reported are also *p*-values from one-tailed Student's *t*-tests. In the last plot, the *p*-value refers to the difference between high and low affinity state with mannose. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Gln^{133}, suggest that the low affinity state allows water molecules to penetrate and hydrate groups involved in hydrogen bonding with mannose [illustrated using two snapshots from the low affinity simulations in Fig. 6(a,b)].

Tyrosine gate

In both, the low and high affinity state, mannose is shielded from the solvent also through contacts with the side chains of Tyr^{48} and Ile^{52} located at the outer edge of the binding pocket [Fig. 6(c,d) and Table III]. These interactions are stabilized by the fact that Ile^{52} is sandwiched between Tyr^{48} and Tyr^{137}, with which it forms stable contacts [Fig. 6(c,d) and Table IV]. In contrast, in the absence of mannose, the contacts between Ile^{52} and the two tyrosine residues were formed much less frequently [Fig. 6(d)], and in particular the side-chain of Tyr^{48} sampled a vast conformational space (visual analysis). Because of the ability of Tyr^{48} to adapt to a ligand

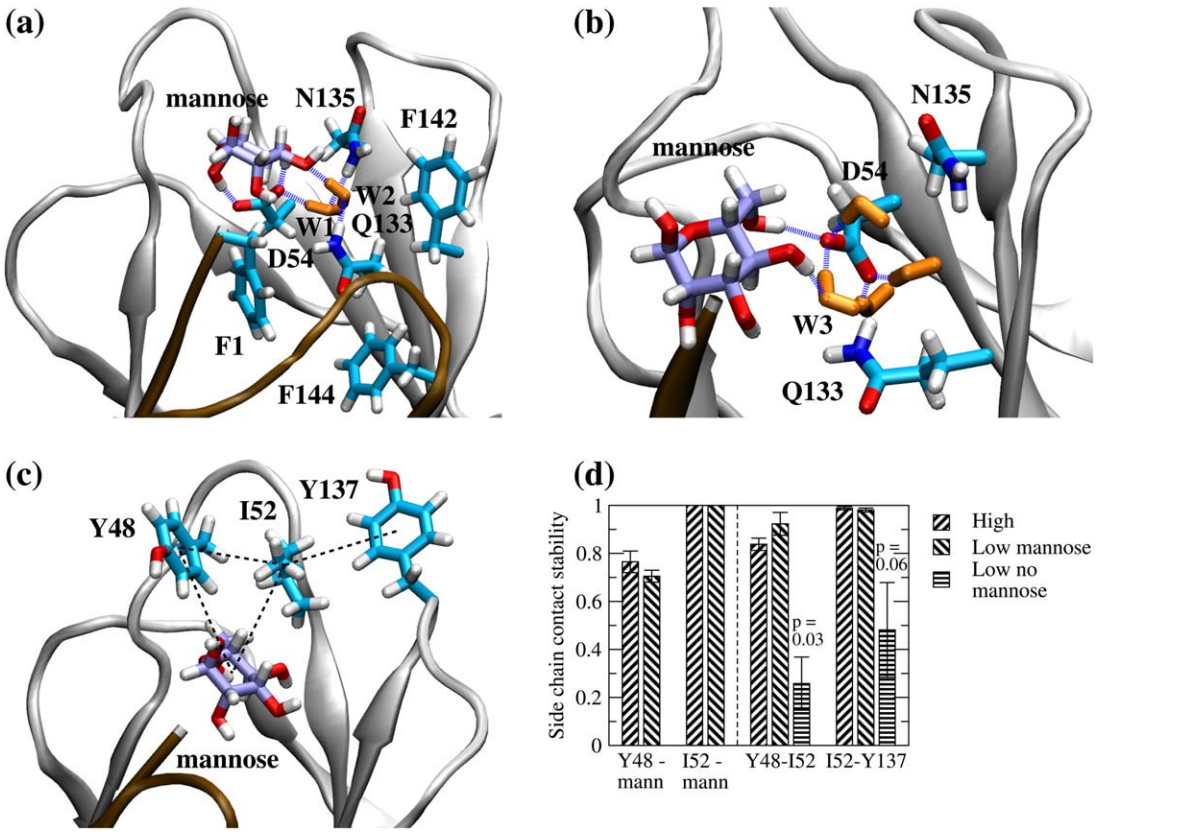

Figure 6

Water penetration into the binding pocket and tyrosine gate. (**a**) Water molecules forming hydrogen bonds with the side-chain of Asp^{54} or the amine group of the Gln^{133} side-chain in the snapshot saved after 29.36 ns of the run Ld_low_mann_1. The side chains that interact with Gln^{133} in high affinity are also displayed. In this snapshot, one water molecule, W1, interacts with Asp^{54} and Asn^{135} side chains, and another water molecule, W2, interacts with mannose and the side-chain of Asn^{133}. Hydrogen bonds are indicated through blue dashed lines and water molecules are colored in orange for better recognition. (**b**) Water molecules forming hydrogen bonds with the side-chain of Asp^{54} or the amine group of the Gln^{133} side-chain in the snapshot saved after 21.14 ns of the run Ld_low_mann_2. One water molecule, W3, forms hydrogen bonds with both, mannose and one of the oxygens of the Asp^{54} side chain. This leads to the loss of one hydrogen bond between Asp^{54} and mannose. Three other water molecules (not labeled) also form hydrogen bonds with the Asp^{54} side chain. Visual analysis shows that the water molecules penetrate the binding pocket through a gap between Phe^{142} and the backbone of the clamp segment and a gap between Gln^{133} and Asn^{135}. (**c**) Residues involved in the tyrosine gate. Side chain contacts observed to be persistent (that is, formed in at least 66% of the frames) in simulations with mannose bound are indicated with dashed lines. (**d**) Ratio of frames where side-chain contacts are formed. The vertical dashed line separates contacts with mannose from contacts within the tyrosine gate. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

and to "open" in the absence of it, the Tyr^{48}-Ile^{52}-Tyr^{137} configuration has been called in the literature the "tyrosine gate".[37,38]

Correlated motions in μs-long simulations

A previous analysis comparing the crystallographic structures of high and low affinity state described that the allosteric conformational change is propagated through the mechanical twist of a continuous large β sheet of the lectin domain [Fig. 7(a)].[10] In particular, transition to low affinity upon docking of the pilin domain involves the tilting of β17-22 which is part of the clamp segment and is located at the edge of the large β sheet (the nomenclature "βxx-yy" indicates a β strand from residue xx to residue yy). However, from the crystallographic structures alone it is not clear whether the pilin domain induces the conformational switch or rather stabilizes a conformation that is accessible in its absence but less populated. In order to investigate this, two multi-μs simulations starting from the low affinity state and two μs-range simulations starting from the high affinity state of the isolated lectin domain were run on Anton, a specialized supercomputer, at a temperature of 330 K (Table I). The slightly higher temperature is used to accelerate sampling. The analysis presented here focuses first on the simulations started from the low affinity state since the absence of the pilin domain (which is docked to the lectin domain in low affinity) likely allows for greater flexibility of the inter-domain region and overall a vaster conformational sampling.

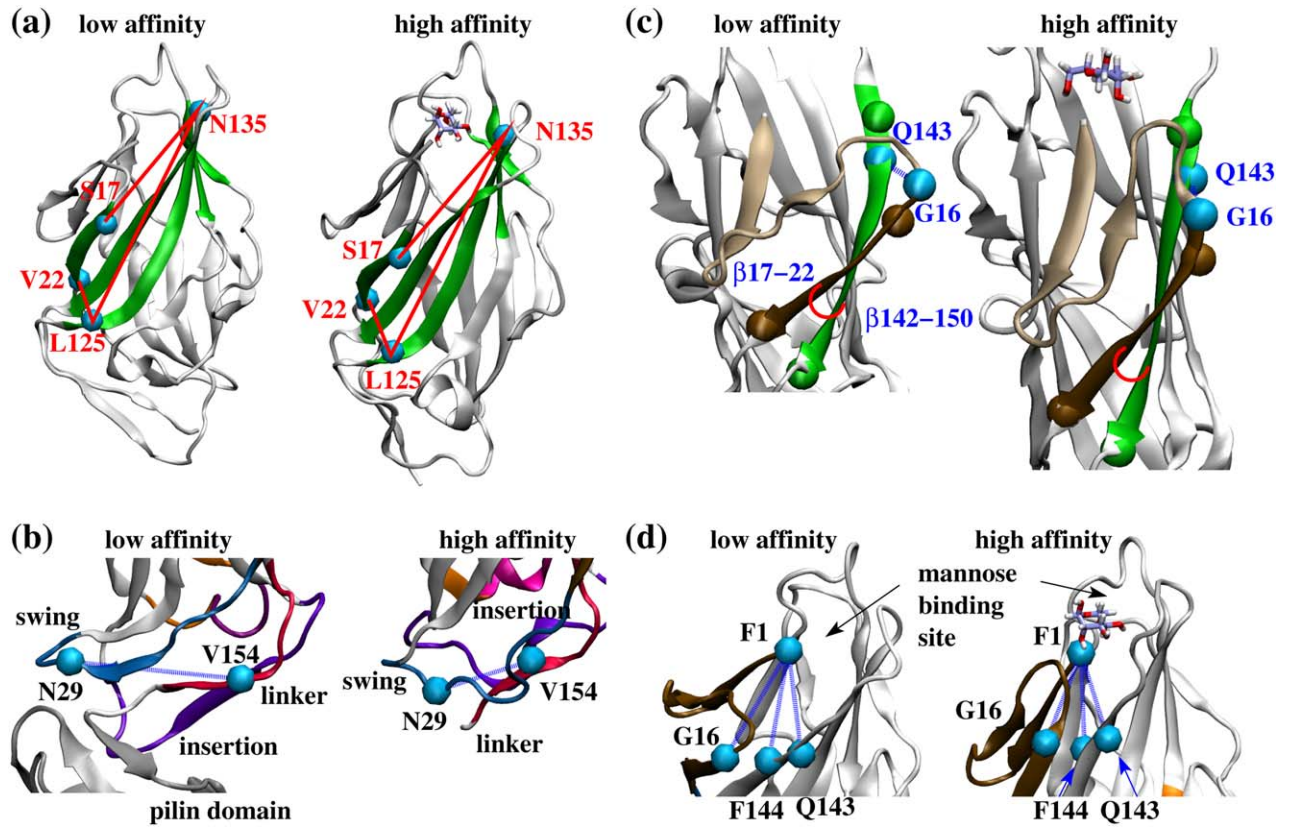

Figure 7

Illustration of angles and distances observed to be correlated in the Anton trajectories. (**a**) Dihedral angle (highlighted in red) defined by the C_α atoms of residues Ser[17], Asn[135], Leu[125], and Val[22] in the large β sheet (colored in green). (**b**) Segments in the inter-domain region used for the calculation of the radius of gyration and distance (highlighted by a blue dashed line) between Asn[29] in the swing segment and Val[154] in the linker segment. Part of the pilin domain is displayed for reference although it was not present in the simulations. (**c**) Angle (highlighted by red semi-circles) between the two adjacent β strands β17-22 and β142-150. (**d**) Distances of three C_α atoms located near the mannose binding site from the C_α atom of Phe[1] (highlighted by blue dashed lines). The average value of these three distances is used for the correlation calculations. All quantities were chosen in such a way that their values are smaller in the high than in the low affinity state in order to facilitate the interpretation of the correlations. Angles and distances are shown for the low (left side of each panel) and the high (right side of each panel) affinity structure. The location of C_α atoms is indicated by spheres. Except in (a), the backbone is colored to highlight segments that change conformation between the two states as described in Figure 1. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Flexibility of the large β sheet

In order to quantify the twist of the large β sheet, the dihedral angle defined by the C_α atoms of residues 17, 135, 125, and 22 was measured along the trajectories [Fig. 7(a)]. The reason for selecting this particular dihedral is that while β17-22 undergoes a significant tilt between the two states, β125-135 does not undergo any major shifts with respect to the rest of the protein [Fig. 1(c)] and presents low fluctuations [Fig. 2(b)] thus constituting a reasonable reference to describe the torsional properties of the large β sheet. Time series calculated from the runs with the low affinity state indicated that the dihedral angle is rather flexible and sampled values similar to both the low and the high affinity state [Fig. 8(a) and Supporting Information Figure S1] indicating that the large β sheet is rather flexible. To corroborate whether the twisting motion of the large β sheet is correlated to movements in other parts of the protein, key distances and angles in the inter-domain region and the binding pocket were monitored along the trajectories [Fig. 8(a)]. The Pearson's linear correlation coefficients were then calculated among all six monitored quantities (including the dihedral angle in the large β sheet) and plotted as a matrix for the combined simulations with the low affinity state [Fig. 8(b)] and for the combined simulations with the high affinity state [Fig. 8(c)].

Correlation between the large β sheet and the inter-domain region

A weak to moderate correlation was observed between angles in the β sheet [Region II in Fig. 8(b)] and two measures in the inter-domain region (Region I): the radius of gyration, which was smaller in the high than in the low affinity state [see Fig. 7(b) and Section "Comparison of backbone

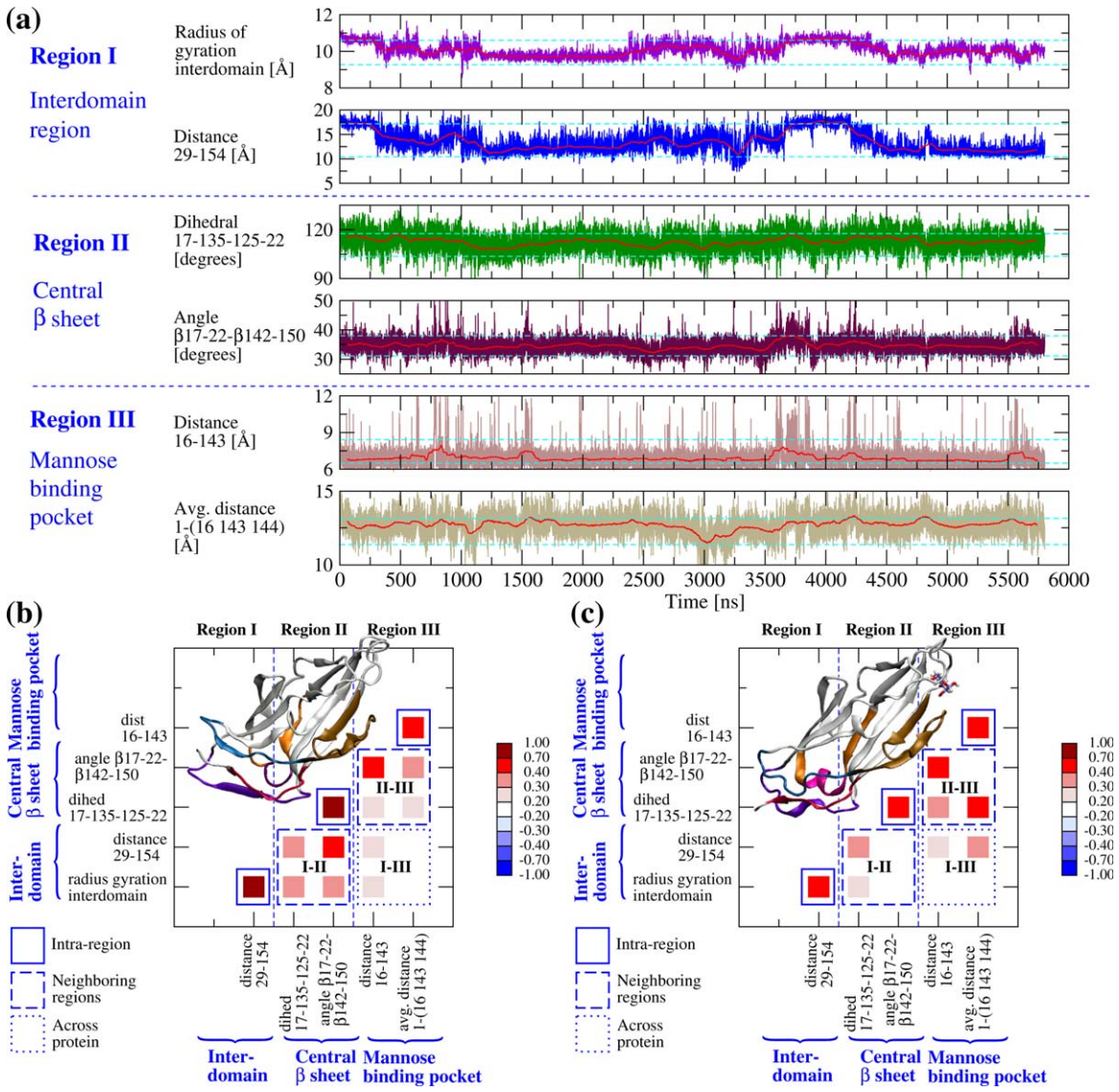

Figure 8

Correlation between key angles and distances measured along the Anton trajectories. (**a**) Time series of key angles and distances (Fig. 7) measured along the Anton simulation started from the low affinity state, Anton_low_330K_1. The red lines are time averages over a 120-ns time window. The cyan dashed lines in each plot indicate average values obtained from 300-K simulations with the low affinity state (larger value) and the high affinity state (smaller value). The averages were calculated from the last 40 ns of two in total 50-ns long simulations with either the low or the high affinity state (Table I). (**b**) Pearsons's linear correlation coefficients between key angles and distances in runs with the low affinity state. The values were obtained after merging the trajectories of two Anton simulations with the low affinity state (Anton_low_330K_1,2) and averaging the time series with a 120-ns time window in order to eliminate noise. The program xmgrace was used to calculate the correlations. The upper half of the plot contains the structure of the lectin domain in the low affinity state. (**c**) Similar as in (b), but using the Anton simulations with the high affinity state (Anton_high_330K_1,2).

flexibility"], and the distance between Asn^{29} in the swing segment and Val^{154} in the linker segment [Figs. 7(b) and 8(b)]. Residues Asn^{29} and Val^{154} are part of the docking site of mAb21, a monoclonal antibody that recognizes only the high affinity state of FimH.[10] Interestingly, visual analysis of the FimH structure in low affinity shows that the pilin domain is wedged between the swing and linker segments, which prevents the two aforementioned residues from moving closer to each other and thus stabilizes the swing segment in its lifted position [Fig. 7(b)]. It is likely that the absence of the pilin domain from the simulations of the low affinity state allowed a larger flexibility of the swing segment.

Correlation between the large β sheet and the binding pocket

The next question is how flattening or twisting of the large β sheet is related to the closing or opening of the

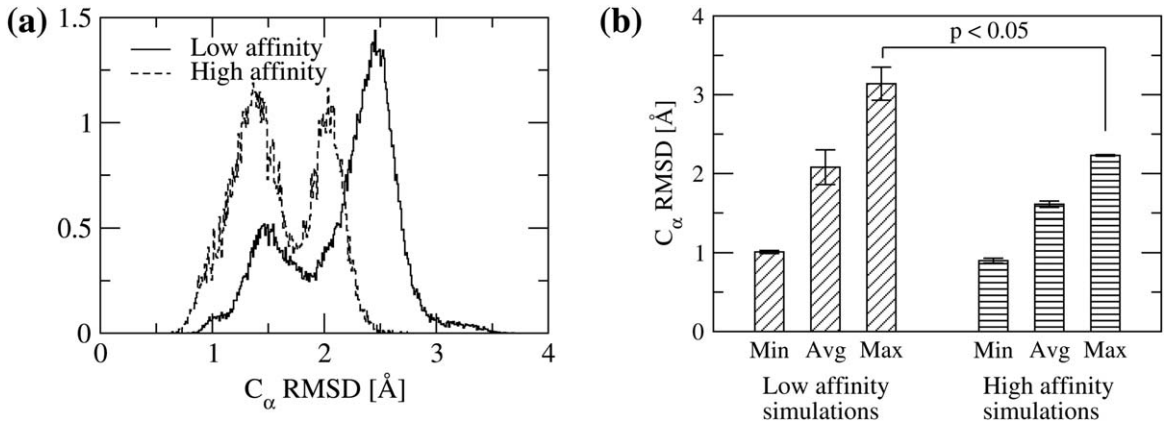

Figure 9

Conformational sampling of the lectin domain in the Anton simulations. (**a**) Normalized histograms of the C_α RMSD (residues 2 − 157) from the initial conformation measured along the Anton simulations with the low and the high affinity state, respectively (Supporting Information Figures S4-S7). (**b**) Minimum, average, and maximum values of the C_α RMSD determined after averaging the time series over a 12-ns time window. Reported are averages over two simulations and the error bars denote standard errors of the mean. The *p*-value between maximum values was calculated using a one-tailed Student's *t* test.

clamp segment and tightening or losening, respectively, of the mannose binding site. A strong correlation was observed between the dihedral angle and the angle between β17-22 strand and neighboring β142-150 [Region II in Fig. 8(b).] The decrease of this angle brings residues Gly^{16} and Gln^{143} closer to each other, which form a backbone hydrogen bond in the high affinity state locking the clamp segment in the tighter conformation [Fig. 7(c)]. In fact, the angle between the two β strands and the distance between these two residues was found to be moderately correlated [Square II-II in Fig. 8(b)]. A weak correlation was also observed between the angle and the average distance between residue Phe^{1} and residues Gly^{16}, Gln^{143}, and Phe^{144} [Square II-III in Fig. 8(b)]. Shortening of these distances tightens the pocket [Fig. 7(d)], which in Section "Affinity for mannose: low versus high affinity state" was shown to prevent water penetration. Interestingly, a very weak correlation was observed between the inter-domain region and one of the key distances in the mannose binding pocket (Square I-III in Fig. 8). These observations suggest that motions in the inter-domain region and in the binding site are coupled with the motion of the β sheet while a loose coupling exists between the two distal sites.

Spring-like characteristics of the large β sheet

Simulations started from the high affinity state also presented similar correlations as the runs with low affinity, although they were generally weaker [Fig. 8(c) and Supporting Information Figures S2 and S3]. This discrepancy is consistent with the fact that the simulations started from the low affinity state generally sampled a vaster conformational space than the runs with the high affinity state as suggested by the distribution of the RMSD from the initial structure [Fig. 9(a,b) and Supporting Information Figures S4-S7]. This could be due to the longer simulation time of the runs with low affinity and to the removal of the pilin domain, which is wedged between the swing and linker segments. It is plausible that the absence of the inhibitory pilin domain from the runs with the low affinity state unlocked the swing segment from its lifted position [Fig. 7(b)], which in turn facilitated an increased mobility of the large β sheet [Fig. 8(a)]. In fact, in the simulations started from the low affinity state, the dihedral angle significantly sampled values representative of the high affinity state [Fig. 10(a)] while its mean was located between the average values calculated from shorter 300-K simulations with either state [Fig. 10(a)]. In contrast, in the simulations with the high affinity state, only a small fraction of the values sampled by the dihedral angle were in the region pertaining to the low affinity state [Fig. 10(b)]. Interestingly, in both cases the distribution was observed to be Gaussian (Fig. 10) indicating that the conformational space of the large β sheet is constrained inside a rather shallow parabolic potential well. Consequently, the β sheet can be described as a rather soft Hookean spring that can be locked in either an over-twisted or an over-stretched conformation, as is the case in the low or high affinity state, respectively. In the simulations with the low affinity state, where the pilin domain had been removed, the β sheet likely sampled values around its equilibrium conformation, that is, neither over-twisted nor over-stretched, according to a Gaussian shaped Boltzmann distribution [Fig. 10(a)].

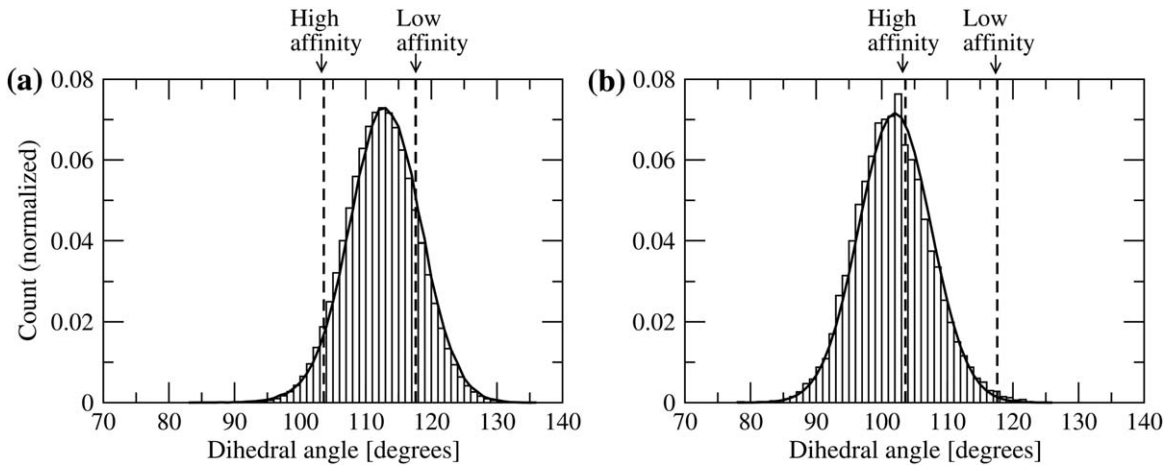

Figure 10

Elasticity of the large β sheet observed in the Anton simulations. Histograms of the dihedral angle defined by the C_α atoms of residues 17, 135, 125, and 22 measured along the simulations with (**a**) the low and (**b**) the high affinity state, respectively. The binning size was one degree. The thick solid line represents fitting with a Gaussian curve. The vertical dashed lines are averages calculated from the last 40 ns of two in total 50-ns long simulations performed at 300 K with either the low or the high affinity state, respectively (Table I).

DISCUSSION

This manuscript presents a model for how allosteric conformational changes propagate across a protein domain that consists mainly of β-sheet structure. The lectin domain of the adhesive protein FimH switches to a high affinity state for binding mannose when separated from the inhibitory pilin domain.[10] The lectin domain can be dissected into three main regions: an inter-domain region consisting mainly of disordered loops that contact the pilin domain in the low affinity state, a mannose binding region also containing flexible loops, and a structured central region composed predominantly of a large β sheet [Figs. 1 and 7(a)]. Understanding how conformational changes transition from one site of the lectin domain to the other requires analyzing how the three regions are connected to each other and how binding or separation of either the pilin domain or mannose affects the respective binding sites.

Entropic effects are likely to play a role in conformational changes occurring in the inter-domain regulatory region upon separation from the pilin domain. In fact, in its docked state, the pilin domain is found wedged between the swing and linker segments located in the inter-domain region of the lectin domain [Fig. 7(b)]. However, after separation, rearrangements occur that reduce the radius of gyration of the inter-domain region and the solvent accessible surface area of key hydrophobic residues. Also, the swing segment in the inter-domain region displays a slightly larger mobility in the high affinity state, which is indicative of higher entropy. This is consistent with chemical shift data, which indicates a relatively high mobility of the inter-domain region in the high affinity state[35] [Fig. 2(b)]. Similarly, entropic effects regulate the function of the mannose binding site. In its open configuration of the low affinity state, the clamp segment [Fig. 3(a)] is relatively flexible and samples conformational space [Fig. 2(a)]. Once mannose is bound to the high affinity state, the clamp segment and nearby side chains are more rigid, preventing water penetration into the binding site. The loss of conformational entropy is compensated by the formation of intra-protein hydrogen bonds and hydrogen bonds with the ligand [Fig. 3(b)]. The function of the clamp segment can be described as a gate that needs to open before mannose can detach, as observed in a 330-K simulation performed here with the high affinity state and mannose initially bound (Supporting Information Fig. S8). Similarly, binding of mannose to the high affinity state also likely requires opening of the clamp segment, explaining the recently experimentally measured slower rate of binding for the high versus the low affinity state.[39]

Simulations in the μs time range started from the isolated lectin domain in low affinity revealed that motions in the inter-domain region and the mannose binding pocket were moderately correlated with the twist-like movement of the centrally located large β sheet. The conformational fluctuations of the two binding sites were weakly correlated to each other. These observations are consistent with a recent mutagenesis study that also showed weak conformational coupling between the two distal regions.[15] The centrally located large β sheet was found here to have the characteristics of a soft spring indicated by the Guassian distribution of a dihedral angle that describes its twisting motion. This suggests that conformational changes occurring at either binding site due

to the interaction with a ligand are transmitted across the lectin domain thanks to the elastic properties of the large β sheet. In fact, the absence of the pilin domain from the μs-long simulations with the low affinity state facilitated structural elements in the inter-domain region and the large β sheet to fluctuate between the low and the high affinity conformations [Figs. 8(a) and 10(a)]. However, a full transition to the high affinity state was not observed in the simulations started from low affinity. In particular, the bulge-helix segment, which is located at the opposite edge of the β sheet than the clamp segment (Fig. 1), generally retained its initial conformation throughout the simulations (Supporting Information Figures S4 and S5). This indicates that the conformational transition is propagated across the lectin domain mainly through twisting of the large β sheet, which has the greatest effect on the clamp and swing loops. In contrast, other elements that lie between the regulatory and active sites, like the bulge-helix segment that are known to be part of the allosteric pathway,[15] do not demonstrate any involvement in propagation between the distal sites in our study, and instead may stabilize the two distinct conformations of the regulatory region.

From a thermodynamic point of view, the allosteric mechanism of FimH can be described by a "population shift" model.[4,18] According to this model, when the lectin domain is in high affinity (that is, separated from the pilin domain) the inter-domain region and the large β sheet can access conformations similar to the low affinity state although higher in energy. Docking of the pilin domain would then stabilize the lifted conformation of the swing segment shifting the equilibrium toward the low affinity state. Similarly, mannose is likely to stabilize the closed conformation of the clamp segment. Classic models of allostery focused on whether allosteric conformational changes happen sequentially[40] or whether they must happen simultaneously in a concerted fashion.[2] The mechanism of the lectin domain is a combination of both. Due to the elasticity of the β sheet, conformational changes at the interface between two neighboring regions are coupled to each other whereas overall conformational changes in the distal regions are weakly related. This allows the presence of transient intermediate states that are difficult to characterize through crystallographic studies alone but for which there is biochemical evidence.[15] Understanding dynamics at atomic level of detail makes it possible to link the experimentally observed intermediate states to the elasticity of the β sheet and to how conformational changes are propagated. During the preparation of this manuscript, a crystal structure was published containing four FimH molecules in the asymmetric unit, three of which locked in an intermediate state with mannose bound to the high affinity conformation of the pocket while the inter-domain region was in the low affinity conformation and docked to the pilin domain (PDB code 4XOB).[41] This is consistent with the mechanism proposed here that the interconversion between the low and the high affinity state consists of a series of intermediate steps.

It would be interesting to evaluate whether β-sheet fluctuations play a role also in the regulation of other proteins. Ten percent of protein domains have a β-sandwich structure, including the binding domains in catch-bond forming cadherins,[42,43] and many bacterial adhesins that mediate a shear-enhanced adhesion that strongly suggests catch bonds.[44–49] It is possible that backbone fluctuations in β-sandwich folds are common in catch-bond forming proteins because the β sandwich would confer the mechanical stability needed to withstand force, while the twisting fluctuations would allow the allostery that provides a mechanism for catch bonds. It would also be interesting to determine whether β sheets in mixed α-helix and β-sheet domains, such as the Rossman folds common in integrins and adhesive proteins such as von Willebrand factor, also provide flexibility that contributes to known or suspected allosterically induced conformational changes.[7,50,51]

In conclusion, the present manuscript illustrates how β sheets can transfer allosteric conformational changes across a protein domain. Thanks to MD simulations, it was possible to distinguish what elements in the lectin domain are key for the propagation of conformational changes and how their motions are related to each other. It would be interesting to apply the methods used here to other proteins that are known or suspected to be regulated allosterically, in particular when the transmission of conformational changes is likely to occur across β sheets. Here, μs-long simulations were run on a specialized supercomputer. However, when such a resource is not available, it is thinkable that accelerated methods, like higher temperature,[52,53] metadynamics[54,55] or accelerated dynamics,[56,57] could be used instead. Molecular dynamics results can then be complemented by experimental assays that have also been successfully used to explore the dynamics behind allostery,[4] like nuclear magnetic resonance[58] or fluorescence resonance energy transfer.[59]

ACKNOWLEDGMENTS

The authors would like to thank Dr. Evgeni Sokurenko and Dr. Jim Pfaendtner for helpful and interesting discussions. The simulations with NAMD were performed on the Trestles supercomputer at the San Diego Supercomputing Center thanks to a XSEDE allocation with grant number TG-MCB060069N, which is made available in part through NSF support. The μs-long simulations were performed through grant number PSCA10085P on the Anton supercomputer located at the Pittsburgh Supercomputing Center. The authors would like to thank Dr. Markus Dittrich and Dr. Nikolay Simakov for their help and support in performing the simulations on Anton.

REFERENCES

1. Changeux J. Feedback control mechanism of biosynthetic L-threonine deaminase by L-isoleucine. Cold Spring Harb Symp Quant Biol 1961;26:313.
2. Monod J, Wyman J, Changeux J. On nature of allosteric transitions—a plausible model. J Mol Biol 1965;12:88.
3. Daily MD, Gray JJ. Local motions in a benchmark of allosteric proteins. Proteins 2007;67:385–399.
4. Cui Q, Karplus M. Allostery and cooperativity revisited. Protein Sci 2008;17:1295–1307.
5. Emberly E, Mukhopadhyay R, Wingreen N, Tang C. Flexibility of alpha-helices: results of a statistical analysis of database protein structures. J Mol Biol 2003;327:229–237.
6. Cecchini M, Houdusse A, Karplus M. Allosteric communication in myosin V: from small conformational changes to large directed movements. PLoS Comput. Biol 2008;4:e1000129.
7. Xiao T, Takagi J, Coller B, Wang J, Springer T. Structural basis for allostery in integrins and binding to fibrinogen-mimetic therapeutics. Nature 2004;432:59–67.
8. Campbell ID, Humphries MJ. Integrin structure, activation, and interactions. Cold Spring Harb. Perspect. Biol 2011;3:pii:a004994.
9. Yakovenko O, Sharma S, Forero M, Tchesnokova V, Aprikian P, Kidd B, Mach A, Vogel V, Sokurenko E, Thomas WE. FimH forms catch bonds that are enhanced by mechanical force due to allosteric regulation. J Biol Chem 2008;283:11596–11605.
10. Le Trong I, Aprikian P, Kidd BA, Forero-Shelton M, Tchesnokova V, Rajagopal P, Rodriguez V, Interlandi G, Klevit R, Vogel V, Stenkamp RE, Sokurenko EV, Thomas WE. Structural basis for mechanical force regulation of the adhesin FimH via finger trap-like β sheet twisting. Cell 2010;141:645–655.
11. Emberly E, Mukhopadhyay R, Tang C, Wingreen N. Flexibility of beta-sheets: principal component analysis of database protein structures. Proteins 2004;55:91–98.
12. Hahn E, Wild P, Hermanns U, Sebbel P, Glockshuber R, Haner M, Taschner N, Burkhard P, Aebi U, Muller S. Exploring the 3D molecular architecture of *Escherichia coli* type 1 pili. J Mol Biol 2002;323:845–857.
13. Aprikian P, Interlandi G, Kidd B, Le Trong I, Tchesnokova V, Yakovenko O, Whitfield M, Bullitt E, Stenkamp R, Thomas W, Sokurenko E. The bacterial fimbrial tip acts as a mechanical force sensor. PLoS Biol 2011;9:e1000617.
14. Aprikian P, Tchesnokova V, Kidd B, Yakovenko O, Yarov-Yarovoy V, Trinchina E, Vogel V, Thomas W, Sokurenko E. Interdomain interaction in the FimH adhesin of *Escherichia coli* regulates the affinity to mannose. J Biol Chem 2007;282:23437–23446.
15. Rodriguez VB, Kidd BA, Interlandi G, Tchesnokova V, Sokurenko EV, Thomas WE. Allosteric coupling in the bacterial adhesive protein fimh. J Biol Chem 2013;288:24128–24139.
16. Ma L, Cui Q. Activation mechanism of a signaling protein at atomic resolution from advanced computations. J Am Chem Soc 2007;129:10261–10268.
17. Kumar S, Ma B, Tsai C, Sinha N, Nussinov R. Folding and binding cascades: dynamic landscapes and population shifts. Protein Sci 2000;9:10–19.
18. Hilser VJ, Wrabl JO, Motlagh HN. Structural and energetic basis of allostery. In: Rees, DC, editor, volume 41 *Annu. Rev. Biophys., vol 41,* of *Annual Review of Biophysics*, 585–609. Annual Reviews, 2012.
19. Bouckaert J, Berglund J, Schembri M, De Genst E, Cools L, Wuhrer M, Hung C, Pinkner J, Slattegard R, Zavialov A, Choudhury D, Langermann S, Hultgren S, Wyns L, Klemm P, Oscarson S, Knight S, De Greve H. Receptor binding studies disclose a novel class of high-affinity inhibitors of the Escherichia coli FimH adhesin. Mol Microbiol 2005;55:441–455.
20. Brooks BR, Brooks CL, III, Mackerell AD, Jr., Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. CHARMM: the biomolecular simulation program. J Comput Chem 2009;30:1545–1614.
21. Kisiela DI, Avagyan H, Friend D, Jalan A, Gupta S, Interlandi G, Liu Y, Tchesnokova V, Rodriguez VB, Sumida JP, Strong RK, Wu XR, Thomas WE, Sokurenko EV. Inhibition and reversal of microbial attachment by an antibody with parasteric activity against the FimH adhesin of uropathogenic *E. coli*. PLoS Pathog 2015;11.
22. Kalé L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K. NAMD2: greater scalability for parallel molecular dynamics. J. Comp. Phys. 1999;151:283–312.
23. 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, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
24. Darden T, York D, Pedersen L. Particle mesh ewald - an N.log(N) method for ewald sums in large systems. J Chem Phys 1993;98:10089–10092.
25. Schneider T, Stoll E. Molecular-dynamics study of a 3-dimensional one-component model for distortive phase-transitions. Phys. Rev. B 1978;17:1302–1322.
26. Feller SE, Zhang YH, Pastor RW, Brooks BR. Constant-pressure molecular-dynamics simulation - the Langevin piston method. J Chem Phys 1995;103:4613–4621.
27. Humphrey W, Dalke A, Schulten K. Vmd: Visual molecular dynamics. J Mol Graph 1996;14:33–38.
28. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossvary I, Klepeis JL, Layman T, Mcleavey C, Moraes MA, Mueller R, Priest EC, Shan Y, Spengler J, Theobald M, Towles B, Wang SC. Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM 2008;51:91–97.
29. Shan Y, Klepeis J, Eastwood M, Dror R, Shaw D. Gaussian split Ewald: A fast Ewald mesh method for molecular simulation. J Chem Phys 2005;122.
30. Nose S. A unified formulation of the constant temperature molecular-dynamics methods. J Chem Phys 1984;81:511–519.
31. Hoover W. Canonical dynamics—equilibrium phase-space distributions. Phys Rev A 1985;31:1695–1697.
32. Martyna G, Tuckerman M, Tobias D, Klein M. Explicit reversible integrators for extended systems dynamics. Mol Phys 1996;87:1117–1157.
33. Miller D, Dill K. Ligand binding to proteins: the binding landscape model. Protein Sci 1997;6:2166–2179.
34. Interlandi G. Backbone conformations and side chain flexibility of two somatostatin mimics investigated by molecular dynamics simulations. Proteins 2009;75:659–670.
35. Vanwetswinkel S, Volkov AN, Sterckx YGJ, Garcia-Pino A, Buts L, Vranken WF, Bouckaert J, Roy R, Wyns L, van Nuland NAJ. Study of the structural and dynamic effects in the FimH adhesin upon alpha-D-Heptyl mannose binding. J Med Chem 2014;57:1416–1427.
36. Nilsson LM, Thomas WE, Sokurenko EV, Vogel V. Beyond induced-fit receptor-ligand interactions: structural changes that can significantly extend bond lifetimes. Structure 2008;16:1047–1058.
37. Wellens A, Lahmann M, Touaibia M, Vaucher J, Oscarson S, Roy R, Remaut H, Bouckaert J. The tyrosine gate as a potential entropic lever in the receptor-binding site of the bacterial adhesin FimH. Biochemistry 2012;51:4790–4799.

38. Fiege B, Rabbani S, Preston RC, Jakob RP, Zihlmann P, Schwardt O, Jiang X, Maier T, Ernst B. The tyrosine gate of the bacterial lectin FimH: a conformational analysis by NMR spectroscopy and X-ray crystallography. ChemBioChem 2015;16:1235–1246.
39. Yakovenko O, Tchesnokova V, Sokurenko EV, Thomas WE. Inactive conformation enhances binding function in physiological conditions. Proc Natl Acad Sci USA 2015;112:9884–9889.
40. Pauling L. The oxygen equilibrium of hemoglobin and its structural interpretation. Proc Natl Acad Sci USA 1935;21:181–191.
41. Sauer MM, Jakob RP, Eras J, Baday S, Eri D, Navarra G, Berneche S, Ernst B, Maier T, Glockshuber R. Catch-bond mechanism of the bacterial adhesin FimH. Nat Commun 2016;7:10738.
42. Rakshit S, Zhang Y, Manibog K, Shafraz O, Sivasankar S. Ideal, catch, and slip bonds in cadherin adhesion. Proc Natl Acad Sci USA 2012;109:18815–18820.
43. Manibog K, Li H, Rakshit S, Sivasankar S. Resolving the molecular mechanism of cadherin catch bond formation. Nat Comm 2014;5.
44. Nilsson L, Thomas W, Trintchina E, Vogel V, Sokurenko E. Catch bond-mediated adhesion without a shear threshold—trimannose versus monomannose interactions with the FimH adhesin of *Escherichia coli*. J Biol Chem 2006;281:16656–16663.
45. Tchesnokova V, McVeigh AL, Kidd B, Yakovenko O, Thomas WE, Sokurenko EV, Savarino SJ. Shear-enhanced binding of intestinal colonization factor antigen I of enterotoxigenic *Escherichia coli*. Mol Microbiol 2010;76:489–502.
46. Liu Y, Esser L, Interlandi G, Kisiela DI, Tchesnokova V, Thomas WE, Sokurenko E, Xia D, Savarino SJ. Tight conformational coupling between the domains of the enterotoxigenic *Escherichia coli* fimbrial adhesin CfaE regulates binding state transition. J Biol Chem 2013;288:9993–10001.
47. Ding AM, Palmer RJ, Jr., Cisar JO, Kolenbrander PE. Shear-enhanced oral microbial adhesion. Appl Environ Microbiol 2010;76: 1294–1297.
48. Stahlhut SG, Tchesnokova V, Struve C, Weissman SJ, Chattopadhyay S, Yakovenko O, Aprikian P, Sokurenko EV, Krogfelt KA. Comparative structure-function analysis of mannose-specific FimH adhesins from klebsiella pneumoniae and *Escherichia coli*. J. Bacteriol. 2009; 191:6592–6601.
49. Kisiela DI, Kramer JJ, Tchesnokova V, Aprikian P, Yarov-Yarovoy V, Clegg S, Sokurenko EV. Allosteric catch bond properties of the FimH adhesin from salmonella enterica serovar typhimurium. J Biol Chem 2011;286:38136–38147.
50. Yago T, Lou J, Wu T, Yang J, Miner JJ, Coburn L, Lopez JA, Cruz MA, Dong JF, McIntire LV, McEver RP, Zhu C. Platelet glycoprotein Ibα forms catch bonds with human WT vWF but not with type 2B von Willebrand disease vWF. J Clin Invest 2008;118:3195–3207.
51. Interlandi G, Thomas W. The catch bond mechanism between von Willebrand Factor and platelets investigated by molecular dynamics simulations. Proteins 2010;78:2506–2522.
52. Interlandi G, Settanni G, Caflisch A. Unfolding transition state and intermediates of the tumor suppressor $p16^{ink4a}$ investigated by molecular dynamics simulations. Proteins 2006;64:178–192.
53. Interlandi G, Wetzel SK, Settanni G, Plückthun A, Caflisch A. Characterization and further stabilization of designed ankyrin repeat proteins by combining molecular dynamics simulations and experiments. J Mol Biol 2008;375:837–854.
54. Laio A, Parrinello M. Escaping free-energy minima. Proc Natl Acad Sci USA 2002;99:12562–12566.
55. Valsson O, Tiwary P, Parrinello M. Enhancing important fluctuations: rare events and metadynamics from a conceptual viewpoint. Annu Rev Phys Chem 2016;67:1150–1155.
56. Hamelberg D, Mongan J, McCammon J. Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys 2004;120:11919–11929.
57. Markwick PRL, McCammon JA. Studying functional dynamics in bio-molecules using accelerated molecular dynamics. Phys Chem Chem Phys 2011;13:20053–20065.
58. Wuthrich K. Protein-structure determination in solution by NMR-spectroscopy. J Biol Chem 1990;265:22059–22062.
59. Helms V. Principles of computational cell Biology. Wiley, Weinheim; 2008.
60. Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF, Tolmie DE, Wenger RK, Yao H, Markley JL, BioMagResBank. Nucleic Acids Res 2008;36:D402–D408.
61. Berjanskii M, Wishart D. A simple method to predict protein flexibility using secondary chemical shifts. J Am Chem Soc 2005;127: 14970–14971.
62. Schwarzinger S, Kroon G, Foss T, Wright P, Dyson H. Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. J Biomol NMR 2000;18:43–48.