Donor-strand exchange in chaperone-assisted pilus assembly revealed in atomic detail by molecular dynamics

Rebecca J. Rose[1,2], Thomas S. Welsh[1,3], Gabriel Waksman[4], Alison E. Ashcroft[1,2], Sheena E. Radford[1,2] and Emanuele Paci[1,3*]

[1]Astbury Centre for Structural Molecular Biology, [2]Institute of Molecular and Cellular Biology, [3]Physics & Astronomy, University of Leeds, Leeds LS2 9JT, UK, [4]Institute of Structural Molecular Biology, UCL and Birkbeck, Malet Street, London WC1E 7HX, UK.

*To whom correspondence should be addressed

Phone +44 113 3433806

Email e.paci@leeds.ac.uk

Running title: Molecular dynamics simulations of pilus assembly

Key words: Molecular dynamics, pili, chaperone, donor-strand exchange, assembly mechanisms

Abstract

Adhesive multi-subunit fibres are assembled on the surface of many pathogenic bacteria *via* the chaperone-usher pathway. In the periplasm, a chaperone donates a β-strand to a pilus subunit to complement its incomplete immunoglobulin-like fold. At the outer membrane, this is replaced with a β-strand formed from the N-terminal extension (Nte) of an incoming pilus subunit by a donor-strand exchange (DSE) mechanism. This reaction has previously been shown to proceed *via* a concerted mechanism, in which the Nte interacts with the chaperone:subunit complex before the chaperone has been displaced, forming a ternary intermediate. Thereafter, the pilus and chaperone β-strands have been postulated to undergo a strand swap by a 'zip-in-zip-out' mechanism, whereby the chaperone strand zips out, residue by residue, as the Nte simultaneously zips in**.** Here, molecular dynamics simulations have been used to probe the DSE mechanism during formation of the *Salmonella enterica* Saf pilus at an atomic level, allowing the direct investigation of the zip-in-zip-out hypothesis. The simulations provide an explanation of how the incoming Nte is able to dock and initiate DSE due to inherent dynamic fluctuations within the chaperone:subunit complex. The chaperone donor-strand is shown to unbind from the pilus subunit residue by residue, in direct support of the zip-in-zip-out hypothesis. In addition, an interaction of a residue towards the N-terminus of the Nte with a specific binding pocket (P*) on the adjacent pilus subunit is shown to stabilise the DSE product against unbinding, which also proceeds by a zippering mechanism. Together, the study provides an in-depth picture of DSE, including the first insights into the molecular events occurring during the zip-in-zip-out mechanism.

Introduction

One of the most fascinating properties of proteins is their ability to adopt well defined structures and to form stable complexes. Intra-protein interactions, in addition to interactions between the protein and solvent, are responsible for the ability of a protein to reach and maintain a folded, functional structure. Inter-protein interactions have diverse roles in living cells, including the assembly of functional complexes and polymeric structures with pathological consequences[1; 2]. One mechanism through which proteins interact to form large complexes consists of the complementation of a missing motif[3]. One particularly clear example of this is the pili that decorate the surface of pathogenic Gram negative bacteria[4; 5]; these fibres are composed of protein subunits polymerised through the exchange of a β-strand (Figure 1a)[6-9]. An adhesin molecule is typically present at the distal end of the fibre which allows binding of the bacteria to a specific surface on the host cell allowing the establishment of infection[10-13]. Knowledge of the molecular details of the pilus assembly mechanism can thus potentially provide a framework for the design of inhibitors of pilus assembly that may allow the development of a new class of antibiotics[14], while such protein polymers themselves may allow the design of novel protein-based materials[15-17].

A major family of adhesive structures, including P pili (encoded by the Pap operon) and Type I pili (encoded by the Fim operon) from *Escherichia coli*, F1 antigens (encoded by the Caf operon) from *Yersinia pestis*, and Saf non-fimbrial adhesins (encoded by the Saf operon) from *Salmonella enterica*, assemble by the so-called chaperone-usher secretion pathway[18; 19]. Here, the subunit 'building blocks' of the pilus are produced in the cytoplasm and are transported, *via* the general secretory pathway (Sec), into the periplasm[5; 20]. Therein, a chaperone binds to the subunit to form a binary complex, which is targeted to an outer membrane usher protein[11; 13; 21; 22]. At this site, the chaperone is released and subunit polymerisation occurs to form the pilus (Figure 1a)[18; 22; 23].

Crystal structures have shown that the overall structures of the pilus subunits and their respective chaperones are highly conserved amongst all pili that assemble *via* the chaperone-usher pathway and have additionally revealed the basis of the interactions involved in pilus assembly[6-9; 24-26]. The pilus subunit (the 'pilin') forms an incomplete immunoglobulin-like fold in which the C-terminal β-strand is absent, effectively leaving an exposed hydrophobic groove[8; 9; 27]. With the exception of the adhesin, each subunit also contains an unstructured N-terminal extension (Nte), usually between 10 and 20 residues in length, which is not involved in its fold. The chaperone is composed of two immunoglobulin domains oriented at right angles to each other (Figure 1b)[7-9; 28-30]. In the chaperone:subunit complex, the G strand from one chaperone domain (G_1) binds to the pilus subunit, effectively filling its hydrophobic groove in a mechanism known as Donor-Strand Complementation (DSC, Figure 1b)[8; 9; 27]. By contrast, within the assembled pilus, the Nte from one subunit forms the missing β-strand in the adjacent subunit, completing its immunoglobulin-like fold, in a reaction known as Donor-Strand Exchange (DSE, Figure 1c)[6-9; 24]. Whilst the donated β-strands in the DSC and DSE reactions have the opposite orientations (parallel and anti-parallel to the subunit's F strand, respectively, Figure 1b,c), they share the common property that strand complementation is mediated by up to 5 hydrophobic residues (termed P1 to P5) that slot into corresponding adjacent hydrophobic binding pockets (P1-P5) in the pilus subunit's groove (Figure 1b,c)[6].

The details of how the β-strands of the chaperone and pilus subunit swap during pilus assembly, until recently, were largely unknown. Models involving dissociation of the chaperone and subunit prior to binding of the Nte, or a concerted mechanism involving the formation of a chaperone:subunit:Nte intermediate, have been proposed[7]. Recent experiments using non-covalent electrospray ionisation mass spectrometry (ESI-MS) on the Saf pilus system from *Salmonella enterica* have allowed the formation of this ternary intermediate during DSE to be observed,

providing direct experimental evidence in support of the concerted mechanism[24]. Similar conclusions have been drawn from a kinetic and thermodynamic study of type I pili[31]. Once the initial chaperone:subunit:Nte intermediate has formed, a second important, yet currently unproven, feature of the concerted mechanism is that chaperone displacement and Nte binding occur in concert in a so-called 'zip-in-zip-out' mechanism[7]. In this model, the Nte is assumed to zip into each subunit hydrophobic pocket as the chaperone G_1 strand zips out. Experimental results have provided some evidence in support of this hypothesis, demonstrating that amino acid substitutions that decrease the hydrophobicity of the P5 residue in the incoming Nte cause a substantial reduction in the rate of DSE, whilst similar changes in the P4 and P3 sites have a much weaker effect[24]. These data support the view that unzippering of the chaperone and binding of the Nte commence at the P5 site, although whether subsequent binding sites become vacated/ occupied sequentially remains unproven, since the kinetic experiments performed to date lack the resolution to monitor the rate of binding to individual sites. A crucial feature of the zip-in-zip-out model thus remains unresolved.

Here, we have used molecular dynamics (MD) simulations to probe the mechanism of DSE in all-atom detail using Saf pili from *Salmonella enterica* as a model. All-atom simulations offer the opportunity to uncover the molecular mechanism of DSE at a level of resolution not possible experimentally[32]. Here, three key steps in the mechanism have been investigated using different protein crystal structures as the starting point for the simulations. First, the mechanism of initiation of DSE was investigated by simulation of the Saf chaperone:subunit (SafB:SafA) complex at room temperature. Second, the unbinding of the chaperone G_1 strand from the chaperone:subunit complex was simulated using steered and biased MD to allow unbinding to occur on a computationally accessible timescale (see Methods for experimental details). Finally, simulations of the unbinding of Nte peptides with different sequences from the subunit:Nte complex were carried out to investigate how stability of the DSE product is imparted and pilus disassembly

disfavoured. Together, the MD simulations present an all-atom view of DSE providing important insights into this protein assembly mechanism in an unprecedented level of detail.

Results

The crystal structures of the SafB:SafA$_{Ntd2}$ DSC and SafA$_{Ntd2}$:Nte DSE complexes, i.e. the subunit (SafA) in complex with the chaperone (SafB) or a synthetic peptide equivalent to the Nte of SafA (19 residues), respectively (Figure 1b,c) have recently been solved[24]. To prevent self-polymerisation, a truncated form of the pilin subunit lacking its Nte was used to model the acceptor subunit; this construct is named SafA$_{Ntd2}$. These crystal structures were used as the starting point for the simulations described here. Control simulations (see Methods) showed that the complexes are stable over nanoseconds at room temperature (300K), suggesting that the implicit solvent model chosen is appropriate for this study. Importantly, this model allows extensive simulations that would not be feasible using explicit solvent, allowing the reproducibility of the results to be assessed.

Simulations of the SafB:SafA$_{Ntd2}$ chaperone:subunit complex

The mechanism of DSE in the Saf system has recently been dissected using X-ray crystallography and ESI-MS[24]. Two crystal structures of SafB:SafA$_{Ntd2}$ were solved (resulting from different crystallisation conditions), one of which (2CO7) showed no electron density around the P5 hydrophobic pocket. We thus chose the better resolved structure (2CO6) as the initial structure for our simulations of the SafB:SafA$_{Ntd2}$ complex; this structure shows electron density around the P5 pocket, although the entire F_1-G_1 loop is not resolved (see Methods). The properties of the SafB:SafA$_{Ntd2}$ complex at room temperature in solution were explored by performing a 205 ns simulation at 300K in implicit solvent. During the simulation residues in SafB not involved in the binding interface were held fixed. Residues in strands A_1, F_1 and G_1 and all others of SafB

involved in the binding interface, as well as all degrees of freedom of $SafA_{Ntd2}$, were free to move under the effect of the inter- and intra-molecular forces; of the total of 343 residues in the complex, 161 were held constant. Significantly, the RMSD of the complex, relative to the X-ray structure, never exceeds 3 Å and is constant across the simulation (Figure 2a). The buried surface area between $SafA_{Ntd2}$ and SafB fluctuates, but remains constant on average (Figure 2b). Together, these data indicate that the complex is stable over the entire 205 ns simulation.

The results of the simulation show that whilst SafB:$SafA_{Ntd2}$ is stable under the conditions employed, the complex nevertheless displays significant positional fluctuations, particularly in the region encompassing the chaperone F_1-G_1 loop and P5 residue (A114). The details of the interaction between A114 of SafB and the $SafA_{Ntd2}$ P5 hydrophobic pocket to which it binds were therefore examined in detail (Figure 2c). Distances between the main-chain nitrogen of A114 and the carboxyl carbon of A35 or A130 in the A and F strands of $SafA_{Ntd2}$, respectively, were used as probes of the location of the P5 residue relative to the pilus subunit. This analysis revealed a fascinating oscillation of the P5 residue between these two sites, commencing with a distance between SafB-A114(N) and $SafA_{Ntd2}$-A130(C) of ~5 Å in the X-ray structure, whilst during the course of the simulation, this distance oscillates between 5.2 Å and 7.8 Å (Figure 2c). This suggests that there are two local minima in the energy landscape. Interestingly, the distances of A114 to the $SafA_{Ntd2}$ A strand and to the $SafA_{Ntd2}$ F strand (Figure 2c) are anti-correlated, i.e., when the residue is close to the A strand, it is distant from the F strand, and *vice versa*.

Closer examination of the trajectory revealed that the P5 residue rotates between occupying these two discrete states (Figure 3a), facilitated by a slight bending of the SafB G_1 strand. When nearest to the subunit A strand, A114 is deeply buried and centrally positioned in the P5 hydrophobic pocket (Figure 3b). In marked contrast to this, at the local energy minimum where A114 is more closely associated with the subunit F strand, A114 is located at the surface of the pocket, which

itself is less well-defined (Figure 3c). This effectively leaves the pocket vacant and therefore accessible to attack from the incoming P5 residue on the Nte (F17). The crystal structure of $SafA_{Ntd2}$ in complex with the Nte used in this study shows that F17, which replaces A114 after DSE, is deeply buried within the hydrophobic pocket (Figure 3d). Thus, the inherent structural fluctuations around the P5 pocket could facilitate the binding of the attacking Nte, leading to a cascade of events that displace the chaperone G_1 β-strand from the hydrophobic groove during DSE.

Although not previously explicitly detected in the crystal structures of SafB:$SafA_{Ntd2}$, the existence of two alternative conformations for this complex is consistent with the experimental data available. In one of the two crystal forms of SafB:$SafA_{Ntd2}$ (2CO7)[24], the region encompassing the P5 residue could not be resolved, suggesting disorder in the same region as that found to exhibit particularly large fluctuations in the simulations presented here. For the crystal structure where the P5 region could be resolved (2CO6)[24], a large B factor suggests positional fluctuations of large amplitude in this region. To explore this further, the root mean square fluctuations of the atomic positions ($< dr_i^2 >$) of $SafA_{Ntd2}$ along the simulated trajectory of the SafB:$SafA_{Ntd2}$ complex were calculated and compared with the fluctuations derived from the experimental B factor ($B=8\pi^3<\delta r_i^2>/3$) (Figure 4a). This analysis revealed that the regions which fluctuate most in the simulations are those which have the largest experimental B factor, strongly suggesting that the dynamics observed in the simulation reflect the intrinsic flexibility of $SafA_{Ntd2}$ when in complex with SafB. Three out of the four main regions of positional fluctuation in both experiment and simulation, centred on residues 38, 80 and 127, play a role in forming, or are directly adjacent to, the P5 binding pocket (see Figure 4b). The region around residue 65 additionally shows high fluctuations, and relates to a presumably intrinsically flexible loop at the back of the subunit with respect to the P5 binding site, that is held away from the main fold in the crystal structure. These

data suggest that not only the location of the chaperone P5 residue relative to SafA, but also inherent dynamics of SafA itself, around and within the binding pocket, are important for destabilising the docking of the chaperone P5 residue to the P5 pocket, allowing the incoming Nte to bind at this site. The detection of two crystal forms, one of which does not show clear electron density in the region characterised by simulation to be of high flexibility, suggests that the two alternative conformations observed in the simulations may well exist in the crystal growth conditions.

The simulations presented strongly suggest that dynamic conformational changes at the P5 pocket of the SafB:SafA$_{Ntd2}$ complex may allow the initiation of DSE, by providing an opportunity for the Nte to dock at the P5 site. Experiments previously performed provide additional support for this hypothesis[24]. Thus, using variants in which the chaperone P5 residue was substituted with a larger hydrophobic side-chain, more complementary to the P5 pocket (A114V or A114F), resulted in a much slower rate of DSE, and the amount of chaperone:subunit:Nte intermediate seen to form was decreased. Similarly, decreasing the size and hydrophobicity of the P5 residue on the incoming Nte by creating the variants F17I, F17V or F17A, also reduced the rate of DSE and the amount of ternary complex observed. The analysis of the simulations presented here support these experimental findings and demonstrate how Nte binding and DSE are made possible through inherent structural fluctuations in the chaperone:subunit complex that alter the accessibility of the P5 pocket to the attacking Nte.

Unbinding of the DSC complex

After the docking of the Nte at the P5 pocket, the event which initiates DSE, the SafB G_1 strand must dissociate from the subunit hydrophobic groove, allowing the binding of the Nte of the incoming pilin subunit to occur simultaneously, in the so-called zip-in-zip-out mechanism of DSE[7]. As the simulations of the complex described above demonstrate, the unbinding of SafA$_{Ntd2}$ and

SafB does not occur spontaneously on a timescale accessible to simulation. To examine the mechanism of DSE in atomic detail, therefore, we employed a non-equilibrium simulation technique in which the two partners are actively pulled apart by applying a force through a harmonic spring (steered molecular dynamics, SMD), to a reaction coordinate which discriminates between bound and unbound states. In this case, the reaction coordinate was chosen to be proportional to the sum of the distances between pairs of atoms which are in contact in the bound complex (see Methods). During the SMD simulation, the value of the time dependent bias potential (or, analogously, the value of the force which is the derivative of such a potential with respect to the reaction coordinate) depends on the resistance opposed to the uniform increase of the reaction coordinate by the complex. Peaks observed in this bias potential are analogous to the force peaks observed in atomic force microscopy simulation of mechanical unfolding or unbinding[33; 34] and reveal the rupture of a "bond" whose strength is proportional to the height of the peak.

A typical profile of the bias energy needed to dissociate the SafB:SafA$_{Ntd2}$ complex in a SMD simulation is shown in Figure 5a. Figure 5b shows the distances of the P1 to P5 residues of SafB from their closest partner in SafA$_{Ntd2}$ (see caption to Figure 5). The data show that unbinding of the SafB G_1 strand from SafA$_{Ntd2}$ occurs by an unzippering mechanism, commencing with the P5 residue of the SafB G_1 strand and continuing down the strand residue by residue. The largest peak in the bias energy (~60 ps) corresponds to a sudden increase in distance between A114 of SafB with A130 of SafA$_{Ntd2}$, relating to the extraction of the P5 residue from its pocket. Smaller peaks in the bias energy occurring at ~190 ps, 350 ps, 720 ps and 2500 ps correspond to the sequential unbinding of the P4, P3, P2 and P1 residues of the SafB G_1 strand from the P4 to P1 pockets. Breaking the contact at the P5 pocket thus allows unzippering to proceed relatively unhindered. To determine whether the order of events occurring during unbinding is typical, or represents a rare unbinding mechanism, SMD was repeated 10 times. In all 10 independent trajectories, starting from different conformations from the control simulation (see Methods), the sequence of events

was identical to that described above, demonstrating that G_1 strand dissociation starts at the P5 pocket and proceeds systematically from the P5 to P1 pockets in a zip-out mechanism.

To test further the robustness of the mechanism described above, additional simulations were performed using a different type of bias, applied to the same reaction coordinate. In biased molecular dynamics (BMD) simulations[35-37], a force is applied when large "obstacles" are encountered. Spontaneous fluctuations are exploited to induce unbinding while an external force is only applied to prevent a spontaneous regression of the reaction (see Methods for details). Biased molecular dynamics simulations of the unbinding of the SafB:SafA$_{Ntd2}$ complex corroborated the SMD simulations by systematically showing the same sequence of events; in all 10 simulations performed, unbinding again occurs through an unmistakable unzipping mechanism, starting at the P5 site (Figure 5c). These results thus demonstrate that the features obtained from simulations using BMD are consistent with those obtained using SMD, confirming the robustness of the results obtained by simulation despite the unbinding being artificially induced by the application of an external force.

Unbinding of the DSE complex

The crystal structure of SafA$_{Ntd2}$:Nte, i.e. the product of DSE, shows that the subunit undergoes a conformational change when it binds the Nte compared with its chaperone-bound counterpart[6; 24]. In the subunit:Nte complex the subunit Ig domain becomes more compact and an extra hydrophobic pocket, named P*, forms by the re-orientation of two aromatic residues (Y142 and W103)[24]. This pocket, which is not present in the DSC complex, accommodates the side-chain of F3 of the Nte in the DSE product (Figure 1c). When F3 is substituted with an alanine, significant dissociation of subunit:Nte complex is observed[24]. Importantly, and by contrast with this, DSE with the wild-type Nte is irreversible[24]. These observations suggest that the binding of the aromatic ring of F3 to the P* pocket provides a thermodynamic capping mechanism which stabilises the

DSE product and prevents subunit:Nte dissociation. To test this hypothesis further and to provide more detailed molecular insights into the role of the P* pocket in DSE, the unbinding of the wild-type Nte, as well as the Nte containing the substitution F3A, from the $SafA_{Ntd2}$:Nte complex was investigated using both SMD and BMD methods (Figure 6). In all 20 simulations performed, the wild-type Nte was seen to unbind from $SafA_{Ntd2}$ by an unzippering mechanism, with contacts at the interface breaking sequentially along the strand. Unbinding of the wild-type Nte started from the P5 site (F17 residue) in 9 out of 10 SMD simulations (Figure 6a and 6c) and in 7 out of 10 BMD simulations (not shown). The bias potential during SMD simulations typically showed that breaking the contact between the F17 (P5) residue and $SafA_{Ntd2}$ at the start of the unbinding required the greatest force (see Figure 6a). Breaking the contact between the F3 (P*) residue and $SafA_{Ntd2}$ at the other end of the strand occurs without giving rise to a clearly detectable peak in the bias energy. Although in the majority of cases, unbinding began at the P5 end, it is also possible for the Nte to unzip from the P* end, the latter occurring less frequently possibly because F3 buries more hydrophobic surface area than F17 in both the X-ray structure (52.9 Å and 40.1 Å, respectively) and at the beginning of the simulation, which may reduce the probability of this residue dissociating first.

In striking contrast to the simulations of unbinding of the wild-type Nte from $SafA_{Ntd2}$, the substitution F3A in the Nte resulted in a complex in which unbinding of the Nte commenced at the P* site in 4 out of the 9 SMD simulations and in 8 of the 9 BMD simulations performed (for both SMD and BMD simulations, in 1 simulation out of 10 the A strand of $SafA_{Ntd2}$ detached from the rest of the protein and the peptide did not fully unbind). However, whilst the majority of simulations show a preference for unbinding commencing from the P* site, the mechanism of unbinding was much less clearly an unzippering one. In all of the simulations where full unbinding occurs, a change in conformation of the Nte during the unbinding causes the formation of a strong interaction between K11 in the Nte (the P2 residue) and L137 in $SafA_{Ntd2}$, resulting in the

unbinding pausing at this site (data not shown). In the equilibrium structure of $SafA_{Ntd2}$:Nte, the P5, P4 and P* residues interact most strongly with their corresponding hydrophobic pockets. By contrast, the P2 and P3 residues are unable to interact strongly with their pockets, due to a slight twist of the peptide in that segment (see Figure 1c). As the peptide starts to unbind, a further twist occurs which changes the orientation of the strand so that the side-chains of the K11 and also Q9 are now directed towards the F strand of the $SafA_{Ntd2}$. The aliphatic chains of these residues can then form a strong interaction with the hydrophobic pockets of $SafA_{Ntd2}$. These interactions could play a role in favouring DSE by effectively increasing the interactions with the P2 and P1 pockets until such time as DSE is completed by forming the capping interactions at P*. Transient, more extensive contacts with the P2 and P1 pockets will thus disfavour dissociation and allow the forward reaction to progress, such that unzippering of the partially docked Nte is disfavoured.

The average time for strand dissociation in BMD simulations is indicative of the strength of binding. In the simulations performed here, full unbinding of the wild-type Nte took 14 ns, on average, whilst the Nte containing the substitution F3A required only an average of 9 ns to unbind (full unbinding is assumed when all the pairwise distances defining the reaction coordinate exceed 20 Å). Together with the analysis presented above, these data indicate that the presence of an aromatic ring at the P* site both decreases the likelihood that unbinding will commence at the N-terminal end of the peptide, and significantly stabilizes the complex with $SafA_{Ntd2}$ formed.

Discussion

The results presented here provide an insightful picture of the mechanism of pilus subunit association which builds on the proposed mechanism of DSE based on experimental methods[7; 24], adding detailed, all-atom information about the molecular steps involved in pilus assembly. New insights into three parts of the mechanism have been revealed by this work, including: (i) a structural basis for the initiation of Nte binding, (ii) further evidence that DSE occurs by a zippering mechanism, and (iii) new insights into the mechanisms by which the final product is stabilised and the reversibility of DSE suppressed.

Here, using MD simulations at ambient temperature, we show that the SafB:SafA$_{Ntd2}$ chaperone:subunit complex exists in two distinct, stable conformations. In one, A114 of SafB (the P5 residue) is buried deeply within the SafA$_{Ntd2}$ hydrophobic pocket. In the second conformation, bending of the chaperone G_1 β-strand results in A114 being displaced from the pocket, such that it now occupies a second binding site at the surface of the pocket. This movement (of ~2.6 Å) increases the accessibility of the P5 pocket, providing an opportunity for the attacking P5 residue in the incoming Nte to dock, thus initiating DSE. The simulations of the unbinding of the SafB G_1 strand show that when the first contacts are broken, subsequent contacts involving the G_1 strand dissociate with relative ease. Unbinding thus commences from the P5 position and then continues residue by residue in an unzippering mechanism until the chaperone G_1 strand is displaced. Binding of the incoming Nte simultaneously to each site, as the G_1 strand is displaced, then leads to the product of DSE *via* the zip-in-zip-out mechanism.

The unbinding of the Nte from SafA$_{Ntd2}$ also indicates unzippering, with contacts being broken sequentially. The direction of unbinding, additionally, allows us to discern the importance of individual interactions in stabilising the DSE product. The peptide representing the wild-type Nte

unbinds from the P5 end in 16 of the total 20 simulations performed, presumably since the interaction at the P5 site is less extensive than that at the P* site. Substitution of the F3 residue for an alanine, however, causes this sequence of unbinding events to be disrupted, unbinding for this variant often commencing from the P* site with concomitant disruption of the unzipping mechanism. A consequence of this disruption is that the F3A peptide unbinds faster than the wild-type equivalent. These data imply that the F3 residue plays an important role in clamping the Nte to the adjacent pilus subunit, preventing dissociation and ensuring a unidirectional mechanism for DSE. In addition, during unbinding of the F3A peptide, a twist in the Nte strand during unzipping allows new hydrophobic contacts to be made, particularly between K11 (the P2 residue) and $SafA_{Ntd2}$, providing a second means of disfavouring dissociation by trapping dissociation mid-stream and presumably enhancing re-zippering and capping by the P* residue. This is indicative of longer range disruptive effects in the process when a single mutation is made. Thus, both the stability of the $SafA_{Ntd2}$:Nte product and the mechanistic details of the unbinding rely inherently on the precise order and identity of the residues involved.

Overall, the simulations presented here highlight the elegant design of the pilus assembly mechanism, that ensures that Nte binding and chaperone release occur in a controlled and concerted manner, such that free pilus subunits are not released from their chaperoning partners until a productive reaction with an incoming Nte can ensue. We show that the dynamics of the P5 residue in the SafB:$SafA_{Ntd2}$ complex provide an opportunity for DSE to commence, the F_1-G_1 loop of the chaperone switching between two conformations in which Nte docking is either blocked or favoured. We anticipate that dynamics at this loop will be a generic feature of DSE in all pili of the FGL class, such as Saf. By contrast, chaperones of the FGS class, of which PapD is a paradigm, have a much shorter F1-G_1 loop[38; 39]. In these complexes the P5 pocket is left unoccupied, providing a permanent opportunity for DSE to commence. Subsequent to docking of the P5 residue, zippering occurs readily and in a concerted manner, ensuring controlled binding of the Nte

and release of the chaperone, whilst capping by the P* residue ensures the irreversibility of DSE. The resulting kinetic and thermodynamic stability of the subunit:Nte product of DSE is essential to form stable, functional pili, able to bind host cells and form strong tethers during bacterial infection. Understanding how pilogenesis is controlled and coordinated *in vivo* presents further challenges for the future, key questions include determining how the ordered array of subunits, characteristic of pili with more complex architectures, is achieved and the role of the membrane-bound usher protein in controlling pilus assembly *in vivo*. The synergy between experiment and simulation will undoubtedly be needed to answer these questions, as well as to design new strategies able to combat disease caused by infection with piliated Gram negative bacteria.

Methods

Force-field and model

Simulations were performed using the CHARMM19 united-atom force-field with an implicit solvent[40] (EEF1) which assumes that the solvation energy of a protein is a sum of group contributions which depend on the group's exposure to the solvent. The temperature of the system was controlled using Langevin dynamics with a friction coefficient of 0.1 ps^{-1}. Integration was performed using the leapfrog algorithm with a 2 fs time-step. Conformations were saved every 500 steps (1 ps) for analysis.

Simulation setup

The crystallographic structures of the SafB:$SafA_{Ntd2}$ (2CO6) and $SafA_{Ntd2}$:Nte (2CO2 and 2CO4 for the F3A and wild-type Nte peptide complexes, respectively)[24] were used as starting conformation for the simulations. The structures were mildly minimised (50 steps steepest descent) and heated from 0 to 300 K by 20 K increments for a total simulation time of 2 ns. During heating, the backbone atoms of both molecules were harmonically constrained to their experimental position. Once the complex was at 300K and equilibrated, all constraints were released and an 8 ns control simulation was performed. During the control simulation positions and velocities were saved every 100 ps to produce independent initial conformations for the biased simulations. In the case of the 2CO2 structure, the region around the Nte alanine substitution at F3 was unresolved. Similarly, in the crystal structure of SafB:$SafA_{Ntd2}$ (2CO6), some residues in the F_1-G_1 loop and the end of the A_1 strand were also not resolved. Before heating, these disordered regions were reconstructed and then, applying a harmonic positional restraint to the backbone atoms but not to those of the disordered residues, the structure was energy minimised and heated using the procedure described above.

Control simulations

In the control simulations, all complexes were stable over 8 ns. The root mean square deviation (RMSD) from the experimental structure after about 2 ns reaches a plateau at about 2.8 Å for 2CO2 and 2CO6 and 3.6 Å for 2CO4. The time series of the total energy and its various components (bonded, electrostatic and van der Waals (not shown)) are also indicative of the stability of the complexes. The native secondary structure is entirely preserved, while all the contacts between $SafA_{Ntd2}$ and the Nte peptide or SafB are also preserved. On the other hand, in the SafB:$SafA_{Ntd2}$ simulation, SafB drifts up to 6 Å RMSD from the initial structure. This does not, however, affect the residues involved in the binding site; moreover, the regions affected during the control simulation were constrained to their experimental position during the simulations of the dissociation of the SafB:$SafA_{Ntd2}$ complex. These findings altogether demonstrate that the force field employed is suitable to study such systems.

Simulations of the SafB:$SafA_{Ntd2}$ complex

Due to the large number of atoms in the complex, simulating the whole SafB:$SafA_{Ntd2}$ complex to full equilibrium was not possible. For this reason, the residues of the SafB molecule not involved in binding to the $SafA_{Ntd2}$ were constrained such that the atoms not directly involved in the binding/unbinding do not move, but are still able to interact with those which do. The important areas of the SafB were deemed to be the A_1, F_1 and G_1 β-strands and the F_1-G_1 loop. This reduces the computation time almost three-fold. A control simulation was carried out for both the constrained and unconstrained models; the results demonstrating that the presence of the constraints does not perturb the stability of the complex (i.e., both in terms of buried surface area and interaction energy no drift was observed). Following the hypothesis that the P5 pocket in the SafB:$SafA_{Ntd2}$ complex is in equilibrium between occupied and unoccupied states[24], this area was observed during the control simulation to reveal any unusually large fluctuations. This was achieved by monitoring the distance between 5 pairs of hydrophobic residues in contact between

SafB and $SafA_{Ntd2}$ and by observing the buried surface area of the complex over time. The buried surface was computed as the sum of the molecular surface of SafA and SafB minus that of the complex; molecular surfaces were computed using the procedure propose by Lee and Richards[41] using a probe radius of 1.6 Å.

Unbinding

Unbinding on a short timescale was simulated using steered molecular dynamics (SMD) and involved applying a harmonic potential to a reaction coordinate; the minimum of the potential was then displaced at speed γ. The biased molecular dynamics (BMD) [35-37] approach employed does not apply a force directly to the reaction coordinate. Instead, a half-quadratic potential is applied to prevent the sizeable decrease of the reaction coordinate while not affecting the system when the coordinate spontaneously increases.

The definition of a reaction coordinate is crucial, since the unbinding observed depends principally on this choice. The following reaction coordinate was used:

$$\rho \propto \sum_{pairs} (r_{ij} - r_{ij}^{0})^2 \qquad (1)$$

where r_{ij} is the distance between atoms i and j, r_{ij}^{o} being this distance at the start of the simulation. The distances summed are those between pairs of atoms in the backbones of the two molecules that are hydrogen bonded and those between C_α atoms of the residues involved in the P1-5 hydrophobic contacts in the crystal structure of the complexes. The pulling speed γ was chosen to be as small as possible in order for the simulation to replicate the natural sequence of events most accurately, but large enough so that simulations could be performed in a reasonable time. Increasing this reaction coordinate induces unbinding but, importantly, without specifically biasing the mechanism or initiation point.

In the case of SMD, the overall time needed to induce the unbinding is determined by the retraction of the harmonic potential; in BMD, the speed of the unbinding depends only on the choice of the force constant α. For the SafB:SafA$_{Ntd2}$ unbinding simulations, the force constant was chosen to be α=0.05 kcal mol^{-1} Å^4 (for SMD simulations) and α=0.04 kcal mol^{-1} Å^4 (BMD simulations) and the speed, γ =0.4 Å ps^{-1} (for SMD simulations). Parameters used for the SafA$_{Ntd2}$:Nte unbinding simulations were α = 0.005 kcal mol^{-1} Å^4 (SMD and BMD) and γ = 0.2 Å ps^{-1} (SMD).

Unbinding using high temperature

Another method often applied to simulations to accelerate unbinding or unfolding is to simulate the events occurring at high temperatures. To investigate this, each complex was heated again after equilibration to a designated temperature and then allowed to undergo dynamics at this constant temperature with no external perturbations. A number of temperatures in the range 350-450K were used. The speed of the unbinding is regulated by the temperature; the higher the temperature, the faster the unbinding is expected to progress. However, at the lower temperatures unbinding did not occur on a reasonably short timescale, reflecting the experimental high stability of the complex. At 450 K, in one simulation of the 2CO4 complex, SafA$_{Ntd2}$ in the SafA$_{Ntd2}$:Nte complex denatured after about 16 ns before unbinding occurred. Similar results were found for the SafB:SafA$_{Ntd2}$ complex (data not shown). These trials demonstrated that to induce the unbinding of such stable complexes simply increasing the temperature is not a suitable alternative to the SMD and BMD procedures described above.

Acknowledgments

TSW acknowledges the BBSRC for a stipend during a summer internship. EP acknowledges INTAS for the grant 05-10000004-7747. RJR is funded by a BBSRC CASE award with Waters UK Ltd.

The authors would like to thank Dr Han Remaut for his many insights and valuable discussions.

Figure Captions.

Figure 1

(a) Saf pilus biogenesis at outer membrane. Chaperone:subunit complexes in periplasm are targeted to the usher assembly platform at the outer membrane. Subunits polymerise by donation of their Nte to an adjacent subunit at the usher pore, which extrudes the growing pilus. For the Saf pili that are the focus of this manuscript, the major pilus subunit SafA is shown in blue; the chaperone, SafB, in brown; the usher, SafC, in grey; and distal adhesin, SafD, in purple.

(b) Crystal structure of $SafA_{Ntd2}$ (blue) in complex with SafB (red) (2CO6)[24]. Indicated are the A_1, F_1 and G_1 β-strands of SafB while the F_1-G_1 loop is not resolved. Also shown are the highly conserved hydrophobic residues P1-P5 of the chaperone that insert into the hydrophobic binding pockets of $SafA_{Ntd2}$.

(c) Crystal structure of $SafA_{Ntd2}$ (blue) in complex with the Nte peptide (red) (2CO4)[24]. The highly conserved hydrophobic residues P1-P5 of the Nte are shown, as is the P* binding residue.

Figure 2

(a) RMSD from the X-ray structure of SafB:$SafA_{Ntd2}$, **(b)** the buried surface area between SafB and $SafA_{Ntd2}$ and **(c)** the distance between the SafB P5 residue (A114 nitrogen) and the $SafA_{Ntd2}$ A strand (A35 carboxyl carbon), in red, or the $SafA_{Ntd2}$ F strand (A130 carboxyl carbon), in black, over the course of the simulation.

Figure 3

(a) Superposition of the SafB:$SafA_{Ntd2}$ complex in the two different states. The conformation of the complex in which A114 (indicated as P5) is closest to $SafA_{Ntd2}$ strand A is shown in red. The conformation in which A114 is closest to $SafA_{Ntd2}$ strand F is shown in green. A114 moves 2.6 Å between the two structures. **(b,c,d)** Enlargement around the P5 pocket where $SafA_{Ntd2}$ is

represented as solvent accessible surface: $SafA_{Ntd2}$ in complex with **(b)** SafB (red) in the conformation in which A114 is close to $SafA_{Ntd2}$ A strand, **(c)** SafB (green) in the conformation in which A114 is closest to $SafA_{Ntd2}$ F strand, and **(d)** the Nte peptide. A114 and F17 are shown in ball and stick representation.

Figure 4

(a) Root mean square fluctuations for $SafA_{Ntd2}$ obtained from the 164 ns simulation of the complex at 300K (red, left scale) compared with the root mean square fluctuations obtained from crystallographic B factors for structure 2CO6[24] (black, right scale). While the magnitude of the latter is smaller, the two curves are similar displaying large fluctuations in the loops which contact P5 (residues 33-40, 78-82, and 123-132), plus one additional loop (residues 63-68).

(b) SafB (red) in complex with $SafA_{Ntd2}$ (dark blue). Regions of high flexibility are indicated in yellow (residues 33-40), pink (residues 63-68), green (residues 78-82) and light blue (123-132).

Figure 5

(a) Energy associated to the bias during the unbinding of SafB:$SafA_{Ntd2}$ complex using SMD, **(b)** distances between SafB residues and $SafA_{Ntd2}$ interaction partners from SMD unbinding simulations and **(c)** from BMD unbinding simulations. The distances shown are between: SafB-A114(N):$SafA_{Ntd2}$-A130(C), black; SafB-L116(N):$SafA_{Ntd2}$-L132(C), red; SafB-L118(N):$SafA_{Ntd2}$-P134(C), green; SafB-L120(N):$SafA_{Ntd2}$-L137(C), blue; SafB-A122(N):$SafA_{Ntd2}$-A138(C), magenta which represent the interactions in the P5, P4, P3 , P2 and P1 pockets.

Figure 6

Unbinding simulations of $SafA_{Ntd2}$ in complex with **(a,c)** wild-type Nte and **(b,d)** F3A Nte. **(a,b)** Energy associated to the bias during SMD simulations. **(c,d)** Distances between N-terminal extension residues and $SafA_{Ntd2}$ interaction partners. The distances shown are between: $SafA_{Ntd2}$-

D131(C):Nte-F17(N), black; $SafA_{Ntd2}$-I135(C):Nte-V13(N), red; $SafA_{Ntd2}$-V139(C):Nte-Q9(N), green; $SafA_{Ntd2}$-G141(C):Nte-F3(N) or A3(N), green.

References

1. Chiti, F. & Dobson, C. M. (2006). Protein misfolding, functional amyloid, and human disease. *Annu Rev Biochem* **75**, 333-66.
2. Dobson, C. M. (2003). Protein folding and misfolding. *Nature* **426**, 884-90.
3. Remaut, H. & Waksman, G. (2006). Protein-protein interaction through beta-strand addition. *Trends Biochem Sci* **31**, 436-44.
4. Thanassi, D. G. & Hultgren, S. J. (2000). Multiple pathways allow protein secretion across the bacterial outer membrane. *Curr Opin Cell Biol* **12**, 420-30.
5. Stathopoulos, C., Hendrixson, D. R., Thanassi, D. G., Hultgren, S. J., St Geme, J. W., 3rd & Curtiss, R., 3rd. (2000). Secretion of virulence determinants by the general secretory pathway in gram-negative pathogens: an evolving story. *Microbes Infect* **2**, 1061-72.
6. Sauer, F. G., Pinkner, J. S., Waksman, G. & Hultgren, S. J. (2002). Chaperone priming of pilus subunits facilitates a topological transition that drives fiber formation. *Cell* **111**, 543-51.
7. Zavialov, A. V., Berglund, J., Pudney, A. F., Fooks, L. J., Ibrahim, T. M., MacIntyre, S. & Knight, S. D. (2003). Structure and biogenesis of the capsular F1 antigen from Yersinia pestis: preserved folding energy drives fiber formation. *Cell* **113**, 587-96.
8. Choudhury, D., Thompson, A., Stojanoff, V., Langermann, S., Pinkner, J., Hultgren, S. J. & Knight, S. D. (1999). X-ray structure of the FimC-FimH chaperone-adhesin complex from uropathogenic Escherichia coli. *Science* **285**, 1061-6.
9. Sauer, F. G., Futterer, K., Pinkner, J. S., Dodson, K. W., Hultgren, S. J. & Waksman, G. (1999). Structural basis of chaperone function and pilus biogenesis. *Science* **285**, 1058-61.
10. Lindberg, F., Lund, B., Johansson, L. & Normark, S. (1987). Localization of the receptor-binding protein adhesin at the tip of the bacterial pilus. *Nature* **328**, 84-7.
11. Kuehn, M. J., Heuser, J., Normark, S. & Hultgren, S. J. (1992). P pili in uropathogenic E. coli are composite fibres with distinct fibrillar adhesive tips. *Nature* **356**, 252-5.
12. Knight, S. D., Berglund, J. & Choudhury, D. (2000). Bacterial adhesins: structural studies reveal chaperone function and pilus biogenesis. *Curr Opin Chem Biol* **4**, 653-60.
13. Remaut, H. & Waksman, G. (2004). Structural biology of bacterial pathogenesis. *Curr Opin Struct Biol* **14**, 161-70.
14. Pinkner, J. S., Remaut, H., Buelens, F., Miller, E., Aberg, V., Pemberton, N., Hedenstrom, M., Larsson, A., Seed, P., Waksman, G., Hultgren, S. J. & Almqvist, F. (2006). Rationally designed small compounds inhibit pilus biogenesis in uropathogenic bacteria. *Proc Natl Acad Sci U S A* **103**, 17897-902.
15. Sanford, K. & Kumar, M. (2005). New proteins in a materials world. *Curr Opin Biotechnol* **16**, 416-21.
16. Fairman, R. & Akerfeldt, K. S. (2005). Peptides as novel smart materials. *Curr Opin Struct Biol* **15**, 453-63.
17. Zhang, S. (2003). Fabrication of novel biomaterials through molecular self-assembly. *Nat Biotechnol* **21**, 1171-8.

18. Thanassi, D. G., Saulino, E. T. & Hultgren, S. J. (1998). The chaperone/usher pathway: a major terminal branch of the general secretory pathway. *Curr Opin Microbiol* **1**, 223-31.
19. Townsend, S. M., Kramer, N. E., Edwards, R., Baker, S., Hamlin, N., Simmonds, M., Stevens, K., Maloy, S., Parkhill, J., Dougan, G. & Baumler, A. J. (2001). Salmonella enterica serovar Typhi possesses a unique repertoire of fimbrial gene sequences. *Infect Immun* **69**, 2894-901.
20. Manting, E. H. & Driessen, A. J. (2000). Escherichia coli translocase: the unravelling of a molecular machine. *Mol Microbiol* **37**, 226-38.
21. Sauer, F. G., Barnhart, M., Choudhury, D., Knight, S. D., Waksman, G. & Hultgren, S. J. (2000). Chaperone-assisted pilus assembly and bacterial attachment. *Curr Opin Struct Biol* **10**, 548-56.
22. Dodson, K. W., Jacob-Dubuisson, F., Striker, R. T. & Hultgren, S. J. (1993). Outer-membrane PapC molecular usher discriminately recognizes periplasmic chaperone-pilus subunit complexes. *Proc Natl Acad Sci U S A* **90**, 3670-4.
23. Saulino, E. T., Thanassi, D. G., Pinkner, J. S. & Hultgren, S. J. (1998). Ramifications of kinetic partitioning on usher-mediated pilus biogenesis. *Embo J* **17**, 2177-85.
24. Remaut, H., Rose, R. J., Hannan, T. J., Hultgren, S. J., Radford, S. E., Ashcroft, A. E. & Waksman, G. (2006). Donor-strand exchange in chaperone-assisted pilus assembly proceeds through a concerted beta strand displacement mechanism. *Mol Cell* **22**, 831-42.
25. Verger, D., Bullitt, E., Hultgren, S. J. & Waksman, G. (2007). Crystal structure of the P pilus rod subunit PapA. *PLoS Pathog* **3**, e73.
26. Verger, D., Miller, E., Remaut, H., Waksman, G. & Hultgren, S. (2006). Molecular mechanism of P pilus termination in uropathogenic Escherichia coli. *EMBO Rep* **7**, 1228-32.
27. Sauer, F. G., Remaut, H., Hultgren, S. J. & Waksman, G. (2004). Fiber assembly by the chaperone-usher pathway. *Biochim Biophys Acta* **1694**, 259-67.
28. Holmgren, A. & Branden, C. I. (1989). Crystal structure of chaperone protein PapD reveals an immunoglobulin fold. *Nature* **342**, 248-51.
29. Pellecchia, M., Guntert, P., Glockshuber, R. & Wuthrich, K. (1998). NMR solution structure of the periplasmic chaperone FimC. *Nat Struct Biol* **5**, 885-90.
30. Knight, S. D., Choudhury, D., Hultgren, S., Pinkner, J., Stojanoff, V. & Thompson, A. (2002). Structure of the S pilus periplasmic chaperone SfaE at 2.2 A resolution. *Acta Crystallogr D Biol Crystallogr* **58**, 1016-22.
31. Vetsch, M., Erilov, D., Moliere, N., Nishiyama, M., Ignatov, O. & Glockshuber, R. (2006). Mechanism of fibre assembly through the chaperone-usher pathway. *EMBO Rep* **7**, 734-8.
32. Vitagliano, L., Ruggiero, A., Pedone, C. & Berisio, R. (2007). A molecular dynamics study of pilus subunits: Insights into pilus biogenesis. *Journal of Molecular Biology* **367**, 935-941.
33. Curcio, R., Caflisch, A. & Paci, E. (2005). Change of the unbinding mechanism upon a mutation: a molecular dynamics study of an antibody-hapten complex. *Protein Sci* **14**, 2499-514.
34. West, D. K., Brockwell, D. J., Olmsted, P. D., Radford, S. E. & Paci, E. (2006). Mechanical resistance of proteins explained using simple molecular models. *Biophys J* **90**, 287-97.
35. Paci, E., Caflisch, A., Pluckthun, A. & Karplus, M. (2001). Forces and energetics of hapten-antibody dissociation: A biased molecular dynamics simulation study. *Journal of Molecular Biology* **314**, 589-605.

36. Paci, E. & Karplus, M. (1999). Forced unfolding of fibronectin type 3 modules: An analysis by biased molecular dynamics simulations. *Journal of Molecular Biology* **288**, 441-459.
37. Paci, E., Smith, L. J., Dobson, C. M. & Karplus, M. (2001). Exploration of partially unfolded states of human alpha-lactalbumin by molecular dynamics simulation. *J Mol Biol* **306**, 329-47.
38. Zav'yalov, V. P., Zav'yalova, G. A., Denesyuk, A. I. & Korpela, T. (1995). Modelling of steric structure of a periplasmic molecular chaperone Caf1M of Yersinia pestis, a prototype member of a subfamily with characteristic structural and functional features. *FEMS Immunol Med Microbiol* **11**, 19-24.
39. Hung, D. L., Knight, S. D., Woods, R. M., Pinkner, J. S. & Hultgren, S. J. (1996). Molecular basis of two subfamilies of immunoglobulin-like chaperones. *Embo J* **15**, 3792-805.
40. Lazaridis, T. & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins* **35**, 133-52.
41. Lee, B. & Richards, F. M. (1971). Interpretation of Protein Structures - Estimation of Static Accessibility. *Journal of Molecular Biology* **55**, 379-400.

Figure 1

Click here to download high resolution image

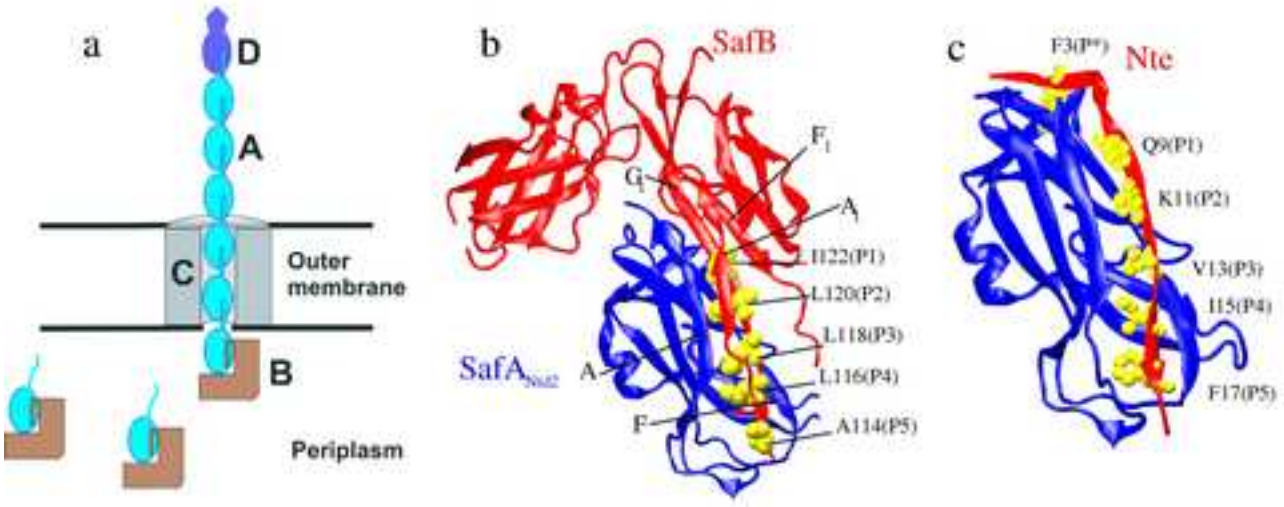

Figure 2

Click here to download high resolution image

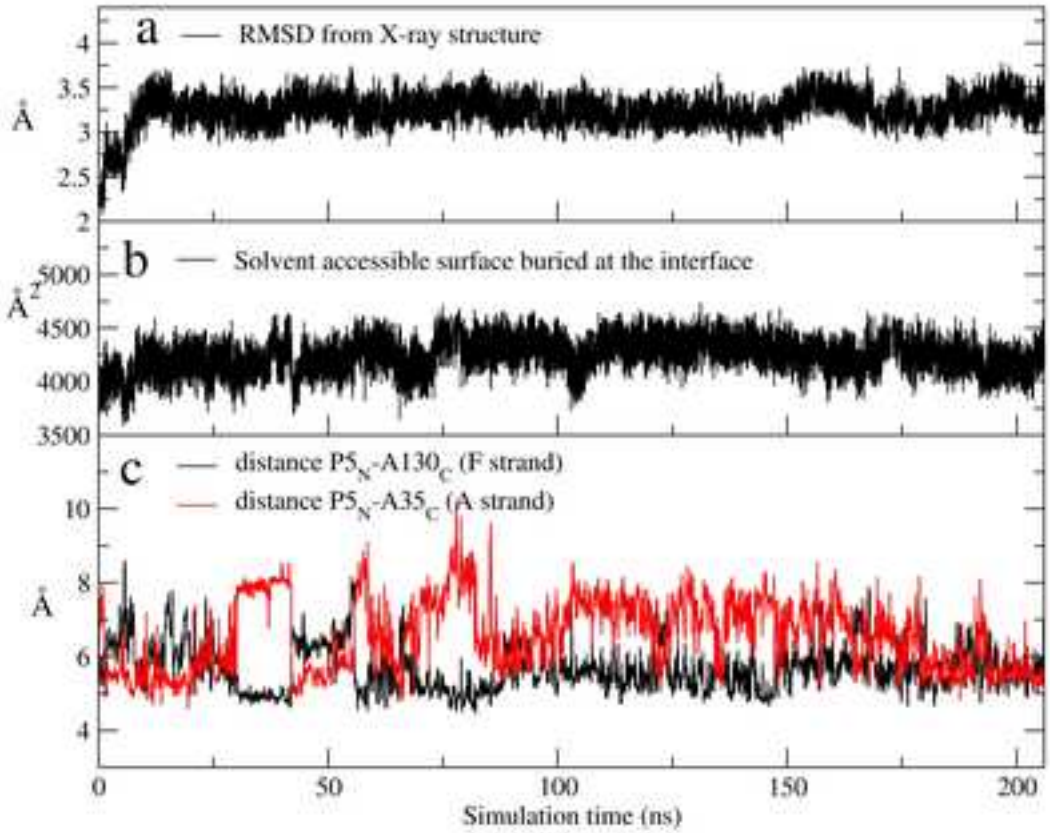

Figure 3

Click here to download high resolution image

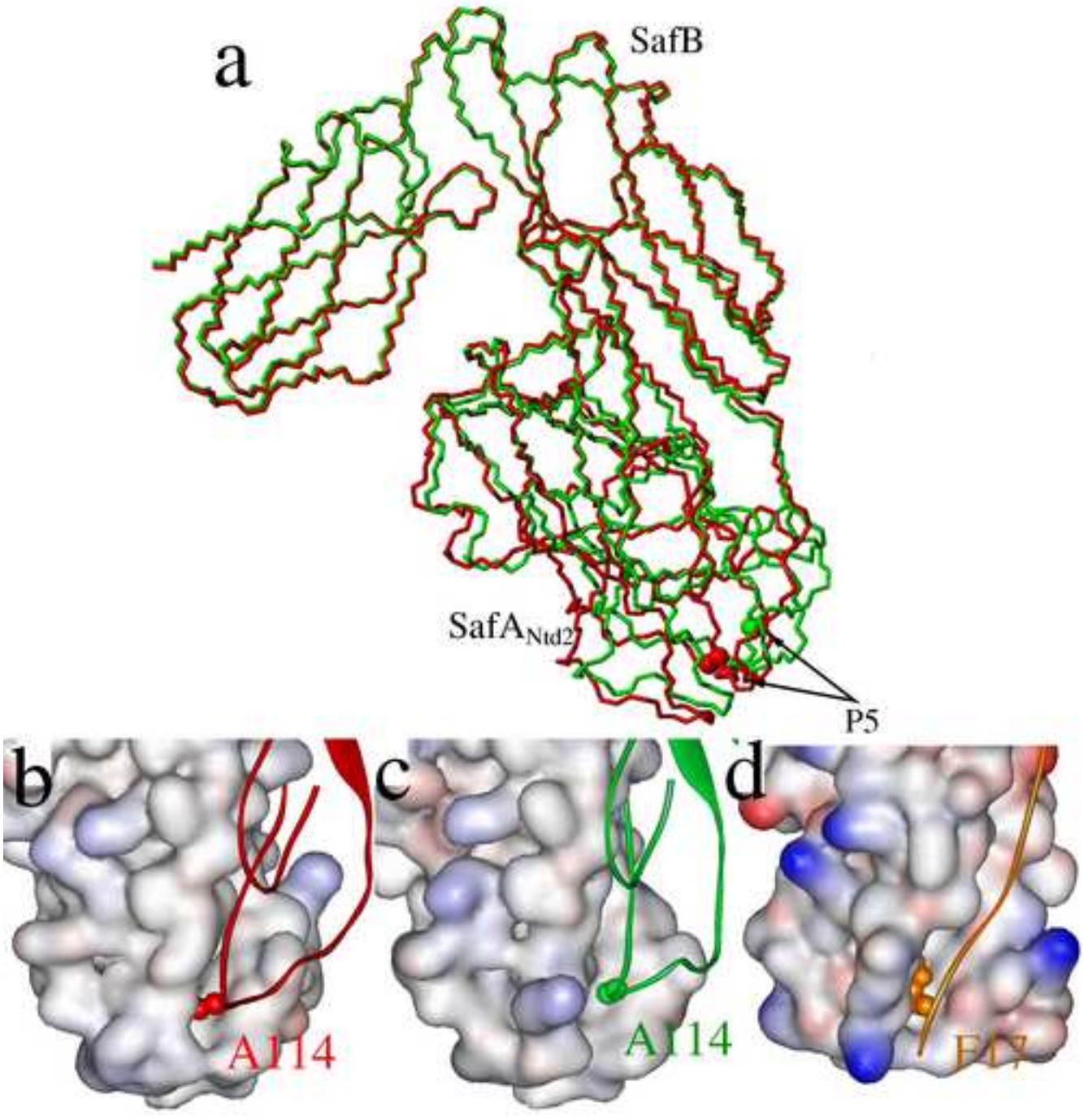

Figure 4

Click here to download high resolution image

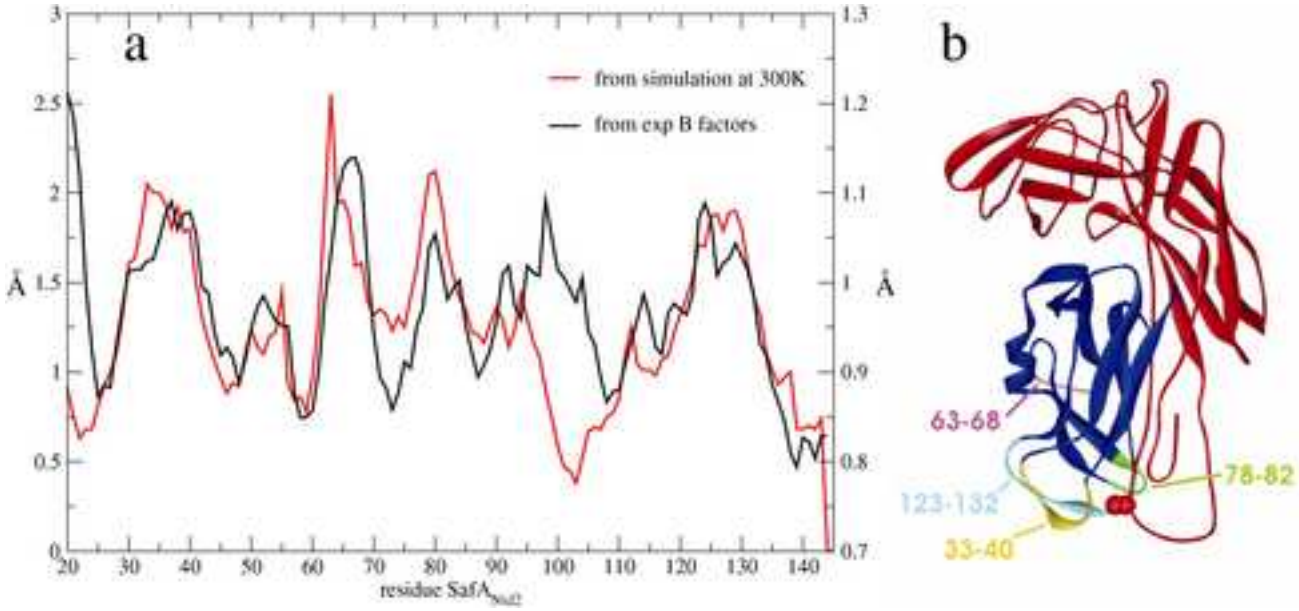

Figure 5

Click here to download high resolution image

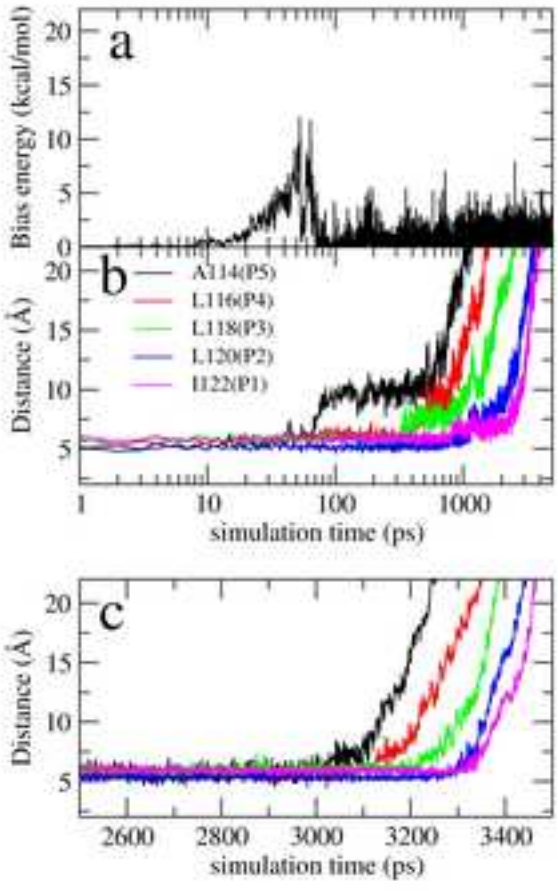

Figure 6

Click here to download high resolution image