Chapter 6

Flagellin refolding before assembly

6.1 How does flagellin refold in the "chamber"?

In the T3SS needle complex of a pathogenic bacterium, effector proteins unfolded for translocation through the hollow needle to the host cell are refolded in the cytoplasm of the host cell [Galán and Wolf-Watz, 2006]. Any misfolded proteins could be repaired or degraded by the host cell's machinery. In contrast, unfolded flagellin might have to complete refolding within a small space below the filament cap (HAP2) measuring 70 Å deep by about 40 Å across that is surrounded by domain D1 of the uppermost turn of assembled flagellin [Yonekura et al., 2000]. Any misfolded flagellin might stall the self-assembly process. The HAP2 chaperone "leg domains" recognize folded D1 [Maki-Yonekura et al., 2003], whose proper folding depends on that of the HVR domains [Honda et al., 1999]. But the "refolding chamber" is probably too small to contain any chaperones that might help refold any misfolded flagellin molecules. Could flagellin refold via a mechanism that reduces the chance of misfolding?

In this chapter, I present results from thermal unfolding simulations of flagellin monomers. I aim to understand the refolding mechanism of flagellin inside the "chamber". A similar order of unfolding as in thermal denaturation experiment was found. I also found contacts within three-stranded β-sheets remaining for long periods of simulation times in the thermally denatured HVR domains, suggestive of folding nucleation sites that would assist in the refolding of flagellin. The simulation results are used to interpret some of the findings in the thermal denaturation experiment [Honda et al., 1999] and in a mutagenesis study on flagellin mechanical stability [Malapaka et al., 2007].

Note: Contents in this chapter has been published in the May 2008 issue of the Bio-

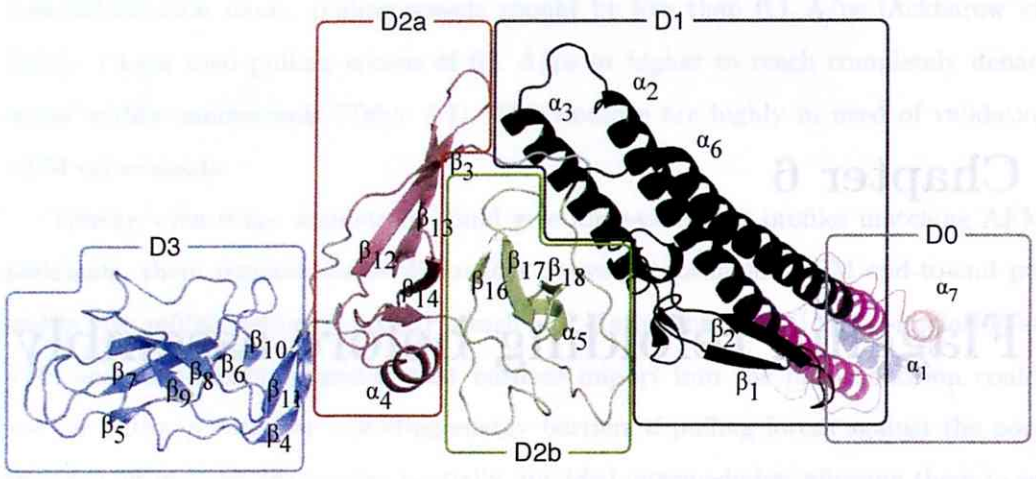


Figure 6.1: **The tertiary structure of flagellin monomer** in solution obtained from MD simulation (starting from *polymeric* flagellin structure in 1UCU.PDB) showing DSSP-assigned α-helices and β-strands. Each (sub)domain is colored differently and labeled, with linkers colored gray. N-terminus (blue) and C-terminus (red) helices are disordered in the monomer. I have neglected the very small β-turn from residues 130 to 135 in my labeling scheme. Segment D_f1 of D1 is in black, with the remaining portion in magenta. Figure rendered with PyMOL [DeLano, 2002].

physical Journal [Chng and Kitao, 2008]. Permission has been granted by the Biophysical Society for use of text, figures and tables from my paper.

6.2 Methods

I have carried out MD simulation in aqueous solvent at 300 K (control) and five runs each at 400, 500 and 600 K to partially and fully denature the flagellin monomer. Simulation details are laid out below. For convenience of discussion, a more detailed model of the flagellin monomer is presented in Fig. 6.1. References to the following subdomains will be made in the text: N-terminal of D0 (ND0) and CD0 are made up of α_1 and α_7, respectively; ND1 includes α_2, α_3, β_1 and β_2 wherease CD1 consists of α_6 only.

6.2.1 HT-MD simulation setup

The equilibrated monomeric flagellin (plus a 5 Å shell of surface water molecules) was re-immersed into a wider rectangular box with at least 10 Å between the protein+surface-water "molecule" and simulation box edges. The new simulation box becomes 216 Å × 105 Å × 84 Å and contains 159,162 atoms. The atoms were then reassigned velocities at 300 K and the system temperature increased to 600 K in stages, following the equilibration protocol as mentioned above. By setting different initial velocities, five independent sets of simulations were set up at 400 K, 500 K and 600 K and monitored for equilibrium in temperature and pressure. After this has been achieved, temperature and pressure controls are turned off so as not to affect energy conservation. The unfolding process at 400 K, 500 K and 600 K were then monitored for 8-ns, 6-ns and 2-ns respectively with a 1-fs time-step. It has been demonstrated that a small set of five to ten simulations are sufficient to capture the average properties and pathway during unfolding [Day and Daggett, 2005]. The 8-ns simulation at 300 K with a 2-fs time-step (with bonds involving H atom constrained) served as the control. I carried out these simulations with the PMEMD module of AMBER 8 on Itanium-2 computer clusters, taking 33 hrs/ns on 16 CPUs. Trajectory frames were saved every 0.5-ps but resampled to 2-ps interval for analysis.

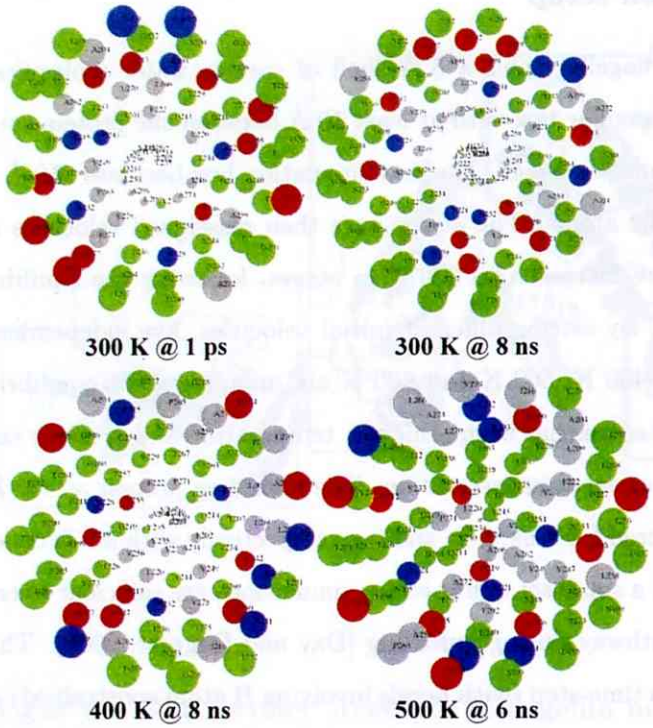
6.3 Results

6.3.1 Order of domain unfolding

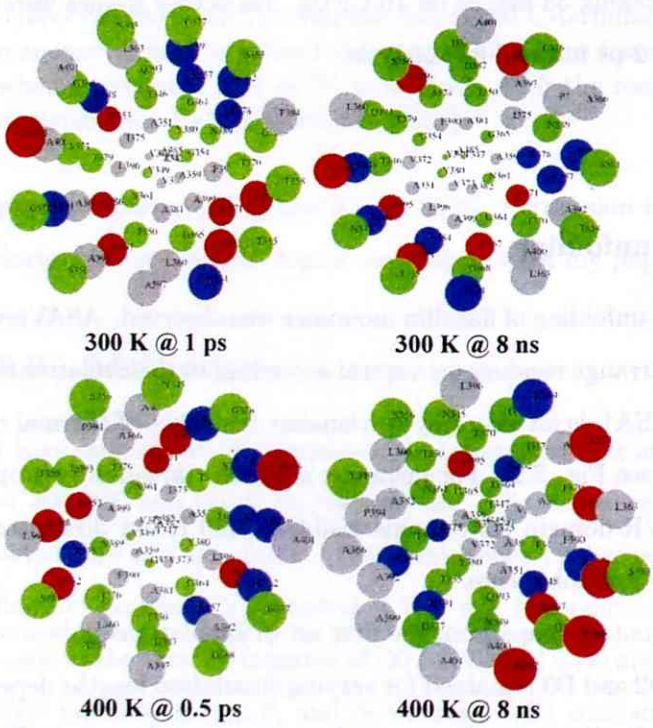
For temperatures above 300 K, unfolding of flagellin monomer was observed. ASAView [Ahmad et al., 2004] plots, which arrange residues on a spiral according to their relative Solvent Accessibility Surface Area (SASA), is an easy way to visualize the effect of thermal denaturation on a protein/domain (see Fig. 6.2). For instance, a narrow thread of hydrophobic residues from the center of HVR domain D3 remains visible at 400 K but disappeared at 500 K, indicating loss of its hydrophobic core.

Time-averaged persistent contact maps from the first set of simulations indicated that pairs of β-strands in domains D2 and D3 remained for varying simulation lengths depending on the temperature (Fig. 6.3 *A*). Corresponding secondary structure changes in Fig. 6.3 *B* also revealed persistent β-strands in D2 and D3 while α-helices in D0 and D1 became denatured.

The fraction of persistent native contacts (defined in Chapter 4) that has been retained



(a) D3



(b) D2b

Figure 6.2: **Solvent exposure during folding.** Denaturation of globular flagellin domain D3 (a) and subdomain D2b (b) as visualized by ASAView spiral plots. Color code: gray (hydrophobic), red (negative charged), blue (positive charged) and green (neutral polar).

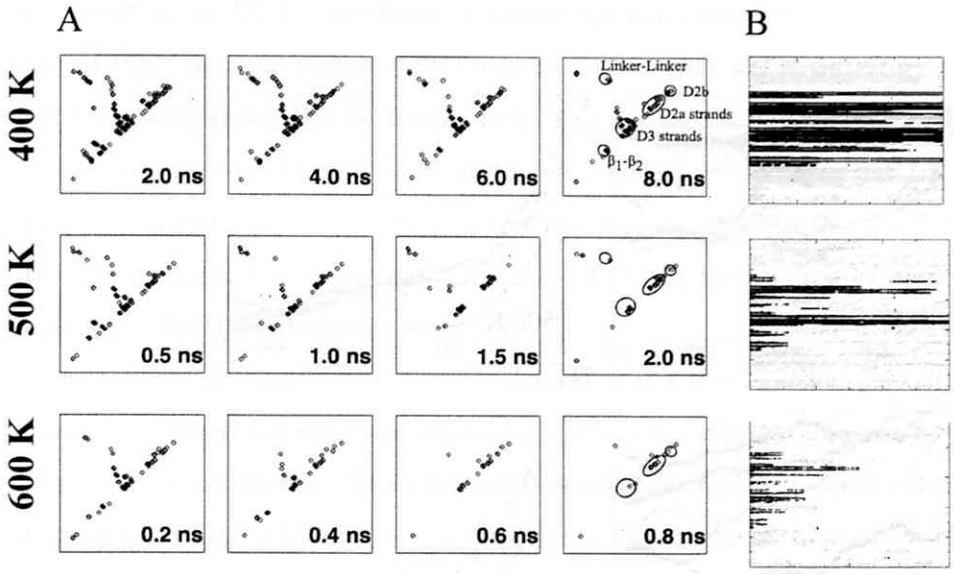


Figure 6.3: **Structural changes during denaturation.** (*A*) Changes to persistent contact maps from the first set of simulations. Here, only persistent contacts within each time window of 2-ns (400 K), 0.5-ns (500 K) or 0.2-ns (600 K) are plotted. (*B*) Changes to DSSP assigned secondary structures from the same simulation set (black=β-strand; gray=α-helix), with plot for 300 K shown as inset to Fig. 4.2. Note that simulation lengths differ: 8-ns (400 K), 6-ns (500 K) and 2-ns (600 K).

under thermal denaturing conditions is a crude indicator of the unfolding order. Fig. 6.4 shows data from the first set of simulations under each temperature, which indicated that D0 helices denatured very easily. D_f1 and D2b were the first structured domains to be denatured. In contrast, for D2 (made up of D2a and D2b) and D3 the fraction remained above 0.1 under 500 K (the value reached by CD0 under 400 K, taken to indicate complete unfolding). Despite a similar decay trend between D2 and D3, D2a contacts were found to be generally better maintained than those of D3.

Although D_f1 remained stable under the control simulation at physiological temperature of 300 K, it became unfolded under "mild" denaturation temperature of 400 K (Fig. 6.4 *B*). Looking at the fractional contacts between D_f1 and D2a/b among the other four sets of 400 K simulation trajectories (Fig. 6.4 *E*), I noticed that D2a fractional contacts remained above or about 0.7 while that for D_f1 dropped to around 0.4 and all plots seem to follow the profiles for D2b that is also marginally stable. Among these trajectories, hydrophobic side-chains of residues on the D1-D2 linker loop remain well packed in the D_f1 core and D_f1 helices suffered minimal distortions. The D2a $\beta_{12}\beta_{13}$ hairpin and ND1a-ND1b helix-turn-helix segment are displaced relative to each other in some trajectories

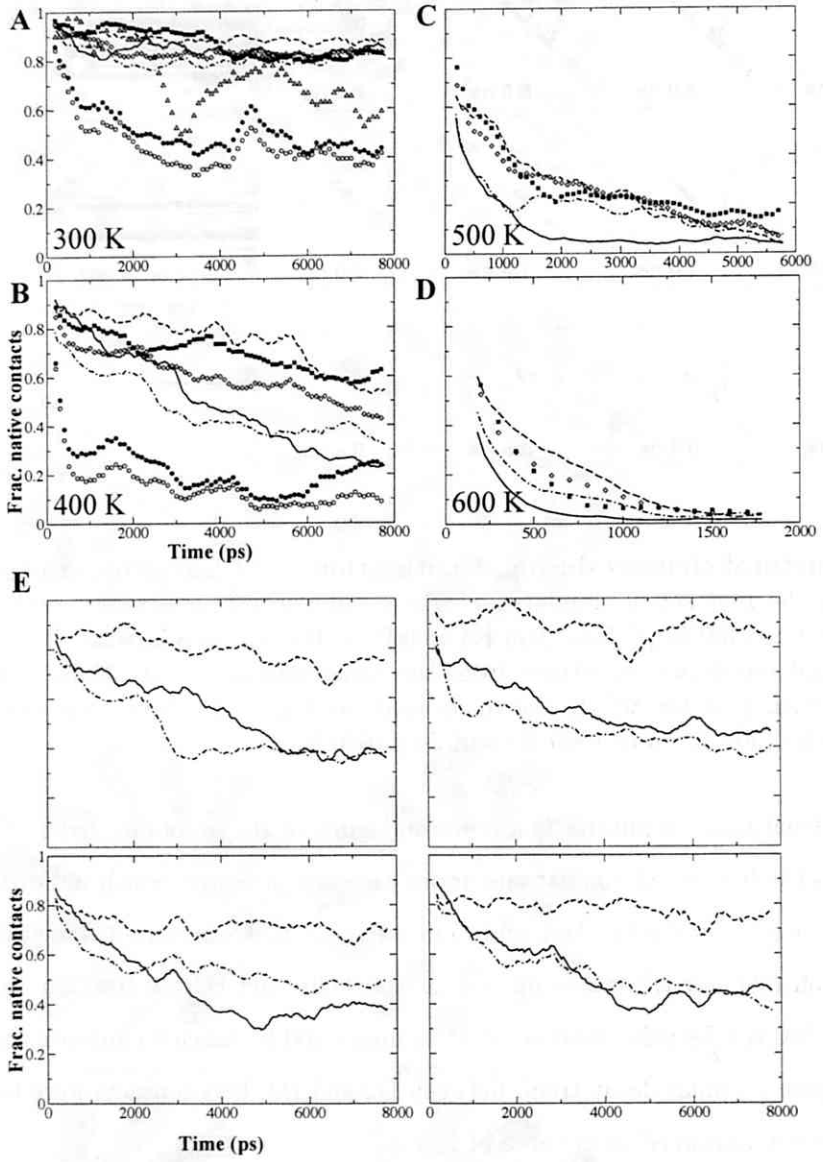


Figure 6.4: **Domain-level fractional contacts.** (A to D) Change of normalized fractional contacts of (sub)domains with simulation time, taken from the first set of simulations. Only contacts present in the starting structure were monitored and resulting fractional contacts normalized for better comparison between subdomains: D_f1 (thick continuous line), D2 (open diamonds), D2a (dashed line), D2b (dot-dashed line), D3 (hatched squares), ND0 (filled circles) and CD0 (open circles). Note that D0 fractional contacts are shown only for 300 and 400 K. Fractional contacts for D1 excluding D_f1 segment are shown only for 300 K (open triangles). (E) Plots from the remaining four 400 K trajectories of 8-ns each, showing only fractional contacts for D_f1, D2a and D2b. Note that abscissa and ordinate for all four plots are the same. In (A to E), data points for D_f1, D2a, D2b have been resampled at 40-ps interval with a 20-ps window running-average applied for smoothing; for the rest, resampling interval is 100-ps with a 10-ps running-averaging.

(thus breaking the D1-D2 interfacial H-bonds) but maintained in others. The further decrease of D2a fractional contacts after 6-ns during the first 400 K trajectory (Fig. 6.4 *B*) could have destabilized D_f1 by further extracting the D1-D2 linker from the hydrophobic core: an increase in SASA for L167 and L169 occurred around 6-ns but L159 remained buried throughout (data not shown). There was also increased distortions to D_f1 helices with loss of regular secondary structures corresponding to the drop of native fractional contacts to a lower value of 0.2.

The sole use of fractional contacts to determine if a (sub)domain is denatured do not work well for D2b as the value can remain higher than 0.1 while loss of secondary structures continues (data not shown). The situation is reversed for D_f1. To capture both aspects of denaturation, I used a hybrid metric: the square-root of the product of fractional native contacts and fractional secondary structure content. When the 200-ps running average of this metric first dropped below 0.2, a value somewhat arbitrarily assigned, I say that the (sub)domain has denatured. Using this criteria to assign the global unfolding events and other criteria for the local unfolding events, I constructed a "state-diagram" for all the simulations as shown in Fig. 6.5. I noticed a general conservation of global unfolding events, such as D2b and D2a being the first and last to denature, respectively.

6.3.2 Partial unfolding/refolding of subdomain D2b

The D2b subdomain has a three-stranded β-sheet and only a weak hydrophobic core composed of residues V373, A385 and T347, according to ASAView (Fig. 6.2). Threonine actually has a small hydropathy of -0.7 according to the Kyte & Doolittle scale, just below that of Ala and Gly. This weak core remains stable under control conditions. The surface-exposed aromatic cluster (Y349, Y380, F390) was found to be even less stable against thermal denaturation than the weak hydrophobic core. The functional role of this cluster is as yet unknown.

From the first 400 K simulation trajectory, I found that the R_g increased gradually from 11 Å and suddenly jumped to 16 Å around 4-ns before dropping back to around 12 Å (Fig. 6.6 *B*). This spike could be due to the partial unfolding and refolding of the hydrophobic core region, with T347 separated from the A385-V373 pair. The "opening-up" of D2b is illustrated in Fig. 6.7 via simulation snapshots. Combined SASAs of selected core residues from all simulations are shown in Fig. 6.8. A threshold of 80 Å^2 was adopted from the averaged SASA of polar residues reported in an early MD study [Zheng et al.,

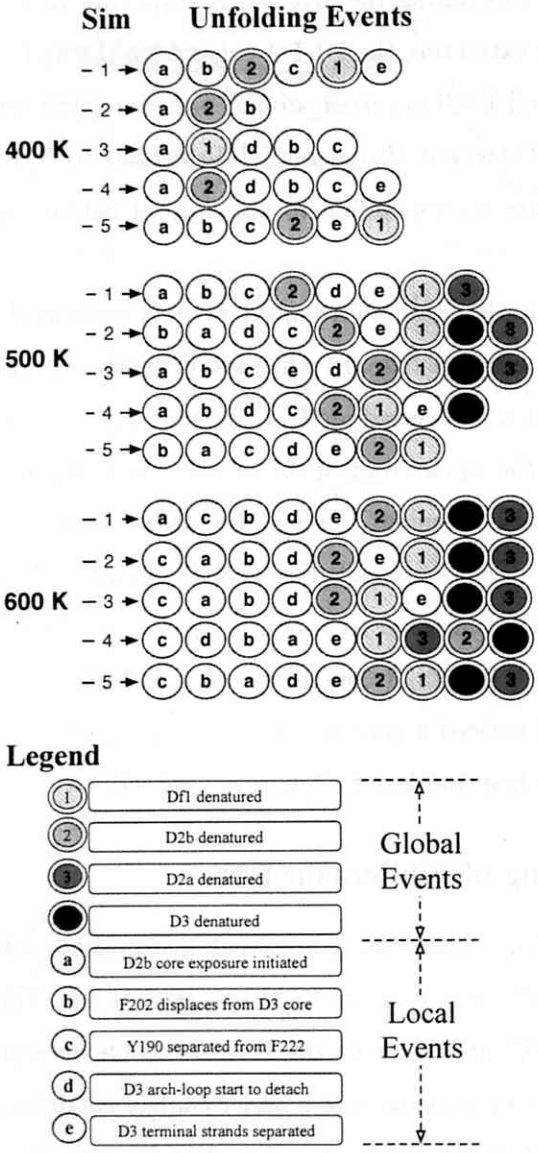


Figure 6.5: **Unfolding events.** A symbolic representation of major events in the thermal unfolding of flagellin from all simulations conducted in this study. The criteria used were as follows. States 1 to 4: when the 200-ps running average of the contact metric (see text for definition) first dropped below the 0.2 threshold. Fractional secondary structure content at time t in a trajectory is defined as the number of residues assigned as helix (D_f1) or strand (D2a, D2b, D3) divided by the maximum number from the whole trajectory. State **a**: when the maximum instantaneous residue SASA among the D2b hydrophobic core residues first exceeded 30 Å^2, which is the maximum observed during the control simulation. State **b**: similar to state **a**, but only one residue is considered and threshold is 25 Å^2. State **c**: when minimum heavy atom separation of Y190 and F222 first exceeded the contact threshold without subsequent fluctuations about the threshold. States **d** and **e**: when the 200-ps running averaged fractional native contact dropped below 0.2.

1990]. Fluctuations of D2b SASA at 400 K are large compared to those in the control or to those of D3 which remain folded. The above observations indicate that D2b has very low stability and would become folded only after D2a and D3.

6.3.3 Unfolding pathway of domain D3

From the thermal unfolding simulations, I identified key events during the unfolding of domain D3. In the "state-diagram" (Fig. 6.5), I present the order of these events (local events **b** to **d**) as observed from each simulation. Conformation snapshots of D3 unfolding are presented in reversed time order in Fig. 6.9.

Event **b** is the displacement of F202 from the hydrophobic core of D3, the precursor to solvent exposure of other core residues. Event **c** is the detachment of terminal anti-parallel β-sheet from the rest of D3 "body", marked by separation of Y190 on the sheet from F222 on $\beta_6\beta_7$ hairpin. This separation allows extension of the β-sheet which then acts as a linker between D3 and D2 which allowed for relative rotations of the domains as observed from my simulations (data not shown). Y190 and F222 side-chains form favorable $CH\cdots\pi$ interactions [Hunter and Sanders, 1990], which together with F229 makes up a surface aromatic cluster. Such clusters have been found to increase thermal stability of thermophiles [Kannan and Vishveshwara, 2000]. The Y190-F222-F229 aromatic cluster might have also protected the linker backbone hydrogen-bonds from solvent attacks which would weaken the linker especially when mechanical stresses are applied to the linker region such as during rotation of the flagellum. Role of mechanical force has been attributed to destabilizing H-bonds to allow for replacement with residue-solvent ones [Pabón and Amzel, 2006]. Lastly, event **d** marks the detachment of the "arch-loop" connecting β_9 with β_{10} and that wraps over the β-folium sheet, through monitoring the separation between Proline (on the loop)-Tyrosine residue pairs (see Fig. 6.12). Subsequent unfolding reduces the size of the β-folium sheet. The consensus order of these events during folding might be: **d** to **c** to **b**. This order suggests that completion of the hydrophobic core is the finishing touch and not the crucial phase in the folding of D3 (Fig. 6.9).

To further analyze the unfolding pathway, I performed property space Principal Component Analysis (PCA) on all five trajectories under 400 to 600 K based on the following six properties: fraction of native contacts, native backbone H-bonds, β-sheet content, R_g, SASA and SASA (hydrophobic) [Kazmirski et al., 1999]. This procedure allows us to assess which physical property changes the most during unfolding, judged by the size of

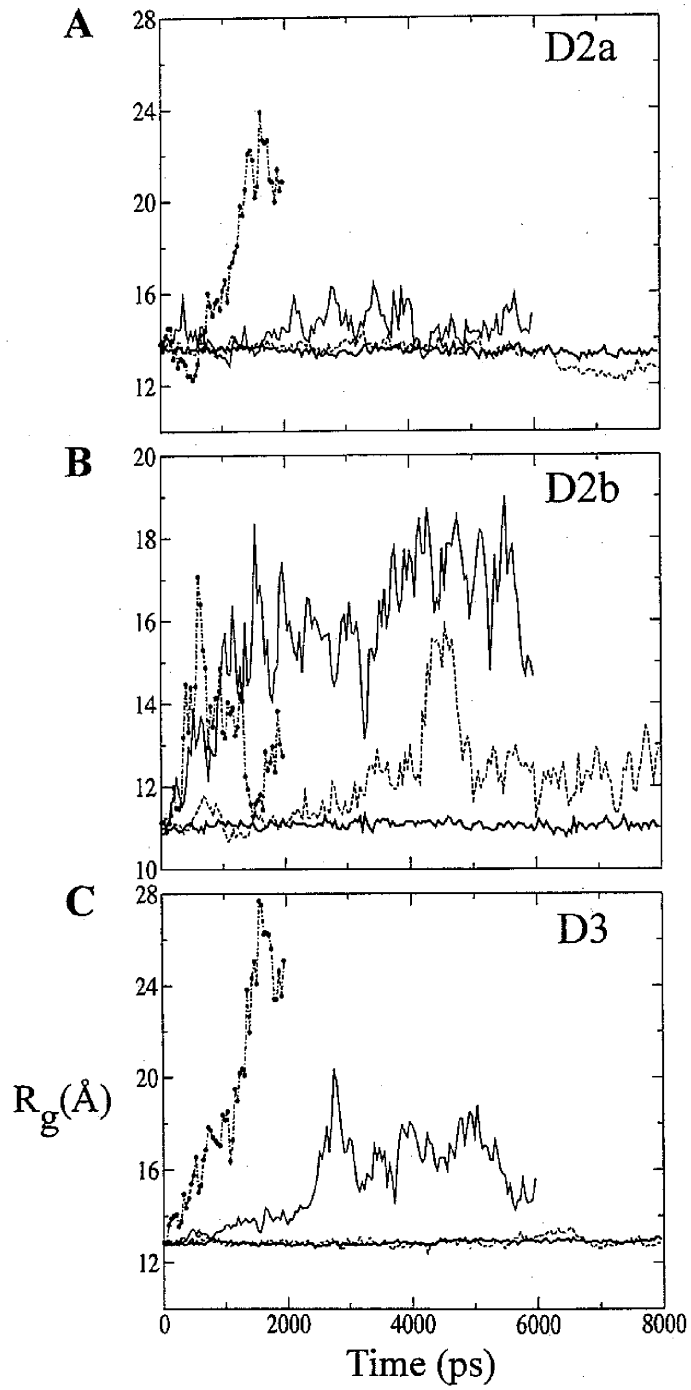


Figure 6.6: **Radius of gyration** of (*A*) subdomain D2a, (*B*) subdomain D2b and (*C*) domain D3 under different temperatures from the first set of simulations: control (bold continuous), 400 K (dashed), 500 K (continuous) and 600 K (dot-dashed line with filled circles). Data points have been re-sampled at an interval of 20-ps.

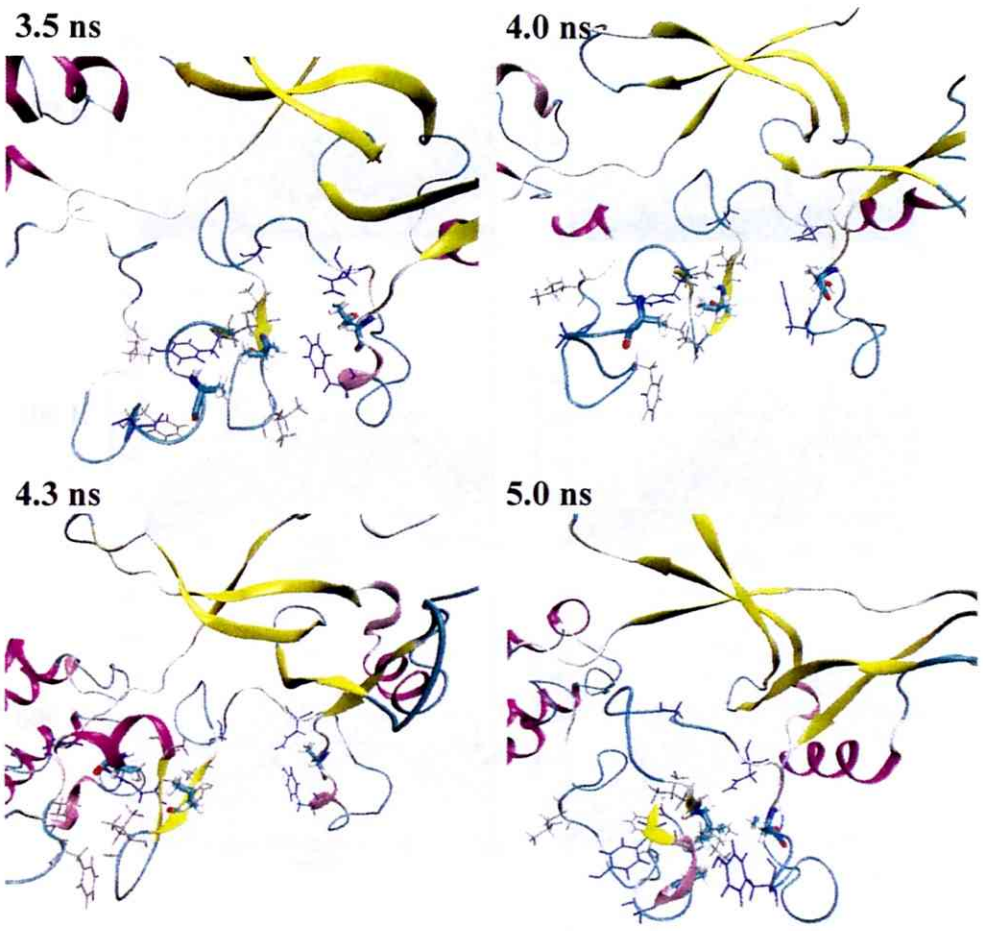


Figure 6.7: **D2b unfolding.** Snapshots taken from the first 400 K simulation showing partial unfolding/refolding of D2b with D2a remaining intact. Weak hydrophobic core residues are shown as sticks colored by element. Residues with low SASA as suggested by ASAView are shown as gray (hydrophobic) and blue (polar) thin sticks. Rendered by Tachyon ray-tracer [Stone, 1998] within VMD.

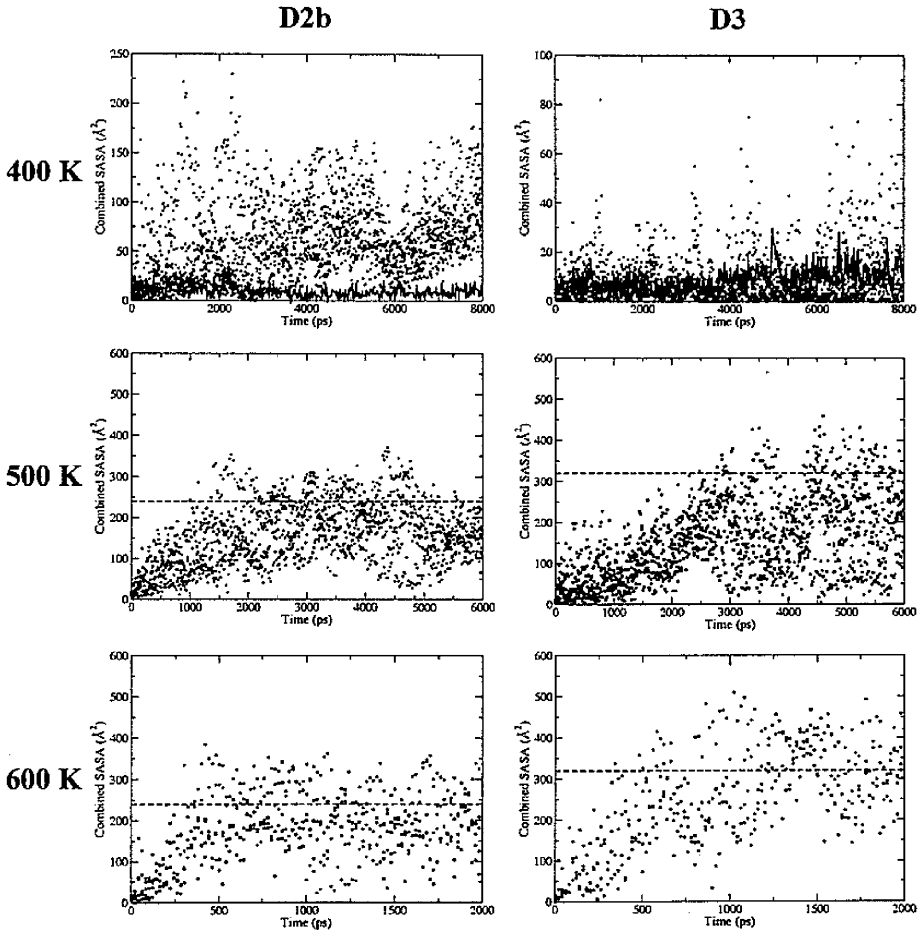


Figure 6.8: **Solvent exposure of D2b and D3 hydrophobic cores.** Total SASA from selected hydrophobic core residues in flagellin D2b and D3 (those with SASA below 80 Å^2 during control simulation). Data from all five simulations are presented as scatterplots (20 ps intervals), with values from control shown as a bold line in the 400 K plots. D2b core residues appeared much perturbed at 400 K compared to D3. For 500 K and 600 K plots, combined thresholds are indicated as dotted lines. Here, the bulk of points passed the threshold later for D3 than D2b under 500 K but around the same time under 600 K. D2b core: T347, V373, A385; D3 core: F202, V233, V247, A231. Note the different scales on the y-axis for the 400 K plots.

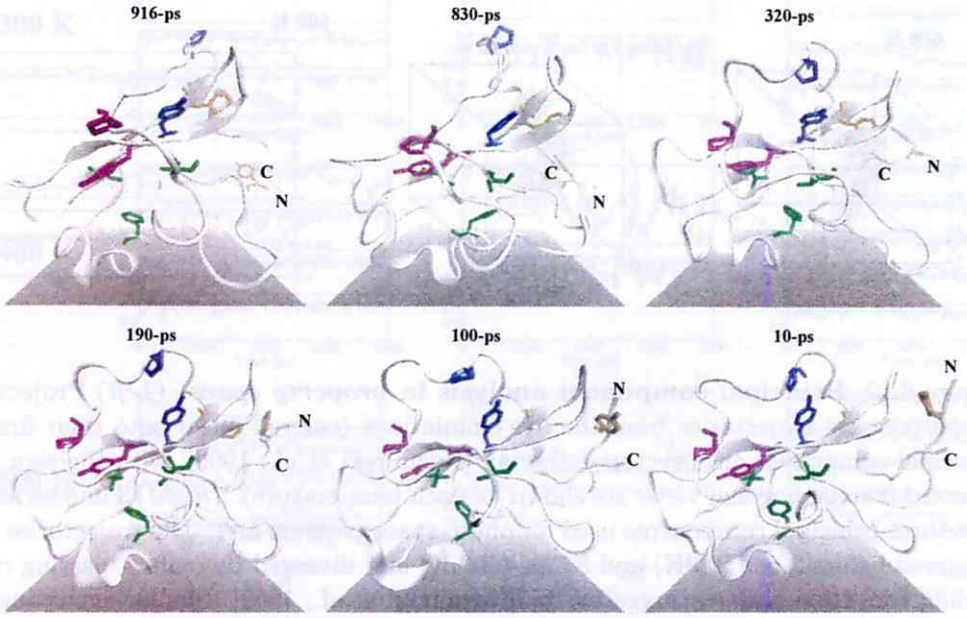


Figure 6.9: **D3 unfolding sequence, reversed.** Representative trajectory snapshots taken from the first 500 K thermal unfolding simulation (after RMS fitting on D3 residues) presented in reversed time order to show folding of domain D3. The N and C termini are labeled. Folding starts from stabilization of the arch-loop conformation by proline-aromatic interactions between pairs P265-Y245 (magenta sticks) and P271-Y230 (blue sticks) with prolines on the loop and tyrosines on the β-folium sheet, denoting event **d** in Fig. 6.5. Aromatic-aromatic interaction between F222 and Y190 (on terminal β-sheet which is also the D2-D3 linker) shown here as orange sticks helps to "bind" the terminal sheet against the rest of the domain (event **c**). The tight packing of F202 in the hydrophobic core (F202, V233 and V247 shown as green sticks) takes the longest time (event **b**). Note the formation of the terminal β-sheet after F222-Y190 interaction becomes stable. Molecular scenes rendered by Tachyon ray tracer [Stone, 1998] within VMD.

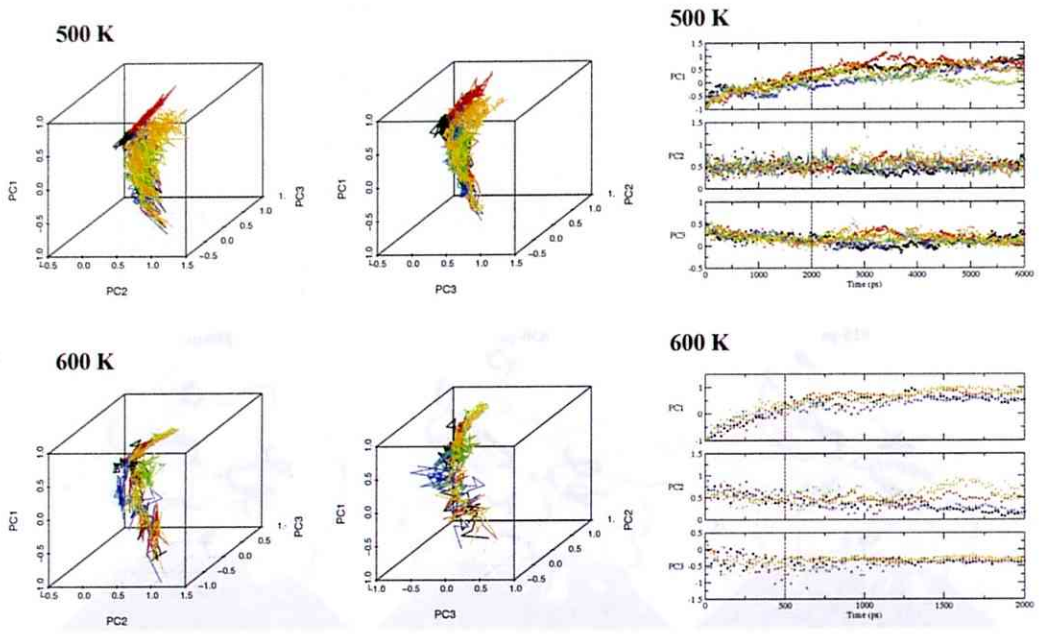


Figure 6.10: **Principal component analysis in *property space*.** (*Left*) Projections of property-space trajectories from the five simulations (colored lines) onto their first three principal component eigenvectors following [Kazmirski et al., 1999], with the sign of PC1 inverted (two orthogonal views are shown for each temperature). (*Right*) Time-series of the first three principal components used for phase-space plots on *Left*. The trajectories appear clustered before 2 ns (500 K) and 0.5 ns (600 K) and diverged thereafter, passing over the folding transition state as suggested by [Kazmirski et al., 1999]. Similar clustering across simulations were also observed in time-series of R_g and SASA under each temperature.

the principal eigenvector components (Table 6.1). Loss of native backbone H-bonds and increase in R_g dominates the second and third principal components under 500 K and 600 K. By projecting the property-space trajectories onto the first three principal eigenvectors, I observed divergence of the trajectories after 2 ns (500 K) and 0.5-ns (600 K) (Fig. 6.10), indicative of a passage through the transition state. Hence, trajectories after these two time-points can be considered to be sampling the denatured ensemble. In some of the trajectories, native secondary structures are observed to be persistent.

6.3.4 Persistent structures in HVR domains D3 and D2a

HVR domains in *S. typhimurium* contain high proportions of β-strands with a unique β-folium fold [Samatey et al., 2001]. D3 and D2a contained similar amounts of β-strands, found to be persistent during thermal denaturation (Fig. 6.3 *B*). The three β-strands in D2b, on the other hand, "dissolved" rapidly (data not shown).

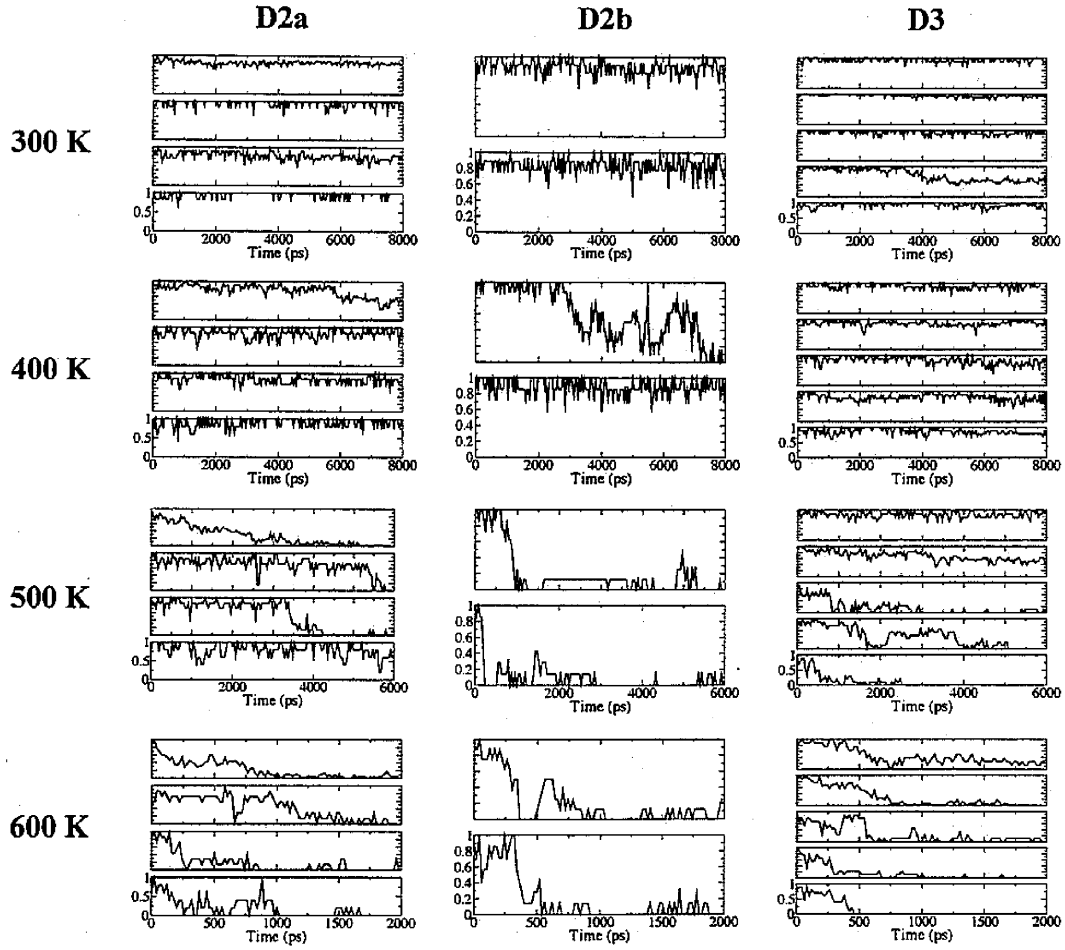


Figure 6.11: **Persistence of β-strand pairs in HVR domains.** Fractional contacts of β-stranded contact clusters, based on the first set of simulation data. There are four sets of multi-plots under each of the columns for each flagellin subdomain, each representing a different temperature. From the top of each multi-plot, the change of fractional contacts with time for contact clusters are listed in the following order: $\beta_{12}\beta_{13}$, $\beta_{13}\beta_{14}$, $\beta_{14}\beta_{15}$ and $\beta_{14}\beta_{16}$ in D2a; $\beta_{16}\beta_{17}$ and $\beta_{17}\beta_{18}$ in D2b; $\beta_6\beta_7$, $\beta_7\beta_8$, $\beta_8\beta_9$, $\beta_5\beta_7$ and $\beta_6\beta_{10}$ in D3. Note that fractional contacts for 300 K were computed with respect to the equilibrated structure while values for 400 K and above were based on persistent native contacts obtained from 300 K trajectory.

	Eigenvalue	Contacts	H-bonds	β-content	R_g	SASA (all)	SASA (hp)
PC1	3.00	−0.48		−0.25	0.45	0.53	0.47
PC2	1.02	0.25		**0.84**	0.39	0.03	0.29
PC3	0.52	−0.55		0.32	−0.63	−0.18	0.42
PC1	4.54	−0.44	−0.33	−0.42	0.40	0.42	0.42
PC2	0.64	0.03	**0.81**	0.18	0.40	0.37	0.10
PC3	0.35	0.37	−0.42	0.31	**0.62**	0.20	−0.41
PC1	4.33	−0.46	−0.37	−0.40	0.36	0.44	0.42
PC2	0.64	0.18	0.37	0.40	**0.74**	0.34	−0.10
PC3	0.47	0.17	**−0.80**	0.17	0.24	0.08	−0.49

Table 6.1: **Property space PCA.** Contributions of each physical property to the first three property-space principal components (as elements in the principal eigenvectors) during denaturation of domain D3 under 400 K (top block of three rows), 500 K (middle block) and 600 K (lower block). Covariance matrix was formed using scaled properties as in [Kazmirski et al., 1999]. The sign of *PC1* components has been inverted for better comparison with reported values for small proteins. "Contacts" means fraction of native contacts. Only selected native backbone H-bonds are monitored here (except for 400 K which showed insignificant changes). Almost equal contributions to the *PC1* eigenvector from the properties indicate that an overall expansion is the dominant unfolding mode for D3. In *PC4* (not shown), β-strand content ("β-content") and SASA of hydrophobic residues ("SASA (hp)") dominates for 500 K and 600 K. I did not separate trajectories into unfolding and denatured phases as in [Kazmirski et al., 1999] but noted that little unfolding occurred under 400 K while denatured states were quickly attained under 600 K. Largest absolute values under *PC2* and *PC3* are highlighted in bold.

Changes in fraction of persistent native contacts for D2a, D2b and D3 under each denaturing temperature from the first set of simulations are presented in Fig. 6.11. For D2a at 500 K, although $\beta_{14}\beta_{16}$ is most persistent in this simulation, the top spot is taken by $\beta_{13}\beta_{14}$ among the other 500 K simulations and also in the first 600 K trajectory. The key stabilizing residue on β_{14} is Y332, which formed aromatic-aliphatic interactions with V325 on β_{13}. The persistent contacts could have restricted the increase in R_g for D2a as compared to D2b at 500 K (Fig. 6.6). In the case of D3, though $\beta_6\beta_7$ is the most persistent at 500 K, $\beta_7\beta_8$ is also persistent at 600 K. Residues F222 and Y229 in $\beta_6\beta_7$ form favorable interaction with their side-chain aromatic rings tilted at a mean angle of 57 degrees (from control simulation). Moreover, the lower Rg at 500 K relative to 600 K could be a result of compact region formation in the denatured state, involving hydrophobic residues 245, 247 and 229 to 231.

To identify which residues are involved in the thermal residual structures, lists of D2 and D3 residues in persistent contacts (>70% occurrence) during each of the 500 K simulations were compared and those residue pairs appearing in at least two of the five lists are selected and shown in Figure 6.12. Of particular mention are the pairs V325-Y332 and D223-K228

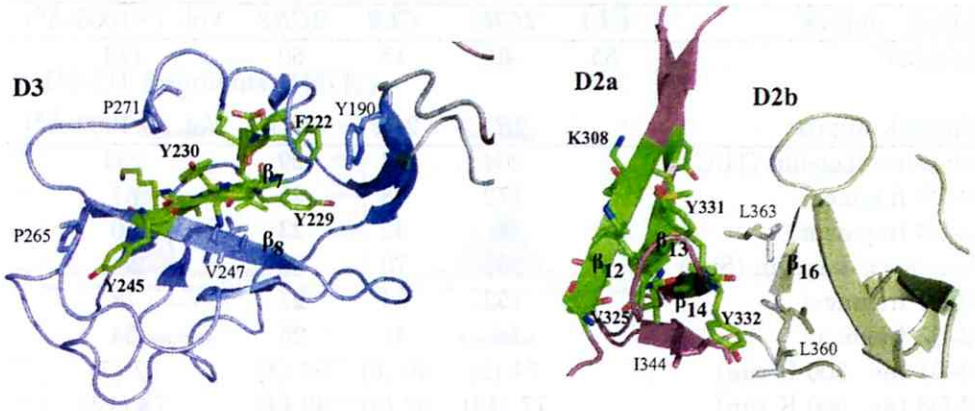


Figure 6.12: **Potential HVR folding cores.** The residues in the HVR domains (D2 and D3) forming persistent pairs of residue contacts in at least two of the five 500 K simulations are shown as sticks and colored by chemical element: carbon (green), oxygen (red) and nitrogen (blue). Some of these residues are labeled in bold. Other labeled residues shown in blue, brown or dark-green sticks are mentioned in the text. The β-sheets $\beta_6\beta_7\beta_8$ and $\beta_{12}\beta_{13}\beta_{14}$ might form folding cores. Molecular structures are rendered with PyMOL [DeLano, 2002].

which appeared in four and five lists, respectively. The latter pair formed a salt-bridge after ~300-ps into two of the simulations, as determined by the VMD [Humphrey et al., 1996] Salt-Bridge plugin. Although there is an abundance of Tyrosine residues in the persistent contacts, Tyr do not appear to be conserved among flagellin homologs' HVR sequences.

6.3.5 Volume of denatured flagellin

By calculating the principal moments of inertia of flagellin molecules and approximating the molecules as ellipsoids, I can get estimates for the dimensions (semi-major radius *R1*, and semi-minor radii *R2* and *R3*) of the equivalent ellipsoids. From the radii, I can get rough estimates for the molecular volume.

The top panel of Table 6.2 shows my estimate of the refolding chamber volume by representing it as a combination of two cylinders. The main panel of the table presents estimated volumes for whole flagellin and its molecular fragments. Twice of the semi-major and semi-minor axes radii are reported for better comparison with chamber dimensions. Both the ellipsoidal approximations of polymeric (1UCU) and monomeric flagellin (starting structure for the denaturing simulations) are too large for the chamber, though their D0-D1 and D2-D3 fragments might fit. The volumes occupied by whole denatured flagellin are also computed. From conformation snapshots taken from the last 1-ns of 500 K and last 0.5-ns of 600 K simulations, I found broad and multi-modal distributions for the volume

Cylindr. approx	*CL1*	2*CR1*	*CL2*	2*CR2*	Vol. (×1000 Å^3)
Chamber*	55	40	15	60	112
Ellipsoid. approx		2*R1*	2*R2*	2*R3*	Vol. (×1000 Å^3)
Polymeric-flagellin (1UCU)		204	74	29	234
D0-D1 fragment		172	28	25	61
D2-D3 fragment		98	42	23	50
Monomeric-flagellin (Sim)		200	70	38	280
D0-D1 fragment		163	34	27	78
D2-D3 fragment		96	41	26	54
D0-D1 (av. 500 K sim)		74 (5)	40 (6)	53 (3)	82 (7)
D2-D3 (av. 500 K sim)		77 (13)	37 (5)	49 (4)	73 (12)

Table 6.2: **Chamber and flagellin volumes.** Estimates of refolding chamber and flagellin volumes. Numbers in brackets denote the standard deviations of the ellipsoidal radii and volume distributions obtained from the combined data set of five 500 K MD trajectories. *The refolding chamber volume is approximated by the sum of the volumes of two cylinders with length(radius) given as *CL1(CR1)* and *CL2(CR2)*.

with values ranging from 140,000 to 250,000 Å^3, much larger than the estimated chamber volume. The ellipsoids representing denatured flagellin are about 120 Å along the "polar" axis and 40 to 54 Å at the "equator", too long for the chamber. Together, the results suggest that maybe half of flagellin might fit inside the chamber.

Physical extents and volumes of denatured flagellin fragments are estimated using 2-ps snapshots taken from the last 1-ns of the five 500 K trajectories. The D0-D1 and D2-D3 fragments are still spatially separable despite being denatured. I present the average and standard deviation (in brackets) of the combined data sets in the lower panel of Table 6.2. The histograms of the fragment volume distributions from all simulations can be found in Fig. 6.13. The results suggest that denatured D0-D1 and D2-D3 fragments could individually fit inside the refolding chamber but not whole flagellin.

6.4 Discussion

6.4.1 Comparison with experimental denaturation study

The folding energetics of flagellin has been investigated by Namba and co-workers by means of analyzing Differential Scanning Calorimetry experiments with a theoretical model formulated for multi-state unfolding transitions [Honda et al., 1999]. They determined from the melting temperatures that the unfolding of (sub)domains follows this order: D_f1 to D2 to D3. From my thermal unfolding simulations of flagellin, I found a similar consensus

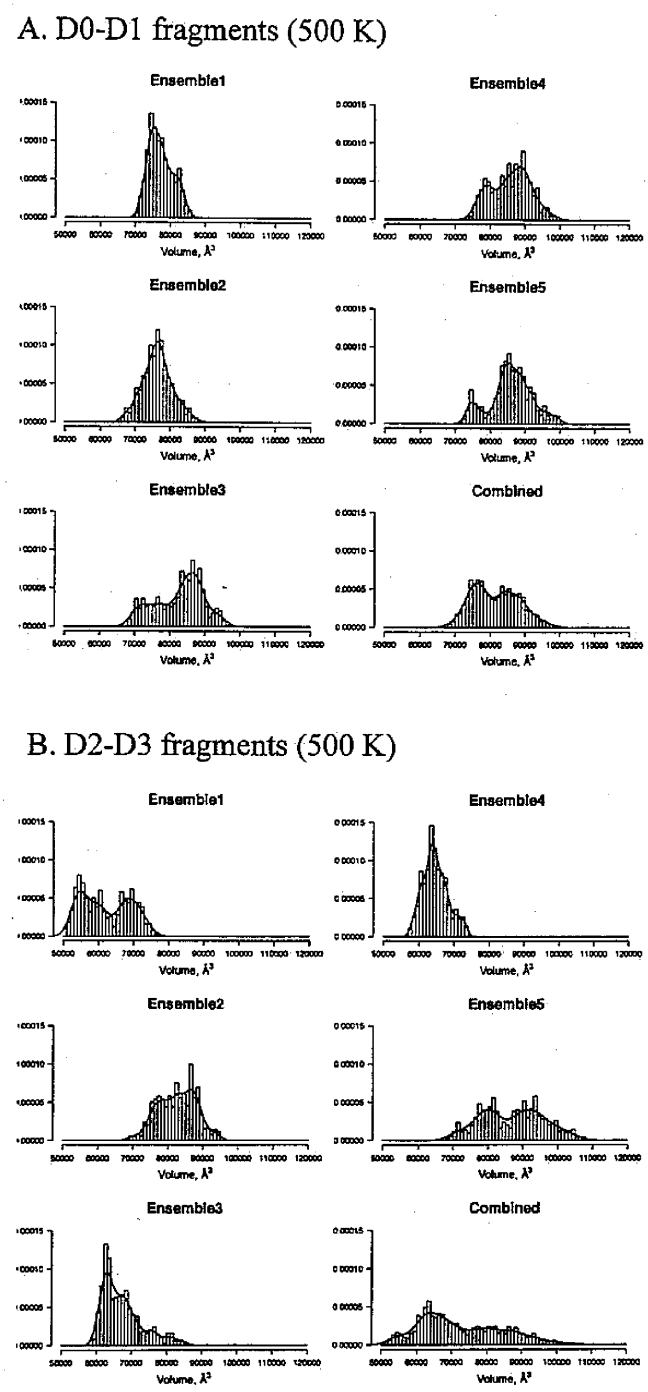


Figure 6.13: **Distributions of flagellin fragment volumes.** Histograms showing the estimated volumes of flagellin molecular fragments based on 2-ps snapshots taken from the last nanosecond of denaturing simulations at 500 K. All volume data are pooled for the histogram under the heading "Combined". All the fragment volumes are below the estimated chamber volume at 111,560 Å^3.

order of unfolding where fragment D_f1 and subdomain D2b unfolded before subdomain D2a and domain D3. The lower thermal stability of subdomain D2b (denaturing under 400 K) observed in my atomistic simulation corroborated that domain D2 should have a lower melting temperature as compared to D3. I have to employ much higher temperatures than used in experiments in order to observe significant changes in the protein over atomistic simulation time-scales of nanoseconds.

The experimental study found an asymmetric DSC curve for D_f1 with steeper slope after the melting temperature as compared to theoretical curve assuming no inter-domain interactions. This steeper drop lead the authors to suggest that unfolding of D2 sped up that of D_f1 [Honda et al., 1999]. They also made the suggestion that D_f1 is marginally stable even under physiological conditions [Honda et al., 1999]. With the flagellin 3D structure (PDB code 1IO1) obtained after the DSC study, we know that (N-terminal) D1-D2 linker is a long backbone segment devoid of any secondary structures that runs under the hydrophobic core of D_f1 and residues L159 (most buried), I162, L167, L169 and L172 on the linker contribute side-chains to the core. This linker is contiguous with β_3 in D2a and has D2b packed beneath it (Fig. 6.1). There is almost no contact region in the interface of D2b-D_f1 and that of D2a-D_f1 was found to be a weak one that is formed by solvent-mediated H-bonds between $\beta_{12}\beta_{13}$ hairpin turn in D2a and the turn across $\alpha_2\alpha_3$ in ND1, as found from the control simulation at 300 K. Hence, stabilization of D_f1 by D2 does seem to be non-enthalpic as suggested [Honda et al., 1999].

From my thermal unfolding simulations at 400 K, I found that in cases where the D_f1 fractional contacts dropped to and stabilized around 0.4 (Fig. 6.4 *E*), the maintenance of the native D2a-ND1 interface is not common to all. This finding supports the suggestion that the stabilization of D_f1 by D2 is not enthalpic. An indication that unfolding of D2a might cause further unfolding of D_f1 is shown in Fig. 6.4 *B*. However, D2b might have a larger influence on D_f1 due to the seeming correlation in the fractional contact profiles, with those of the latter lagging behind the former (Fig. 6.4 *E*). If the correlations are real and assuming that the unfolding pathway is conserved [Day et al., 2002], D2 unfolding might indeed hasten D_f1 unfolding as suggested by experiment [Honda et al., 1999], through the D1-D2 linker. A possible way to assess the role of the linker is to constrain its conformational flexibility in further thermal unfolding simulations.

It was also demonstrated in the denaturation study that D2 and D3 interact both enthalpically and entropically to stabilize each other [Honda et al., 1999]. The unfolding

as monitored by fractional contacts do indicate similar profiles for the two domains among 400 K trajectories (data not shown) but the similarity alone might not be a convincing indicator of unfolding cooperativity.

To determine melting temperatures of individual domains and domain-domain folding cooperativity from simulation, we need to monitor changes in contacts over a wide range of temperatures and also increase the temperature within a simulation to mimic DSC experiments. The all-atom nature of my models limited *temperature-scan* types of equilibrium simulations to only a few selected temperatures due to the high computation cost. *Temperature-ramp* MD simulations might also be very time-consuming due to the need to increase the temperature in small steps and with sufficient equilibrium at each step in order to keep the system at quasi-equilibrium for proper simulation behavior. A simplified or coarse-grained model is thus desirable with the advantage of lower computation cost due to faster dynamics which allows much longer simulations into micro-seconds, the realm of biological processes [Ding and Dokholyan, 2005]. Such an approach was employed by Chen and Dokholyan, who performed *temperature-scan* and *temperature-ramp* simulations to look at domain interactions and unfolding kinetics of vinculin [Chen and Dokholyan, 2006].

6.4.2 HVR domains fold via nucleation?

Persistent secondary structures were detected in the denatured states of flagellin from my simulations at 500–600 K: β_6 to β_8 in D3 and β_{12} to β_{14} in D2a. These β-sheets might account for the relatively high melting temperatures of D2 and D3 found in experiment [Honda et al., 1999]. Residue-residue contacts persistent under unfolding were identified using all five sets of 500 K trajectories and involved the following residues: 217, 220–223, 228–232, 245 in D3 and 305–309, 321–326, 331–332 and 334 in D2a. These residues are located in the persistent β-sheets mentioned above (Fig. 6.12).

The K-Fold server, which uses a machine-learning algorithm trained on 63 proteins with known folding characteristics, predicts both folding rate and kinetics of proteins given the 3D structure [Capriotti and Casadio, 2007]. The residues in the persistent secondary structures were found to be major contributors to the total contact order (data not shown). Bahar and co-workers [Demirel et al., 1998] have suggested that strong dynamic coupling (high contact order) to all other residues implies that such residues are hubs in the intra-protein network. Such hubs have been associated with the folding nucleus in a theoretical

study on the small-world nature of protein structures [Vendruscolo et al., 2002].

I thus propose that the persistent three-stranded anti-parallel β-sheets observed in D2 and D3 might form early in the (re)folding process, serving as spatially-diffused folding nucleation sites or cores to promote folding of the rest of the domains, following the *nucleation-condensation* model proposed by Alan Fersht [Fersht, 1997]. Folding cores are like the foundation pillars of a building, which needs to be present before the rest of the building can come together. They also define the shape of the building. Similar suggestions of persistent three-stranded sheets being folding cores have been made by other researchers via thermal unfolding simulations [Akanuma and Yamagishi, 2005,Sham et al., 2002]. Due to their local nature, such folding cores might even be formed during the flagellin transport phase. Sato and Raleigh reported the detection of significant backbone hydrogen bonds in the transition state of the ribosomal protein L9 N-terminal domain by means of amide H/D isotope exchange experiment [Sato and Raleigh, 2007]. Their work showed the existence of secondary structures in the transition/denatured state of proteins.

It has been deduced that new flagellin is added to the growing filament every second [Ikeda et al., 1993], implying that flagellin has one second to refold if the chamber is always occupied. According to a linear regression relationship between logarithm of experimental folding rate and Absolute Contact Order (ACO, a measure of topological complexity defined for native structure) established for two- and multi-state folders [Ivankov et al., 2003], the ACO of a domain should be less than about 20 for a folding time of less than one second. Based on the experimental structure of flagellin (PDB code 1UCU), helical domains D0 and D1 have ACO around 3 while D2 has a value close to 10. The globular domain D3 has the largest value of about 15, meaning that it could be the rate-limiting step in the refolding process. The numbers also suggests that D2-D3 might either fold cooperatively or fold independently but simultaneously and then "dock" together to reduce folding time. The presence of folding cores in HVR domains deduced from my simulations would enable fast folding of D2-D3 followed by "docking" of folded D1-D0 domains. Such a mechanism could help flagellin to complete refolding within a second, overcoming the limitation of high topological complexity.

A study comparing folding mechanisms of Immunoglobulin(Ig)-like β-sandwich proteins from diverse families found that interactions defining the structural topology also guided folding [Clarke et al., 1999]. Clarke and co-workers have also found that the folding nucleus can be significantly deformed to preserve the same folding mechanism despite sequence

variation in Ig-like domains [Geierhaas et al., 2004]. I suspect that HVR domains are under evolutionary pressure to adopt structurally stable folds despite sequence divergence. Alas, to-date we only have the structures from *S. typhimurium* in the PDB[1]. It is interesting to see if HVR domains from diverse flagellin species adopt similar folds. If so, HVR domains from other flagellin might also harbor folding cores.

6.4.3 Importance of folding cores to flagellin stability

In a mutagenesis study [Malapaka et al., 2007] screening for functional (retention of mobility) deletion flagellin mutants, I found that the three deletion mutants showing the same mechanical stability as wild-type retained residues I have identified to be in the folding cores (except for Y245 that is absent in one of the mutants). Also, the folding and mechanical stability of mutants were not affected if residues 250 to around 290 are missing: starting from the β-turn of $\beta_8\beta_9$, through the "arching-loop" to β_{10} and reaching the α-helix in D2a. These residues are not part of the D2 or D3 folding cores.

For mutants with more brittle filaments, they either have a reduced D2a or in the case of "C11" in the paper, missing residues A231 and K232 from the potential folding cores and V233 from the native hydrophobic core. This mutant flagellin highlighted the greater importance of retaining folding core residues (especially the salt-bridge D223-K228) over residues in the hydrophobic core. This observation also indicated that the salt-bridge is not an artifact of simulating at elevated temperatures, since salt-bridges become more important at higher temperatures and confer thermo-stability to thermophilic proteins [Missimer et al., 2007]. Interestingly, in the native state under physiological conditions K228 forms a salt-bridge with E246 on β_8 instead of D223 on β_6, which suggests a switch in partner after the folding core has been formed and more extensive interactions are desired to form the rest of D3 (note: potential folding core ends at Y245). Lastly, functional mutants which could not swim but showed swarming after 12–16 hours of growth were missing all of the folding core residues or all of D2 [Malapaka et al., 2007], affecting the formation of the native HVR domains. This mutagenesis study thus highlighted the existence and importance of the folding cores to flagellin mechanical stability.

With the identification of potential folding cores, amino-acid substitution experiments

[1]The terminal-truncated structure of a flagellin homolog with a very small HVR domain (about 100 residues) was published in Feb 2008 [Maruyama et al., 2008], after our paper became accepted in Jan 2008. The HVR domain contains β-sheets but with different folds.

can now be directed at replacing D223 (breaking only the folding core salt-bridge) or V233 to investigate their effects on D3 folding and stability. In the case of D2a, Valine pair 306–325 and the Tyrosine pair 331–332 in the folding core might be good candidates for stability studies. Destabilization of the D2a folding-core might be a way to test if folding of D_f1 is indeed dependent on that of D2 as suggested in [Honda et al., 1999]. The effect would be manifested as a slow down or abolishment of filament growth since natively folded D1 domains are needed to build the outer tube of the filament.

6.4.4 Refolding of flagellin may happen in stages

Flagellin is likely to be substantially unfolded in order to be exported by the apparatus located at the entrance to the filament channel (Fig. 1.6). I can also assume that unfolded flagellin molecules travel in single-file along the filament channel, as used in a mathematical model of filament growth [Keener, 2006]. At the end of the channel under the filament cap (formed by HAP2 chaperone) is a cylindrical space surrounded by D1 domains from the last round of polymerized flagellin [Yonekura et al., 2000], in which refolding of flagellin is believed to occur. The depth of the "chamber" at about 70 Å is close to the span of the longest D1 helices. Because D1 of flagellin is conserved across bacterial species [Beatson et al., 2006], the chamber size might also be conserved and independent of HVR domain sizes. Whether flagellin enters the chamber as a linear chain or as a loop (half-folded chain) remains unknown. Investigation into the mechanical forces needed to generate these two *transport* states/forms have been performed and reported in Chapter 5.

Based on estimates of denatured flagellin volumes from my simulations and comparing with an estimated chamber volume, I suggest that only fragments D2-D3 or D0-D1 can be accommodated but not the whole molecule. Yonekura has commented that the chamber is large enough for at most one molecule [Yonekura et al., 2000]. The refolding process may therefore be a two-stage one: chamber occupancy and folding of HVR (D2-D3) domains precedes the filament-core (D0-D1) domains (the more likely scenario). Furthermore, HVR domains have been suggested to fold cooperatively from DSC experiment [Honda et al., 1999] and might fold rapidly with the help of the folding cores I have identified, before being extruded to the filament exterior through the HAP2 gap [Yonekura et al., 2000,Maki-Yonekura et al., 2003]. As D2b was found to fold late, it could still be rather "fluid" and makes flagellin extrusion easier. After filament-core domains become folded, "docking" of D_f1 to folded D2 would complete the refolding process for flagellin. In bacterial species

with short or negligible HVR segment, refolding of filament-core domains would be the only step.

6.5 Conclusion: flagellin refolds from HVR domain nucleation sites?

Persistent native structures (three-stranded Z-like β-sheets) detected in the thermally denatured ensemble of multi-domain *S. typhimurium* flagellin from simulation suggest that they might constitute folding initiation sites or folding cores. The Z-like β-sheets is one of the known folding nuclei motifs [Garbuzynskiy and Kondratova, 2008]. Interestingly, the fastest folding protein is also made up of a three-stranded anti-parallel β-sheet, with a folding time of about 140-ns at 308 K [Xu et al., 2006]. Proper folding of HVR domains would lead to proper folding of filament-core domain D1 which is recognized by the assembly chaperone HAP2 [Maki-Yonekura et al., 2003].

If each HVR domain fold via the nucleation-condensation mechanism for small globular proteins, formation of the folding nucleus would mean that the correct folding transition state has been reached and refolding then proceeds "downhill" in free energy from that point: a signature of efficient two-state (Denatured to Native) folding (Section 3.3.1). Explicit identification of the transition state ensemble was not carried out in this study due to the need for a large number of simulation trajectories. In fact, detection of two-state folding kinetics for domain D3 in experiments would lend support to the nucleation-condensation mechanism. This has indeed been suggested in a folding study of isolated domain D3, showing that it could fold stably (presence of a well-formed hydrophobic core) and reversibly via an all-or-none transition [Sebestyén et al., 2008].

Chapter 7

Conclusions and outlook

7.1 Insights into flagellin unfolding/refolding for transport

In this thesis, I have used molecular dynamics simulations to investigate the unfolding and refolding processes that enable bacterial flagellar filament protein (flagellin) to be transported from within the bacterium to the assembly site at the filament tip. The main findings are summarized below in response to the aims of the thesis set out in the Introduction:

1. What is the transport form of flagellin?

    - Two forms (*wire* and *hairpin*) has been proposed and the mechanical cost to obtain them has been computed and compared via mechanical unfolding simulations that pulled a flagellin monomer along two different directions. Results suggest that the *wire* form would require lower peak forces but would require higher mechanical work (consume more ATP molecules, say). This is suitable for a steady, low-powered unfolding machinery as are most ATP-powered unfoldases.

2. How flagellin spontaneously refold in the cavity under the filament cap?

    - Multiple thermal unfolding simulations at different denaturing temperatures were conducted to probe the folding process in reverse. Domain-level unfolding sequence reflects thermal resistance indicated by experiment. Lower thermal stability of a subdomain D2b can help to explain why D2 unfolded earlier than D3 despite similar stability between D2a and D3.

- Residual native-like interactions found in the denatured sequence-variable HVR domains D2 and D3 indicate possible folding initiation sites (Z-like β-sheet). These domains might fold via the nucleation-condensation mechanism (see section 3.3.1) which could reduce the chance of misfolding due to concurrent formation of secondary and tertiary interactions.
- Volume estimates suggest that refolding in the filament chamber has to occur in two-stages, since there is sufficient space only for either denatured HVR domains (D2-D3) or denatured filament-core domains (D1-D0). Radial arrangement of flagellin in the filament requires HVR domains to be extruded to the filament surface first, hence they should refold before the filament-core domains.

3. Can we generalize the findings to other flagellin homologs and other flagellar export proteins?

   - Secondary structure predictions on the putative HVR domains of flagellin homologs indicate high β-sheet contents. Large HVR domains in flagellin homologs might contain folding nuclei though the motif may not be Z-like β-sheet but dictated by the structural topology.
   - Several other export proteins (HAP2, hook) also have adjacent termini. Unravelling them from the N-terminal might also require less mechanical effort rather than pulling both their termini into the channel.

7.2 Tentative model of flagellin export

Based on current understanding in the literature and from the findings presented in this thesis, we have the following tentative model of flagellin export:

At the export gate in the flagellar basal body under the inner membrane, flagellin is recognized by the export and sorting signals at its unstructured N-terminal [Stafford et al., 2007]. A mechanical force generated from an as-yet unclear mechanism would pull flagellin into the channel from its N-terminal, unraveling β-sheets by "lateral shearing" and breaking salt-bridge interactions in domains D1 and D3 in the creation of a *wire* flagellin. Upon entering the filament channel, the negatively charged flagellin (charge of -12 at pH of 7 to 9, according to titration curve from the H++ webserver [Gordon et al., 2005]) would travel through the channel via diffusion. The electric potential set up by the Proton

Motive Force might further assist in the transport (or might even drive the entire export process as suggested by [Minamino and Namba, 2008, Paul et al., 2008]) and would also prevent back-slippage of the translocating molecule [Schiebel et al., 1991]. Upon arrival at the distal "refolding chamber", the formation and growth of the folding nuclei in D2 and D3 (in the form of three-stranded β-sheets) coupled with formation of native salt-bridges guides the refolding process. Folded D2–D3 exits the "chamber" through the gap formed by a symmetry mismatch between HAP2 filament cap and the tip of the filament [Yonekura et al., 2000]. Folding of D1–D0 next commences. Folded D1 is then recognized by the tail domains of HAP2 filament cap protein which assists in the polymerization of the newly arrived flagellin into the filament [Maki-Yonekura et al., 2003].

7.3 Outlook

Molecular simulations and experiments should be viewed as complementary tools to gain insights into biological processes. I hope that experimental work would be carried out in the near future to determine if flagellin indeed (re)fold via the nucleation-condensation mechanism.

The mechanical unfolding simulations performed in this study are to mimic two-way pulling in AFM experiments. However, the pulling mode for the unfoldase might be one-way. The hexameric ATPase FliI (sharing structural similarity with rotary motor F_1-ATPase) could be the unfoldase in the flagellar export system, whose monomer has been solved [Imada et al., 2007]. It would be interesting to compare the threading of the flagellin terminal/termini into the ATPase pore with two-way AFM-like unfolding reported in this thesis. However, a realistic model of flagellar protein unfolding for transport can be constructed only after a complete understanding of the workings of the type III export apparatus (common to the flagellum and needle complex) has been gleaned from experiments.

Lastly, it is tempting to propose that the particular topology of β-sheets in the HVR domains of *S. typhimurium* flagellin have been evolutionally tuned for easy unraveling from the terminal region: strands in β-sheets get detached one-by-one via low effort 'lateral shear'. Optimally oriented β-sheets that could also be folding nuclei might be a general 'design principle' to allow flagellar export proteins to meet the dual requirements of mechanical stability and ease of mechanical unfolding for transport. It would be interesting to test if the above 'principle' applies to the unfolding/refolding of a flagellin homolog whose

3D structure has recently been solved [Maruyama et al., 2008]. Because Z-like β-sheets are absent in the single HVR domain, any folding nucleus if present might take on a different form. Other flagellar export proteins with high resolution 3D structures (e.g. the hook subunit) are also interesting subjects in future studies to gain further insights into the fascinating flagellar protein transport system.

Appendix A

Use of implicit solvent model

A.1 The Generalized Born implicit solvent model

Here I wish to give a brief introduction to the Generalized Born implicit solvent model which was used for the mechanical unfolding simulations. Part of the contents here are based on the AMBER8 manual. The advantage of GB method is a less costly way to get the electrostatic part of the solvation free-energy than to solve the nonlinear Poission-Boltzmann equation.

The idea stems from the Born model of ion solvation. The solvation free-energy (Born equation) can be expressed as:

$$\triangle G_{Born} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon_{solvent}}\right)\frac{q^2}{R}$$

This is the work needed to transfer an ion of charge q and radius R from vacuum (dielectric constant of 1) to a solvent ($\epsilon_{solvent} = 80$ for water). The derivation can be found on page 321 of the text Molecular Biophysics by Michel Daune published by Oxford University Press (2004).

This model is extended to the case of a charged atom inside an arbitrarily-shapred cavity, such as in a protein. Assuming that we can compute the associated solvation free energy and equate the expression to the Born equation above, we can get the "effective Born radius" α. If the atom is exposed to solvent, α will be close to but larger than the van der Waals radius [1]. For buried atoms, α can be quite large. We next consider the case of many

[1] The vdW radius of an atom can be determined from measuring atomic spacings between pairs of unbonded atoms in crystals (from Wikipedia). Essentially, if we represent atoms as hard spheres, half the interatomic separation between identical atoms can be taken as the vdW radius of the atom.

charges inside an arbitrary cavity, where the electrostatic component of the total solvation free-energy is expressed as a sum of the individual Born energies and contributions from pair-wise interactions (Generalized-Born approximation [Still et al., 1990]):

$$\begin{aligned}
\Delta G_{elec} &\approx \Delta G_{GB} = \sum_i \Delta G_i^{self} + 2\sum_i \sum_{j>i} \Delta G_{ij}^{pair} \\
\Delta G_{ij}^{pair} &= -\frac{1}{2}\left(1 - \frac{1}{\epsilon_{solvent}}\right)\frac{q_i q_j}{f_{GB}} \\
f_{GB} &= \sqrt{r_{ij}^2 + \alpha_i \alpha_j exp\left(-r_{ij}^2/4\alpha_i\alpha_j\right)}
\end{aligned}$$

where r_{ij} is the distance between atoms i and j.

A simple way to compute the effective Born radius for each atom is to invert the Born equation:

$$\alpha_i^{-1} = 2\Delta G_{elec} q_i^{-2}\left(\frac{1}{\epsilon_{solvent}} - 1\right)^{-1}$$

which involves getting a good estimate of ΔG_{elec} for each atom (self-energy). Classicial electrostatics relates the self-energy to the volume integral of the electric displacement vector $\mathbf{D}$, which is $\epsilon\mathbf{E}$ for a linear dielectric medium where $\mathbf{E}$ is the electrostatic field. By using the so-called Coulomb Field Approximation, we can write $\mathbf{E}$ using the coulomb field and get $\mathbf{D} = (q_i/r^3)\mathbf{r}$. With this approximation, we can derive an approximation for the effective Born radius as:

$$\frac{1}{\alpha_i} \approx \frac{1}{\rho_i} - \frac{1}{4\pi}\int_{r>\rho_i} \frac{1}{r^4} dr^3$$

where ρ_i is the vdW radius of the charge and the volume integral is taken over spatial locations beyond the vdW radius but still within the solute (or protein in our case).

In the first GB model in AMBER, GB^{HCT} by Hawkins, Cramer and Truhlar [Hawkins et al., 1995], a set of vdW spheres are used to approximate the true molecular surface of the solute for evaluating the 3D integral. This was found to underestimate the value for α in the case of large molecules such as proteins because vacuum-filled cavities between the vdW spheres of protein atoms are taken as solvent-filled (hence outside the solute). A newer model GB^{OBC} by Onufriev, Bashford and Case [Onufriev et al., 2004] was proposed to address this shortcoming.

A.2 Checks on salt-bridge over-stabilization

Salt-bridges in implicit solvent MD simulations may be over-stabilized relative to its counterparts in explicit solvent. Simmerling and co-workers found that among the GB implicit solvent models in the AMBER8 package, GB^{HCT} over-estimated TIP3P salt-bridge stability by 4 kcal/mol but the GB^{OBC} model (used in our study) only over-estimated by 0.5 kcal/mol [Geney et al., 2006]. Hence salt-bridge over-stabilization might only be marginal in our simulations.

In the 8-ns TIP3P simulation that served as the control for our thermal-unfolding study, salt-bridge Glu246-Lys228 in domain D3 remained stable as found by the VMD salt-bridge plugin used with an additional cutoff of 4.0 Å between side-chain centroids (see Fig. A.1). Though Asp217-Lys232 in D3 was not as stable, it may be part of the D3 folding nucleus as suggested by our thermal unfolding study (Chapter 6). To critically assess the contribution of salt-bridges to mechanical resistance, we may create in silico mutants where the key salt-bridges are replaced with residue pairs which can form side-chain H-bonds but are neutral in charge, such as homo- or hetero- Asn/Gln pairs.

Alternatively, we could immerse the load-bearing elements into explicit solvent and investigate their stability in the absence and presence of constant forces. This was in fact performed for the non-native Arg431-Glu153 salt-bridge. A fragment (residues 137 to 162 in ND1 and 416 to 436 from CD1) from the *UB1* trajectory snapshot with the non-native salt-bridge was solvated with TIP3P waters in a 51 Å × 44 Å × 147 Å simulation box. Simulation at 300 K and 1 atm without external force showed that the salt-bridge is stable in the presence of solvent molecules. Pulling CD1 helix in the Z-direction while restraining ND1b hairpin at two different speeds resulted in restraint forces of around 400-pN before the salt-bridge H-bond was replaced by H-bonds to solvent molecules. Forces drop to low values when the salt-bridges are disrupted. See Fig. A.2.

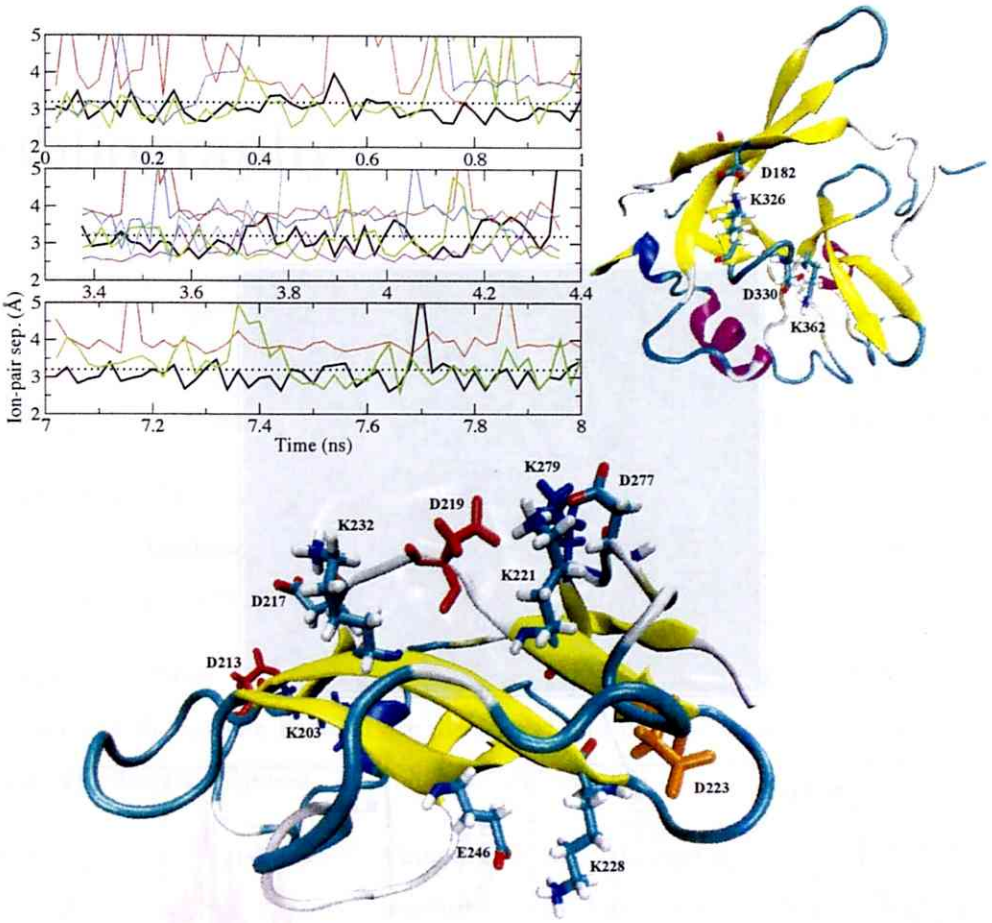


Figure A.1: **Salt-bridges in flagellin HVR domains D2 and D3.** Salt-bridges were identified with VMD from an 8-ns explicit solvent MD simulation at 300 K that served as the control in our previous thermal unfolding study (17). The center-of-mass separations between oxygen and nitrogen atoms on side-chains of oppositely charged residues in D2a and D3 are shown with the default cutoff of 3.2 Å indicated by a dotted line. Three time-segments are investigated: 0 to 1-ns, 3.3 to 4.3-ns and 7 to 8-ns. Salt-bridges found are: D182-K326 (black), D217-K232 (red), E246-K228 (green), D277-K221 (blue), D219-K279 (violet) and D213-K203 (turquoise). Only D182-K326 and E246-K228 satisfy the criteria for salt-bridge during the last 1-ns. Separations between D330-K362 (at D2a-D2b interface) undergo large fluctuations and are not shown. Pairs D219-K279 and D213-K203 which only appears during the middle time-segment are shown as sticks: red for acidic and blue for basic residues. D223 which interacts with K228 in the thermally denatured state is shown as orange stick. Cartoon representations of the structure at 0-ns are rendered by VMD.

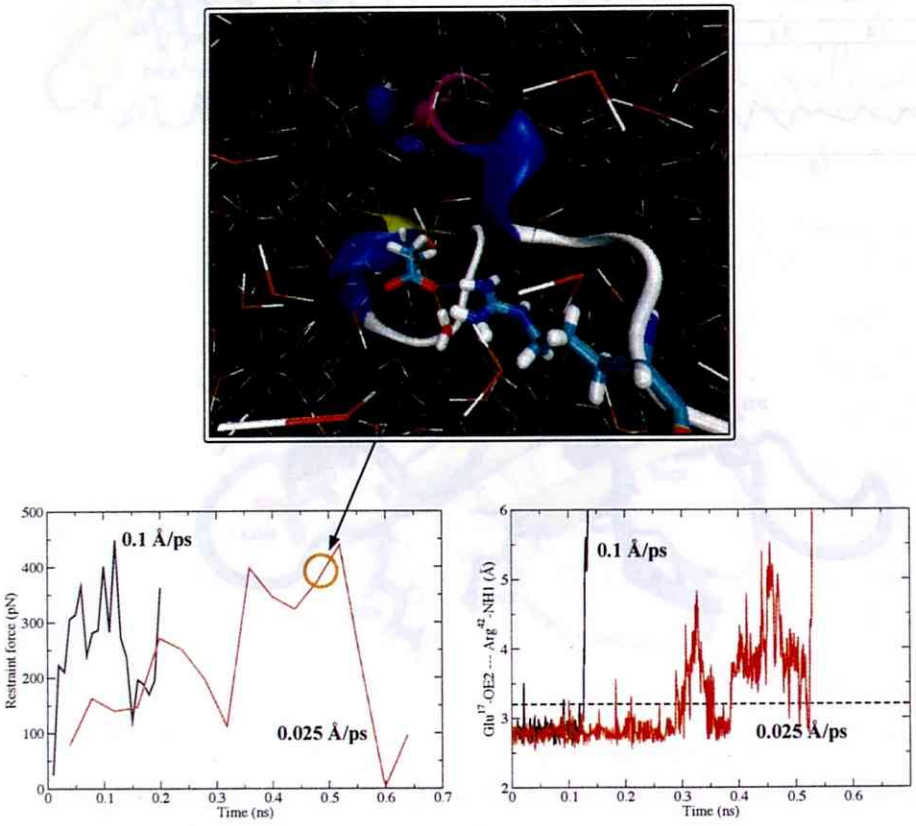


Figure A.2: **Immersing D_f1 non-native salt-bridge in solvent.** A load-bearing salt-bridge found from implicit solvent simulation is also stable in explicit solvent and produced significant restraint forces against disruption, monitored using the separation between side-chain oxygen on the Glu and side-chain NH on the Arg.

Bibliography

[Ackbarow et al., 2007] Ackbarow, T., Chen, X., Keten, S., and Buehler, M. J. (2007). Hierarchies, multiple energy barriers, and robustness govern the fracture mechanics of α-helical and β-sheet protein domains. *Proc. Natl. Acad. Sci. USA*, 104(42):16410–16415.

[Ahmad et al., 2004] Ahmad, S., Gromiha, M., Fawareh, H., and Sarai, A. (2004). ASAView: database and tool for solvent accessibility representation in proteins. *BMC Bioinformatics*, 5:51.

[Aizawa et al., 1990] Aizawa, S.-I., Vonderviszt, F., Ishima, R., and Akasaka, K. (1990). Termini of Salmonella flagellin are disordered and become organized upon polymerization into flagellar filament. *J. Mol. Biol.*, 211:673–677.

[Akanuma and Yamagishi, 2005] Akanuma, S. and Yamagishi, A. (2005). Identification and characterization of key substructures involved in the early folding events of a $(\beta/\alpha)_8$-barrel protein as studied by experimental and computational methods. *J. Mol. Biol.*, 353(5):1161–1170.

[Akeda and Galán, 2005] Akeda, Y. and Galán, J. E. (2005). Chaperone release and unfolding of substrates in type III secretion. *Nature*, 437(7060):911–915.

[Alder and Wainwright, 1957] Alder, B. J. and Wainwright, T. (1957). Phase transition for a hard sphere system. *J. Chem. Phys.*, 27:1208–9.

[Bayas et al., 2007] Bayas, M. V., Kearney, A., Avramovic, A., van der Merwe, P. A., and Leckband, D. E. (2007). Impact of salt bridges on the equilibrium binding and adhesion of human CD2 and CD58. *J. Biol. Chem.*, 282(8):5589–5596.

[Bayas et al., 2003] Bayas, M. V., Schulten, K., and Leckband, D. (2003). Forced detachment of the CD2-CD58 complex. *Biophys. J.*, 84(4):2223–2233.

[Beatson et al., 2006] Beatson, S. A., Minamino, T., and Pallen, M. J. (2006). Variation in bacterial flagellins: from sequence to structure. *Trends Microbiol.*, 14(4):151–155.

[Beck and Daggett, 2004] Beck, D. A. and Daggett, V. (2004). Methods for molecular dynamics simulations of protein folding/unfolding in solution. *Methods*, 34(1):112–120.

[Berendsen et al., 1984] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Dinola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690.

[Bond and Sansom, 2006] Bond, P. J. and Sansom, M. S. (2006). Insertion and assembly of membrane proteins via simulation. *J. Am. Chem. Soc.*, 128(8):2697–2704.

[Brockwell et al., 2003] Brockwell, D. J., Paci, E., Zinober, R. C., Beddard, G. S., Olmsted, P. D., Smith, D. A., Perham, R. N., and Radford, S. E. (2003). Pulling geometry defines the mechanical resistance of a β-sheet protein. *Nat. Struct. Biol.*, 10(9):731–737.

[Capriotti and Casadio, 2007] Capriotti, E. and Casadio, R. (2007). K-Fold: a tool for the prediction of the protein folding kinetic order and rate. *Bioinformatics*, 23(3):385–386.

[Carrion-Vazquez et al., 2003] Carrion-Vazquez, M., Li, H., Lu, H., Marszalek, P. E., Oberhauser, A. F., and Fernandez, J. M. (2003). The mechanical stability of ubiquitin is linkage dependent. *Nat. Struct. Biol.*, 10(9):738–743.

[Case et al., 2004] Case, D., T.A., D., Cheatham, T. I., Simmerling, C., Wang, J., Duke, R., Luo, R., Merz, K., Wang, B., Pearlman, D., Crowley, M., Brozell, S., Tsui, V., Gohlke, H., Mongan, J., Hornak, V., Cui, G., Beroza, P., Schafmeister, C., Caldwell, J., Ross, W., and Kollman, P. (2004). *AMBER 8*. University of California, San Francisco.

[Chen et al., 2008] Chen, Y., Ding, F., Nie, H., Serohijos, A. W., Sharma, S., Wilcox, K. C., Yin, S., and Dokholyan, N. V. (2008). Protein folding: then and now. *Arch. Biochem. Biophys.*, 469(1):4–19.

[Chen and Dokholyan, 2006] Chen, Y. and Dokholyan, N. V. (2006). Insights into allosteric control of vinculin function from its large scale conformational dynamics. *J. Biol. Chem.*, 281(39):29148–29154.

[Chng and Kitao, 2008] Chng, C.-P. and Kitao, A. (2008). Thermal unfolding simulations of bacterial flagellin: insight into its refolding before assembly. *Biophys. J.*, 94(10):3858–3871.

[Chng and Yang, 2008] Chng, C.-P. and Yang, L.-W. (2008). Coarse-grained models reveal functional dynamics – II. Molecular dynamics simulation at the coarse-grained level – theories and biological applications. *Bioinformatics and Biology Insights*, 2:171–185.

[Clarke et al., 1999] Clarke, J., Cota, E., Fowler, S. B., and Hamill, S. J. (1999). Folding studies of immunoglobulin-like β-sandwich proteins suggest that they share a common folding pathway. *Structure*, 7(9):1145–1153.

[Cole et al., 2008] Cole, C., Barber, J. D., and Barton, G. J. (2008). The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*, 36:W197–W201.

[Daggett, 2006] Daggett, V. (2006). Protein folding – simulation. *Chem. Rev.*, 106(5):1898–1916.

[Day et al., 2002] Day, R., Bennion, B. J., Ham, S., and Daggett, V. (2002). Increasing temperature accelerates protein unfolding without changing the pathway of unfolding. *J. Mol. Biol.*, 322(1):189–203.

[Day and Daggett, 2005] Day, R. and Daggett, V. (2005). Ensemble versus single-molecule protein unfolding. *Proc. Natl. Acad. Sci. USA*, 102(38):13445–13450.

[Day and Daggett, 2007] Day, R. and Daggett, V. (2007). Direct observation of microscopic reversibility in single-molecule protein folding. *J. Mol. Biol.*, 366(2):677–686.

[DeLano, 2002] DeLano, W. L. (2002). *The PyMOL molecular graphics system.* DeLano Scientific, San Carlos, CA, USA.

[Demirel et al., 1998] Demirel, M., Atilgan, A., Jernigan, R., Erman, B., and Bahar, I. (1998). Identification of kinetically hot residues in proteins. *Protein Sci.*, 7:2522–2532.

[Desvaux et al., 2006] Desvaux, M., Hebraud, M., Henderson, I. R., and Pallen, M. J. (2006). Type III secretion: what's in a name? *Trends Microbiol.*, 14(4):157–160.

[Dill et al., 2007] Dill, K. A., Ozkan, B. S., Weikl, T. R., Chodera, J. D., and Voelz, V. A. (2007). The protein folding problem: when will it be solved? *Curr. Opin. Struct. Biol.*, 17(3):342–346.

[Ding and Dokholyan, 2005] Ding, F. and Dokholyan, N. V. (2005). Simple but predictive protein models. *Trends Biotechnol.*, 23(9):450–455.

[Dodson et al., 2008] Dodson, G. G., Lane, D. P., and Verma, C. S. (2008). Molecular simulations of protein dynamics: new windows on mechanisms in biology. *EMBO Reports*, 9(2):144–150.

[Dougan et al., 2008] Dougan, L., Feng, G., Lu, H., and Fernandez, J. M. (2008). Solvent molecules bridge the mechanical unfolding transition state of a protein. *Proc. Natl. Acad. Sci. USA*, 105(9):3185–3190.

[Duan and Kollman, 1998] Duan, Y. and Kollman, P. A. (1998). Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science*, 282(5389):740–744.

[Essmann et al., 1995] Essmann, U., Perera, L., Berkowitz, M., Darden, T., Lee, H., and Pedersen, L. (1995). A smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577–8593.

[Eyal et al., 2006] Eyal, E., Yang, L.-W., and Bahar, I. (2006). Anisotropic network model: systematic evaluation and a new web interface. *Bioinformatics*, 22:2619–2627.

[Feig et al., 2004] Feig, M., Karanicolas, J., and Brooks, C. (2004). MMTSB Tool Set: Enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.*, 22:377–395.

[Fersht, 1997] Fersht, A. R. (1997). Nucleation mechanisms in protein folding. *Curr. Opin. Struct. Biol.*, 7(1):3–9.

[Frishman and Argos, 1995] Frishman, D. and Argos, P. (1995). Knowledge-based protein secondary structure assignment. *Proteins*, 23(4):566–579.

[Galán, 2008] Galán, J. E. (2008). Energizing type III secretion machines: what is the fuel? *Nature*, 15(2):127–128.

[Galán and Wolf-Watz, 2006] Galán, J. E. and Wolf-Watz, H. (2006). Protein delivery into eukaryotic cells by type III secretion machines. *Nature*, 444(7119):567–573.

[Gao et al., 2006] Gao, M., Sotomayor, M., Villa, E., Lee, E. H., and Schulten, K. (2006). Molecular mechanisms of cellular mechanics. *Phys. Chem. Chem. Phys.*, 8(32):3692–3706.

[Gao et al., 2002] Gao, M., Wilmanns, M., and Schulten, K. (2002). Steered molecular dynamics studies of titin I1 domain unfolding. *Biophys. J.*, 83(6):3435–3445.

[Garbuzynskiy and Kondratova, 2008] Garbuzynskiy, S. O. and Kondratova, M. S. (2008). Structural features of protein folding nuclei. *FEBS Lett.*, 582(5):768–772.

[Geierhaas et al., 2004] Geierhaas, C. D., Paci, E., Vendruscolo, M., and Clarke, J. (2004). Comparison of the transition states for folding of two Ig-like proteins from different superfamilies. *J. Mol. Biol.*, 343(4):1111–1123.

[Geney et al., 2006] Geney, R., Layten, M., Gomperts, R., Hornak, V., and Simmerling, C. (2006). Investigation of salt bridge stability in a Generalized Born solvent model. *J. Chem. Theory Comput.*, 2(1):115-127.

[Gophna et al., 2003] Gophna, U., Ron, E. Z., and Graur, D. (2003). Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene*, 312:151–163.

[Gordon et al., 2005] Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S., and Onufriev, A. (2005). H++: a server for estimating pKas and adding missing hydrogens to macromolecules. *Nucleic Acids Res.*, 33(Web Server issue).

[Gouet et al., 2003] Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.*, 31(13):3320–3323.

[Greenleaf et al., 2007] Greenleaf, W. J., Woodside, M. T., and Block, S. M. (2007). High-resolution, single-molecule measurements of biomolecular motion. *Annu. Rev. Biophys. Biomol. Struct.*, 36:171–190.

[Grubmüller et al., 1996] Grubmüller, H., Heymann, B., and Tavan, P. (1996). Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271(5251):997–999.

[Hawkins et al., 1995] Hawkins, G., Cramer, C., and Truhlar, D. (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.*, 246:122–129.

[Henzler-Wildman and Kern, 2007] Henzler-Wildman, K. and Kern, D. (2007). Dynamic personalities of proteins. *Nature*, 450(7172):964–972.

[Homma et al., 1990] Homma, M., DeRosier, D. J., and Macnab, R. M. (1990). Flagellar hook and hook-associated proteins of Salmonella typhimurium and their relationship to other axial components of the flagellum. *J. Mol. Biol.*, 213(4):819–832.

[Honda et al., 1999] Honda, S., Uedaira, H., Vonderviszt, F., Kidokoro, S., and Namba, K. (1999). Folding energetics of a multidomain protein, flagellin. *J. Mol. Biol.*, 293(3):719–732.

[Humphrey et al., 1996] Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD – Visual Molecular Dynamics. *J. Mol. Graph. Model.*, 14:33–38.

[Hunter and Sanders, 1990] Hunter, C. and Sanders, J. (1990). The nature of π-π interactions. *J. Am. Chem. Soc.*, 112:5525–5534.

[Ikeda et al., 1993] Ikeda, T., Yamaguchi, S., and Hotani, H. (1993). Flagellar growth in a filament-less Salmonella fliD mutant supplemented with purified hook-associated protein 2. *J. Biochem. (Tokyo)*, 114(1):39–44.

[Imada et al., 2007] Imada, K., Minamino, T., Tahara, A., and Namba, K. (2007). Structural similarity between the flagellar type III ATPase FliI and F1-ATPase subunits. *Proc. Natl. Acad. Sci. USA.*

[Ishima et al., 1991] Ishima, R., Akasaka, K., Aizawa, S., and Vonderviszt, F. (1991). Mobility of the terminal regions of flagellin in solution. *J. Biol. Chem.*, 266(35):23682–23688.

[Israelewitz et al., 2001] Israelewitz, B., Baudry, J., Gullingsrud, J., Kosztin, D., and Schulten, K. (2001). Steered molecular dynamics investigations of protein function. *J. Mol. Graph. Model.*, 19(1):13–25.

[Itzhaki et al., 1995] Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995). The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.*, 254(2):260–288.

[Ivankov et al., 2003] Ivankov, D. N., Garbuzynskiy, S. O., Alm, E., Plaxco, K. W., Baker, D., and Finkelstein, A. V. (2003). Contact order revisited: influence of protein size on the folding rate. *Protein Sci.*, 12(9):2057–2062.

[Jones, 1999] Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, 292:195–202.

[Jorgensen and Tirado-Rives, 2005] Jorgensen, W. L. and Tirado-Rives, J. (2005). Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proc. Natl. Acad. Sci. USA*, 102(19):6665–6670.

[Kabsch and Sander, 1983] Kabsch, W. and Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637.

[Kannan and Vishveshwara, 2000] Kannan, N. and Vishveshwara, S. (2000). Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng.*, 13(11):753–761.

[Kazmirski et al., 1999] Kazmirski, S. L., Li, A., and Daggett, V. (1999). Analysis methods for comparison of multiple molecular dynamics trajectories: applications to protein unfolding pathways and denatured ensembles. *J. Mol. Biol.*, 290(1):283–304.

[Keener, 2006] Keener, J. P. (2006). How Salmonella typhimurium measures the length of flagellar filaments. *Bull. Math. Biol.*, 68(7):1761–1778.

[Kenniston et al., 2003] Kenniston, J. A., Baker, T. A., Fernandez, J. M., and Sauer, R. T. (2003). Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell*, 114(4):511–520.

[Kirmizialtin et al., 2004] Kirmizialtin, S., Ganesan, V., and Makarov, D. E. (2004). Translocation of a β-hairpin-forming peptide through a cylindrical tunnel. *J. Chem. Phys.*, 121(20):10268–10277.

[Kitao and Go, 1999] Kitao, A. and Go, N. (1999). Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.*, 9(2):164–169.

[Kitao et al., 2006] Kitao, A., Yonekura, K., Maki-Yonekura, S., Samatey, F. A., Imada, K., Namba, K., and Gō, N. (2006). Switch interactions control energy frustration and multiple flagellar filament structures. *Proc. Natl. Acad. Sci. USA*, 103(13):4894–4899.

[Kohl et al., 1988] Kohl, N. E., Emini, E. A., Schleif, W. A., Davis, L. J., Heimbach, J. C., Dixon, R. A., Scolnick, E. M., and Sigal, I. S. (1988). Active human immunodeficiency virus protease is required for viral infectivity. *Proc. Natl. Acad. Sci. USA*, 85(13):4686–4690.

[Kostyukova et al., 1988] Kostyukova, A. S., Pyatibratov, M. G., Filimonov, V. V., and Fedorov, O. V. (1988). Flagellin parts acquiring a regular structure during polymerization are disposed on the molecule ends. *FEBS Lett.*, 241(1-2):141–144.

[Kubelka et al., 2004] Kubelka, J., Hofrichter, J., and Eaton, W. A. (2004). The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.*, 14(1):76–88.

[Kumara et al., 2006] Kumara, M. T., Srividya, N., Muralidharan, S., and Tripp, B. C. (2006). Bioengineered flagella protein nanotubes with cysteine loops: self-assembly and manipulation in an optical trap. *Nano. Lett.*, 6(9):2121–2129.

[Kuwajima et al., 1989] Kuwajima, G., Kawagishi, I., Homma, M., Asaka, J., Kondo, E., and Macnab, R. M. (1989). Export of an N-terminal fragment of Escherichia coli flagellin by a flagellum-specific pathway. *Proc. Natl. Acad. Sci. USA*, 86(13):4953–4957.

[Leach, 2001] Leach, A. R. (2001). *Molecular Modelling: Principles and Applications (2nd Edition)*. Prentice Hall.

[Lee et al., 2001] Lee, C., Schwartz, M. P., Prakash, S., Iwakura, M., and Matouschek, A. (2001). ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Molecular Cell*, 7(3):627–637.

[Linke and Grützner, 2008] Linke, W. A. and Grützner, A. (2008). Pulling single molecules of titin by AFM – recent advances and physiologial implications. *Pflügers Arch - Eur. J. Physiol.*, 456:101–115.

[Lu et al., 1998] Lu, H., Isralewitz, B., Krammer, A., Vogel, V., and Schulten, K. (1998). Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.*, 75(2):662–671.

[Macnab, 2003] Macnab, R. M. (2003). How bacteria assemble flagella. *Annu. Rev. Microbiol.*, 57:77–100.

[Maki-Yonekura et al., 2003] Maki-Yonekura, S., Yonekura, K., and Namba, K. (2003). Domain movements of HAP2 in the cap-filament complex formation and growth process of the bacterial flagellum. *Proc. Natl. Acad. Sci. USA*, 100(26):15528–15533.

[Malapaka et al., 2007] Malapaka, R. R., Adebayo, L. O., and Tripp, B. C. (2007). A deletion variant study of the functional role of the Salmonella flagellin hypervariable domain region in motility. *J. Mol. Biol.*, 365(4):1102–1116.

[Maruyama et al., 2008] Maruyama, Y., Momma, M., Mikami, B., Hashimoto, W., and Murata, K. (2008). Crystal structure of a novel bacterial cell-surface flagellin binding to a polysaccharide. *Biochemistry*, 47(5):1393–1402.

[McGuffin et al., 2000] McGuffin, L. J., Bryson, K., and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16:404–405.

[Minamino and Namba, 2008] Minamino, T. and Namba, K. (2008). Distinct roles of the FliI ATPase and proton motive force in bacterial flagellar protein export. *Nature*, 451(7177):485–488.

[Missimer et al., 2007] Missimer, J. H., Steinmetz, M. O., Baron, R., Winkler, F. K., Kammerer, R. A., Daura, X., and van Gunsteren, W. F. (2007). Configurational entropy elucidates the role of salt-bridge networks in protein thermostability. *Protein Sci.*, 16(7):1349–1359.

[Moraes et al., 2008] Moraes, T. F., Spreter, T., and Strynadka, N. C. J. (2008). Piecing together the Type III injectisome of bacterial pathogens. *Curr. Opin. Struct. Biol.*, 18(2):258–266.

[Ng et al., 2005] Ng, S. P., Rounsevell, R. W., Steward, A., Geierhaas, C. D., Williams, P. M., Paci, E., and Clarke, J. (2005). Mechanical unfolding of TNfn3: The unfolding pathway of a FNIII domain probed by protein engineering, AFM and MD simulation. *J. Mol. Biol.*, 350(4):776–789.

[Oberhauser and Carrión-Vázquez, 2008] Oberhauser, A. F. and Carrión-Vázquez, M. (2008). Mechanical biochemistry of proteins one molecule at a time. *J. Biol. Chem.*, 283(11):6617–6621.

[Oberhauser et al., 2001] Oberhauser, A. F., Hansma, P. K., Carrion-Vazquez, M., and Fernandez, J. M. (2001). Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc. Natl. Acad. Sci. USA*, 98(2):468–472.

[Ohta et al., 2004] Ohta, S., Alam, M. T., Arakawa, H., and Ikai, A. (2004). Origin of mechanical strength of bovine carbonic anhydrase studied by molecular dynamics simulation. *Biophys. J.*, 87(6):4007–4020.

[Onufriev et al., 2004] Onufriev, A., Bashford, D., and Case, D. A. (2004). Exploring protein native states and large-scale conformational changes with a modified Generalized Born model. *Proteins*, 55(2):383–394.

[Oroguchi et al., 2005] Oroguchi, T., Ikeguchi, M., Saeki, K., Kamagata, K., Sawano, Y., Tanokura, M., Kidera, A., and Kuwajima, K. (2005). Atomically detailed description of the unfolding of α-lactalbumin by the combined use of experiments and simulations. *J. Mol. Biol.*, 354(1):164–172.

[Pabón and Amzel, 2006] Pabón, G. and Amzel, L. M. (2006). Mechanism of titin unfolding by force: insight from quasi-equilibrium molecular dynamics calculations. *Biophys. J.*, 91(2):467–472.

[Pallen and Matzke, 2006] Pallen, M. J. and Matzke, N. J. (2006). From the origin of species to the origin of bacterial flagella. *Nature Reviews Microbiol.*, 4(10):784–790.

[Parot et al., 2007] Parot, P., Dufrêne, Y. F., Hinterdorfer, P., Le Grimellec, C., Navajas, D., Pellequer, J.-L., and Scheuring, S. (2007). Past, present and future of atomic force microscopy in life sciences and medicine. *J. Mol. Recognit.*, 20(6):418–431.

[Paul et al., 2008] Paul, K., Erhardt, M., Hirano, T., Blair, D. F., and Hughes, K. T. (2008). Energy source of flagellar type-III secretion. *Nature*, 451(7177):489–492.

[Prakash and Matouschek, 2004] Prakash, S. and Matouschek, A. (2004). Protein unfolding in the cell. *Trends Biochem. Sci.*, 29(11):593–600.

[Raman et al., 2007] Raman, E. P., Takeda, T., Barsegov, V., and Klimov, D. K. (2007). Mechanical unbinding of Aβ peptides from amyloid fibrils. *J. Mol. Biol.*, 373(3):785–800.

[Rohs et al., 1999] Rohs, R., Etchebest, C., and Lavery, R. (1999). Unraveling proteins: a molecular mechanics study. *Biophys. J.*, 76(5):2760–2768.

[Ryckaert et al., 1977] Ryckaert, J., Ciccotti, G., and Berendsen, H. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.*, 23:327–341.

[Salazar-Gonzalez and McSorley, 2005] Salazar-Gonzalez, R. and McSorley, S. J. (2005). Salmonella flagellin, a microbial target of the innate and adapative immune system. *Immun. Lett.*, 101:117–122.

[Samatey et al., 2001] Samatey, F. A., Imada, K., Nagashima, S., Vonderviszt, F., Kumasaka, T., Yamamoto, M., and Namba, K. (2001). Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature*, 410(6826):331–337.

[Samorì et al., 2005] Samorì, B., Zuccheri, G., and Baschieri, R. (2005). Protein unfolding and refolding under force: methodologies for nanomechanics. *Chemphyschem*, 6(1):29–34.

[Sato and Raleigh, 2007] Sato, S. and Raleigh, D. P. (2007). Kinetic isotope effects reveal the presence of significant secondary structure in the transition state for the folding of the N-terminal domain of L9. *J. Mol. Biol.*, 370(2):349–355.

[Sato et al., 2005] Sato, T., Esaki, M., Fernandez, J. M., and Endo, T. (2005). Comparison of the protein-unfolding pathways between mitochondrial protein import and atomic-force microscopy measurements. *Proc. Natl. Acad. Sci. USA*, 102(50):17999–18004.

[Schaeffer et al., 2008] Schaeffer, D. R., Fersht, A., and Daggett, V. (2008). Combining experiment and simulation in protein folding: closing the gap for small model systems. *Curr. Opin. Struct. Biol.*, 18(1):4–9.

[Schiebel et al., 1991] Schiebel, E., Driessen, A. J., Hartl, F. U., and Wickner, W. (1991). $\Delta\mu_H^+$ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell*, 64(5):927–939.

[Scott et al., 2006] Scott, K. A., Randles, L. G., Moran, S. J., Daggett, V., and Clarke, J. (2006). The folding pathway of spectrin R17 from experiment and simulation: using experimentally validated md simulations to characterize states hinted at by experiment. *J. Mol. Biol.*, 359(1):159–173.

[Sebestyén et al., 2008] Sebestyén, A., Muskotál, A., Végh, B. M., and Vonderviszt, F. (2008). The hypervariable D3 domain of Salmonella flagellin is an autonomous folding unit. *Protein Peptide Lett.*, 15(1):54–57.

[Sham et al., 2002] Sham, Y. Y., Ma, B., Tsai, C. J., and Nussinov, R. (2002). Thermal unfolding molecular dynamics simulation of Escherichia coli dihydrofolate reductase: thermal stability of protein domains and unfolding pathway. *Proteins*, 46(3):308–320.

[Shih et al., 2007] Shih, A. Y., Freddolino, P. L., Arkhipov, A., and Schulten, K. (2007). Assembly of lipoprotein particles revealed by coarse-grained molecular dynamics simulations. *J. Struct. Biol.*, 157(3):579–592.

[Sickmeier et al., 2007] Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J., Uversky, V. N., Obradovic, Z., and Dunker, A. K. (2007). DisProt: the Database of Disordered Proteins. *Nucleic Acids Res.*, 35(Database issue).

[Stafford et al., 2007] Stafford, G. P., Evans, L. D., Krumscheid, R., Dhillon, P., Fraser, G. M., and Hughes, C. (2007). Sorting of early and late flagellar subunits after docking at the membrane ATPase of the type III export pathway. *J. Mol. Biol.*, 374(4):877–882.

[Still et al., 1990] Still, W., Tempczyk, A., Hawley, R., and Hendrickson, T. (1990). Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, 112(16):6127–6129.

[Stone, 1998] Stone, J. (1998). An efficient library for parallel ray tracing and animation. Master's thesis, Computer Science Department, University of Missouri-Rolla.

[Sulkowska and Cieplak, 2008] Sulkowska, J. I. and Cieplak, M. (2008). Stretching to understand proteins – A survey of the Protein Data Bank. *Biophys. J.*, 94:6–13.

[Tamura et al., 1997] Tamura, Y., Gekko, K., Yoshioka, K., Vonderviszt, F., and Namba, K. (1997). Adiabatic compressibility of flagellin and flagellar filament of Salmonella typhimurium. *Biochim. Biophys. Acta.*, 1335(1-2):120–126.

[Tian and Andricioaei, 2005] Tian, P. and Andricioaei, I. (2005). Repetitive pulling catalyzes co-translocational unfolding of barnase during import through a mitochondrial pore. *J. Mol. Biol.*, 350(5):1017–1034.

[Trachtenberg and DeRosier, 1988] Trachtenberg, S. and DeRosier, D. J. (1988). Three-dimensional reconstruction of the flagellar filament of Caulobacter crescentus. a flagellin lacking the outer domain and its amino acid sequence lacking an internal segment. *J. Mol. Biol.*, 202(4):787–808.

[Trylska et al., 2007] Trylska, J., Tozzini, V., Chang, C.-E., and McCammon, J. A. (2007). HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophys. J.*, 92(12):4179–4187.

[Vendruscolo et al., 2002] Vendruscolo, M., Dokholyan, N. V., Paci, E., and Karplus, M. (2002). Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E*, 65(6 Pt 1).

[Vonderviszt et al., 1989] Vonderviszt, F., Kanto, S., Aizawa, S., and Namba, K. (1989). Terminal regions of flagellin are disordered in solution. *J. Mol. Biol.*, 209(1):127–133.

[Wasmer et al., 2008] Wasmer, C., Lange, A., Van Melckebeke, H., Siemer, A. B., Riek, R., and Meier, B. H. (2008). Amyloid fibrils of the HET-s(218-289) prion form a β solenoid with a triangular hydrophobic core. *Science*, 319(5869):1523–1526.

[Weber-Ban et al., 1999] Weber-Ban, E. U., Reid, B. G., Miranker, A. D., and Horwich, A. L. (1999). Global unfolding of a substrate protein by the Hsp100 chaperone ClpA. *Nature*, 401(6748):90–93.

[Xu et al., 2006] Xu, Y., Purkayastha, P., and Gai, F. (2006). Nanosecond folding dynamics of a three-stranded β-sheet. *J. Am. Chem. Soc.*, 128(49):15836–15842.

[Yang and Chng, 2008] Yang, L.-W. and Chng, C.-P. (2008). Coarse-grained models reveal functional dynamics – I. Elastic Network Models – theories, comparisons and perspectives,. *Bioinformatics and Biology Insights*, 2:25–45.

[Yonekura et al., 2000] Yonekura, K., Maki, S., Morgan, D. G., DeRosier, D. J., Vonderviszt, F., Imada, K., and Namba, K. (2000). The bacterial flagellar cap as the rotary promoter of flagellin self-assembly. *Science*, 290(5499):2148–2152.

[Yonekura et al., 2003] Yonekura, K., Maki-Yonekura, S., and Namba, K. (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature*, 424(6949):643–650.

[Zheng et al., 1990] Zheng, C., Wong, C. F., and McCammon, A. J. (1990). Fluctuation of the solvent-accessible surface area of tuna ferrocytochrome c. *Biopolymers*, 29(14):1877–1883.