Nanoscale

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Nanoscale*, 2020, **12**, 5422

Protein unfolding by SDS: the microscopic mechanisms and the properties of the SDS-protein assembly†

David Winogradoff, [a,b] Shalini John[b] and Aleksei Aksimentiev *[a,b,c]

The effects of detergent sodium dodecyl sulfate (SDS) on protein structure and dynamics are fundamental to the most common laboratory technique used to separate proteins and determine their molecular weights: polyacrylamide gel electrophoresis. However, the mechanism by which SDS induces protein unfolding and the microstructure of protein–SDS complexes remain largely unknown. Here, we report a detailed account of SDS-induced unfolding of two proteins—I27 domain of titin and β-amylase—obtained through all-atom molecular dynamics simulations. Both proteins were found to spontaneously unfold in the presence of SDS at boiling water temperature on the time scale of several microseconds. The protein unfolding was found to occur *via* two distinct mechanisms in which specific interactions of individual SDS molecules disrupt the protein's secondary structure. In the final state of the unfolding process, the proteins are found to wrap around SDS micelles in a fluid necklace-and-beads configuration, where the number and location of bound micelles changes dynamically. The global conformation of the protein was found to correlate with the number of SDS micelles bound to it, whereas the number of SDS molecules directly bound to the protein was found to define the relaxation time scale of the unfolded protein. Our microscopic characterization of SDS–protein interactions sets the stage for future refinement of SDS–enabled protein characterization methods, including protein fingerprinting and sequencing using a solid-state nanopore.

Received 25th October 2019,
Accepted 11th February 2020

DOI: 10.1039/c9nr09135a

rsc.li/nanoscale

1. Introduction

The three-dimensional shape of a protein is fundamentally important to its biological function. In living cells, nascent proteins begin to fold as soon as they are synthesized by a ribosome,[1,2] a process that is often aided by chaperones such as GroEL/ES[3] or by translocons[4] that direct newly synthesized peptide chains into designated membranes or compartments. Protein unfolding or misfolding is known to promote protein aggregation,[5] which in turn is associated with a range of health conditions, for example, neurodegenerative diseases.[6] Not surprisingly, determining exactly how disordered peptide chains fold into well-defined, three-dimensional structures has long been a topic of scientific inquiry,[7–12] alongside characterizing the modes of protein dynamics.[13–16] The reverse process, protein unfolding, has received its fair share of interest from both experimental[17] and simulation[18] investigations that were primarily directed at gaining insights into the protein folding mechanisms. Protein unfolding can be realized by a variety of means, including a change in the solution temperature,[19–21] pH,[22,23] or pressure,[24] the application of a mechanical force,[25–27] or the introduction of chemical denaturants, such as urea[28–30] or sodium dodecyl sulfate (SDS).[31]

SDS–induced unfolding of proteins deserves particular attention as it underlies the most common experimental protocol for determining the molecular weight of a protein, SDS polyacrylamide gel electrophoresis.[32] In a typical measurement, a protein sample is first mixed with an SDS buffer at high temperature, which unfolds the proteins in a process that is known to depend, in part, on the protein's secondary structure.[33] Subject to an applied electric field, a protein–SDS complex migrates towards the positive electrode because SDS unfolding imparts the protein with a negative charge. Conducting such electrophoresis experiments on a polyacrylamide gel sorts the proteins from a mixture according to molecular weight, with smaller proteins traveling farther. Staining the proteins with a dye[34] allows direct visualization of the pro-

[a]*Center for the Physics of Living Cells, University of Illinois at Urbana–Champaign, Urbana, IL, USA. E-mail: aksiment@illinois.edu; Tel: +(217) 333-6495*
[b]*Department of Physics, University of Illinois at Urbana–Champaign, Urbana, IL, USA*
[c]*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana–Champaign, Urbana, IL, USA*

†Electronic supplementary information (ESI) available. See DOI: 10.1039/C9NR09135A

teins' locations in a gel after electrophoresis. A quantitative estimate of molecular weight is obtained through comparison to a ladder of protein markers of known molecular weights. SDS-unfolding of proteins is also central to Western Blot analysis,[35] where proteins are identified from a complex biological mixture with the help of antibody–antigen interaction. Recently, SDS treatment of proteins has been explored for nanopore protein sequencing, where the amino acid sequence of a protein is determined by the modulations of the ionic current flowing through a nanopore.[36,37]

Despite being an essential part of the most commonly used protein characterization techniques, the very mechanism by which SDS induces protein unfolding has not yet been unequivocally established. Based on rheological,[38] binding,[39] scattering,[40–44] and spectroscopic studies,[45–47] several structural models of a protein–SDS complex have been proposed, including a rigid rods model,[38,48] a necklace-and-beads model[32,49] where SDS micelles cover several protein regions spatially separated by uncovered regions of protein, and a flexible helix model,[50] in which the protein wraps around a single, cylindrical SDS micelle. At extremely low concentrations of SDS, proteins have been reported to adopt a compact state.[51] Over the past decade, several scattering studies provided significant evidence for protein-decorated SDS micelle complex formation,[42–44,52–55] although the presence of intact disulfide bonds may alter the configuration. Questions remain over how exactly SDS denatures proteins in the presence of heat, the degree to which SDS–protein complexes change dynamically, and the relative preference of specific categories of protein amino-acid residues to interact with SDS molecules.

Molecular dynamics (MD) simulations provide a means to explore the fine details of protein–SDS structure and dynamics. Coarse-grained and atomistic MD studies have examined the self-assembly of detergent molecules into micelles,[56–58] including an examination of hydrogen bonds at the micelle–water interface.[59,60] Furthermore, several MD simulations have been performed to explore protein–SDS interactions. An MD study provided an atomistic-level description of SDS aggregation and helix association of transmembrane protein GpA in the presence of SDS.[61] Another MD study observed individual protein helices to partially unfold after being placed within an SDS micelle.[62] Studies combining experiment and MD simulation investigated the effects of detergents DPC[63,64] and DDM[65,66] on membrane proteins, observing the proteins not to undergo major unfolding, but that immersion in DPC micelles would alter a membrane protein's structure and function.[64] MD simulations examined specific interactions of SDS with charged amino acids of model peptides[67] and with the common elements of protein structure: α-helices and β-sheets.[68] By combining MD simulations with nanopore measurements, we recently showed that SDS-treated proteins can translocate through a nanopore in an unfolded state, moving through the pore by electrophoresis.[69]

Here, we report a detailed microscopic account of spontaneous (*i.e.*, not triggered by the application of external force or electric field) protein unfolding induced by SDS at boiling water temperature based on the results of multiple all-atom MD simulations. We characterize the mechanisms of the unfolding process, the global features of protein–SDS assembly, and the local properties of protein–SDS interactions using two proteins, titin I27 domain and β-amylase, that differ in molecular weight and secondary structure composition. The unfolding process was found to depend on specific interactions between individual SDS molecules and the proteins, through two distinct mechanisms that will be described in further detail in the Results and Discussion. Upon unfolding, we observe the proteins to wrap around SDS micelles in a fluid necklace-and-beads configuration, in which the location and the number of bound micelles can change on the microsecond timescale.

2. Results and discussion

2.1. Spontaneous unfolding of the I27 domain of titin induced by SDS

To determine the molecular mechanism by which SDS induces protein unfolding under conditions typical to sample preparation of SDS-PAGE and nanopore translocation experiments, we built an all-atom system, Fig. 1a, containing one folded I27 domain of titin surrounded by 400 mM NaCl solution and 120 singly dissolved SDS molecules randomly dispersed throughout the system, corresponding to a 110 mM SDS concentration —a typical SDS concentration used for protein unfolding.[69,70] Starting from that configuration, three equilibration MD simulations were run differing only by the random seed used to initialize atom velocities. Each system was simulated at 1 bar pressure and boiling (373 K) temperature for 6 μs. Based on nanopore experiments performed under similar conditions,[69] we estimate the critical micelle concentration to be about 3.5 mM SDS. Fig. 1b illustrates one such trajectory; Fig. S1–S4 and Movies S1–S3† illustrate the course of all three trajectories. Structural details about I27 domain of titin are provided in ESI Fig. S5.†

Despite starting from a dispersed configuration, the majority of SDS molecules formed micelles within the first 40 ns of each MD simulations, Fig. S6.† During and after this time period, SDS monomers and micelles associated with the folded protein, both at the protein's disordered termini and at the loops connecting its folded core. Upon such binding, the protein was observed to unfold, which we characterized quantitatively by plotting the root mean squared deviation (RMSD) of the protein coordinates with respect to the folded structure, Fig. 1c, and the protein's radius of gyration R_g, Fig. S7.†

Clearly, the unfolding process in the three systems did not follow the same path: two of the three systems became unfolded within the first 0.5 μs, whereas the third system remained stable for approximately 1.5 μs followed by an abrupt transition to an unfolded state. Differences in the unfolding behavior also occurred after the protein lost most of its initial folded structure: the protein was seen to undergo a spontaneous transition between compact and extended confor-

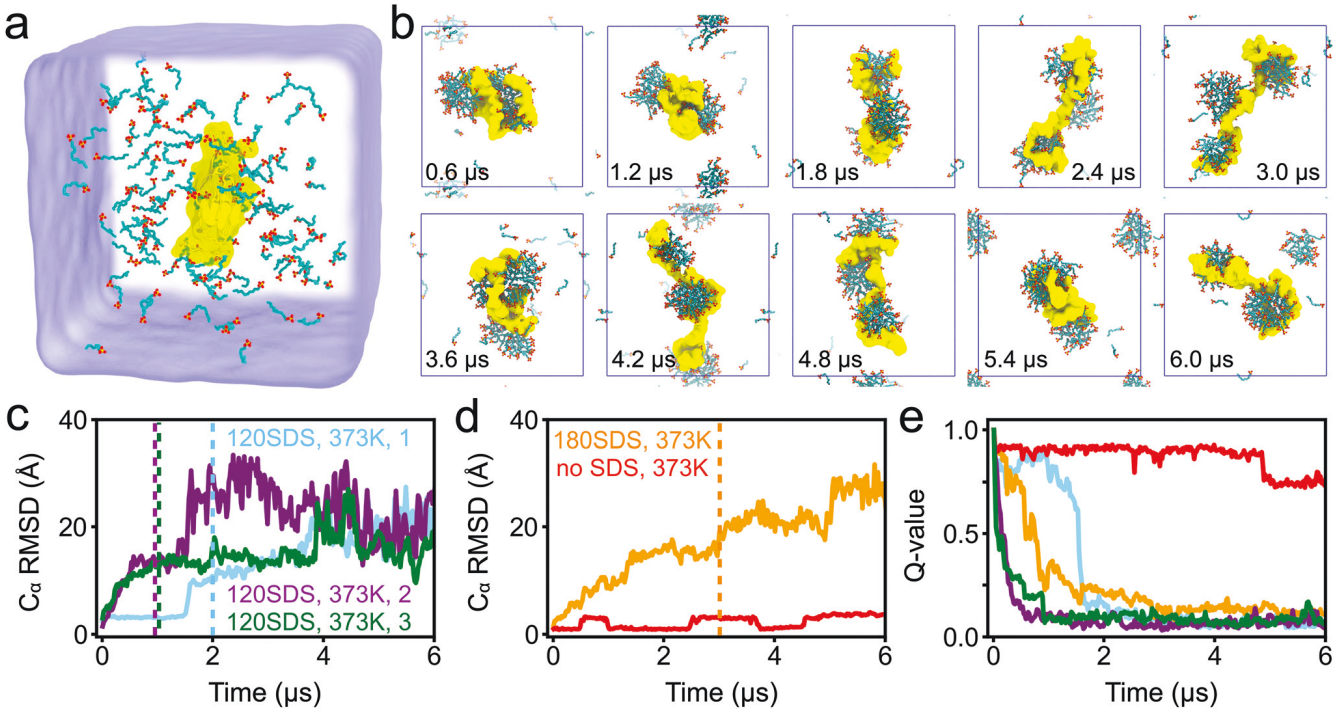

Fig. 1 Spontaneous unfolding of titin I27 domain by SDS. (a) Simulation setup: the I27 domain of a titin protein in its folded conformation (yellow molecular surface) is surrounded by 120 randomly-placed SDS molecules (red and cyan bonds) and 0.4 M NaCl solution (blue semi-transparent surface). (b) Spontaneous unfolding of titin I27 by SDS and the conformational dynamics of the SDS-I27 assembly. Images illustrate the state of the simulation system over the course of a 6 μs equilibrium MD simulation at 373 K. Blue squares denote the system's dimensions. Periodic images of some SDS molecules are shown to highlight micelle formation. (c) RMSD of the protein Cα atoms from their crystal structure coordinates in three independent 6 μs simulations (blue, purple, green) performed at 373 K. Each system contained 120 SDS molecules. Panel b illustrates the trajectory of the second replica with 120 SDS, purple curve in panel c. (d) RMSD of the protein Cα atoms from their crystal structure coordinates in a 6 μs simulation performed in the absence of SDS molecules (red) and when 180 SDS molecules were present in the system (orange). RMSD calculations omit intrinsically disordered N- and C-terminal tails. (e) The fraction of native contacts Q *versus* simulation time for all five titin I27 systems. The color coding is the same as in panels c and d. The Q value calculations were done only for those residues that were identified as β sheets or bridges in the folded structure. Vertical dashed lines in panels c and d indicate the moment the protein unfolds completely (*i.e.* $Q < 0.1$). Note, purple and green vertical lines overlap in panel c.

mations, Fig. 1b, which is reflected by the stochastic changes in the RMSD value, Fig. 1c. Spontaneous unfolding of titin's I27 domain was also observed in another MD simulation carried out at a 50% higher SDS concentration, where 180 SDS molecules were initially dispersed within the same simulation volume, Fig. 1d and S8 and Movie S4.† To directly show that the spontaneous unfolding originates from the protein–SDS interactions, we simulated titin I27 in the absence of SDS under the same simulation conditions, Fig. 1d, Fig. S9 and Movie S5.† Over the course of this 6 μs trajectory, titin I27 maintained its folded conformation, Fig. 1d, transitioning between two stable states that differed from one another by the conformation of the protein loop formed by residues 53 to 65: the loop was either folded over and in contact with β-strands C & D or not in contact with strands C & D and more free to flop around.

To gain more insight into kinetics of the protein unfolding, we plotted the fraction of native contacts Q [71] *versus* simulation time for all five MD trajectories, Fig. 1e, see Methods for the mathematical definition of Q. Values of Q range from 0 and 1, with $Q = 1$ indicating a native, folded state and $Q = 0$ indicating a state with no resemblance to the native state. In practice, observing a Q value close to 0 is highly unlikely for proteins and, hence, structures of $Q < 0.1$ are considered to be unfolded. Plots of the Q value, Fig. 1e, indicate that the protein unfolding transitions occur rather abruptly, that reduction of the protein's RMSD during the simulation trajectory does not indicate partial protein refolding, and that some abrupt unfolding transitions, such as in the 180 SDS system, are not evident from the RMSD or R_g plots. Experiments have suggested that refolding is possible after unfolding by SDS, especially in the presence of other surfactants.[43,44,53]

2.2. Mechanisms of SDS–induced protein unfolding

The multiple MD trajectories captured the process of SDS–induced unfolding in all-atom detail, providing an opportunity to determine the molecular mechanism by which SDS induces protein unfolding. We found specific interactions between individual SDS molecules and the protein to play a critical role in promoting protein unfolding *via* one of the following two mechanisms. An individual SDS molecule can become inserted directly in between of two neighboring β strands, with

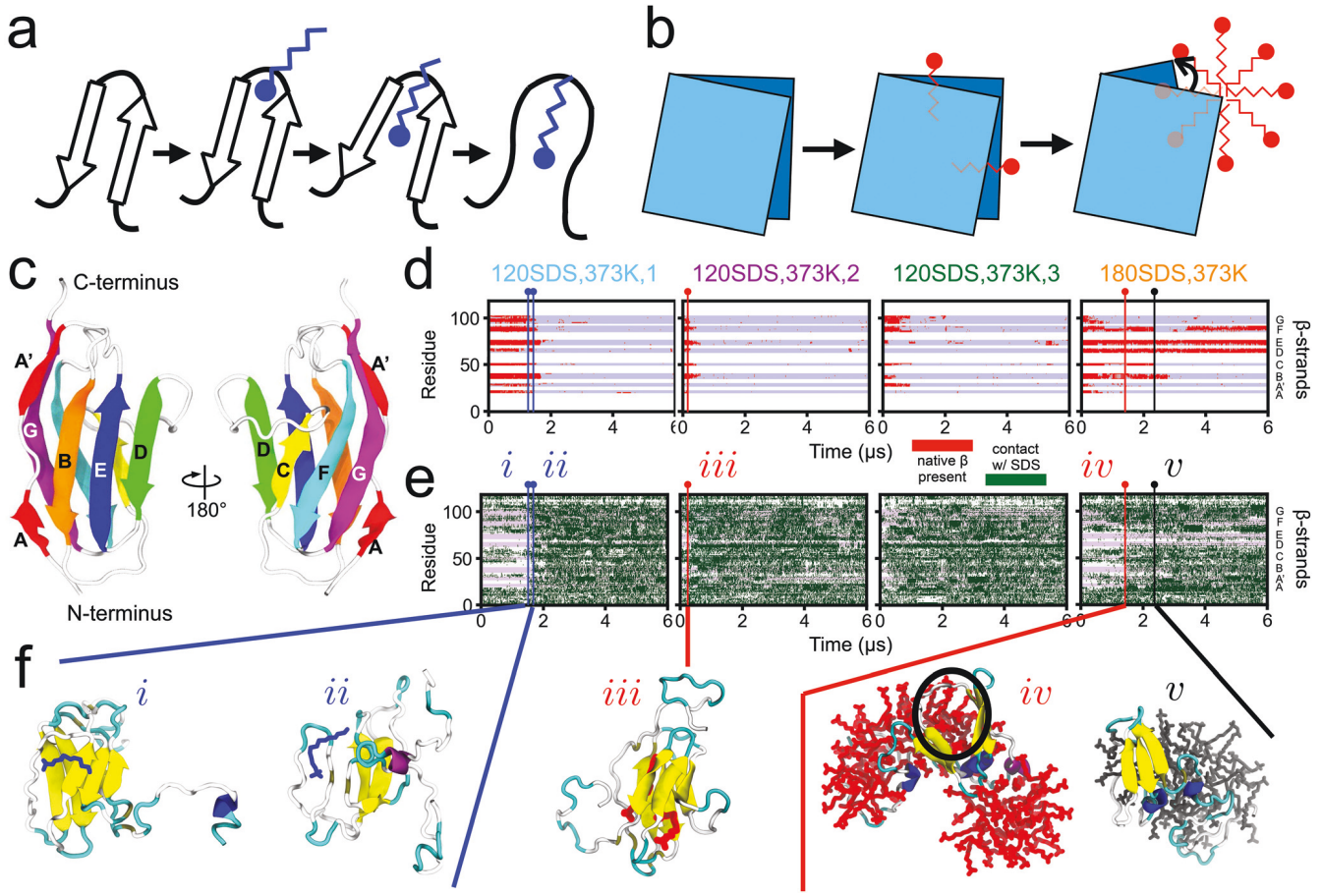

Fig. 2 Mechanisms of protein unfolding induced by SDS. (a) Illustration of an unfolding mechanism where an SDS molecule (blue) inserts between two neighboring β strands at the surface of a protein, disrupting the interactions between the strands. (b) Illustration of a protein unfolding mechanism where an SDS molecules (red) inserts between two β sheets, allowing an SDS micelle to wedge the two sheets apart. (c) Cartoon representation of the I27 domain of titin, β strands distinguished by color and labeled. (d, e) Local structure of titin I27 as a function of simulation time. Residues preserving their native β fold are shown in red (panel d), whereas those in direct contact with SDS are shown in green (panel e). Light blue background highlights β-sheet parts of the folded structure. (f) Examples of SDS–induced protein unfolding events. The time of each event is indicated by a vertical line in panels d and e. SDS molecules are shown in blue (*i, ii*) or red (*iii, iv*) according to the mechanisms depicted in panels a and b, respectively. The black oval in snapshot *iv* indicates SDS molecules that push the protein apart. Snapshot *v* shows a partially unfolded conformation stabilized by SDS molecules (shown in gray).

the SDS molecule's headgroup leading the way to disrupt the protein's inter-strand bonds, Fig. 2a. Alternatively, an individual SDS molecule can penetrate tail-first into the central fold of the protein, making it possible for an SDS micelle to wedge the two β sheets apart, Fig. 2b. Below we provide examples of how these two mechanisms produce protein unfolding.

Titin's I27 domain consists of two β sheets—ABED and A′GFC—that are adjacent in the native, folded state, Fig. 2c. To monitor the process of SDS–induced unfolding, we identify the protein residues that preserve their native fold along the simulation trajectory, Fig. 2d, and the residues that are in direct contact with SDS, Fig. 2e. The superposition of the two plots provides the timeline of the unfolding process. Fig. S4† illustrates the secondary structure changes in all four MD trajectories.

The major unfolding events in the trajectory of the first replica with 120 SDS occurred at 1.48 and 1.56 μs and were initiated by the mechanism schematically shown in Fig. 2a. Being a part of an SDS micelle, one SDS molecule inserted between two β strands at the protein surface, Fig. 2f*i*. Subsequently, three neighboring β strands were reduced to two, and then to a single β bridge, Fig. 2f*ii*. Thus, the contacts between the protein and the SDS molecule disrupted the inter-protein contacts needed to maintain the local secondary structure. A more detailed look at the unfolding mechanism observed for replica one of titin with 120 SDS at boiling temperature is provided in ESI Fig. S10a.†

In the other three systems, protein unfolding proceeded by the mechanism schematically shown in Fig. 2b. For the second replica with 120 SDS, two specific SDS molecules inserted tail-first into titin I27's central fold, *i.e.* in between the two adjacent β sheets, Fig. 2f*iii*. At the time of insertion (about 0.05 μs), these two SDS molecules were not part of a larger micelle. Once inserted, the molecules pushed the two β sheets apart

slightly. Following that, an entire SDS micelle attached to the protein surface between the two sheets, wedging them apart and producing a major disruption of the titin I27's hydrophobic core. A detailed visualization at the key steps along the unfolding pathway for replica two is provided in Fig. S10b.†

Major protein unfolding in the third replica with 120 SDS (at 1.0 μs) and the system containing 180 SDS (at 3.0 μs) proceeded *via* a similar mechanism. Fig. 2f*iv* illustrates the wedging step of the unfolding process in molecular detail. In the 180 SDS system, after significant unfolding, several native β strands settled on the surface of a single SDS micelle, Fig. 2f*v*, and persisted throughout the remainder of the simulation, Fig. 2d and S8 and Movie S4.† Interestingly, association with SDS micelles was observed to facilitate the formation of non-native helical structure for two of the three replicas containing titin I27 plus 120 SDS at boiling temperature, see Fig. S11.† Conversion from β to α elements was also observed in small angle X-ray scattering (SAXS) experiments.[44] In general, experimental studies have demonstrated that SDS can stabilize proteins in the presence of denaturing urea or heat,[72–74] one proposed explanation being that SDS displaces bound water.[75]

One important distinction between the two unfolding mechanisms is that, for the second mechanism, the protein's global arrangement is disrupted initially and then local elements dissolve, whereas, in the first mechanism, a local structural disruption leads to a global structural unfolding. As the occurance of specific interactions at the center of each mechanism is stochastic, so is the probability of observing either unfolding pathway. Note that starting our simulations with singly distributed SDS molecules might have increased the likelihood of observing SDS unfolding *via* the first mechanism. Indeed, a recent MD study reported an ACBP protein to lose about half of its native contacts over the course of a 1 μs room-temperature simulation when the simulation was started with randomly dispersed SDS monomers and only one quarter of such contacts over the same time period when the simulation was started in the presence of pre-formed SDS micelles.[76]

2.3. Global features of protein–SDS assembly

To characterize the structure of a protein–SDS complex that forms after SDS unfolds the protein, we computed the number of SDS molecules bound to the protein in our MD trajectories. For these calculations, we used only those portions of the trajectories where the protein was unfolded. Fig. 3a shows that the number of SDS molecules that form direct contacts with the protein converges to about 50 for all four titin I27 systems. On the other hand, the total number of SDS bound, directly or through SDS-mediated contacts, varied significantly within a single trajectory as well as across the systems.

The abrupt jumps in the total number of SDS molecules bound to the protein reflect stochastic changes in the number of SDS micelles bound to the protein, Fig. 3b. In our simulations, SDS molecules were found to either form micelles at the surface of the protein or in free solution, or be dispersed in solution, Fig. 3c. The number of micelles bound to the protein could change over the course of the trajectory through either binding of an additional micelle to the protein, unbinding of a micelle off the protein, or fusion of the micelles. The average number of SDS molecules comprising a single micelle ranged from about 30 to 50, Fig. 3d, which agrees with the previous experimental estimate of about 36 SDS molecules per micelle.[77] The protein–bound micelles were found to consist of slightly more SDS molecules than the unbound micelles, Fig. S12.† Control simulations carried out in the absence of any protein indicate that the average size of the micelles does not strongly depend on the presence of the protein, Fig. S12.† The concentration of dispersed SDS molecules in simulation also remained unaffected by the presence of a protein and was about 7 mM on average. Experiments have shown that the SDS micelle aggregation number goes down as temperature increases,[78] which explains, in part, why the SDS micelles at 373 K are smaller than those observed experimentally at room temperature. The ionic salt concentration also matters, as a globular-to-rod transition occurs at a high enough concentration,[79] which we expect our simulation conditions to be below. However, the limited sampling of micelle fusion and fission events does not allow us to make a quantitative statement about the equilibrium average micelle size in the absence of protein.

In a complex with SDS molecules, the unfolded protein was found to adopt two types of global conformations—either a compact globule or a more extended tubule, Fig. 3c. Accordingly, two peaks are observed in the protein's R_g distribution, Fig. 3e; please see Fig. S7† for plots of R_g as a function of simulation time. Thus, the unfolded protein was more likely to adopt an extended state when more micelles were bound to it, Fig. 3e. In our simulations, we observed one, two or three micelles forming a complex with the unfolded protein, with two bound micelles being the most common (69%) state, Fig. 3e inset.

2.4. Temperature dependence of protein–SDS interactions

To examine the effects of temperature on protein–SDS interactions, we performed two independent simulations of a system containing 120 SDS molecules and unfolded titin I27 starting from two distinct configurations observed in the 373 K runs: a compact, globule-like state having two SDS micelles bound to the protein, Fig. 4a, and a more extended protein conformation with three SDS micelles bound, Fig. 4b. From the four structures shown in Fig. 4a and b, we can see that the protein is often found at the SDS micelle's amphiphilic interface, weaving between SDS head groups, consistent with the conclusions of an experimental study.[80] During these room temperature (300 K) simulations, the number of SDS molecules bound to the protein either directly, or *via* SDS–mediated contacts, did not change significantly, Fig. 4c, averaging 0.67 and 1.0 SDS molecules per amino acid residue in replica one

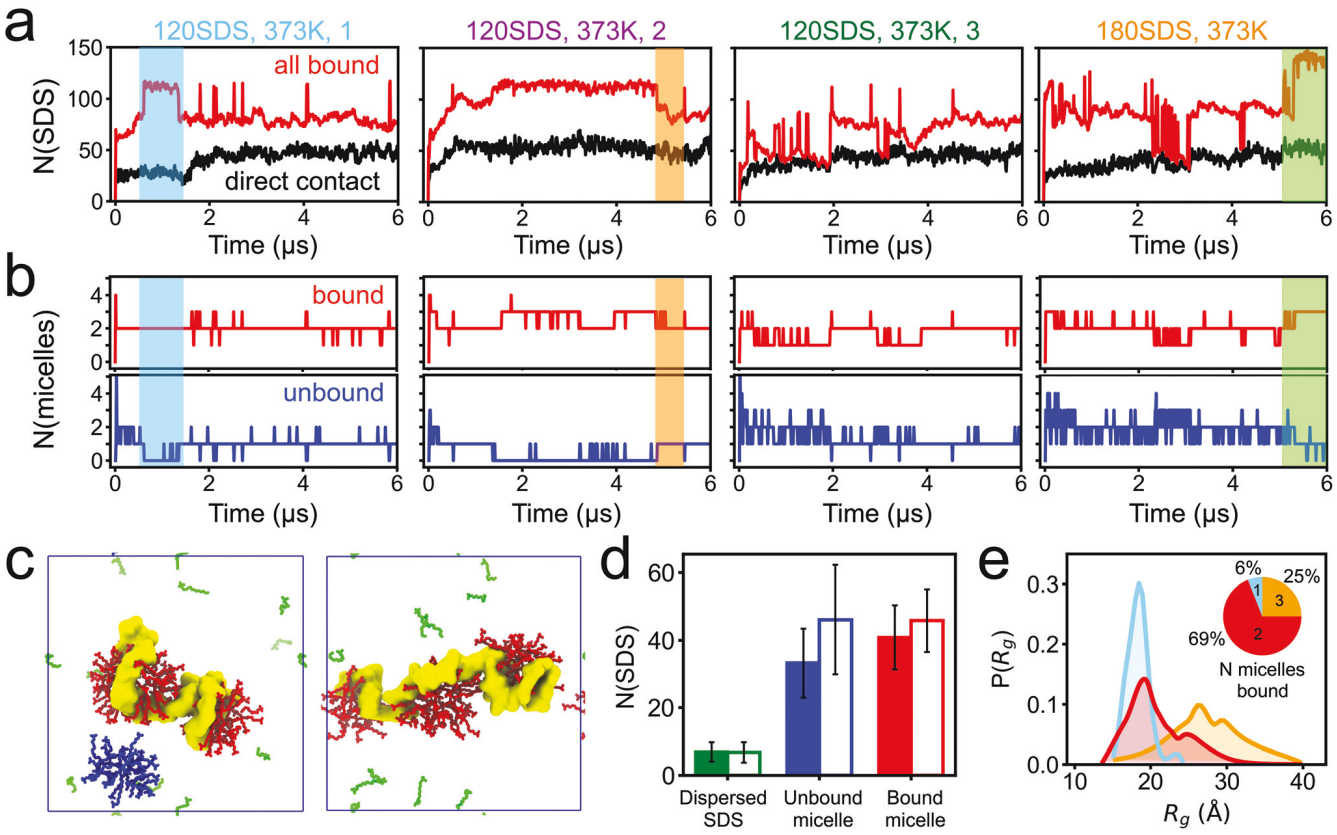

Fig. 3 Global features of protein–SDS assembly. (a) The number of SDS molecules in direct contact with the protein (black), and of all SDS molecules bound to the protein (directly or *via* SDS-mediated contacts, red) *versus* simulation time. (b) The number of protein-associated (red) and free-standing (blue) micelles *versus* simulation time. Shaded boxes indicate the fusion of an unbound and bound micelle (light blue), the unbinding of a micelle off the protein (orange), and binding of an additional micelle to the protein (lime). Fig. S1–S3, S8 and Movies S1–S4† illustrate these global rearrangements. (c) Examples of a system's configurations where SDS molecules form protein-bound (red) and free-standing (blue) micelles. Dispersed SDS molecules are shown in green, the protein in yellow. (d) The average number of SDS molecules found in the protein-bound micelles (red), free-standing micelles (blue), and dispersed in solution (green) in the simulation systems containing 120 (filled bars) and 180 (open bars) SDS molecules. Error bars represent standard deviations. The average number of dispersed SDS corresponded to a concentration of about 7 mM. An SDS molecule was counted to be in direct contact with the protein if any SDS atom was within 3 Å of any protein atom. (e) Normalized distributions of titin I27's radius of gyration when the unfolded protein is in a complex with one (light blue), two (red), or three (orange) SDS micelles. The pie chart in the inset illustrates the percentage of time each number of micelles were bound. All systems analyzed in this figure were simulated at 373 K. For panels d and e, analysis was performed on the parts of the simulation trajectories featuring the protein in the unfolded state (after 2, 1, 1 and 3 μs for replicas 1, 2, 3 containing 120 SDS and the 180 SDS system, respectively).

and two, respectively. These values are in good overall agreement with the range (0.79–1.08) predicted from experiments performed under similar conditions.[81] In one of the room-temperature systems, the unfolded protein encircled the two micelles bound to it for the entire duration of the simulation, Fig. S13 and Movie S6.† In the second simulation, however, two of the three micelles initially bound to the protein merged into one micelle containing 80 SDS molecules, Fig. S14 and S15 and Movie S7.† The shape of the 80 SDS micelle was seen to deviate considerably from a spherical configuration, in agreement with the results of scattering experiments.[54] The simulation snapshots shown in Fig. 4a, b and ESI Fig. S13, S14† indicate that the protein does not cover the bound SDS micelles uniformly in our atomistic MD simulations, whereas previous SAXS measurements and coarse-grained MD simulations of a different protein/detergent system[82] suggested a close-to-uniform micelle coverage by the proteins.

The plots of the unfolded protein's R_g show that both a lower temperature and being bound to more micelles favors a more extended conformation of the protein, Fig. 4d. This behavior qualitatively matches the results of previous experimental studies that found larger SDS aggregates at lower temperatures.[77,83,84] The larger size of SDS-protein aggregates at room temperature correlates with the larger number of SDS molecules directly bound to the unfolded protein, Fig. 4e. Interestingly, decreasing from boiling to room temperature increases the average number of SDS molecules directly bound to a protein residue by about 0.1, Fig. 4e, while keeping the total number of SDS molecules bound the same, Fig. 4c. At both room and boiling temperature, the micelles bound to a

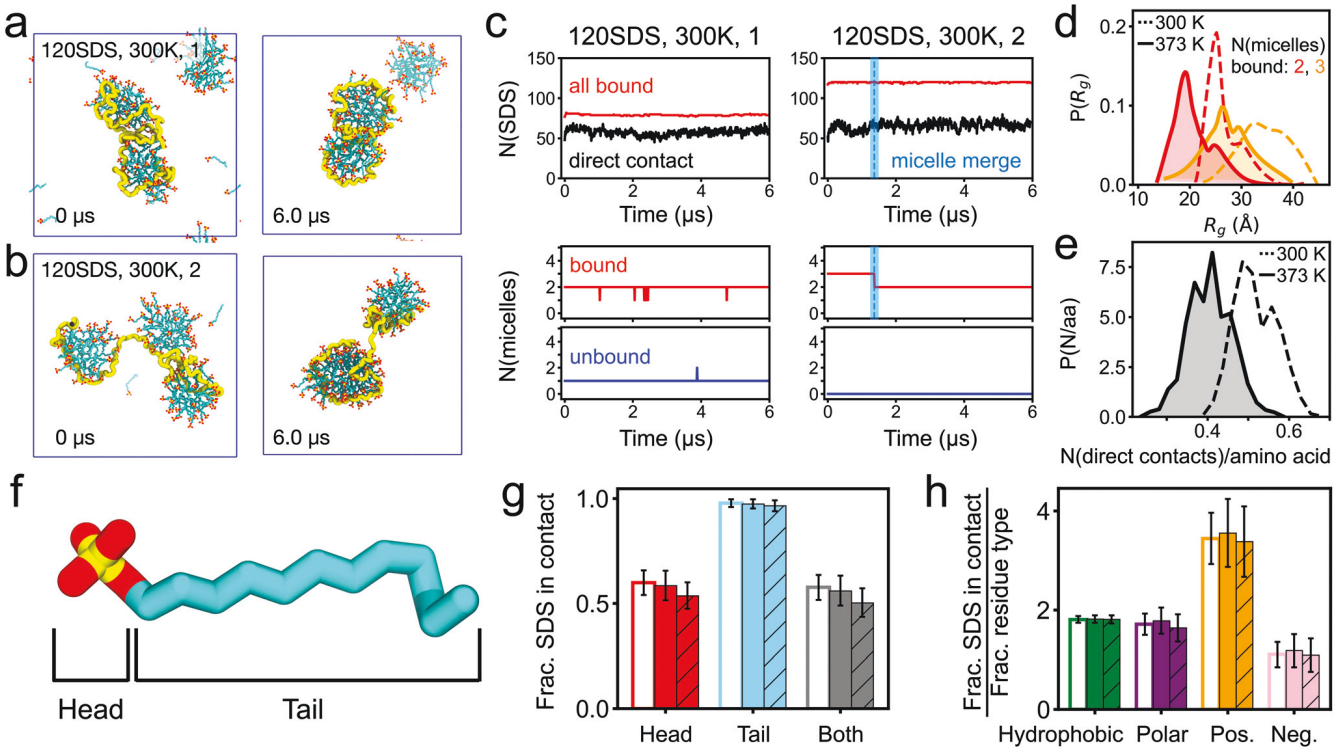

Fig. 4 The effects of temperature on protein–SDS interaction. (a, b) Initial and final configurations of two titin I27–SDS assemblies simulated at room temperature, 300 K. The I27 domain of titin is shown as a yellow tube, the 120 SDS molecules as red and cyan bonds; 400 mM NaCl solution is not shown, for clarity. (c) The number of SDS molecules bound to the protein (top) and the number of protein-bound and free-standing micelles (bottom) as a function of simulation time. The vertical blue rectangle highlights the moment the two protein-bound micelles merged into one. (d) Normalized distribution of the protein's R_g values for the states where the unfolded protein was found to form a complex with two (red) or three (orange) SDS micelles. Dashed and solid lines correspond to data obtained at 300 and 373 K. (e) Normalized distribution of the number of SDS molecules in direct contact with an amino acid of unfolded titin I27 at 300 and 373 K. (f) Schematic of an SDS molecule: a negatively charged SO_4^- head group is connected to a hydrophobic C_{12} tail. (g) The average fraction of SDS-protein contacts formed by the head (red), tail (blue) or both (gray) domains of SDS. Open, filled and hatch-filled bars correspond to the 120 SDS in 300 K; 120 SDS in 373 K; and 180 SDS in 373 K simulations. A contact was defined as having any SDS or SDS domain atom located within 3 Å of any protein atom. (h) The average fraction of direct protein–SDS contacts formed by the hydrophobic (green), polar (purple), positively-charged (orange) or negatively-charged (pink) residues divided by the fraction of the corresponding residues in titin I27. Bar styles match those in panel g. Error bars in panels g and h represent standard deviations.

protein were found to be slightly larger than in control simulations carried out in the absence of the protein, Fig. S12.† The concentrations of SDS dispersed in simulation, however dropped from about 7 mM at boiling temperature to about 3 mM at room temperature, regardless of the total SDS concentration or the presence of a protein. From the simulation trajectories, we also observe individual SDS molecules more readily dissociate and re-associate with micelles at boiling compared to room temperature.

Being a polar molecule, Fig. 4f, the local interactions of SDS with the protein can be expected to depend on the biophysical properties of the amino acids. Considering binding of the head and tail parts of SDS to the protein separately, we found almost all SDS–titin I27 contacts to include a tail–protein interaction, Fig. 4g, regardless of temperature or SDS concentration. The endpoint simulation snapshots, Fig. 4a and b, reveal that the unfolded protein wraps around the SDS micelles at the level of the head–tail interface. Interestingly, SDS molecules had a similar likelihood of forming direct contacts with hydrophobic and polar amino-acid residues, while they formed many more contacts with positively-charged residues, and significantly fewer contacts with negatively-charged residues, after taking the abundance of each residue type into account, Fig. 4h.

2.5. Conformational dynamics of the protein–SDS assembly

To characterize the dynamics of the protein–SDS assemblies after complete unfolding of the protein, we computed the auto-correlation function (ACF) of the protein end-to-end distance, Fig. 5a. The plots of ACF assessed how rapidly (or slowly) the protein conformation changes in time, with a steeper decaying ACF (a shorter relaxation time) indicating faster protein dynamics. For a random polymer coil, the characteristic end-to-end autocorrelation time τ is inversely proportional to the temperature.[85]

Our analysis reveals that the end-to-end relaxation of the unfolded protein in complex with SDS is not determined by the temperature alone and depends also on the specific con-

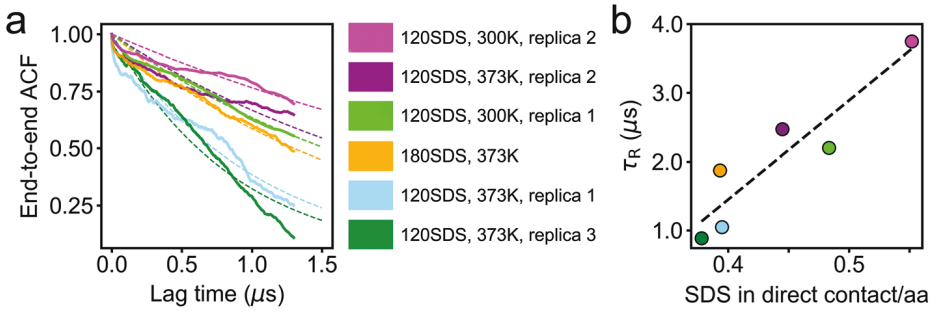

Fig. 5 Dynamics of unfolded protein–SDS assemblies. (a) Auto-correlation of the protein's end-to-end distance as a function of lag time. Dashed lines show an exponential decay fit, $\exp(-t/\tau_R)$, where t and τ_R are the lag and the relaxation time, respectively. Labels to the right indicate the number of SDS present, temperature and replica (if applicable). (b) End-to-end distance relaxation time, τ_R, as a function of trajectory-averaged number of SDS molecules forming direct contacts with an amino acid of the unfolded I27 domain of titin. The black dashed line shows a linear fit to the data.

figuration of the protein-SDS assembly. Thus, faster relaxation was observed in three out of four systems simulated at boiling temperature in comparison to the systems simulated at room temperature, Fig. 5a. The most pronounced exception is the second replica simulation of the I27 domain system carried out in the presence of 120 SDS at boiling temperature: its ACF closely follows the curves observed for unfolded titin I27 at room temperature, Fig. 5a. Among the three replicas of the 120 SDS systems simulated at 373 K, the second replica stands out by the furthest deviation from the initial, folded conformation, Fig. 1c, and the greatest average R_g, Fig. S7a.† Thus, the dynamics of individual protein-SDS assemblies is highly heterogeneous, and appears to be affected by the protein's configuration, with a more extended state showing slower dynamics.

Further analysis revealed a correlation between the end-to-end relaxation time τ_R and the average number of SDS molecules directly bound to the unfolded protein, Fig. 5b. A larger number of SDS molecules in direct contact with the unfolded protein generally corresponded to a longer relaxation time, and therefore slower conformational dynamics. Altogether, our results indicate that the dynamics of individual unfolded proteins bound to SDS is affected by both temperature and the protein conformation, with the dynamics of individual assemblies being correlated with the number of SDS molecules forming direct contacts with the protein.

2.6. Spontaneous unfolding of β-amylase by SDS

Thus far, we have investigated SDS–induced unfolding of the I27 domain of titin, whose folded state is made predominantly of β sheets. Here, we computationally investigate the effects of SDS on β-amylase, shown in Fig. 6a, the folded structure of which is predominantly α-helical.

Fig. 6b illustrates the initial state of the simulation system: a folded β-amylase structure[86] surrounded by 400 mM NaCl solution and 500 SDS monomeric molecules (110 mM) dispersed randomly through the system. The system was simulated at 1 bar pressure and boiling (373 K) temperature for 15 μs, Fig. 6b. As in the case of titin I27, β-amylase was observed to spontaneously unfold in the presence of SDS. The plots of the protein's RMSD, Fig. 6c, and R_g, Fig. S7d,† indicate gradual deterioration of the β-amylase structure over the first ten microseconds of the simulation. The system's trajectory, visualized in Fig. S16, S17 and Movie S8,† indicate that the unfolding process largely consisted of α helices spreading apart. Structural details about β-amylase are provided in ESI Fig. S18.†

Quantifying the unfolding process further, the plot of the β-amylase's Q value, Fig. 6d, displays two abrupt drops, and gradual decline thereafter. From about 2 to 3 μ, the protein can be considered to visit a molten-globular state characterized by locally preserved secondary structure but globally distorted overall configuration. The protein became unfolded ($Q < 0.1$) after about 12 μs, when the protein's central β barrel dissolved, Fig. S17 and S19.† Accordingly, the β-amylase's RMSD, Fig. 6c, and R_g, Fig. S7d,† increase considerably after 12 μs. Several native α helices, on the other hand, persisted throughout the trajectory, Fig. S17, S19 and Movie S8.† The 12 μs unfolding time is significantly longer than that observed for titin I27 (1–3 μs), which makes sense considering that β-amylase (498 residues) is significantly larger than titin I27 (119 residues). Fig. 6e and f plots the number of SDS molecules and micelles bound to β-amylase during the unfolding trajectory. At its highest, about 400 SDS molecules were bound to β-amylase, Fig. 6e, about four fifths times the number of amino acid residues (498), corresponding a binding ratio of about 1.9 g SDS per g Protein, within the range measured from simulations of titin I27 and experiments performed under similar conditions.[81] The number of SDS molecules directly in contact with β-amylase increased, however, to only about 200 during the simulation, Fig. 6e, corresponding to an average of about 0.4 SDS molecules bound to a single protein amino acid, matching the value for titin I27 at 373 K, Fig. 4e. In the unfolded state (*i.e.* after 12 μs), between four and nine micelles were bound to the protein, Fig. 6f, which is a significantly wider range than the one to three micelles bound to unfolded titin I27, Fig. 3b.

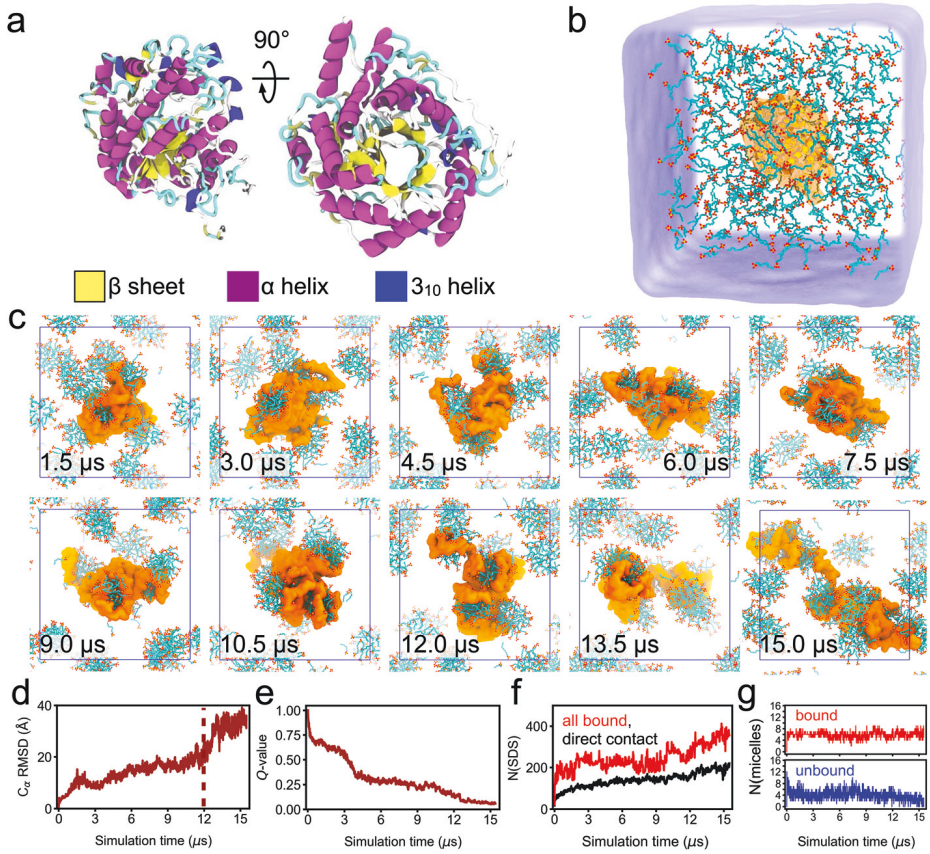

Fig. 6 Spontaneous unfolding of β-amylase by SDS. (a) Cartoon representation of the β-amylase, secondary-structural elements distinguished by color. (b) Simulation setup: a β-amylase protein in its folded conformation (orange molecular surface) is surrounded by 500 randomly-placed SDS molecules (red and cyan bonds) and 0.4 M NaCl aqueous solution (semi-transparent blue) at 373 K. (c) Snapshots illustrating the time evolution of the protein–SDS complex over 15.0 μs. Blue squares denote the system's dimensions. Periodic images of some SDS molecules are shown to highlight micelle formation. (d) Protein RMSD with respect to the initial, folded structure. (e) Fraction of native contacts Q *versus* simulation time. The Q value calculations were done only for residues that were identified as α helical or β sheet or bridge in the folded structure. RMSD and Q were calculated based on Cα positions. The vertical dashed line indicates the moment the protein unfolds completely (*i.e.* $Q < 0.1$). (f) The number of SDS molecules in direct contact with β-amylase (black), and of all SDS molecules bound to the protein (directly or *via* SDS-mediated contacts, red) *versus* simulation time. (g) The number of protein-associated (red) and free-standing (blue) micelles *versus* simulation time.

3. Conclusions

Here, we reported the molecular mechanisms of spontaneous protein unfolding induced by SDS. Previously, all-atom MD simulations[25,26,87–89] and experiment[26,87] examined titin I27 unfolding induced by a mechanical force, finding the unfolding process to follow a largely deterministic pathway in which β-strand pairs A/B and A′/G rupture first, and then a sequence of β strand pairs break apart moving from the protein termini inward. In our simulation of SDS–induced unfolding, the first β-strand pair of titin I27 to break varied from one simulation to another whereas the remaining β strands could either abruptly dissolve or remain structured for a significant amount of time after the first β-strand pair breakage. A similar two-stage unfolding mechanism was observed in an MD simulation study[30] of urea-induced protein unfolding, where urea molecules were found to first penetrate the protein's hydrophobic core and then flood into the protein's interior, disrupting protein–protein hydrogen bonds. In contrast to previous studies that found detergent-induced unfolding to occur at the millisecond to second time scale[33,70] at low millimolar SDS concentrations and near room temperature, we found SDS unfolding of protein to occur much faster (in tens of microseconds) at the conditions realized by SDS-PAGE sample preparation.

Equilibrium dialysis experiments[38] performed on a range of proteins at low SDS concentration (2–4 mM) determined a remarkably constant ratio of the mass of SDS molecules bound to a protein to the mass of the protein (1.4 to 1), corresponding to the commonly referenced stoichiometry of "one SDS molecule per two amino acids". Follow-up experiments found that the mass ratio could vary between 1.2 and 1.5 [90] and depend on the SDS concentration.[81] In our simulations, the average number of SDS molecules forming direct contacts with unfolded titin I27 at room temperature was 0.52 ± 0.06 per amino acid, Fig. 4e, which matches the experimentally-derived number at low SDS concentration.[38,90] When we considered all SDS molecules bound to the protein, either directly or through

micelle formation, the binding ratio observed in our simulations falls within the range of 0.6 to 1.0 SDS molecule per amino acid, or, equivalently, 1.4 to 2.4 grams of SDS per gram of protein. This range overlaps with the mass ratio range measured in experiment at comparable SDS concentration (1.9 to 2.6 grams of SDS per gram of protein).[81] The variability observed in the simulations results from stochastic fluctuations in the number of SDS micelles bound to the protein, which could change on the microsecond time scale. Overall, our simulations provide a direct support to the fluid necklace-and-beads model previously suggested by scattering[41,91] and spectroscopy[31] measurements.

According to our simulations, the number of SDS micelles bound to a protein determines both the overall conformation and the relaxation kinetics of the protein-SDS assembly. A larger number of bound micelles favors a more extended conformation of the protein-SDS assembly, an effect that becomes even more pronounced for larger molecular weight proteins (compare Fig. 6b to Fig. 1b). We hypothesize that, subject to a hydrodynamic flow, a tethered protein-SDS assembly may unravel to adopt an extended beads-on-a-string conformation. Driving such a linearized SDS-protein assembly through a nanopore large enough to admit SDS micelles may enable single-molecule protein characterization and fingerprinting.[36,69,92,93] In such measurements, blockades of nanopore ionic current,[36,69] or fluorescence of labeled amino acids[92,93] provide a sequence of signals that can be used to determine the molecular weight of or identify the passing protein.

4. Methods

4.1. General MD protocols

All MD simulations were performed using the CHARMM36 force field[94] for proteins and sodium dodecyl sulfate (SDS), the TIP3P water model,[95] and custom CUFIX corrections for non-bonded interactions between charged groups.[96] Periodic boundary conditions were employed in all simulations, and long-range electrostatics were calculated using the Particle Mesh Ewald[97] method over a 0.12 nm spaced grid. Non-bonded Coulomb and Lennard-Jones interactions were truncated at 1.2 nm, and the non-bonded list was updated every 10 steps. Covalent bonds to hydrogen in water and in non-water molecules were constrained using SETTLE[98] and LINCS[99] algorithms, respectively. Systems were minimized using steepest descent for 10 000 steps followed by a 1 ns equilibration simulation during which harmonic restraints were applied to the protein backbone (k = 1000 kcal mol^{-1} nm^{-2}). Minimization, restrained equilibration, and initial unrestrained production MD simulations of at least 100 ns were performed using the gromacs 4.5.5 package,[100] with a 2 fs integration timestep in the constant number of particles N, pressure P, and temperature T ensemble; specified temperature maintained using the Nosé–Hoover thermostat,[101,102] pressure of 1.0 bar maintained using the Parrinello–Rahman barostat.[103] Production runs in the NPT ensemble were completed on the D. E. Shaw Research Anton 2 supercomputer.[104] The simulations on Anton 2 employed a set of parameters equivalent to those listed above, except for using the Martyna–Tobias–Klein barostat[105] and the k-space Gaussian split Ewald method[106] to calculate the electrostatic interactions.

4.2. MD simulation of titin I27

The initial all-atom model of titin I27 was built by combining the crystal structure of the folded titin I27 domain[107] (PDB ID: 1TIT) with the disordered His (MRGSHHHHHHGLVPRGS) and Ssr-A (RSAANDENYALAA) peptide tags added to the N- and C-termini of the protein, respectively,[69] a construct containing 119 amino acids. 120 or 180 SDS molecules were placed randomly within a cubic volume measuring 12 nm on each side (0.11 and 0.17 M SDS concentration, respectively), ensuring no SDS molecules overlapped with the protein. The volume was filled with pre-equilibrated water molecules using VMD.[108] The systems were then neutralized, and ions were added to achieve a 0.4 M concentration of NaCl. Energy-minimization, restrained equilibration, and production MD followed, as described in the previous paragraph. Five independent simulations were performed at 373 K of the folded titin I27 structure: three replica simulations in the presence of 120 SDS molecules, one simulation with 180 SDS molecules and one simulation with no SDS molecules added. Simulation configurations from two 373 K simulation trajectories containing 120 SDS (the first replica at 6.0 μs and the second replica at 2.5 μs) were chosen to initiate simulations of room (300 K) temperature that lasted 6.0 μs. Prior to room temperature simulation, the systems were cooled down from 373 to 300 K over the course of 50 ns. Table 1 summarizes all MD simulations reported in this study.

4.3. MD simulations of β-amylase

The sole β-amylase system was prepared starting from its crystal structure[86] (PDB ID: 1FA2). The *psfgen* tool was used to add any missing atoms to this 498-residue protein. Five hundred SDS molecules were randomly added to a cubic volume measuring 14 nm on each side (0.11 M SDS concentration), ensuring that no SDS molecules overlapped with the protein. The resulting system was solvated and ions were added to produce a 0.4 M NaCl solution. The energy-minimization, restrained equilibration, and production MD was done following the same protocols as for the titin I27 systems. The production simulation of β-amylase lasted 15.0 μs and was carried out in the NPT ensemble at 373 K and 1.0 bar.

4.4. MD simulations of SDS-water mixtures

Systems were also prepared consisting only of SDS molecules, 0.4 M NaCl, and water. 120 or 180 SDS molecules were placed randomly within a cubic volume measuring 12 nm on each side (0.110 and 0.165 M SDS concentration, respectively), and the remaining volumes were filled with water, and then the ions needed to reach a neutral 0.4 M NaCl solution. Nine

Table 1 Summary of performed simulations

System name	N(SDS)	NaCl (M)	SDS (M)	Time (μs)
Titin I27, 120 SDS, 373 K, replica 1	120	0.4	0.11	6.0
Titin I27, 120 SDS, 373 K, replica 2	120	0.4	0.11	6.0
Titin I27, 120 SDS, 373 K, replica 3	120	0.4	0.11	6.0
Titin I27, 180 SDS, 373 K	180	0.4	0.17	6.0
Titin I27, 0 SDS, 373 K	0	0.4	0.0	6.0
Titin I27, 120 SDS, 300 K, replica 1	120	0.4	0.11	6.0
Titin I27, 120 SDS, 300 K, replica 2	120	0.4	0.11	6.0
β-Amylase, 500 SDS, 373 K	500	0.4	0.11	15.0
120 SDS, 300 K, replicas 1, 2, 3	120	0.4	0.11	4.5
120 SDS, 373 K, replicas 1, 2, 3	120	0.4	0.11	4.5
180 SDS, 373 K, replicas 1, 2, 3	180	0.4	0.17	4.5

systems in total: three replicas with 120 SDS molecules at 1.0 bar and 300 K; three replicas with 120 SDS molecules at 1.0 bar and 373 K; and three replicas with 180 SDS molecules at 1.0 bar and 373 K. Within each condition, the replicas differed only by the random seed used to generate initial atomic velocities. The simulations were performed for 4.5 μs each, and only simulation snapshots after 3 μs were considered for the purpose of analysis.

4.5. Analysis

To quantify protein unfolding, we measured the fraction of native contacts Q:

$$Q = \frac{1}{N_{\text{pairs}}} \sum_{i<j-2} \exp[-((r_{ij} - r_{ij}^{\text{native}})^2/2\sigma_{\text{ij}}{}^2)], \quad (1)$$

where N_{pairs} is the number of pairs of residues considered, r_{ij} is the instantaneous distance between the Cα atoms of residues i and j, r_{ij}^{native} is the same distance in the experimentally determined structure, and $\sigma_{ij} = (1 + |i - j|)^{0.15}$.

When protein secondary structure is shown, it was determined for each conformation separately using STRIDE[109] in VMD.[108]

An SDS molecule was considered to be in direct contact with the protein if any SDS atom was within 3 Å of any protein atom. Layers of SDS molecules surrounding the protein (*i.e.* the first layer being SDS in direct contact with the protein, the second layer consisting of SDS molecules in contact with the first SDS layer, and so forth) were found for each simulation snapshot to determine the total number of SDS bound to the protein.

To identify SDS micelles, we used the DBSCAN[110] clustering algorithm, using the positions of the C6 atoms of SDS molecules, counting a pair of C6 atoms within 9.5 Å to be in the same cluster, and requiring a cluster to have a minimum of three SDS molecules.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported *via* grants from the National Science Foundation (PHY-1430124, DMR-1827346) and the National Institutes of Health (P41-GM104601). The supercomputer time was provided through the Extreme Science and Engineering Discovery Environment (XSEDE) allocation MCA05S028, the Blue Waters petascale supercomputer at the University of Illinois, and Anton 2 allocation PSCA00052. Anton 2 computer time was provided by the Pittsburgh Supercomputing Center through grant R01-GM116961 from the National Institutes of Health. The Anton 2 machine was made available by D. E. Shaw Research.

References

1 M.-J. Gething and J. Sambrook, *Nature*, 1992, **355**, 33–45.
2 C. M. Dobson, *Nature*, 2003, **426**, 884–890.
3 J. Martin, T. Langer, R. Boteva, A. Schramel, A. L. Horwich and F.-U. Hartl, *Nature*, 1991, **352**, 36.
4 D. J. Schnell and D. N. Hebert, *Cell*, 2003, **112**, 491–505.
5 M. Bucciantini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson and M. Stefani, *Nature*, 2002, **416**, 507.
6 S. Wang and R. J. Kaufman, *J. Cell Biol.*, 2012, **197**, 857–867.
7 C. Levinthal, *J. Chim. Phys. Phys.-Chim. Biol.*, 1968, **65**, 44.
8 K. A. Dill, *Biochemistry*, 1990, **29**, 7133–7155.
9 J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes, *Annu. Rev. Phys. Chem.*, 1997, **48**, 545–600.
10 N. V. Dokholyan, L. Li, F. Ding and E. I. Shakhnovich, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 8637–8641.
11 P. L. Freddolino, F. Liu, M. Gruebele and K. Schulten, *Biophys. J.*, 2008, **94**, L75–L77.
12 K. Lindorff-Larsen, S. Piana, R. O. Dror and D. E. Shaw, *Science*, 2011, **334**, 517–520.
13 P. I. Zhuravlev and G. A. Papoian, *Q. Rev. Biophys.*, 2010, **43**, 295–332.
14 A. T. Van Wart, J. Durrant, L. Votapka and R. E. Amaro, *J. Chem. Theory Comput.*, 2014, **10**, 511–517.

15 J. R. Lewandowski, M. E. Halse, M. Blackledge and L. Emsley, *Science*, 2015, **348**, 578–581.
16 D. Winogradoff, H. Zhao, Y. Dalal and G. A. Papoian, *Sci. Rep.*, 2015, **5**, 17038.
17 T. E. Creighton, *Prog. Biophys. Mol. Biol.*, 1979, **33**, 231–297.
18 V. Daggett and M. Levitt, *J. Mol. Biol.*, 1993, **232**, 600–619.
19 R. L. Baldwin, *Proc. Natl. Acad. Sci. U. S. A.*, 1986, **83**, 8069–8072.
20 J. Sabelko, J. Ervin and M. Gruebele, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 6031–6036.
21 V. S. Pande and D. S. Rokhsar, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9062–9067.
22 Y. GoTo, L. J. Calciano and A. L. Fink, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 573–577.
23 W. Gu, T. Wang, J. Zhu, Y. Shi and H. Liu, *Biol. Chem.*, 2003, **104**, 79–94.
24 C. Dumont, T. Emilsson and M. Gruebele, *Nat. Methods*, 2009, **6**, 515–519.
25 H. Lu and K. Schulten, *Proteins: Struct., Funct., Bioinf.*, 1999, **35**, 453–463.
26 P. E. Marszalek, H. Lu, H. Li, M. Carrion-Vazquez, A. F. Oberhauser, K. Schulten and J. M. Fernandez, *Nature*, 1999, **402**, 100.
27 D. K. Klimov and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 7254–7259.
28 C. Tanford, *J. Am. Chem. Soc.*, 1964, **86**, 2050–2059.
29 B. J. Bennion and V. Daggett, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 5142–5147.
30 L. Hua, R. Zhou, D. Thirumalai and B. Berne, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 16928–16933.
31 N. J. Turro, X.-G. Lei, K. Ananthapadmanabhan and M. Aronson, *Langmuir*, 1995, **11**, 2525–2533.
32 K. Weber and M. Osborn, *J. Biol. Chem.*, 1969, **244**, 4406–4412.
33 M. M. Nielsen, K. K. Andersen, P. Westh and D. E. Otzen, *Biophys. J.*, 2007, **92**, 3674–3685.
34 S. F. d. S. Groth, R. Webster and A. Datyner, *Biochim. Biophys. Acta*, 1963, **71**, 377–391.
35 W. N. Burnette, *Anal. Biochem.*, 1981, **112**, 195–203.
36 E. Kennedy, Z. Dong, C. Tennant and G. Timp, *Nat. Nanotechnol.*, 2016, **11**, 968–976.
37 L. Restrepo-Pérez, C. Joo and C. Dekker, *Nat. Nanotechnol.*, 2018, **13**, 786–796.
38 J. A. Reynolds and C. Tanford, *Proc. Natl. Acad. Sci. U. S. A.*, 1970, **66**, 1002–1007.
39 M. N. Jones and A. Wilkinson, *Biochem. J.*, 1976, **153**, 713–718.
40 S.-H. Chen and J. Teixeira, *Phys. Rev. Lett.*, 1986, **57**, 2583.
41 X. Guo, N. Zhao, S. Chen and J. Teixeira, *Biopolymers*, 1990, **29**, 335–346.
42 K. K. Andersen, C. L. Oliveira, K. L. Larsen, F. M. Poulsen, T. H. Callisen, P. Westh, J. S. Pedersen and D. Otzen, *J. Mol. Biol.*, 2009, **391**, 207–226.
43 D. Saha, D. Ray, J. Kohlbrecher and V. K. Aswal, *ACS Omega*, 2018, **3**, 8260–8270.
44 H. Ø. Rasmussen, J. J. Enghild, D. E. Otzen and J. S. Pedersen, *Biochim. Biophys. Acta, Gen. Subj.*, 2020, **1864**, 129434.
45 W. Parker and P.-S. Song, *Biophys. J.*, 1992, **61**, 1435–1439.
46 T. K. Das, S. Mazumdar and S. Mitra, *Eur. J. Biochem.*, 1998, **254**, 662–670.
47 E. Gelamo, C. Silva, H. Imasato and M. Tabak, *Biochim. Biophys. Acta*, 2002, **1594**, 84–99.
48 J. A. Reynolds and C. Tanford, *J. Biol. Chem.*, 1970, **245**, 5161–5165.
49 K. Tsujii and T. Takagi, *J. Biochem.*, 1975, **77**, 511–519.
50 P. Lundahl, E. Greijer, M. Sandberg, S. Cardell and K.-O. Eriksson, *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol.*, 1986, **873**, 20–26.
51 A. Dutta, T.-Y. Kim, M. Moeller, J. Wu, U. Alexiev and J. Klein-Seetharaman, *Biochemistry*, 2010, **49**, 6329–6340.
52 Y. Sun, P. L. O. Filho, J. C. Bozelli, J. Carvalho, S. Schreier and C. L. Oliveira, *Soft Matter*, 2015, **11**, 7769–7777.
53 J. D. Kaspersen, A. Søndergaard, D. J. Madsen, D. E. Otzen and J. S. Pedersen, *Biophys. J.*, 2017, **112**, 1609–1620.
54 H. G. Mortensen, J. K. Madsen, K. K. Andersen, T. Vosegaard, G. R. Deen, D. E. Otzen and J. S. Pedersen, *Biophys. J.*, 2017, **113**, 2621–2633.
55 Y. Sun, P. L. O. Filho and C. L. Oliveira, *Food Hydrocolloids*, 2017, **62**, 10–20.
56 S. J. Marrink, D. P. T. Tieleman and A. E. Mark, *J. Phys. Chem. B*, 2000, **104**, 12165–12173.
57 M. Sammalkorpi, M. Karttunen and M. H. Haataja, *J. Phys. Chem. B*, 2007, **111**, 11722–11733.
58 S. Kawada, M. Komori, K. Fujimoto, N. Yoshii and S. Okazaki, *Chem. Phys. Lett.*, 2016, **646**, 36–40.
59 S. Balasubramanian, S. Pal and B. Bagchi, *Phys. Rev. Lett.*, 2002, **89**, 115505.
60 S. Pal, S. Balasubramanian and B. Bagchi, *J. Chem. Phys.*, 2002, **117**, 2852–2859.
61 R. Braun, D. M. Engelman and K. Schulten, *Biophys. J.*, 2004, **87**, 754–763.
62 V. Krishnamani and J. K. Lanyi, *Biochemistry*, 2012, **51**, 1061–1069.
63 A. Berthaud, J. Manzi, J. Pérezz and S. Mangenot, *J. Am. Chem. Soc.*, 2012, **134**, 10080–10088.
64 M. Zoonens, J. Comer, S. Masscheleyn, E. Pebay-Peyroula, C. Chipot, B. Miroux and F. Dehez, *J. Am. Chem. Soc.*, 2013, **135**, 15174–15182.
65 J. D. Kaspersen, C. M. Jessen, B. S. Vad, E. S. Sørensen, K. K. Andersen, M. Glasius, C. L. P. Oliveira, D. E. Otzen and J. S. Pedersen, *ChemBioChem*, 2014, **15**, 2113–2124.
66 A. J. Wolfe, W. Si, Z. Zhang, A. R. Blanden, Y.-C. Hsueh, J. F. Gugel, B. Pham, M. Chen, S. N. Loh, S. Rozovsky, A. Aksimentiev and L. Movileanu, *J. Phys. Chem. B*, 2017, **121**, 10228–10241.
67 H. Dominguez, *J. Mol. Model.*, 2017, **23**, 210.
68 G. Roussel, Y. Caudano, A. Matagne, M. S. Sansom, E. A. Perpète and C. Michaux, *Spectrochim. Acta, Part A*, 2018, **190**, 464–470.

69 L. Restrepo-Pérez, S. John, A. Aksimentiev, C. Joo and C. Dekker, *Nanoscale*, 2017, **9**, 11685–11693.
70 D. E. Otzen and M. Oliveberg, *J. Mol. Biol.*, 2002, **315**, 1231–1240.
71 M. Eastwood, C. Hardin, Z. Luthey-Schulten and P. Wolynes, *IBM J. Res. Dev.*, 2001, **45**, 475–497.
72 E. L. Duggan and J. M. Luck, *J. Biol. Chem.*, 1948, **172**, 205–220.
73 Y. Moriyama, Y. Sato and K. Takeda, *J. Colloid Interface Sci.*, 1993, **156**, 420–424.
74 Y. Moriyama, N. Kondo and K. Takeda, *Langmuir*, 2012, **28**, 16268–16273.
75 S. Mukherjee, P. Sen, A. Halder, S. Sen, P. Dutta and K. Bhattacharyya, *Chem. Phys. Lett.*, 2003, **379**, 471–478.
76 A. H. Poghosyan, N. P. Schafer, J. Lyngsø, A. A. Shahinyan, J. S. Pedersen and D. E. Otzen, *Protein Eng., Des. Sel.*, 2019, **32**(4), 175–190.
77 N. J. Turro and A. Yekta, *J. Am. Chem. Soc.*, 1978, **100**, 5951–5952.
78 B. Kronberg and B. Lindman, *Surfactants and polymers in aqueous solution*, John Wiley & Sons Ltd, Chichester, 2003.
79 G. V. Jensen, R. Lund, J. Gummel, T. Narayanan and J. S. Pedersen, *Angew. Chem., Int. Ed.*, 2014, **53**, 11524–11528.
80 C. Højgaard, H. V. Sørensen, J. S. Pedersen, J. R. Winther and D. E. Otzen, *Biophys. J.*, 2018, **115**, 2081–2086.
81 C. A. Nelson, *J. Biol. Chem.*, 1971, **246**, 3895–3901.
82 J. N. Pedersen, P. W. J. M. Frederix, J. S. Pedersen, S. J. Marrink and D. E. Otzen, *ChemBioChem*, 2018, **19**, 263–271.
83 J. Hayter and J. Penfold, *Colloid Polym. Sci.*, 1983, **261**, 1022–1030.
84 A. Malliaris, J. Le Moigne, J. Sturm and R. Zana, *J. Phys. Chem.*, 1985, **89**, 2709–2713.
85 P.-G. De Gennes, *Scaling concepts in polymer physics*, Cornell University Press, 1979.
86 C. G. Cheong, S. H. Eom, C. Chang, D. H. Shin, H. K. Song, K. Min, J. H. Moon, K. K. Kim, K. Y. Hwang and S. W. Suh, *Proteins: Struct., Funct., Bioinf.Proteins: Struct., Funct., Genet.*, 1995, **21**, 105–117.
87 H. Lu, B. Isralewitz, A. Krammer, V. Vogel and K. Schulten, *Biophys. J.*, 1998, **75**, 662–671.
88 H. Lu and K. Schulten, *Biophys. J.*, 2000, **79**, 51–65.
89 M. Gao, M. Wilmanns and K. Schulten, *Biophys. J.*, 2002, **83**, 3435–3445.
90 T. Takagi, K. Tsujii and K. Shirahama, *J. Biochem.*, 1975, **77**, 939–947.
91 J. M. Chen, T. M. Su and C. Y. Mou, *J. Phys. Chem.*, 1986, **90**, 2418–2421.
92 J. van Ginkel, M. Filius, M. Szczepaniak, P. Tulinski, A. S. Meyer and C. Joo, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 3338–3343.
93 S. Ohayon, A. Girsault, M. Nasser, S. Shen-Orr and A. Meller, *PLoS Comput. Biol.*, 2019, **15**, e1007067.
94 A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
95 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
96 J. Yoo and A. Aksimentiev, *Phys. Chem. Chem. Phys.*, 2018, **20**, 8432–8449.
97 T. A. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
98 S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.
99 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
100 S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.
101 S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.
102 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
103 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
104 D. E. Shaw, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. P. Grossman, C. R. Ho, D. J. Lerardi, I. Kolossváry, J. L. Klepeis, T. Layman, C. McLeavey, M. M. Deneroff, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, S. C. Wang, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson and K. J. Bowers, *Commun. ACM*, 2008, **51**, 91.
105 G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177–4189.
106 Y. Shan, J. L. Klepeis, M. P. Eastwood, R. O. Dror and D. E. Shaw, *J. Chem. Phys.*, 2005, **122**, 054101.
107 S. Improta, A. S. Politou and A. Pastore, *Structure*, 1996, **4**, 323–337.
108 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, 1996, **14**, 33–38.
109 M. Heinig and D. Frishman, *Nucleic Acids Res.*, 2004, **32**, W500–W502.
110 M. Ester, H.-P. Kriegel, J. Sander and X. Xu, *Kdd*, 1996, pp. 226–231.

