Rotational dynamics and transition mechanisms of surface-adsorbed proteins

Shuai Zhang[a,b,1] [iD], Robbie Sadre[c,1], Benjamin A. Legg[a,b,1], Harley Pyles[d,e], Talita Perciano[c] [iD], E. Wes Bethel[c,f] [iD], David Baker[d,e,g] [iD], Oliver Rübel[c,2], and James J. De Yoreo[a,b,2]

Assembly of biomolecules at solid–water interfaces requires molecules to traverse complex orientation-dependent energy landscapes through processes that are poorly understood, largely due to the dearth of in situ single-molecule measurements and statistical analyses of the rotational dynamics that define directional selection. Emerging capabilities in high-speed atomic force microscopy and machine learning have allowed us to directly determine the orientational energy landscape and observe and quantify the rotational dynamics for protein nanorods on the surface of muscovite mica under a variety of conditions. Comparisons with kinetic Monte Carlo simulations show that the transition rates between adjacent orientation-specific energetic minima can largely be understood through traditional models of in-plane Brownian rotation across a biased energy landscape, with resulting transition rates that are exponential in the energy barriers between states. However, transitions between more distant angular states are decoupled from barrier height, with jump-size distributions showing a power law decay that is characteristic of a nonclassical Levy-flight random walk, indicating that large jumps are enabled by alternative modes of motion via activated states. The findings provide insights into the dynamics of biomolecules at solid–liquid interfaces that lead to self-assembly, epitaxial matching, and other orientationally anisotropic outcomes and define a general procedure for exploring such dynamics with implications for hybrid biomolecular–inorganic materials design.

orientational energy landscapes | rotational dynamics of protein | high-speed atomic force microscopy | machine learning | Levy-flight transition

Living systems create a wide range of extended biomolecular arrays with complex functions that are inherently connected to array architecture and often comprise interfaces between disparate materials, as with S-layer cell envelopes (1), bacterial microcompartments (2, 3), and other vesicles (4), and insoluble matrices that direct formation of mineralized tissues (5–7). Inspired by such systems, numerous synthetic biomolecular systems have been developed, many of which assemble through epitaxial relationships with mineral surfaces (8–15). Both the dynamics and final architectures are determined by the energy landscape arising from an interplay between protein–protein and protein–substrate interactions (16, 17), which often exhibit a degree of rotational symmetry that creates multiple energy minima separated by small barriers. While living systems use such multiwell landscapes to alter array architecture in response to environmental conditions (18), similar capabilities in artificially engineered biomolecular arrays have rarely been realized, in part, because the energy landscapes remain largely unexplored. In particular, the dynamics and mechanisms by which biomolecules at inorganic surfaces undergo transitions between metastable states in the energy landscape to reach the final stable state, as well as the response of that landscape to environmental stimuli, are unknown. This lack of knowledge presents a fundamental obstacle to the design and application of artificial hybrid biomaterials.

In this work, we used in situ high-speed atomic force microscopy (HS-AFM) (19) and machine learning (ML)–based analysis to investigate the energy landscapes and interminima transitions associated with in-plane rotation of the de novo designed repeat protein, DHR10-micaN (MicaN; N denotes the number of repeat subunits per molecule), on the muscovite mica (m-mica) (001) surface under a variety of conditions. As these molecules each display preferred orientations with the m-mica substrate, they provide the opportunity to test how molecules transition between preferred states.

The predominant paradigm for understanding nanoscale biomolecular motion is derived from the theory of Brownian motion developed by Einstein (20) and Smoluchowski (21). If this theory holds, transitions rates between angular states could be predicted through numerical implementation of Brownian drift–diffusion

Significance

The exquisite organization exhibited by hybrid biomolecular–inorganic materials in nature has inspired the development of synthetic analogues for numerous applications. Nevertheless, a mechanistic picture of the energetic controls and response dynamics leading to organization is lacking. Here, we pair high-speed atomic force microscopy with machine learning and Monte Carlo simulations to analyze the rotational dynamics of rod-like proteins on a crystal lattice, simultaneously quantifying the orientational energy landscape and transition probabilities between energetically favorable orientations. Although rotations largely follow Brownian diffusion, proteins often make large jumps in orientation, thus rapidly overcoming barriers that usually inhibit rotation. Moreover, the rotational dynamics can be tuned via protein size and solution chemistry, providing tools for controlling biomolecular assembly at inorganic interfaces.

equations. However, studies on nanoparticle transitional motion are increasingly pointing toward the possibility of nonclassical "superdiffusional" processes involving intermittent large jumps or Levy-flight jump-size distributions (22–27). If such processes also occur during biased angular motion, they may have important effects on the transition rates between angular states. Thus, in this study, we utilize kinetic Monte Carlo (kMC) simulations to compare our state-of-the-art experimental observations with classical expectations. These comparisons show that transitions between neighboring angular states can be largely reproduced through classical models of Brownian motion, with transition rates that are exponential in the energy barrier between states in both cases, as expected. However, transitions between nonadjacent minima are more frequent than expected and show a power law decay in jump size that is indicative of a nonclassical Levy-flight process. We suggest that these large jumps may be facilitated by enhanced rotation via a spectrum of quasithree-dimensional motions involving transient desorption from the substrate. Our findings provide fundamental insights into the dynamics and energetics of biomolecules at liquid–solid interfaces, confirming that nonclassical diffusion processes, like Levy-flight motion, occur in soft matter systems, and they suggest approaches to biomaterials design that exploit nonclassical motion as a means for biomolecules to rapidly bypass high-energy barriers and reach stable minima.

Results

Impact of Rod Length and Cation Type on Orientation. MicaN has a programmed binding surface on which glutamate side chains are geometrically matched to the K^+ sublattice of m-mica (001) (Fig. 1A) (15). Hence, the minimum energy orientation was expected to occur when the long axis is aligned parallel to one of the close-packed directions of the m-mica (001) lattice ([100], [110], or [1$\bar{1}$0]), which are separated by 60° intervals. Based on the protein design, these three directions were expected to be equally favorable, but atomic force microscopy (AFM) imaging showed that one direction was generally preferred (15). This reduced symmetry reflects complex interactions between the protein and the mica surface, which arise from structural features in the underlying crystal lattice that distort its symmetry (28). This asymmetry provides an opportunity to quantify both the relative depths and intervening barriers of the minima, as well as the rates and mechanisms of transitions between them.

In situ HS-AFM data clearly show that, for Mica18 and Mica34 on m-mica (001) in 10 mM [K^+] solution (Mica18@KCl10mM and Mica34@KCl10mM, respectively), most proteins share the same preferred orientation, and proteins tend to retain that general orientation for extended periods of time (Fig. 1 B and C and Movies S1 and S2). We consider only a 180° rotational space as opposed to a full 360° rotation because the N-terminal and C-terminal ends of the protein nanorod are indistinguishable in the AFM images. However, because the pattern of contact between mica-lattice sites and glutamate side chains is similar upon 180° rotations (15), we anticipate that rotations of 180° will produce a configuration with similar energy. Orientation tracking of selected Mica18 nanorods (Fig. 1 F and G) further illustrates that, while the protein nanorods have one thermodynamically preferred orientation, they occasionally transit to other higher-energy orientations with relatively long lifetimes. Mica6 (Mica6@KCl10mM), which has fewer glutamate residues per molecule and a lower binding affinity to the substrate, is more mobile on

m-mica (001), with the consequence that the dominant orientations are less visually distinguishable (Fig. 1D and Movie S3).

When Na^+ is used instead of K^+, the surficial K^+ ions of mica are replaced by Na^+ (29). Unlike K^+, Na^+ has been found to inhibit self-assembly of proteins on m-mica (001) (10, 15). HS-AFM data show that Mica18@NaCl10mM has much faster in-plane dynamics and much lower coverage on m-mica (001). Moreover, whether a dominant orientation even exists is unclear without careful statistical analysis of distributions over long time periods (Fig. 1E and Movie S4).

Orientation-Dependent Energy Landscapes. The HS-AFM data record details of angular distributions and rotational dynamics for MicaN on m-mica (001). However, the amount of data generated (hundreds of frames in each condition with tens of protein nanorods in each frame), as with other big datasets, creates a severe obstacle to quantification with traditional data-processing approaches (30, 31). Hence, we developed an ML approach to track and quantify the distributions and dynamics. Using the segmentation routine from U-NET (32), the orientational distributions of MicaN (n = 34, 18, 6) across the time span of the experiments (Fig. 2, Movie S5, and *SI Appendix*, Fig. S1) were extracted. For all values of N and both K^+ and Na^+, the orientational distributions of protein nanorods are stable throughout the experiments (Fig. 2 A, *Lower*, B, *Lower*, C, *Lower*, and D, *Lower*), indicating they reflect equilibrium distributions (histograms in Fig. 2), from which the orientation-dependent energy landscape $G(\theta)$ can be determined (curves in Fig. 2). The resulting landscapes all show minima at several angles with intervening barriers on the order of 1 to 6 k_BT.

For the case with K^+, the expectation that energy minima should occur when the long axis of the nanorod is aligned parallel to one of the K^+ close-packed directions of m-mica (001) or possibly at an intermediate angle close to 30°, which allows a high ratio of carboxylate side chains of glutamates to simultaneously match with the K^+ sublattice, is partially confirmed. With Mica18@KCl10mM (Fig. 2A), we observe one primary minimum at around −20° (defined arbitrarily), a secondary minimum at around −80° (−60° relative to the primary minimum), and a third at around 70°/−110° (90° relative to the primary minimum). However, the lack of a stable minima at 40° (+60° relative to the primary minimum) suggests that interactions with the mica surface involve more complex phenomena than simple geometric coincidence between the potassium lattice and the glutamate side chains.

This conclusion is reinforced by the behavior of other designs; Mica34@KCl10mM shows a similar pattern with angles at 80° and 20° and another minimum between them, but the minimum at −40°/140° is missing; also, the angular dependence of $G(\theta)$ is less pronounced with Mica6@KCl10mM, for which the energy barriers between local minima are smaller (Fig. 2C), consistent with the lower binding affinity of Mica6 to m-mica (001) relative to that of Mica18 and Mica34 (15).

In contrast, Mica18@NaCl10mM does not show a clear, dominant orientation on m-mica (001) but has three roughly equal orientations (Fig. 2D), two of which are about 60° apart and one of which is at an intermediate value. The results demonstrate that ions such as K^+ and Na^+, which have their own distinct binding affinities and adsorption geometries on mica (001) (33, 34), play a critical role in mediating the binding configurations of the proteins on m-mica.

Energy Barriers and Transition Rates. So far, we have discussed the stationary angular distributions of the system. To
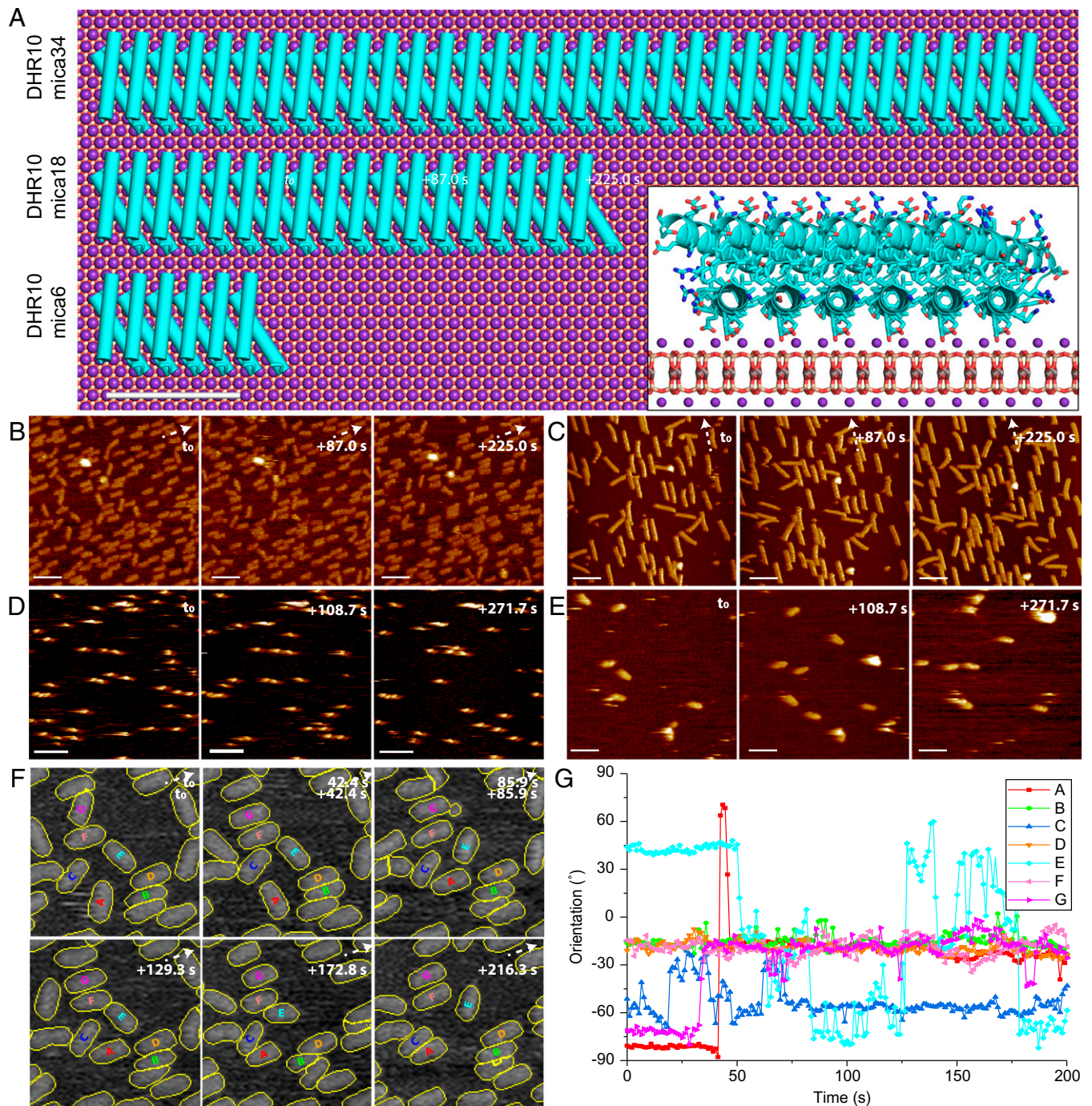
Fig. 1. Overview of design and in-plane dynamics of MicaN on *m*-mica (001). (*A*) Models of MicaN (*N*= 34, 18, and 6) on *m*-mica (001). The purple spheres indicate K⁺ sublattice, occupying the negatively charged sites on *m*-mica (001). *Inset* shows the side view of Mica6 on *m*-mica (001) with glutamate side chains interacting with the K⁺ sublattice. (Scale bar: 5 nm.) (*B–E*) Selected frames of four consecutive HS-AFM datasets, representing the in-plane dynamics of MicaN (*N* = 18, 34, and 6) with 10 mM KCl and Mica18 with 10 mM NaCl on *m*-mica (001), respectively. (Scale bars: 50 nm.) (*F*) Selected frames tracking Mica18 with 10 mM KCl on *m*-mica (001). (*G*) Direction tracking of the selected Mica18 molecules in *F*. The arrows in the upper right corners indicate the dominated orientations of MicaN in each dataset. The times in the upper right corners indicate the relative capturing time of each frame in the corresponding dataset. All solutions are buffered with 20 mM Tris (pH 7). The frame speeds are 0.92 Hz.

investigate dynamics, we compiled trajectory heat maps for the Mica18@KCl10mM, as shown in Fig. 3*A* and Movie S6. These figures, which were generated by identifying nanorods with a specific orientation and plotting a heat map of their subsequent orientational distributions as a function of elapsed time, help reveal several distinct types of dynamic behavior. First, there is a clear tendency for proteins to follow the local slope of the orientational energy landscape and migrate toward local orientational minima. Fig. 3 *A, i* shows that Mica18@KCl10mM

nanorods found with an initial orientation of −7.5° ± 1.5° (i.e., within the deep primary basin in the energy landscape) clearly migrate toward the optimal orientation at −20°. Particles already located at their local minima (such as Fig. 3 *A, ii*) tend to retain their initial orientation but may make small fluctuations within the well. However, given enough time, some of these particles are seen to escape the local minimum and transition into other energy wells, as evidenced by the increasing population in the major minimum at −20°.
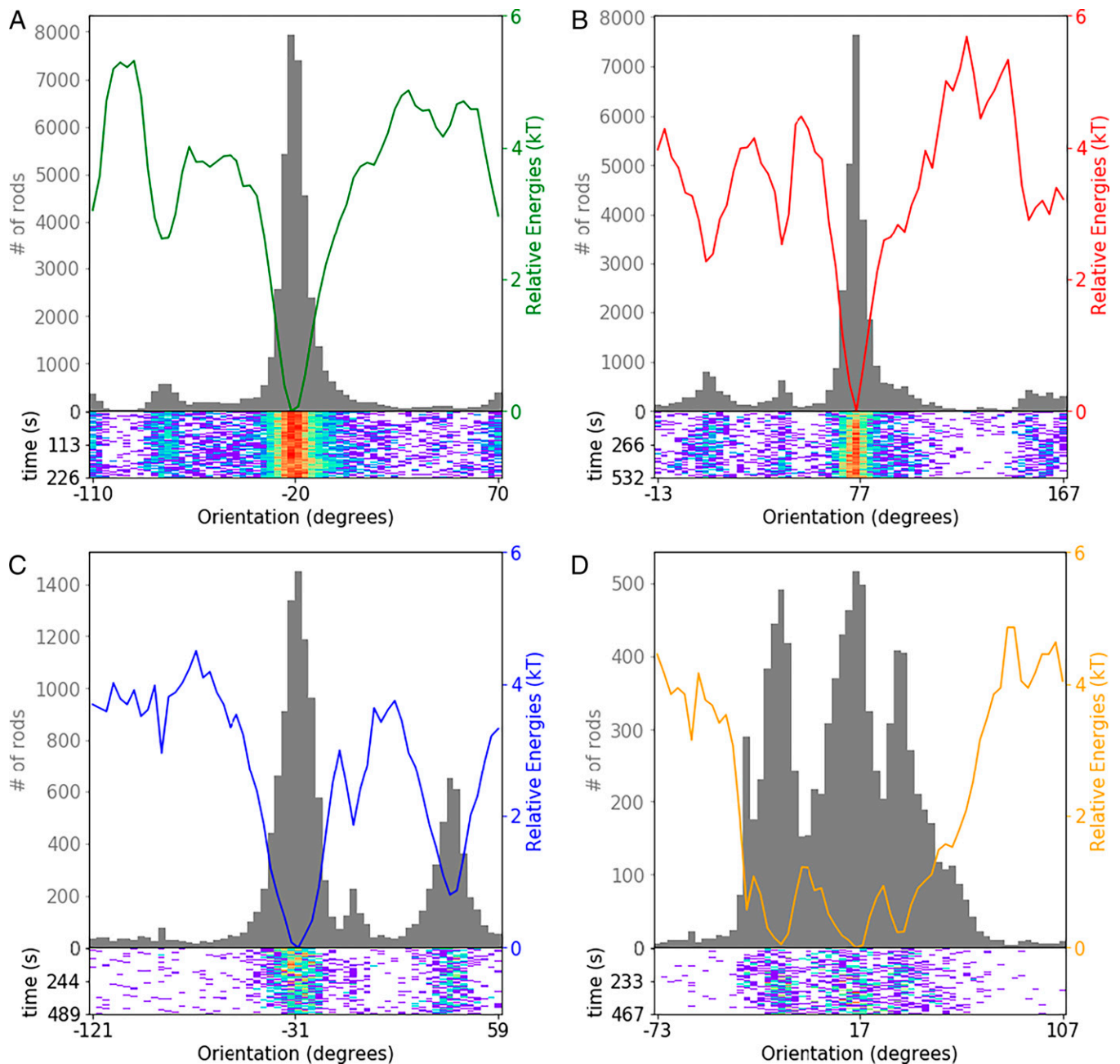
Fig. 2. Orientation distributions and relative energy landscapes, $G(\theta)$, of MicaN. Accumulated orientation distributions over time (histograms) and corresponding derived relative energy landscapes, $G(\theta)$ (colored curves; *Upper*) (details are in *SI Appendix*), and instant orientation distribution at each time step (*Lower*) are visualized for (*A*) Mica18@KCl10mM, (*B*) Mica34@KCl10mM, (*C*) Mica6@KCl10mM, and (*D*) Mica18@NaCl10mM. The orientation distributions are computed from extracted region properties of the instance segmentation. These distributions are normalized relative to the largest numbers in the histograms. The choice of $\theta = 0$ represents horizontal with respect to the image orientation but is arbitrary with respect to the orientation of the m-mica (001) lattice. The orientation range is 180° and centered at the largest minima in each dataset. There is no correlation between the angles (and hence, the states) of different datasets.

In order to efficiently quantify the transition behavior for each dataset, we conducted a detailed analysis on the angular transitions of protein nanorods between consecutive frames in each HS-AFM dataset (Fig. 1*G* shows an example). By computing the frequencies and probabilities for rods to transition between angular states, we constructed transition frequency matrices (Fig. 3*B* and *SI Appendix*, Fig. S2), which give the number of times a rod starting in an angular state X' in one frame is observed transitioning to a state X in the next frame. We found that the transition–frequency matrices of all datasets are roughly symmetric with respect to forward and reverse transitions between states X and X', demonstrating agreement with

the principle of detailed balance (e.g., that every elementary process is equilibrated by its reverse process) and indicating that the dynamics are compatible with a system at thermodynamic equilibrium driven by random motions. Thus, we can approximately treat the motions of proteins as random Markov chains.

To model these dynamics as a Markovian process, the probabilities for transitioning between states were established. Fig. 3*C* shows the state transition probability (P) matrix for Mica18@KCl10mM, where $P(S_{t+1} = X|S_t = X')$ is the probability that a rod will transition to state X given that it started in state X'. The diagonal of the matrix indicates the probability for proteins to
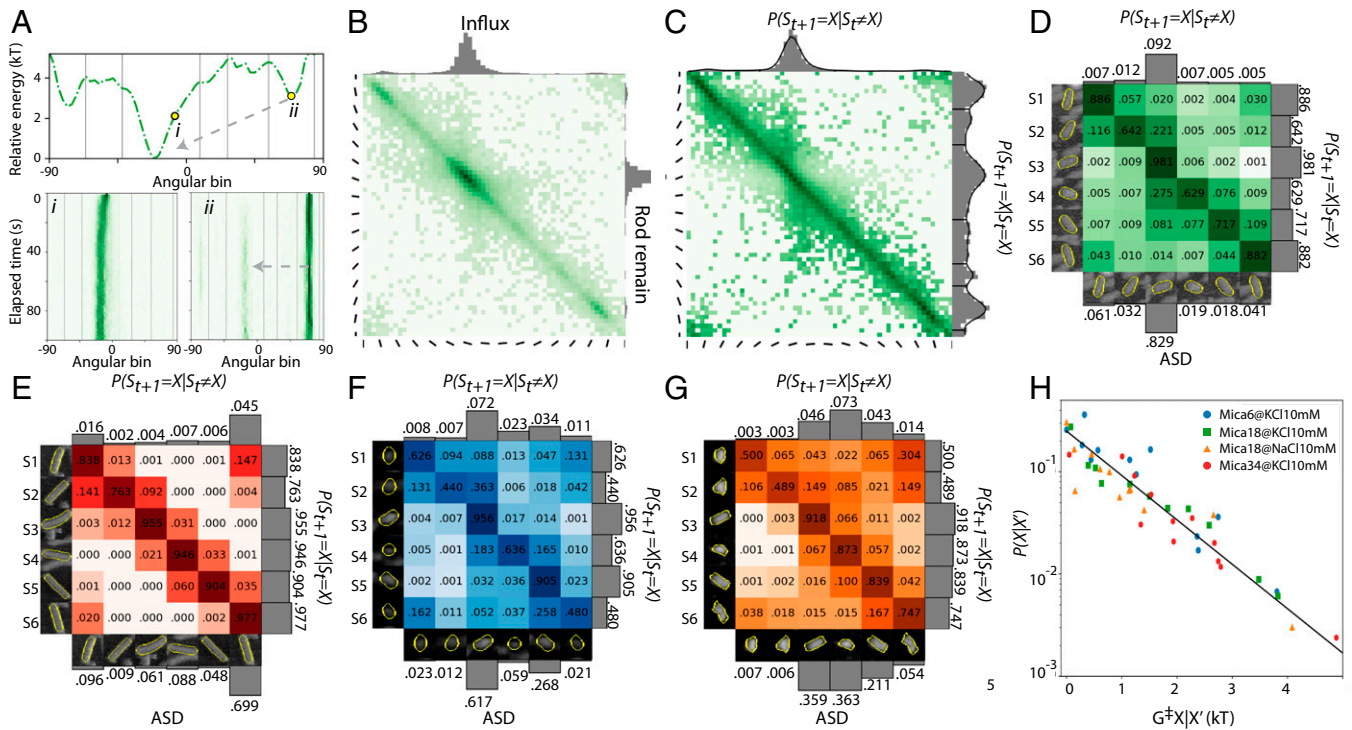
Fig. 3. Trajectory heat maps, degree quantized transition tables, and Markov chain models of MicaN. *A, Upper* shows $G(\theta)$ of Mica18@KCl10mM with two selected cases (*A, i* and *ii* in *Lower*) at different orientations. *A, Lower* shows the corresponding heat maps of orientation distribution evolution as a function of elapsed time. (*B*) The 3° state transition table of Mica18@KCl10mM. (*C*) The state transition probability table of Mica18@KCl10mM constructed for 3° increments in orientation. The upper histogram shows the computed probability (*P*) to enter each state from any other state [i.e., $P(S_{t+1} = X | S_t \neq X)$]. The right histogram shows *P* of remaining in each state [i.e., $P(S_{t+1} = X | S_t = X)$]. Using the peaks and valleys of these two probabilities as marked by the black lines, the orientational states of Mica18@KCl10mM can be split into six states. (*D*) The six-state matrix derived from *C*. As in *C*, the upper histogram shows $P(S_{t+1} = X | S_t \neq X)$, and the right histogram shows $P(S_{t+1} = X | S_t = X)$. The lower bar chart shows the ASD of the hidden Markov model. The zoomed-in AFM images show examples of each state. (*E–G*) The Markov models of Mica34@KCl10mM, Mica6@KCl 10 mM, and Mica18@NaCl10mM derived by the same method as in *D*. (*H*) The correlation of transition probabilities $P(X|X')$ and energy barriers in $G(\theta)$, $G^{\ddagger}_{X|X}$, of Mica18@KCl10mM, Mica34@KCl10mM, Mica6@KCl10mM, and Mica18@NaCl10mM for transition between adjacent states. There is no correlation between the angles (and hence, the states) of different datasets.

remain in the same state [Fig. 3*C*, $P(S_{t+1} = X | S_t = X)$]. From the matrix, we directly computed the influx probability for a protein molecule to rotate into X at time $t + 1$ if it was not in X at time t [$P(S_{t+1} = X | S_t \neq X)$]. This analysis allows us to identify long-lived states that are not immediately visible as local minima in the stationary distributions (Fig. 2*A*). For example, in contrast to the stationary orientation distributions of Mica18@KCl10mM in Fig. 2*A*, there are six, instead of three, distinct peaks in both the remaining and influx probability histograms in Fig. 3*C*.

Using the minima of these distributions, the appropriate boundaries for six populated angular states were defined (Fig. 3*C*, histogram on the right) to obtain the resulting six-state transition probability matrix for Mica18@KCl10mM (Fig. 3*D*). The expected angular stationary distribution (ASD) for the transition probability matrix was then computed with the assumption of a random Markov model (Fig. 3*D*, histogram along the bottom). Compared with the observed distributions of HS-AFM data (Fig. 2*A*), the computed stationary distributions match in all cases with low absolute errors (*SI Appendix,* Table S1). Following the same procedure (*SI Appendix,* Figs. S2 and S3), the six-state Markov models for Mica34@KCl10mM, Mica6@KCl10mM, and Mica18@NaCl10mM were also computed (Fig. 3 *E–G*). These models provide detailed insights into how the transition possibilities between states in these systems give rise to the observed ASDs.

In aggregate, the Markov probability tables reveal a strong correlation between the probability to remain in a state and the depth of the corresponding energy well in $G(\theta)$. More

specifically, when the probability of transitioning into an adjacent state, $P(X|X)$, is plotted vs. the energy barrier for that transition, $G^{\ddagger}_{X|X'}$ (Fig. 3*H*), we clearly observe the expected exponential relationship predicted by transition state theory, with $P(X|X) \sim P_0 \exp(-G^{\ddagger}_{X|X}/k_B T)$. This relationship demonstrates that knowledge of the equilibrium angular-dependent energy landscape allows us to account for a significant portion of the differences in angular dynamics between systems. For example, Mica34, which has the strongest binding affinity to the mica surface and the highest-energy barriers between states, generally displays the lowest frequency of transitions, while the smaller Mica6 tends to display the lowest-energy barriers between states and the highest frequency of transitions.

Transitions between Adjacent States Reproduced in the Brownian Model. The above analysis indicates that the equilibrium orientational energy landscape exhibits significant control over state transition dynamics. Here, we consider the ability of existing classical models of Brownian motion to simulate the observed nanorod rotation and predict rates of transitions between states. To represent Brownian motion on a biased energy landscape, $G(\theta)$, we employ a Markovian random walk-through angular space, assuming thermally activated, nonballistic, diffusive motion.

The use of a nonballistic model is justified on the grounds that the momentum persistence time for nanoscale particles in aqueous fluids is typically on the order of picoseconds, making it much shorter than the timescale of observation. Moreover,

the kinetic energies associated with observable rotational motions are negligibly small in comparison with k_BT (*Materials and Methods* has details). The nonballistic behavior is further confirmed by phase-space diagrams (Fig. 4*B*), derived from the orientation tracking (Fig. 4*A*). In these diagrams, the magnitude of the angular velocities is found to increase as the rods transition from one stable state to another. This behavior is contradictory to ballistic motion, for which one would expect to see velocities slow during transitions as kinetic energy is transformed into potential energy near the maximum of the energy barrier, but it is consistent with thermally activated motion, for which the largest observed motions occur after particles traverse local maxima and are driven into new energy wells.

We implemented the Markovian random walk through a coarse-grained kMC simulation, in which particles undergo a series of incremental rotations of either $+1°$ or $-1°$. The relative frequencies of forward and backward rotations are weighted to reflect the experimentally determined angular free energy landscape using a standard rejection-free kMC residence time criteria, and after simulation, we calibrate the rate constant (i.e., the unbiased rotation attempt frequency) so that the rms angular deviation between sampling time steps matches that of the experimental data (*SI Appendix* has details). The model's ability to reproduce small angular motions on the order of a few degrees is limited by its coarse graining, and efforts to increase the angular resolution are impeded by our limited knowledge of the underlying energy landscape, which is only

determined to $3°$ resolution. Smaller angular motions must be based on an interpolated energy landscape, and we thus expect some systematic error when modeling small-scale angular motions within a given energy well. However, the simulation's resolution should still be sufficient to predict the rates of Brownian transitions between different wells given that the angular distances between local minima in the energy landscape are typically tens of degrees apart.

Comparing the resulting simulation trajectories (Fig. 4*C*) and phase-space diagrams (Fig. 4*D*) with the experimental trajectories (Fig. 4*A*) and phase-space diagrams (Fig. 4*B*), we find that the simulations reproduce much but not all of the experimental behaviors. Given that the nanoparticle motion is stochastic, it is not possible to reproduce any particular trajectory but only to generate trajectories with qualitatively similar behavior. We find that the simulations reproduce the stationary distributions as designed (*SI Appendix*, Fig. S7), and they produce qualitatively similar trajectories, in which proteins spend extended periods of time making small motions about preferred angles (reflecting local minima in the energy landscape); however, they occasionally make large and rapid angular motions as they overcome energy barriers to enter neighboring local minima (Fig. 4 *B* and *D* and *SI Appendix*, Fig. S4). The rapidness of these large jumps is reflected by the appearance of "steps" in both the experimental and simulated trajectories (Fig. 4 *A* and *C*) and the appearance of higher angular velocities as proteins transition between local minima in the phase-space diagrams (Fig. 4 *B* and *D*).
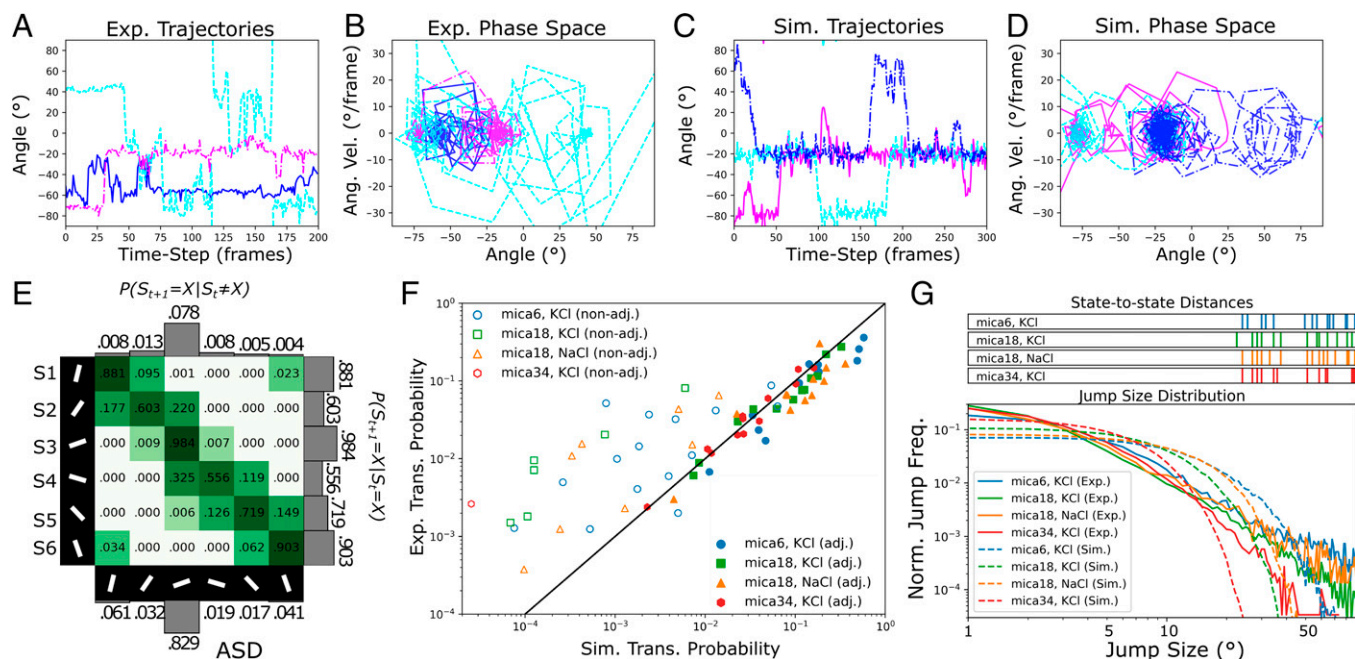

Fig. 4. Trajectories and kMC simulations of rotational dynamics of MicaN. (*A*) Angular trajectories for three protein nanorods selected from Mica18@KCl10mM experimental (Exp.) data shown in Fig. 1. Continuous trajectories were constructed within the cyclic boundary conditions under the assumption that the trajectory between observations followed the minimum angular path length. (*B*) Phase-space diagrams of angular velocity vs. angle for the same experimental data of *A*. (*C* and *D*) Corresponding angular trajectories and phase-space plots generated from kMC simulated (Sim.) trajectories of Mica18@KCl10mM. (*E*) A six-state probability transition table for the Mica18@KCl10mM simulation for comparison with the experimental probability transition table in Fig. 3*B*. (*F*) Comparison of state-to-state transition probabilities per experiment vs. kMC simulations. The aggregated data show good 1:1 agreement for transitions between adjacent states (solid markers), but the experimental probability for transition between nonadjacent states (open markers, non-adj.) is significantly higher than in simulation. The black line shows the ideal 1:1 correlation. (*G*) Jump-size distribution plot. *G, Lower* plots the normalized frequency of observed jumps as a function of jump size. A key feature is that the experimental data (solid lines) show a power law decay, whereas simulated data (dashed lines) show an exponential decay. As a consequence of the power law decay, the experimental data show more transitions at larger distances than are seen in simulation. Vertical lines in *G, Upper* (which shares an *x* axis with *G, Lower*) mark the characteristic center-to-center distances between pairs of states (in degrees). These state-to-state transitions primarily occur at angular distances greater than 20°, highlighting the fact that the region of power law decay seen in the experimental data corresponds to the larger distances where state-to-state transitions are accessible. Note that discrepancies between the experimental and simulated curves at small angles are due, in part, to the approximate nature of the energy landscape derived from the experiments and the coarse graining of that landscape used in the simulation.

There are subtle differences in the size and intensity of the nanorod motion within local minima, as the simulated nanorods tend to make larger and more rapid excursions about the local minima in the angular energy landscape. However, as previously noted, this is expected since it is tied to our limited knowledge of the angular energy landscape at scales of a few degrees or less. Unfortunately, it may not prove feasible to attain a perfect reproduction of nanoscale motion using a single energy landscape since each real protein exists in a slightly different local environment (e.g., local crowdedness or variations in local mica surface chemistry) that may slightly perturb its behavior from what would be expected based on the averaged energy landscape. Despite slight discrepancies at small scale, the simulations show remarkable success at predicting the likelihood that proteins either remain within or exit a given state. When comparing experimental histograms (Fig. 3B) with simulated histograms (Fig. 4E), we see similar patterns in both cases. This confirms that $G(\theta)$ imposes the primary control over protein dynamics and that the Markovian random walk comprises a good first approximation for the primary mechanism of angular motion. This success is most dramatically demonstrated when comparing the probability for observing transitions between various pairs of adjacent states, as shown in Fig. 4F (solid symbols), where we see a clear 1:1 relationship between the experimental and simulated probabilities.

These simulations thus provide a valuable tool for understanding particle motion. They produce trajectories that have many of the qualitative behaviors of real particles (demonstrating periods of intrawell motion punctuated by occasional interwell jumps), and they can quantitatively predict the frequency of transitions between adjacent wells. Moreover, important knowledge can be obtained from these simulations. In particular, the unbiased rotation attempt frequency rate constant calibration for each dataset can be used to estimate rotational diffusion coefficients for each protein and thus, provides an important link to the hydrodynamic properties of proteins at the interface (*Materials and Methods* has details). In fact, estimated diffusion coefficients are found to range from 0.006 rad^2/s for Mica34 to 0.048 rad^2/s for Mica6, and thus, they follow the expected trend that larger particles show lower diffusivity.

Deviations from Brownian Behavior during Transitions between Nonadjacent States. Despite the simulation's success in reproducing jump frequencies between neighboring states, there are important discrepancies when considering large-angle jumps between nonneighboring states. This is most clear in Fig. 4F (open symbols), where we see that the 1:1 relationship between experiment and simulation jump probabilities (which was observed for jumps between adjacent states) is broken for jumps between nonadjacent states. Instead, the simulations systematically predict fewer jumps between nonadjacent states than are observed in the experiment by as much as two orders of magnitude, and this trend exists across all datasets. This behavior can also be examined in detail by comparing the experimental-state probability transition matrices (Fig. 3) with the simulated-state probability transition matrices (*SI Appendix*, Fig. S6) or by comparing the 3° transition tables and transition probability tables (*SI Appendix*, Figs. S2, S3, and S5). In all cases, the experimental tables show a greater probability for "large jumps" occurring far off the diagonal than do the simulations.

The nature of the discrepancy can be better understood through Fig. 4G, which shows a histogram of jump probability vs. jump size. In this figure, the Brownian kMC simulations display an exponential decay in jump probability at large jump size for all proteins. This exponential decay in the simulations differs slightly from simple Gaussian decay, but it is consistent with Brownian behavior under the influence of a complex energy landscape (27). In contrast, the experimental data display a long tail at large jump sizes, with a power law decay that is a key indicator of "Levy-flight" motion through angular space (35, 36). Importantly, this power law decay cannot be reproduced by Brownian simulations, even when accounting for the complex angular energy landscape. Since classical Brownian motion universally emerges when motion is driven by the statistical combination of many smaller, independent events, these long-tail deviations indicate that motion is partially driven by rare, large rotational events. The observed Levy-flight motion is a particularly important case of non-Brownian diffusion that occurs when the underlying jump-size distribution possesses a scale-invariant long-tail distribution (35, 36). This prevents the normally expected convergence toward Brownian behavior (20, 21), and Levy-flight random walks are often referred to as superdiffusion due to the associated enhanced probability for large jumps. The observation of non-Brownian motion here is consistent with numerous other recent in situ transmission electron microscopy and AFM studies of nanoparticle lateral and rotational motion, where both bimodal jump distributions (26) or Levy-flight jump distributions (22, 23, 37) have been observed.

Discussion

Although past studies have revealed complex modes of nanoparticle motion, potentially involving quasithree-dimensional motion of concerted rotational and translational motion or motion through the bulk (26, 35, 37), the results reported here are unique in that we have observed such motions on a complex energy landscape that possesses distinct barriers and local minima in two dimensions. While our analyses show that these barriers are directly linked to the diffusional dynamics—and thus, may provide a tool for tuning the kinetics of rotation and facilitating the controlled assembly of two-dimensional hierarchical architectures at solid–liquid interfaces—they also imply that the more complex motions leading to Levy-flight behavior can provide a pathway for molecules to bypass these barriers to achieve biomaterials with nonequilibrium states.

Levy flights were proposed to explain anomalies in the lateral motion of inorganic nanoparticles (22, 23) as well as polymer chains (24, 25); however, although non-Gaussian Brownian diffusion may be common in soft materials (27), such non-Fickian behavior has not, to our knowledge, been previously identified for the rotational motion of biomolecules. Given the mundane nature of both the substrate (i.e., a heterogeneously structured mineral) and the interacting functional groups of the protein, namely glutamate side chains, the results here suggest that Levy flights may frequently be responsible for large-angle rotations of bio-macromolecules and numerous biological processes at interfaces and play a significant role in sampling the energy minima in the process of attaining equilibrium.

With respect to the mechanistic emergence of Levy-flight dynamics, it has been demonstrated that power law–distributed jump sizes emerge when the underlying events are also power law distributed. Thus, its observation implies that additional modes of rotation involving large jumps exist in addition to the small incremental motions represented by a Brownian model. As we consider how these large jumps are facilitated, a likely

possibility is that large jumps occur when proteins momentarily dissociate from the surface and temporarily enter an excited state, where they encounter conditions of decreased angular drag that enable a brief period of more rapid motion. If a single excited state was accessible, it could lead to a double-exponential jump-size distribution. If, however, there is a spectrum of accessible excited states, with each state being increasingly difficult to access but of increasingly higher angular diffusivity, a power law jump-size distribution could in principle be achieved (*SI Appendix*, Fig. S9). For a bio-macromolecule like MicaN, this spectrum of states may reflect, for example, the number of carboxyl groups that are instantaneously available to bind with the surface with low cation concentration. Thus, we propose that the Levy-flight diffusion and associated superdiffusional jumps between nonadjacent states may be facilitated by a spectrum of quasithree-dimensional motions.

We also consider whether the dynamics of nanorod rotation discussed above are influenced by interactions between neighboring proteins. In our experiments, this effect should be most significant for Mica18@KCl10mM, in which a sizeable fraction of proteins has a nearest neighbor within 16 nm (*SI Appendix,* Fig. S10), which is obtained when particles lie side by side. Surface coverages for the other conditions are much lower, and few particles have nearby neighbors. Thus, we expect that protein–protein interactions are more negligible in these cases. Examining the energy landscapes for subpopulations of Mica18@KCl10mM, we see that that proteins with nearest neighbors closer than 16 nm are slightly more likely to be found in the primary minima and less likely to exist in one of the secondary minima. Thus, collective protein–protein interactions can modify the energy landscape experienced by individual proteins. However, jump-size distributions remain similar, and the existence of the Levy-flight distribution is clearly observed no matter whether the proteins possess a nearby neighbor (*SI Appendix*, Fig. S11).

Finally, the approach reported here quantifies positions and orientations in an HS-AFM image series by ML. It further establishes energy landscapes and transition rates that are then seamlessly connected to the underlying physics via computational models. Thus, the findings both broaden the physical understanding of biomolecular dynamics at solid–liquid interfaces and define a general procedure for using in situ visualization and ML to explore such dynamics. Thus, the approach promises to enable discoveries of physical phenomena in such systems and introduces a feasible sampling procedure for the design of hybrid biomolecular–inorganic materials.

Materials and Methods

Protein Design and Synthesis. The computational design of DHR10-mica6 (Mica6), DHR10-mica18 (Mica18), and the DHR10 protein they are based on is described elsewhere (15, 38). A DHR10-mica34 (Mica34) model was produced by copying the backbone and sequence of internal repeats in Mica18 from 16 to 32 repeats. A plasmid containing a gene encoding Mica34 was cloned from a Mica18-encoding plasmid via recursive directional ligation by plasmid reconstruction (39). All proteins were expressed and purified as described (15), except that Mica18 and Mica34 were expressed without His tags and the Ni–nitriloacetic acid purification and His tag cleavage steps were replaced with the following procedure. First, the clarified lysate was incubated at 80 °C for 1 h and centrifuged at 14,000 relative centrifugal force (RCF) for 30 min, and the supernatant was retained. Second, solid ammonium sulfate was added to 30% saturation; the mixture was rocked at room temperature for 1 h and then, centrifuged at 10,000 RCF for 15 min. Finally, the readily soluble white pellets formed by the MicaN proteins were resuspended in Tris(hydroxymethyl)aminomethane (Tris)–buffered saline solution (150 mM NaCl, 20 mM Tris, pH 8). All subsequent purification steps were completed as described previously (15).

HS-AFM. Protein stock solutions were diluted to 0.05 μM with 20 mM Tris buffer (pH 7) having 10 mM KCl or 10 mM NaCl. A 20-μL diluted protein solution was dropped onto freshly cleaved *m*-mica (001) (SPI Supplies) and characterized by Cypher Video-Rate AFM (Asylum Research) in liquid amplitude–modulation mode. The probe USC-F1.2-K0.15 (NanoWorld) was used. The imaging force was adjusted to minimize any interruption. The estimated energy loss of the AFM cantilever within each oscillation cycle is 0.26 to 0.55 k_BT (19), smaller than the calculated energy barriers of the relative energy landscapes, $G(\theta)$, in Fig. 2. The standard offline data processing was done by Asylum analyzing software written on IgorPro (WaveMetrics) and software SPIP (Image Metrology). Tris·HCl buffer (pH 7, 1 M), KCl, and NaCl were bought from Sigma-Aldrich. Nuclease-free water was bought from Ambion.

Denoising. Before feeding the images into our deep learning segmentation pipeline, we used two techniques for the preprocessing step to correct unbalanced illumination, reduce the amount of noise, and improve the contrast of the input images: bilateral filter (40) and the contrast limited adaptive histogram equalization (CLAHE) (41). The bilateral filter was an edge-preserving and noise-reducing filter. It averaged pixels based on their spatial closeness and radiometric similarity. In other words, it smoothed homogeneous regions of the image and preserved details (such as borders of objects). After improving the signal to noise ratio using the bilateral filter, the next step was to correct illumination and emphasize targeted structures for segmentation. We applied the CLAHE method, which used histograms computed over different tile regions of the image. In doing so, local details were enhanced even in regions that were darker or lighter than most of the image.

Deep Learning Segmentation. The first step in the pipeline involved segmenting the nanorods into blobs. For this, we employed U-NET (42), a deep neural network architecture that is commonly used for microscopy and medical image segmentation. One common issue with deep learning segmentation tasks is the challenge of getting good separation between objects in cluttered microscopy images. The high level of noise makes learning boundaries difficult at times. To combat this issue, we used a strategy similar to that used in Zhou et al. (43), where an additional neural network was trained to learn centroids of rods. For our purposes, the results of this neural network were used as seeds for a watershed postprocess to get better instance segmentation/separation. This was important for extracting as much data as possible from the images. The example pipeline is illustrated as *SI Appendix,* Fig. S1.

Angular Free Energy Calculation. To compute the relative angular free energy landscape, $G(\theta)$, for each condition, we used the distributions of rod orientations throughout all images in the corresponding AFM datasets. We generated a histogram of orientations using 3° bin widths, $N(\theta)$, and normalized these histograms relative to the number of rods in the dominant orientation, $N(\theta^*)$. Taking the negative natural logarithm of these values gave us the $G(\theta)$ in terms of Boltzmann's constant and the temperature, k_BT, as $G(\theta) = -k_BT \ln(N(\theta)/N(\theta^*))$.

Tracking. For performing tracking, we employed a graph-based method similar to that of Yang et al. (44). While the seed/blob segmentation method we had implemented had yielded strong results that offered us the ability to maximize the amount of information extractable from any given frame, this did not guarantee that the boundaries were correctly discerned for all frames for all rods. For some blobs, we still saw false merges that we cannot separate easily using morphology or other methods. This is an issue when trying to do tracking on cluttered images. We first constructed a graph of blob associations for every pair of consecutive frames. Each blob in each frame was a node. For every blob in frame t, we found the blob in the following frame $t + 1$ with the highest intersection over union (IoU) and created an edge between these nodes. Then, for every blob in frame $t + 1$, we found the blob in frame t with the highest IoU and created an edge here. If no parent is found for any of the blobs in frame $t + 1$ and the size of the blob meets some area threshold requirement, this is marked as a "starting point" node for the traversal. We additionally impose a minimum set value for the maximum IoU for potential associations as well.

Once this was completed for all pairs of consecutive frames, we performed a traversal from each starting node (nodes with no connection to anything in the previous frame chronologically). When the traversal hits a node with two children, a decision is made based on the average n former values of the centroid on which child should be considered to be part of that track. From this, we were

able to construct trajectories of rod motions. Once rod trajectories are formed, we visualize the tracking of rods on the original video of the rods. We then filter out tracks with errors from the final tracking analysis.

Markov Analysis. For the Markov model analysis, the first step was to take the segmentation results and perform frame to frame data associations. We used IoU for this association to find the corresponding blobs in frame $t + 1$ to t; for each rod i at a given time t with angle $\theta_i(t)$, we found the corresponding rod at time $t + 1$ to determine $\theta_i(t + 1)$. Since the stationary distributions of angles only revealed the preferred angular states rather than all six expected potential states, we began again by quantizing the data regularly into 3° bins. With this, we could compute the frequency of transitioning between angular states: that is, we compute $| \theta(S_l \rightarrow S_m) | = | \{ i \mid S_l \leq \theta_i(t) < S_l + 3, \; S_m \leq \theta_i(t + 1) < S_m + 3 \} |$, and we do this $\vee \; S_l, S_m \in \{ S \mid -90 \leq S \leq 87, \; a \bmod 3 = 0 \}$. From the transition frequency matrix, we computed the probability $P(S_l \rightarrow S_m)$ of a rod to transition from angle S_l to S_m between two frames by dividing each row of the matrix by the number of rods that started in S_l, which was simply the sum of the corresponding row. This analysis was helpful because it allowed us to 1) find angles that rods prefer to move toward as well as remain in and 2) create a statistical, Markov-based model for how rods move within 1-s intervals (i.e., between frames).

kMC Simulations. kMC was used to simulate the dynamics as a Markovian random walk on a biased energy landscape using a standard rejection-free "residence time" kMC algorithm (45). We considered potential rotations in angle of $\pm 1°$, where rates of forward and backward rotation are given as $k_{+1}(\theta) = k_0 \exp \frac{G(\theta) - G(\theta + 1°)}{2kT}$ and $k_{-1}(\theta) = k_0 \exp \frac{G(\theta) - G(\theta - 1°)}{2kT}$, and $G(\theta)$ is the angle-dependent relative free energy. $G(\theta)$ was estimated from the experimental AFM data as described above and was resampled to 1° intervals using piecewise cubic spline interpolation. The resulting trajectories were resampled at 1-s intervals (corresponding to the AFM video frame rate) after calibrating k_0 such that frame-to-frame rms angular deviation in the simulations reproduced that of the experiment.

Justification of Random Walk Model. The dynamics of particles in dissipative (e.g., viscous) environments can often be broken into fast and slow processes, with short-time behaviors following "ballistic" motion (i.e., motion involving momentum terms) and long-time behavior emerging from the accumulated effect of many incremental motions. As discussed in Einstein's work on Brownian motion (20), it is often the case that the short-time behavior cannot be observed on experimental timescales. In such cases, it is appropriate to model the dynamics using course-grained models that are not sensitive to the detailed behavior on short timescales. Here, we confirm the reasonableness of this approach through two analysis.

First, we consider that the relaxation time, T_r, for perturbations of angular velocity of a nanoparticle in a viscous fluid can be estimated as $T_r \approx I / f$, where I is the moment of inertia and f is the rotational friction drag coefficient. For protein nanorods, we can estimate $I \approx 1/12 \; \text{mass} \cdot \text{length}^2$ using the standard equation for the rotation of a rod about its centroid, and we can approximate $f \approx \pi \cdot \eta \cdot \text{length}^3$ using Kirchoff's formula for the rotational friction drag coefficient of a sphere immersed in a fluid of dynamic viscosity η (44). The resulting relaxation times are estimated to be $T_r \approx 0.2$ to 0.3 ps. These values will be reduced even further in the vicinity of the interface where the effective drag coefficient will be higher. Thus, the timescales for ballistic motion (picoseconds) are much smaller than the experimental timescales (seconds), and we conclude that we can employ course-grained models.

Alternatively, we can consider the energetic contributions of the rotational motion observed in AFM. The maximum angular velocities, ω, that can be observed in AFM using a 1-Hz sampling rate are ~1.5 rad/s. The associated rotational kinetic energy can be estimated as $KE_{\text{rot}} = \frac{1}{2} I \omega^2$, giving maximum observable values of ~$1 \cdot 10^{-40}$, $5 \cdot 10^{-39}$, and $3 \cdot 10^{-38}$ J for Mica6, Mica18, and Mica34, respectively. This is a clearly negligible contribution to the system energy when compared with the thermal energies, which are on the order of $k_B T = 4 \cdot 10^{-21}$ J at room temperature.

These analyses confirm that a course-grained model neglecting ballistic motion, such as a Markovian random walk, provides an appropriate strategy for modeling the observable system dynamics. As noted in the text, we further confirm this through phase-space plots of the nanorod trajectories that show no evidence for a reduction in kinetic energy as the protein traverses the energy barriers between stable states, which would typify ballistic motion.

Checking Ergodicity. Simulated trajectories of angle vs. time were obtained by running the kMC simulations for 20,000,000 iterations. We confirmed that the resulting trajectory is ergodic by comparing the simulation's output population distributions with the experimental angular population distributions that were used to generate the input energy landscape (*SI Appendix*, Fig. S7). We obtain close agreement with the input population distribution and the output populations and can thus assume that the simulation time period has run long enough to provides a statistically robust approximation of the system energetics, for which our analysis will be limited by the statistics of the input data rather than the statistics of the simulation.

Calibration of Time Constant. The kMC simulations utilize an arbitrary rate constant k_0, which must be calibrated in order to directly compare simulation dynamics with experimental dynamics. We perform this calibration for each system by choosing the rate constant that best reproduces the rms angular deviation from frame to frame in the experimental data. The calibration is demonstrated in *SI Appendix*, Fig. S8, where it is determined by finding the cross-over points between the horizontal line (which represents the rms angular deviation that we observe in experimental data between frames: that is, after a 1-s lag time) and the corresponding thicker arc (which shows rms deviation of the simulation data as a function of the number of computational time steps).

The value of k_0, which provides the best agreement between experiment and simulation, is tabulated below for each condition:

p06K: $k_0 = 159$ attempts per second (unbiased diffusion coefficient, $D = 0.048$ rad^2/s);
p18N: $k_0 = 68$ attempts per second (unbiased diffusion coefficient, $D = 0.021$ rad^2/s);
p18K: $k_0 = 52$ attempts per second (unbiased diffusion coefficient, $D = 0.016$ rad^2/s); and
p34K: $k_0 = 19$ attempts per second (unbiased diffusion coefficient, $D = 0.006$ rad^2/s).

We may also estimate the unbiased angular diffusion coefficient (i.e., the angular diffusion coefficient that would be expected for that attempt frequency in the absence of a biasing energy landscape) as $D = k_0 \; _{\Delta \theta 0}{}^2$. This analysis provides physically intuitive results. We find that the attempt frequency (and thus, the angular diffusivity) is inversely related to particle size and is slightly higher with sodium (an ion that binds less strongly to the mica surface than potassium).

Data Availability. The tracking verification of each dataset and the synthetic datasets by the simulation can be found at Zenodo (DOI: 10.5281/zenodo.6300127).

Author affiliations: [a]Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195; [b]Physical Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352; [c]Scientific Data Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720; [d]Department of Biochemistry, University of Washington, Seattle, WA 98195; [e]Institute for Protein Design, University of Washington, Seattle, WA 98195; [f]Department of Computer Science, San Francisco State University, San Francisco, CA 94132; and [g]HHMI, University of Washington, Seattle, WA, 98195

1. D. Moll *et al.*, S-layer-streptavidin fusion proteins as template for nanopatterned molecular arrays. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14646–14651 (2002).
2. C. A. Kerfeld *et al.*, Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936–938 (2005).
3. T. O. Yeates, C. A. Kerfeld, S. Heinhorst, G. C. Cannon, J. M. Shively, Protein-based organelles in bacteria: Carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6**, 681–691 (2008).
4. F. Pfeifer, Distribution, formation and regulation of gas vesicles. *Nat. Rev. Microbiol.* **10**, 705–715 (2012).
5. I. Weissbuch, L. Addadi, M. Lahav, L. Leiserowitz, Molecular recognition at crystal interfaces. *Science* **253**, 637–645 (1991).
6. A. Gal *et al.*, Macromolecular recognition directs calcium ions to coccolith mineralization sites. *Science* **353**, 590–593 (2016).
7. F. Nudelman *et al.*, The role of collagen in bone apatite formation in the presence of hydroxyapatite nucleation inhibitors. *Nat. Mater.* **9**, 1004–1009 (2010).
8. M. Sarikaya, C. Tamerler, A. K. Y. Jen, K. Schulten, F. Baneyx, Molecular biomimetics: Nanotechnology through biology. *Nat. Mater.* **2**, 577–585 (2003).
9. Y. Yang, C. Wang, Hierarchical construction of self-assembled low-dimensional molecular architectures observed by using scanning tunneling microscopy. *Chem. Soc. Rev.* **38**, 2576–2589 (2009).
10. S. Ido *et al.*, Immunoactive two-dimensional self-assembly of monoclonal antibodies in aqueous solution revealed by atomic force microscopy. *Nat. Mater.* **13**, 264–270 (2014).
11. S. Woo, P. W. K. Rothemund, Self-assembly of two-dimensional DNA origami lattices using cation-controlled surface diffusion. *Nat. Commun.* **5**, 4889 (2014).
12. M. Sleutel, J. Lutsko, A. E. S. Van Driessche, M. A. Durán-Olivencia, D. Maes, Observing classical nucleation theory at work by monitoring phase transitions with molecular precision. *Nat. Commun.* **5**, 5598 (2014).
13. W. Mamdouh, M. Dong, S. Xu, E. Rauls, F. Besenbacher, Supramolecular nanopatterns self-assembled by adenine-thymine quartets at the liquid/solid interface. *J. Am. Chem. Soc.* **128**, 13305–13311 (2006).
14. J. Chen *et al.*, Building two-dimensional materials one row at a time: Avoiding the nucleation barrier. *Science* **362**, 1135–1139 (2018).
15. H. Pyles, S. Zhang, J. J. De Yoreo, D. Baker, Controlling protein assembly on inorganic crystals through designed protein interfaces. *Nature* **571**, 251–256 (2019).
16. J. J. De Yoreo, P. G. Vekilov, Principles of crystal nucleation and growth. *Rev. Mineral. Geochem.* **54**, 57–93 (2003).
17. S.-H. Shin *et al.*, Direct observation of kinetic traps associated with structural transformations leading to multiple pathways of S-layer assembly. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 12968–12973 (2012).
18. Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl, F. U. Hartl, Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* **82**, 323–355 (2013).
19. T. Ando, T. Uchihashi, S. Scheuring, Filming biomolecular processes by high-speed atomic force microscopy. *Chem. Rev.* **114**, 3120–3188 (2014).
20. A. Einstein, *Investigations on the Theory of the Brownian Movement*, R. Furth, Ed. (Dover Publications, Mineola, NY, 1956).
21. J. Philibert, One and a half century of diffusion: Fick, Einstein before and beyond. *Diffusion Fundamentals* **4**, 6.1–6.19 (2006).
22. S. W. Chee, U. Anand, G. Bisht, S. F. Tan, U. Mirsaidov, Direct observations of the rotation and translation of anisotropic nanoparticles adsorbed at a liquid-solid interface. *Nano Lett.* **19**, 2871–2878 (2019).
23. S. W. Chee, Z. Baraissov, N. D. Loh, P. T. Matsudaira, U. Mirsaidov, Desorption-mediated motion of nanoparticles at the liquid–solid interface. *J. Phys. Chem. C* **120**, 20462–20470 (2016).
24. M. J. Skaug, J. Mabry, D. K. Schwartz, Intermittent molecular hopping at the solid-liquid interface. *Phys. Rev. Lett.* **110**, 256101 (2013).
25. M. J. Skaug, J. N. Mabry, D. K. Schwartz, Single-molecule tracking of polymer surface diffusion. *J. Am. Chem. Soc.* **136**, 1327–1332 (2014).
26. H. Zheng, S. A. Claridge, A. M. Minor, A. P. Alivisatos, U. Dahmen, Nanocrystal diffusion in a liquid thin film observed by in situ transmission electron microscopy. *Nano Lett.* **9**, 2460–2465 (2009).
27. B. Wang, J. Kuo, S. C. Bae, S. Granick, When Brownian diffusion is not Gaussian. *Nat. Mater.* **11**, 481–485 (2012).
28. E. W. Radoslovich, The structure of muscovite, KAl2(Si3Al)O10(OH)2. *Acta Crystallogr.* **13**, 919–932 (1960).
29. R. M. Pashley, Hydration forces between mica surfaces in aqueous electrolyte solutions. *J. Colloid Interface Sci.* **80**, 153–162 (1981).
30. S. V. Kalinin *et al.*, Big, deep, and smart data in scanning probe microscopy. *ACS Nano* **10**, 9068–9086 (2016).
31. M. Husain, T. Boudier, P. Paul-Gilloteaux, I. Casuso, S. Scheuring, Software for drift compensation, particle tracking and particle analysis of high-speed atomic force microscopy image series. *J. Mol. Recognit.* **25**, 292–298 (2012).
32. J. Akeret, C. Chang, A. Lucchi, A. Refregier, Radio frequency interference mitigation using deep convolutional neural networks. *Astron. Comput.* **18**, 35–39 (2017).
33. I. C. Bourg, S. S. Lee, P. Fenter, C. Tournassat, Stern layer structure and energetics at mica–water interfaces. *J. Phys. Chem. C* **121**, 9402–9412 (2017).
34. M. Ricci, P. Spijker, K. Voïtchovsky, Water-induced correlation between single ions imaged at the solid-liquid interface. *Nat. Commun.* **5**, 4400 (2014).
35. B. B. Mandelbrot, *The Fractal Geometry of Nature* (Henry Holt and Company, 1983).
36. D. H. Zanette, P. A. Alemany, Thermodynamics of anomalous diffusion. *Phys. Rev. Lett.* **75**, 366–369 (1995).
37. L. R. Parent *et al.*, Tackling the challenges of dynamic experiments using liquid-cell transmission electron microscopy. *Acc. Chem. Res.* **51**, 3–11 (2018).
38. T. J. Brunette *et al.*, Exploring the repeat protein universe through computational protein design. *Nature* **528**, 580–584 (2015).
39. J. R. McDaniel, J. A. Mackay, F. G. Quiroz, A. Chilkoti, Recursive directional ligation by plasmid reconstruction allows rapid and seamless cloning of oligomeric genes. *Biomacromolecules* **11**, 944–952 (2010).
40. C. Tomasi, R. Manduchi, "Bilateral filtering for gray and color images" in *Sixth International Conference on Computer Vision*, S. Chandran, U. Desai (Narosa Publishing House, New Delhi, India, 1998), pp. 839–846.
41. K. Zuiderveld, "Contrast limited adaptive histogram equalization" in *Graphics Gems IV*, P. S. Heckbert, Ed. (Academic Press Professional, Inc., 1994), pp. 474–485.
42. O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional networks for biomedical image segmentation" in *Medical Image Computing and Computer-Assisted Intervention* (Springer International Publishing, Munich, Germany, 2015), pp. 234–241.
43. Z. Zhou *et al.*, "Joint multi-frame detection and segmentation for multi-cell tracking" in *International Conference on Image and Graphics* (Springer International Publishing, Beijing, China, 2019), pp. 435–446.
44. F. Yang, M. A. Mackey, F. Ianzini, G. Gallardo, M. Sonka, "Cell segmentation, tracking, and mitosis detection using temporal context" in *Proceedings of the 8th International Conference on Medical Image Computing and Computer-Assisted Intervention–Volume Part I*, J. S. Duncan, G. Gerig, Eds. (Springer-Verlag, Palm Springs, CA, 2005), pp. 302–309.
45. A. F. Voter, *Introduction to the Kinetic Monte Carlo Method* (Springer Netherlands, Dordrecht, The Netherlands, 2007), pp. 1–23.