Structural and dynamics analysis of intrinsically disordered proteins by high-speed atomic force microscopy

Noriyuki Kodera, Daisuke Noshiro, Sujit K Dora, Tetsuya Mori, Johnny Habchi, David Blocquel, Antoine Gruet, Marion Dosnon, Edoardo Salladini, Christophe Bignon, et al.

Structural and dynamics analysis of intrinsically disordered proteins by high-speed atomic force microscopy

Noriyuki Kodera[1†], Daisuke Noshiro[1†], Sujit K. Dora[2†], Tetsuya Mori[2], Johnny Habchi[3], David Blocquel[3], Antoine Gruet[3], Marion Dosnon[3], Edoardo Salladini[3], Christophe Bignon[3], Yuko Fujioka[4], Takashi Oda[5], Nobuo N. Noda[4], Mamoru Sato[5], Marina Lotti[6], Mineyuki Mizuguchi[7], Sonia Longhi[3]* and Toshio Ando[1]*

[1]Nano Life Science Institute (WPI-NanoLSI), Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan
[2]Department of Physics, Institute of Science and Engineering, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan
[3]CNRS and Aix-Marseille Univ, Laboratoire Architecture et Fonction des Macromolecules Biologiques (AFMB), UMR 7257, 13288, Marseille, France
[4]Institute of Microbial Chemistry, Tokyo 141-0021, Japan
[5]Graduate School of Nanobioscience, Yokohama City University, Yokohama 230-0045, Japan
[6]Department of Biotechnology and Biosciences, State University of Milano-Bicocca, Milano, Italy
[7]Faculty of Pharmaceutical Sciences, University of Toyama, 2630 Sugitani, Toyama 930-0194, Japan

[†]These authors equally contributed to this work.
*Correspondence and requests for materials should be addressed to T.A. (tando@staff.kanazawa-u.ac.jp) or S.L. (sonia.longhi@afmb.univ-mrs.fr).

Intrinsically disordered proteins (IDPs) are ubiquitous proteins that are disordered entirely or partly and play important roles in diverse biological phenomena. Their structure dynamically samples a multitude of conformational states, thus rendering their structural analysis considerably difficult. Here we explore the potential of high-speed atomic force microscopy (HS-AFM) for characterizing the structure and dynamics of IDPs. Successive HS-AFM images of an IDP molecule can not only identify constantly folded and constantly disordered regions in the molecule but also can document disorder-to-order transitions. Moreover, the number of amino acids contained in these disordered regions can roughly be estimated, enabling a semi-quantitative, realistic description of the dynamic structure of IDPs.

IDPs constitute nearly a half of the entire protein realm[1,2] and function as hubs in signalling and regulation of transcription, translation and cell cycle[3]. Our understanding of IDPs however falls far behind that of well-structured proteins. X-ray crystallography and electron microscopy (EM) have no practical application for IDPs because intrinsically disordered regions (IDRs) hamper crystallization and are too thin to be visualized by EM. NMR is so far most instrumental in the structural analysis of IDPs[4,5] but provides only distance restrains between amino-acid (a.a.) residues separated by < 2 nm. To complement NMR short-range structural information and estimate the overall dimensions of IDPs, small-angle X-ray scattering (SAXS) is used[6]. However, both are ensemble-averaging methods that cannot examine individual conformations and their populations. To mitigate this problem, computational ensemble generation of disordered conformations is attempted by simulations[7,8] or using statistical distribution of coil conformations from structure databases[9,10]. Single-molecule Förster resonance energy transfer spectroscopy, which can resolve structural and dynamic heterogeneity, can be used for the structural analysis of chemically denatured proteins and IDPs[11-13]. Nevertheless, the donor and acceptor have to be in close proximity (< ~10 nm). Thus, we still lack versatile techniques enabling simultaneous and temporal assessment of both local (even with semi-residue resolution) and overall structures of IDPs.

Here we explored the potential of HS-AFM[14,15] for structural and dynamics analysis of IDPs. All imaging experiments were performed in amplitude modulation mode, where a cantilever was vertically oscillated at its resonant frequency of ~1 MHz in buffer solution. The cantilever free oscillation amplitude and the set point amplitude were set at appropriate values so that the loss of cantilever's oscillation energy *per* tap was adjusted to be 2−3 k_BT on average (k_B, Boltzmann constant; T, room temperature in kelvin) (Supplementary Note 1). This energy, which is considered to be transferred to the specimen, is partitioned into many degrees of freedom in the specimen and dissipates quickly into solution. As demonstrated previously for an IDP, Facilitates Chromatin Transcription (FACT) protein, HS-AFM can visualise the thin and flexible structure of IDRs[16,17]. These studies, along with numerous previous HS-AFM studies performed on structured proteins[e.g.,18-21], have ruled out possible effects of the tip-sample contact on the structural properties and function of the protein under study. The present study on various IDPs shows that HS-AFM can identify at the near residue level the fully disordered regions and the

border between disordered and folded regions. Moreover, it provides information on the structural transition dynamics and the structural nature of transiently folded conformations.

Structural analysis of PQBP-1 constructs

We first imaged polyglutamine tract binding protein-1 (PQBP-1) and its four deletion constructs with shorter IDRs placed on mica, at 13.1–23.8 frames per second (fps) using HS-AFM (Fig. 1 and Supplementary Table 1). The HS-AFM movie (Supplementary Video 1) revealed that wild type (WT) PQBP-1 has a small globule at one end and a long flexible, thin structure (i.e., IDR) (Fig. 1b, top), consistent with previous NMR data[22]. The N-terminal segment (1−47), for which no NMR data are available, appeared as being part of the globule containing the ordered WW domain (48−81). In the five constructs (Fig. 1a), the IDRs with average diameter 0.5 ± 0.1–0.3 nm (mean ± standard deviation, s.d.) (Fig. 1e and Supplementary Table 1) appeared constantly disordered. To measure the two-dimensional (2D) end-to-end distance (R_{2D}) of the IDRs, the direct distance D between the N-terminal globule and the C-terminal end was measured, followed by measurements of the N-terminal globule height H_1 and the C-terminal end height H_2. R_{2D} was estimated as $R_{2D} = D - H_1/2 - H_2/2$ (Fig. 1c and Supplementary Note 2).

The estimated values of $\langle R_{2D} \rangle$, $\langle H_1 \rangle$ and $\langle H_2 \rangle$ of PQBP-1 (1–214) were independent of the imaging rate within the tested range (6.7–50 fps; Extended Data Fig. 1 and Supplementary Table 2), suggesting no significant structural deformation of the molecule by its contact with the tip. Moreover, in WT PQBP-1 the C-terminus position was circularly distributed around the WW domain (Fig. 2a), and its displacement in the fast scan (X) direction occurring during different time intervals were symmetrically distributed around their zero mean displacement (Fig. 2b). The averaged absolute values of displacements in the X-direction were nearly identical to those in the slow scan (Y) direction (Fig. 2c). Since the HS-FM images were captured during tip scanning relative to the sample stage from −X to +X direction, these data showing no directional bias in the C-terminus displacements ascertain no effect of tip-molecule contact on the IDR motion. Besides, the autocorrelation functions of time series of R_{2D} showed no correlation of its change over time even when the HS-AFM images were captured at 50 fps, suggesting no impediment of the mica surface and tip to IDR diffusional motion in this time scale (Fig. 2d–f).

The $\langle R_{2D} \rangle$ values for these constructs were 17.2 ± 0.09, 14.4 ± 0.14, 11.0 ± 0.04, 9.3 ± 0.07, and 7.1 ± 0.02 nm (mean ± standard error, s.e.; Supplementary Note 3) in the order of

decreasing construct length (Fig. 1f), approximately following a scaling law $\langle R_{2D} \rangle = \beta_{2D} \times N_{aa}^{\nu}$ with β_{2D} = 1.18 ± 0.11 nm (best fit value ± s.e.) and ν = 0.51 ± 0.02. This Flory exponent ν is close to 0.5, which holds for 2D and three-dimensional (3D) ideal and Gaussian chains[23], implying that these constantly disordered segments have similar values of persistence length (L_p) that can be estimated from

$$\langle R_{2D}^2(L) \rangle = 4L_pL\left[1 - \frac{2L_p}{L}\left\{1 - exp\left(-\frac{L}{2L_p}\right)\right\}\right], \quad (1)$$

where L represents the contour length of a chain[24]. For an unstructured polypeptide chain, $L = (N_{aa} - 1) \times \langle d_{aa} \rangle$, where N_{aa} is the number of a.a. contained in the chain and $\langle d_{aa} \rangle$ is the average distance between two adjacent a.a., with the most likely value of $\langle d_{aa} \rangle$ being 0.36 nm[25]. The L_p values estimated for these constructs were 1.16, 1.15, 1.11, 1.35, and 1.45 nm in the order of decreasing construct length.

The chain conformations of these IDRs are very likely to be affected to some extent by their confinement to the liquid-mica interface because of their significantly flexible nature. To assess this influence, we performed SAXS measurements in solution (i.e., not on mica) using four IDRs of PQBP-1 (82–265, 82–214, 82–164 and 82–134), yielding values of radius of gyration $\langle R_g \rangle$ of 3.62 ± 0.05, 3.10 ± 0.18, 2.69 ± 0.04, and 2.14 ± 0.04 nm, respectively (red circles in Fig. 3n; Extended Data Fig. 2 and Supplementary Table 3). In Gaussian chains, $\langle R_{2D}^2 \rangle$ is theoretically identical to $2 \times \langle R_{3D}^2 \rangle$, where R_{3D} is the end-to-end distance in solution, and the value of $\langle R_{3D}^2 \rangle / \langle R_g^2 \rangle$ is 6.0 when $L >> L_p$[26]. Therefore, we expect $\langle R_{2D}^2 \rangle / \langle R_g^2 \rangle \approx 12.0$. However, the $\langle R_{2D}^2 \rangle / \langle R_g^2 \rangle$ values of the fully disordered IDRs were significantly larger than 12.0 (20.0 on average), indicating that the mica surface expands the IDR 2D dimensions roughly by a factor $\sqrt{20/12} \approx 1.3$. Its possible cause could be frictional forces locally exerted from mica against fast Brownian motion of the IDR chain, which would increase the IDR chain's undulation wavelength and/or decrease the undulation amplitude, resulting in swelling of its 2D dimensions[27].

Structural analysis of autophagy proteins

Next, we imaged autophagy proteins containing longer IDRs, Atg1 (D211A kinase-dead mutant) and Atg13 (Fig. 3 and Supplementary Videos 2 and 3). The IDR of Atg1 (D211A) showed a double Gaussian distribution of R_{2D} (Fig. 3h), suggesting interactions between the IDR

and the dead kinase domain, in agreement with phosphorylation of the IDR by the N-terminal kinase domain of the WT protein (Extended Data Fig. 3b). The IDR of Atg13 exhibited temporarily appearing and disappearing small globules (Fig. 3d, 9.2 s and 30.0 s), resulting in a double Gaussian distribution of R_{2D} (Fig. 3l). Since the four small regions within the IDR (red bars in Fig. 3c) are known to interact with Atg1 or Atg17[28], the small globules probably correspond to these regions. The $<R_{2D}>$ values in the longer IDR states of Atg1 and Atg13 were 20.5 ± 1.42 and 28.1 ± 4.62 nm (Fig. 3h,l), respectively, and remained nearly unchanged upon pH increase from 6 to 8 for both proteins and upon modification of both salt type and concentration for Atg1 (Fig. 3i,m). In contrast, the $<R_{2D}>$ values in the shorter states varied (Extended Data Figs. 3 and 4 and Supplementary Tables 4 and 5). Since the IDRs in the longer states are considered to be fully disordered, we estimated their L_p values using (N_{aa}, $<R_{2D}>$) = (255, 20.5 nm) for Atg1 and (N_{aa}, $<R_{2D}>$) = (477, 28.1 nm) for Atg13, yielding L_p of 1.18 and 1.17 nm, respectively. These "measured L_p values" close to each other are also close to those of the IDRs in the PQBP-1 constructs. The height of the C-terminal end of Atg13 was 0.5 ± 0.2 nm (mean ± s.d.) (Fig. 3k).

Structural analysis of N_{TAIL}

Next, we observed the C-terminal domain of the measles virus (MeV) nucleoprotein (N_{TAIL}). Its Box1 (401–420) and Box2 (486–502) regions were previously shown to correspond to a Molecular Recognition Element[29,30] of helical nature (α-MoRE) in equilibrium among completely unfolded and folded helical conformations[30–34]. We used N_{TAIL} constructs fused at their N- or C-terminus to GFP or thioredoxin (Trx) to impede the otherwise very fast diffusion of N_{TAIL}. N_{TAIL}-GFP (Fig. 4a) showed a thin and flexible tail-like structure and a C-terminal GFP globule with mean height of 3.2 nm[35] (Fig. 4d,g and Supplementary Video 4). The IDR showed a temporarily appearing and disappearing small globule at the N-terminus corresponding to Box1 (Fig. 4d). Except for this, the IDR was observed to be fully disordered (with mean height of 0.4−0.5 nm) (Extended Data Fig. 5a). Box2 was indiscernible due to its close proximity to GFP but appeared in the images of Trx-N_{TAIL} (Extended Data Fig. 6, Supplementary Fig. 1 and Video 5).

The R_{2D} histogram of the IDR in N_{TAIL}-GFP showed a single Gaussian peak at 11.6 ± 0.1 nm (Fig. 4h and Supplementary Fig. 2), consistent with the very small population of Box1 in the

158 higher state, as observed in both N_{TAIL}-GFP (Fig. 4k) and N_{TAIL}-Trx (Extended Data Fig. 7f,i and
159 Supplementary Video 6), and the large population of Box2 in the higher state as observed in Trx-
160 N_{TAIL} (Extended Data Fig. 6e). From $<R_{2D}>$ = 11.6 nm and N_{aa} = 88 in the IDRs spanning 421–
161 485 and 503–525, L_p was estimated to be 1.16 ± 0.02 nm, a value close to those described above.
162 HS-AFM images of N_{TAIL}-Trx also gave similar $<R_{2D}>$ values at pH 6.0 and 7.0 (11.2 ± 0.1 and
163 11.1 ± 0.1 nm, respectively) (Extended Data Fig. 7e,h), yielding L_p values around 1.06 nm.
164 However, HS-AFM images of Trx-N_{TAIL} gave a slightly smaller value, $<R_{2D}>$ = 10.1 ± 0.1 nm
165 (Extended Data Fig. 6d), resulting in a smaller L_P value (0.86 ± 0.02 nm), possibly due to an
166 increased order propensity of Box1 induced by interaction with the adjacent Trx. To gain more
167 insights into the structural transitions of Box1 and Box2, we measured their height H_2. The H_2
168 histograms of Box1 in N_{TAIL}-GFP and N_{TAIL}-Trx (at pH 6.0 and 7.0), as well as that of Box2 in
169 Trx-$N_{TAIL,}$ showed double Gaussian peaks at 0.8 and 1.1 nm (Fig. 4k, Extended Data Figs. 6e and
170 7f,i and Supplementary Tables 6 and 7). Moreover, despite the single Gaussian distribution of
171 R_{2D}, these H_2 distributions were correlated with the corresponding R_{2D} distributions (Fig. 4n and
172 Supplementary Fig. 3), suggesting that the IDR length partly changes due to order-disorder
173 transitions in these globules. The order propensity of Box1 and Box2 was estimated from the
174 population ratios (K_e) between their higher and lower states, yielding K_e = 0.19−0.28 (depending
175 on the pH value) and K_e = 1.91 (at pH 6.0), respectively (Supplementary Tables 6 and 7). The H_2
176 value of 0.8 nm is larger than the height of fully disordered regions, indicating that Box1 and
177 Box2 are both partially ordered even in their lower height states. The H_2 value of 1.1 nm is
178 identical to the diameter of a single α-helix, suggesting that Box1 and Box2 are α-MoREs,
179 consistent with previous studies[29,32]. From these features, the dynamic structure of N_{TAIL} could
180 be sketched (Fig. 4q).
181
182 **Power law for $<R_{2D}>$**
183 In the above observations, we used proteins with IDR boundaries previously identified at
184 the residue level. The $<R_{2D}>$ values of these IDRs in their fully disordered state were plotted
185 against N_{aa} (Fig. 3n), together with a data point (N_{aa}, $<R_{2D}>$) = (121, 14.3 ± 0.18 nm) obtained by
186 our previous HS-AFM study for a fully disordered IDR in unphosphorylated FACT protein[17].
187 The nine points could be fitted to a power law $<R_{2D}> = \beta_{2D} \times N_{aa}^{\nu}$ with β_{2D} = 1.16 ± 0.057 nm and

188 ν = 0.52 ± 0.009 (blue solid line in Fig. 3n). The mean observed L_p value was 1.18 nm
189 (Supplementary Table 8).
190
191 **Structural analysis of PNT**
192 Although the N-terminal domain of the MeV phosphoprotein (PNT) is mainly disordered,
193 the regions 1–37, 87–93, and 189–198 were shown by NMR to have α-helical propensities, and
194 hence named $\alpha_{1/2}$, α_3 and α_4, respectively[36]. α_3 and α_4 are interspaced by an acidic IDR (124–168).
195 The resistance of the N-terminal 27–99 region of PNT to proteolysis in the presence of
196 trifluoroethanol[37] (a secondary structure stabilizer[33,34]), cannot be rationalized from the NMR
197 studies, thus suggesting the existence of a large-scale structure undetectable by NMR. To shed
198 light onto this issue, we imaged a PNT segment (1–229) fused to GFP (PNT-GFP; Fig. 4b). PNT-
199 GFP exhibited a thin and flexible tail-like structure with a temporarily appearing and
200 disappearing small globule at the N-terminal end, together with the C-terminal GFP (Fig. 4e and
201 Supplementary Video 7). In addition, a small globule close to GFP could be occasionally
202 detected (middle panel in Fig. 4e), very likely corresponding to α_4. The R_{2D} histogram of the IDR
203 was best fitted to a double Gaussian with peaks at 8.9 ± 0.33 and 14.3 ± 9.26 nm (Fig. 4i). The
204 mean height of the IDR was 0.4−0.5 nm, in both metastable shorter and longer states (Extended
205 Data Fig. 5b), indicating that the IDR is fully disordered even upon disorder-to-order transition of
206 the small globule. Unlike R_{2D}, the H_2 histogram of the N-terminal small globule undergoing
207 transitions was best fitted to a single Gaussian with a peak at 1.1 ± 0.003 nm (Fig. 4l). To
208 estimate N_{aa} in the IDR, we used both the power law $\langle R_{2D} \rangle = 1.16 \times N_{aa}^{0.52}$ and $\langle L_p \rangle$ = 1.18 nm
209 as obtained above, yielding N_{aa} = 52 and N_{aa} = 129 from the power law, and N_{aa} = 54 and N_{aa} =
210 128 from $\langle L_p \rangle$ for the shorter and longer states, respectively. Therefore, the number of a.a.
211 contained in the small globule is roughly estimated to be ~175 and ~100 for the shorter and
212 longer tail states, respectively (α_4 does not much affect the estimate). The number ~100 suggests
213 that $\alpha_{1/2}$, α_3 (87−93) and the IDR segment between them (38−86) are contained in the small
214 globule in both longer and shorter IDR states. Therefore, the region encompassing the N-
215 terminus and helix α_3 always forms a compact structure, explaining the resistance to proteolysis
216 of the N-terminal segment extending to a.a. 99. The number ~175 suggests that the structural
217 transitions between the two states mainly occur by association and dissociation between the N-

terminal compact structure and the acidic IDR (124–168) (Fig. 4r). Note that $\langle H_2 \rangle$ = 1.1 nm possibly corresponds to the height of α-helices contained in the small globule.

Structural analysis of Sic1

Finally, using Sic1-GFP (Fig. 4c) we studied this IDP that possesses a propensity to adopt a collapsed chain-like form[38]. Sic1-GFP exhibited a thin and flexible tail-like structure with a temporarily appearing and disappearing small globule at the N-terminal end, together with the C-terminal GFP (Fig. 4f and Supplementary Video 8). The R_{2D} histogram for the IDR was best fitted to a double Gaussian with peaks at 12.2 ± 0.66 and 20.5 ± 2.14 nm (Fig. 4j). The IDR is fully disordered in both shorter and longer states, as judged from its mean height of 0.4−0.5 nm in both metastable states (Extended Data Fig. 5c,d). The H_2 histogram for the N-terminal small globule undergoing transitions was best fitted to a double Gaussian with peaks at 0.9 ± 0.05 and 1.3 ± 0.28 nm (Fig. 4m). The height difference (0.4 nm) between the two states is small, compared to the large difference (8.3 nm) between the two $\langle R_{2D} \rangle$ values, suggesting that the partially folded state has a flattened, collapsed structure, consistent with a previous study[38]. As done above for PNT-GFP, we estimated N_{aa} contained in the shorter and longer IDRs; we obtained N_{aa} = 95 and N_{aa} = 259 from the power law and N_{aa} = 95 and N_{aa} = 255 from $\langle L_P \rangle$ = 1.18 nm for the shorter and longer states, respectively. Therefore, from the total number of a.a. contained in Sic1 (284), the number of a.a. contained in the small globule was estimated to be ~30 and ~190 for its lower and higher states, respectively. The large number ~190 indicates that as much as ~65% of the whole protein is loosely folded, with only the remaining C-terminal region (encompassing roughly 95 a.a) being fully extended. The remarkable difference between ~30 and ~190 indicates that a large segment containing ~160 residues undergoes a transition between fully unstructured and loosely folded conformations (Fig. 4s). The folding propensity derived from the H_2 distribution (K_e = 1.6) is close to that estimated from the R_{2D} distribution (K_e = 1.5).

Order propensity and structural transition dynamics

From all K_e values obtained above, we can infer, for all small globules undergoing transitions, the following relative order propensity: PNT > Box2 (N_{TAIL}) > Sic1 >> Box1 (N_{TAIL}) (Fig. 4q−s and Supplementary Tables 6 and 7). Disorder predictions failed to detect these subtle

differences (Supplementary Figs. 4 and 5), underlining the importance of experimental analysis of order-disorder transitions. The order propensity and the transiently ordered structures very likely dictate key features in binding of these IDPs to their targets (Supplementary Note 4). Next, we calculated auto-correlation functions $G(\tau)$s for time-series $H_2(t)$ or $R_{2D}(t)$ data of these IDPs to estimate the order-to-disorder and disorder-to-order transition rates (k_{OD} and k_{DO}, respectively). The rate constants can be expressed as $k_{OD} = \lambda/(1 + K_e)$ and $k_{DO} = \lambda K_e/(1 + K_e)$, where $\lambda \equiv k_{OD} + k_{DO}$ is the decay rate of $G(\tau)$. All $G(\tau)$s were best fitted to single exponential functions (Fig. 2g–m). Time correlations in the $G(\tau)$s for $R_{2D}(t)$ (Fig. 2k,m) reflect the order/disorder transitions, because Brownian undulation motion of constantly disordered IDRs in PQBP-1 constructs is too fast to appear in $G(\tau)$ (Fig. 2d–f). In fact, in Sic1 the $G(\tau)$s for $H_2(t)$ and $R_{2D}(t)$ were nearly identical (Fig. 2l,m). In contrast to the widely varying lifetime of the ordered state (0.16–1.89 s) among these IDPs, the lifetime of disordered state does not vary much (0.49–0.94 s) (see Fig. 4q–s and Supplementary Table 6).

Refinement of mica effect, power law and L_p

SAXS data of non-canonical IDPs obtained by truncation of foldable proteins yielded ν = 0.54, despite their hydrophobic nature[39]. SAXS data of chemically denatured proteins yielded ν = 0.598[40], indicating that chemically denatured proteins are more extended than native IDRs[11,41] and behave as polymer chains in good solvent (theoretically ν = 0.6[23] or 0.588[42]). In contrast, $\langle R_g \rangle$ vs. N_{aa} plots for 72 naturally occurring IDPs without permanent secondary structures were found to be largely scattered[43], possibly due to residual structure leading to varied compactness. In fact, this inference is corroborated by detailed structural analyses for some of the 72 IDPs[44–48] (Supplementary Note 5). Here we should emphasize that unlike SAXS, our HS-AFM method can single out IDRs in the fully disordered state even when they undergo disorder-order transitions. We here selected ten tau-protein constructs that are not significantly affected by the extended repeat domains and phosphor-mimic mutation[47]. Then, we combined those $\langle R_g \rangle$ data with those of the four PQBP-1 constructs herein reported. Fitting this data set to a power law yielded $\langle R_g \rangle = (0.25 \pm 0.038) \times N_{aa}^{0.53 \pm 0.028}$ (Supplementary Fig. 6). A very close power law ($\langle R_g \rangle = 0.254 \times N_{aa}^{0.522}$) was previously obtained for several IDRs from their theoretical $\langle R_g \rangle$ values calculated using the coil database[49]. This agreement prompted us to use these 14 experimental (N_{aa}, $\langle R_g \rangle$) data to quantitate more accurately the expansion effect of the mica surface on $\langle R_{2D} \rangle$, while

assuming that this effect is independent of N_{aa}. The measured $<R_{2D}>$ values of the IDRs (in the longer state when undergoing transitions) were converted to corresponding $<R_g>$ values as $<R_g> = <R_{2D}>/(2\sqrt{3}u)$, where u is the expansion factor. Then, these nine (N_{aa}, $<R_g>$) data containing the parameter u were combined with the above fourteen SAXS data of (N_{aa}, $<R_g>$). Finally, the combined data were fitted to a power law, yielding $u = 1.24 \pm 0.01$ and a fitted curve of $<R_g> = (0.26 \pm 0.02) \times N_{aa}^{0.52 \pm 0.015}$, from which all data points only slightly deviated (Extended Data Fig. 8a). From the $<R_g(u)>$ values at $u = 1.24$ for the nine IDRs, their L_p values in solution were calculated (Supplementary Table 8), yielding $<L_p> = 0.78 \pm 0.06$ nm. Similar L_p values can be found in various IDPs[12,13,43] (Extended Data Fig. 8b and Supplementary Table 9).

Despite the rather small deviations of data points from the power laws of $<R_g> = 0.26 \times N_{aa}^{0.52}$ and $<R_{2D}> = 1.16 \times N_{aa}^{0.52}$, it is advised to avoid generalization of these power laws or the mean L_p values of 0.78 nm in solution and 1.18 nm on mica to all naturally occurring IDRs in their fully disordered state. Indeed, they cannot be applied to IDRs with a high charge density, a large fraction of charged residues (FCR) or largely segregated oppositely charged residues characterized by large values of κ (see Supplementary Note 6 and Extended Data Fig. 9). The global structure of IDRs has been indeed shown to be modulated by electrostatic effects, i.e. a high charge density swells the volume of an IDR[11,50], whereas a combination of large FCR and large κ values results in protein compaction[51,52]. All IDPs randomly chosen in this study do not possess high charge densities or large values of FCR and κ (Supplementary Table 10); their charge densities (ρ) range between –0.087 and +0.038. Small charge densities $|\rho| \leq 0.1$ are often seen in numerous naturally occurring IDRs (50.9% of all IDRs deposited in the Database of Protein Disorder; Extended Data Fig. 9). Although the a.a. sequence and composition of an IDR modulate its conformational sampling at various extents, we resorted to use the power law and the mean L_p value on mica to make a rough estimate of the number of a.a. (N_{aa}) encompassed by the fully disordered IDRs contained in PNT and Sic1. The delineation from this N_{aa} estimation of the global structure of PNT and its transitions appear to be consistent with the domain diagram given by the recent NMR analysis of PNT[36] and with its proteolytic pattern[37]. In the case of Sic1, the reliability of our quantitative estimation of its global structure and transitions cannot be assessed because structural information is currently limited. With the above-mentioned caveats, the power laws and mean L_p values on mica and in solution obtained above appear to be useful

for semi-quantitative delineation of the structure and dynamics of IDPs, especially taking into account the current paucity of more useful means for this analysis.

Conclusion

HS-AFM imaging of IDPs on mica can distinguish fully and partially disordered regions as well as fully ordered regions at the near residue level, without fragmentation of the entire chain into shorter pieces. Moreover, HS-AFM imaging can provide not only length and height distributions for distinct regions in an IDP but also information on the structural nature of the transiently folded structures (e.g., α-helix and flattened, collapsed structure) and their order propensity and transition dynamics. Thus, HS-AFM imaging is an extremely powerful technique for characterizing IDPs, in particular allowing resolving conformational heterogeneity and avoiding the complications arising from ensemble averaging. Thus, when all molecular features revealed by HS-AFM are combined with the folded local structure given by NMR, the combined information allows a quantitative delineation of the structural and dynamic characters of IDPs, in a more realistic manner compared to the pictures depicted individually, as demonstrated for PNT.

Methods

Methods and any associated references are available in the online version of the paper.

Data availability

All data that support the findings of this study have been included in the main text, Extended Data Figures and Supplementary Information. The original source data associated with all figures are attached to respective figures.

Code availability

The original code for AFM image analyses is opened to the public at the following web site: https://elifesciences.org/content/4/e04806/article-data#fig-data-supplementary-material

Additional information

Supplementary information and videos are available in the online version of the paper. Reprints and permission information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to T. Ando or S. Longhi.

REFERENCES

1. Romero, P. et al. Thousands of proteins likely to have long disordered regions. Pac. Symp. Biocomp. 3, 437−448 (1998).

2. Uversky, V. N., Gillespie, J. R. & Fink A. L. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins 41, 415−427 (2000).

3. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signaling and regulation. Nat. Rev. Mol. Cell Biol. 16, 18−29 (2015).

4. Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. Chem. Rev. 104, 3607−3622 (2004).

5. Jensen, M. R., Zweckstetter, M., Huang, J. & Blackledge, M. Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy. Chem. Rev. 114, 6632–6660 (2014).

6. Kikhney, A. G. & Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. FEBS Lett. 589, 2570–2577 (2015).

7. Dedmon M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M., Dobson, C. M. Mapping long-range interactions in alphasynuclein using spin-label NMR and ensemble molecular dynamics simulations. J. Am. Chem. Soc. 127, 476–477 (2005).

8. Henriques, J., Cragnell, C. & Skepö, M. Molecular dynamics simulations of intrinsically disordered proteins: force field evaluation and comparison with experiment. J. Chem. Theory. Comput. 11, 3420-3431 (2015).

9. Bernadó, P. et al. A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. Proc. Natl. Acad. Sci. U. S. A. 102,17002–17007 (2005).

10. Ozenne, V. et al. Flexible-meccano: A tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28,1463-1470 (2012).

11. Schuler, B., Soranno, A., Hofmann, H. & Nettels, D. Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. Annu. Rev. Biophys. 45, 207–231 (2016).

12. Borgia, A., et al. Consistent view of polypeptide chain expansion in chemical denaturants from multiple experimental methods. J. Am. Chem. Soc. 138, 11714–11726 (2016).

13. Fuertes, G. et al. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. Proc. Natl. Acad. Sci. U. S. A. 114, E6342–E6351 (2017).

14. Ando, T. et al. A High-speed atomic force microscope for studying biological macromolecules. Proc. Natl. Acad. Sci. U. S. A. 98, 12468–12472 (2001).

15. Ando, T., Uchihashi, T. & Fukuma, T. High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. Prog. Surf. Sci. 83, 337–437 (2008).

16. Miyagi, A. et al. Visualization of intrinsically disordered regions of proteins by high-speed atomic force microscopy, Chem. Phys. Chem. 9, 1859–1866 (2008).

17. Hashimoto, M. et al. Phosphorylation-coupled intramolecular dynamics of unstructured regions in chromatin remodeler FACT. Biophys. J. 104, 2222–2234 (2013).

18. Uchihashi, T., Iino, R., Ando, T. & Noji, H. High-speed atomic force microscopy reveals rotary catalysis of rotorless F_1-ATPase. Science 333, 755–758 (2011).

19. Kodera, N., Yamamoto, T., Ishikawa, R. & Ando, T. Video imaging of walking myosin V by high-speed atomic force microscopy. Nature 468, 72–76 (2010).

20. Ando, T., Uchihashi, T. & Scheuring, S. Filming biomolecular processes by high-speed atomic force microscopy. Chem. Rev. 114, 3120–3188 (2014).

21. Ando, T. High-speed atomic force microscopy. Curr. Opin. Chem. Biol. 51, 105–112 (2019).

22. Takahashi, M. et al. Polyglutamine tract binding protein-1 is an intrinsically unstructured protein. Biochim. Biophys. Acta 1794, 936–943 (2009).

392 23. de Gennes, P. –G. Scaling concepts in polymer physics. Cornell University Press, Ithaca and
393 London (1979).

394 24. Frontali, C. Excluded-volume effect on the bidimensional conformation of DNA molecules
395 adsorbed to protein films. Biopolymers 27, 1329–1331 (1988).

396 25. Pérez, J., Vachette, P., Russo, D., Desmadril, M. & Durand, D. Heat-induced unfolding of
397 neocarzinostatin, a small all-beta protein investigated by small-angle X-ray scattering. J. Mol.
398 Biol. 308, 721–743 (2001).

399 26. Iwao, T. Polymer Solutions: An Introduction to Physical Properties, John Wiley & Sons, Inc.,
400 New York (2002).

401 27. Kirk, J. & Ilg, P. Chain dynamics in polymer melts at flat surfaces. Macromolecules 50,
402 3703−3718 (2017).

403 28. Yamamoto, H. et al. The intrinsically disordered protein Atg13 mediates supramolecular
404 assembly of autophagy initiation complexes. Dev. Cell 38, 86–99 (2016).

405 29. Mohan, A. et al. Analysis of molecular recognition features (MoRFs). J. Mol. Biol. 362,
406 1043–1059 (2006).

407 30. Jensen, M. R. et al. Intrinsic disorder in measles virus nucleocapsids. Proc. Natl. Acad. Sci. U.
408 S. A. 108, 9839–9844 (2011).

409 31. Gely, S. et al. Solution structure of the C-terminal X domain of the measles virus
410 phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the
411 nucleoprotein. J. Mol. Recognit. 23, 435–447 (2010).

412 32. Oldfield, C. J. et al. Coupled folding and binding with α-helix-forming molecular recognition
413 elements. Biochemistry 44, 12454–12470 (2005).

414 33. Morin, B. et al. Assessing induced folding of an intrinsically disordered protein by site-
415 directed spin-labeling electron paramagnetic resonance spectroscopy. J. Phys. Chem. B 110,
416 20596–20608 (2006).

34. Belle, V. et al. Mapping α-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. Proteins 73, 973–988 (2008).

35. Ormö, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y. & Remington, S. J. Crystal structure of the Aequorea victoria green fluorescent protein. Science 273, 1392–1395 (1996).

36. Milles, S. et al. An ultraweak interaction in the intrinsically disordered replication machinery is essential for measles virus function. Sci. Adv. 4, eaat7778 (2018).

37. Habchi, J., Mamelli, L., Darbon, H. & Longhi, S. Structural disorder within henipavirus nucleoprotein and phosphoprotein: From predictions to experimental assessment. PLoS One 5, e11684 (2010).

38. Brocca, S. et al. Compaction properties of an intrinsically disordered protein: Sic1 and its kinase-inhibitor domain. Biophys. J. 100, 2243–2252 (2011).

39. Riback, J. A. et al. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. Science 358, 238–241 (2017).

40. Kohn, J. E. et al. Random-coil behavior and the dimensions of chemically unfolded proteins. Proc. Natl. Acad. Sci. U. S. A. 101, 12491–12496 (2004).

41. Meier, S., Grzesiek, S. & Blackledge, M. Mapping the conformational landscape of urea-denatured ubiquitin using residual dipolar couplings. J. Am. Chem. Soc. 129, 9799–9807 (2007).

42. Le Guillo, J. C. & Zinn-Justin, J. Critical exponents for the n-vector model in three dimensions from field theory. Phys. Rev. Lett. 39, 95–98 (1977).

43. Cordeiro, T. N. et al. Structural characterization of highly flexible proteins by small-angle scattering. Adv. Exp. Med. Biol. 1009, 107–129 (2017).

44. Boze, H. et al. Proline-rich salivary proteins have extended conformations. Biophys. J. 99, 656–665 (2010).

45. Kate M. Nairn, K. M. et al. A synthetic resilin is largely unstructured. Biophys. J. 95, 3358–3365 (2008).

46. Salmon, L. et al. NMR characterization of long-range order in intrinsically disordered proteins. J. Am. Chem. Soc. 132, 8407–8418 (2010).

47. Mylonas, E. et al. Domain conformation of tau protein studied by solution small-angle X-ray scattering. Biochemistry 47, 10345–10353 (2008).

48. Lanza, D. C., Silva, J. C., Assmann, E. M., Quaresma, A. J., Bressan, G. C., Torriani, I. L. & Kobarg, J. Human FEZ1 has characteristics of a natively unfolded protein and dimerizes in solution. Proteins 74, 104–21 (2009).

49. Bernadó, P. & Blackledge, M. A. Self-consistent description of the conformational behavior of chemically denatured proteins from NMR and small angle scattering. Biophys. J. 97, 2839–2845 (2009).

50. Müller-Späth, S. et al. Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc. Natl. Acad. U. S. A. 107, 14609–14614 (2010).

51. Das, R. K., Huang,Y., Phillips, A. H., Kriwacki, R. W. & Pappu, R. V. Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signalling. Proc. Natl. Acad. Sci. U. S. A. 113, 5616–5621 (2016).

52. Sherrya, K. P., Das. R. K., Pappu, R. V. & Barricka, D. Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. Proc. Natl. Acad. Sci. U. S. A. 114, E9243–E9252 (2017).

ACKNOWLEDGMENTS

We thank Ms. Maya Hakozaki (Kanazawa University) for her technical assistance, Dr. Hitoshi Okazawa (Tokyo Medical and Dental University) for providing the human PQBP-1 cDNA, and Antoine Schramn (Lab. AFMB UMR 7257, Aix-Marseille University and CNRS) for providing the statistics of IDRs deposited in the DisProt database. This work was supported by grants from the MEXT, Japan, Grants-in-Aid for Scientific Research on Innovative Areas: Research in a

Proposed Research Area, #21113002 and #26119003 (T.A.), #21113003 (M.M.), #15H01651 (Y.F.) and #25111004 (N.N.N.), Grant-in-Aid for Basic Research S, #24227005 and #17H06121 (T.A.), CREST program of JST (T.A. and N.N.N.) and PRESTO program of JST (N.K.), and partly supported by grants from the Agence Nationale de la Recherche, specific programs 'Microbiologie et Immunologie' ANR-05-MIIM-035-02 (S.L.) and 'Physico-Chimie du Vivant' ANR-08-PCVI-0020-01 (S.L.). D.B. was supported by a joint doctoral fellowship from the Direction Générale de l'Armement (DGA) and the CNRS. E.S. was supported by a joint doctoral fellowship from the Direction Générale de l'Armement (DGA) and Aix-Marseille Université. M.D. was supported by a PhD fellowship from the French Ministry of National Education, Research and Technology.

AUTHOR CONTRIBUTIONS

S.L. and T.A. designed the project. M.L. provided the original constructs encoding GFP fusions. M.M, D.B., J.H., A.G, M.D., E.S., C.B., Y.F. and N.N.N. prepared constructs and/or protein samples used in this study. T.A. and N.K. developed the HS-AFM system. N.K., D.N., S.K.D. and T.M. performed the HS-AFM experiments. N.K., T.A., D.N., S.K.D. and T.M. analyzed HS-AFM data. T.O. and M.S. performed the SAXS experiments and analyzed SAXS data. T.A. made all theoretical formulations and wrote the draft of the manuscript. S.L., N.K. and T.A. prepared the final manuscript based on the discussions performed among all authors.

Competing interests

There is no financial competing interest.

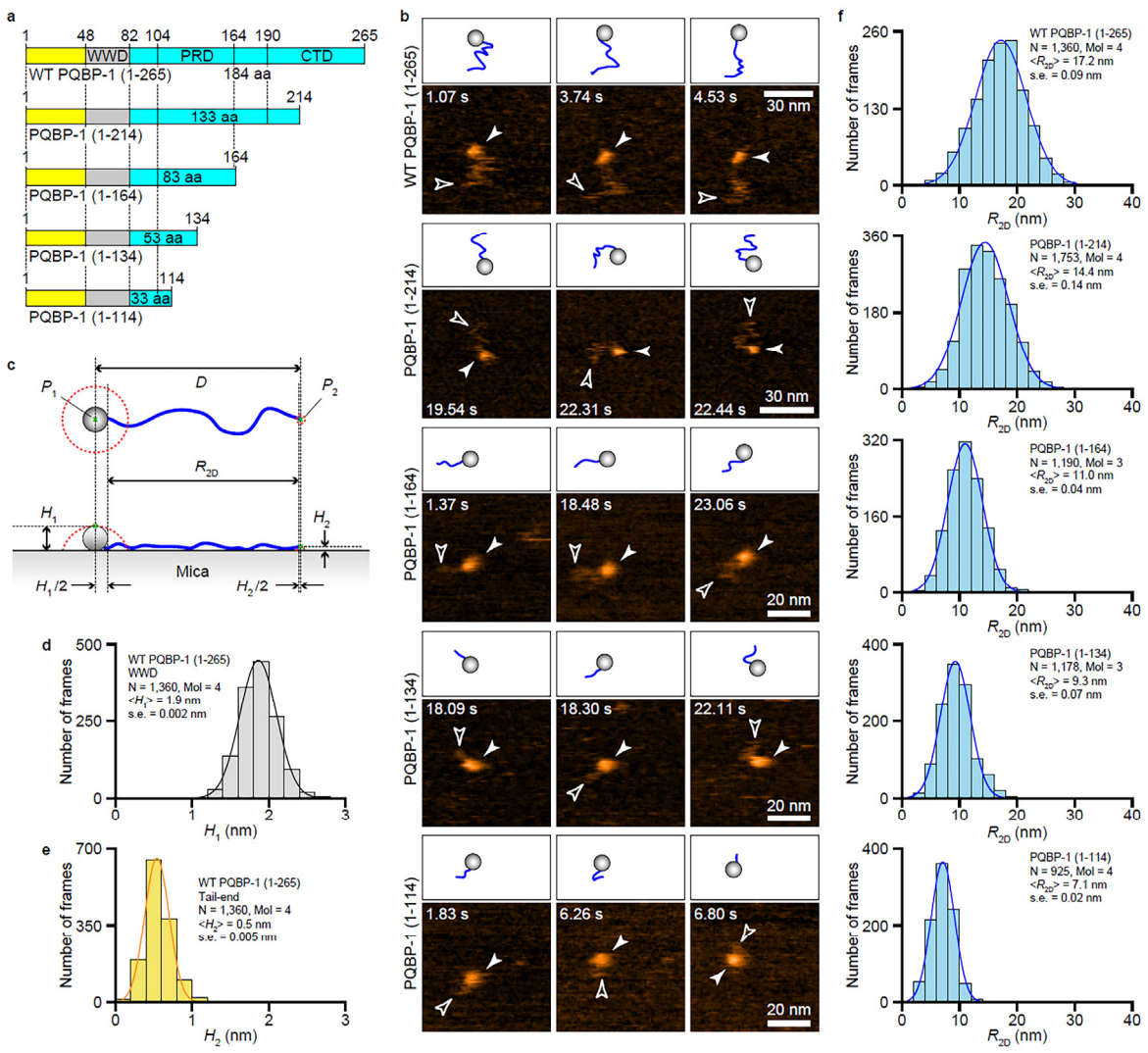


Figure 1 | Structural features of PQBP-1 constructs revealed by HS-AFM imaging. a, Domain diagrams of five PQBP-1 constructs containing IDRs (light blue), folded WWD (gray) and the N-terminal segment not investigated with NMR (yellow). **b**, HS-AFM images and schematized molecular features of PQBP-1 constructs. Z-scale, 0.0–3.5 nm. **c**, Schematics showing the observed molecular characteristics of the PQBP-1 constructs (Top, top-view; Bottom, side-view): Gray spheres, N-terminal globular domain; blue thick solid lines, IDRs; dashed red lines, schematic topographies of the globular domain convoluted with the finite size of the tip apex in lateral contact with the globule. **d**,**e**, Height distributions of globule (**d**) and tail end (**e**) in WT PQBP-1. **f**, R_{2D} distributions of tail-like segments in PQBP-1 constructs. Solid blue lines show most probable fitting curves. The values obtained by Gaussian distribution fitting are mean ± standard error (s.e.). This definition is the same for all Gaussian fitting results in this paper, unless otherwise stated. "Mol" in (**d**–**f**) indicates the number of molecules analysed. This is the same for **Figs. 3** and **4**.

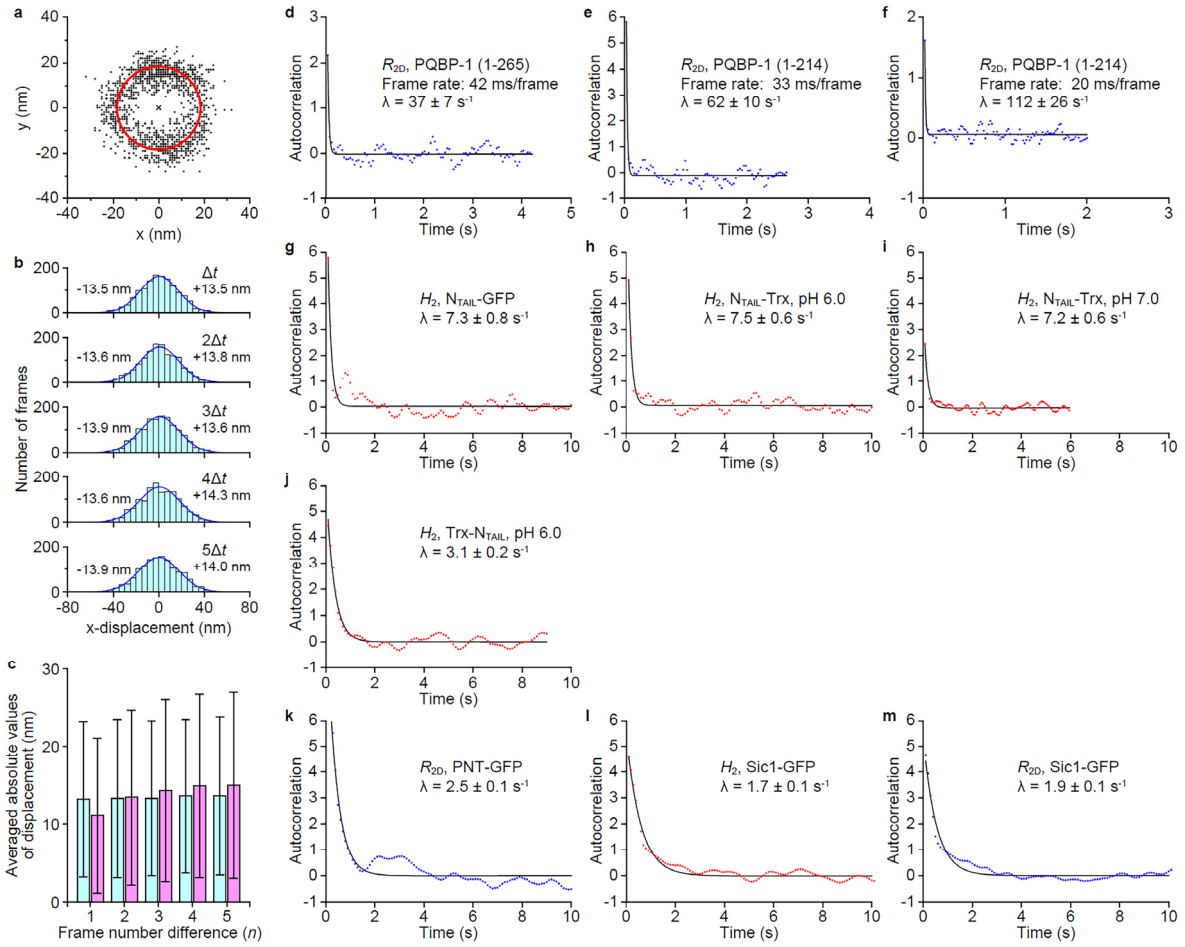


Figure 2 | Absence of directional bias by scanning tip and of significant impediment by tip and mica to diffusional motion of IDR, and autocorrelation functions for time-series $R_{2D}(t)$ or $H_2(t)$ data. a, Two-dimensional scatter plots of the C-terminal end position relative to the WWD (the cross mark) in WT PQBP-1. The red circle indicates the mean direct distance between the WWD and the C-terminal end (i.e., radius = $\langle R_{2D} \rangle$ + $\langle H_1 \rangle/2$ + $\langle H_2 \rangle/2$). **b**, Distributions of C-terminal end displacements in the X-direction occurred during time periods of $n\Delta t$ (n = 1, 2, 3, 4, 5), where Δt is one frame acquisition time (42 ms). The numbers shown on each panel are the mean displacements in the +X and –X directions. **c**, Averaged absolute values of displacements in the X- (light blue) and Y- (pink) directions, during time periods of $n\Delta t$. The thin vertical lines represent s.d. The large s.d. values reflect the flexible nature of IDR. **d–m**, Autocorrelation functions of time-series data of $R_{2D}(t)$ (**d–f**, **k**, **m**) for IDRs contained in PQBP-1 (1–265) (**d**), PQBP-1 (1–214) (**e**, **f**), PNT-GFP (**k**) and Sic1-GFP (**m**), and autocorrelation functions of $H_2(t)$ data (**g–j, l**) for Box1 in N_{TAIL}-GFP (**g**), N_{TAIL}-Trx at pH 6.0 (**h**) and N_{TAIL}-Trx at pH 7.0 (**i**), Box2 in Trx-N_{TAIL} at pH 6.0 (**j**), and the N-terminal collapsed structure in Sic1-GFP (**f**). In the case of the constantly disordered IDRs in the PQBP-1 constructs, the characteristic time for the C-terminal end movement relative to the N-terminus is much shorter than the one frame acquisition time (20 ms).

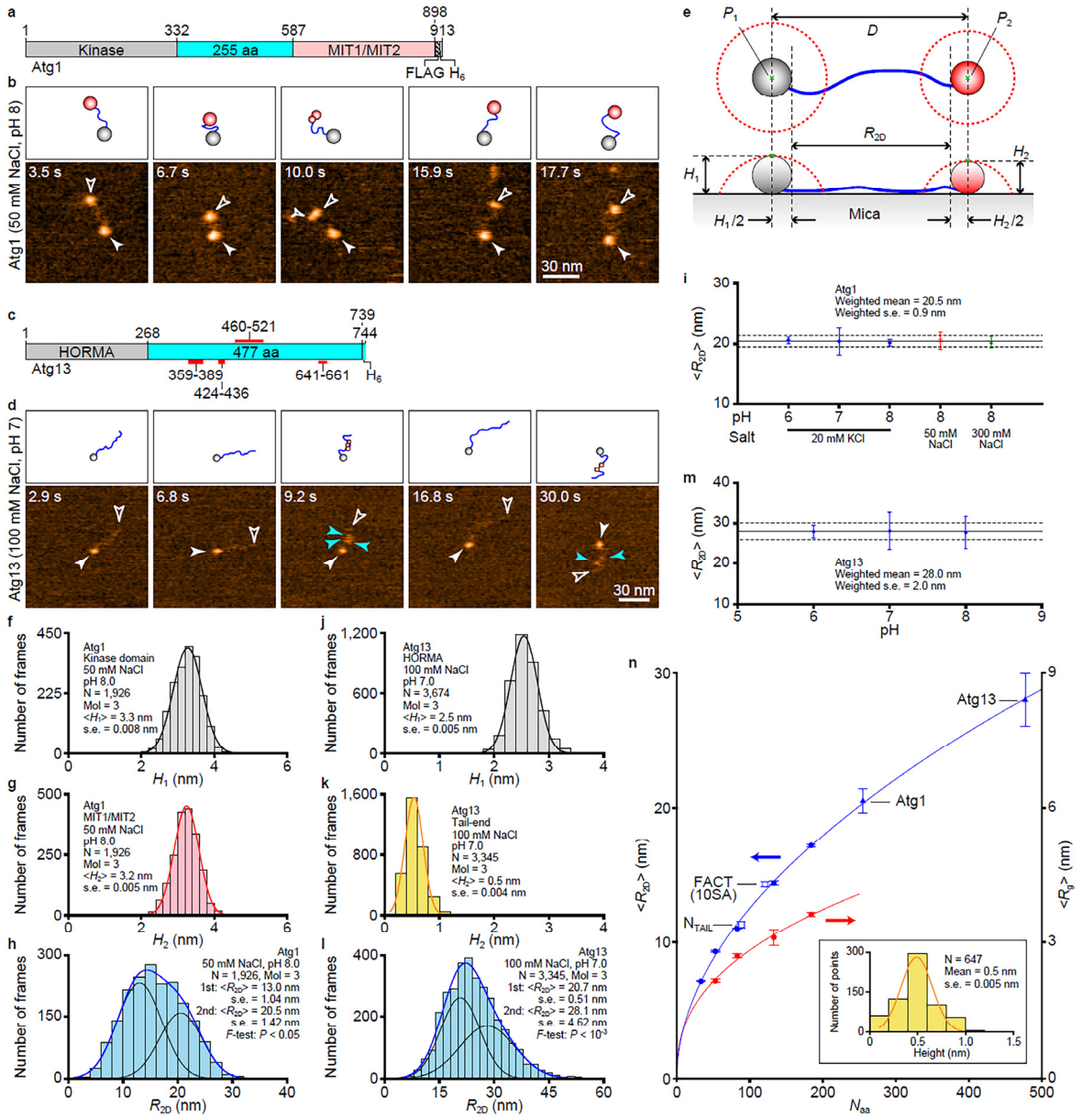


Figure 3 | Structural features of Atg1 and Atg13 revealed by HS-AFM imaging and $<R_{2D}>$–N_{aa} and $<R_g>$–N_{aa} relationships. a,**c**, Domain diagrams of Atg1 (D211A) (**a**) and Atg13 (**c**); the top and bottom red bars in **c** are Atg1- and Atg17-binding regions, respectively. **b**,**d**, HS-AFM images and their schematics of Atg1 (D211A) (**b**) and Atg13 (**d**). Z-scale, 0.0–3.5 nm. Light-blue arrow-heads (**d**, 9.2 s and 30.0 s) indicate temporarily appearing small globules. **e**, Schematics of observed molecular features of Atg1 (D211A) (Top, top-view; Bottom, side-view). **f**,**g**,**j**,**k**, Height distributions of kinase (**f**), MIT1/MIT2 (**g**) and HORMA (**j**) domains and C-terminal end of Atg13 (**k**). **h**,**l**, R_{2D} distributions of Atg1 (D211A) (**h**) and Atg13 (**l**). **i**, Dependence of $<R_{2D}>$ in the longer state of Atg1 (D211A) on pH and salt type and concentration. **m**, pH dependence of $<R_{2D}>$ in the longer state of Atg13. **n**, $<R_{2D}>$–N_{aa} relationship (blue line) from $<R_{2D}>$ data of PQBP-1 (closed blue circles), FACT (IDR in the SSRP1 subunit)[21], N_{TAIL}, Atg1 (D211A) and Atg13, and $<R_g>$–N_{aa} relationship from SAXS data of IDRs in PQBP-1 (closed red circles). The $<R_{2D}>$ values for the Atg proteins are those in the longer states. Inset: height distribution of constantly disordered IDR in WT PQBP-1.

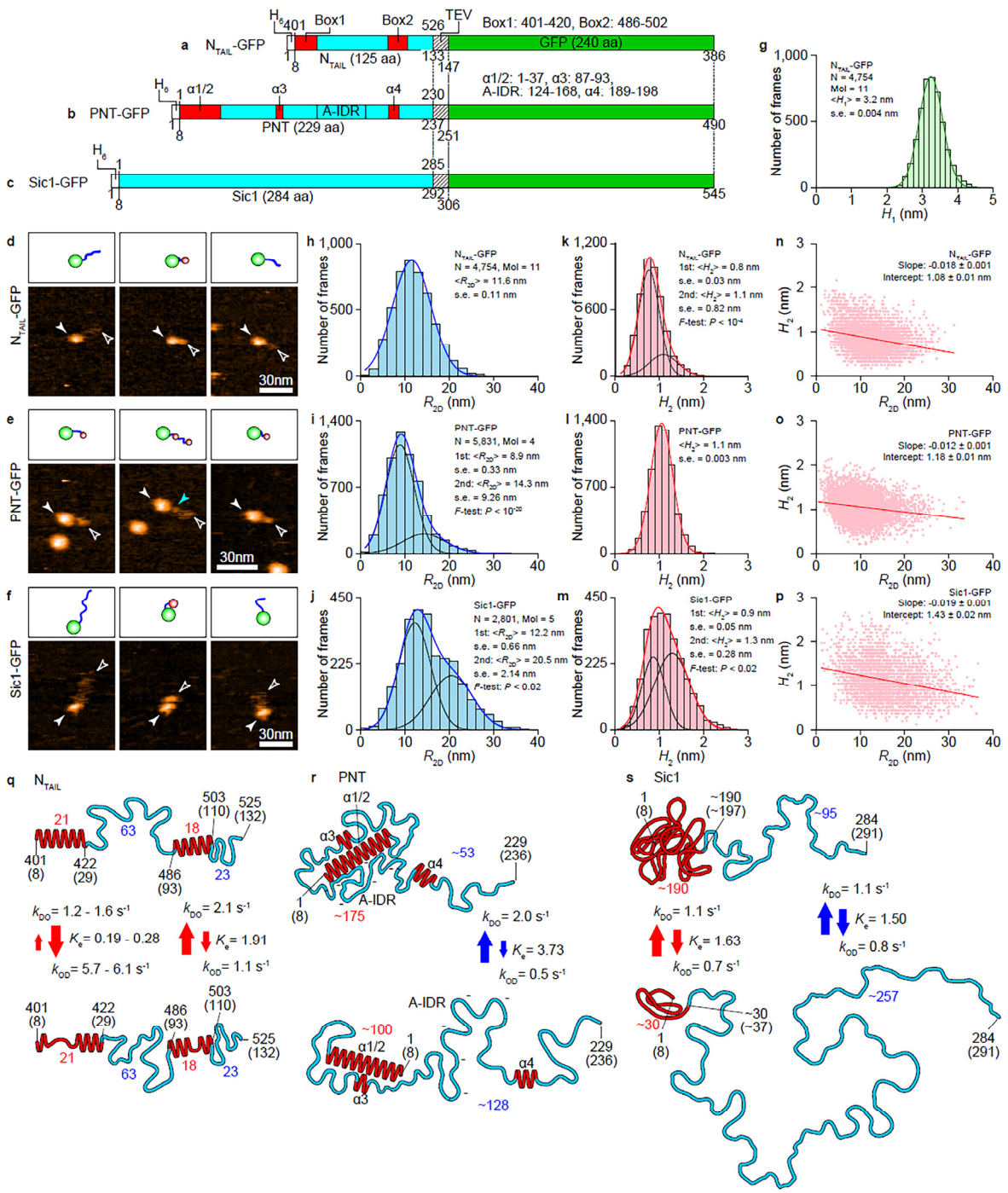
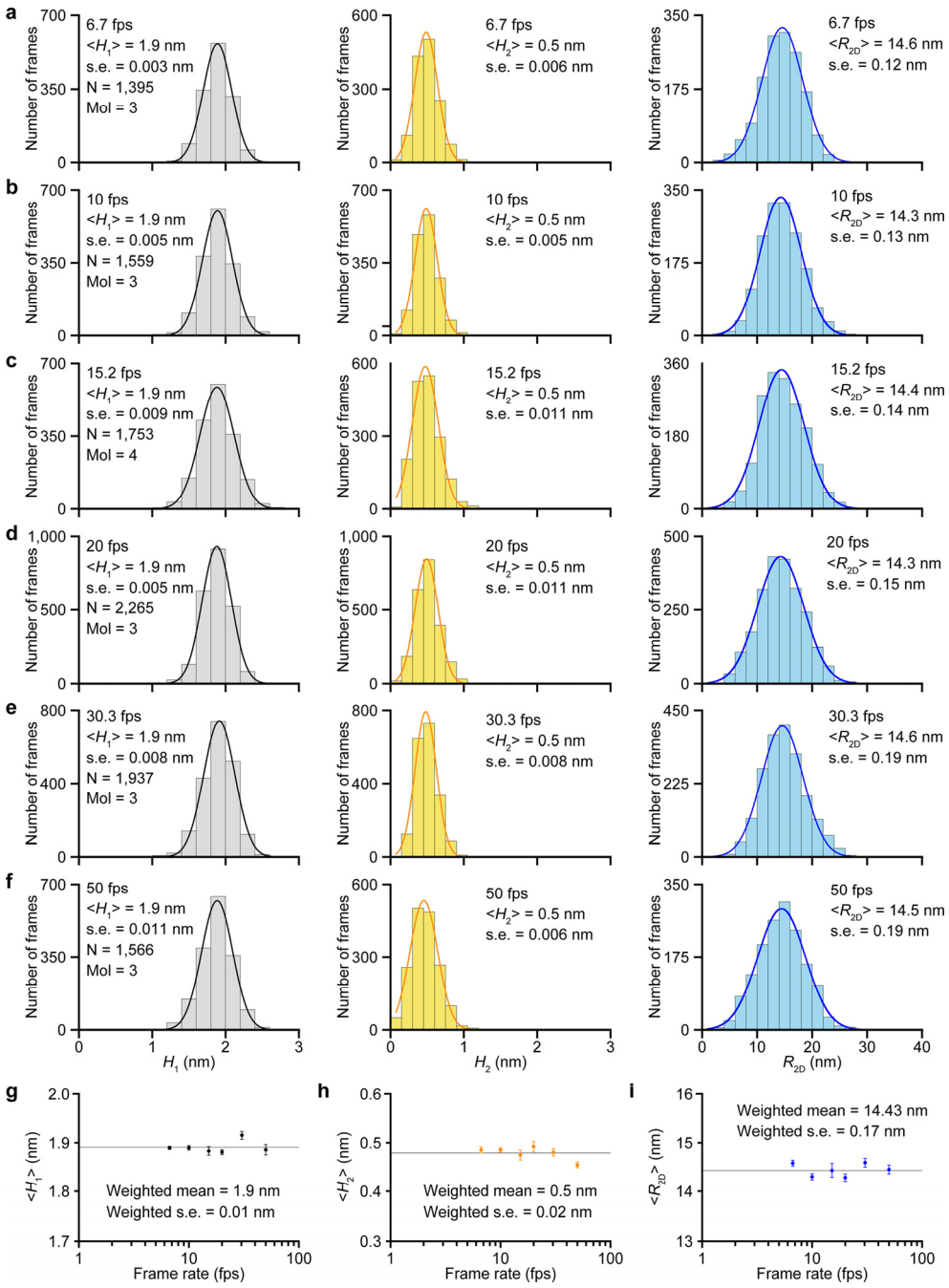
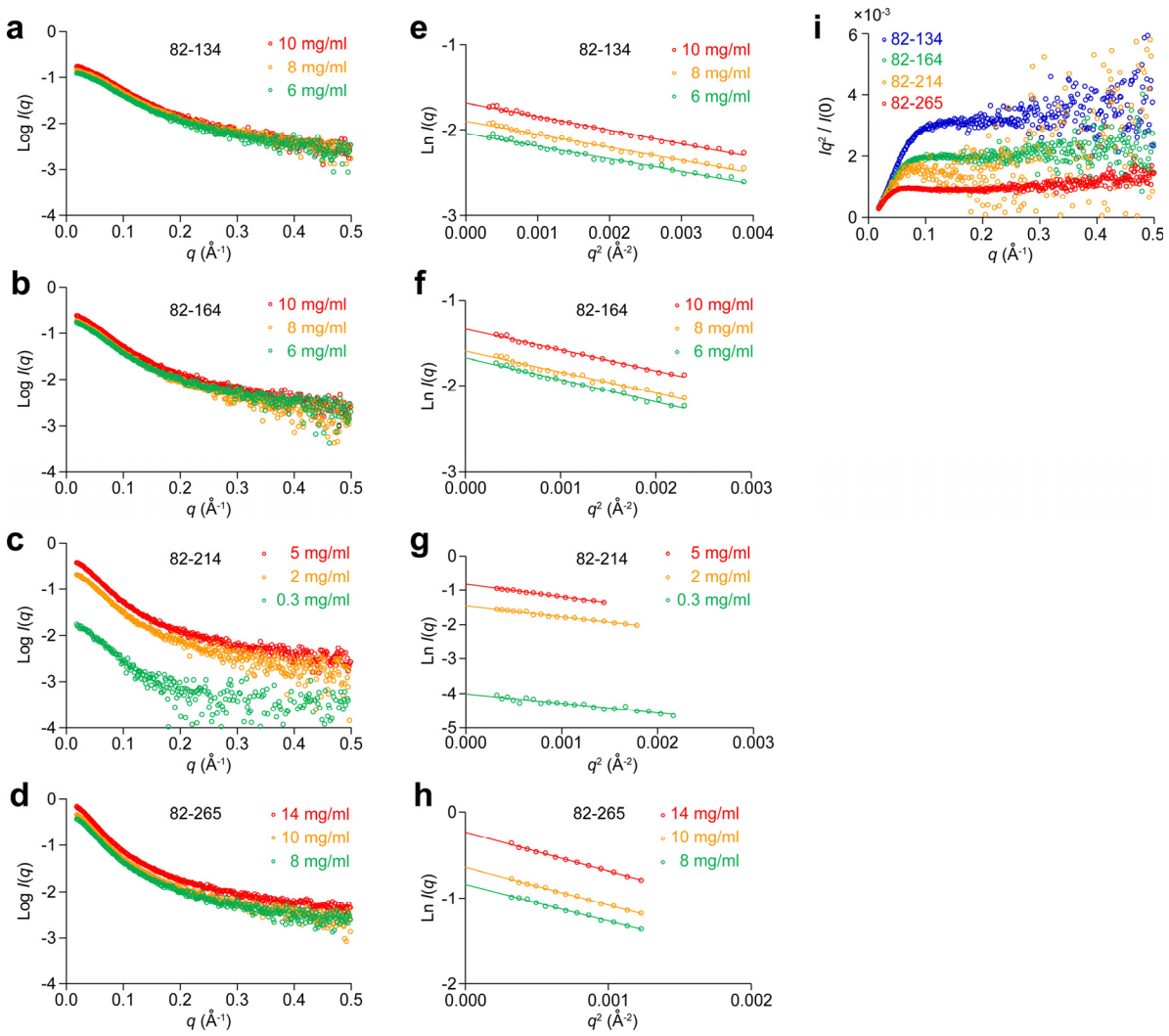


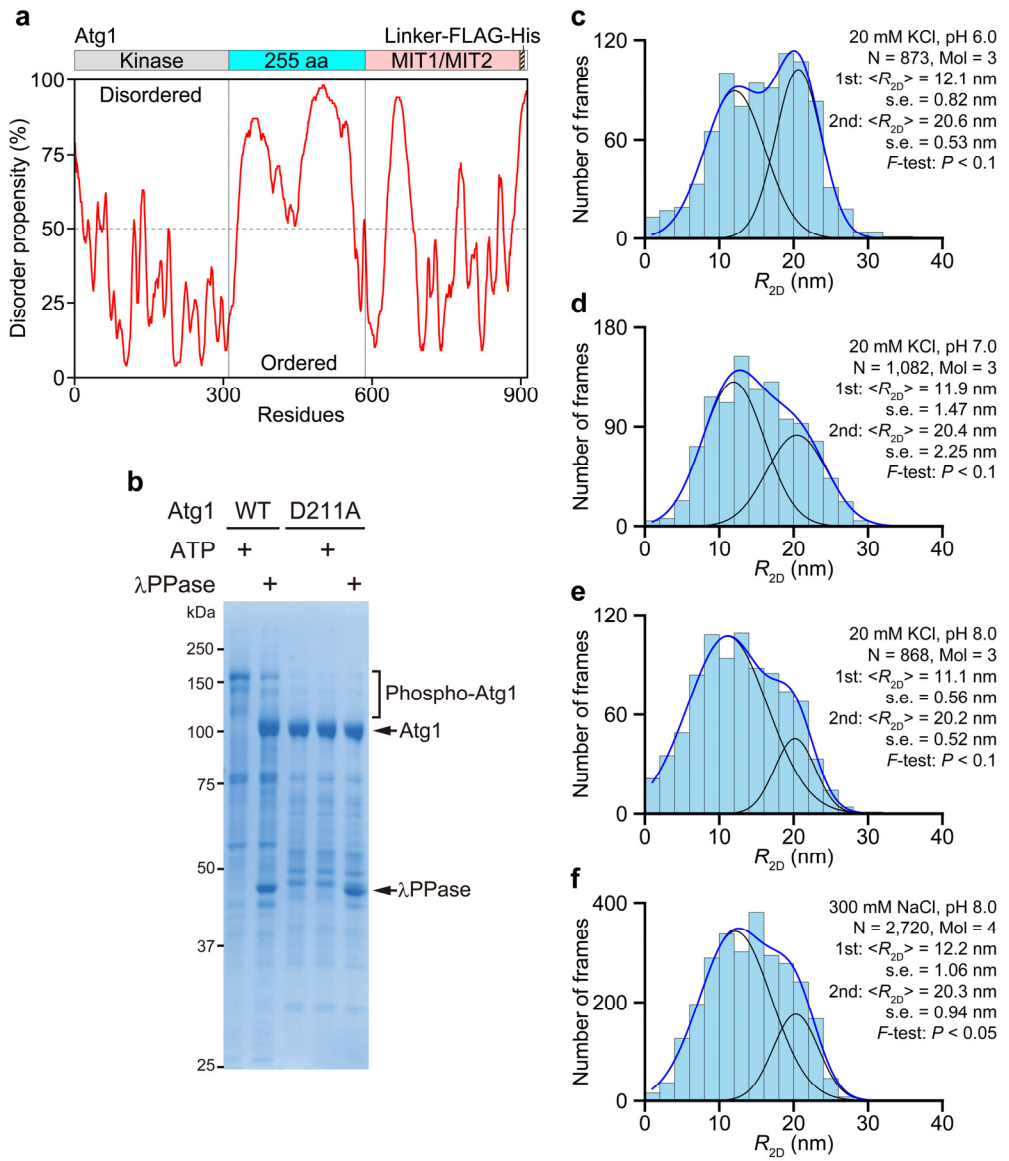
Figure 4 | Structural and dynamic features of N_{TAIL}, PNT and Sic1 revealed by HS-AFM imaging. **a**–**c**, Domain diagrams of three IDP-GFP fusions. The numbers (top and bottom) are original a.a. sequential numbers and those of the constructs, respectively. **d**–**f**, HS-AFM images and their schematics of three IDP-GFP fusions. Z-scale, 0.0–3.5 nm. The light-blue arrow-head (**e**, middle) indicates a rarely appearing tiny globule, likely to be α_4. **g**, Height distribution of GFP in N_{TAIL}-GFP. **h**–**j**, R_{2D} distributions of N_{TAIL}-GFP (**h**), PNT-GFP (**i**) and Sic1-GFP (**j**). **k**–**m**, H_2 distributions of N-terminal globules in N_{TAIL}-GFP (**k**), PNT-GFP (**l**) and Sic1-GFP (**m**). **n**–**p**, Relationship between H_2 and R_{2D}. **q**–**s**, Schematics showing structural and dynamic features of three IDPs. The top and bottom panels correspond to the more-ordered and less ordered states, respectively. The numbers in red and blue represent the numbers of a.a. contained in the respective small globules and fully disordered IDRs, respectively. The red and blue arrows indicate the kinetic nature of changes in H_2 and R_{2D}, respectively.



Extended Data Fig. 1 | Molecular features of PQBP-1 (1−214) observed at various imaging rates. **a**–**f**, H_1, H_2 and R_{2D} distributions measured from HS-AFM images captured at 6.7 (**a**), 10.0 (**b**), 15.2 (**c**), 20.0 (**d**), 30.3 (**e**) and 50.0 (**f**) fps. The most probable fitting curves are drawn with the solid lines. **g**, $<H_1>$ at various imaging rates. **h**, $<H_2>$ at various imaging rates. **i**, $<R_{2D}>$ at various imaging rates. Each plot corresponds to mean values ± s.e. measured at each frame rate. The horizontal lines indicate the weighted mean values. Details of these analyses are summarized in **Supplementary Table 2**. Note that the intermittent tip-sample contact force becomes larger with increasing imaging rate, whereas the number of contacts *per* frame increases with decreasing imaging rate. Nevertheless, the values of $<R_{2D}>$, $<H_1>$ and $<H_2>$ are nearly constant, irrespective of the imaging rate, indicating no significant impact of the tip-sample contact on the structure of this protein.

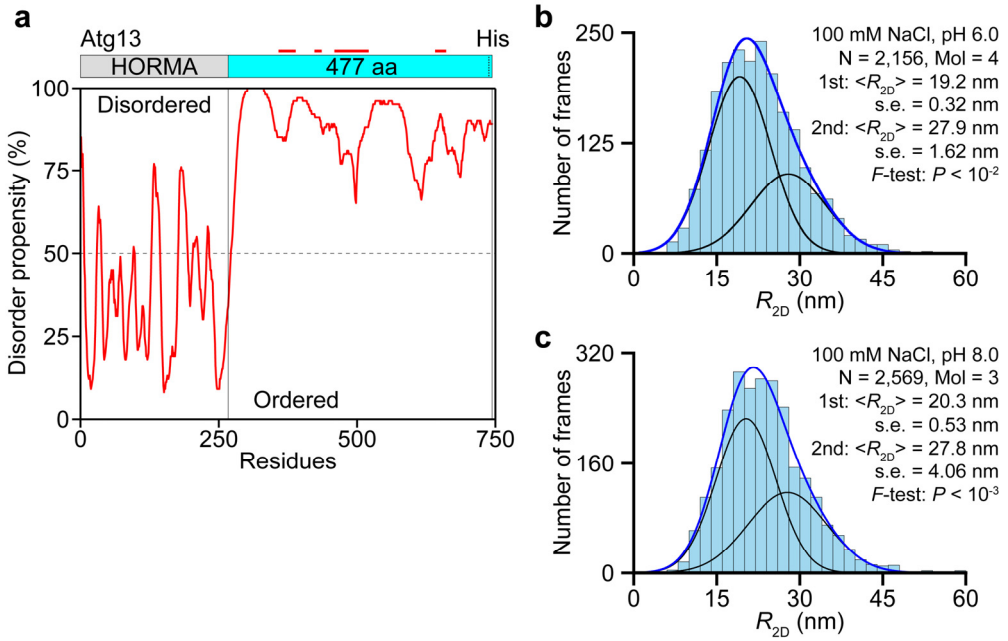


Extended Data Fig. 2 | *I*(*q*) curves, Guinier plots, and Kratky plots obtained from SAXS measurements of IDR segments of PQBP-1 (82−134, 82−164, 82−214 and 82−265). a−d, *I*(*q*) curves displayed in a *q* range from 0.018 to 0.500 Å^{-1}. **e−h**, Guinier plots displayed in the smaller region of $qR_g < 1.3$. **i**, Kratky plots normalized by *I*(0) indicating the fully disordered nature of the four IDRs. A summary of these analyses is presented in **Supplementary Table 3**.



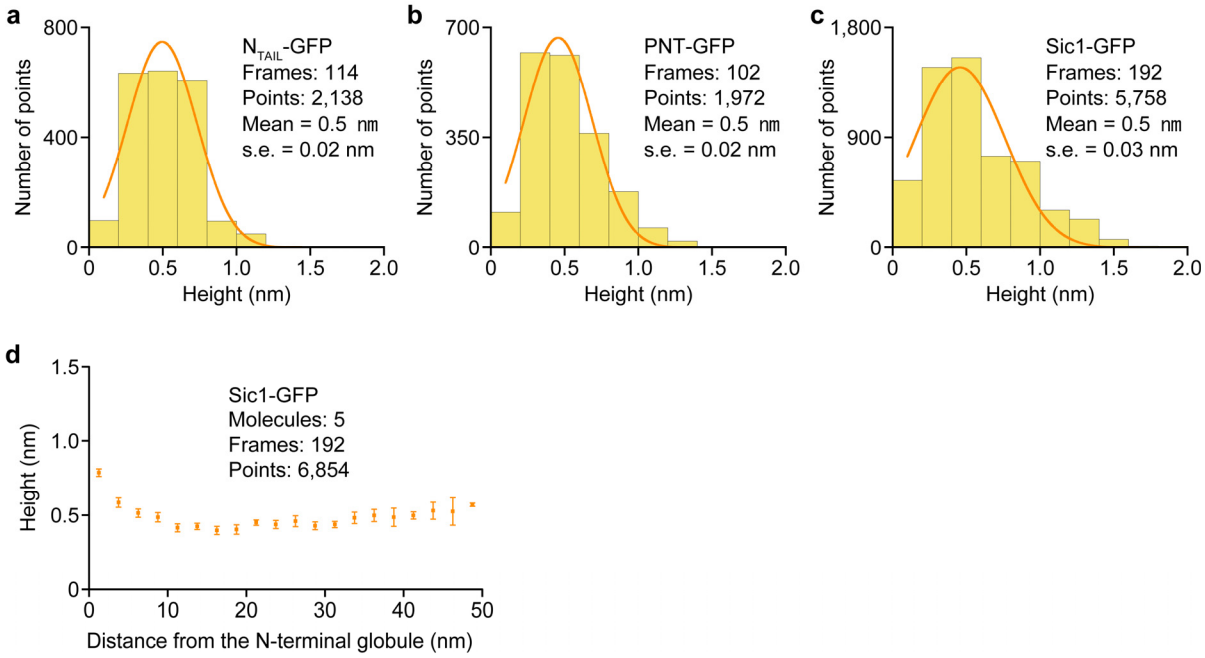
Extended Data Fig. 3 | Characterization of Atg1 (D211A) and R_{2D} distributions of its IDR measured under various solution conditions. a, Domain diagram of the Atg1 (D211A) construct (light blue, IDR) and its order/disorder map along its length as predicted by DISPROT (VSL2). **b**, Coomassie Brilliant Blue-stained SDS-PAGE (10%) of WT Atg1 and Atg1 (D211A) showing autophosphorylation of WT Atg1 and no phosphorylation of Atg1 (D211A). The autophosphorylation reaction was performed by incubating 2 μM Atg1 with 2 mM ATP and 5 mM $MgCl_2$ in 20 mM Tris-HCl, pH 8.0, 150 mM NaCl buffer. The dephosphorylation reaction was performed by incubating 2 μM Atg1 with 1 μM lambda protein phosphatase (λPPase), 2 mM dithiothreitol and 1 mM $MnCl_2$ in the same buffer. These reactions were terminated by the addition of Laemmli SDS sample buffer. **c-f**, R_{2D} distributions of IDR measured under various solution conditions. The most probable fitting curves are shown with the solid blue lines. The black lines represent the Gaussian components in double-Gaussian fitting. Note that the $<R_{2D}>$ values at the second peaks (longer states) are nearly identical irrespective of the solution condition, whereas the $<R_{2D}>$ values at the first peaks (shorter states) vary depending on the solution condition. A summary of these analyses is presented in **Supplementary Table 4**.

514
515

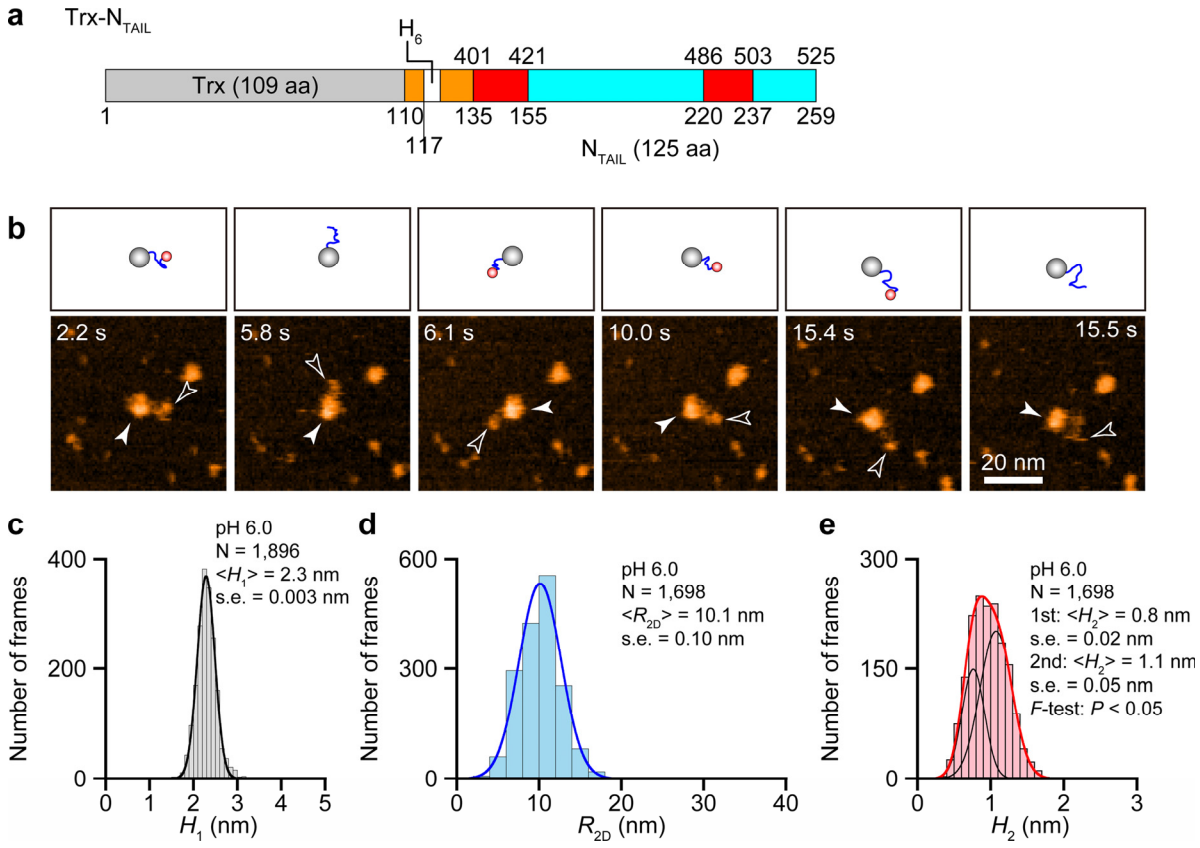


Extended Data Fig. 4 | Order/disorder map along the length of Atg13 and R_{2D} distributions of its IDR measured at two pH values. a, Domain diagram of the Atg13 construct used in this study and order/disorder map along its length as predicted by DISPROT (VSL2). The regions indicated by the red bars are 359–389, 424–436 and 641-661 (Atg17-binding regions) and 460–521 (Atg1-binding region). **b**,**c**, R_{2D} distributions of IDR measured at pH 6.0 (**b**) and pH 8.0 (**c**). The most probable fitting curves are shown with the solid blue lines. The black lines represent the Gaussian components in double-Gaussian fitting. Note that the $<R_{2D}>$ values at the second peaks (longer states) are nearly identical irrespective of the solution condition, whereas the $<R_{2D}>$ values at the first peaks (shorter states) slightly vary depending on pH. A summary of these analyses is presented in **Supplementary Table 5**.

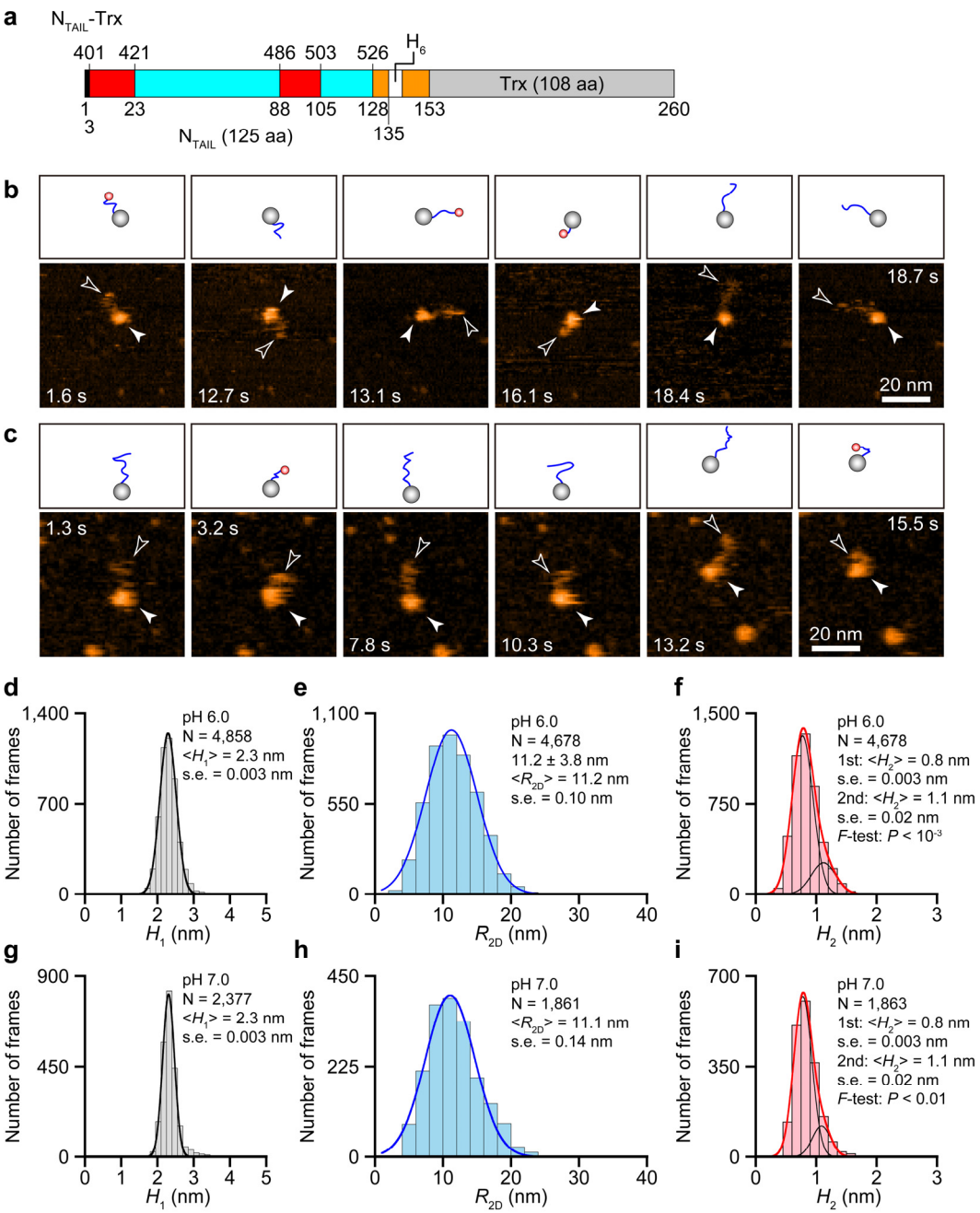
516
517



Extended Data Fig. 5 | Height of fully disordered IDRs contained in three IDP-GFP fusions. To make sure that the IDRs under analysis are fully disordered, HS-AFM images showing longer IDRs were chosen. **a**, N_{TAIL}-GFP. **b**, PNT-GFP. **c**, Sic1-GFP. **d**, Height of IDR in Sic1-GFP as a function of its distance from the N-terminus. AFM images of Sic1-GFP molecules showing IDR longer than 25 nm were chosen in this height analysis to ensure that the IDR under analysis was formed upon order-to-disorder transition of the N-terminal small globule. The height of IDR was measured as a function of the distance from the N-terminal globule with a bin width of 2.5 nm. The height values for different frames and molecules measured within the bin width at each lateral distance were averaged and the mean height was plotted together with s.e. The mean height of the N-terminal end (0.8 nm) is slightly smaller than the height of peak 1 (0.9 nm) shown in **Fig. 4m** in the main text. This is due to the relatively large bin width (2.5 nm) of the lateral distance used in this analysis.

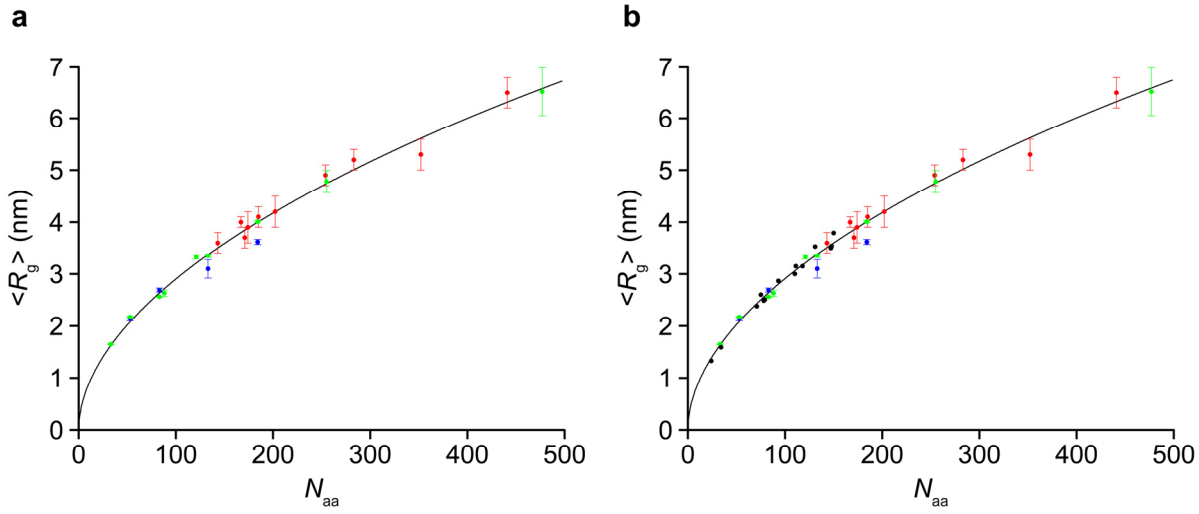


Extended Data Fig. 6 | Structural features of Trx-N_{TAIL} observed by HS-AFM imaging. a, Domain diagram of Trx-N_{TAIL} (the a.a. sequence is given in **Supplementary Fig. 1**). Color codes: Gray, Trx; black, Met and Ser; orange, linkers; white, His_6; red, α-More; cyan, IDR. **b**, Successive HS-AFM images of Trx-N_{TAIL} molecules observed at pH 6.0. The molecular features of observed molecules are schematized just above the respective AFM images. The closed and open arrow heads point at the Trx moiety and the tail end, respectively. **c**, Height distribution of the Trx domain. **d**, R_{2D} distribution of the IDR. **e**, Height distribution of the C-terminal globule (Box2). The most probable fitting curves are shown with the thick solid lines. The thin black lines in (**e**) indicate two Gaussian components in double-Gaussian fitting. A summary of these analyses is presented in **Supplementary Table 7**.



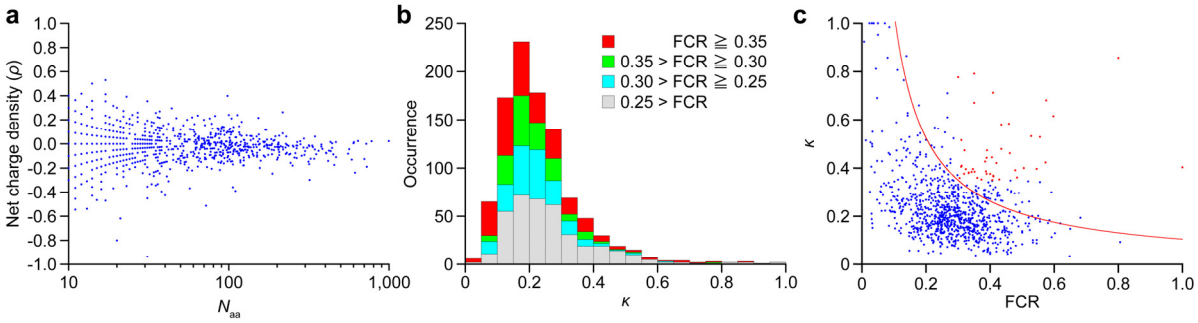
Extended Data Fig. 7 | Structural features of N_{TAIL}-Trx observed by HS-AFM imaging. a, Domain diagram of N_{TAIL}-Trx (the a.a. sequence is given in **Supplementary Fig. 1**). Color code: same as in **Extended Data Fig. 6**. **b**,**c**, Successive HS-AFM images of N_{TAIL}-Trx molecules observed at pH 6.0 (**b**) and pH 7.0 (**c**). The closed and open arrow heads point at the Trx moiety and the tail end, respectively. **d**,**g**, Height distributions of the Trx domain measured at pH 6.0 (**d**) and pH 7.0 (**g**). **e**,**h**, R_{2D} distributions measured at pH 6.0 (**e**) and pH 7.0 (**h**). **f**,**i**, Height distributions of the N-terminal small globule (Box1) measured at pH 6.0 (**f**) and pH 7.0 (**i**). The most probable fitting curves are shown with thick solid lines (**d**–**i**). The thin black lines in (**f**,**i**) are two Gaussian components in double-Gaussian fitting. A summary of these analyses is presented in **Supplementary Table 7**.

522
523



Extended Data Fig. 8 | Refinement of expansion effect of mica on the 2D dimensions of IDR and power law for <R_g>. a, <R_g>–N_{aa} relationship for 23 data. The 10 data plotted with the red circles are those from the study by Mylonas *et al*. for tau protein constructs[47], with phosphor-mimic constructs and those significantly affected by the extended repeat domain[43,49] having been removed here. The constructs herein analyzed are tau ht40, tau K32, tau K16, tau ht23, tau K27, tau K27, tau K44, tau K10, tau K25, and tau K23. The data plotted with the blue circles are those from the present SAXS study for four segments of PQBP-1 IDR (82−265, 82−214, 82−164, and 82−134). The nine data plotted with the green circles are those of (N_{aa}, <R_{2D}>/($2\sqrt{3}u$) at u = 1.24) of the five PQBP-1 constructs, Atg1, Atg13, FACT protein and N_{TAIL} (<R_{2D}> data highlighted with orange color in **Supplementary Table 8** are used). The value of u were determined by fitting all 23 data to a power law, <R_g> = $\beta_g \times N_{aa}^{\nu}$, in which the parameter u was contained in the nine data as one of the variables to be determined. The black solid line indicates the best fitting result; u = 1.24, β_g = 0.26 nm, and ν = 0.52. **b**, <R_g>–N_{aa} relationship for 37 data. The 37 data points include 14 data found from literature search (black circles; **Supplementary Table 9**) and those 23 data shown in (**a**). Fitting of the 37 data points to a power law, <R_g> = $\beta_g \times N_{aa}^{\nu}$, yielded β_g = 0.260 ± 0.021 nm and ν = 0.524 ± 0.015 (black line).

524
525



Extended Data Fig. 9 | Statistics of charge-associated parameters for many naturally occurring IDRs. a, 2D plot of (N_{aa}, ρ) (ρ, charge density) of naturally occurring IDRs whose information is deposited in the Database of Protein Disorder that gathers information of IDRs that have been confirmed to be disordered. The paired values (N_{aa}, ρ) of 1,011 IDRs are plotted in this graph. Half of these data points (50.9%) are within the range of $|\rho| \leq 0.1$. Note that shorter IDRs are enriched in the database, reflecting the fact that short IDPs tend to be more experimentally characterized than longer IDRs. **b**, Histogram of κ for the IDRs contained in the database; IDRs with $N_{aa} \leq 19$ were omitted. Each bar of the histogram is divided into four regions according to the value of fraction of charged residues (FCR). The CIDER software program (opened to public use) was used to calculate the values of κ and FCR (see **Supplementary Note 6**). **c**, Scatter diagram of (FCR, κ) for the IDRs contained in the database; IDRs with $N_{aa} \leq 19$ were omitted. The points shown in red (4.5%) correspond to IDRs that meet the combined condition of $\kappa \geq 0.35$ & $FCR \geq 0.3$ that is considered to be required for the appearance of a distinct volume compaction effect of oppositely charged residues[52]. It would be plausible that a volume compaction effect could be manifest even for IDRs possessing a FCR value less than (for example) 0.25, when they have a significantly large value of κ. To depict such a possibility, the red line is drawn ($\kappa \times FCR = 0.35 \times 0.30$). It should be kept in mind however that this is just tentative and awaits future experimental confirmation.

Methods

Preparation of PQBP-1 constructs. The DNA encoding full-length PQBP-1 was inserted into a pGEX6P-1 plasmid that was used to transform the *E. coli* BL21 [DE3] strain. Protein expression was induced by the addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). PQBP-1 was expressed as a glutathione S-transferase (GST)-fusion protein. After cultivation, BL21 [DE3] cells were collected by centrifugation. The cells were resuspended in a buffer containing 50 mM sodium phosphate (pH 7.0) and 2 mM dithiothreitol. After cell lysis by sonication and removal of insoluble cell debris by centrifugation, the supernatant was applied to a glutathione sepharose 4B column equilibrated with 50 mM Tris/HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol. After washing the column with the same buffer, the GST-fusion protein was eluted from the column by adding 50 mM Tris/HCl (pH 8.5), 50 mM NaCl, and 10 mM reduced glutathione. To obtain PQBP-1, the GST fusion protein was cleaved with a recombinant rhinovirus 3C protease (PreScission protease; GE Healthcare). The cleavage mixture was passed through the glutathione sepharose 4B column to remove GST and 3C protease. PQBP-1 with shorter IDRs and four IDR segments were obtained in the same manner as the full-length PQBP-1. These constructs including WT PQBP-1 were finally purified by reverse phase high-performance liquid chromatography using a 5C8-AR-300 column or a Protein-R column (Nacalai Tesque), followed by dialysis against 20 mM sodium phosphate (pH 7.0). The molecular mass of PQBP-1 and its fragments was confirmed by matrix-assisted laser desorption/ionization time-of-flight mass (MALDI-TOF/MS) spectrometry.

Preparation of Atg1 and Atg13. Expression and purification of Atg1 (D211A) were performed using the baculovirus expression system and tandem chromatography as reported previously[28]. The DNA encoding C-terminal His_6-tagged full-length Atg13 was inserted into a pET11a plasmid (Novagen) that was used to transform the E. coli BL21 [DE3] strain. Protein expression was induced by the addition of 0.1 mM IPTG. After cultivation, BL21 [DE3] cells were collected by centrifugation. The cells were resuspended in a buffer containing 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole, 5% (v/v) glycerol, 5mM 2-mercaptoethanol, and Protease Inhibitor Cocktail (Nacalai). After cell lysis by sonication and removal of insoluble cell debris by centrifugation, the supernatant was applied to a Ni-NTA column (Qiagen) equilibrated with 50

mM Tris-HCl (pH 8.0), 500 mM NaCl, and 10 mM imidazole. After washing the column with the same buffer, the protein was eluted from the column by adding 50 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 200 mM imidazole. Eluents were further purified by gel filtration using a HiLoad 26/60 Superdex 200 PG column (GE, Healthcare) and 20 mM Tris-HCl (pH 8.0), 150 mM NaCl as elution buffer. Purified Atg1 (D211A) and Atg13 were concentrated using Centriprep (molecular cutoff of 50,000 Da) (Millipore) and then stored at –80°C.

Preparation of GFP-fused proteins. The constructs encoding the GFP fusion proteins used in this study have been already described[53]. They all encode for proteins consisting of an N-terminally hexahistidine tagged IDP (N_{TAIL}, PNT or Sic1), followed by a linker of 14 residues containing the tobacco etch virus (TEV) protease cleavage sequence (Glu-Asn-Leu-Tyr-Phe-Gln-Gly-Ser) and the GFP, in this order (**Fig. 4a–c**). The *E. coli* strain Rosetta [DE3] pLysS (Novagen) was used for expression of GFP-fusions. Expression and purification was carried out as already described[53] except that immobilized metal affinity chromatography was performed by supplementing both equilibration and elution buffer with 1 M NaCl to get rid of contaminating DNA. Eluents were further purified by gel filtration using a Superdex 200 HR 16/60 column (GE, Healthcare) and 50 mM sodium phosphate pH 7.5, 200 mM NaCl as elution buffer. Proteins were concentrated using Centricon Plus-20 (molecular cutoff of 10,000 Da) (Millipore) and then stored at –20°C.

Preparation of Trx-fused N_{TAIL}. The thioredoxin-6His-N_{TAIL} (Trx-6His-N_{TAIL}) construct was generated as follows. The sequence encoding N_{TAIL} followed by two stop codons was sub-cloned by Gateway technology from an entry clone, whose sequence had been checked by DNA sequencing[54], into the Gateway expression vector pETG-20A (a gift of Dr. Arie Geerlof, EMBL Heidelberg, Germany) by LR reaction. In this construct, the Trx-6His and N_{TAIL} moieties are separated by a linker (LYKKAGS) encoded by the Attb1 cloning site borne by the pETG-20A vector. In order to obtain an N_{TAIL} fusion protein only differing for the position of the Trx tag, we designed a construct (referred to as N_{TAIL}-6His-Trx) bearing the same Attb1-encoded sequence between the 6His tag and Trx. Since another Attb1 cloning site is also present in the cloning vector (pDEST14, see below), the Attb1 sequence added between the 6His tag and Trx was

mutated (in italic in the primer sequences given below) in such a way that it could not be recognized by the LR enzyme while still encoding the same linker as *wild-type* attb1.

The coding sequence of N_{TAIL} was PCR amplified using the same entry clone bearing the N_{TAIL} coding sequence as above[54], and primers Attb1-SD-NT
(GTACAAAAAAGCAGGCTAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG
TCGACTACTgaggacaagatcag)
and NT-link
(*GT*CATGGTACCATGATGATGATGATGGTGCATATGGCCAGAACCAGAACCGTCTAG
AAGATTTCTGTCATTG). The coding sequence of Trx was PCR amplified using pETG-20A as template and primers link-TRX
(CCATCATCATCATCATGGTACCATG*AAGccTGTAtAAgAAgGCgGGtagc*AGCGATAAAAT
TATTCACC)
and TRX-stop-Attb2
(ACCACTTTGTACAAGAAAGCTGGGTttattaGGCCAGGTTAGCGTCGAGGAAC).
After DpnI digestion, 1µl of each of the first two PCR product were mixed and then used as template in a third PCR amplification with primers attL1a
(ACCTGTTCGTTGCAACAAATTGATGAGCAATGCTTTTTTATAATGCCAAGTTTGTAC
AAAAAAGCAGGC)
and attL2a
(TTTTGACTGATAGTGACCTGTTCGTTGCAACAAATTGATAAGCAATGCTTTCTTATA
ATGCCCACTTTGTACAAGAAAGCTGG)[55]. The third PCR amplification allows to *(i)* fuse the first two PCR products, *(ii)* insert a 6His tag (underlined in the above primer sequences) and the Attb1–encoded linker between N_{TAIL} and Trx, *(iii)* add respectively Attl1a and Attl2a sequences at the 5' and 3'end of the PCR product required for the subsequent cloning step by LR reaction. The product of the third PCR was then inserted into the Gateway vector pDEST14 by LR reaction and checked by DNA sequencing. A detailed description of the two fusion proteins is provided in **Supplementary Fig. 1**.

The *E. coli* strain T7pRos[54] was used for expression of Trx-N_{TAIL} and N_{TAIL}-Trx fusion proteins. Cultures were grown overnight to saturation in LB medium containing 100 µg/mL ampicillin. An aliquot of the overnight culture was diluted 1/25 in LB medium and grown at 37 °C. At OD_{600} of 0.7, IPTG was added to a final concentration of 0.5 mM, and the cells were

grown at 37 °C for 3 hours. The induced cells were harvested by centrifugation (6,000 g for 20 minutes at 4 °C) and resuspended in 30 mL of buffer A (50 mM Tris-HCl, pH 8, 500 mM NaCl, 20 mM Imidazole, 1 mM PMSF), supplemented with lysozyme 0.1 mg/mL, DNase I 10 µg/mL, protease inhibitor cocktail (Roche). After a 20-minutes incubation with gentle agitation, the cells were disrupted by sonication (using a 750 W sonicator and 3 cycles of 30 s each at 45% power output). Following the removal of cell debris by centrifugation at 14,000 g for 30 min, the supernatant was applied onto a Ni(II)-loaded 5 mL His-Trap HP column (GE Healthcare) previously equilibrated in buffer A. The affinity resin was washed with 20 column volumes of buffer A containing 1 M NaCl. After this step, the affinity column was connected in tandem to a Superdex 75 16/60 column (GE Healthcare) pre-equilibrated in NaP 10 mM, NaCl 200 mM pH 7 buffer. The two connected columns were eluted first using 20 mL of buffer A containing 500 mM imidazole and subsequently using 10 mM phosphate buffer at pH 7, containing 200 mM NaCl. The fractions containing the recombinant protein were collected, lyophilized and stored at –20 °C. For both proteins, mass spectrometry analysis yielded a value that fits perfectly with the expected molecular mass for a form in which the initial methionine has been cleaved off.

SAXS analysis. SAXS measurements were performed at 23°C with a BioSAXS-1000 system (Rigaku) mounted on a MicroMx007HF X-ray generator (Rigaku). A PILATUS100K detector (DECTRIS) at a distance of 462 mm from the sample, was used to measure scattering intensities. The IDR segments of PQBP-1 (82−134, 82−164, 82−214 and 82−265) in 10 mM phosphate buffer (pH6.0) were used for SAXS measurements. Circular averaging of the scattering intensities was carried out to obtain one-dimensional scattering data *I(q)* as a function of q ($q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle and λ is the X-ray wavelength 1.5418 Å) by using the program SAXSlab (Rigaku). The $I(q)$ data were processed by using the software applications embedded in the ATSAS package[56]. The radius of gyration R_g and forward scattering intensity *I(0)* were estimated from the Guinier plot of *I(q)* in the smaller angle region of $qR_g < 1.3$. The molar mass of the IDR segments were estimated by comparing *I(0)*/*c* (where *c* is the protein concentration) of the IDR segments to that of ovalbumin, or estimated from Q_R[57]. To correct interparticle interference, *I(q)* data were collected at three different protein concentrations (10, 8, and 6 mg/ml for 82−134 and 82−164. 5, 3, and 0.3 mg/ml for 82−214. 14, 10, and 8 mg/ml for 82−265). Because the intensity profile of 82−134, 82−164 and 82−265 did not indicate a

concentration effect, the correction for interparticle interference was not applied. 82−214 fragment exhibited weak self-association at 5 mg/ml and 3 mg/ml. Therefore, a diluted sample (0.3 mg/ml) was used for structural analysis. SAXS data are summarized in **Extended Data Fig. 2** and **Supplementary Table 3**.

HS-AFM observation. The procedure for HS-AFM observation of protein dynamics at the single-molecule level has already been described[58]. In brief, a glass sample stage (diameter, ~2 mm; height, ~2 mm) with a thin mica disc (1 mm in diameter and 0.05 mm thick) glued to the top by epoxy was attached onto the top of the Z-scanner by a drop of nail polish. A freshly cleaved mica surface was prepared by removing the top layers of mica using Scotch tape. Then a drop (~2 µl) of the diluted sample (~3 nM) was deposited onto the mica surface. After an incubation of about 3 minutes, the mica surface was rinsed with 20 µl of an observation buffer shown below for each sample to remove floating samples. The sample stage was then immersed in a liquid cell containing ~60 µl of the same observation buffer. HS-AFM observation was performed in the tapping mode using a laboratory built apparatus[14,15]. The short cantilevers used are custom-made by Olympus (Tokyo, Japan): resonant frequency, ~1 MHz in water; quality factor, ~2 in water; spring constant, 0.1−0.15 N/m. The cantilever's free oscillation amplitude A_0 and set point amplitude A_s were set at 1−2 nm and ~0.9 × A_0, respectively, so that the loss of cantilever's oscillation energy per tap was adjusted to be 1−3 k_BT on average. The imaging rate, scan size and the pixel size for each sample are described in the main text and Supplementary information. The observation buffers used for the respective IDPs are as follows:

PQBP-1	10 mM Sodium phosphate buffer, pH 6.0
Atg1	20 mM KCl, and 20 mM Tris-HCl, pH 6.0
	20 mM KCl, and 20 mM Tris-HCl, pH 7.0
	20 mM KCl, and 20 mM Tris-HCl, pH 8.0
	50 mM NaCl, and 20 mM Tris-HCl, pH8.0
	300 mM NaCl, and 20 mM Tris-HCl, pH8.0
Atg13	100mM NaCl and 20 mM Tris-HCl, pH 6.0
	100mM NaCl and 20 mM Tris-HCl, pH 7.0
	100mM NaCl and 20 mM Tris-HCl, pH 8.0

FACT-10SA	50 mM KCl, 10 mM $MgCl_2$, 20 mM Tris-HCl, pH 7.5 and 0.5% glycerol (vol/vol)
N_{TAIL}-GFP	10 mM Sodium phosphate buffer, pH 6.0
N_{TAIL}-Trx	10 mM Sodium phosphate buffer, pH 6.0
	10 mM Sodium phosphate buffer, pH 7.0
Trx- N_{TAIL}	10 mM Sodium phosphate buffer, pH 6.0
PNT-GFP	10 mM Sodium phosphate buffer, pH 6.0
Sic1-GFP	10 mM Sodium phosphate buffer, pH 6.0

Analysis of AFM images. To measure topographical parameters (R_{2D}, H_1 and H_2), we used a pixel-search software program that we had made for a previous HS-AFM study[59]. This program is opened to the public at the following web site:
https://elifesciences.org/content/4/e04806/article-data#fig-data-supplementary-material
First, AFM images are briefly pretreated (mostly flattened for scan line non-horizontality) by our AFM-specific data acquisition software. From an AFM image, the operator roughly finds an end region of IDR and then clicks the mouse pointer at a molecule-free position on the surface that appears very close to the end region. Then, the pixel search program automatically finds a pixel position (X, Y) having the largest height Z, by searching over $n \times n$ pixels around the operator-specified pixel position (the value of n is appropriately chosen; typically n = 5). The height H (H_1 or H_2) is obtained by subtracting the average height of the substrate surface from the Z value. The same procedure is repeated at the other end of the IDR. From the obtained two sets of (X, Y) coordinates, the direct distance D between the two pixels is calculated, which is followed by a calculation of R_{2D} value as $R_{2D} = D - H_1/2 - H_2/2$. The autocorrelation functions $G(\tau)$ of time series data $x(n\Delta t)$ of R_{2D} and H_2 were calculated as

$$G(\tau) \equiv \sum_{n=1}^{N-h}[x(n\Delta t) - \hat{\mu}]\left[x\big((n+h)\Delta t\big) - \hat{\mu}\right] / \sum_{n=1}^{N}[x(n\Delta t) - \hat{\mu}]^2,$$

where $\tau \equiv h\Delta t$ and $\hat{\mu}$ is the mean value of $x(n\Delta t)$ (n = 1, 2, 3, - - - , N).

References for Methods

689 53. Sambi, I., Gatti–Lafranconi, P., Longhi, S. & Lotti, M. How disorder influences order and
690 vice versa – mutual effects in fusion proteins containing an intrinsically disordered and a
691 globular protein. *FEBS J.* **277**, 4438–4451 (2010).

692 54. Gruet, A., Longhi, S. & Bignon, C. One-step generation of error-prone PCR libraries using
693 Gateway® technology. *Microb Cell Fact*. **11**, 14 (15 pp) (2012).

694 55. Gruet, A. et al. Fuzzy regions in an intrinsically disordered protein impair protein-protein
695 interactions. FEBS J. 283, 576–594 (2016).

696 56. Petoukhov, M. V. *et al*. New developments in the ATSAS program package for small-angle
697 scattering data analysis. *J. Appl. Cryst*. **45**, 342–350 (2012).

698 57. Rambo, R. P. & John A. Tainer, J. A. Accurate assessment of mass, models and resolution by
699 small-angle scattering. *Nature* **496**, 477–481 (2013).

700 58. Uchihashi, T., Kodera, N. & Ando, T. Guide to video recording of structure dynamics and
701 dynamic processes of proteins by high-speed atomic force microscopy. *Nat. Protoc.* **7**, 1193–
702 1206 (2012).

703 59. Ngo, K. X., Kodera, N., Katayama, E., Ando, T. & Uyeda, T. Q. P. Cofilin-induced
704 unidirectional cooperative conformational changes in actin filaments revealed by high-speed
705 AFM. *e-Life* **4**, e04806 (2015).

706