nature research

https://doi.org/10.1038/s41565-020-00798-9

Supplementary information

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

In the format provided by the authors and unedited

Supplementary Information for

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]Aix-Marseille University and CNRS, 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 Medical Life Science, 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@univ-amu.fr).

This PDF file includes:

Supplementary Notes 1 to 6

Supplementary Figs. 1 to 6

Supplementary Tables 1 to 10

References

Additional Supplementary Information includes:

Supplementary Videos 1 to 8

Supplementary Note 1 | HS-AFM imaging under controlled tip-sample contact strength. All imaging experiments in this study were performed with HS-AFM in the amplitude modulation mode (also referred to as "tapping mode"); a cantilever is vertically oscillated at its resonant frequency in buffer solution, so that the tip intermittently contacts with the sample. We used short cantilevers with spring constant k_c = 0.1−0.15 N/m, quality factor $Q_c \approx 2$ in water and resonant frequency $f_c \approx 1$ MHz in water. The cantilever free oscillation amplitude A_0 and the set point amplitude $A_s = r \times A_0$ ($r < 1$) were set at appropriate values so that the loss of cantilever's oscillation energy *per* tap was adjusted approximately at 1−3 k_BT on average (k_B, Boltzmann constant; T, room temperature in kelvin), as described below. The amount of energy loss *per* tap is expressed as

$$\Delta E = \frac{1}{2} k_c A_0^2 (1 - r^2)/Q_c, \qquad (S1)$$

when the shift of cantilever resonant frequency (Δf_c) is neglected. For example, in our typical experimental conditions, Condition-1 (k_c = 0.1 N/m, Q_c = 2, A_0 = 1.0 nm, and r = 0.85) and Condition-2 (k_c = 0.15 N/m, Q_c = 2, A_0 = 1.5 nm, and r = 0.9), we obtain $\Delta E \approx 1.7$ and 3.7 k_BT, respectively. In reality the amount of energy loss is smaller than these values because of Δf_c. The Δf_c is caused by the force gradient $k \equiv \partial F / \partial z$ acting between tip and sample, and can be approximately expressed as

$$\Delta f_c \approx -\frac{f_c}{2k_c} \times k. \qquad (S2)$$

This is because an elastic tip-sample interaction (i.e., energy-conservative interaction) acts as a spring, changing the cantilever spring constant to an effective value of $k_c - k$, where $k < 0$ for repulsive interactions. Since the small cantilevers we used for HS-AFM imaging has high f_c and small k_c values, we can expect large Δf_c values. Although the value of Δf_c cannot be directly measured in our system, it can be estimated from the phase shift $\Delta\theta_c$ (> 0 for repulsive interactions) at the excitation frequency. Under the conditions of A_0 = 1.5 nm and r = 0.9, the measured $\Delta\theta_c$ was ~11.5°, which corresponds to $\Delta f_c \approx 57.4$ kHz, a value estimated from the second-order transfer function of the cantilever (i.e., frequency response). This Δf_c lowers the excitation efficiency because the excitation frequency is fixed at the resonant frequency in free oscillation. The amplitude reduction resulting from the lowered excitation efficiency is estimated to be ~3% from the cantilever's transfer function. Therefore, this amplitude reduction by the elastic interaction contributes by ~30% to the measured amplitude reduction. Thus, the actual energy loss *per* tap is 1.2 and 2.8 k_BT under Condition-1 and Condition-2, respectively, approximately 20−30% smaller than the values estimated using ***Eq.* (*S1*)**.

The oscillation energy that is lost at every tip-sample contact is considered to be transferred to the sample. This transferred energy is partitioned into many degrees of freedom of the sample, as the tip end (with a radius of ~5 nm) establishes contacts with many residues and atoms. Note that for a given number of residues the degree of freedom in disordered proteins is larger than that in ordered proteins. Therefore, IDPs can distribute the transferred energy over more degrees of freedom, and hence, are less susceptible

to the tip-sample contact. The energy partitioned into many degrees of freedom dissipates quickly into solution. If it did not dissipates quickly but were accumulated, the protein sample would be quickly destroyed as the cantilever is oscillating at ~1 MHz. In fact, in the typical HS-AFM imaging of the $\alpha_3\beta_3$ subcomplex of F_1-ATPase (Ref. 1), its conformational changes driven by the ATPase reaction continuously appeared on the HS-AFM images without deterioration of the conformational activity over the imaging period of 40 s.

The setting of A_0 at 1−2 nm and r at 0.85−0.9, which can give a small energy loss (1−3 k_BT) *per* tip-sample contact, is made possible by our specialized feedback controller called the "dynamic PID controller" (Ref. 2). In this controller, the gain parameters can be automatically adjusted during imaging. In the downhill region of the sample, the gain is increased to avoid complete detachment of the tip from the sample surface. In conventional HS-AFM systems, the controller's gain parameters are fixed during imaging. Therefore, under the condition of ΔE = 1−3 k_BT the tip is completely detached from the sample surface in its downhill region, and thus, imaging becomes impossible.

Supplementary Note 2 | Two-dimensional end-to-end distance measurement of IDRs. To measure the two-dimensional (2D) end-to-end distance R_{2D} of the IDRs in the five PQBP-1 constructs, first the direct distance D between the centre of the N-terminal globule and the C-terminal end position was measured. Second, the height of the N-terminal globule H_1 and the height of the C-terminal end H_2 were measured. Then, R_{2D} was estimated as

$$R_{2D} = D - H_1/2 - H_2/2 \qquad (S3)$$

(see **Fig. 1c** in the main text). We employed this procedure because generally AFM has a high precision in height measurement (± ~0.15 nm in the case of high-speed imaging with our instrument), whereas it has a low precision in the lateral size measurement due to a tip-size effect, i.e., an AFM image of a particle appears larger than its actual size in the lateral direction by the extent of the diameter of the tip apex region in lateral contact with the particle, as illustrated with the dashed red lines in **Fig. 1c** in the main text. Since no globule exists in the C-terminal end, it may not have been necessary to subtract $H_2/2$ from D. We however decided to determine R_{2D} in this way, as we reasoned that an error caused by this subtraction is negligible because the value of $H_2/2$ for the PQBP-1 constructs (half of ~0.5 nm) is much smaller than D. Similarly, we estimated R_{2D} of the tail-like structures of Atg1, Atg13, the three GFP-fused IDPs, Trx-N_{TAIL} and N_{TAIL}-Trx (see **Fig. 3e** in the main text), using ***Eq.* (*S3*)**, in which H_1 is now the heights of GFP in the GFP-fused IDPs, kinase domain in Atg1, HORMA domain in Atg13 and Trx in Trx-N_{TAIL} and N_{TAIL}-Trx.

Supplementary Note 3 | Usage of standard error of the mean. The two-dimensional end-to-end distance (R_{2D}) of an IDR distributes widely because of the flexible nature of IDR. Therefore, its Gaussian distribution fitting produces a large standard deviation (s.d). Since this s.d. does not represent an error for

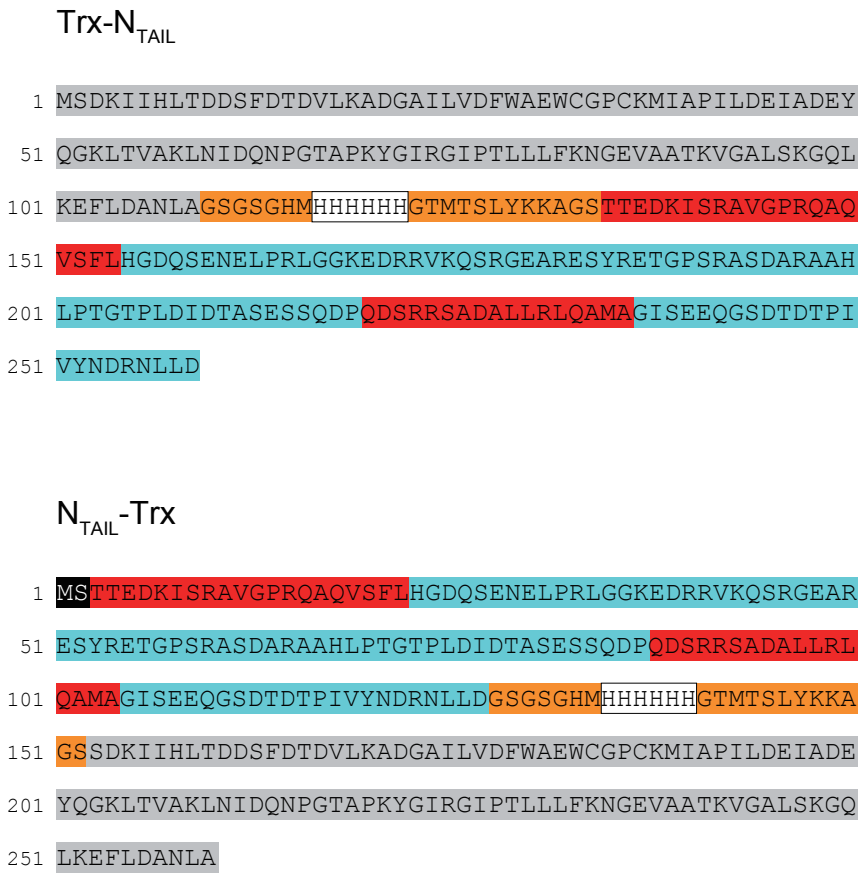
the estimation of the mean value of R_{2D} (i.e., $\langle R_{2D} \rangle$), we use "standard error of the mean" (s.e.m.) as an estimation error. The s.e.m. represents an accuracy range of the obtained median (i.e., $\langle R_{2D} \rangle$) of the Gaussian distribution; the probability that the true mean value is in the range of mean ± 2 × s.e.m. is 0.95. The height of a fully disordered region also distributes widely because the height differs among a.a. residues contained therein. The height of metastable small globules also distributes widely owing to their disorder/order transitions. Even in the case of stable globules such as the WWD of PQBP-1, their height thermally fluctuates. Therefore, the estimated values for these heights are also expressed as mean ± s.e.m., unless otherwise stated.

Supplementary Note 4 | Order propensity of IDR and its affinity for partner proteins. An IDR containing a binding site often folds into an ordered structure upon binding to its target. Without performing kinetics and/or NMR relaxation studies, it cannot be anticipated however whether before binding the partner the IDR is in its fully disordered state, as hypothesized in the 'fly-casting model', or whether it adopts a partially ordered state (Ref. 3). It is plausible that the IDR adopts a loosely folded conformation before binding to the target, thereby facilitating IDR recognition by the target via reduction of the entropic penalty associated to the disorder-to-order transition. In N_{TAIL}, PNT and Sic1 studied here, their IDRs undergoing structural transitions contain binding sites for partners, which may suggest the conformational selection-then-fitting mechanism. This model posits that partial folding of an IDR is a prerequisite for its recognition by the partner, implying that an IDR that adopts constantly and fully disordered conformations would be a poor binder. However, even an IDR that looks constantly and fully disordered, may transiently visit a loosely folded state. In addition, and in spite of the partial preconfiguration of the α-MoRE, N_{TAIL} binding to the X domain of the viral phosphoprotein (XD) was shown to occur according to an induced fit mechanism (Refs. 4 and 5). The low extent of pre-configuration of binding sites would lead to binding of the target with a low affinity. This view seems consistent with the low affinity binding, for example, (i) between the disordered C-terminal region of PQBP-1 and its target, a spliceosomal protein U5-15kD (dissociation constant, 56 μM) (Ref. 6) and (ii) between the KIX domain of cAMP-response element-binding protein (CREB)-binding protein (CBP) and the unphosphorylated form of the kinase-inducible transactivation domain (KID) from CREB (dissociation constant, ~100 μM) (Refs. 7 and 8). On the other hand, N_{TAIL}, which contains a pre-structured α-MoRE, has a relatively high affinity for the X domain of the viral phosphoprotein (dissociation constant, 3 μM) (Ref. 5). Although systematic studies are awaited, in particular to assess the combined effects of preconfiguration of MoREs and of "fuzzy" appendages, the propensity of an IDR to adopt a loosely folded state is very likely to provide a high affinity of the IDR for its target.

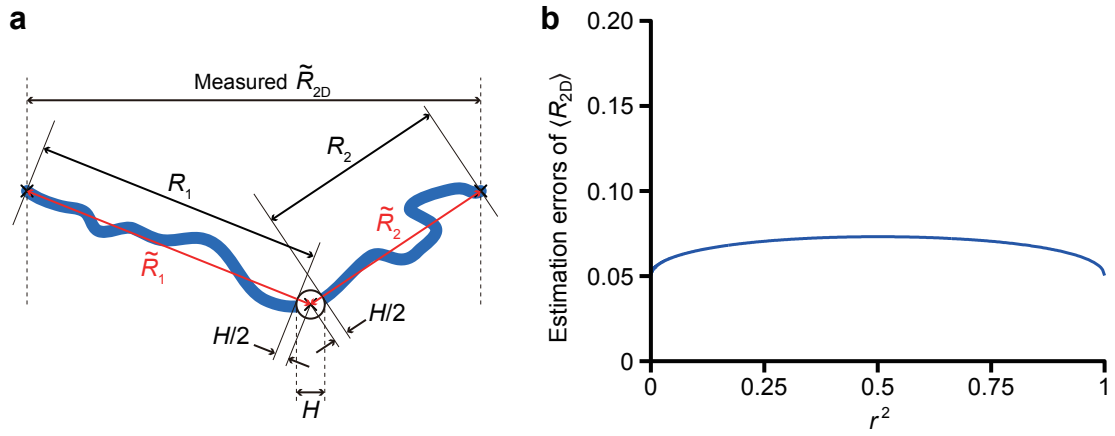
Supplementary Note 5 | Residual structures and sequences found by detailed structural analysis. Previous studies of IDPs have identified the following sequence features responsible for residual structures that even if transient lead to variations in compactness with respect to the random coil: high contents of proline residues leading to polyproline structures (Ref. 9), a specific repeat sequence

responsible for hydrogen bonds enabling self-association (Ref. 10), specific sequences allowing transient end-to-end association (Ref. 11), extended repeat domains forming β-structure (Ref. 12), and sequences with a propensity to form a coiled-coil (Ref. 13).

Supplementary Note 6 | Compaction or swelling effect of charged residues on the dimension of IDRs. The power laws and the constant values of L_p for IDRs in solution and on mica obtained in this study 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 (Ref. 14–17) (**Extended Data Fig. 9**). A software program called CIDER (opened to public use) is available to calculate the values of charge density (ρ), FCR and κ (Ref. 18). All IDPs randomly chosen in this study do not possess high charge densities or large values of FCR and κ (**Supplementary Table 10**), as seen in numerous naturally occurring IDRs (**Extended Data Fig. 9**; Refs. 19 and 20).



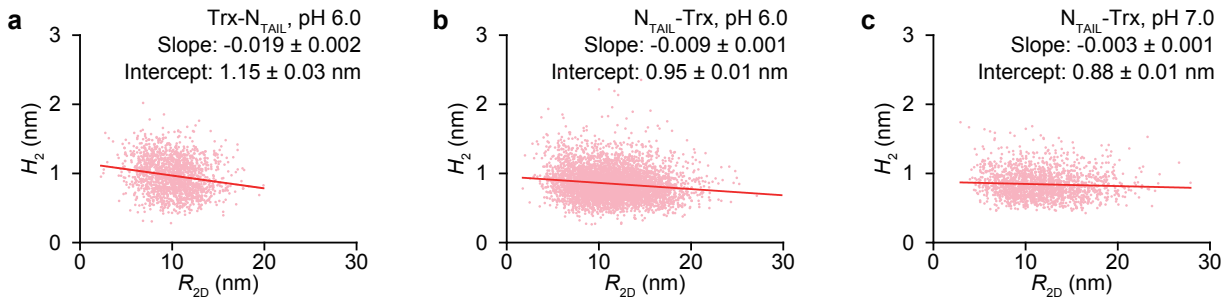
Supplementary Fig. 1 | Amino-acid sequences of Trx-N$_{TAIL}$ and N$_{TAIL}$-Trx. Color codes: Gray, Trx; black, Met and Ser; orange, linkers; white, His$_6$; red, α-More; cyan, IDR.



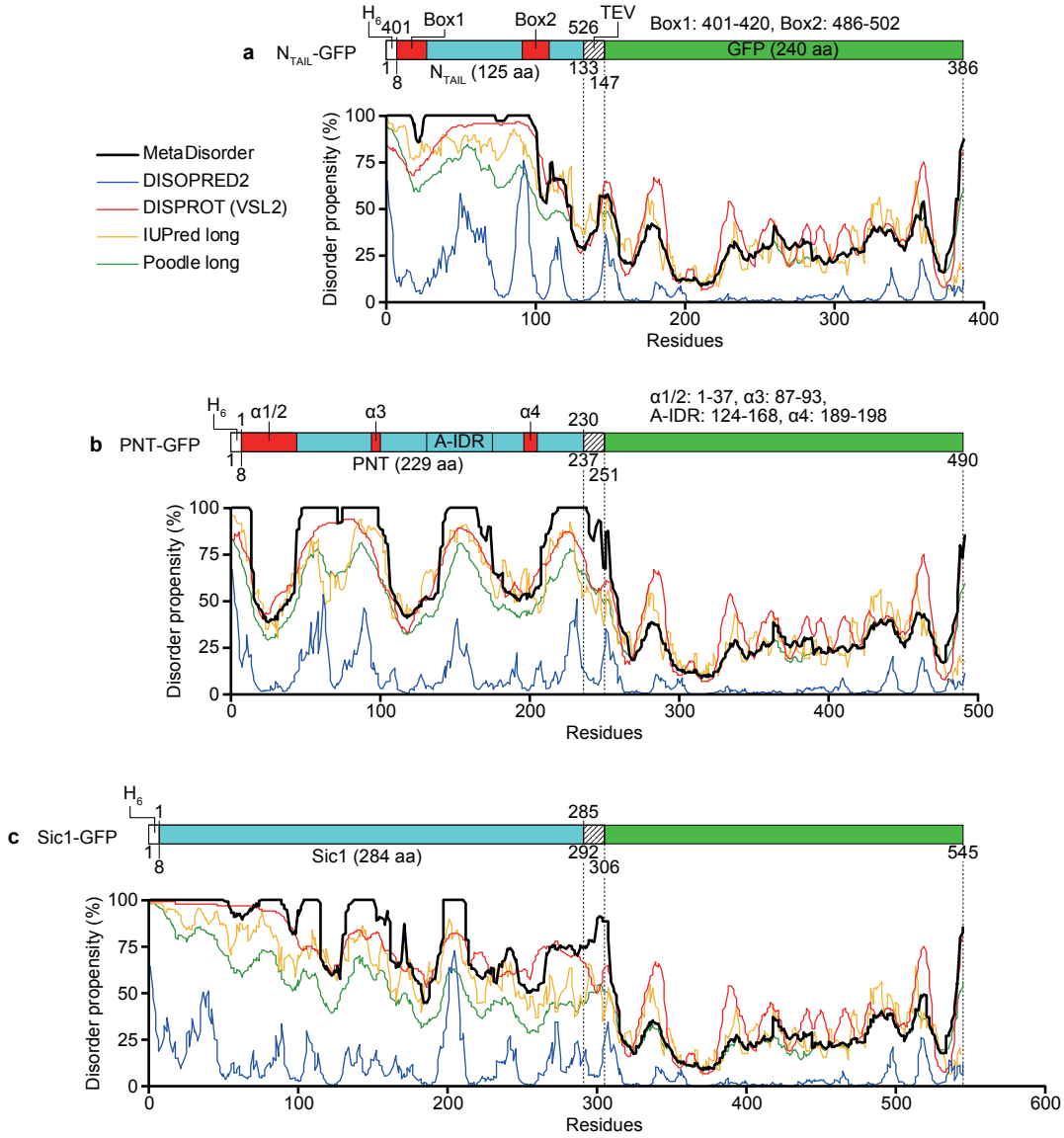
Supplementary Fig. 2 | Effect of small globule intervening between two IDR segments on the estimation of $\langle R_{2D} \rangle$. **a**, Schematic of two IDRs connected via a small globule with height H. **b**, Estimation errors of $\langle R_{2D} \rangle$ of total IDR consisting of two IDRs (IDR-1 and IDR-2) as a function of the position of the small globule. The error was calculated using ***Eq.* (*S4*)** shown below. The mean square end-to-end distance $\langle R_{2D}^2 \rangle$ of the total IDR is expressed as $\langle R_{2D}^2 \rangle = \langle R_1^2 \rangle + \langle R_2^2 \rangle$, where R_1 and R_2 are the end-to-end distances of IDR-1 and IDR-2, respectively. When there is a small globule with height H that intervenes between the two IDR segments, R_1 and R_2 are expressed as $R_1 = \tilde{R}_1 - H/2$ and $R_2 = \tilde{R}_2 - H/2$ respectively, where $\tilde{R}_1$ and $\tilde{R}_2$ represent the respective end-to-end distances of IDR-1 and IDR-2 when the existence of the small globule is neglected. Therefore, when it is neglected, the mean value of the measured end-to-end distance $\langle \tilde{R}_{2D} \rangle$ is given by $\langle \tilde{R}_{2D}^2 \rangle = \langle \tilde{R}_1^2 \rangle + \langle \tilde{R}_2^2 \rangle$. The estimation error ($\varepsilon$) of $\langle R_{2D} \rangle$ caused by this disregard is well approximated as $\varepsilon = H(\langle \tilde{R}_1 \rangle + \langle \tilde{R}_2 \rangle)/(2\langle R_{2D}^2 \rangle)$ under the condition of $H \ll \langle R_{2D} \rangle$. This error size depends on the position of the small globule; when R_1 is expressed as $\langle R_1 \rangle = r \times \langle R_{2D} \rangle$ ($0 < r < 1$), ε is well approximated by

$$\varepsilon = H[r + \sqrt{1 - r^2}\,]/(2\langle R_{2D} \rangle) \quad (S4).$$

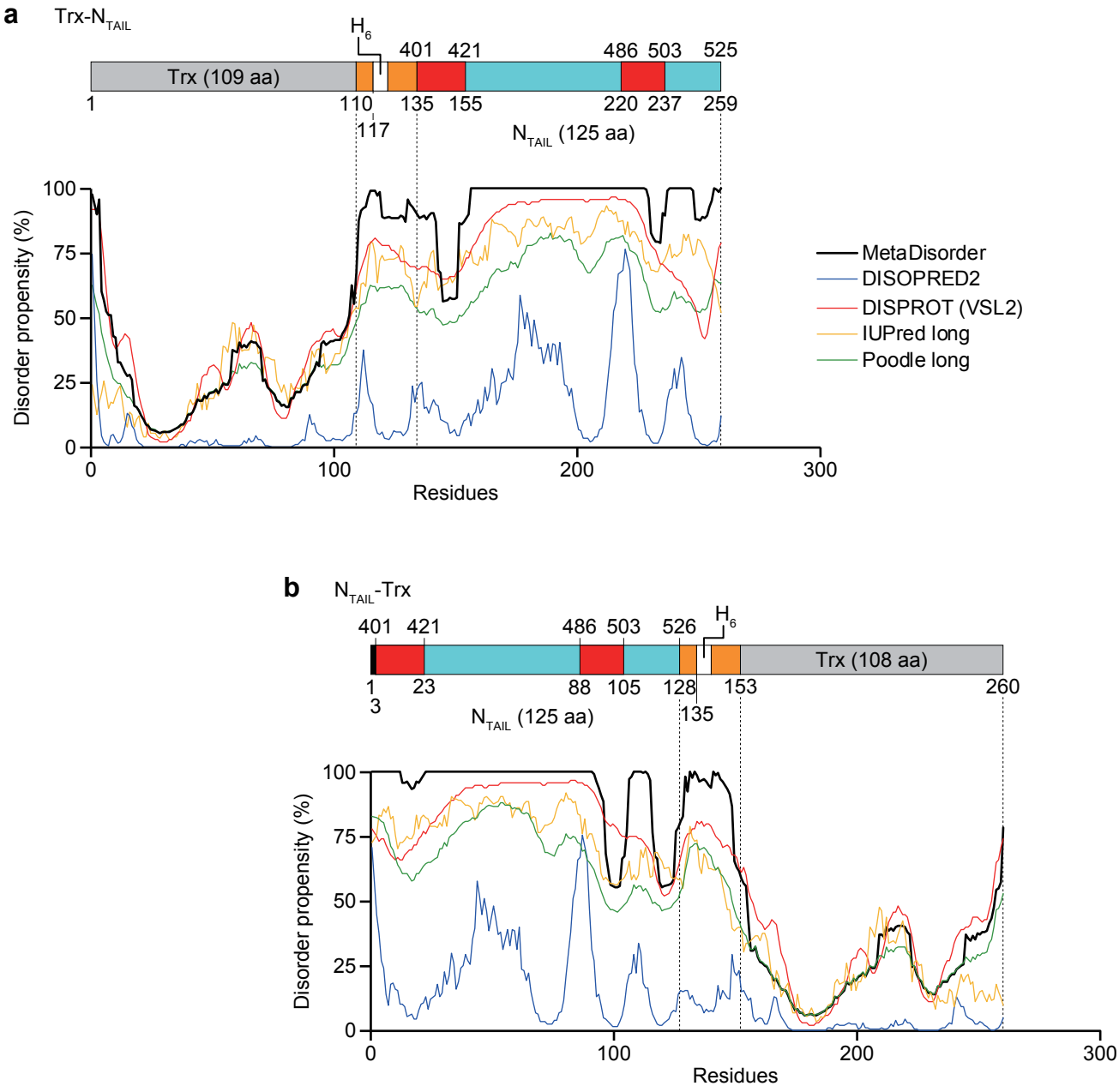
Using ***Eq.* (*S4*)**, height $\langle H \rangle$ = 1.1 nm and $\langle R_{2D}^2 \rangle$ = 11.6 nm measured for N_{TAIL}-GFP and $r^2 \sim 0.76$ (see **Fig. 4a** in the main text), the estimation error of $\langle R_{2D}^2 \rangle$ is ~6% when estimated without considering the presence of the C-terminal small globule (i.e., Box2).



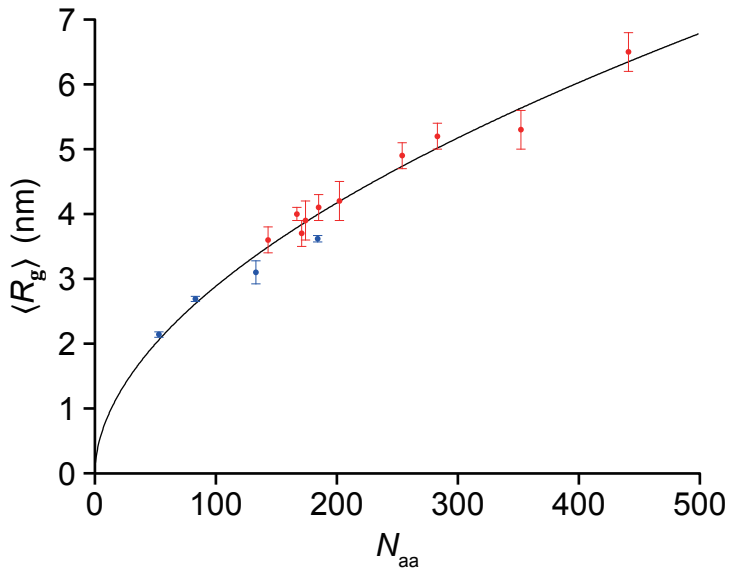
Supplementary Fig. 3 | Scattered plots showing correlation between R_{2D} and height of small globules within Box 1 and Box 2 in N_{TAIL}. **a**, Correlation between R_{2D} and H_2 (Box 2) in Trx-N_{TAIL} measured at pH 6.0. **b**, Correlation between R_{2D} and H_2 (Box 1) in N_{TAIL}-Trx measured at pH 6.0. **c**, Correlation between between R_{2D} and H_2 (Box 1) in N_{TAIL}-Trx measured at pH 7.0. The small correlation observed for N_{TAIL}-Trx at pH 7.0 reflects that Box 1 is mostly in the lower height state (**Extended Data Fig. 7j).**



Supplementary Fig. 4 | Order/disorder maps of three GFP-fusions along their length. Five publicly available prediction programs were used: MetaDisorder, DISPRED2, DISPROT (VSL2), IUPred long, and Poodle-long (Ref. 21). **a**–**c**, Prediction results for N_{TAIL}-GFP (**a**), PNT-GFP (**b**) and Sic1-GFP (**c**).



Supplementary Fig. 5 | Order/disorder maps of Trx-N_{TAIL} and N_{TAIL}-Trx along their length. Five publicly available prediction programs were used: MetaDisorder, DISPRED2, DISPROT (VSL2), IUPred long, and Poodle-long (Ref. 21). **a**,**b**, Prediction results for Trx-N_{TAIL} (**a**) and N_{TAIL}-Trx (**b**).



Supplementary Fig. 6 | Relationship between N_{aa} and $\langle R_g \rangle$ estimated by SAXS measurements for ten tau protein constructs and four PQBP-1' s IDR segments. The data plotted with the red circles are those from the study by Mylonas *et al.* for tau protein constructs (Ref. 12), with phosphor-mimic constructs and those significantly affected by the extended repeat domain (Refs. 22 and 23) having been removed here. The constructs herein analyzed are tau ht40, tau K32, tau K16, tau ht23, tau K27, tau K17, tau K44, tau K10, tau K25, and tau K23. The data plotted with the blue circles are from the present SAXS study for four segments of PQBP-1 IDR (82−265, 82−214, 82−164, and 82−134). The black line indicates the best fitting result obtained by fitting the combined data to a power law, $\langle R_g \rangle = \beta_g \times N_{aa}^{\nu}$: $\beta_g = 0.25 \pm 0.038$ nm, and $\nu = 0.53 \pm 0.028$.

Supplementary Table 1 | Summary of the AFM image analysis for five PQBP-1 constructs. The standard deviation (s.d.) of R_{2D} indicates the width of the Gaussian distribution that mostly originates from the flexible nature of IDRs.

Parameter	WT PQBP-1 (1-265)	PQBP-1 (1-214)	PQBP-1 (1-164)	PQBP-1 (1-134)	PQBP-1 (1-114)
Number of a.a. contained in the unstructured chain	184	133	83	53	33
Frame rate (fps)	23.8	15.2	13.2	13.9	13.2
$\langle R_{2D} \rangle$, Single Gaussian fit (mean ± s.d.)	17.2 ± 4.4 nm	14.4 ± 4.1 nm	11.0 ± 3.0 nm	9.3 ± 2.6 nm	7.1 ± 2.0 nm
Estimation error of $\langle R_{2D} \rangle$ (s.e.m.)	0.09 nm	0.14 nm	0.04 nm	0.07 nm	0.02 nm
$\langle H_1 \rangle$, height of N-terminal globule Single Gaussian fit (mean ± s.d.)	1.9 ± 0.2 nm	1.9 ± 0.2 nm	1.9 ± 0.2 nm	1.9 ± 0.2 nm	1.9 ± 0.2 nm
Estimation error of $\langle H_1 \rangle$ (s.e.m.)	0.002 nm	0.004 nm	0.002 nm	0.003 nm	0.006 nm
$\langle H_2 \rangle$, height of tail end Single Gaussian fit (mean ± s.d.)	0.5 ± 0.2 nm	0.5 ± 0.2 nm	0.5 ± 0.3 nm	0.5 ± 0.2 nm	0.5 ± 0.1 nm
Estimation error of $\langle H_2 \rangle$ (s.e.m.)	0.005 nm	0.006 nm	0.002 nm	0.001 nm	0.002 nm
Number of molecules analyzed	4	4	3	3	4
Number of frames analyzed	1,360	1,753	1,190	1,178	925

Supplementary Table 2 | Summary of the image analysis for PQBP-1 (1-214) captured at various frame rates. The standard deviation (s.d.) of R_{2D} indicates the width of the Gaussian distribution that mostly originates from the flexible nature of IDRs.

Parameter	50 fps	30 fps	20 fps	15.2 fps
$\langle R_{2D} \rangle$, Single Gaussian fit (mean ± s.d.)	14.5 ± 4.3 nm	14.6 ± 3.8 nm	14.3 ± 4.2 nm	14.4 ± 4.1 nm
Estimation error of $\langle R_{2D} \rangle$ (s.e.m.)	0.19 nm	0.19 nm	0.15 nm	0.14 nm
$\langle H_1 \rangle$, height of N-terminal globule Single Gaussian fit (mean ± s.e.m.)	1.9 ± 0.011 nm	1.9 ± 0.008 nm	1.9 ± 0.005 nm	1.9 ± 0.009 nm
$\langle H_2 \rangle$, height of tail end Single Gaussian fit (mean ± s.e.m.)	0.5 ± 0.006 nm	0.5 ± 0.005 nm	0.5 ± 0.011 nm	0.5 ± 0.011 nm
Area of fitted Gaussian distribution (arbitrary unit)	3,150	3,820	4,540	3,530
Number of molecules analyzed	3	3	3	4
Number of frames analyzed	1,566	1,937	2,265	1,753

Parameter	10 fps	6.67 fps	Averaged over the six conditions
$\langle R_{2D} \rangle$, Single Gaussian fit (mean ± s.d.)	14.3 ± 3.7 nm	14.6 ± 3.7 nm	14.4 ± 4.0 nm
Estimation error of $\langle R_{2D} \rangle$ (s.e.m.)	0.13 nm	0.12 nm	0.17 nm
$\langle H_1 \rangle$, height of N-terminal globule Single Gaussian fit (mean ± s.e.m.)	1.9 ± 0.005 nm	1.9 ± 0.003 nm	1.9 ± 0.01 nm
$\langle H_2 \rangle$, height of tail end Single Gaussian fit (mean ± s.e.m.)	0.5 ± 0.005 nm	0.5 ± 0.006 nm	0.5 ± 0.02 nm
Area of fitted Gaussian distribution (arbitrary unit)	3,110	3,150	-
Number of molecules analyzed	3	3	-
Number of frames analyzed	1,559	1,395	-

The areas of respective Gaussian distributions are used as weights for the calculation of weighted mean values of $\langle R_{2D} \rangle$, $\langle H_1 \rangle$ and $\langle H_2 \rangle$ averaged over the six imaging-rate conditions.

Supplementary Table 3 | Summary of structural parameters of IDR segments of PQBP-1 as determined from SAXS measurements

Segment	Protein concentration (mg/ml)	Exposure Time (h)	R_g (Å) mean ± s.d.	$I(0)$ (a.u.) mean ± s.d.	Molar mass [a] (kDa)	Molar mass [b] (kDa)	Theoretical molar mass [c] (kDa)
82−134	10	4	21.6 ± 0.2	0.184±0.00	4.0	6.0	6.4
	8	4	20.9 ± 0.2	0.147±0.00	4.0	5.2	
	6	4	21.8 ± 0.2	0.128±0.00	4.7	5.1	
82−164	10	4	26.9 ± 0.3	0.261±0.00	5.7	10.1	10.3
	8	4	26.4 ± 0.3	0.199±0.00	5.5	11.4	
	6	4	27.4 ± 0.4	0.186±0.00	6.8	8.7	
82−214	5	2	33.4 ± 0.4	0.437±0.00	33.4	18.3	16.2
	3	2	30.6 ± 0.5	0.232±0.00	29.5	17.2	
	0.3	12	29.0 ± 0.2	0.018±0.00	23.3	16.8	
82−265	14	4	36.8 ± 0.2	0.791±0.00	17.4	23.8	21.6
	10	4	36.1 ± 0.3	0.526±0.00	16.3	25.0	
	8	4	35.7 ± 0.3	0.432±0.00	16.7	23.1	

[a] Estimated by comparison of $I(0)/c$ with that from the reference solutions of ovalbumin.

[b] Estimated from Q_R.

[c] Calculated from amino acid sequence.

Supplementary Table 4 | Summary of the AFM image analysis for Atg1 (D211A) at various buffer conditions. The standard deviation (s.d.) of R_{2D} indicates the width of the Gaussian distribution that mostly originates from the flexible nature of IDRs.

Parameter	20 mM KCl pH 6.0	20 mM KCl pH 7.0	20 mM KCl pH 8.0	50 mM NaCl pH 8.0	300 mM NaCl pH 8.0
$\langle R_{2D} \rangle$, 1st peak of double Gaussian fit (mean ± s.d.)	12.1 ± 4.1 nm	11.9 ± 4.1 nm	11.1 ± 5.4 nm	13.0 ± 4.0 nm	12.2 ± 4.9 nm
Estimation error of the 1st peak of $\langle R_{2D} \rangle$ (s.e.m.)	0.82 nm	1.47 nm	0.56 nm	1.04 nm	1.06 nm
$\langle R_{2D} \rangle$, 2nd peak of double Gaussian fit (mean ± s.d.)	20.6 ± 3.1 nm	20.4 ± 4.0 nm	20.2 ± 2.7 nm	20.5 ± 3.9 nm	20.3 ± 3.1 nm
Estimation error of the 2nd peak of $\langle R_{2D} \rangle$ (s.e.m.)	0.53 nm	2.25 nm	0.52 nm	1.42 nm	0.94 nm
F-test (Double Gaussian vs. Single Gaussian)	$P < 0.1$	$P < 0.1$	$P < 0.1$	$P < 0.05$	$P < 0.05$
Area ratio of two Gaussian components of $\langle R_{2D} \rangle$ distribution (1st : 2nd), giving order propensity K_e	913 : 788 $K_e = 1.16$	1357 : 883 $K_e = 1.54$	1,451 : 309 $K_e = 4.70$	2,337 : 1,539 $K_e = 1.52$	4,173 : 1,354 $K_e = 3.08$
$\langle R_{2D} \rangle$ of the 2nd peak averaged over the five conditions (mean ± s.e.m.)	20.5 ± 0.9 nm				
Number of a.a. contained in the unstructured chain (332 – 586 a.a.)	255				
$\langle H_1 \rangle$, height of kinase domain (mean ± s.d.)	3.7 ± 0.4 nm	3.5 ± 0.4 nm	3.5 ± 0.4 nm	3.3 ± 0.4 nm	3.4 ± 0.5 nm
Estimation error of $\langle H_1 \rangle$ (s.e.m.)	0.007 nm	0.006 nm	0.005 nm	0.005 nm	0.010 nm
$\langle H_1 \rangle$ averaged over the five conditions (mean ± s.e.m.)	3.5 ± 0.2 nm				
$\langle H_2 \rangle$, height of MIT domain (mean ± s.d.)	3.2 ± 0.3 nm	3.3 ± 0.4 nm	3.3 ± 0.4 nm	3.2 ± 0.3 nm	3.2 ± 0.3 nm
Estimation error of $\langle H_2 \rangle$ (s.e.m.)	0.012 nm	0.011 nm	0.005 nm	0.003 nm	0.006 nm
$\langle H_2 \rangle$ averaged over the five conditions (mean ± s.d.)	3.2 ± 0.2 nm				
Number of molecules analyzed	3	3	3	3	4
Number of frames analyzed	873	1,082	868	1,926	2,720

The areas of respective Gaussian distributions are used as weights for the calculation of weighted means, $\langle R_{2D} \rangle$, $\langle H_1 \rangle$ and $\langle H_2 \rangle$, averaged over the five ionic conditions.

Supplementary Table 5 | Summary of the AFM image analysis for Atg13 at various buffer conditions. The standard deviation (s.d.) of R_{2D} indicates the width of the Gaussian distribution that mostly originates from the flexible nature of IDRs.

Parameter	100 mM NaCl pH 6.0	100 mM NaCl pH 7.0	100 mM NaCl pH 8.0
$\langle R_{2D} \rangle$, 1st peak of double Gaussian fit (mean ± s.d.)	19.2 ± 5.6 nm	20.7 ± 5.3 nm	20.3 ± 5.4 nm
Estimation error of the 1st peak of $\langle R_{2D} \rangle$ (s.e.m.)	0.32 nm	0.51 nm	0.53 nm
$\langle R_{2D} \rangle$, 2nd peak of double Gaussian fit (mean ± s.d.)	27.9 ± 6.7 nm	28.1 ± 7.5 nm	27.8 ± 7.1 nm
Estimation error of the 2nd peak of $\langle R_{2D} \rangle$ (s.e.m.)	1.62 nm	4.62 nm	4.06 nm
$\langle R_{2D} \rangle$ of the 2nd peak averaged over the three conditions (mean ± s.e.m.)	28.0 ± 2.0 nm		
Number of a.a. contained in the unstructured chain (268 – 744 a.a.)	477		
F-test (Double Gaussian vs. Single Gaussian)	$P < 10^{-2}$	$P < 10^{-3}$	$P < 10^{-3}$
Area ratio of two Gaussian components in R_{2D} distribution (1st : 2nd), giving order propensity K_e	2,810 : 1,513 $K_e = 1.86$	3,498 : 3,173 $K_e = 1.10$	3,055 : 2,087 $K_e = 1.46$
$\langle H_1 \rangle$, height of HORMA domain (mean ± s.d.)	2.7 ± 0.3 nm	2.5 ± 0.2 nm	2.7 ± 0.4 nm
Estimation error of $\langle H_1 \rangle$ (s.e.m.)	0.003 nm	0.005 nm	0.007 nm
$\langle H_1 \rangle$ averaged over the three conditions (mean ± s.d.)	2.6 ± 0.3 nm		
$\langle H_2 \rangle$, height of tail end (mean ± s.d.)	0.5 ± 0.2 nm	0.5 ± 0.2 nm	0.5 ± 0.2 nm
Estimation error of $\langle H_2 \rangle$ (s.e.m.)	0.006 nm	0.004 nm	0.004 nm
$\langle H_2 \rangle$ averaged over the three conditions (mean ± s.d.)	0.5 ± 0.2 nm		
Number of molecules analyzed	4	3	3
Number of frames analyzed	2,156	3,345	2,569

The areas of respective Gaussian distributions are used as weights for the calculation of weighted mean values of $\langle R_{2D} \rangle$, $\langle H_1 \rangle$ and $\langle H_2 \rangle$ averaged over the three pH conditions.

Supplementary Table 6 | Summary of AFM image analysis of three GFP-fused IDPs. The standard deviation (s.d.) of R_{2D} indicates the width of the Gaussian distribution that mostly originates from the flexible nature of IDRs. The standard error of the mean (s.e.m.) for R_{2D}, H_1 and H_2 indicate estimation errors for these parameters.

Parameter	N_{TAIL}-GFP	PNT-GFP	Sic1-GFP
$\langle R_{2D} \rangle$, Single Gaussian fit (mean ± s.d.)	11.6 ± 4.4 nm	–	–
Estimation error of $\langle R_{2D} \rangle$ (s.e.m.)	0.11 nm	—	—
$\langle R_{2D} \rangle$, 1st peak of double Gaussian fit (mean ± s.d.)	–	8.9 ± 3.2 nm	12.2 ± 3.6 nm
Estimation error of the 1st peak of $\langle R_{2D} \rangle$ (s.e.m.)	—	0.33 nm	0.66 nm
$\langle R_{2D} \rangle$, 2nd peak of double Gaussian fit (mean ± s.d.)	–	14.3 ± 4.8 nm	20.5 ± 4.8 nm
Estimation error of the 2nd peak of $\langle R_{2D} \rangle$ (s.e.m.)	—	9.26 nm	2.14 nm
F-test (Double Gaussian vs. Single Gaussian)	–	$P < 10^{-20}$	$P < 0.02$
Area ratio of two Gaussian components in R_{2D} distribution (1st : 2nd), giving order propensity K_e	–	9,260 : 2,480 $K_e = 3.73$	3,310 : 2,210 $K_e = 1.50$
λ, decay rate of autocorrelation function of time series R_{2D} data (best fit value ± s.e.m.)	–	2.50 ± 0.14 s^{-1}	1.87 ± 0.08 s^{-1}
k_{OD}, order-to-disorder transition rate estimated from R_{2D} dynamics (best fit value ± s.e.m.)	–	0.53 ± 0.23 s^{-1}	0.75 ± 0.15 s^{-1}
k_{DO}, disorder-to-order transition rate estimated from R_{2D} dynamics (best fit value ± s.e.m.)	–	1.97 ± 0.25 s^{-1}	1.12 ± 0.16 s^{-1}
$\langle H_1 \rangle$, height of GFP moiety Single Gaussian fit (mean ± s.d.)	3.2 ± 0.3 nm	3.1 ± 0.3 nm	3.2 ± 0.4 nm
Estimation error of $\langle H_1 \rangle$ (s.e.m.)	0.004 nm	0.008 nm	0.013 nm
$\langle H_2 \rangle$, height of N-terminal globule Single Gaussian fit (mean ± s.d.)	–	1.1 ± 0.3 nm	–
Estimation error of $\langle H_2 \rangle$ (s.e.m.)	—	0.003 nm	—
$\langle H_2 \rangle$, height of N-terminal globule 1st peak of double Gaussian fit (mean ± s.d.)	0.8 ± 0.2 nm	–	0.9 ± 0.3 nm
Estimation error of the 1st peak of $\langle H_2 \rangle$ (s.e.m.)	0.03 nm	—	0.05 nm
$\langle H_2 \rangle$, height of N-terminal globule 2nd peak of double Gaussian fit (mean ± s.d.)	1.1 ± 0.3 nm	–	1.3 ± 0.4 nm
Estimation error of the 2nd peak of $\langle H_2 \rangle$ (s.e.m.)	0.82 nm	—	0.28 nm
F-test (Double Gaussian vs. Single Gaussian)	$P < 10^{-4}$	–	$P < 0.02$
Area ratio of Gaussian components in H_2 distribution (1st : 2nd), giving order propensity K_e	561 : 156 $K_e = 0.28$ (Box 1)	–	160 : 260 $K_e = 1.63$
λ, decay rate of autocorrelation function of time series H_2 data (best fit value ± s.e.m.)	7.31 ± 0.84 s^{-1}	–	1.71 ± 0.06 s^{-1}
k_{OD}, order-to-disorder transition rate estimated from H_2 dynamics (best fit value ± s.e.m.)	5.72 ± 0.67 s^{-1}	–	0.65 ± 0.34 s^{-1}
k_{DO}, disorder-to-order transition rate estimated from H_2 dynamics(best fit value ± s.e.m.)	1.59 ± 0.23 s^{-1}	–	1.06 ± 0.34 s^{-1}
Average height of IDR (mean ± s.e.m.) (not including small globules)	0.5 ± 0.02 nm (N = 2,138)	0.5 ± 0.02 nm (N = 1,972)	0.5 ± 0.03 nm (N = 5,758)
$N_{aa}{}^T$	125	229	284
N_{aa}	88	~53 (sorter state) ~128 (longer state)	~95 (shorter state) ~257 (longer state)
ΔN_{aa} between order and disorder states	–	~75	~160
Number of molecules analyzed	11	4	5
Number of frames analyzed	4,754	5,831	2,801
Number of frames used	5,193	6,316	2,888

$N_{aa}{}^T$ indicates the number of a.a. contained in each IDP (the first Met and the following His-tag are not included).

Supplementary Table 7 | Summary of the AFM image analysis of two Trx-fusion N_{TAIL} constructs. The standard deviation (s.d.) of R_{2D} shown is the width of the Gaussian distribution that mostly originates from the flexible nature of IDRs.

Parameter	N_{TAIL}-Trx (at pH 6.0)	N_{TAIL}-Trx (at pH 7.0)	Trx-N_{TAIL} (at pH 6.0)
$\langle R_{2D} \rangle$, Single Gaussian fit (mean ± s.d.)	11.2 ± 3.8 nm	11.1 ± 3.6 nm	10.1 ± 2.5 nm
Estimation error of of $\langle R_{2D} \rangle$ (s.e.m.)	0.10 nm	0.14 nm	0.10 nm
$\langle H_1 \rangle$, height of Trx moieties Single Gaussian fit (mean ± s.d.)	2.3 ± 0.2 nm	2.3 ± 0.2 nm	2.3 ± 0.2 nm
Estimation error of $\langle H_1 \rangle$ (s.e.m.)	0.003 nm	0.003 nm	0.003 nm
$\langle H_2 \rangle$, height of N-terminal globule 1st peak of double Gaussian fit (mean ± s.d.)	0.8 ± 0.2 nm	0.8 ± 0.2 nm	0.8 ± 0.2 nm
Estimation error of the 1st peak of $\langle H_2 \rangle$ (s.e.m.)	0.003 nm	0.003 nm	0.02 nm
$\langle H_2 \rangle$, height of N-terminal globule 2nd peak of double Gaussian fit (mean ± s.d.)	1.1 ± 0.2 nm	1.1 ± 0.2 nm	1.1 ± 0.2 nm
Estimation error of the 2nd peak of $\langle H_2 \rangle$ (s.e.m.)	0.02 nm	0.02 nm	0.05 nm
F-test (Double Gaussian vs. Single Gaussian)	$P < 10^{-3}$	$P < 0.01$	$P < 0.05$
Area ratio of two Gaussian components in H_2 distribution (1st : 2nd), giving order propensity K_e	572 : 131 K_e = 0.23 (Box1)	233 : 45 K_e = 0.19 (Box1)	58 : 111 K_e = 1.91 (Box2)
λ, decay rate of autocorrelation function of time series H_2 data (best fit value ± s.e.m.)	7.49 ± 0.64 s^{-1}	7.16 ± 0.64 s^{-1}	3.13 ± 0.18 s^{-1}
k_{OD}, order-to-disorder transition rate estimated from H_2 dynamics (best fit value ± s.e.m.)	6.09 ± 0.67 s^{-1}	6.00 ± 0.78 s^{-1}	1.07 ± 0.17 s^{-1}
k_{DO}, disorder-to-order transition rate estimated from H_2 dynamics (best fit value ± s.e.m.)	1.40 ± 0.44 s^{-1}	1.16 ± 0.58 s^{-1}	2.06 ± 0.20 s^{-1}
$N_{aa}{}^{T}$	125	125	125
N_{aa} (from a.a. sequence)	88	88	88
Number of molecules analyzed	3	5	3
Number of frames analyzed	4,678	1,861	1,698
Number of frames used	4,858	2,377	2,888

$N_{aa}{}^{T}$ indicates the number of a.a. contained in N_{TAIL} (the N-terminal Met and Ser in N_{TAIL}-Trx are not included).

Supplementary Table 8 | Estimation from AFM images of persistence length (L_p) of IDRs contained in IDPs whose local structures have been characterized at the residue level.

Proteins	IDR region	N_{aa}	$\langle R_{2D} \rangle$ (nm) [a] mean ± s.e.m.	$\langle R_g \rangle$ (nm) [b] mean ± s.e.m.	L_p on mica (nm) mean ± s.e.m.	L_p in solution [c] (nm) mean ± s.e.m.
PQBP-1 (1−265)	82−265	184	17.2 ± 0.09	4.01 ± 0.021	1.16 ± 0.01	0.76 ± 0.0082
PQBP-1 (1−214)	82−214	133	14.4 ± 0.14	3.35 ± 0.033	1.15 ± 0.02	0.74 ± 0.015
PQBP-1 (1−164)	82−164	83	11.0 ± 0.04	2.56 ± 0.009	1.11 ± 0.01	0.72 ± 0.006
PQBP-1 (1−134)	82−134	53	9.3 ± 0.07	2.17 ± 0.016	1.35 ± 0.02	0.86 ± 0.015
PQBP-1 (1−114)	82−114	33	7.1 ± 0.02	1.65 ± 0.005	1.45 ± 0.01	0.89 ± 0.006
Atg1 (20 mM KCl, pH 6.0)	332−586	255	20.6 ± 0.53	4.80 ± 0.11	1.19 ± 0.06	0.77 ± 0.037
Atg1 (20 mM KCl, pH 7.0)	332−586	255	20.4 ± 2.25	4.75 ± 0.52	1.17 ± 0.26	0.76 ± 0.172
Atg1 (20 mM KCl, pH 8.0)	332−586	255	20.2 ± 0.52	4.71 ± 0.12	1.14 ± 0.06	0.74 ± 0.039
Atg1 (50 mM NaCl, pH 8.0)	332−586	255	20.5 ± 1.42	4.78 ± 0.33	1.18 ± 0.16	0.77 ± 0.110
Atg1 (300 mM NaCl, pH 8.0)	332−586	255	20.3 ± 0.94	4.73 ± 0.22	1.16 ± 0.11	0.75 ± 0.071
Atg1: averaged over the five conditions (mean ± s.d.)[d]			20.5 ± 0.9	4.78 ± 0.21	1.17 ± 0.02	0.77 ± 0.070
Atg13 (100 mM NaCl, pH 6.0)	268−744	477	27.9 ± 1.62	6.50 ± 0.38	1.15 ± 0.14	0.75 ± 0.088
Atg13 (100 mM NaCl, pH 7.0)	268−744	477	28.1 ± 4.62	6.55 ± 1.08	1.17 ± 0.39	0.76 ± 0.253
Atg13 (100 mM NaCl, pH 8.0)	268−744	477	27.8 ± 4.06	6.48 ± 0.95	1.14 ± 0.34	0.74 ± 0.220
Atg13: averaged over the three conditions (mean ± s.d.)[d]			28.0 ± 2.0	6.52 ± 0.47	1.15 ± 0.02	0.75 ± 0.109
FACT-10SA (dSSRP1 subunit)	433−554	121	14.3 ± 0.18	3.33 ± 0.042	1.26 ± 0.03	0.81 ± 0.022
N_{TAIL}-GFP	421−485, 503−525	88	11.6 ± 0.11	2.70 ± 0.026	1.16 ± 0.02	0.75 ± 0.015
N_{TAIL}-Trx (pH 6.0)	421−485, 503−525	88	11.2 ± 0.10	2.61 ± 0.023	1.08 ± 0.02	0.70 ± 0.013
N_{TAIL}-Trx (pH 7.0)	421−485, 503−525	88	11.1 ± 0.14	2.59 ± 0.033	1.05 ± 0.03	0.68 ± 0.018
N_{TAIL}: averaged over the above three cases (mean ± s.d.)[d]			11.3 ± 0.26	2.63 ± 0.061	1.10 ± 0.06	0.71 ± 0.035
Trx-N_{TAIL}	421−485, 503−525	88	10.1 ± 0.10	2.35 ± 0.023	0.86 ± 0.02	0.56 ± 0.012
					Average* 1.18 ± 0.08 mean ± s.d.	Average* 0.78 ± 0.06 mean ± s.d.

The nine $\langle R_{2D} \rangle$ data highlighted with orange color were applied to the analysis for determining the expansion factor u (see **Extended Data Fig. 8**).

[a]For IDPs whose IDRs undergo transitions, $\langle R_{2D} \rangle$ values shown here are those in the longer state.

[b]The $\langle R_g \rangle$ values in solution were estimated from corresponding measured $\langle R_{2D} \rangle$ values on mica using an expansion factor u = 1.24 by which mica-IDR interactions expand R_{2D}.

*Data of Trx-N_{TAIL} were omitted.

[c]The persistence length L_p in solution was calculated from $\langle R_g \rangle$ using *Eq*. (*S5*) shown below:

$$\langle R_g \rangle = \frac{L_p L}{3} - L_p^2 + \frac{2L_p^3}{L} - \frac{2L_p^4}{L^2} \times \left[1 - Exp\left(-L/L_p\right)\right]. \quad (S5)$$

[d]Weighted mean obtained using the areas of respective Gaussian distributions as weights.

Supplementary Table 9 | IDPs with L_p values in a range of 0.78 ± 0.08 nm.

Protein	N_{aa}	$\langle R_g \rangle$ (nm)	L_p (nm)	References
Osteopontin[a]	150	3.79	0.84	Ref. 25
Pig Calpastatin domain[a]	148	3.54	0.74	Ref. 26
HrpO[a]	147	3.50	0.73	Ref. 27
Human NHE ctd (45°C)[a]	131	3.53	0.84	Ref. 28
Neuroligin 3 (731−848)[a]	118	3.15	0.74	Ref. 29
Jaxtanodin (172−282)[a]	111	3.15	0.80	Ref. 30
P53 (1−93)[a]	93	2.87	0.80	Ref. 31
Activator for thyroid hormone and retinoid receptors (ACTR) (45°C)[a]	71	2.38	0.74	Ref. 28
Histatin 5[a]	24	1.33	0.86	Ref. 32
Nuclearporin Nup49 (121−154)	34	1.59	0.77	Ref. 33
Nuclearporin 153 kDa (884−993)	110	3.0	0.73	Ref. 33
Nuclearporin 153 kDa (1313−1390)	78	2.49	0.72	Ref. 33
ACTR (25°C)	79	2.51	0.73	Ref. 34
Phosphorylated insulin receptor (PID) interaction region in Grb14[b]	75	2.6	0.84	Ref. 35

[a]These IDPs were searched from the list compiled by Cordeiro et al (Ref. 24).

[b]Essentially unstructured but with a potentially structured short stretch encompassing residues 399–407.

By combining these 14 data of (N_{aa}, $\langle R_g \rangle$) with those of 10 tau-protein constructs and four IDR segments of PQBP-1 described in **Extended Data 8a (**red and blue circles, respectively**)** and 9 data shown in **Supplementary Table 8**, we obtained a power law of $\langle R_g \rangle = (0.260 \pm 0.021) \times N_{aa}^{(0.524 \pm 0.015)}$ (See **Extended Data Fig. 8b**).

Supplementary Table 10 | Parameters of charged residues contained in the IDRs (or segments thereof) used in this study.

Proteins	Start residue	End residue	Number of residues	Number of negative charges	Number of positive charges	Number of charged residues	FCR	Net charge	ρ	κ
PQBP-1	82	265	184	41	48	89	0.483	7	0.038	0.093
	82	214	133	37	39	76	0.571	2	0.015	0.089
	82	164	83	24	26	50	0.602	2	0.024	0.050
	82	134	53	12	14	26	0.491	2	0.038	0.102
N_{TAIL}	401	525	125	23	17	40	0.32	–6	–0.048	0.146
PNT	1	229	229	38	21	59	0.258	–17	–0.074	0.230
	1	102	102	17	11	28	0.275	–5	–0.059	0.240
	103	229	127	21	10	31	0.244	–11	–0.087	0.250
	177	229	53	6	8	14	0.264	2	0.038	0.257
Sic1	1	284	284	42	43	85	0.299	1	0.004	0.272
	30	284	255	42	40	82	0.322	–2	–0.008	0.269
	30	190	161	20	20	40	0.248	0	0	0.306
	191	284	94	22	20	42	0.447	–2	–0.021	0.256
	1	190	190	20	23	43	0.226	3	0.016	0.310
Atg1	332	586	255	28	24	52	0.204	–4	–0.016	0.175
Atg13	268	744	477	47	49	96	0.201	2	0.004	0.274

References

1. 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).
2. Kodera, N., Sakashita, M. & Ando, T. Dynamic proportional-integral-differential controller for high-speed atomic force microscopy. Rev. Sci. Instrum. **77**, 083704 (7 pages) (2006).
3. Wright, P. E. & Dyson, H. J. Linking folding and binding. *Curr. Opin. Struct. Biol*. **19**, 31–38 (2009).
4. Wang, Y., Chu, X., Longhi, S., Roche, P., Han, W., Wang, E. & Wang, J. Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. *Proc. Natl. Acad. Sci. U. S. A.* **110**, E3743–E3752 (2013).
5. Dosnon, M., Bonetti, D., Morrone, A., Erales, J., di Silvio, E., Longhi, S. & Gianni, S. Demonstration of a folding after binding mechanism in the recognition between the measles virus NTAIL and X domains. *ACS Chem. Biol*. **10**, 795–802 (2015).
6. Takahashi, M., Mizuguchi, M., Shinoda, H., Aizawa, T., Demura, M., Okazawa, H. & Kawano, K. Polyglutamine tract-binding protein-1 binds to U5-15kD via a continuous 23-residue segment of the C-terminal domain. *Biochim. Biophys. Acta* **1804**, 1500–1507 (2010).
7. Zor, T., Mayr, B. M., Dyson, H. J., Montminy, M. R. & Wright, P. E. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem*. **277**, 42241–42248 (2002).
8. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025 (2007).
9. Boze, H. et al. Proline-rich salivary proteins have extended conformations. Biophys. J. 99, 656–665 (2010).
10. Kate M. Nairn, K. M. et al. A synthetic resilin is largely unstructured. Biophys. J. 95, 3358–3365 (2008).
11. Salmon, L. et al. NMR characterization of long-range order in intrinsically disordered proteins. J. Am. Chem. Soc. 132, 8407–8418 (2010).
12. Mylonas, E. et al. Domain conformation of tau protein studied by solution small-angle X-ray scattering. Biochemistry 47, 10345–10353 (2008).
13. 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).
14. Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc. Natl. Acad. Sci. U. S. A*. **110**, 13392–13397 (2013).
15. 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).
16. 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).

17. 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).

18. Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. and Pappu, R. V. CIDER: Resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. *Biophys J.* **112**, 16–21 (2017).

19. http://www.disprot.org/.

20. Piovesan, D. et al. DisProt 7.0: a major update of the database of disordered proteins. Nucl. Acid Res. **4**, 219–D227 (2017).

21. Hirose, S., Shimizu, K., Kanai, S., Kuroda, Y. & Noguchi, T. POODLE-L: a two-level SVM prediction system for reliably predicting long disordered regions. Bioinformatics 23, 2046–2053 (2007).

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

23. Bernado, 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).

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

25. Lenton, S., Seydel,T., Nylander, T., Holt, C., Hartlein, M., Teixeira, S., Zaccai. G. Dynamic footprint of sequestration in the molecular fluctuations of osteopontin. *J. R. Soc. Interface* **12**, 0506 (2015).

26. Konno, T., Tanaka, N., Kataoka, M., Takano, E., Maki, M. A circular dichroism study of preferential hydration and alcohol effects on a denatured protein, pig calpastatin domain I. *Biochim. Biophys. Acta* **1342**, 73–82 (1997).

27. Gazi, A. D., Bastaki, M., Charova, S. N., Gkougkoulia, E. A., Kapellios, E. A., Panopoulos, N. J., Kokkinidis, M. Evidence for a coiled-coil interaction mode of disordered proteins from bacterial type III secretion systems. *J. Biol. Chem*. **283**, 34062–14068 (2008).

28. Kjaergaard, M., Nørholm, A. B., Hendus-Altenburger, R., Pedersen, S. F., Poulsen, F. M., Kragelund, B. B. Temperature-dependent structural changes in intrinsically disordered proteins: formation of alpha-helices or loss of polyproline II? *Protein Sci.* **19**, 1555–1564 (2010).

29. Paz, A., Zeev-Ben-Mordehai, T., Lundqvist, M., Sherman, E., Mylonas, E., Weiner, L., Haran, G., Svergun, D. I., Mulder, F. A. A., Sussman, J. L., Silman, I. Biophysical characterization of the unstructured cytoplasmic domain of the human neuronal adhesion protein neuroligin 3. *Biophys. J*. **95**, 1928–1944 (2008).

30. Ruskamo, S., Chukhlieb, M., Vahokoski, J., Bhargav, S. P., Liang, F., Kursula, I., Kursula, P. Juxtanodin is an intrinsically disordered F-actin-binding protein. *Sci. Rep*. **2**, 899 (2012).

31. Wells, M., Tidow, H., Rutherford, T. J., Markwick, P., Jensen, M. R., Mylonas, E., Svergun, D. I., Blackledge, M., Fersht, A. R. Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation

domain. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 5762–5767 (2008).

32. Cragnell, C., Durand, D., Cabane, B., Skepö, M. Coarse-grained modelling of the intrinsically disordered protein Histatin 5 in solution. Monte Carlo simulations in combination with SAXS. *Proteins* **84**, 777–791 (2016).

33. 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).

34. 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).

35. Moncoq, K. *et al*. SAXS study of the PIR domain from the Grb14 molecular Adaptor: A natively unfolded protein with a transient structure primer? *Biophy. J*. **87**, 4056–4064 (2004).