Supplementary Information

Materials and Methods

Expression and purification of Npm-N: Proteins were expressed and purified as described previously[1]. Briefly, all constructs were cloned into pET28a(+) (Novagen) vector, between the *NdeI* and *XhoI* sites, in frame with a N-terminal poly-histidine tag and a Thrombin cleavage site. Point mutations were introduced using the QuikChange kit (Agilent). Proteins were expressed in *E. Coli* BL21(DE3) (Novagen) and purified by Ni-NTA affinity and size exclusion chromatography (HiLoad 26/60 Superdex 200, GE Healthcare) in 10 mM Tris, 150 mM NaCl, 2 mM DTT, pH 7.5 buffer. All the purified proteins had three additional residues (GSH) at the N-terminus, and we used an asterisk (*) in the text to distinguish the first residue (i.e., G1*) from the endogenous N-terminus of the protein.

Arf peptides production: Arf6 and Arf14 peptides were synthesized using standard solid phase peptide synthesis chemistry by the Molecular Synthesis resource at Hartwell Center for Bioinformatics and Biotechnology at St. Jude Children's Research Hospital. Arf37 peptide was recombinantly expressed in *E. Coli* BL21(DE3) (Novagen) and purified as described previously[2].

Labeling: The mono- and dual-labeling of Npm-N for the ensemble and smFRET experiments, respectively, were carried out using cys-maleimide chemistry as described previously[3] with some modifications. For ensemble experiments, Alexa594 and Alexa488 fluorophores were used to mono-label Npm-N at C104 in separate batches using the same conditions as dual labeling. The native sequence of Npm-N has 2 cysteines, C21 and C104. C21 is completely buried in the pentameric core, and hence it is sterically inaccesible for labeling, while C104 is positioned at a solvent exposed loop and available for conjugation in the pentameric structure[1, 4]. Therefore, we took advantage of this differential steric accessibility of these two native cysteins, and selectively labeled C104 in presence of C21 almost quantitatively. The mono-labeling at C104 was confirmed by LC-MS (Scripps Center for Mass Spectrometry). However, since C21 becomes exposed in the disordered S48E construct, we introduced an additional C21T mutation in the latter, thereby ensuring selective labeling of the C104 position in this case. Dual labeling was achieved with an excess of the donor (Alexa488 C5-maleimide derivative, Molecular Probes), and the acceptor dyes (Alexa594 flurophore C5-maleimide derivative, Molecular Probes) in one pot in 10 mM potassium phosphate, 150 mM NaCl, pH 7.5 at 4 ^{o}C, with overnight incubation, in the dark. The excess dye in all labeling reactions was removed by multiple rounds of washing with the labeling buffers using a 3K MWCO centrifugal filter device (Millipore). The labeling efficiency for all the mono-labeled samples was ≥ 90% (UV-Vis absorption measurements), and these were used without any further purification. The dual-labeled samples were further purified by HPLC with a C4 column. The purity of all the samples (including non-specific dye-binding) were checked by SDS-PAGE and LC-MS (Scripps Center for Mass Spectrometry).

CD spectroscopy and data analysis: CD measurements were performed using a Jasco J-815 Spectropolarimeter equipped with a temperature controller in a 1 mm path-length cuvette. All the measurements were performed at 25^{o} C. The unfolding measurements were done in 10 mM

potassium phosphate, 7.5 mM NaCl, pH 7.5, and the folding measurements were carried out in 10 mM potassium phosphate, 150 mM NaCl, pH 7.5. Variable protein concentrations were used to determine the concentration dependence of the the folidng and unfolding kinetics. The folding reaction was initiated by manual addition of 150 mM NaCl (final concentration) from a 5.08 M aqueous stock solution. The kinetic traces for folding and unfolding were fitted to a single-exponential model, which resulted in a reasonable fitting judged by the residual analysis (Supplementary Figures. 2a and 7a).

Bulk fluorescence spectroscopy and data analysis: Bulk fluorescence measurements were performed using an automated temperature controlled PC1 spectrofluorometer (ISS, Champaign, IL) at 25^{o} C. The ratio of donor:acceptor labeled protein was 1:1 in our measurements. Varying this ratio up to 1:4 did not have any significant effect on the kinetics. Association and dissociation kinetics were measured in *high salt* and *low salt buffer*, respectively. The association process was initiated by manual addition of the appropriate amount of NaCl from a 5.08 M aqueous stock solution to acheive the desired final concentration. For assembly experiments with Arf6, [Npm-N] = 10 μM, and [Arf6] = 500 μM were used. For disassembly experiments in presence of various Arf fragments, the concentrations of WT, T95D and S125E Npm-N were 0.5 μM, and 2 μM for the double mutant (T95D/S125E). The protein:peptide ratio for the samples was 1:1000 for Arf6 (such a large excess is used since K_D for Npm-N/Arf6 complex is ~ 1 mM [1]), 1:20 for Arf14 and 1:9 for Arf37.

The disassembly kinetics showed first order kinetics (concentration independent) and were poorly fitted by a single-exponential model. To improve the fit, we tested multi-exponential models. Residual analysis showed that an exponential model with at least three phases resulted in a reasonable fit (Supplementary Figure-7b). Therefore, we used a triple-exponential model to fit all our disassembly data. On the other hand, the assembly kinetics were fitted by a quotient of exponential functions given by [5]:

$$Y = (Y_{Eq}) \cdot (1 - e^{-\frac{t}{\tau}})/(1 + \omega \cdot e^{-\frac{t}{\tau}}) \quad (1)$$

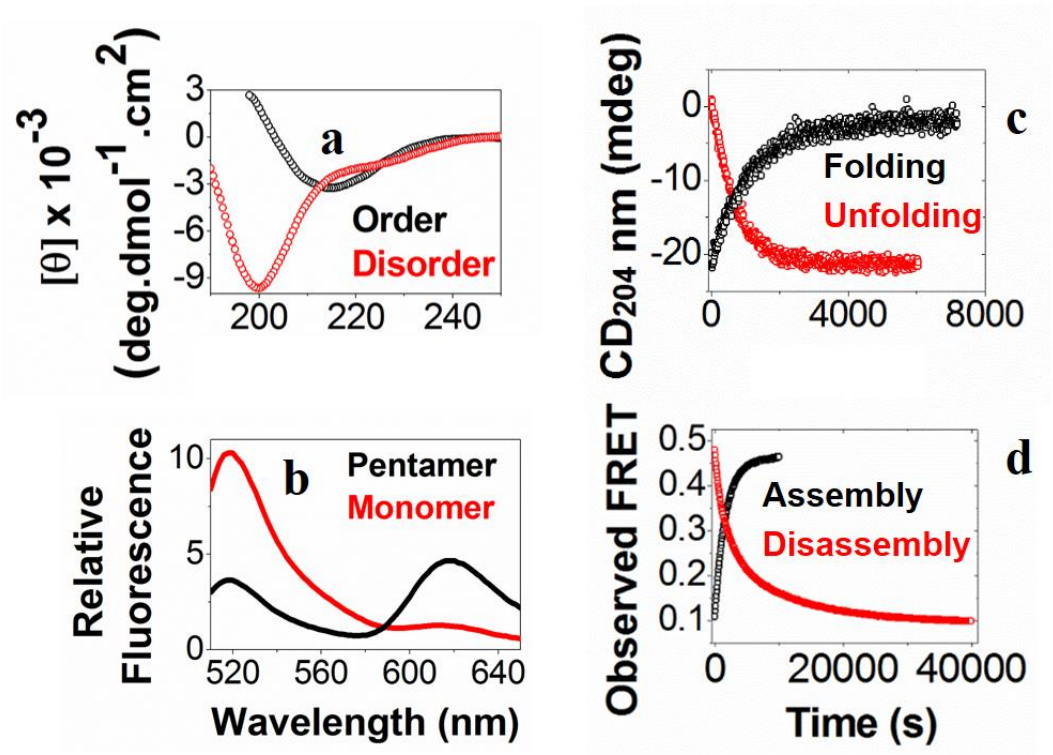
Y_{Eq} = equilibrium value of the observable (i.e., FRET in our case); τ = time constant.

Here, ω parameter describes the deviation of the experimental results from a first order kinetics, and assumes values in the $-1 < \omega < 1$ range[5]. For Npm-N concentrations ≥7.5 μM, we observed that the value of ω approached ~ 0, indicating a first order kinetics. This is in line with our observation that the assembly kinetics becomes independent of the protein concentration at ≥7.5 μM Npm-N (Figure 1c).

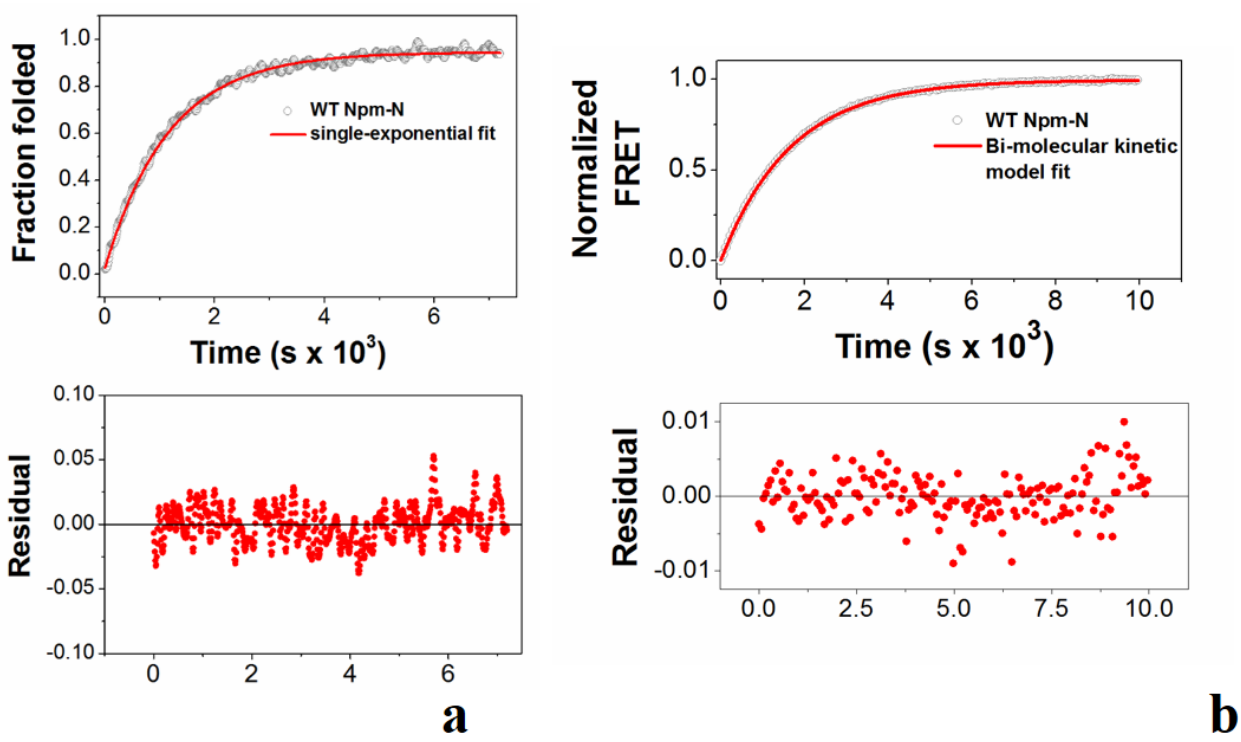
Single-molecule fluorescence spectroscopy: The details of smFRET instrumentation, data collection and data analysis have been described elsewhere[6]. All the experiments were performed at room temperature either in high salt buffer for folding (10 mM potassium phosphate, 150 mM NaCl, pH 7.5) or in low salt buffer (10 mM potassium phosphate, 7.5 mM NaCl, pH 7.5) for

unfolding using dual-labeled (at positions 15 and 104; 1* and 104) Npm-N at a concentration of ~ 100 pM in the presence of 10 µM unlabeled Npm-N. The folding reaction was initiated by addition of 150 mM NaCl (final concentration) manually from a 5.08 M aqueous stock solution.

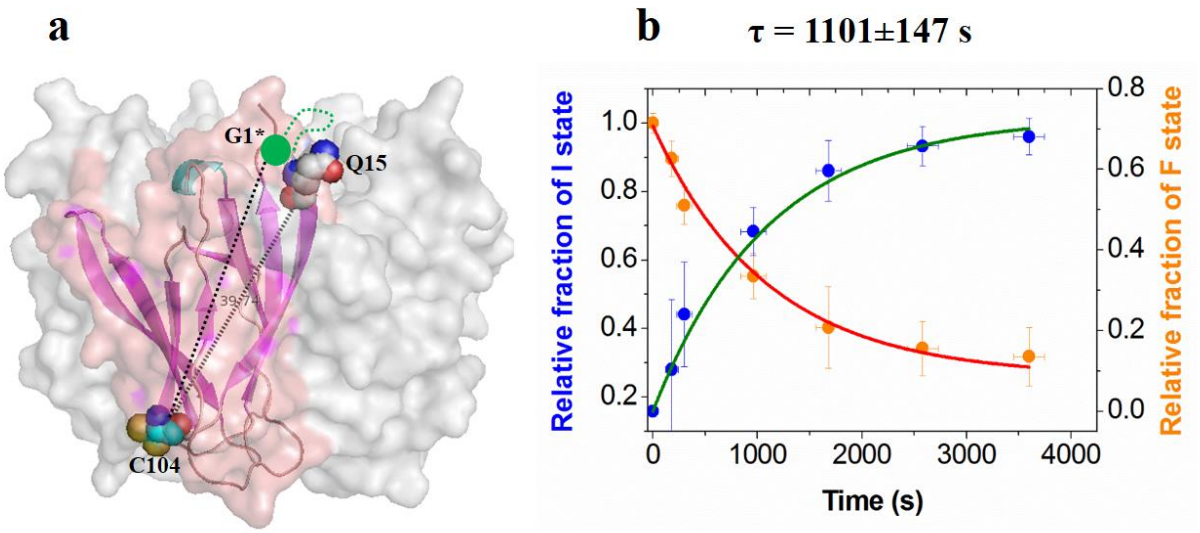
Supplementary Figures



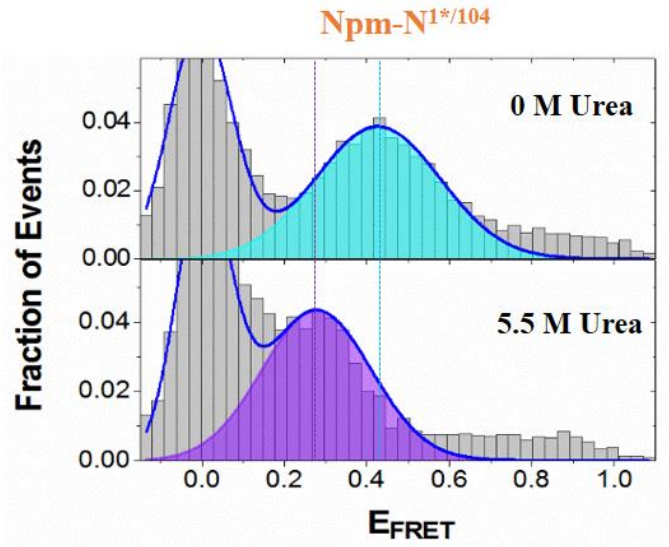
Supplementary Figure-1: Order-disorder and monomer-pentamer transitions of Npm-N: spectra and unproccessed kinetic data. (a) Far UV CD spectra of Npm-N in the unfolded monomeric state (red; recorded at 7.5 mM NaCl), and folded pentameric state (black; recorded at 150 mM NaCl). (b) Representative fluorescence spectra showing high inter-molecular FRET in the pentameric Npm-N (black; recorded at 150 mM NaCl), compared to the monomeric Npm-N (red; recorded at 7.5 mM NaCl). For the FRET experiments, individual donor (Alexa488) and acceptor (Alexa594) labeled Npm-N were mixed at low salt, then switched to high salt. The spectra corresponding to the end points are shown here. (c) Unprocessed kinetic data showing representative traces for unfolding (order-to-disorder) and folding (disorder-to-order) of Npm-N, at 7.5 mM and 150 mM NaCl, respectively, using CD experiments. (d) Unprocessed kinetic data showing representative traces for assembly (monomer-to-pentamer) and disassembly (pentamer-to-monomer) of Npm-N, at 150 mM NaCl and 7.5 mM , respectively, using ensemble FRET (inter-molecular) experiments. These CD and FRET assays were used to extract the time-constants of folding/unfolding and assembly/disassembly of Npm-N, as described in the *main text* and Figures-1b and 2a. Please see Methods for further experimental details.



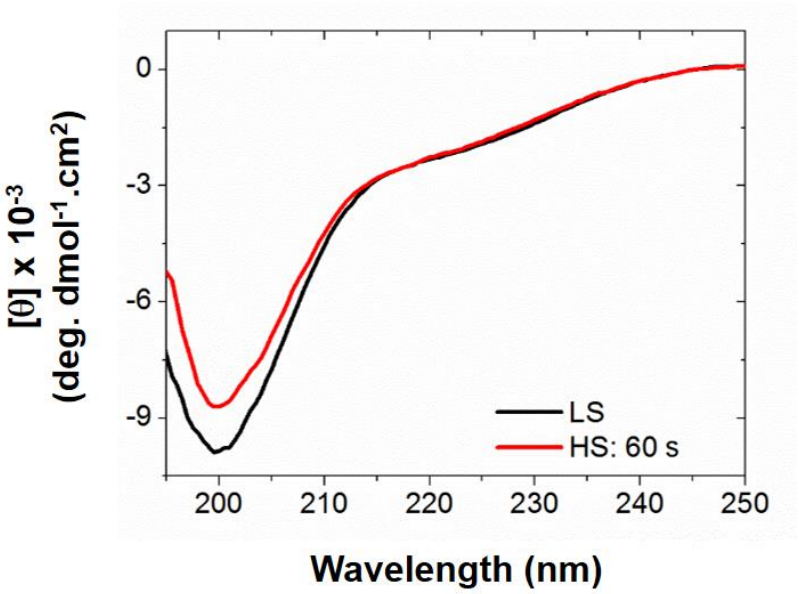
Supplementary Figure-2: Plot of the residuals from the data fitting of folding and oligomerization kinetics of Npm-N. (a) shows that folding kinetics of Npm-N can be well described by a single-exponential model, while the association data in (b) were fitted with the equation-1 using the kinetics model described in the Methods section.



Supplementary Figure-3: smFRET labeling sites and the single-molecule time constant of I_M to F conversion of Npm-N. (a) Cartoon representation of the crystal structure of Npm-N (PDB ID: 4N8M) showing the relative distance between the sites 15 and 104 in a monomeric subunit in the pentamer. The G1* is also displayed as a cartoon in this structure to indicate that the relative distance between C104 and G1* is similar to that between C104 and C15 (40 Å). This is based on our observation of very similar FRET efficiency of these two constructs (see Figure 1d and Supplementary Figure-6) (b) Time constant of salt (150 mM)-induced folding of Npm-N extracted from smFRET histograms. Considerable overlap of the peaks corresponding to I and F states (see Figure 1d) resulted in an increased error in the estimated time constant of folding. Nevertheless, the estimated time constant agrees reasonably well with the time constant of folding (1167±17 s) determined using CD spectroscopy (Figure 1b; Supplementary Table-2).

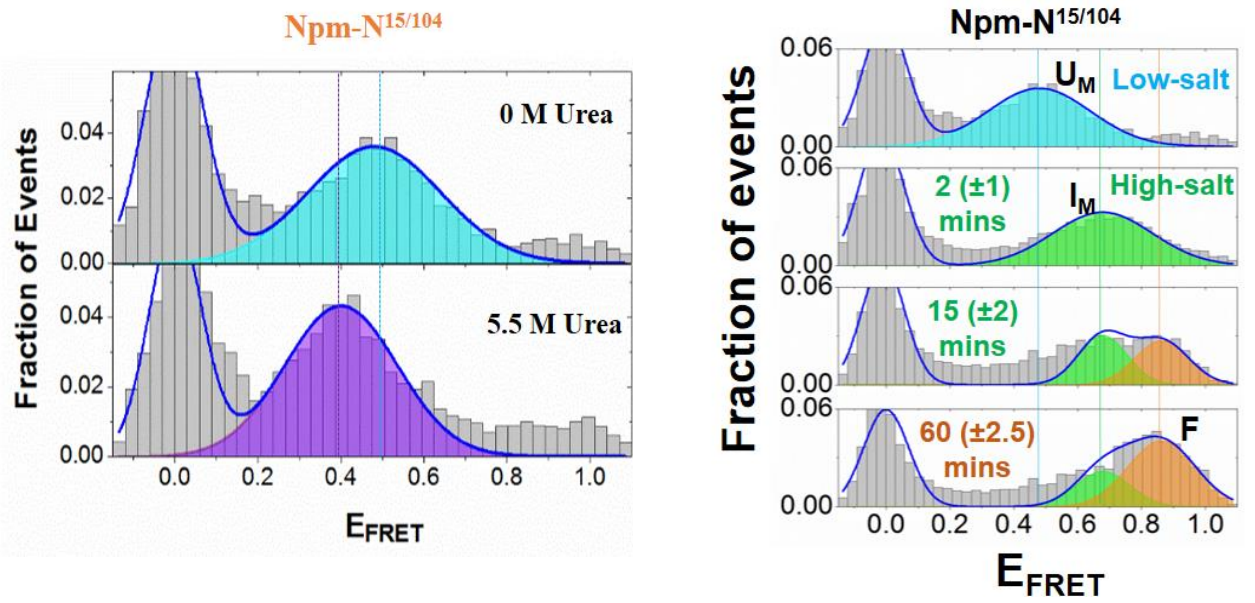


Supplementary Figure-4: Denaturation effect on the disordered state of Npm-N. smFRET histograms of Npm-N showing the natively disordered state at low salt condition and in presence of 5.5 M urea (dual-labeled at 1 and 104 positions) at 10 mM potassium phosphate buffer, 7.5 mM NaCl, pH 7.5. The E_{FRET} shifts from 0.45 (0 M urea) to 0.28 (5.5 M urea).



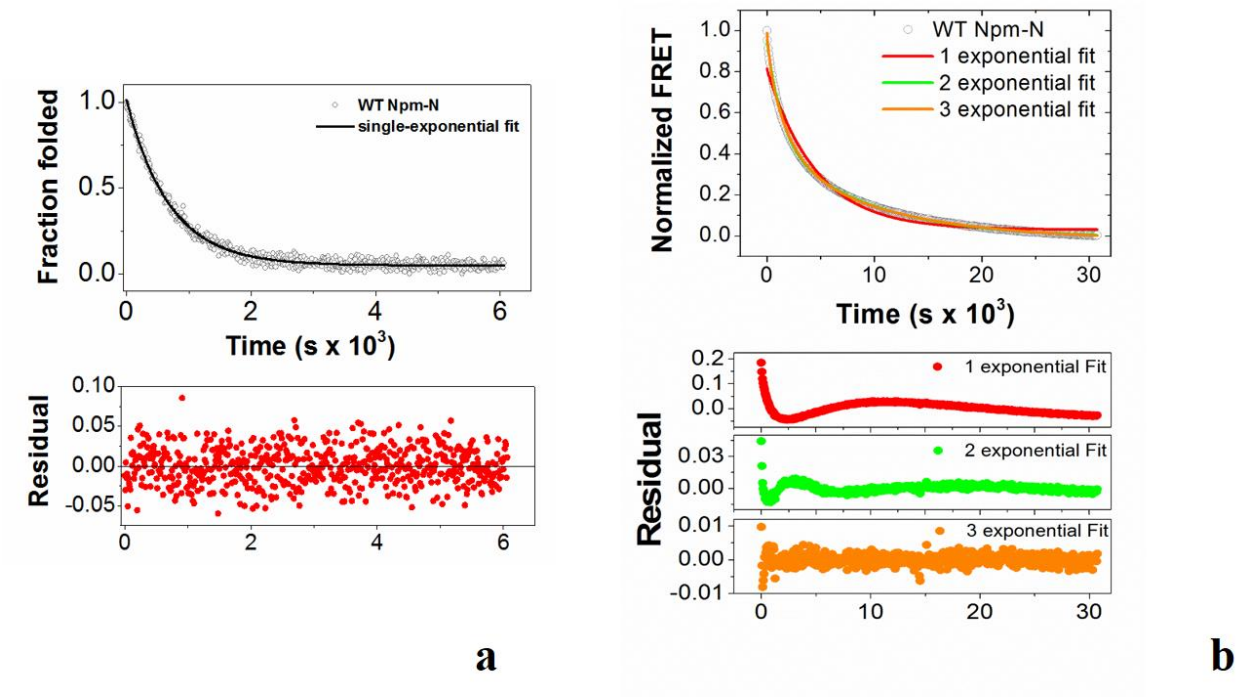
Supplementary Figure-5: Salt-induced collapse of the Npm-N during folding (compare with 2nd panel from the top in Figure 1d and Supplementary Figure-6b). Far UV CD spectra of Npm-N under low-salt condition (7.5 mM NaCl, black), and after 1 min of addition of 150 mM NaCl (red). These data show only very minor alteration in the secondary structure content of Npm-N in the salt-induced collapsed state, relative to the disordered state. Using the online server Bestsel[7], we have quantitatively analyzed this difference, which revealed ~2% increase in β-sheet content in the HS spectrum at 60 s.

a **b**

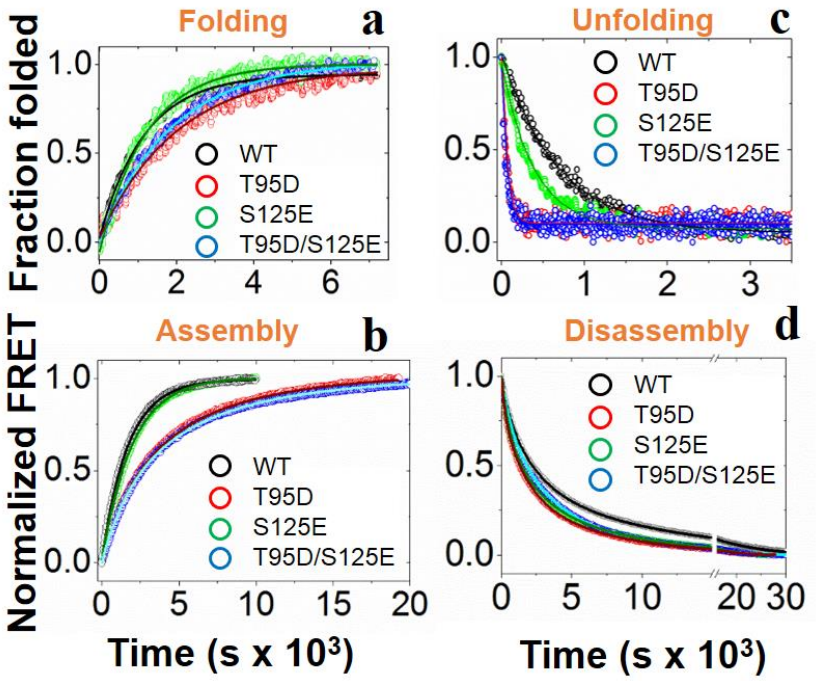


Supplementary Figure-6: Coupled folding–assembly kinetics of Npm-N using smFRET: Here, we used dual-labeled Npm-N^{Q15C} with a FRET pair (Alexa488/Alexa594) covalently attached at cysteine residues at positions 15 (introduced by mutation) and 104 (present in the native sequence; Supplementary Figure-3a). This labeling was done using conventional cys-maleimide chemistry. Although widely used in single-molecule fluorescence experiments, the problem with this approach is as follows: since both positions were labeled using identical chemistry, an unavoidable result is scrambling of the donor and acceptor fluorophores between these two sites, and the dual-donor and dual-acceptor labeled protein were also formed as a side product. While this scrambling was not a substantial problem in our measurements, since only protein molecules with both donor and acceptor attached provided smFRET information, and dual-donor actually contributed to the "zero-peak", further optimization of reaction condition and additional purification steps were needed (see methods for details). As described in Supplementary Note-1, we developed an orthogonal approach of labeling with PLP mediated N-terminal modification, and resolved these issues of dual labeling of proteins for SM fluorescence experiments (see Supplementary Note-1 for details on this scheme). Nevertheless, with this dual-labeled construct (termed as Npm-N 15/104), the disordered monomeric state (U) in low salt buffer is characterized by a FRET efficiency (E_{FRET}) value of 0.48. **(a)**. **(b)** Transitioning from low to high salt conditions,

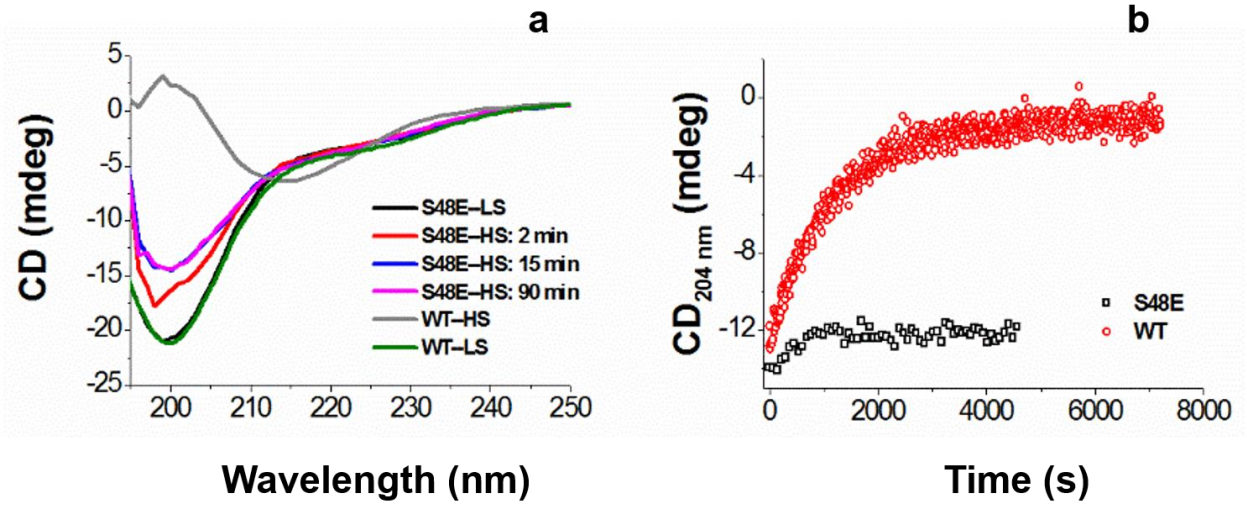
a relatively rapid shift in the smFRET peak to a higher E_{FRET} value of ~ 0.65 (2nd panel from the top) was recorded. This is likely due to collapse of the disordered state of Npm-N as a response to the increased concentration of salt, also observed for the other smFRET construct Npm-N 1*/104 (Figure-1d). We note that previously, similar salt induced fast collapse of disordered polypeptide chains with substantial net charge has been demonstrated using smFRET by Schuler and colleagues[8]. Consistent with this notion, since the net charge of Npm-N at pH 7.0 is -16[9], a similar charge screening by salt is likely to be the cause of the rapid collapse of the protein backbone without any significant formation of secondary structure. In addition, the folded conformation was signaled by the appearance of a new peak at a higher E_{FRET} value of ~ 0.85 with a time constant (τ ~ 17 mins) similar to that obtained in the bulk Npm-N folding measurements (19.45 ± 0.3 mins, by CD). The high value of E_{FRET} ~ 0.85 is expected for the folded, pentameric state (F_P state) based on the crystal structure of Npm-N (Supplementary Figure-3a). The Förster distance (R_o) for Alexa488/Alexa594 dye pair is 54 Å, while the distance (r) between the FRET sites in the protein (Q15 and C104) is calculated as 40 Å (Supplementary Figure-3a). A simple calculation, based on the equation $E_{FRET} = {1}/{(1 + \frac{r}{R_o})}$ yields an estimate for the value of E_{FRET} of 0.91 (considering γ to be 1.0[3]), which is consistent within error of the value measured (0.85) given that linkers and dyes would result in an increase in the distance and therefore somewhat lower E_{FRET}. These data therefore reinforce the findings reported in the *main text* using a different dual-labeled Npm-N construct (Npm-N 1*/104). Thus, our smFRET data directly reveal that the folding of the monomeric disordered Npm-N involves at least three states, among which the I_M state was largely undetected in the ensemble folding experiments using CD spectroscopy. We note that both Npm-N 15/104 and Npm-N 1*/104 did not show any significant changes in E_{FRET} between F_M and F_P states, suggesting a lack of major conformational changes in Npm-N monomers during assembly process. This is also consistent with the folding-induced-assembly mechanism as discussed in the *main text*. The solid lines in this figure represent fitting of the experimental data to a Gaussian model. The peak at zero is due to molecules lacking an active acceptor dye. The time values in the parenthesis indicate uncertainties due to finite data acquisition time.



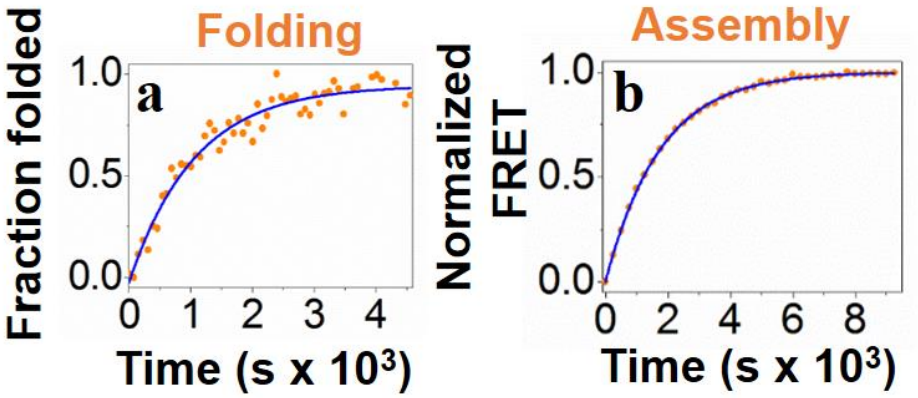
Supplementary Figure-7: Plot of the residuals from the data fitting of unfolding and disassembly kinetics of Npm-N. (a) shows that unfolding kinetics of Npm-N can be well described by a single-exponential model, while a three exponential model was required to achieve an acceptable fitting of the dissociation data (b).



Supplementary Figure-8: Effect of phospho-mimic mutations on the folding-assembly and unfolding-disassembly kinetics of Npm-N. (a) and (b) show the folding-coupled-assembly kinetics and (c) and (d) show unfolding-coupled-disassembly kinetics for the T95D, S125E and T95D/S125E mutants, relative to the wild-type Npm-N. The folding data were obtained from CD measurements, while the assembly kinetics were extracted from inter-molecular FRET experiments (Supplementary Fig. 1). These data indicate that phospho-mimic mutations impede assembly more efficiently than folding in the forward pathway, while they accelerate unfolding more dramatically than disassembly in the reverse pathway. The time-constants extracted from these traces are summarized in Supplementary Table-2.



Supplementary Figure-9: Salt-induced collapse of the S48E mutant of Npm-N. (a) Overlay of the far UV CD spectra of WT and S48E mutant at low salt (LS; 7.5 mM NaCl) and high salt (HS; 150 mM NaCl) to indicate that the S48E mutation prevents Npm-N from adopting the folded, β-sheet rich native state, and traps the protein in a disordered state. Also shown are the three additional time points for S48E mutant. The online server Bestsel[7] was used to analyze the differences in the secondary structure content between S48E HS and LS states, and indicated only ~ 4% gain in the β-sheet and a concomitant decrease in the random coil structure at HS. (b) Temporal changes in the CD signal of the phospho-mutant S48E of Npm-N, as compared to the WT protein, during folding reaction at 150 mM salt.

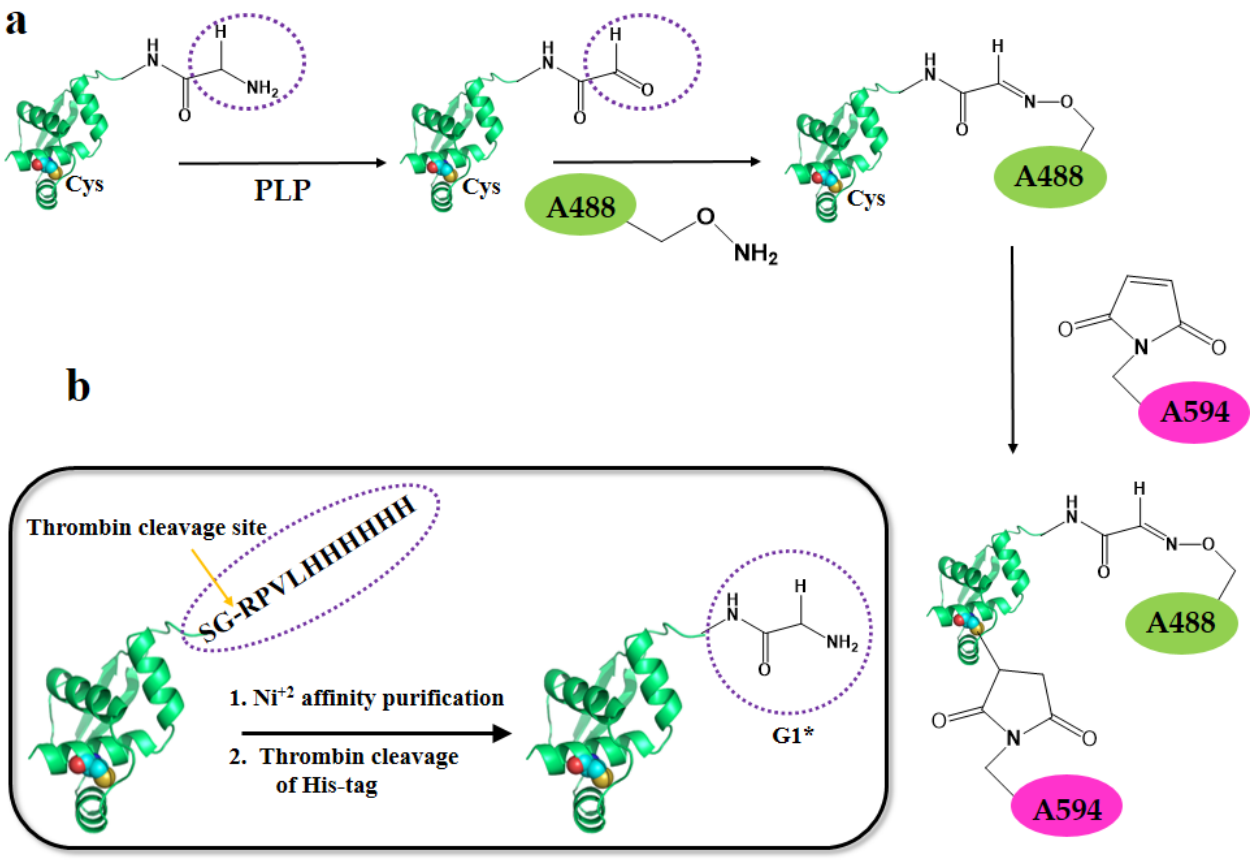


Supplementary Figure-10: Effect of binding partner on the order-disorder/assembly-disassembly transitions of Npm-N. R-rich linear binding motifs in the protein binding partners were previously shown to modulate the conformational equilibrium of Npm-N by favoring the folded pentameric conformation[1]. Using Far UV CD (Supplementary Figure-1a) and bulk fluorescence measurements (Supplementary Figure-1b) at low salt (7.5 mM NaCl), we observed that a six-residue R-rich fragment of Arf, corresponding to the N-terminus of Arf (Arf6), induced folding-assembly of Npm-N following the same pathway as observed at high salt, i.e., oligomerization-upon-monomer folding. Shown here are **(a)** folding kinetics, and **(b)** assembly kinetics. In the presence of 500 µM Arf6 (1:50 molar ratio), the time-constant for folding was τ_F = 1060±84 s and the observed time-constant for pentamer formation was τ_A^{obs}= 1764±15 s. These time-constants are comparable to the time-constants obtained at 150 mM NaCl in the absence of Arf (Supplementary Table-2).

Supplementary Note-1: Site-specific labeling of Npm-N for smFRET using orthogonal chemistry

The basic requirement of protein sample preparation for single-molecule fluorescence experiments is the site-specific covalent attachment of fluorescence dyes. The most commonly used practice is site-specific introduction of cysteine(s) by mutagenesis (if the protein of interest (POI) does not contain any native cysteine(s) at the specific structural site(s) under investigation), which can be subsequently labeled with thiol reactive fluorescence dye(s) of choice[10]. However, this technique poses one principal concern of whether the Cys-variants of the POI will have native-like properties, especially for smFRET measurements where dual labeling is necessary. Therefore, for smFRET measurements, the incorporation of two Cys residues in the structure of POI are often necessary. Also, site-specificity is limited by the identical chemistry between the dyes and the side chain of Cys, hence scrambling of the donor and acceptor dyes cannot be controlled[10a]. Alternatively, in order to achieve site-specificity, a more complex approach can be taken which would involve site-specific introduction of unnatural amino acids, and labeling via orthogonal chemistries[11].

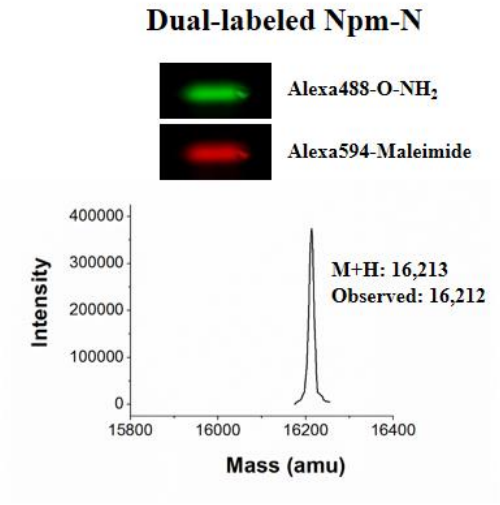
Herein, we report a much simpler scheme to generate site-specifically dual-labeled protein samples for smFRET experiments using orthogonal chemistries. Our approach is based on the application of pyridoxal-5'-phosphate (PLP)-mediated N-terminal amine modification of proteins which was originally developed by Francis et al.[12], in conjunction with conventional thiol-maleimide chemistry. This approach has the advantage of one less Cys mutation in the POI compared to conventional methods involving either dual Cys variants or a combination of one Cys mutation and one unnatural amino acid constructs[11]. Moreover, we demonstrate that the proposed labeling scheme can be coupled with a prior affinity-based protein purification technique, and is easy to implement for a diverse class of proteins with a general applicability. The general scheme of our method (Supplementary Figure-11a) is as follows: (i) modification of N-terminus of the POI by PLP using biomimetic transamination chemistry to generate a ketone functional moiety, followed by labeling with a ketone-reactive fluorophore, and (ii) labeling at a site-specific cysteine using thiol-reactive dye. Since the PLP-catalyzed transamination reaction has been shown to have some selectivity among various amino acids[13], and Gly being one of the better amino acids in terms of reaction yield, we have devised a strategy that couples the recombinant protein purification after expression in *E. Coli* cells with the subsequent fluorescence labeling. In this scheme, we genetically engineer the plasmid containing the POI sequence to incorporate an N-terminal His-tag, followed by a highly specific protease (thrombin) cleavage site (LVPR-GS) (the cleavage site is indicated by "–"). The POI can be expressed in bacterial cells, followed by one step purification by affinity chromatography using Ni-NTA column. The exogenous His-tag in the purified POI can subsequently be removed using the protease treatment, which will result in an N-terminus Gly (Supplementary Figure- 11b). Finally, the site specific two-step protein labeling can be carried out using the reaction scheme as described in Supplementary Figure-11a.



Supplementary Figure-11: General scheme for the coupled purification and site-specific dual labeling of proteins using an orthogonal N-terminal/Cys approach: (a) the two-step labeling of a POI (in this case Npm-N) which involves generation of a ketone functional group at the N-terminus by PLP action, followed by coupling with hydroxylamine derivative of Alexa488 dye (donor fluorophore). The 2nd step involves coupling of maleimide derivative of Alexa594 (acceptor fluorophore) with the thiol group of a cysteine residue in the POI. (b) The prior purification strategy of the recombinant POI using an exogenous His-tag, flanked by a thrombin cleavage site. The protease-induced cleavage of the His-tag after one step affinity purification leads to a Gly residue at the N-terminus, which has a better reactivity towards PLP-mediated transamination reaction[14].

Using this approach, we dual labeled Npm-N in two steps. First, we introduced a ketone group by PLP-action at the N-terminus of Npm-N, followed by the covalent attachment of the hydroxylamine derivative of Alexa488 (donor). In the final step of the labeling reaction, the native

C104 was labeled using maleimide derivative of Alexa594 fluorophore (acceptor). We note that Npm-N has another Cys (C21) in its native state, which is not solvent accessible in the folded, pentameric state, while C104 is completely solvent accessible (PDB ID: 4N8M). We took advantage of these differential steric accessibility properties of native cysteines in the folded Npm-N during selective labeling of C104 (see Methods). The introduction of the two dyes was checked independently using SDS-PAGE and ESI-mass spectrometry (ESI-MS) (Supplementary Figure-12). The smFRET experimental data using this construct has been shown in Figure 1d in the *main text*, and Supplementary Figure-4.



Supplementary Figure-12: Characterization of the 1*/104 dual-labeled Npm-N: SDS-PAGE showing incorporation of the donor and acceptor dyes in Npm-N (*top*), while ESI-MS confirms the presence of the dual labeled Npm-N (*bottom*). The respective calculated and observed masses for the dual-labeled protein are indicated. The gel images (top panel) were collected using a FluorChem M (ProteinSimple) using the following settings: blue channel (Alexa 488 fluorophore): excitation 475nm/emission 537nm (26nm); green channel (Alexa 594 fluorophore): excitation 534nm/emission 607nm (36nm).

Supplementary Note-2: A kinetic model for the effect of phospho-modifications on the order-disorder transition pathways of Npm-N

We describe each individual observed step in the folding-assembly and unfolding-disassembly pathways as a two state process with a corresponding transition state (TS) (Supplementary Figures-13; 14a,b). In that case, one can use the Arrhenius equation to estimate the relative alteration in the TS energy, $\Delta\Delta G^{\neq}$, as described below:

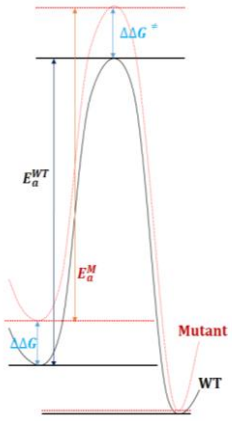
Arrhenius equation: $$\ln(k) = -\frac{E_a}{R \cdot T} + \ln(A)$$

Therefore: $$\ln(\frac{k^{WT}}{k^{T95D}}) = -\frac{{E_a}^{WT}}{R \cdot T} + \ln(A^{WT}) - (-\frac{{E_a}^{T95D}}{R \cdot T} + \ln(A^{T95D}))$$

$$\ln(\frac{k^{WT}}{k^{T95D}}) = \frac{1}{R \cdot T} \cdot ({E_a}^{T95D} - {E_a}^{WT})$$

(Assuming $A^{WT} = A^{T95D}$)

$$(\mathrm{R} \cdot \mathrm{T}) \cdot \ln(\frac{k^{WT}}{k^{T95D}}) = ({E_a}^{T95D} - {E_a}^{WT}) \qquad (1)$$



Supplementary Figure-13: Schematic representation of a two state process for WT and mutant Npm-N, highlighting the relative changes in the ground state and transition state energies due to mutation.

In Supplementary Figure-13, we have shown a schematic representation of a two state process following Arrhenius kinetics, where the relative destabilization in the ground state of the protein due to mutation is given by $\Delta\Delta G$, while the same for the transition state (TS) is given by $\Delta\Delta G^{\neq}$. Therefore, the $\Delta\Delta G^{\neq}$ can be expressed as:

$$\Delta\Delta G^{\neq} = ({E_a}^{T95D} + \Delta\Delta G) - {E_a}^{WT}$$

$$\Delta\Delta G^{\neq} = ({E_a}^{T95D} - {E_a}^{WT}) + \Delta\Delta G \qquad (2)$$

Using the experimentally measured time constants from Supplementary Table-2, and the calculated $\Delta\Delta G$ values for the disordered monomer using eScape algorithm[15], and folded monomer and folded pentamer using FoldX[16] (phospho-mutants were used; values are tabulated in Supplementary Figure-14), we have estimated the $\Delta\Delta G^{\neq}$ for the unfolding of pentameric Npm-N,

as well as for the association of monomeric Npm-N to form pentamers. The results are listed in Supplementary Figure-14 (highlighted in blue).

We next define an empirical normalized "order" parameter for Npm-N, O, as a product of the degree of secondary structure and the degree of oligomeric structure.

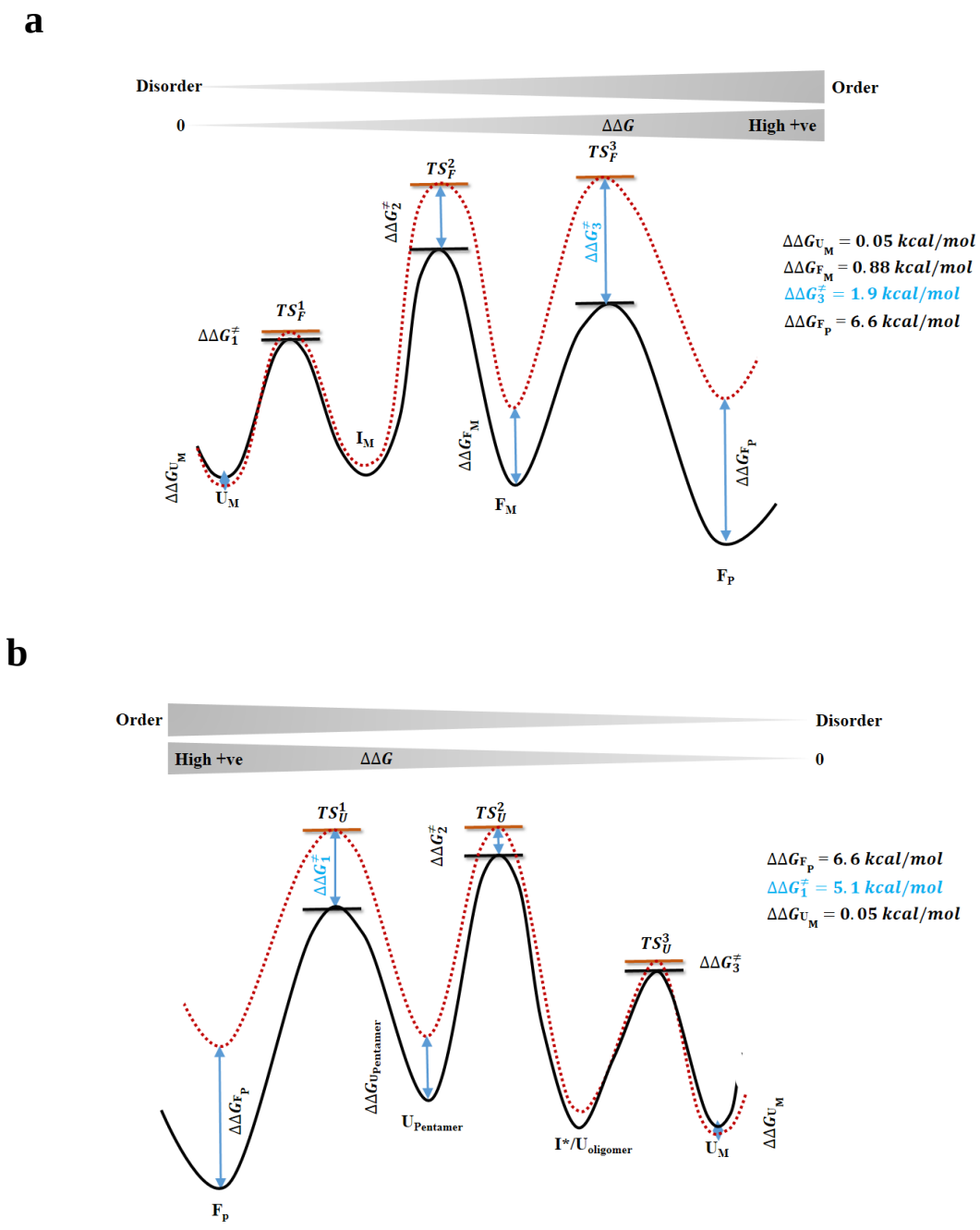
$$O = f_1\,(secondary\ structure).\,f_2\,(oligomeric\ state)$$

The upper and lower bounds of O are following: O is 1 for the ordered pentameric state (F_P), and 0 for disordered monomeric Npm-N (U_M). Conformational states of Npm-N have higher thermodynamic destabilization as a result of phospho-modifications as the value of O approaches 1 and vice versa. This is based on our energy calculation which reveals that the phospho-mutants affect the stability of the ordered pentameric Npm-N (F_P state) most, the disordered, monomeric structure is least affected, while the monomeric folded Npm-N (F_M) falls in the middle (Supplementary Figure-14a,b).

Therefore, in the folding-coupled-oligomerization pathway, O increases from left to right, implying that the TS_F^2 is more ordered than the I_M state, and TS_F^3 has higher value of O than the F_M state, as suggested by the estimated value of $\Delta\Delta G^{\neq}{}_3$ (Supplementary Figure-14a). Hence, phospho-mutations alter (increase) the relative free energy barrier of the assembly process more substantially than for folding, meaning that the destabilization of the pentameric structure by these mutations (as previously estimated computationally[17]) is reflected more strongly by destabilizing the transition-state (TS) of the association process relative to the folding process, i.e., $(\Delta\Delta G^{\neq}{}_3 - \Delta\Delta G(F_M)) \gg (\Delta\Delta G^{\neq}{}_2 - \Delta\Delta G(U_M))$ (Supplementary Figure-14a). Now, since the rates for different mutants correlate well with the thermodynamic stability of their folded states[1], we can reasonably argue that the TS (TS_F^2) that connects the I_M and F_M states in the folding pathway is structurally more similar to the F_M state rather than the I_M state, and the TS (TS_F^3) that connects the F_M and F_P states is structurally more similar to the F_P state rather than the F_M state. Together, this explains the observed differential effects of phospho-mutations on the folding *vs.* assembly processes in the forward pathway.

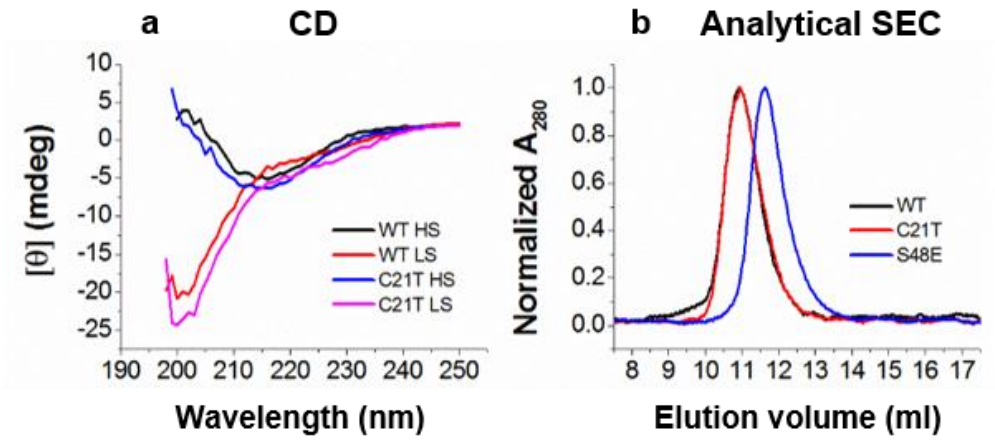
Conversely, for the reverse pathway, O decreases from left to right (Supplementary Fig. 14b). Estimation of our $\Delta\Delta G^{\neq}{}_1$ for TS_U^1 in this pathway supports this argument as it appears to be lower than $\Delta\Delta G$ value for the F_P state. In addition, phospho-modifications lowers the activation energy barrier of unfolding relatively more than for the disassembly process, i.e., $((\Delta\Delta G(F_p)) - \Delta\Delta G^{\neq}{}_1) \gg ((\Delta\Delta G(U_p)) - \Delta\Delta G^{\neq}{}_2)$. In other words, the relative degree of destabilization between the initial state and the TS decreases as the disorder in Npm-N system increases (Supplementary Figure-14b). We note that the effects of phospho-mutations in this pathway are different than what is expected based on the analysis presented for the forward pathway (i.e., it is not a mirror image of the forward pathway). This is consistent with the fact that the reverse pathway is distinct compared

to the forward pathway, suggesting that the observed effects of phospho-mutations are pathway specific.



Supplementary Figure-14: Schematic free energy diagram of the order-disorder transition pathways of Npm-N. (a) Coupled folding and assembly pathway, and (b) unfolding and disassembly pathway of Npm-N showing different states and the relative energy barriers for the WT (black line) and the phospho-mutant T95D (red dotted line).

In conclusion, using the experimentally observed states of Npm-N, the measured time-constants of individual processes for the phospho-mutants and the WT Npm-N, and the estimated relative energy changes ($\Delta\Delta G$) of each state as well as the transition states, we have presented a basis of the differential effect of phospho-mutations on the folding-coupled-assembly, and the unfolding-disassembly pathways of Npm-N.



Supplementary Figure-15: C21T mutation does not alter the secondary structure and pentamer formation of Npm-N. (a) Overlay of the far UV CD spectra of WT and C21T showing similarity in their folding at low and high salt (7.5 mM and 150 mM, respectively). (b) Analytical size exclusion (using Superose 12 column) chromatogram showing that the pentameric WT and C21T elute at the same volume at high salt, while S48E mutant that is monomeric at high salt, elutes at a higher volume.

Supplementary Tables

Supplementary Table-1: Folding and assembly time constants of Npm-N at 10 μM protein concentration at different salt concentrations*.

[NaCl] mM	τ_F (s)	τ_A^{obs} (s)
50	1874±45	2566±169
150	1167±17	1711±31
300	1125±48	1605±33

*Increasing [NaCl] from 50 mM to 300 mM resulted in a substantial decrease of both folding (τ_F) and assembly time-constants (τ_A^{obs}), indicating that Na^+ ions reduce the energy barrier of both the folding and assembly processes. This observed effect of [Na^+] is likely to be manifested by relieving electrostatic repulsion between the evolutionarily conserved clustered acidic tracts A1 (residues D34-E39) and A2 (residues E120-E130)[1, 4] in Npm-N, thereby stabilizing the transition states relative to the disordered state.

Supplementary Table-2: Folding and assembly time constants, and unfolding and disassembly time constants for the Npm-N and phospho-mutants, from CD and bulk fluorescence measurements*. These measurements were performed at 10 μM protein concentration.

Protein	τ_F (s)	τ_A^{obs} (s)	τ_U (s)	τ^1_D (s)	τ^2_D(s)	τ^3_D(s)
WT	1167±7	1711±11	691±13	463±16	2198±25	11167±67
T95D	1865±12	7952±189	62±4	296±12	1658±37	6469±90
S125E	1211±7	1948±22	391±22	450±9	1923±31	7783±79
T95D/S125E	2028±21	9274±145	61±4	296±15	1875±34	6017±59

*Our unfolding/disassembly kinetics data suggest that during the structural transition from the ordered pentamer to the disordered monomer, Npm-N undergoes a relatively fast transition to a disordered oligomeric state, followed by a multi-step disassembly process into monomers, and navigates slowly through intermediate disordered oligomeric states along this pathway. Strong repulsive interactions between acidic A1 and A2 tracts in individual monomers of Npm-N are likely to drive the unfolding of the polypeptide chain of Npm-N at low salt[1, 4]. In contrast, ~ 63% of the atoms in the monomer-monomer interface in pentameric Npm-N are non-polar, and hydrophobic and hydrogen bonding interactions substantially contribute to the assembly of individual subunits[4]. Therefore, the disassembly process under low salt conditions is likely to be primarily driven by unfolding of individual subunits rather than direct electrostatic repulsive forces at the subunit interfaces (although electrostatic repulsion between unfolded monomers can indirectly influence this process), providing an explanation for the lower rate of dissociation relative to the rate of unfolding.

Supplementary Table-3: Unfolding and disassembly time constants for the Npm-N from CD and bulk fluorescence measurements at different protein concentrations (nd* = not determined).

[Npm-N] μM	τ_U (s)	τ^1_D (s)	τ^2_D(s)	τ^3_D(s)
0.5	nd*	373±36	1951±40	10576±114
5	636±7	569±20	2298±49	11192±163
7.5	687±6	568±33	2411±37	11765±89
10	691±13	463±16	2198±25	11167±67
22	803±20	nd	nd	nd

Supplementary Note-3: The U_M to I_M shift in Npm-N: effect of salt concentrations on the dye properties and E_{FRET}

In our salt-induced folding experiments, we initiated a jump in [NaCl] from 7.5 mM to 150 mM. A valid question is whether this alteration can cause a potential change in the dye photophysics/environment that may result in a spurious effect in the smFRET histogram by changing relative fluorescence properties of donor and acceptor dyes. Since two different labeling constructs (1^*/104 and 15/104) and two different conditions (LS to HS, and LS-no Arf to LS with Arf) both show very similar U to I shifts in E_{FRET}, it is unlikely that the shift was due to a labeling position effect or due specifically to a salt-induced effect on dye photophysics.

Supplementary Table-4: Effect of [NaCl] jump from LS (7.5 mM) to HS (150 mM) on the dye properties (relative fluorescence intensity and anisotropy) at different positions of Npm-N primary structure.

*Intensity ratio was not altered substantially between donor (D; Alexa488) and acceptor (A; Alexa594).

Dye position	Intensity ratio* (LS/HS)	Anisotropy (LS)	Anisotropy (HS)
104	1.36	0.092 (D) 0.134 (A)	0.094 (D) 0.139 (A)
15	1.33	0.1265 (D) 0.1284 (A)	0.14 (D) 0.135 (A)
1	1.21	0.067(A)	0.077(A)

Moreover, to confirm this experimentally, we carried out control measurements to check for changes in fluorescence intensity and anisotropy, which could contribute to possible changes in E_{FRET}. In our control experiments, we measured the fluorescence anisotropy of the donor and acceptor dyes (Alexa488 and Alexa594, respectively) covalently attached to Npm-N at 1*, 15 and 104 (each construct was measured separately), and observed relatively low values and no significant change as a result of a change in the salt concentration from 7.5 to 150 mM (Supplementary Table-4). These data indicate unperturbed environment for both these fluorophores at the three positions tested. In addition, we compared their individual intensities, which revealed a mild decrease in the fluorescence intensity for all three positions investigated in this paper (1*, 15 and 104) for both Alexa488 and Alexa594 fluorophores on going from LS to

HS. This effect was observed to be insensitive to the identity of individual dyes. We also measured the relative absorbances of each these fluorophores attached to Npm-N at low and high salts, which revealed no significant change in response to the increased salt concentration. Since increasing salt concentration did not change the intensities of the free dyes, we believe that the decrease in the fluorescence intensity may be due to a more general intra-molecular quenching effect involving aromatic side chains, as previously observed[18].

We next estimated relative changes in E_{FRET}, based on the difference between the relative intensity decrease at positions 104 and 15, and positions 104 and 1*. From the data shown in Supplementary Table-4 (column-b), it is ≤ 0.03, and is within the experimental error of our reported data. Therefore, we can conclude that the observed shift in the smFRET histogram due to a jump in the salt concentration from 7.5 to 150 mM (U_M to I_M transition; Figure 1d, Supplementary Figure-6b) is not a mere effect of changes in the dye environment, but represents a collapse of the Npm-N backbone.

Supplementary References

[1] D. M. Mitrea, C. R. Grace, M. Buljan, M. K. Yun, N. J. Pytel, J. Satumba, A. Nourse, C. G. Park, M. Madan Babu, S. W. White, R. W. Kriwacki, *Proc Natl Acad Sci U S A* **2014**, *111*, 4466-4471.

[2] B. Bothner, W. S. Lewis, E. L. DiGiammarino, J. D. Weber, S. J. Bothner, R. W. Kriwacki, *J Mol Biol* **2001**, *314*, 263-277.

[3] A. C. Ferreon, Y. Gambin, E. A. Lemke, A. A. Deniz, *Proc Natl Acad Sci U S A* **2009**, *106*, 5645-5650.

[4] H. H. Lee, H. S. Kim, J. Y. Kang, B. I. Lee, J. Y. Ha, H. J. Yoon, S. O. Lim, G. Jung, S. W. Suh, *Proteins* **2007**, *69*, 672-678.

[5] F. Malatesta, *Biophysical chemistry* **2005**, *116*, 251-256.

[6] A. C. Ferreon, J. C. Ferreon, P. E. Wright, A. A. Deniz, *Nature* **2013**, *498*, 390-394.

[7] A. Micsonai, F. Wien, L. Kernya, Y. H. Lee, Y. Goto, M. Refregiers, J. Kardos, *P Natl Acad Sci USA* **2015**, *112*, E3095-E3103.

[8] S. Muller-Spath, A. Soranno, V. Hirschfeld, H. Hofmann, S. Ruegger, L. Reymond, D. Nettels, B. Schuler, *Proc Natl Acad Sci U S A* **2010**, *107*, 14609-14614.

[9] H. C. Gasteiger E., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A., *The Proteomics Protocols Handbook* (Ed.: J. M. Walker), Humana Press, **2005**, pp. 571-607.

[10] a) A. C. Ferreon, C. R. Moran, Y. Gambin, A. A. Deniz, *Methods Enzymol* **2010**, *472*, 179-204; b) R. Roy, S. Hohng, T. Ha, *Nat Methods* **2008**, *5*, 507-516; c) S. Weiss, *Science* **1999**, *283*, 1676-1683.

[11] E. M. Brustad, E. A. Lemke, P. G. Schultz, A. A. Deniz, *J Am Chem Soc* **2008**, *130*, 17664-17665.

[12] J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi, M. B. Francis, *Angew Chem Int Ed Engl* **2006**, *45*, 5307-5311.

[13] L. S. Witus, T. Moore, B. W. Thuronyi, A. P. Esser-Kahn, R. A. Scheck, A. T. Iavarone, M. B. Francis, *J Am Chem Soc* **2010**, *132*, 16812-16817.

[14] L. S. Witus, M. Francis, *Curr Protoc Chem Biol* **2010**, *2*, 125-134.

[15] a) J. Gu, V. J. Hilser, *Structure* **2008**, *16*, 1627-1637; b) J. Gu, V. J. Hilser, *Molecular biology and evolution* **2009**, *26*, 2217-2227.

[16] a) J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, *Nucleic Acids Res* **2005**, *33*, W382-388; b) R. Guerois, J. E. Nielsen, L. Serrano, *J Mol Biol* **2002**, *320*, 369-387.

[17] D. M. Mitrea, R. W. Kriwacki, *Pacific Symposium on Biocomputing* **2012**, 152-163.

[18] S. Mukhopadhyay, R. Krishnan, E. A. Lemke, S. Lindquist, A. A. Deniz, *Proc Natl Acad Sci U S A* **2007**, *104*, 2649-2654.