Transient Non-local Interactions Dominate the Dynamics of Measles Virus N_{TAIL}

Lillian Otteson[1,2&], Gabor Nagy[3&], John Kunkel[1,2], Gerdenis Kodis[1,2], Wenwei Zheng[4], Christophe Bignon[5], Sonia Longhi[5], Helmut Grubmüller[3*], Andrea C. Vaiana[3,6*], and Sara M. Vaiana[1,2*]

[1]Center for Biological Physics, Arizona State University, Tempe, AZ, USA,

[2]Department of Physics, Arizona State University, Tempe, AZ 85287, USA,

[3]Theoretical and Computational Biophysics, Max Planck Institute for Multidisciplinary Sciences, Göttingen, Germany,

[4]College of Integrative Sciences and Arts, Arizona State University, Mesa, AZ 85212, USA

[5]Aix Marseille Univ, CNRS, AFMB, UMR 7257, Marseille, France

[6]Present address: Nature's Toolbox, Inc. (NTx), Rio Rancho, NM 87144, USA

[&]These authors have contributed equally

*To whom correspondence may be addressed

Abstract.

The RNA genome of measles virus is encapsidated by the nucleoprotein within a helical nucleocapsid that serves as template for both transcription and replication. The intrinsically disordered domain of the nucleoprotein (N_{TAIL}), partly protruding outward from the nucleocapsid, is essential for binding the polymerase complex responsible for viral transcription and replication. As for many IDPs, binding of N_{TAIL} occurs through a short molecular recognition element (MoRE) that folds upon binding, with the majority of N_{TAIL} remaining disordered. Though N_{TAIL} regions far from the MoRE influence the binding affinity, interactions between them and the MoRE have not been investigated in depth. Using an integrated approach, relying on photo-induced electron transfer (PET) experiments between tryptophan and cysteine pairs placed at different positions in the protein under varying salt and pH conditions, combined with simulations and analytical models, we identified transient interactions between two disordered regions distant in sequence, which dominate N_{TAIL} dynamics, and regulate the conformational preferences of both the MoRE and the entire N_{TAIL} domain. Co-evolutionary analysis corroborates our findings, and suggests an important functional role for the same intramolecular interactions. We propose mechanisms by which these non-local interactions may regulate binding to the phosphoprotein, polymerase recruitment, and ultimately viral transcription and replication. Our findings may be extended to other IDPs, where non-local intra-protein interactions affect the conformational preferences of intermolecular binding sites.

Introduction

The measles virus (MeV) is a non-segmented, single-stranded, negative sense RNA virus, in which the viral genome is wrapped within a helical nucleocapsid, made of thousands of repeats

of the nucleoprotein (N) [1]. This structure is shared among all Paramyxo- and Pneumoviruses, which include common human viruses (e.g., MeV, mumps, respiratory syncytial virus, metapneumovirus, and other para-influenza viruses) and animal viruses (e.g., Sendai virus, Newcastle disease virus and rinderpest virus), as well as zoonotic biosafety level 4 agents of great concern to public health (e.g., Hendra and Nipah viruses, respectively HeV and NiV)[2]. The N protein not only provides the interaction with the RNA and holds the nucleocapsid in its helical structure, but it is the substrate used for viral transcription and replication. Viral transcription and replication in Paramyxoviruses are ensured by a complex formed by the large-polymerase (L) and the phosphoprotein (P), where the latter enables tethering L to the nucleoprotein-RNA (N-RNA) template for transcription and replication[2,3].

In MeV, N consists of a folded 400 amino-acid (a.a.) domain (N_{CORE}) which binds the RNA and forms the structured helical nucleocapsid, and of an intrinsically disordered 125 a.a. C-terminal domain (N_{TAIL}), which partly protrudes radially outward from the helical nucleocapsid structure[4–7]. The intrinsically disordered N_{TAIL} domain binds to the folded X domain (P_{XD}) of the phosphoprotein (P) and is essential for MeV transcription and replication. P in turn,binds as a tetramer to both L and N_{CORE} [6,8–11]. A similar structure and binding of N_{TAIL} to P_{XD} are found in NiV and HeV[3,12]. Binding of N_{TAIL} to P_{XD} is thought to assist in positioning the L appropriately on the nucleocapsid to read the RNA, while allowing efficient transcription re-initiation at intergenic regions[13] and hence progression from one reading frame to the next. It may also assist in maintaining contact between L and the nucleocapsid template, preventing it from leaving before transcribing all the viral genome[11,14]. In addition, the presence of N_{TAIL} loosens the nucleocapsid helical structure, possibly facilitating transcription and replication by L[11].

Because of its crucial role in transcription and replication, binding of MeV N_{TAIL} to P_{XD} has been extensively studied[15–21]. It occurs through coupled folding and binding of a short molecular recognition region (MoRE) of 18 a.a. (grey in Fig. 1A,B)[21]. This folds into an α-helix, forming a four-helix bundle with three helices from P_{XD} (PDB: 1t6o[8,9]). However, the majority of N_{TAIL} (i.e., the remaining 109 a.a. on the two sides of the MoRE) remains disordered in the bound complex[21]. The N_{TAIL}-P_{XD} complex is structurally similar in MeV, NiV and HeV[3,12]. While the regions outside the MoRE (from here on referred to as "flanking regions") do not fold, or engage in stable binding interactions[16,17], they have been clearly shown to contribute to binding affinities, with the binding affinity increasing in truncated variants of N_{TAIL} containing the intact MoRE[22,23] . This effect could be due to a pure excluded volume (entropic) repulsion of the flanking regions, as well as to potential interactions involving the flanking regions of N_{TAIL} (either at the intra- or inter-protein level)[22,24,25].

For N_{TAIL} free in solution, the secondary structure propensity of the MoRE region has been characterized by NMR on full length N_{TAIL}[7] , and by an all-atom simulation study of the MoRE fragment[19,26]. In contrast, we know very little about potential interactions between residues that are distant in sequence (non-local interactions), despite their particular importance for N_{TAIL} binding to P_{XD}[24,27] and, more generally, for IDP function. In particular, we expect interactions involving the regions of N_{TAIL} flanking the MoRE on either side, to confer specificity to the IDP by dictating large-scale reconfiguration dynamics. In fact, large-scale reconfiguration dynamics can regulate IDP binding mechanisms, rates, and affinities. Examples of how protein conformational dynamics can affect binding mechanisms and rates can be found even in simple models of proteins undergoing two-state conformational transitions[28]. In addition, non-local interactions can potentially act as allosteric regulators in IDPs by stabilizing certain

conformational states, therefore shifting the IDP conformational ensemble. This population shift can lead to allosteric regulation, as described in the ensemble allosteric model for IDPs[29,30].

Probing non-local intra-chain interactions is therefore essential for a molecular understanding of IDP dynamics, binding, and function in general[22,31–33]. It is however particularly difficult to directly probe such interactions experimentally, due to their transient nature, and the fact that IDPs undergo large conformational changes on sub-microsecond time scales. Atomistic simulations of IDPs, on the other hand, are equally challenging due to the wide conformational space sampled by IDPs. To overcome these experimental and theoretical challenges, we leveraged the high sensitivity of IDP dynamics to non-local interactions. We experimentally quantified the dynamics of full-length N_{TAIL} in solution, by probing intra-molecular contact formation times, using photo-induced electron transfer (PET) between a tryptophan (W) and a cysteine (C) placed in the sequence[34–37]. Using this technique, we measured rates along the well-defined C-W distance reaction coordinate, probing large-scale backbone motions in the hundreds of nanoseconds to microsecond internal reorganization time scale[36], which is most relevant to IDP function[38]. With minimal perturbation to the sequence, compared to introducing large prosthetic dyes, we obtained data with high sensitivity at short spatial distances. To obtain a structural interpretation and a causal description, we then compared the measurements to corresponding observables derived from both simulations and analytical models. Our integrated approach allowed us to identify key intra-protein interactions that extend beyond the MoRE region, and which may play an important role in N_{TAIL} dynamics and function.

Results and Discussion

Photo-induced electron transfer reveals the heteropolymeric nature of N_{TAIL} dynamics

To characterize the dynamics of N_{TAIL} in solution, we probed intra-molecular contact formation, using a technique based on photo-induced electron transfer (PET) between a tryptophan (W) and a cysteine (C)[35–37] placed in different regions of the protein (see Supporting Method S2). For this purpose, we created a series of N_{TAIL} variants, each with a unique C-W pair, substituting native tyrosine (Y) residues with W, and introducing C in place of native serine (S) residues (see Supporting Method S1). Note that wild-type N_{TAIL} does not contain native W or C residues. Fig. 1A and B illustrate the positions of our W and C labels within the protein (see Supporting Table S1). After exciting W to the triplet state, using a nanosecond UV laser pulse, we monitor the excited W triplet state population in time, via transient absorption measurements. As N_{TAIL} samples its conformational states in solution, W comes into contact with C via stochastic collisions, and transfers an excited state (triplet) electron to C, forming a W radical cation which eventually relaxes to the ground state. The C-W relaxation time, τ_{CW}, due to PET between W and C, was obtained by fitting W triplet state relaxation curves as illustrated in Supporting Fig. S1 and S2, and taking into account the natural lifetime of the triplet state in the absence of C (as explained in Supporting Method S2). All raw data and fits may be found in Supporting Fig. S3 to S16, and all fitting parameters in Supporting Table S2**.**

We first carried out PET experiments on each N_{TAIL} variant under near physiological conditions (i.e. 15mM Tris, 150mM NaCl, 1mM TCEP, pH 7.6, in the following referred to as "reference conditions"). The C-W relaxation times, τ_{CW}, due to PET upon C-W contact, measured under these neutral pH conditions, are plotted in Fig. 1C (black circles), for each variant. Each data point is the result of globally fitting multiple, repeated measurements, as explained in Supporting Method S2. To ascertain whether we could explain these relaxation times in terms

of C-W sequence separation, we first compared them to the times expected for a simple polymer model. Our estimated times for a homopolymer model are plotted for comparison in Fig. 1C (dashed line) as a function of sequence separation $|i\text{-}j|$ between the W (electron donor) and C (electron acceptor). For a simple homopolymer, the distance in space, R_{ij}, between the i-th and j-th amino acids in the sequence scales with the sequence separation, $|i\text{-}j|$, according to Flory's scaling law $R_{ij} \propto |i-j|^{\nu}$,[39] where ν is Flory's scaling exponent. Because the time for contact formation between the i-th and j-th amino acids is related to the R_{ij} distance between them (i.e. smaller distances correspond to shorter contact times), in such a homopolymer model the time for contact formation also scales with the sequence separation $\tau \propto |i-j|^{\nu_\tau}$. We estimated the scaling exponent ν_τ for C-W relaxation times, from the (Flory) scaling exponent ν for distances, using a homopolymer model (SAW-ν model[40]), as described in Supporting Method S4. A Flory distance scaling exponent ν of 0.52, estimated from small-angle X-ray scattering (SAXS) measurements of N_{TAIL}[6], corresponds to a relaxation time scaling exponent ν_τ of 1.69 (see Supporting Fig. S17). Accordingly, the dashed line in Fig 1C corresponds to the scaling equation:

$$\tau = 0.011|i-j|^{1.69}$$

where the prefactor was obtained by fitting the experimental data. This represents the relaxation times we would expect if N_{TAIL} behaved as a simple homopolymer, i.e. if the only non-local interactions were excluded volume interactions.

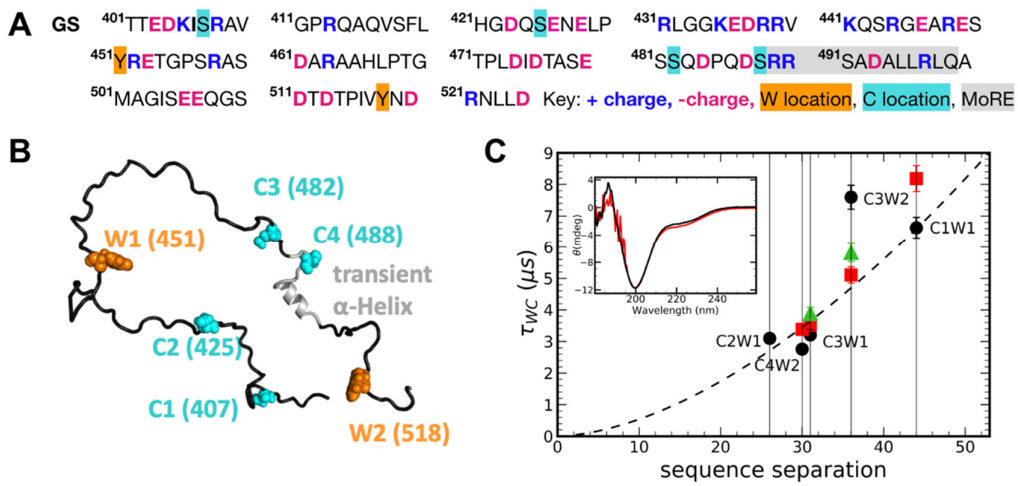


Figure 1. *Results of photo-induced electron transfer (PET) experiments at T=20 °C, on N_{TAIL} variants containing a unique C-W pair. A) N_{TAIL} sequence, showing charged residues at pH 7.6, location of Y to W, and S to C substitutions, and location of MoRE, which folds into a helix upon binding to P_{XD}. B) W and C positions. Each variant contains a unique C-W pair. C3W2 and C4W2 variants encompass the MoRE. C) C-W relaxation time τ_{CW}, due to PET occurring upon contact between W and C, measured for each variant and plotted as a function of C-W sequence separation. Experimental conditions: pH 7.6 and 150 mM NaCl (black circles); pH 4.0 and 150 mM NaCl (red squares); pH 7.6 and 500mM NaCl (green triangles). Dashed line: expected scaling of τ_{CW}, from SAW-v homopolymer model at pH 7.6 (black). Inset: Comparison of N_{TAIL} circular dichroism spectra at pH 7.6 (black) and pH 4 (red).*

While the general trend of the data at physiological pH (black circles) follows the estimated polymer scaling law (black dashed curve), most of the data points deviate from the curve, with some contacts forming faster and others slower than the expected times, depending on the region of the protein probed. This dynamic heterogeneity (i.e. deviation from homopolymer model) observed along the sequence is most striking when comparing the two variants with C-W pairs encompassing the MoRE region (C3W2 and C4W2) (Fig. 1C), which differ by just six residues in sequence separation (Figs. 1A and B). Here, the measured τ_{CW} is a factor 2.8 larger for C3W2 than for C4W2 (Fig. 1C) as opposed to the factor 1.3 predicted from polymer scaling (ν_τ=1.69). This increase is mostly due to the 7.6 µs measured relaxation time for C3W2, which is much slower than the 4.7 µs homopolymer estimate. Such an anomalously slow relaxation

time suggests the presence of interactions in N_{TAIL} that deviate from those of a simple homopolymer. Indeed, this simple model lacks both secondary structure interactions and electrostatic interactions (as well as more complex solvent-induced interactions). To test whether additional interactions could explain the observed dynamic heterogeneity, we turned to coarse-grained (CG) molecular dynamics simulations, using a more complex model, which includes both secondary structure interactions and charge contributions from amino acids placed at specific positions in the sequence. We first performed additional experiments to modulate the interactions and test some simple hypotheses.

Charged amino acids partially account for N_{TAIL} dynamic heterogeneity

We first looked at the effect of interactions between charged amino acids. Inspection of the N_{TAIL} sequence in fact shows a significant number of charged residues that are often clustered together in certain regions of the sequence (Fig. 1A). If electrostatic interactions between charged residues were significantly contributing to the observed dynamic heterogeneity, we would expect that screening such interactions, by increasing the salt concentration, would decrease or eliminate the heterogeneity, bringing the data points closer to the polymer curve. Indeed, in line with expectations, when we increase the salt concentration from 150mM (black circles) to 500mM (green triangles) the C3W2 relaxation time decreases from 7.6 μs to 5.8 μs, and the C3W1 relaxation time increases from 3.2 μs to 3.9 μs (Fig. 1C, green triangles), bringing both these data points close to the polymer curve. This suggests that electrostatic interactions are a major contributor to the observed heteropolymeric behavior of N_{TAIL}. The effect is likely due to the different distribution of charged amino acids in different regions of N_{TAIL}, and requires further investigation. However, without a more elaborate model it is hard to predict the different behavior of the two variants[41,42], namely that the C3W2 relaxation time decreases (suggesting a compaction of this portion of the chain) whereas the C3W1 relaxation

time increases (suggesting an expansion of this other portion of the chain) at increased salt concentration.

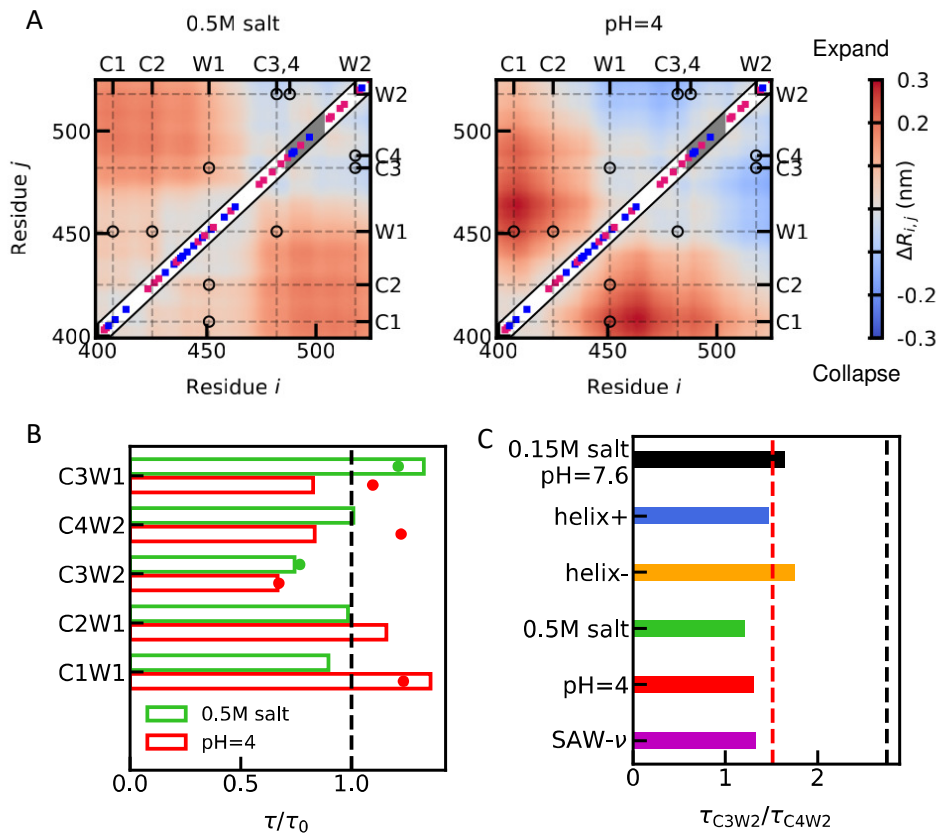


Figure 2. *Coarse-grained (CG) simulation of N_{TAIL}. A) Differences between the pairwise distances ($R_{i,j}$) of the CG simulations at either 0.5 M salt (Left) or pH 4.0 (Right), and the simulations at the reference condition (0.15 M salt and pH 7.6). The diagonal shows the charged amino acid positions with positively charged amino acids in blue and negatively charged amino acids in magenta. The MoRE is highlighted in grey. B) Ratio of C-W relaxation times due to PET, calculated from simulations (bars) and measured in experiments (dots) at different conditions (0.5 M salt: green, or pH 4.0: red), relative to the reference condition (0.15M salt and pH 7.6). C) Ratio of C-W relaxation times between the two pairs of labelling positions, C3W2 and C4W2, from simulations (bars) and experiments (dash lines, reference conditions: black, pH 4.0: red).*

We started with a simple CG model that allowed us to explicitly include electrostatic interactions from the specific charge content and patterning of N_{TAIL}, while also including secondary structure contributions, which may be significant in the MoRE region. The model is

a modified version of the recently developed HPS model[44] for IDPs, which is a residue-based CG model. We adjusted this to reproduce the experimental radius of gyration of N_{TAIL} obtained from SAXS[6] and the helical propensity of the MoRE region from NMR[7] (see details in Supporting Method S5). The salt dependence of the CG model is described by Debye-Hückel electrostatic screening[43]**.** We performed CG simulations, and generated equilibrium distance distribution plots for each pair of amino acids, including the C-W distances probed in PET experiments.

To determine the effect of increased salt concentration (150 mM to 500 mM) on the compaction/expansion of each segment within N_{TAIL}, we mapped the difference in the equilibrium distances R_{ij} between each pair of amino acids in the simulations in Fig. 2A (left). A corresponding map of scaling exponents [40,44–46] can be found in Fig. S18. In general, the figure shows that the N-terminal part of N_{TAIL} before C3 expands (red areas), whereas the C-terminal part becomes more compact (blue areas). This explains the opposite behavior observed experimentally for C3W1 compared to C3W2 when increasing the salt concentration. Specifically, expansion of the W1-C3 segment under high-salt conditions explains the increase in relaxation time observed for the C3W1 variant, while compaction of the C3-W2 segment explains the decrease in relaxation time observed for C3W2. The reason for the different behavior of these regions can be found in the particular charge distribution of N_{TAIL} (diagonal of Fig. 2A left). The positively and negatively charged amino acids are well-blended in the N-terminal region, with a close to zero net charge. The N-terminal half therefore behaves like a polyampholyte, expanding when electrostatic interactions between charged amino acids are screened by salt concentrations. On the other hand, the C-terminal region under physiological conditions has more negatively charged amino acids than positively charged amino acids, leading to net repulsive electrostatic interactions and keeping this region expanded. Increasing

the electrostatic screening at high salt concentrations reduces the repulsion, which leads to the compaction of the C-terminus of N_{TAIL}.

We then estimated how the salt-induced compaction/expansion of N_{TAIL}, observed in CG simulations, would impact the predicted PET relaxation times, and we compared these to the experimentally observed changes in relaxation times. The estimated relaxation times from PET were computed from the C-W equilibrium distance distributions generated by the CG simulations, as described in Supporting Method S5. In Fig. 2B, we plotted the ratio between estimated C-W relaxation times at the reference condition (low salt, 150 mM) and at high salt concentrations (500 mM) for each variant (green bars). A larger than one ratio indicates an increase of C-W relaxation time, exclusively due to the expansion of that segment. The experimental ratios (green dots) in Fig. 2B show, that the expansion of W1-C3 distances in CG simulations correctly predict the increase in measured C3W1 relaxation times, whereas the compaction of the C-terminus accurately predicts the decreased C3W2 relaxation times.

To further test the effect of electrostatic interactions on the dynamic heterogeneity of N_{TAIL}, we measured C-W relaxation times due to PET, upon decreasing the pH from 7.6 to 4.0, while maintaining a near-physiological salt concentration (150mM) (Fig. 1C, red squares). Unlike increasing the salt concentration, decreasing the pH only alters the charge on specific residues that titrate in the 7.6-4.0 pH range (*i.e.* negatively charged Asp and Glu become more neutral, and neutral His become more positively charged). If electrostatic interactions involving these specific charged residues were the main contributor to the dynamic heterogeneity observed at physiological pH, we would expect a reduction in heterogeneity when decreasing the pH to 4.0. Specifically, we expect the N-terminal part, which acts as a balanced polyampholyte at pH 7.6, to become positively charged and to expand at pH 4.0. Conversely, we expect the negatively charged C-terminal part to become more neutral, allowing for more compact conformations.

The measurements at pH 4.0 (red squares, Fig. 1B) are in line with our expectations, with almost all the new data points falling closer to the homopolymer curve. This includes a significant reduction in the heterogeneity for the C-W pairs encompassing the MoRE, with the ratio of relaxation times for C3W2 and C4W2 reducing from 2.8 to 1.5, which is closer to the homopolymer prediction of 1.3 (this is mostly due to the decrease in anomalously large relaxation time of C3W2 from 7.6 μs to 5.1 μs, which is closer to the 4.7 μs homopolymer prediction). Since the intrinsic bimolecular quenching rate between C and W triplet state is similar at pH 7.6 and 4.0 (see Supporting Fig. S19 and Method S2), we can conclude that the observed changes in C-W relaxation times are entirely due to changes in the protein, rather than to intrinsic pH-induced changes of the electron transfer efficiency between W and C. The experimental data of Fig. 1C clearly show that dynamic heterogeneity is greatly reduced at low pH, particularly in the MoRE region, further supporting the hypothesis that this is due in large part to the interactions between charged amino acids in N_{TAIL}.

To understand the individual behavior of each C-W pair studied, we again resorted to our CG model, and tested how well compaction or expansion of different regions of N_{TAIL} predicted by the model explain the changes in measured C-W relaxation times at different pH values. The effects of pH were represented in the CG model by changing the charge of protonatable amino acids (Asp, Glu, and His) as detailed in Supporting Method S5**.** The expansion or compaction of N_{TAIL} segments were mapped in Fig. 2A (right). Similar, though more pronounced, to the changes at high salt concentrations, the complex charge patterning gives rise to a different behavior in different regions of the chain, with the N terminus mainly expanding and the C terminus mainly compacting.

Akin to CG simulations at high salt concentrations, the C-W relaxation times due to PET were estimated from low pH (4.0) and high pH (7.6) CG simulations. Then, the ratios of estimated

relaxation times (red bars) were compared to those of measured relaxation times from PET experiments (red dots) in Fig. 2B. Looking at individual variants, the CG model correctly captures the trend of C1W1 (expansion) and C3W2 (compaction), but not that of C4W2 (close to C3W2) and of C3W1. The CG model suggests that variants C3W1 and C4W2 collapse upon reducing the pH, which would lead to a decrease in relaxation times, but PET experiments indicate a small increase in both relaxation times. Possible explanations for such discrepancy include: 1) N_{TAIL} is rich in Asp and Glu, amino acids that reach protonation equilibrium close to pH 4.0[47], therefore a fixed charge CG model does not capture the charge distribution accurately at this pH; 2) specific interaction modes such as salt-bridges are missing in the CG model; and 3) our estimates of relaxation times from CG simulations are based solely on equilibrium distance distributions, and do not include additional contributions from chain dynamics. However, the fact that such a simple CG model can capture most features when varying salt concentration and pH, suggests that the intramolecular dynamics of N_{TAIL} is strongly dependent on its charged amino acids and how they are distributed along the sequence.

In addition, we investigated how well the CG model can capture the experimentally observed pronounced difference in relaxation times between the variants C3W2 and C4W2, encompassing the MoRE of N_{TAIL}. We quantified this by using the ratio of C-W relaxation times, τ_{C3W2}/τ_{C4W2}. In Fig. 2C, we compare this ratio for experimental C-W relaxation times (dashed lines), with the ratio of estimated relaxation times computed from various CG models, as well as from our simple homopolymer model. First, we note that (as indicated by the dashed lines) the strong dynamic heterogeneity seen in PET measurements under the reference condition leads to a relaxation time ratio of 2.8 (black line, 150 mM NaCl at pH 7.6) between C3W2 and C4W2. This ratio is diminished to 1.5 upon lowering the pH (red line, pH 4.0). We expect that increasing the salt concentration would have a similar (but perhaps smaller)

reducing effect on the dynamic heterogeneity, and relaxation time ratios. Further, the 1.5 ratio measured at low pH is very similar to the ratio of 1.3 predicted by the simple homopolymer model, which suggests that electrostatic interactions indeed play a key role in the dynamic heterogeneity of C3W2 and C4W2. Secondly, we note that the CG simulations also predict a decrease in the dynamic heterogeneity for both increased salt concentrations and lowered pH, in agreement with the experimental findings. However, the CG model (ratio of 1.6) does not capture the magnitude of the dynamic heterogeneity measured in PET experiments (ratio 2.8) under reference conditions. This suggests that additional interactions and/or dynamic effects, not represented in our CG model, must be contributing to the heteropolymeric dynamics of N_{TAIL}.

One of the possible additional interactions that can affect the C3W2 and C4W2 relaxation times is the presence of a folded helix in the MoRE region. Upon changing pH and salt, CG models suggest a negligible difference of helical propensity for the entire chain with a maximum variation of less than 2% (Supporting Fig. S20**).** However, our CG model provides a convenient way to test if the helical propensity of the MoRE plays a role in C-W contact formation times, by adjusting residue-specific dihedral terms imposed only on the MoRE residues. This approach allowed us to increase or decrease the MoRE helical propensity between 35% (helix-) and 60% (helix+). As the estimated ratio of C3W2 and C4W2 relaxation times shown in Fig. 2C indicate, even considerable changes in helix propensity have only a small effect (±0.1) on the dynamic heterogeneity of N_{TAIL}, which is not sufficient to explain the large, measured heterogeneity around the MoRE.

To validate the predictions of CG simulations with respect to the MoRE helical propensities, we carried out control measurements to test the changes of N_{TAIL} helix propensity at each of the PET measurement conditions. In particular, we measured far-UV circular dichroism (CD)

spectra of N_{TAIL} solutions at pH 7.6 and 4.0. Note that due to the high UV absorbance of chloride ions, we were unable to measure high-quality CD spectra at high salt concentrations. However, as seen in the inset of Fig. 1C, the CD spectra at pH 7.6 (black) and pH 4.0 (red) are almost identical, implying that there is no significant change in helical propensity as the pH is lowered. We therefore conclude that the strong dynamic heterogeneity at the reference conditions is likely not due to the presence of the MoRE helix. This is in line with the CG model predictions. We also note that, based on the secondary structure estimates of the CD spectra by SESCA (see Supporting Fig. S21 and Method S13), the helical propensity of the full length N_{TAIL} is fairly low (below 8%), which would imply a helical propensity of at most 58% for the MoRE, which is consistent with the upper limit of 60% we used in CG simulations based on NMR chemical shifts.

Overall, our results show that N_{TAIL} dynamics is heteropolymeric in nature, and does not follow simple homopolymer scaling, especially in the region encompassing the MoRE. In this region, C-W relaxation times due to PET (which reflect intra-molecular contact formation times), but not helical propensity, are dramatically affected by solution conditions that affect the electrostatic interactions between charged residues (i.e., salt and pH). This finding suggests that electrostatic interactions are the main cause of the heteropolymeric behavior of N_{TAIL}. Although we cannot exclude the possibility that formation of transient helices might affect some specific contacts, both our CD experiments and CG simulations suggest that this is not the main factor. Consistent with PET experiments, our CG models, which account for charge patterning in N_{TAIL}, indeed predict a reduction in dynamic heterogeneity across the MoRE, when increasing the salt concentration or reducing the pH. Moreover, they predict a complex behavior of compaction and expansion within specific regions of N_{TAIL}. However, two aspects observed in the experiment cannot be captured by the current CG model. First, the ratio between relaxation

times of C3W2 and C4W2 variants (where the C and the W encompass the MoRE) is much larger in the experiments than in the CG simulations under reference conditions. Second, when reducing the pH, the experimentally observed slowdown of the two labelling positions C3W1 and C4W2, is not captured by the CG simulations. It is possible that transient interactions between two regions with large sequence separation (one in the N terminal half of N_{TAIL}, close to W1, the other overlapping with the MoRE, close to C3/C4) affect N_{TAIL} dynamics in such a way that neither a simple polymer scaling model (Fig. 2B), nor a CG model including explicit interactions between charged amino acids, can fully explain. Since neither PET experiments alone, nor polymer or CG models, can provide an accurate description of these transient interactions, we turned to all-atom explicit-solvent molecular dynamics simulations.

Disordered N_{TAIL} is characterized by a conformational ensemble with distinct states

All-atom explicit-solvent MD simulations capture specific interactions between amino acids such as salt-bridges, and hydrogen bonds[48], which we expect to be crucial to accurately model the transient structures and conformational dynamics of N_{TAIL}. Additionally, these simulations include chain dynamics (including full side chain dynamics) required to more accurately compute the contact formation rates and relaxation times due to PET. However, IDP simulations are particularly sensitive to variations in the used force field[49–51], and force field accuracy for IDPs often varies in a system-dependent manner[52]. We therefore chose to first validate several force fields and water models against existing independent experimental data on the wild-type (WT) N_{TAIL} sequence (shown in Supporting Table S3). N_{TAIL} WT simulations performed using various force fields were evaluated for their accuracy in compactness using SAXS data, and for their secondary structure propensities by comparing them to CD and NMR chemical shift measurements. SAXS curves, CD spectra, and chemical shift profiles were calculated from the simulation trajectories using the software CRYSOL[53], SESCA[54], and

Sparta+[55], respectively. A detailed discussion of this force field validation is provided in the Supporting Method S8.

The force fields that were in line with the WT experimental data were tested against CD and PET measurements of the C3W2 and C4W2 variants as well (see Supporting Table S4). In summary, the CHARMM36m (C36M)[56] force field with optimal point charge (OPC) water model[57] was the only parameter set that reproduced both variant and WT measurements, within the experimental and computational uncertainty. Therefore, we chose C36M-OPC trajectories to interpret the microscopic details of the N_{TAIL} dynamics. Since the relaxation curves predicted from these simulations reproduce the PET dynamic heterogeneity (Supporting Fig. S22), we assume the simulations contain the necessary atomic level information to explain why C3W2 has a slower relaxation rate compared to C4W2.

To this aim, we computed the free energy landscape of N_{TAIL} variants as a function of the C-W distance, for positions C3W2 and C4W2 respectively. These distances, which determine the C-W relaxation times due to PET, serve as natural reaction coordinates to explain the observed N_{TAIL} dynamic heterogeneity found in Fig. 1C. For this analysis, we assumed that the variants C3W2 and C4W2 have similar dynamics, since the difference between these variants is the exchange of one Oxygen and one Sulphur atom between two nearby residues (C3 at 482 and C4 at 488, respectively). This assumption allowed us to combine the variant simulation trajectories of C36M-OPC to maximize the sampling efficiency and calculate both donor-acceptor distances for each conformation in the trajectory.

Fig. 3 shows the resulting free-energy landscape along C3-W2 (482-518) and C4-W2 (488-518) distances. Surprisingly, this landscape is characterized by four local minima (S1 to S4) corresponding to four N_{TAIL} conformational states. N_{TAIL} spends roughly 60% of the time close

to these minima. The relative population of conformational states and the height of free-energy barriers separating them are potential sources of the dynamic heterogeneity observed in PET experiments for these variants. S1 corresponds to the conformational state around the absolute minimum of the landscape and comprises approximately 24% of all conformations. S2 and S3 are less densely populated conformational states, accounting for 14% and 8% of all simulation frames, respectively. S4 is the least populated state, with approximately 2% of conformations. The landscape shows only small barriers (< 1 kT) separating S1, S2, and S3, and a low free energy barrier between S1 and S4 (< 1.5 kT). This flat and accessible free energy landscape suggests nearly free diffusion between conformational states. Such behavior is expected for a largely disordered protein domain like N_{TAIL}[38], and is in line with previous SAXS results suggesting N_{TAIL} to be a pre-molten globule with residual structure[6].

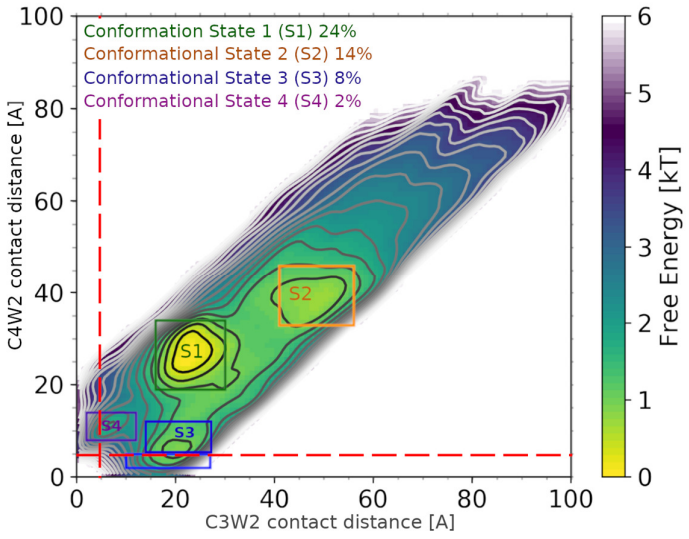


Figure 3. *Conformational states revealed by full atom MD. The free-energy landscape of sampled N_{TAIL} conformations as a function of calculated C-W distances from C36M/OPC molecular dynamics simulations. Contour lines denote a 0.3 kT difference in the estimated free energy landscape. Conformational states on the landscape are highlighted by*

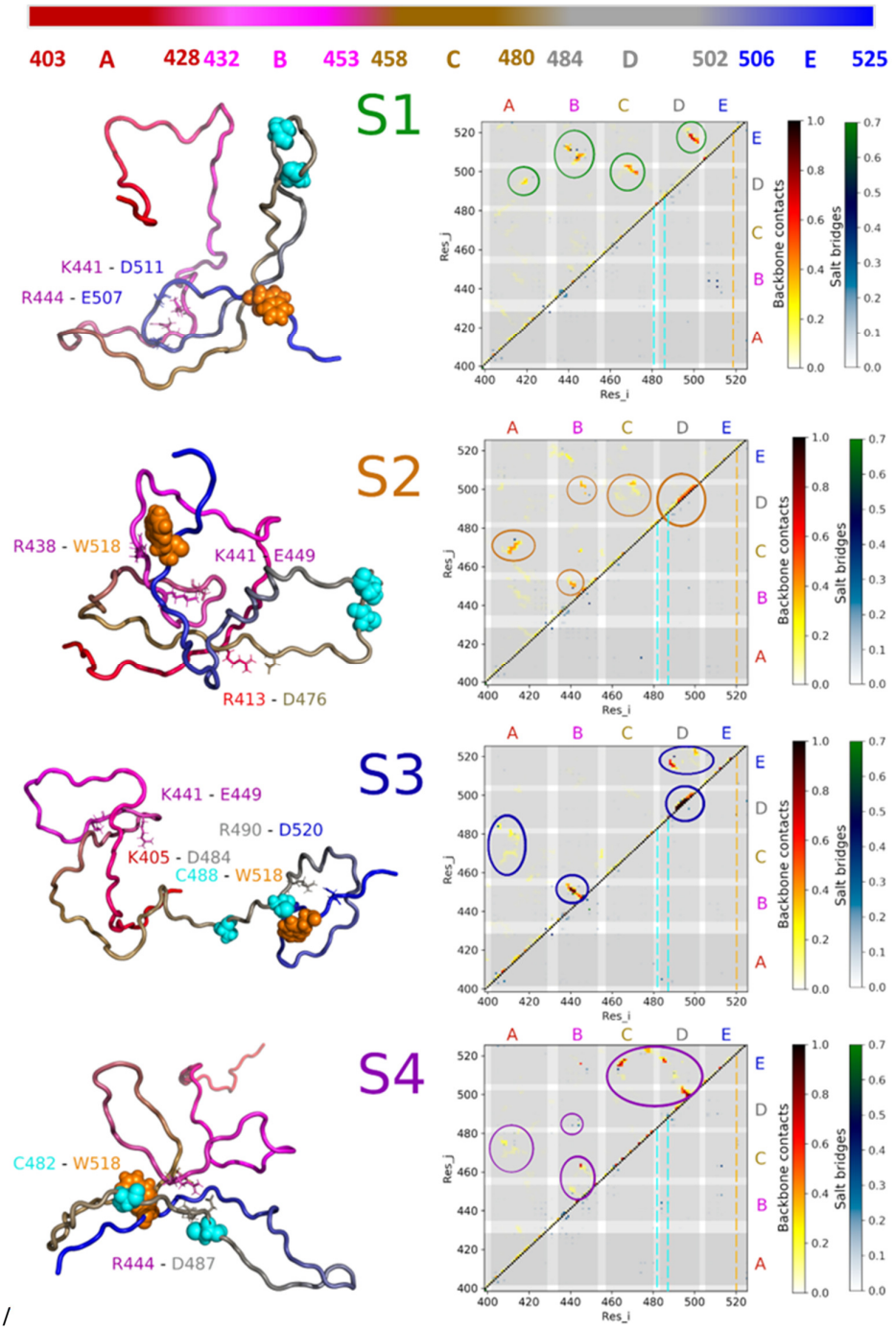
colored labels and rectangles. The population of conformations within the rectangles is shown in the legend. Red dashed lines indicate a contact distance of 5 Å. Conformations below these cutoffs would allow quenching of the W triplet state, due to electron transfer to C, for the appropriate variant (C4W2 or C3W2 respectively).

The two most populated states (S1 and S2) do not allow electron transfer to cysteines at either position C4 (488) or C3 (482), because the electron acceptor would be more than 20 Å away from the W2 (518) electron donor. In contrast, S3 would allow a cysteine at position C4 (488) to come within contact distance (< 4 Å) of W2 (518), promoting electron transfer in the C4W2 variant. State S4, on the other hand, would also allow electron transfer to a cysteine at position C3 (482) in the C3W2 variant. The fact that S3 is more easily accessible from S1, and more populated in our simulations than S4, suggests that electron transfer to cysteine at position C4 (488) is more likely. This is consistent with the observed shorter C-W relaxation times due to PET in the C4W2 variant compared to that of C3W2.

Non-local N_{TAIL} interactions explain dynamic heterogeneity

Next, we determined specific interactions that stabilize the four conformational states identified above. These interactions can provide atomic-level insights into what causes the observed N_{TAIL} dynamic heterogeneity. To this aim, we computed intramolecular contact probability maps of N_{TAIL} conformations within each conformational state, as shown in the right side of Fig. 4. The maps contain two different contact probabilities, backbone (C_α) contacts (in red to yellow), and side chain contacts between charged amino acids (salt bridges, in green to blue). The latter type of contact was considered, because our pH and salt concentration dependent PET measurements (Fig. 1C) and CG simulations suggested that electrostatic interactions play an important role in N_{TAIL} dynamics. To ease the interpretation of the contact data, we also defined five interaction regions along the N_{TAIL} sequence (color coded and labelled A-E) as shown in Fig. 4. These regions are also indicated by grey shaded areas in the contact maps. Regions A-C are subdivisions of the polyampholyte N-terminal part identified previously in CG simulations,

whereas regions D and E are parts of the negatively charged C-terminal part, with the MoRE located in region D.



/ /

Figure 4. *Contact maps and representative conformations in donor-acceptor conformational states. Representative conformations (left) and contact probability maps (right) are shown for each conformational state identified in Fig. 3 (S1-S4). The maps show C_α contact probabilities (in yellow to red) and salt bridge side chain contact probabilities (blue to green), in the plot region above the diagonal. In the plot region below the diagonal, only salt bridge probabilities are shown. The position of the cysteine acceptors C3 and C4 (residues 482 and 488) and tryptophan donor W2 (residue 518) are shown as cyan and orange dashed lines in the plot region below the diagonal respectively. The identified interaction regions are indicated by letters A-E, and their boundaries and color-coding for representative conformations are shown on the top of the figure. The same regions are indicated by the shaded grey areas of the contacts maps. The most prominent interactions in each state are also encircled. On the left side, the backbone of representative conformations is shown in cartoon representation, colored according to regions. The donor and acceptor atoms are highlighted as orange and cyan spheres, respectively. Residues with prominent interactions in the conformational states are labelled and shown in stick representation.*

The panels of Fig. 4 clearly show that the four states are stabilized by different sets of interactions, although some of the interactions occur in more than one conformational state. The contact maps also provide information on the secondary and tertiary structure preferences of N_{TAIL} in these conformational states. Stable non-local interactions are shown as clusters of contacts in darker colors. Helical secondary structure elements and tight turns appear on the maps as continuous sets of interactions that run parallel and close (within 5 residues) to the diagonal. β-strands, β-hairpins or longer loops appear as sets of interactions that run further away and are either parallel or perpendicular to the diagonal. Disordered regions typically have few or sporadic interactions and appear as gray horizontal and vertical bands in the contact maps.

The four conformational states show markedly different structures, including the MoRE, which is located in region D of the contact maps. In state S1, region D is largely unstructured, but forms several transient local salt bridges between charged residues. In addition, the S1 contact map shows several non-local interactions (encircled in green) that may prevent donor-acceptor contacts both for the C3W2 and C4W2 variants. S1 conformations are further stabilized by a set of salt bridges formed between regions B and E. Salt bridges involving K441, R444 of

region B and E507 and D511 are particularly frequent and probably play a significant role in stabilizing this conformational state.

S2 includes conformations where region D is disordered, as well as conformations with a folded α-helix between residues 492-502. The helical structures appear on the S2 contact map as an interaction patch near the diagonal of region D (encircled in orange). Non-local interactions in S2 appear to be less stable than in S1 and include contacts between regions A and C, C and D, as well as B and D. The more stable interactions in S2 also involve two stable salt bridges between R413 and D476 (A and C) and K441 and E449 (local interaction in region B). This S2 contact map also shows a large number of sporadic interactions (both backbone and side chain) within the N-terminus (between regions A and B). Taken together, interactions in S2 suggest the existence of multiple conformational substates and most of the interactions can be described as the central regions (B and C) forming contacts with both the N_{TAIL} termini (A, D, and E).

Conformations of S3 comprise fewer non-local interactions, and are instead characterized by two very stable local interaction sites. The local interactions show a shorter helix between residues 492-497 present in all S3 conformations, as well as a stable hairpin or zipper in region B, stabilized by a salt bridge (K441-E449) observed previously in S2. Further, there are two sets of notable non-local interactions. First, a patch of interactions between regions D and E folds the C-terminus into a loop, stabilized by the salt bridge between R490 and D520. These non-local contacts also bring the electron donor W2 (518) very close to a potential acceptor at position C4 (488), and thus promote electron transfer in the C4W2 N_{TAIL} variant. At the same time, the zipper in region B brings the N-terminal close to regions C and D, where another salt bridge between K405 and D484 may prevent electron transfer to position C3 (482) in the C3W2 variant.

Finally, the S4 state, which allows electron transfer to position C3 (482) in the C3W2 variant, is structurally very different from the other three conformational states. The most defining structural feature of this state is a long loop reminiscent of an anti-parallel β-sheet that stretches over the entire C-terminus (regions D and E). This loop is supported by interactions between regions C and E including a salt bridge between R463 and D513. The N-terminus of N_{TAIL} in this conformational state has only weaker interactions, mostly limited to contacts between regions A and C as well as between regions B and C.

The contact analysis revealed several interactions that appear to stabilize the observed N_{TAIL} conformational states, and therefore may promote or hinder the C-W electron transfer for either the C3W2 or C4W2 variants. To quantify correlations between C-W contacts required for PET relaxation and other intra-molecular interactions within N_{TAIL}, we computed a normalized pointwise mutual information (NPMI) between contact distances as presented in Supporting Method S21. Here we provide a brief summary of the NPMI analysis described in more detail in Supporting Method S23. Non-local interactions prominent in the S1 conformational state, including all contacts listed in Supporting Table S5**,** are negatively correlated with the formation of C-W contacts and likely inhibit PET relaxation in both C3W2 and C4W2 variants. In contrast, prominent contacts in the S3 and S4 states are positively correlated (and likely promote) C-W contact formation for the C4W2 and C3W2 variants respectively. The analysis also highlights several charged residues in regions B, D, and E that form prominent stabilizing interactions in multiple conformational states.

Taken together, the charged residues in region B highlighted in Supporting Method S23 apparently can form either non-local salt bridges with residues close to the MoRE region, or local salt bridges with other charged residues within region B. We observe that many of these salt bridges are accompanied by clusters of interactions (see encircled clusters in Fig. 4). It is

possible that the formation of these clusters is nucleated by the salt bridges. In any case, switching between these sets of interactions is necessary for the system to transition between the most populated conformational states S1 and S2 in Fig. 3, and competition between these interactions is what determines the populations of the conformational states themselves.

Functional relevance of non-local interactions is supported by coevolutionary analysis

Having identified the most important interactions that dictate N_{TAIL} structure and dynamics, we attempted to ascertain whether these might also play a functional role. To this aim, we performed a coevolutionary analysis on N_{TAIL} using EVcouplings [58] (see Supporting Method S22 for details). Coevolutionary analysis is a sequence-based method which quantifies the correlated mutations of homologous residue pairs within proteins from related species, and thus highlights pairwise-interactions of potential functional importance. A total of 12,072 sequences similar to N_{TAIL} have been aligned and selected for this analysis.

As shown in Supporting Fig. S23, most non-local interactions with positive EVcouplings scores connect region B (residues 440-450) and region D (485-505), suggesting evolutionarily conserved interactions between these two regions. We would like to specifically highlight four non-local interactions with the highest EVcoupling scores, on the top left of Fig. S23 and summarized at the bottom of Table S5**.** These include three hydrophobic contacts involving leucine residues, which are part of the binding interface with P_{XD}, and one salt bridge between R444 and D493. These non-local interactions suggest a biologically relevant, functional connection between distant regions B and D. Given that the function of region D (MoRE) is to bind P_{XD}, it is possible that the function of B-D interactions is to regulate binding. We note that the contact analysis in the all-atom MD simulations does not identify the coevolving contacts as prominent interactions, but it does corroborate the putative allosteric connection between regions B and D. Specifically, all-atom MD identifies several key amino acids in these regions

whose interactions act as conformational switches between the most populated states, governing large scale N_{TAIL} dynamics in near-physiological conditions, as illustrated schematically in Fig. 5 and further discussed below.

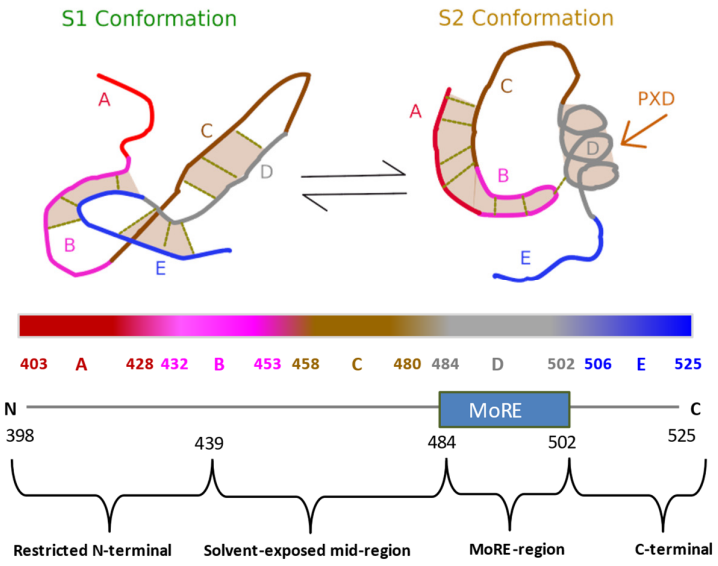


Figure 5. *Schematic representation of the two major NTAIL conformational states (S1 and S2). Different regions of N_{TAIL} are color-coded according to Figure 4. Non-local interactions between the regions are represented as dashed lines. Key interactions between regions are highlighted by orange shading. The helical binding site of the phosphoprotein X-domain (P_{XD}) in region D is marked by an orange arrow.*

Conclusions

To understand the role of disordered regions flanking the MoRE on either side, we probed non-local interactions of full-length N_{TAIL} in solution, using photo-induced electron transfer experiments between tryptophan and cysteine residue-pairs introduced in the N_{TAIL} sequence. With our PET experiments we measured and compared C-W contact formation times between different regions of the protein, and under varying solution conditions. Our experiments

revealed a pronounced dynamic heterogeneity within the protein, in contrast with what would be expected for a homopolymer. Based on our molecular modeling efforts and molecular dynamics simulations, we attribute this heterogeneity to the presence of non-local interactions between specific regions of N_{TAIL}. We observed a particularly strong dynamic heterogeneity close to the MoRE binding region (Fig. 1) under physiological conditions. This anomalous behavior was reduced or eliminated in PET experiments under conditions that weaken electrostatic interactions, such as increased salt concentration, or lowered pH.

Interpretation of the general role of electrostatics

Consistent with experiments, our simulation results (both coarse-grained and all-atom) point to an important role of electrostatic interactions between charged residues of N_{TAIL}. Our CG simulations qualitatively capture the important role of electrostatic interactions. In fact, even though the model does not include atomic-level details, it suggests that specific regions of the protein behave differently (compact or expand) upon changing the CG model charges to emulate increased ionic screening or pH related protonation. In addition, CG simulations performed with different helical propensities in the MoRE combined with circular dichroism measurements indicate that changes in helicity of the MoRE do not give rise to the large differences observed between contact formation times in this region (i.e. C3W2 *versus* C4W2).

Specific non-local interactions dictate N_{TAIL} conformation and dynamics

Our all-atom MD simulations, performed with a carefully validated force-field, include interactions with atomic level detail. These simulations capture the observed experimental PET behavior, specifically the large difference in W triplet-state relaxation times between C-W pairs encompassing the MoRE (C3W2 *versus* C4W2). We found that while N_{TAIL} is clearly disordered, a few key non-local interactions stabilize four distinct conformational states. The two most populated states do not allow PET for C3W2 or C4W2, while the third and fourth

states allow electron transfer from W2 to C4 and C3, respectively. Analysis of the free energy landscape (Fig. 3) between the four states reveals nearly free diffusion between the first three states. However, structural rearrangements required to reach the C3W2 contact state are limited by a small (~3.7 kJ/mol) free-energy barrier under physiological conditions. This free-energy barrier and the lower equilibrium population of the C3W2 contact state contributes to the dynamic heterogeneity of the N_{TAIL} C-terminal region.

Analysis of the state-specific contacts (Fig. 4) reveals different, competing clusters of interactions that stabilize the four conformational states. These interaction clusters often involve (and are possibly nucleated by) prominent salt bridges between charged amino acid side chains. Mutual information analysis confirms a direct correlation between the presence of prominent salt bridges and C-W contact formation in MD trajectories, suggesting that electrostatic interactions indeed play an important role in N_{TAIL} dynamics.

Further, focusing on the two most populated conformational states S1 and S2, we found that specific charged residues in region B (residues 435-451) can form either non-local salt bridges with C-terminal regions D-E (484-525, close to the MoRE), or local salt bridges with other charged residues within region B. Switching between these sets of interactions (as sketched in Fig. 5) is necessary for the system to transition between the two dominant states, affecting large-scale N_{TAIL} dynamics. Competition between these interactions also determines the populations of the conformational states themselves.

Co-evolutionary analysis independently identifies regions B and D as the only regions that co-evolve. Given the distance in sequence between these two regions, this not only corroborates the presence of non-local interactions between region B and D, but also suggests that they play a functionally important role.

In summary, three independent methods highlighted region B of N_{TAIL} as a potential allosteric interaction partner to regions D and E, which in turn include the binding site (MoRE) for the phosphoprotein X-domain (P_{XD}). Additionally, based on experimental evidence obtained on truncated N_{TAIL} variants[23], it appears that intramolecular interactions of N_{TAIL} involving the residues 435-451 (region B) have a significant impact on the P_{XD} binding in vitro, despite these residues being located far away from the identified molecular recognition element.

Identified interactions in the context of N_{TAIL} to P_{XD} binding

We hypothesize that the intramolecular N_{TAIL} interactions observed in our simulations may not only govern the N_{TAIL} dynamics, but also play a role in regulating the MoRE-P_{XD} binding process. Experimental studies by Bignon *et al*[59] have shown that the helix content of the MoRE is positively correlated with the binding affinity, and that N-terminal truncation of N_{TAIL} increases its affinity for P_{XD}[22,24,27]. Similarly, Gruet et al[27] have identified several regions in the N_{TAIL} sequence, on either side of the MoRE (enhancer boxes) that increase P_{XD} binding affinity upon mutation. The crystallographic structure of the N_{TAIL} MoRE bound to P_{XD} (PDB: 1t6o, Fig. 6 blue inset) shows that the binding interface consists mainly of a hydrophobic edge along the MoRE helix in region D, between S491 and M501 (hereafter D1-site, shown in dark green), which interacts via intermolecular contacts with the α2-α3 face of the P_{XD} triple helix. The other side of the MoRE helix comprises a hydrophilic edge formed by charged and polar residues (D2-site, dark red) between Q486 and R497. The N_{TAIL}-P_{XD} complex is further stabilized by intermolecular interactions involving R497 and salt bridges formed by D487 and R490 which are located at the beginning of the MoRE helix between the two edges (D3-site, light green)[27]. Many of these key amino acids are involved in intramolecular N_{TAIL} interactions in one or more conformational states, as shown by the contact probability data (Fig. 4) obtained from our all-atom MD simulations. Based on these interactions, we propose two independent

mechanisms that provide a possible explanation for the observed changes in the binding affinity of the MoRE to P_{XD} upon N_{TAIL} mutation or truncation.

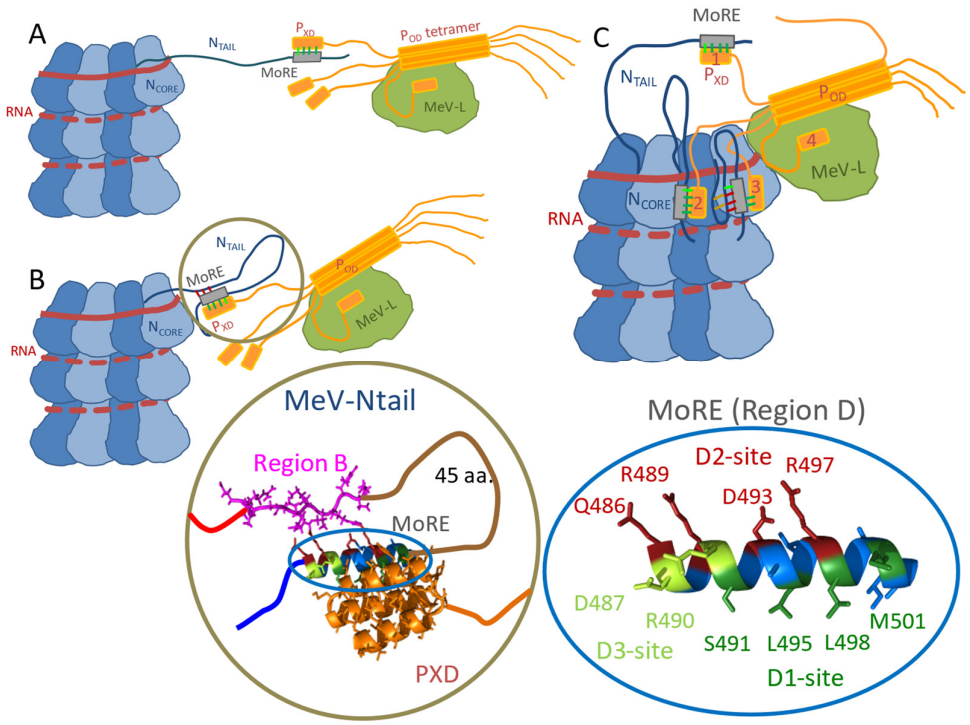


Figure. 6*. Speculated biological context of non-local N_{TAIL} interactions. (**A**) During the initial steps of the virus replication the MeV Nucleocapsid protein (N, blue) recruits the large polymerase (L, green), by binding to the phosphoprotein (P, orange). Here N_{TAIL} is extended. (**B**) When Brownian motion allows L to approach the RNA (red) in the nucleocapsid groove, N_{TAIL} adopts a hairpin conformation, which promotes intramolecular interactions between the N_{TAIL} termini. C) Once contact between the nucleocapsid and L is established, additional complexes can be formed between free P_{XD} and downstream N monomers. Further, intramolecular N_{TAIL} interactions help release bound P to allow polymerase progression, by weakening the binding to P_{XD} when L is nearby. Numbers (1-4) on P_{XD} denote which step individual domains are in a "cartwheeling" L progression mechanism. In panels A to C, only 1-3 N_{TAIL} domains have been depicted for the sake of clarity. The RNA molecule is shown as a solid line on its first helix turn around the nucleocapsid to indicate that it is more solvent exposed, and as a dashed line on downstream turns to highlight that it is partially buried in the RNA-binding ridge.* ***Yellow inset:*** *Cartoon representation of the P_{XD}-MoRE complex based on the available crystal structure (PDB: 1t6o).* ***Blue inset:*** *N_{TAIL} MoRE residues in region D on the hydrophobic P_{XD}-binding interface are highlighted in dark green, charged residues that may be involved in intramolecular N_{TAIL} interactions are shown in dark red. Residues that may be involved in both intra- and intermolecular interactions are shown in light green.*

Direct competition for MoRE: In our all-atom MD simulations, the D1-site residues (Fig. 6 inset, dark green) form intramolecular interactions with amino acids of regions B and C in the

S2 state, and with residues in region E in the S3 state (see Supporting Table S7), both of which are states with considerable helical propensity of the MoRE region. In addition, the co-evolution analysis indicates an evolutionary link between the MoRE D1-site and residues in region B (see Supporting Table S5). Therefore, assuming that a folded MoRE helix is a necessary part of the biologically active N_{TAIL} conformation, our results suggest that these intramolecular interactions would directly compete with the intermolecular interactions of P_{XD} to bind the MoRE D1-site. This competition would affect both the thermodynamics and the kinetics of the MoRE-P_{XD} binding process. First, intramolecular interactions would reduce the free helix population of the MoRE, thereby decreasing the binding rate of P_{XD}. Second, intramolecular interactions may also displace P_{XD} from the binding site without exposing the hydrophobic edge of the MoRE to water, lowering the P_{XD} unbinding activation barrier and increasing the unbinding rate. Both of these effects would regulate MoRE-P_{XD} binding by reducing the binding affinity, and explain why the binding affinity increases when flanking regions of the MoRE are removed[27,59].

Indirect allosteric regulation: Helical conformations of the N_{TAIL} MoRE appear to have higher affinity for P_{XD}[59]. Therefore, shifting the population of helical MoRE conformations via intramolecular interactions would also regulate MoRE-P_{XD} binding affinity. This mechanism is similar to the ensemble allosteric model for IDPs described by White *et al*[29]. Specifically, intramolecular interactions that stabilize non-helical MoRE conformations (such as the contacts seen in state S1) would decrease the population of helical conformations, reducing both N_{TAIL} binding affinity and the rate of binding to P_{XD}. On the basis of the contacts observed in Fig. 4, it is likely that the S1–S2 equilibrium in N_{TAIL} is controlled by the competing interactions of region B, which acts as a conformational switch. As illustrated in Fig. 5, the intramolecular interactions observed in S1 between regions B and E stabilize a loop-like, non-helical

conformation of the N_{TAIL} C-terminus, which is likely binding-incompetent towards P_{XD}. However, in S2, region B residues engage in local B-B interactions, allowing the N_{TAIL} C-terminus to form the high-affinity helical structure in the MoRE at region D.

In both mechanisms proposed above, intramolecular interactions between region B and the MoRE in region D would regulate P_{XD} binding, either by direct competition for the P_{XD} binding site, or by indirect competition between N_{TAIL} conformational states.

Potential functional role of intra-protein N_{TAIL} interactions in viral replication

Non-local interactions between regions B and D/E of N_{TAIL} may also play a functional role in RNA polymerization and viral replication. Indeed, our two mechanisms proposed above both reduce the P_{XD}-N_{TAIL} binding affinity either by altering the helix population (allosteric regulation) or competing for hydrophobic P_{XD} binding site on the folded MoRE helix (direct competition). The literature suggests that unbinding is one of the rate-limiting steps of the replication process[10]. Therefore, a reduction in binding affinity may be beneficial for viral replication, especially if it is coupled with an increase in the rate of P_{XD} unbinding (e.g. direct competition). In addition to the two regulatory mechanisms for P_{XD}-MoRE binding mentioned above, we also speculate on two additional mechanisms where intramolecular interactions between the B and D regions of N_{TAIL} may play an important role in optimizing the kinetics of viral replication.

Tether shortening: Increasing the rate at which the viral polymerase complex is recruited to the nucleocapsid is another way in which intramolecular N_{TAIL} interactions may increase the rate of viral replication. Initial polymerase complex recruitment likely involves the formation of one or more tethers between the polymerase complex and the nucleocapsid. By limiting the volume available for diffusion, tethering increases the local concentration of the polymerase complex

around the nucleocapsid. It is probable that one or more of the 145 amino acid tethers formed by an N_{TAIL} and a phosphoprotein C-terminal domain (P_{LOOP}) as shown in Fig. 6A also play a role in the initiation process. After such a tether is assembled by forming the N_{TAIL} MoRE-P_{XD} complex, intramolecular N_{TAIL} interactions between the MoRE D2-site and region B (as seen in the coevolution analysis) may shorten the tether as depicted in Fig. 6B by 45 amino acids. This would further reduce the volume available to the polymerase complex, and increase the rate of initiation of viral replication. We note that a study by Jensen et al[7] suggested that region B may be partially restricted in the RNA binding ridge of the nucleocapsid. However, the formation of the ridge requires the presence of upstream N monomers that are missing at the beginning of the nucleocapsid, giving region B of the initial N monomers greater potential range and mobility.

The work of Guseva et al. suggested another possible tether consisting of an approximately 105 amino acid long disordered segment of a phosphoprotein N-terminal tail (P_{TAIL}) binding with its helical linear motif (HELL) to N_{CORE}[11]. The two tethers may coexist and in fact work in concert to securely attach the polymerase complex to the nucleocapsid during the replication initiation step.

P_{XD} unbinding during polymerase progression: A study from Sourimant *et al* outlined a mechanism for polymerase progression based on a deletion mutant of the N protein that removed regions B and C of N_{TAIL}[10]. The study revealed that deletion of this region considerably slowed down polymerase progression. The proposed mechanism is based on a newly discovered binding site on N_{CORE} for the $\alpha1$-$\alpha2$ face of P_{XD}. This site likely binds the P_{XD}-MoRE complex, facilitating unbinding, and subsequent binding of P_{XD} to an N_{TAIL} MoRE of a downstream N monomer.

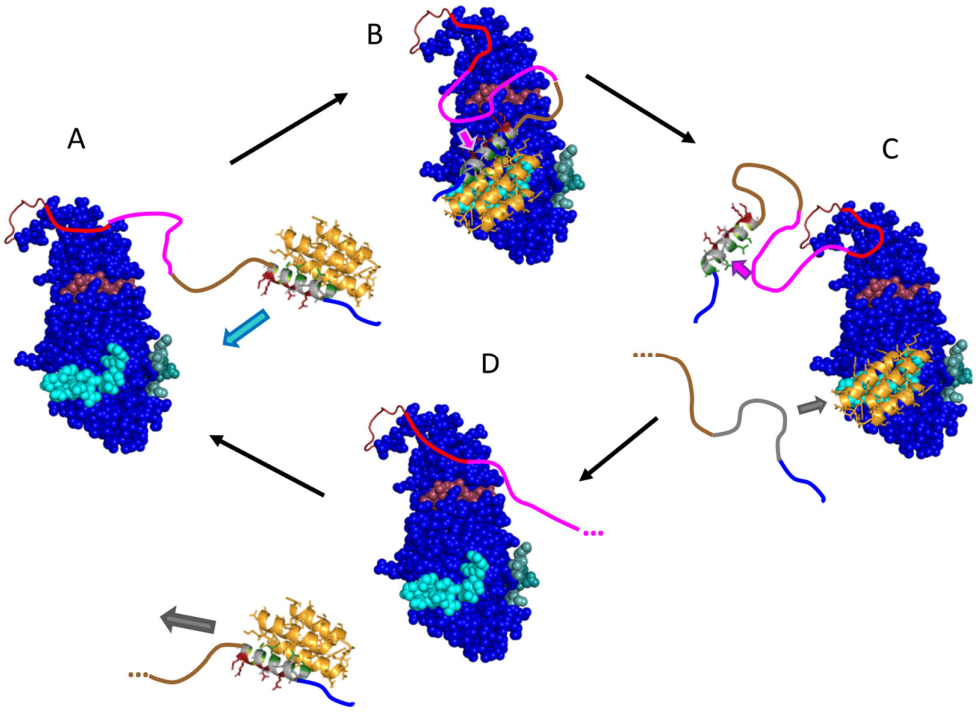


Figure 7: *Proposed mechanism of MoRE-P_{XD} unbinding during polymerase progression. A) N_{TAIL} protrudes from the RNA binding grove, binds P_{XD} and brings it to the N_{CORE} binding site (cyan). B) Region B (in magenta) interacts with MoRE, and weakens the P_{XD}-MoRE complex. C) the N_{TAIL} of the current monomer unbinds from P_{XD} and N_{TAIL} from a downstream monomer binds P_{XD}. D) P_{XD} is detached from N_{CORE} and the polymerase complex can progress further. Panels of the Figure show N_{CORE} in a blue space filling representation with the N_{CORE} binding site shown in cyan space filling. The viral RNA on the figure is shown as a red space filling representation, regions of N_{TAIL} are shown as ribbons in similar colors as shown Fig. 5, and the P_{XD} triple helix is shown in orange cartoon representation. Side chains for the N_{TAIL} MoRE and P_{XD} are shown as sticks with the colors shown in Fig 6.*

Our findings on the intramolecular interactions of N_{TAIL} may indeed corroborate the mechanism of Sourimant *et al* as summarized in Fig. 7. Briefly, N_{TAIL} binds P_{XD} to tether the polymerase to the nucleocapsid, then the P_{XD}-MoRE complex is docked to the binding site on N_{CORE}. Next, the interactions between the MoRE and region B can weaken or displace the MoRE-P_{XD} interactions to facilitate unbinding, exposing P_{XD} to the solvent. The exposed P_{XD} is then free to bind to a downstream N_{TAIL} monomer or to its binding site on the L polymerase[60].

This proposed mechanism implies that deleting regions B and C doesn't just shorten N_{TAIL}, but also eliminates B-D intramolecular interactions that promote MoRE-P_{XD} unbinding. This in turn would explain why the deletion hampers viral replication, and provides a possible reason for region B residues to co-evolve with the MoRE (Table S5). In addition, based on the available cryo-EM structure of the nucleocapsid (PDB: 4uft), region B would emerge from the RNA binding ridge close to the N_{CORE} binding site, with sufficient space for region B to fold into a loop and interact with the MoRE region, as shown in the illustrative model in Fig. S24.

Our findings may be similarly compatible with the cartwheeling mechanism[61] of polymerase progression shown in Fig. 6c. Here, the P is recruited (1) by N_{TAIL} and docked to the N_{CORE} binding site (2) of the same monomer as previously described. However, in this mechanism P_{XD} remains bound to the same N monomer until the L polymerase approaches. Then, due to steric hinderance from L or conformational changes at the N-N interface region B of N_{TAIL} adopts a loop-like confirmation (3) and weakens P_{XD}-N_{TAIL} interactions. This in turn, allows P_{XD} to be transferred to its binding site on L (4) as L progresses. Subsequently, P_{XD} can either be released to the solvent or directly bind to the N_{TAIL} of a downstream N monomer in the next cartwheeling cycle.

As discussed above, experimental and computational evidence suggests that the intramolecular N_{TAIL} interactions weaken P_{XD} binding. Tight binding to P_{XD} is beneficial for tethering the viral polymerase to the nucleocapsid during polymerase recruitment and for ensuring that the transcription/replication of the genome is completed. However, unbinding of individual P_{XD}-N_{TAIL} complexes is required for polymerase progression during viral replication[62]. Therefore, MeV and related viruses may be at an evolutionary disadvantage if the binding between P_{XD} and N_{TAIL} is so tight as to limit the rate of replication, as well illustrated by mutational studies that targeted P_{XD}[62]. It is tempting to hypothesize that a possible functional role of non-local

intra-protein N_{TAIL} interactions is to provide a conformation-dependent optimization of P_{XD} binding affinity

Summary

Our combined experimental and simulation approach shows that, while N_{TAIL} is clearly disordered, contains very little secondary structure, and shows nearly diffusive dynamics, it still displays some structural features which cause its dynamics to deviate significantly from that of a homopolymer. In particular, we identified several key transient interactions between disordered regions distant in sequence as the main reason for such behavior. These interactions also affect the overall conformation and dynamics of N_{TAIL} in solution. Interestingly, similar interactions also emerge from independent co-evolutionary analysis, corroborating our simulation results and suggesting that they not only govern the conformational dynamics of the essential N_{TAIL} domain, but also are functionally important. We therefore propose possible mechanisms by which these non-local interactions regulate binding to the phosphoprotein X domain (P_{XD}), and consequently, recruitment and progression of the polymerase complex onto the nucleocapsid template. Our results for N_{TAIL} corroborate the importance of flanking regions in IDPs that bind via short molecular recognition elements (MoREs)[63], and suggest a regulatory role of regions that may be far in the sequence from such MoREs. These non-local interactions occurring within N_{TAIL} can regulate binding *via* an ensemble allosteric model[29,30] without requiring the binding of a third molecule. Further studies targeting the N_{TAIL}-P_{XD} interaction may provide additional insight into the underlying mechanisms and effects of these interactions.

It is plausible that our findings and our proposed mechanisms can be extended to other IDPs that share sequence characteristics with N_{TAIL}. Specifically, non-local intra-protein interactions in IDPs may regulate the conformational preferences of both the MoRE and of the entire protein, either via direct competition or indirectly by shifting the population of free-energy minima.

Methods

Protein expression, purification, and sample preparation.

Wild type (WT) N_{TAIL} and the seven variants listed in Supporting Table S1, were expressed in *E. coli*, and further purified as described in detail in Supporting Method S1. Briefly, after expression of tagged protein in *E. coli*, cell pellets were resuspended in 8 M urea, 50 mM Tris pH8, 0.3 M NaCl and frozen. After thawing, the solution was sonicated, spun, and tagged proteins were purified using Ni-Sepharose fast flow beads (Cytiva). After dialyzing the eluent in 50 mM Tris pH8, 0.3 M NaCl, TEV protease was used to remove tags from the desired protein. TEV protease as well as uncut tagged proteins were eventually removed using Ni-sepharose beads. Purified N_{TAIL} proteins, free of histidine tags, were dialyzed against 15 mM Tris, 150 mM NaCl, 1 mM TCEP buffer at neutral pH, checked using SDS-PAGE, UV absorption, and far-UV circular dichroism (CD), and frozen for shipping. After shipping, samples were further purified via HPLC (semipreparative Vydac C18 column) and lyophilized. Before PET experiments, the lyophilized protein was dissolved directly into filtered buffers (for reference conditions: 15 mM Tris, 150 mM NaCl, 1 mM TCEP pH 7.6; for low pH measurements: 20 mM NaAc, 150 mM NaCl, 1 mM TCEP pH 4.0; for high salt concentrations: 15 mM Tris, 500 mM NaCl, 1mM TCEP pH 7.6). Protein concentration was adjusted to ~100 μM as determined by UV-absorbance at 280 nm (extinction coefficient 6990 cm^{-1} M^{-1}). Samples (~350 μL) were placed in 5x10x30(h) mm Spectrocell quartz gas tight cuvettes with screwcap (equipped with Teflon coated silicon membrane), and bubbled with USP grade nitrous oxide for at least one hour, to reduce the concentration of dissolved oxygen (a quencher of excited tryptophan triplet state), and to introduce solvated electron scavenger (to prevent reactions with electrons, produced by water decomposition under UV laser pulses).

Photo-induced electron transfer (PET) experiments.

To probe intra-molecular contact formation, we used a technique based on photo-induced electron transfer (PET) between a tryptophan (W) and a cysteine (C) [34–36] placed at different positions within the sequence (sequences in Supporting Table S1). Details are described in Supporting Method S2. Briefly, we used a homebuilt nanosecond transient absorption apparatus to excite the W to the triplet state, and to monitor the excited state population as a function of time. When, C comes into contact with W (within Van der Waals distance) due to stochastic collisions, an excited-state electron is transferred from the triplet state of W to C. As explained in Supporting Method S2, and in Sizemore et al. 2015[36], the measured relaxation time is related to the intra-molecular contact formation time between W and C in the protein. The contribution to the triplet state relaxation times due to C-W quenching via electron transfer, are reported in Fig. 1C. These were obtained by fitting W triplet state relaxation curves as illustrated in Supporting Figs. S1 and S2, and taking into account the natural lifetime of the triplet state in the absence of C, measured on single W variants (see Supporting Method S2). Each value is the result of globally fitting multiple, repeated measurements, under given solution conditions. All raw data and their fits are shown in Supporting Figs. S3 through S14. The corresponding fitting parameters, obtained for each variant under each solution condition, are reported in Supporting Table S2. Solution conditions used in PET experiments of Fig. 1 were, for reference conditions: 15 mM Tris, 150 mM NaCl, 1 mM TCEP pH 7.6; for low pH measurements: 20 mM NaAc, 150 mM NaCl, 1 mM TCEP pH 4.0; for high salt concentrations: 15 mM Tris, 500 mM NaCl, 1mM TCEP pH 7.6

Circular Dichroism

Circular Dichroism (CD) spectra reported in Fig. 1C inset were measured as described in Supporting Method S3. To improve signal to noise, and access low wavelengths which are

particularly important for the interpretation of IDP spectra, at each pH we carried out measurements for a series of protein concentrations and combined them. Measurements at high salt concentration were not possible due to the high absorbance of these samples below 210nm, a range which is most important for the interpretation of IDP spectra. Solution conditions used for spectra of Fig. 1 were: 10mM $NaPO_4$, 150 mM, NaF for pH 7.6 measurements, and 20 mM NaAc, 150 mM NaF for pH 4.0 measurements.

Coarse-grained molecular dynamics

We applied a coarse-grained model based on the original HPS model[64] to N_{TAIL}. Each amino acid is represented by a bead with charge (+1, 0, -1) and hydropathy. There are three types of interactions in the HPS model: bonded interactions, electrostatic interactions, short-range pairwise interactions. The electrostatic interactions are modeled using a Coulombic term with Debye-Hückel[43] electrostatic screening to account for salt concentration. The short-range pairwise potential accounts for both protein-protein and protein-solvent interactions, which was optimized using the experimental radius of gyration of N_{TAIL}[6]. We further added additional terms for angle and dihedral preference so that the secondary structure of the MoRE region can be captured (see Supporting Method S5). The HOOMD-Blue software v2.9.3[65] together with the azplugins (https://github.com/mphowardlab/azplugins) were used for running the molecular dynamics simulations. All simulations were run using a Langevin thermostat with a friction coefficient of 0.01 ps^{-1}, a time step of 10 fs and a temperature of 298 K for 5 μs. The first 0.5 μs were dumped for equilibration.

All atom molecular dynamics

We performed all molecular dynamics (MD) simulations on three full length N_{TAIL} variants. The wild type (WT) N_{TAIL} (401-525) simulations were performed including an N-terminal hexa-histidine (His6) tag, while the latter was not included in the case of variants C3W2 (i.e. S482C-

Y518W) and C4W2 (i.e. S488C-Y518W) to match the available experimental data as much as possible. All simulations were performed using the GROMACS 2019 simulation package[66]. Each N_{TAIL} variant was simulated using two different force fields (CHARMM36m[56] and AMBER99SB-disp[50]) and at least three replicas per force field started from different initial conformations. The WT N_{TAIL} simulations were performed only with the CHARMM36m force field. For the Amber force field validation, we used previously published wild type N_{TAIL} simulation trajectories[50]. The simulations of the two variants in Amber was carried out using force field parameters adapted to the GROMACS simulation package. Both the WT trajectories and the force field parameters were kindly provided by Piana et al. Each simulation was carried out using periodic boundary conditions in a box filled with explicit solvent molecules consisting of either optimal point charge (OPC[57], for CHARMM36m) or a modified four point-charge (TIP4P[51] for AMBER99SB-disp) water models as well as Na+ and Cl- ions corresponding to an ion concentration of 150 mM. The total size of simulated systems was approximately 120,000 to 200,000 atoms. Total simulation time was more than 40 μs per variant with conformations recorded after every nanosecond. Hydrogen bond vibrations were constrained using virtual atom sites to enable a 4 fs time step during simulations. All simulations were kept at 1 atm pressure and 298 K temperature. Further details on the MD simulation parameters and system preparation are provided in the Supporting Methods S7 and S9. Analysis methods related to the comparison to experimental data, calculation of mutual information, contact maps, and free-energy landscapes are described in the Supporting Methods S11, S19-21.

Data availability.

Data supporting the findings of this study are available from the corresponding authors upon reasonable request.

Code availability.

The implementation of the coarse-grained HPS model can be downloaded at https://www.public.asu.edu/~wzheng38/tools.html.

Acknowledgements

L.O, J.K., G.K., S.L. and S.M.V. acknowledge the support from the National Institutes of Health (R01GM120537). A.C.V., H.G. and G.N. were supported by the Max Planck Society and by the German Science Foundation (DFG), Excellence Strategy Grant MBExC 2067/1; computer time was provided by the Max Planck Computing and Data Facility. G.N. was additionally supported by the Alexander von Humboldt-Foundation. A.C.V. was additionally supported by the European Union's Horizon 2020 Framework Programme for Research and Innovation, Specific Grant Agreement No. 945539 (Human Brain Project SGA3). W.Z. acknowledges the support from the National Science Foundation (MCB-2015030), the National Institutes of Health (R35GM146814) and the research computing at Arizona State University.

References.

1. Gutsche, I. *et al.* Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science (1979)* **348**, (2015).

2. Bloyet, L. M. The nucleocapsid of paramyxoviruses: Structure and function of an encapsidated template. *Viruses* **13**, 2465 (2021).

3. Longhi, S., Bloyet, L. M., Gianni, S. & Gerlier, D. How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication. *Cellular and Molecular Life Sciences* **74**, 3091 (2017).

4. Karlin, D., Ferron, F., Canard, B. & Longhi, S. Structural disorder and modular organization in Paramyxovirinae N and P. *Journal of General Virology* **84**, 3239–3252 (2003).

5. Kingston, R. L., Baase, W. A. & Gay, L. S. Characterization of Nucleocapsid Binding by the Measles Virus and Mumps Virus Phosphoproteins. *J Virol* **78**, 8630–8640 (2004).

6. Longhi, S. *et al.* The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *Journal of Biological Chemistry* **278**, 18638–18648 (2003).

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

8. Johansson, K. *et al.* Crystal Structure of the Measles Virus Phosphoprotein Domain Responsible for the Induced Folding of the C-terminal Domain of the Nucleoprotein. *Journal of Biological Chemistry* **278**, 44567–44573 (2003).

9. Kingston, R. L., Hamel, D. J., Gay, L. S., Dahlquist, F. W. & Matthews, B. W. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc Natl Acad Sci U S A* **101**, 8301–8306 (2004).

10. Sourimant, J., Thakkar, V. D., Cox, R. M. & Plemper, R. K. Viral evolution identifies a regulatory interface between paramyxovirus polymerase complex and nucleocapsid that controls replication dynamics. *Sci Adv* **6**, eaaz1590 (2020).

11. Guseva, S., Milles, S., Blackledge, M. & Ruigrok, R. W. H. The nucleoprotein and phosphoprotein of measles virus. *Front Microbiol* **10**, 1832 (2019).

12. Gondelaud, F. *et al.* Functional benefit of structural disorder for the replication of measles, Nipah and Hendra viruses. *Essays Biochem* **66**, 915 (2022).

13. Bloyet, L. M. *et al.* Modulation of Re-initiation of Measles Virus Transcription at Intergenic Regions by PXD to NTAIL Binding Strength. *PLoS Pathog* **12**, e1006058 (2016).

14. Cox, R. & Plemper, R. K. The paramyxovirus polymerase complex as a target for next-generation anti-paramyxovirus therapeutics. *Front Microbiol* **6**, 459 (2015).

15. Huang, W., Ravikumar, K. M., Chance, M. R. & Yang, S. Quantitative mapping of protein structure by hydroxyl radical footprinting-mediated structural mass spectrometry: A protection factor analysis. *Biophys J* **108**, 107–115 (2015).

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

17. 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: Structure, Function and Genetics* **73**, 973–988 (2008).

18. Kavalenka, A. *et al.* Conformational analysis of the partially disordered measles virus N TAIL-XD complex by SDSL EPR spectroscopy. *Biophys J* **98**, 1055–1064 (2010).

19. Wang, Y. *et al.* Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles èirus nucleoprotein. *Proc Natl Acad Sci U S A* **110**, E3743 (2013).

20. Dosnon, M. *et al.* Demonstration of a folding after binding mechanism in the recognition between the measles virus NTAIL and X domains. *ACS Chem Biol* **10**, 795 (2015).

21. Bourhis, J.-M. *et al.* The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Science* **14**, 1975 (2005).

22. Gruet, A. *et al.* Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. *FEBS Journal* **283**, 576–594 (2016).

23. Bignon, C., Troilo, F., Gianni, S. & Longhi, S. Modulation of measles virus NTAIL interactions through fuzziness and sequence features of disordered binding sites. *Biomolecules* **9**, 8 (2019).

24. Troilo, F., Bonetti, D., Bignon, C., Longhi, S. & Gianni, S. Understanding Intramolecular Crosstalk in an Intrinsically Disordered Protein. *ACS Chem Biol* **14**, 337 (2019).

25. Mukhopadhyay, R. & Hoh, J. H. AFM force measurements on microtubule-associated proteins: The projection domain exerts a long-range repulsive force. *FEBS Lett* **505**, 374 (2001).

26. Robustelli, P., Piana, S. & Shaw, D. E. Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. *J Am Chem Soc* **142**, 11092–11101 (2020).

27. Gruet, A. *et al.* Dissecting partner recognition by an intrinsically disordered protein using descriptive random mutagenesis. *J Mol Biol* **425**, 3495 (2013).

28. Cai, L. & Zhou, H. X. Theory and simulation on the kinetics of protein-ligand binding coupled to conformational change. *Journal of Chemical Physics* **134**, 105101 (2011).

29. White, J. T., Li, J., Grasso, E., Wrabl, J. O. & Hilser, V. J. Ensemble allosteric model: Energetic frustration within the intrinsically disordered glucocorticoid receptor. *Philosophical Transactions of the Royal Society B: Biological Sciences* **373**, 20170175 (2018).

30. Berlow, R. B., Dyson, H. J. & Wright, P. E. Hypersensitive termination of the hypoxic response by a disordered protein switch. *Nature* **543**, 447 (2017).

31. Chowdhury, A., Nettels, D. & Schuler, B. Interaction Dynamics of Intrinsically Disordered Proteins from Single-Molecule Spectroscopy. *Annu Rev Biophys* **52**, 433–462 (2023).

32. Staby, L. *et al.* Flanking Disorder of the Folded αα-Hub Domain from Radical Induced Cell Death1 Affects Transcription Factor Binding by Ensemble Redistribution: Flanking disorder tunes the RCD1-RST ensemble. *J Mol Biol* **433**, 167320 (2021).

33. Peng, Y. *et al.* A Metastable Contact and Structural Disorder in the Estrogen Receptor Transactivation Domain. *Structure* **27**, 229–240 (2019).

34. Buscaglia, M., Schuler, B., Lapidus, L. J., Eaton, W. A. & Hofrichter, J. Kinetics of Intramolecular Contact Formation in a Denatured Protein. *J Mol Biol* **332**, 9–12 (2003).

35. Vaiana, S., Best, R. B., Yau, W.-M., Eaton, W. A. & Hofrichter, J. Evidence for a partially structured state of the amylin monomer. *Biophys. J.* **97**, 2948–2957 (2009).

36. Sizemore, S. M., Cope, S. M., Roy, A., Ghirlanda, G. & Vaiana, S. M. Slow Internal Dynamics and Charge Expansion in the Disordered Protein CGRP: A Comparison with Amylin. *Biophys J* **109**, 1038 (2015).

37. Lapidus, L. J., Eaton, W. A. & Hofrichter, J. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 7220–7225 (2000).

38. Schuler, B. & Hofmann, H. Single-molecule spectroscopy of protein folding dynamics—expanding scope and timescales. *Curr. Opin. Struct. Biol.* **23**, 36–47 (2013).

39. Flory, P. J. *Principles of Polymer Chemistry*. *Principles of Polymer Chemistry* (Cornell University Press, 1953).

40. Zheng, W. *et al.* Inferring properties of disordered chains from FRET transfer efficiencies. *Journal of Chemical Physics* **148**, 123329 (2018).

41. Sawle, L. & Ghosh, K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. *J. Chem. Phys.* **143**, 85101 (2015).

42. 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).

43. Debye, P. & Hückel, E. De la theorie des electrolytes. I. abaissement du point de congelation et phenomenes associes. *Physikalische Zeitschrift* **24**, 185–206 (1923).

44. Gomes, G.-N. W. N. W. *et al.* Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET. *J Am Chem Soc* **142**, 15697–15710 (2020).

45. Hofmann, H. *et al.* Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences* **109**, 16155–16160 (2012).

46. Wiggers, F. *et al.* Diffusion of a disordered protein on its folded ligand. *Proceedings of the National Academy of Sciences* **118**, e2106690118 (2021).

47. Fossat, M. J. & Pappu, R. V. Q-Canonical monte carlo sampling for modeling the linkage between charge regulation and conformational equilibria of peptides. *Journal of Physical Chemistry B* **123**, (2019).

48. Zheng, W. *et al.* Molecular Details of Protein Condensates Probed by Microsecond Long Atomistic Simulations. *Journal of Physical Chemistry B* **124**, 11671 (2020).

49. Best, R. B., Zheng, W. & Mittal, J. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J. Chem. Theor. Comput.* **10**, 5113–5124 (2014).

50. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E4758–E4766 (2018).

51. Piana, S., Donchev, A. G., Robustelli, P. & Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *Journal of Physical Chemistry B* **119**, 5113–5123 (2015).

52. Henriques, J. & Skepö, M. Molecular dynamics simulations of intrinsically disordered proteins: On the accuracy of the TIP4P-D water model and the representativeness of protein disorder models. *J Chem Theory Comput* **12**, 3407–3415 (2016).

53. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL–a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* **28**, 768–773 (1995).

54. Nagy, G., Igaev, M., Jones, N. C., Hoffmann, S. V. & Grubmüller, H. SESCA: Predicting Circular Dichroism Spectra from Protein Molecular Structures. *J Chem Theory Comput* **15**, 5087 (2019).

55. Shen, Y. & Bax, A. SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. *J Biomol NMR* **48**, 13–22 (2010).

56. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **14**, 71–73 (2017).

57. Shabane, P. S., Izadi, S. & Onufriev, A. V. General Purpose Water Model Can Improve Atomistic Simulations of Intrinsically Disordered Proteins. *J Chem Theory Comput* **15**, 2620 (2019).

58. Hopf, T. A. *et al.* The EVcouplings Python framework for coevolutionary sequence analysis. *Bioinformatics* **35**, 1582–1584 (2019).

59. Bignon, C., Troilo, F., Gianni, S. & Longhi, S. Partner-Mediated Polymorphism of an Intrinsically Disordered Protein. *J Mol Biol* **430**, 2493–2507 (2018).

60. Pont, V. Du, Jiang, Y. & Plemper, R. K. Bipartite interface of the measles virus phosphoprotein X domain with the large polymerase protein regulates viral polymerase dynamics. *PLoS Pathog* **15**, e1007995 (2019).

61. Kolakofsky, D., Le Mercier, P., Iseni, F. & Garcin, D. Viral RNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology* **318**, 463 (2004).

62. Brunel, J. *et al.* Sequence of Events in Measles Virus Replication: Role of Phosphoprotein-Nucleocapsid Interactions. *J Virol* **88**, 10851 (2014).

63. Bugge, K. *et al.* Interactions by Disorder – A Matter of Context. *Front Mol Biosci* **7**, 110 (2020).

64. Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B. & Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* **14**, e1005941 (2018).

65. Anderson, J. A., Glaser, J. & Glotzer, S. C. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Comput Mater Sci* **173**, 109363 (2020).

66. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25 (2015).