

Molecular simulations of enzymatic phosphorylation of disordered proteins and their condensates

Emanuele Zippo^{1,2}, Dorothee Dormann^{2,3}, Thomas Speck⁴, Lukas S. Stelzl^{2,3,5}

1 Institute of Physics, Johannes Gutenberg University Mainz, Germany

2 Institute of molecular Physiology (ImP), Johannes Gutenberg University Mainz, Germany

3 Institute of Molecular Biology (IMB), Mainz, Germany

4 Institute for Theoretical Physics IV, University of Stuttgart, Germany

5 KOMET1, Institute of Physics, Johannes Gutenberg University Mainz, Germany

Abstract

Understanding the condensation and aggregation of intrinsically disordered proteins in a non-equilibrium environment is crucial for unraveling many biological processes. Active enzymes catalyse many processes by consuming chemical fuels such as ATP. Enzymes called kinases phosphorylate disordered regions of proteins and thus profoundly affect their properties and interactions. Protein phosphorylation is implicated in neurodegenerative diseases and may modulate pathogenesis. However, how protein sequence and molecular recognition of a disordered protein by kinases determine phosphorylation patterns is not understood. In principle, molecular dynamics simulations hold the promise to resolve how phosphorylation affects disordered proteins and their assemblies. In practice, chemically-detailed simulations of enzymatic reactions and the dynamics of enzymes are highly challenging, in particular it is difficult to verify whether implementations of driven simulations are thermodynamically consistent. We can now address this problem with residue-level coarse-grained molecular dynamics simulations, integrating Metropolis Monte Carlo steps to model chemical reactions. Importantly, we show how to verify by Markov-state modeling that the realisation of a non-equilibrium steady state satisfies local-detailed balance. We investigate TDP-43 phosphorylation by the kinase CK1 δ in simulations, examining patterns of phosphorylation and assessing its preventive role in chain aggregation, which may be a cytoprotective mechanism in neurodegenerative diseases. We find that the degree of residue phosphorylation is determined by sequence preference and charges, rather than by the position in the chain. The phosphorylation frequency is also affected by the phosphorylation patterns, since the interactions between CK1 δ and TDP-43 actively change after each reaction. For TDP-43, our simulations show condensates dissolution through phosphorylation with kinases binding to the condensates and phosphorylating TDP-43 in the condensates.

1 Introduction

2 Biological systems operate far from equilibrium[1]. The functionalities of cells and of their organelles
3 and compartments are possible only through a very precise self-organization, driven by a continuous
4 injection of energy from the external environment [2]. In the cell, chemical energy is stored, e.g. in
5 the form of ATP molecules among others[3]. This energy is then used to synthesize and degrade
6 molecules through biological cycles. On time scales shorter than physiological changes, microscopic
7 rates are approximately constant and the system enters a non-equilibrium steady states (NESS)[3–
8 5].

9 Cellular compartmentalisation underpinning biological function is achieved not only by lipid
10 membranes and organelles surrounded by such membranes, but also by phase separation of pro-
11 teins, giving rise to biomolecular condensates[6]. Membrane-less compartments of phase-separated
12 proteins can concentrate or exclude molecules and thus organize biochemical processes in time
13 and space, which is analogous to the compartmentalisation provided by lipid membranes. These
14 phase-separated condensates can often act as chemical reactions organizers[7]. However, these con-
15 densates of proteins can also age into solid aggregates, which are believed to contribute to neuronal
16 dysfunction and neurodegeneration[8, 9]. As condensates age and become less liquid-like, they
17 frequently lose their biochemical functionalities[10]. Aggregates of intrinsically disordered proteins
18 (IDPs) are often linked to neurodegenerative diseases. Some examples are Tau protein aggregates,
19 associated with Alzheimer’s disease[11], α -synucleic aggregates, associated with Parkinson’s disease
20 [12], or TAR DNA-binding protein 43 (TDP-43) aggregates, mostly found in patients with amy-
21 otrophic lateral sclerosis (ALS) [13], frontotemporal dementia [14], but also in many patients with
22 Alzheimer’s disease [15].

23 Proteins within condensates can also undergo chemical reactions themselves [2], driving the
24 system out of equilibrium by dissipating a biochemical fuel, such as ATP. The modification of
25 those proteins by addition of chemical groups, such as phosphate groups, are referred to as post-
26 translational modifications (PTMs). IDRs are not only essential in driving the condensation of
27 proteins, but they are also prime targets of PTMs [16]. PTMs can drastically change the properties
28 of individual proteins[17] and collectively of condensates [18], enhancing[11, 19] or suppressing
29 the condensation and aggregation of IDPs[20, 21]. For instance, it has been shown that chemical
30 reactions can stabilize the size of liquid droplets by suppressing Ostwald ripening [22, 23].

31 To connect these advances in the understanding of active processes in condensates to the biolog-
32 ical roles of proteins, it will be important to elucidate how ATP driven phosphorylation shapes the
33 interactions of intrinsically disordered protein regions (IDRs) of neurodegeneration-linked proteins
34 such as TDP-43. The disordered low-complexity domain (LCD) of TDP-43 is hyper-phosphorylated
35 in disease, and in experiments such a hyper-phosphorylation has been found to suppress TDP-43
36 condensation and aggregation[24]. Enzymes can add PTMs to IDPs in dilute solution, but en-
37 zymatic addition of PTMs may also occur in protein condensates. Recently, it was shown that
38 phase-separated condensates can speed up phosphorylation of Tau protein[25]. Phosphorylation of
39 the TDP-43 C-terminal residues Ser 379, Ser 403, Ser 404, Ser 409, and Ser 410 in patient samples
40 is associated with neurodegenerative disease[26]. TDP-43 is phosphorylated by Casein kinase 1 δ
41 (CK1 δ). How the enzymatic phosphorylation of TDP-43 is modulated in dilute solution and how it
42 is affected by protein condensates is not known. The disordered tail of CK1 δ is auto-inhibitory[27,
43 28], but how it inhibits TDP-43 phosphorylation is unclear on the molecular scale. IDRs of enzymes
44 have multiple functions, such as auto-inhibition by binding to the active site. IDRs are involved in
45 substrate binding, for instance IDRs can speed up reactions via fly casting effect, where the IDR

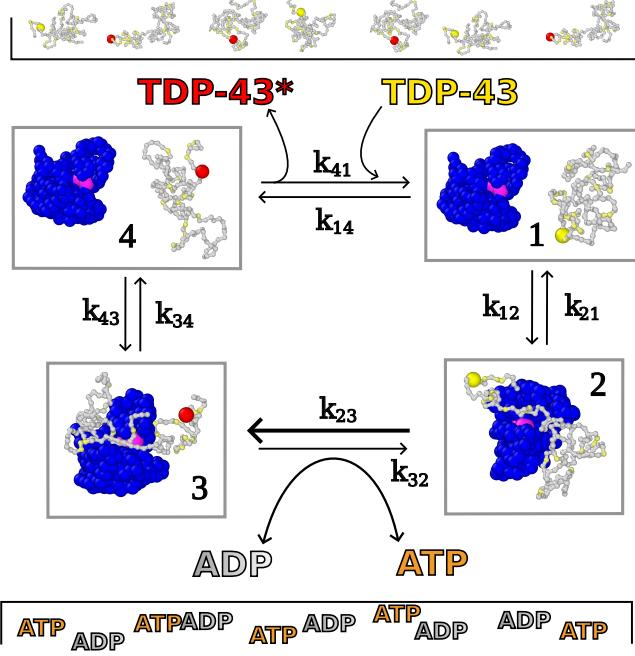


Figure 1: Enzymatic phosphorylation cycle driven by the consumption of the chemical fuel ATP. In state **1** TDP-43 (grey) is unphosphorylated and is not bound to the kinase CK1 δ (blue, active site in pink). In state **2** TDP-43 binds to CK1 δ . In state **3** the reactive serine is phosphorylated by kinase, converting one ATP into one ADP. In state **4** phosphorylated TDP-43 dissociates from CK1 δ . Phosphorylated and unphosphorylated TDP-43 are supplied through reservoirs and we consider exchanges between these reservoirs and our simulation box. Serines are colored in yellow, while phospho-serines in red.

46 increases the search volume for the binding of partner proteins[29].

47 Phase behaviour of intrinsically disordered proteins (IDPs) and the biological functionalities
 48 of protein condensates have been studied in the past years using multi-scale molecular dynamics
 49 (MD) simulations. Such simulations capture the spontaneous condensation of hundreds or more
 50 proteins while maintaining enough chemical detail in the simulations to elucidate sequence-specific
 51 interactions of proteins[24, 30]. Comparison to more highly-resolved coarse-grained methods [30–
 52 34] or atomistic molecular simulations [35, 36] can then highlight important drivers of protein
 53 condensation[24].

54 However most of these studies assume thermodynamic equilibrium, neglecting the dynamical
 55 changes in the properties of individual proteins and protein condensates, as well as the dissipation
 56 caused by chemical fluxes. Much progress has already been made in the simulations of mechanically-
 57 driven non-equilibrium steady state (NESS), where external mechanical forces give rise to driven
 58 dynamics [37]. An important step was the construction of Markov state models to better under-
 59 stand the effects of driving on the molecular scale[38]. Analogously, a biological chemically-driven
 60 NESS, such as molecular motors, can be simulated by maintaining a chemical potential difference,
 61 i.e. by fixing the ATP to ADP concentration ratio. [4, 39] Chemical reactions could in principle also
 62 be modelled via quantum mechanical approaches[40], but these are computationally very demand-

63 ing, which can preclude their application to large-scales dynamics in complex biochemical systems.
64 Recently, exciting progress has been made in integrating chemical reactions in molecular dynamics
65 simulations via neural networks[41]. Even in the case of coarse-grained simulations, chemical reac-
66 tions have been modelled through the use of reactive beads that can form bonds between molecules
67 [42]. In many cases, one could model chemical reactions in complex system by combining MD with
68 a suitably chosen Monte Carlo (MC) step [39, 43]. Arguably, the absence of a straightforward
69 approach of validating the thermodynamic consistency of simulations of NESS has held back the
70 widespread application of MD/MC approaches to biochemical reactions on the molecular scale.

71 Here we demonstrate how to validate the thermodynamic consistency of simulations of enzymatic
72 phosphorylation of proteins using TDP-43 LCD and its phosphorylation by CK1 δ as an example.
73 We do so by constructing a Markov state model (MSM), which is a generally applicable approach.
74 Our coarse-grained simulations of enzymatic phosphorylation of TDP-43, show how the sequence
75 specific interactions of CK1 δ with TDP-43 LCD affects the phosphorylation frequency of serines
76 residues in the TDP-43 LCD in dilute solution and in condensates. In particular the C-terminal
77 domain is more phosphorylated than the N-terminus, in agreement with experiments. Indeed,
78 multiple serines of TDP-43 LCD have been found phosphorylated in patient samples, in particular
79 in the C-terminal region[44, 45], with Ser 409/Ser 410 phosphorylation being established as a
80 hallmark of TDP-43 pathology in disease [26] and detected, together with Ser 403/Ser 404 and
81 Ser 379, by phospho-specific antibodies[46]. The phosphorylation frequency is also affected by the
82 phosphorylation patterns, since the interactions between CK1 δ and TDP-43 actively change after
83 each reaction, enhancing further phosphorylations[47]. Moreover we study the role of the CK1 δ IDR
84 (residues from 295 to 415) in phosphorylating TDP-43 both in condensate and dilute regime. CK1 δ
85 IDR strongly interacts with TDP-43 LCD, reducing its contacts with active site of the enzyme in
86 dilute regime. In dense regime, the CK1 δ tail anchors of the enzyme to the droplet surface.

87 Results

88 Markov-state modeling demonstrates thermodynamic consistency of sim- 89 ulations of chemically-driven dynamics

90 Molecular dynamics (MD) simulations together with a thermostat holding the temperature fixed
91 can be employed to sample from the canonical equilibrium distribution. However, introducing
92 phosphorylation reactions in MD simulations generally inject energy into the system, thus breaking
93 detailed balance and displacing the system away from thermal equilibrium. We simulate the action
94 of the kinase CK1 δ (truncated at residue 294 for the purposes of this section) on the substrate
95 protein TDP-43 by combining one-bead-per-residue implicit-solvent MD with MC phosphorylation
96 steps and validate the thermodynamic consistency of our simulations by making use of Markov state
97 models (MSMs). We assume that only the serines (Ser) of TDP-43 LCD can be phosphorylated
98 into phospho-serines (pSer). The phosphorylation reaction is the following:



99 Whenever a Ser (or pSer) is in contact with the active site of the kinase, we try to swap it with a
100 pSer (or the opposite) with acceptance probability given by

$$A(\text{Ser}, \text{pSer}) = \min(1, \exp(-\beta \Delta U_P - \beta \Delta \mu_P)) \quad (2)$$

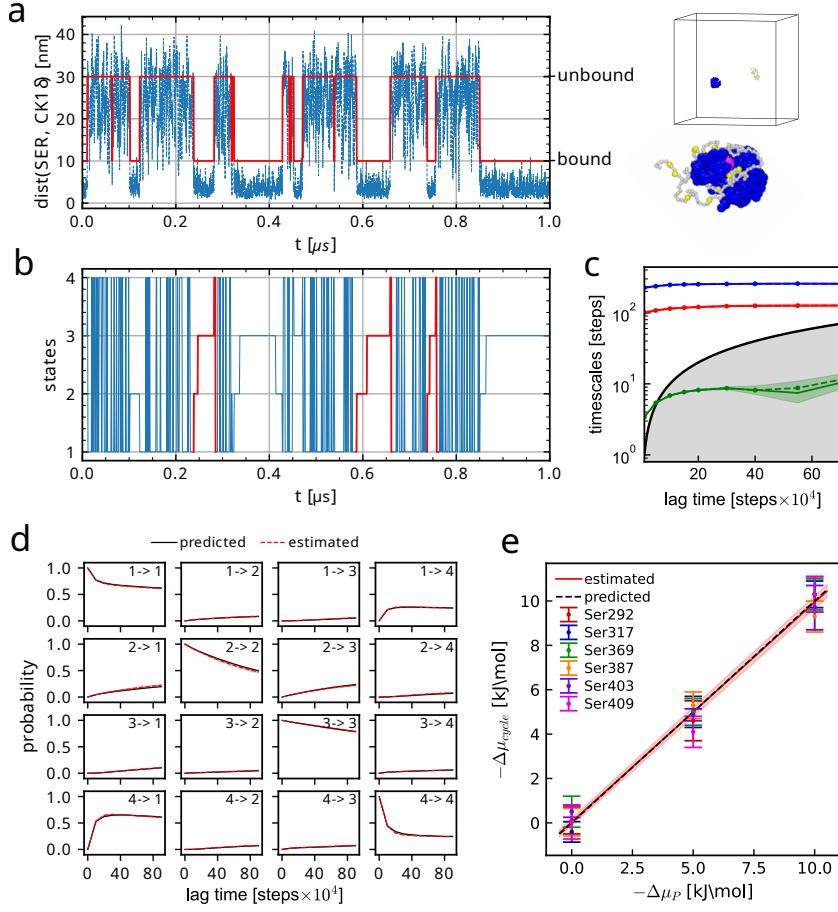


Figure 2: Thermodynamic consistency of simulations with phosphorylation step using MSMs. (a) Example of trajectory of the distance between Ser 403 of TDP-43 LCD and the active site of CK1 δ for the simulation at $\Delta\mu_P = -5$ kJ/mol. In red the two output states of the neural network (bound and unbound) with respective illustrative examples on the right. (b) Example of discretized 4-state MSM trajectory related to the trajectory in a; we highlight complete cycles in red. (c) Example of implied timescales from the 4-state MSM trajectory shown in b; they remain stable for $\tau \gtrsim 10^5$ MD steps (10 Markov chain steps). (d) Example of Chapman-Kolmogorov test from the 4-state MSM trajectory shown in b; prediction and model are in agreement. (e) We plot $\Delta\mu_{cycle}$ vs $\Delta\mu_P$; for all the six different phosphorylation sites, the chemical potential computed from Eq. 3 matches the applied chemical potential $\Delta\mu_P$. Errorbars on $\Delta\mu_{cycle}$ are obtained via bootstrapping of the total simulation trajectory collected.

where $\beta = 1/(k_B T)$, ΔU_P is the difference between the potential energy of the configuration with pSer and the one of the configuration with Ser, and $\Delta\mu_P$ is the chemical potential difference between the ATP and ADP molecules involved in the phosphorylation reaction (Eq. 1). ATP and ADP are modelled implicitly and are not explicitly simulated, with concentrations kept fixed and fully characterized through the choice of $\Delta\mu_P$. Indeed biological reactions, such as the phosphorylation

106 reaction, in living cells happen in open systems in which the concentrations of substrates, products
 107 and the chemical fuel are kept approximately constant over relevant timescales by e.g. metabolic
 108 processes. A non-zero value of $\Delta\mu_P$ biases the chemical reaction, pushing the simulation away from
 109 thermodynamic equilibrium.

110 The first step is to validate the thermodynamic consistency of our simulations by showing that
 111 the energy gain in a phosphorylation cycle (referred to as $\Delta\mu_{cycle}$ in the following) is equal to the
 112 chemical potential difference $\Delta\mu_P$ in the phosphorylation step. In order to compute $\Delta\mu_{cycle}$, we
 113 employ a discretization of the MD trajectory in a MSM. The simplest example of a phosphorylation
 114 cycle that we can build is a system with one enzyme and one substrate protein in which only one
 115 residue is reactive. In order to get complete phosphorylation cycles, we assume the exchange
 116 between TDP-43 and phosphorylated TDP-43 happens when substrate and enzyme are far away
 117 from each other without chemical driving and with equilibrium concentrations, through another
 118 MC step (Methods). This naturally happens in cells through the action of phosphatases that can
 119 catalyze a dephosphorylation reaction.

120 To gain insights into the effects of including phosphorylation through Eq. 2, we build an MSM
 121 from simulated MD trajectories. Firstly we distinguish between bound and unbound state using a
 122 neural network called VAMPnet [48]. VAMPnet is able to map molecular coordinates to Markov
 123 states through a score function called VAMP-2 score based on the Koopman's theory. Finding the
 124 transformation of the input variables that maximizes the VAMP-2 score is equivalent to optimizing
 125 the Markovianity of the output states. In this way we can easily distinguish between the two slowest
 126 processes, binding and unbinding, without arbitrarily choosing an a priori criterion of contact. As
 127 input for the neural network, we use the 154 distances between each residue of TDP-43 LCD and
 128 the active site of CK1 δ , while as output we ask for 2 states (ideally bound and unbound). We then
 129 filter spurious transitions using transition-based state assignment[49]. As an example, we show in
 130 Fig. 2a the trajectory of the distance between Ser 403 (the reactive residue) of TDP-43 LCD and
 131 the active site of CK1 δ for the simulation at $\Delta\mu_P = -5$ kJ/mol (SI Movie 1). We can see that
 132 the two states predicted by the neural network comprise bound configurations (when the distance
 133 between Ser and CK1 δ active site is smaller) and unbound configurations (when the distance is
 134 larger).

135 By distinguishing between Ser and pSer along the trajectory, we coarse-grain the system dynamics
 136 into the 4 states sketched in Fig. 1. Assuming that our system is a NESS, we can then compute
 137 the time-independent transition probabilities $T_{ij}(\tau)$ from state i to state j using the non-reversible
 138 Maximum Likelihood estimator [50, 51]. We report in Fig. 2b the resulting MSM discretized tra-
 139 jectory referred to the simulation in panel A. Complete cycles $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$ are highlighted
 140 in red. Every step of the Markov chain corresponds to 10^4 MD steps, or 0.1 ns in simulation time.
 141 For all our simulations, we choose a lag time $\tau = 10$ Markov chain steps (Methods). We show in
 142 Fig. 2c the implied timescales for the example case of reactive Ser 403 and $\Delta\mu_P = -5$ kJ/mol. In
 143 the end, we estimate the goodness of the MSM by looking at the Chapman-Kolmogorov test (CK
 144 test) [52, 53]. In all the validation simulations, the CK tests suggest good agreement between model
 145 and prediction for a wide range of lag times, as shown in Fig. 2d for the example case of reactive
 146 Ser 403 and $\Delta\mu_{in} = -5$ kJ/mol (SI Table S1, Table S2, Fig. S1 for complete data, Methods).

147 If our system is a NESS, the local detailed balance condition must be satisfied [3, 5]:

$$\Delta\mu_{cycle} = -k_B T \ln \left(\frac{T_{12}T_{23}T_{34}T_{41}}{T_{14}T_{43}T_{32}T_{21}} \right) \quad (3)$$

148 where $\Delta\mu_{cycle}$ is the energy injected into the system in one forward cycle $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 1$. It

is interesting to observe that the logarithm contains the ratio between the forward and backward transition probabilities. In many formulations, Eq. 3 is via rate coefficients rather than transition probabilities. For the short lag times considered here, we can estimate a rate matrix from the transition probability matrix and find virtually indistinguishable results for $\Delta\mu_{\text{cycle}}$ (SI Text). Since the transitions $1 \rightleftharpoons 2$, $3 \rightleftharpoons 4$ (the binding/unbinding of the enzyme with TDP-43 or phosphorylated TDP-43) and $4 \rightleftharpoons 1$ (the reservoir exchange step) satisfy detailed balance, while the phosphorylation reaction $2 \rightleftharpoons 3$ breaks detailed balance injecting into the system an amount of energy equal to $\Delta\mu_P$, we expect $\Delta\mu_{\text{cycle}}$ to be equal to $\Delta\mu_P$ (SI Text). We compute the estimated energy gain $\Delta\mu_{\text{cycle}}$ from the transition probabilities T_{ij} and plot them against the parameter $\Delta\mu_P$ of the phosphorylation step for different reactive Ser and $\Delta\mu_P$. Encouragingly, for all the six different phosphorylation sites, the chemical potential computed from Eq. 3 matches the applied chemical potential $\Delta\mu_P$ (Fig. 2e).

We repeated the estimate of $\Delta\mu_{\text{cycle}}$ using a 3-states MSM, in which the unbound states 1 and 4 are merged into the new state 1. The results are in agreement with the 4-states MSM (SI Table S1, SI Fig. S1). Indeed, the transition between state 1 and 4 has a very high rate and can be associated with the smallest implied timescale, that is lower than the lag time for $\tau = 10$ Markov chain steps or larger.

We also checked the reliability of VAMPnet by using considerably more input distances (4620 distances) and a different architecture for the case of reactive Ser 403 and $\Delta\mu_P = -5$ kJ/mol (Methods). The estimated $\Delta\mu_{\text{cycle}}$ with the new version of VAMPnet is $\Delta\mu_{\text{cycle}} = 4.7 \pm 0.6$ kJ/mol (implied timescales and CK test in SI Fig. S2).

Phosphorylation preferences are determined by sequence-specific interactions

Having established a model of chemically-driven dynamics, we investigate how sequence context determines the phosphorylation of the disordered protein TDP-43 LCD by the enzyme CK1 δ , so that we can begin to rationalize sequence-specificity of TDP-43 phosphorylation in experiments[24] and why C-terminal Ser residues such as Ser 410 are frequently found to be phosphorylated in experiments[24, 26, 44]. In our simulations, we follow directly the dynamics of TDP-43 LCD and CK1 δ folded domain (truncated at residue 294) on the single molecule level (Fig. 3a). We run 100 simulations of TDP-43 LCD in presence of CK1 δ and at physiological ATP/ADP ratio ($\Delta\mu_P = -48$ kJ/mol), which mimics in vitro kinase assays. In the simulations, unphosphorylated TDP-43 LCD will eventually encounter CK1 δ and give rise to different phosphorylation patterns, as shown for an example simulation on Fig. 3a. In this simulation TDP-43 LCD is initially phosphorylated in the C-terminal region. The kinase dissociates after two phosphorylation events and then binds again to the substrate. Multiple Ser residues in the C-terminus of TDP-43 LCD are phosphorylated, including Ser 410, which gets phosphorylated after ten other residues. In our simulations, Ser residues towards the C-terminus of TDP-43 LCD (Ser 369 to Ser 410) are more readily phosphorylated than Ser residues in the N-terminal region of the LCD (Ser 266 to Ser 350), with the phosphorylation rate r_P on average roughly 3-4 times larger in the C-terminal segment than in the N-terminal segment (Fig. 3b). In mass spectrometry analysis of TDP-43 from ALS patient samples, the phosphorylated sites (the 12 residues Ser 373, Ser 375, Ser 379, Ser 387, Ser 389, Ser 393, Ser 395, Ser 403, Ser 404, Ser 407, Ser 409 and Ser 410 [24, 44] and also Ser 369 [45]) are mostly in the C-terminal region and, interestingly, they are among the ones with largest phosphorylation rate r_P in our simulations. In particular, Ser 409/Ser 410 phosphorylation has long

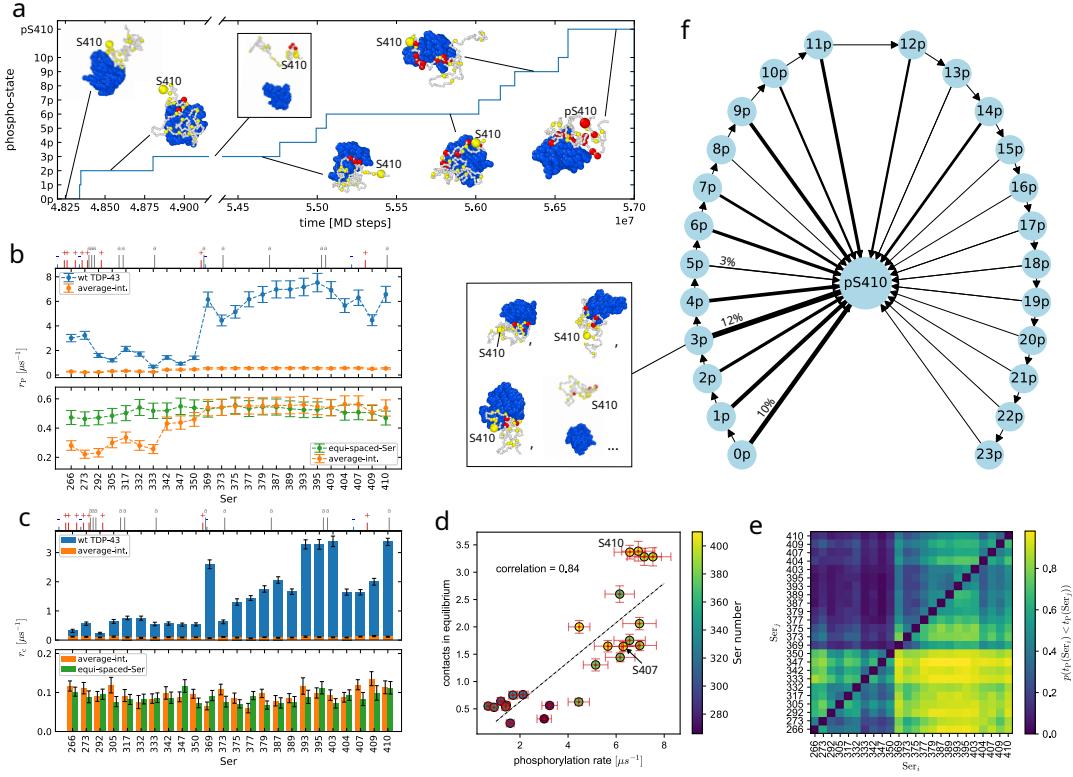


Figure 3: Analyzing the sequence dependence in phosphorylation dynamics of TDP-43. **(a)** Example trajectory discretized in phosphorylation-states relative to Ser 410. np is the state with Ser 410 not phosphorylated and n other phosphorylated Ser, $pS410$ are all the states with phosphorylated Ser 410. **(b)** We compare the phosphorylation rates r_P for every Ser for the wild type TDP-43 (blue), the averaged-interaction sequence (orange) and the averaged-interaction with equally-spaced Ser (green). The C-terminal is more phosphorylated. The ticks on top show the position of the charged (red '+') positive, blue '-' negative) and aromatic (grey 'a') residues. **(c)** Same comparison for the rates of contact r_c between Ser residues of TDP-43 LCD and the active site of CK1 δ in equilibrium simulations without phosphorylations. The positive charges in the N-terminal screen the interaction with the enzyme. Contact frequency are constant for the averaged-interaction sequence. **(d)** Correlation plot of contact frequency in equilibrium and phosphorylation rates for the wild type TDP-43. Ser 407 and Ser 410 have similar phosphorylation rate, but the probability of contact of Ser 410 is larger. **(e)** Probability $p(t(\text{Ser}_i) < t(\text{Ser}_j))$ of Ser_i being phosphorylated before Ser_j , data from 100 trajectories. C-terminal residues are much more likely to be phosphorylated before N-terminal residues. **(f)** Phosphorylation pattern representation for Ser 410. The thickness of the arrows represent the percentage of simulations in which Ser 410 was phosphorylated after n other Ser residues (e.g. 12% of simulations go from state 3p to pS410). We show in the inset some examples of 3p states.

been established as a hallmark of TDP-43 pathology in disease [26]. This qualitative agreement with simulations tentatively suggests that sequence specific interactions of TDP-43 LCD with the CK1 δ could explain why these residues are frequently found phosphorylated in experiments and in patient samples.

197 The differences in the phosphorylation rates can be largely accounted for by how readily Ser
198 residues engage in contacts with the CK1 δ active site (Fig. 3c), with the phosphorylation rates
199 strongly correlated with a sample Pearson correlation coefficient of 0.84 (Fig. 3d). In order to com-
200 pare the phosphorylation rates with the frequency of making contacts at equilibrium, we performed
201 MD simulations of the same system without phosphorylation MC steps. To establish to what extent
202 contacts predicts the relative phosphorylation rates, we consider a contact whenever all the three
203 distances to residues Asp 149, Phe 150 and Gly 151 close to the active site are less than 1 nm, in
204 the same way as for the MC phosphorylation step. By contrast, the acceptance probability for the
205 phosphorylation MC step for Ser residues once they are in contact is > 0.97 for the entire sequence
206 and the variations in the acceptance probability of the phosphorylation step are not correlated with
207 the variation of the phosphorylation rates (SI Fig. S8). Ser residues in the C-terminal segment of
208 the LCD, including Ser 369, Ser 393, Ser 395, Ser 403, and Ser 410, have the largest tendencies to
209 form contacts, as tracked by r_c , which is the rate at which a residue forms contacts with the CK1 δ
210 active site (Fig. 3c). At the same time these residues have within the statistical uncertainty the
211 fastest phosphorylation rates of the TDP-43 LCD (Fig. 3b). Ser residues in the N-terminal part
212 of the LCD (Ser 266 to Ser 350) form fewer contacts than serines in the C-terminal segment (Ser
213 369 to Ser 410), with the exception of Ser 373 in the latter segment, which also forms few contacts
214 with the active site of CK1 δ . The N-terminus is enriched in charged amino acids (mostly positive)
215 (Fig. 3b and c), which may hinder its binding to the CK1 δ active site, since the active site features
216 multiple charged residues and is overall positively charged (SI Fig. S7). On the other hand, the
217 C-terminus has more aromatic residues, which increase the attraction through cation-pi and pi-pi
218 interactions [54] (Fig. 3b and c). This difference between the N- and C-terminal segments of the
219 TDP-43 LCD is also apparent on the correlation plot in Fig. 3d, where the N-terminal residues
220 have both low rates and low number of contacts, whereas the C-terminal residues have mostly high
221 phosphorylation rates and many contacts with the active site.

222 Dynamics of TDP-43 serine phosphorylation is influenced by preceding 223 phosphorylation events

224 Although the correlation between the relative rates for CK1 δ and TDP-43 contact formation and
225 the phosphorylation rates is strong, there are deviations from the this simple relationship (Fig. 3d),
226 which could hint at structural correlations and possible correlations between phosphorylation events.
227 For instance, Ser 410 forms contacts more than two times more readily than Ser 407 but their
228 phosphorylation rates are the same within the statistical uncertainty (Fig. 3b and Fig. 3c). To
229 better understand the underlying correlations, we expanded our analysis of the phosphorylation
230 kinetics. To estimate the phosphorylation rates r_P , we assume that the phosphorylation process
231 is a memory-less process, which follows single-exponential kinetics [55]. In this case, observing a
232 single event is in principle sufficient to estimate the rates of a process. In addition to the number
233 of events one observes, the time spent waiting before an event happens also contributes to the rate
234 estimate. We checked the results by fitting the cumulative histograms of phosphorylation time
235 for each Ser with a simple single-exponential process and an exponential process conditioned to
236 another exponential process (e.g. the binding of TDP-43 to CK1 δ) (Methods). Most of the times
237 the conditioned exponential process fits perfectly. We found that the rate extrapolations from
238 the two fits are in agreement with the Bayesian estimates (SI Fig. S9). It is interesting to notice
239 that the fastest rate is different for every Ser (Ser 266 with a second rate coeffiecent of $13.5\mu\text{s}^{-1}$
240 and Ser 393 of $56\mu\text{s}^{-1}$) suggesting that the phosphorylation of some serines could involve other

processes than the binding to CK1 δ , e.g. the previous phosphorylation of another Ser. For Ser 410, the two fit extrapolation and the single-exponential fit are in agreement, with differences in the phosphorylation rate of about 2%, while for Ser 403 the conditioned process fit leads to an 8% smaller rate compared to the single-exponential fit. For Ser 407, the conditioned process yields a 10% larger rate. These comparison suggests that the phosphorylation of Ser 403 and Ser 407 could actually follow a more complex process.

We determined the most likely order of phosphorylation to understand correlation between phosphorylation events and differences from what the contact statistics at equilibrium would predict better. In order to study more deeply the phosphorylation pattern of TDP-43, we count for each Ser couple (Ser_i, Ser_j) how many times Ser_i is phosphorylated before Ser_j by aggregating data from our 100 trajectories to compute the probability $p(t_P(Ser_i) < t_P(Ser_j))$, where $t_P(Ser_i)$ is the time of phosphorylation for Ser_i from the start of the simulation. We show $p(t_P(Ser_i) < t_P(Ser_j))$ as a heatmap in Fig. 3e. We see again that, on the single-molecule level, C-terminal residues are typically phosphorylated first. The lower right corner shows that on average C-terminal residues are much more likely to be phosphorylated before N-terminal residues and as a corollary, the upper left sub-matrix shows that C-terminal residues are rarely phosphorylated after N-terminal residues. Instead, looking at the lower left block, we see that Ser 266 and Ser 273 are usually the first phosphorylated in the N-terminal region, while the serines within residues 333 and 350 are the last ones. In the end, by focusing in the C-terminus on the upper right block, we see that the first phosphorylations occur on Ser 369, Ser 393, Ser 395, Ser 403 and Ser 410, followed by Ser between 377 and 389 and Ser 407. In Fig. 3f we aggregate the data from the different trajectories and illustrate the likelihood for Ser 410 of getting phosphorylated after n other Ser through the thickness of the arrows. In the figure, the state pS410 includes all the possible configurations in which Ser 410 is phosphorylated, while np are the configurations with n pSer different from Ser 410, as shown in the inset for four different examples of state 3p. Very often Ser 410 is among the first three residues to be phosphorylated. Only in a few trajectories, Ser 410 is phosphorylated after nine or eleven other Ser residues are already phosphorylated. Ser 395 shows similar behaviour to Ser 410 (SI Fig. S11). While Ser 403 and Ser 407 are also phosphorylated early on by this analysis, they are less frequently the first Ser residues to be phosphorylated compared to Ser 410 (SI Fig. S11), which is in line with the deviations from single-exponential behaviour (SI Fig. S9). A possible influence of prior phosphorylation can also be detected for Ser 373. Ser 373 forms few contacts but is readily phosphorylated. The phosphorylation rate of Ser 373 is just slightly lower than for Ser 375, which has twice as many contacts. Indeed Ser 369 engages in many more contacts and the rates are just slightly higher than for Ser 373 and Ser 375. Fig. 3e shows that the probability $p(t_P(Ser_i) < t_P(Ser_j))$ of Ser 369 to be phosphorylated before Ser 373 and Ser 375 is approximately 0.8 and 0.7. Fig. S11 (SI) demonstrates that Ser 373 and Ser 375 are phosphorylated when multiple Ser residues are already phosphorylated. Changes in the interaction of CK1 δ with TDP-43 LCD as more residues are phosphorylated could explain why phosphorylation rates are not fully accounted for by the interaction propensities of the LCD with the active site of the enzyme. By analyzing long equilibrium MD simulations with VAMPnet[48], we find that the phosphorylation facilitates the binding of CK1 δ to the substrate TDP-43 LCD, with the binding free energy going from 5.0 kJ/mol in the case of wild type TDP-43 LCD to 1.2 kJ/mol for a chain with pSer 395, pSer 403 and pSer 410 (SI Text). As a result, the first phosphorylation events speed up further phosphorylation events, in agreement to what suggested by experiments [47], and we find that, in the simulations of enzymatic phosphorylation of TDP-43, phosphorylated TDP-43 LCD stays attached to CK1 δ (SI Fig. S3).

287 **Phosphorylation dynamics is determined by sequence context not relative**
288 **position to N- and C-termini**

289 The relative position of the Ser residues to the N- and C-termini does not affect the phosphorylation
290 rates. It has been hypothesized that the tendency of C-terminal residues to get phosphorylated
291 could be due to the greater accessibility of residues close to the N- and C-termini of a disordered
292 protein chain [24]. In order to understand whether the phosphorylation pattern is affected by the
293 position of the Ser residues along the TDP-43 LCD chain and not only by the neighboring residues,
294 we repeated the same simulation but replacing all the residues of TDP-43 LCD different from Ser
295 with an averaged interaction strength bead, 1) leaving the serines at their original positions and 2)
296 spreading them equally spaced along the chain. From the contact frequency r_c in the lower panel
297 in Fig. 3c, we can see that in equilibrium, before any phosphorylation occurs, the probability of
298 contact is uniform along the chain, suggesting that the ends are not a priori more accessible and
299 hence that sequence context and its effects on molecular recognition likely explain the prominence
300 of C-terminal TDP-43 phosphorylation. By looking at the phosphorylation rates r_P (lower panel
301 in Fig. 3b), we can see that the C-terminal domain is more phosphorylated in the case of simple
302 averaged-interaction beads. We also computed the probability $p(t_P(Ser_i) < t_P(Ser_j))$ for the case
303 of averaged-interaction chain (SI Fig. S10 left), which also demonstrated that the C-terminus is
304 phosphorylated before the N-terminus. This suggests that the negative charges of the pSer also
305 plays a role. These are denser in the C-terminus when TDP-43 gets hyper-phosphorylated. Indeed
306 by distributing the Ser residues at equal distances, phosphorylation rates are constant within the
307 statistical uncertainty. A similar behavior has already been found in experiments for the case of
308 cyclin-dependent kinases phosphorylation of multisite targets[56]. Overall, the phosphorylation
309 rates, as well as the contact frequency in equilibrium, are one order of magnitude smaller in the
310 case of averaged interaction sequence compared to the wild type TDP-43, highlighting once more
311 the importance of the sequence context.

312 **CK1 δ binds to TDP-43 condensates and dissolves condensates by hyper-**
313 **phosphorylation**

314 In our simulations, CK1 δ folded domain binds to TDP-43 LCD condensates and the LCD condensates
315 dissolve when they are hyper-phosphorylated (Fig. 4a). In cells, TDP-43 often phase-separates
316 into liquid-like droplets which has been linked to the formation of toxic aggregates. Recent exper-
317 iments have shown that hyper-phosphorylation of TDP-43 LCD can prevent phase separation and
318 aggregation by increasing the solubility of TDP-43 [24]. However it remains unclear whether ki-
319 nases, such as CK1 δ , actually bind to TDP-43 condensates, or only phosphorylate TDP-43 in dilute
320 solution. Snapshots from an example simulation with five CK1 δ enzymes are shown in Fig. 4a (SI
321 Movie 3), with the first snapshot depicting the starting configuration with 200 TDP-43 chains
322 phase-separated in a condensate and the enzymes (blue molecules) randomly placed in the box. Af-
323 ter 1 μ s of simulation time the enzymes are all attached at the surface of the condensate and they
324 are in the process of phosphorylating several serine residues (pSer in red). Hyper-phosphorylated
325 TDP-43 chains start to disassociate from the condensate, which appears almost entirely dissolved
326 after 5 μ s of simulation time in the last snapshot. We report in Fig. 4b the percentage of chains
327 in the condensate (blue, left y-axis) and the percentage of phosphorylated Ser (red, right y-axis)
328 over time from simulations with 1,3 and 5 enzymes, averaged over 4 independent replicas. The
329 percentage of TDP-43 chains in condensate drops over time as the phosphorylation count increases.

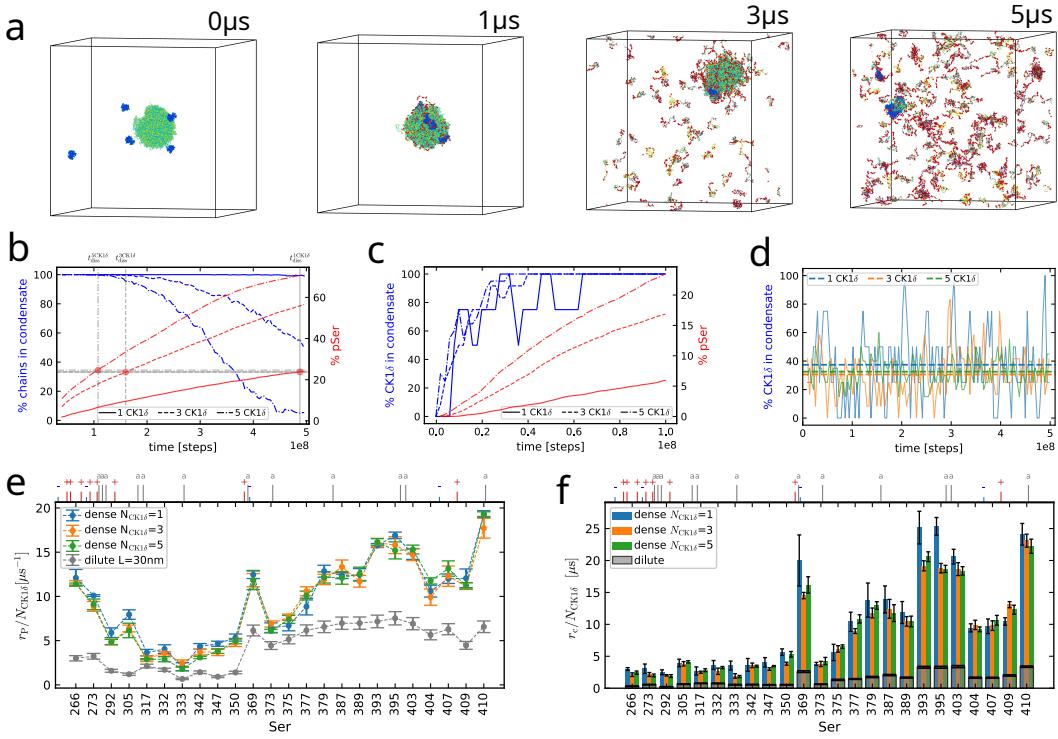


Figure 4: Analyzing the effect of hyper-phosphorylation of a TDP-43 LCD condensate and the interaction of CK1 δ with the condensate. All the simulations involved in the plots are performed in a cubic box of 100 nm side length with 200 TDP-43 LCD chains. (a) Snapshots from simulation with 5 CK1 δ in cubic box of 100 nm side length at times 0,1,3 and 5 μ s showing the dissolution of the TDP-43 condensate. The enzymes are colored in blue and the phospho-serine in red. (b) Percentage of TDP-43 chains in the condensate (blue, left y-axis) and percentage of phosphorylated Ser (red, right y-axis) in time for simulations with 1,3 and 5 CK1 δ . The condensate starts to dissolve after about 24 % pSer. (c) Percentage of CK1 δ attached to the condensate (blue, left y-axis) and percentage of phosphorylated Ser (red, right y-axis) in time for simulations with 1,3 and 5 CK1 δ . The enzymes remain attached to the condensate after some phosphorylations. (d) Percentage of CK1 δ attached to the condensate in time for equilibrium simulations without phosphorylation with 1,3 and 5 CK1 δ . In absence of pSer, only about 35% of the enzymes stay attached to the condensate in average. (e) Comparison of phosphorylation rates r_p for every Ser of TDP-43 LCD divided by the number of CK1 δ chains in dilute regime (grey) and in condensate in presence of 1,3 and 5 enzymes. In dense regime, phosphorylation of the ends of TDP-43 LCD is enhanced. The ticks on top show the position of the charged (red '+', blue '-') and aromatic (grey 'a') residues. (f) Contact rates r_c for every Ser of TDP-43 LCD divided by the number of CK1 δ chains in dilute regime (grey) and in condensate in presence of 1,3 and 5 enzymes at equilibrium.

We computed the size of the condensate using a standard clustering analysis algorithm, thanks to which we were able to distinguish the chains in the larger condensate at every frame (Methods). In every simulation, the condensate starts to lose TDP-43 chains when about 24-25% of Ser are phosphorylated. As a result, the speed of phosphorylation decreases with time, as the TDP-43 chains start to migrate in the dilute regime, far from the action of the enzymes. This effect is

335 particularly evident in the case of 5 CK1 δ after about 3.5 μ s. It is interesting to notice that the
336 speed of phosphorylation decreases slightly with time even before the beginning of the dissolution,
337 with C-terminal Ser being the most affected (SI Fig. S5 upper panel). The slowing down of the
338 phosphorylation rates of the most accessible Ser suggests a possible saturation effect. Moreover, we
339 notice that, at least after about 5% of Ser are phosphorylated, most of the TDP-43 chains feature
340 only 1 or 2 phosphates, with a small minority of chains being hyper-phosphorylated (SI Fig. S6
341 grey), supporting the idea of an early saturation of the most accessible phosphosites.

342 TDP-43 phosphorylation facilitates CK1 δ binding to the condensate, compared to unphosphorylated
343 condensates at equilibrium. In Fig. 4c we show the percentage of enzymes attached to the
344 condensate in time compared to the percentage of phosphorylated Ser, averaged again over 4 replicas.
345 With increasing phosphorylations, CK1 δ binds more stably to the condensate, suggesting that
346 the negative charges of the pSer residues enhance the binding to the enzyme positively charged
347 residues. In equilibrium simulations without phosphorylation only about 35 % of enzymes are
348 attached to the droplet in average (Fig. 4d).

349 In our simulations, the protein sequence context determines how much a given Ser residue is
350 phosphorylated in the condensates. We compute the phosphorylation rates r_P for each Ser of
351 TDP-43 LCD from the counts of phosphorylations and we compare them with the single-chain
352 simulations results. For this computation, we use only the part of the simulations before the
353 start of the condensate dissolution. We can see from Fig. 4e that the phosphorylation rates scale
354 proportionally to the number of enzymes in the box. Moreover, the phosphorylation of the N- and
355 C- terminal serines (namely Ser 266 and Ser 410) is enhanced, as well as for Ser 393, Ser 395 and
356 Ser 403, compared to the single-chain case (see correlation plot in SI Fig. S13 left).

357 The probability of contact with the active site of CK1 δ in the condensate for every Ser of TDP-
358 43 LCD differs from the single-chain case roughly by a factor 6. The C-terminus is more accessible,
359 in particular Ser 369, Ser 393, Ser 395, Ser 403 and Ser 410, as shown in Fig. 4f, similar to what
360 occurs in dilute regime. For the dense phase, the phosphorylation rates seem very well correlated
361 to contact rates in equilibrium for the C-terminal serines (sample Pearson correlation 0.91), while
362 the end of the N-terminus is more phosphorylated compared to what one would expect based on
363 the contact statistics from equilibrium simulations (SI Fig. S14 left).

364 **The role of CK1 δ disordered domain in the phosphorylation of TDP-43 365 in dilute solution and condensates**

366 In simulations with full-length CK1 δ , we find that the disordered region of CK1 δ (residues 295 to
367 415) slows down the phosphorylation of TDP-43 in accordance with experiments. In experiments
368 truncated CK1 δ is more active than the full-length enzyme[27]. First, we run simulations in dilute
369 solution. In our simulations CK1 δ is unphosphorylated and we do not allow possible autophosphorylations
370 of the CK1 δ IDR[28] (Fig. 5a, SI Movie 4). With the full-length enzyme, phosphorylation
371 is more restricted to a few residues in the C-terminal region and N-terminal serines are almost never
372 phosphorylated (Fig. 5c). This is due to a reduction of the contacts between TDP-43 residues and
373 the active site of CK1 δ , located in the folded domain of the enzyme. Indeed, in Fig. 5d we can see
374 that also the rates of contact in equilibrium without pSer r_c are reduced by one order of magnitude.
375 However, the Ser residues from residue 379 to 410 (10 of the 12 Ser residues found phosphorylated
376 in patient samples) have at least three times the phosphorylation rate of the N-terminus also in
377 the case of full-length CK1 δ , while Ser 373 and Ser 375 are only slightly more phosphorylated than
378 the N-terminal serines. The rate of active-site contact formation and the phosphorylation rates are

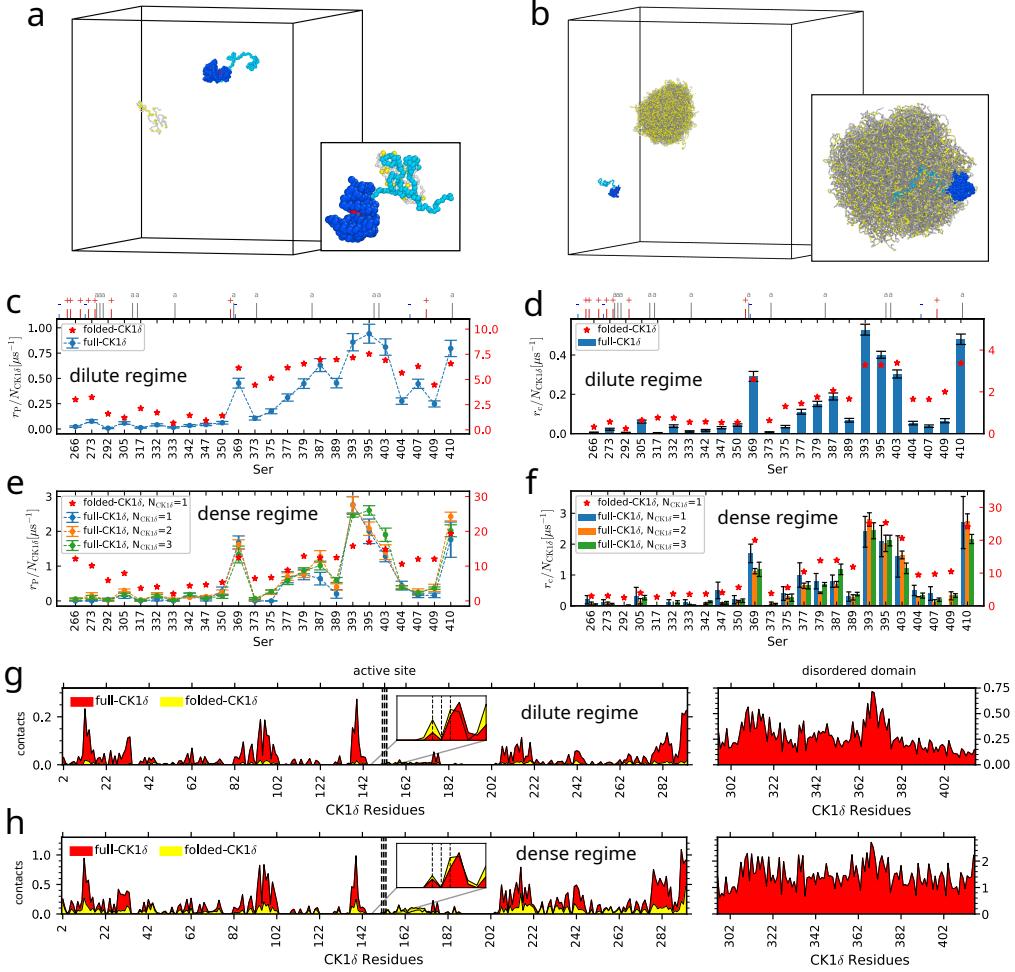


Figure 5: Assessing the role of CK1 δ disordered domain in phosphorylating TDP-43 LCD both in dilute and dense regime. **(a)** Example of simulation setup of full-length CK1 δ (blue) and TDP-43 LCD (grey, Ser in yellow) in dilute regime. The inset shows TDP-43 interacting with CK1 δ IDR (light blue, active site in red). **(b)** Example of simulation setup of full-length CK1 δ and condensate of 200 TDP-43 chains. The inset shows CK1 δ IDR (light blue) anchoring the folded domain (dark blue) to the surface of the condensate. **(c-d)** Phosphorylation rates r_P (c) and contact frequency in equilibrium r_c (d) for every Ser of TDP-43 LCD in presence of full-length CK1 δ in dilute regime. The ticks on top show the position of the charged (red '+' positive, blue '-' negative) and aromatic (grey 'a') residues. Results from simulations without tail are reported in red on the right y-axis. **(e-f)** Phosphorylation rates r_P (e) and contact frequency in equilibrium r_c (f) for every Ser of TDP-43 LCD in presence of full-length CK1 δ in condensate. Results from simulations without tail are reported in red on the right y-axis. **(g-h)** Contact rates for every residue of full-length CK1 δ in dilute (g) and dense (h) regime at equilibrium. The disordered region (residues 295 to 415) has more contacts. Simulations with full-length CK1 δ (red) have more contacts than without IDR (yellow), the insets show that contacts of the active site (Asp 149, Phe 150, Gly 151) are comparable.

more correlated in this case than for the folded domain alone (Fig.S12). However Ser 404, Ser 407 and Ser 409 have similar contacts in equilibrium compared to the N-terminal serines, but higher rate of phosphorylation. As neighboring Ser residues become phosphorylated, binding to the active site, and thus phosphorylation, is likely enhanced.

The disordered tail of CK1 δ partially auto-inhibits the enzyme not by occluding the active site, but by sequestering the substrate. TDP-43 LCD interacts more strongly with the tail of CK1 δ than with the folded domain of CK1 δ (inset in Fig. 5a). Consequently, full-length CK1 δ binds more strongly to TDP-43 LCD than the folded domain on its own. From equilibrium simulation of the enzyme and TDP-43 LCD as analysed by VAMPnet[48], the binding free energy goes from 5 kJ/mol to -4 kJ/mol. The tail sequesters the substrate in the dilute phase allowing less contacts with the active site, that is located on the opposite side of the enzyme surface, and thus resulting in fewer TDP-43 LCD phosphorylation events. We can see in of Fig. 5g that the disordered domain of CK1 δ (residues from 295 to 415) has more contacts with TDP-43 LCD compared to the folded domain surface residues in equilibrium simulations without pSer. By looking instead to the residues in the folded domain (from 0 to 294), we notice that the full-length CK1 δ (red) has in general more contacts than the CK1 δ without tail (yellow), due to the stronger binding of TDP-43 LCD with the CK1 δ IDR. However, the active site features a comparable amount of contacts in two cases, with residue 149 having even more contacts in the simulations without tail. By contrast, we find in our simulations that the IDR does not inhibit the CK1 δ by occluding the active site. The disordered tail of CK1 δ rarely forms close contacts with the active site and any close contacts are lost very rapidly (Fig. S15).

We run simulations with 200 chains of TDP-43 LCD in a cubic box of 100 nm side length, adding 1,2 and 3 chains of full-length CK1 δ (Fig. 5b, SI Movie 5). In the dense regime there is high amount of chains, so both contacts and phosphorylation counts increase compared to the dilute case (Fig. 5 e and f). Contacts and phosphorylation counts are also highly correlated in this case, since the abundance of chains allows the enzyme to neglect the less accessible phosphosites and phosphorylate the most accessible ones from every chain. This constitutes a disadvantage for Ser 389, Ser 404, Ser 407 and Ser 409 that are less phosphorylated compared to the dilute case. We note that the phosphorylation rates are directly proportional to the number of enzymes acting on the condensate, as well as the contact statistics (SI Fig. S14). By comparing the phosphorylation rates in condensate with the simulations without IDR, we notice that they are in general lower for the full-length case. The tail acts as a filter, allowing only the phosphorylation of Ser 369, Ser 393, Ser 395, Ser 403, Ser 410 and to a lesser extent Ser 377, Ser 379, Ser 387 and Ser 389, apart from some other very rare phosphorylation events. The effect of the tail on the relative phosphorylation rates is even more pronounced than what we observed in the simulations of CK1 δ and single chains of TDP-43 LCD. When more than 5% of Ser residues in TDP-43 LCD are phosphorylated, most of the TDP-43 chains feature 1 or 2 phosphates. The distribution of the number of phosphorylated Ser residues per TDP-43 LCD chains is even narrower than in the simulations without the CK1 δ disordered tail (SI Fig. S6 orange). As for the case without tail, the phosphorylation rate of the most accessible phosphosites decreases with time (SI Fig. S5. This saturation effect of the most reactive Ser residues is apparent even before the eventual dissolution of the condensate (SI Fig. S5 lower panel).

The disordered tail of CK1 δ facilitates the binding of CK1 δ to TDP-43 condensates. We observe that the tail recruits the condensate and keeps the folded region anchored to its surface, as illustrated in the inset of Fig. 5b. Also in this case the disordered domain of CK1 δ has more contacts compared to the folded domain surface residues, as shown in Fig. 5h. Thanks to the disordered tail, the enzyme

425 remains bound to the condensate surface even in absence of phosphorylations. As a consequence,
426 the residues of the folded domain of CK1 δ form more contacts with TDP-43 compared to the case
427 without the tail (Fig. 5h, residues from 0 to 294). However, the number of contacts of TDP-43
428 with the active site are again comparable in the two cases, explaining why the rates of binding to
429 the active site shown in Fig. 5f are not greater than the ones in Fig. 4f. As for the single-chain
430 simulation, the inaccessibility of the active site seems to be due to its opposite location on the
431 enzyme surface compared to the disordered tail. Despite the stable anchoring to the condensate,
432 the enzyme active site faces outwards, making it less accessible to the TDP-43 serines. Even in this
433 case the auto-inhibitory and self-regulatory effects of the enzyme tail do not seem to be due to an
434 obstruction of the active site, since CK1 δ IDR interacts strongly with the condensate (SI Fig. S16).

435 Discussion

436 We have demonstrated how Markov-state models enable us to straightforwardly validate molecular
437 simulations of chemically-driven non-equilibrium steady states (NESS). Chemically-driven NESS
438 are essential in cell biology[1]. Cells require the constant turnover of fuels and metabolites to
439 grow and thrive. Chemically-driven NESS are likely also essential in the function of biomolecular
440 condensates in the cells[2].

441 We envisage that our approach to establish the thermodynamic consistency of simulations and
442 the combination of molecular dynamics and Monte Carlo can be readily applied to more complex
443 systems and simulations of such systems in high resolution [41]. For more complex systems, extracting
444 kinetically meaningful states becomes even more challenging. In this respect advances based on
445 neural networks and Koopman theory are highly encouraging[48, 53, 57, 58].

446 PTMs such as phosphorylation of proteins are a fundamental regulatory mechanism in cells
447 and with molecular simulations we can start to investigate how protein sequence and structure de-
448 termine substrate-enzyme interactions and PTM patterns. Our simulations demonstrate that the
449 IDR of CK1 δ could have important roles in TDP-43 phosphorylation, 1) by facilitating the binding
450 to condensates and 2) by auto-inhibiting the enzyme, which our simulations capture in line with
451 experiments [27, 28]. It is important to note that details of the conformations of proteins will be
452 critical for the molecular recognition of potential phosphorylation sites by kinases and more detailed
453 molecular simulations [59] will be required to fully understand the recognition mechanisms. A more
454 detailed description of conformational flexibility will be particularly important to understand in
455 detail how the disordered tail of CK1 δ inhibits phosphorylation and whether sequestering of the
456 disordered substrate rather than binding to the active site really underpins auto-inhibition by the
457 CK1 δ IDR. Overall our simulations point to a potential preference for the C-terminal residues of
458 TDP-43 on account of its sequence. Aggregated TDP-43 in patient samples is frequently phospho-
459 rylated at, e.g., Ser 379, Ser 403/Ser 404 and Ser 409/Ser 410 [26, 46], which are among the most
460 phosphorylated residues also in our simulations. Due to the high concentration of substrates in
461 condensates, proteins are readily phosphorylated. The phosphorylation rates for Ser residues are
462 larger for TDP-43 in condensates than in the dilute phase. Interestingly, phosphorylation patterns
463 are overall similar in dilute solution and condensates. While there are differences in the phospho-
464 rylation propensities, sequence context still determines which sites can be phosphorylated. One
465 can speculate that differences between phosphorylation in dilute and dense solution could be partly
466 explained by the overall higher phosphorylation level in condensates, which means that some sites
467 will effectively be more readily phosphorylated than in dilute solutions[25], while the kinase may
468 retain sequence-dependent recognition of substrates in the condensates.

469 **Methods**

470 **Coarse-grained MD simulations**

471 In our work, we simulated TDP-43 LCD and the kinase CK1 δ using a one-bead-per-residue coarse-
472 grained model called hydrophobicity scale model [31] (HPS model) and a modified version of it,
473 referred to as modified HPS model in the text (SI Text). In these models, the water solvent and
474 the ions concentration are implicit in the pair potential definition. We used the original HPS for
475 the thermodynamic consistency validation simulations, in which we preferred to give priority to
476 the frequency of the binding and phosphorylation events at the expense of having a more realistic
477 force field, in order to get better statistics. For the other simulations, we used the modified HPS
478 model [30] in which cation-pi interactions are enhanced [60] and folded domains interaction are
479 reduced by 30% [61, 62] (SI Text). Simulations were conducted using Langevin dynamics at a
480 temperature of 300 K and friction coefficient of 0.001 ps⁻¹ and in a cubic box with periodic boundary
481 conditions of side length of 30 nm for the single TDP-43 chain simulations and 100 nm for the
482 condensate simulations. The simulated TDP-43 LCD includes residues from 261 to 414 of the
483 full-length TDP-43. The folded domain of CK1 δ (residues from 1 to 294) follows a rigid body
484 dynamics with rotational drag coefficient of 4 ps⁻¹ for every axis, the structure is provided by
485 <https://alphafold.ebi.ac.uk/entry/P48730>.

486 For the dilute regime, 100 simulations with phosphorylation reaction step and without reservoir
487 exchange step were run, 2×10^8 MD steps long (2 μ s in simulation time) for the case with 1 wild
488 type TDP-43 LCD and 1 CK1 δ folded-domain, 4×10^8 MD steps long (4 μ s in simulation time)
489 both for the case with 1 averaged-interaction polymer and 1 CK1 δ folded-domain and for the case
490 with 1 wild type TDP-43 LCD and 1 full-length CK1 δ . To characterize the intrinsic affinity of the
491 enzyme for TDP-43 LCD, we repeated the same simulations, but without phosphorylation reactions
492 at thermodynamic equilibrium. We collected in total 450 μ s of simulation time for the case of wild
493 type TDP-43 LCD and CK1 δ folded-domain and 900 μ s for the averaged-interaction polymer and
494 for full-length CK1 δ . The averaged interaction polymer is built by substituting the TDP-43 LCD
495 residues different from Ser with a bead having average TDP-43 LCD mass, size parameter σ and
496 hydropathy parameter λ (SI Text) 1) leaving the serines at their original positions and 2) spreading
497 them equally spaced along the chain.

498 We also simulated a condensate of 200 TDP-43 LCD chains. We ran 4 simulations 5×10^8 MD
499 steps long (5 μ s in simulation time) with phosphorylation steps without reservoir exchange step (as
500 for the single chain simulations) in presence of 1,3 or 5 CK1 δ folded-domain chains and of 1,2 or 3
501 full-length CK1 δ chains. We repeated the same simulations, but without phosphorylation reactions
502 at thermodynamic equilibrium, collecting a total of 20 μ s of simulation time for each case.

503 All the simulations involved in this paper were performed using the Python package HOOMD-
504 blue version 3.8.1. The code used for the simulations is available at https://github.com/ezippo/hoomd3_phosphorylation. The Ashbaugh-Hatch pair potential for the non-bonded interactions is
505 available at https://github.com/ezippo/ashbaugh_plugin as a HOOMD-blue plugin.

507 **Phosphorylation reaction through a Monte Carlo step**

508 In addition to the standard MD simulation, we added a Monte Carlo step to mimic the phosphory-
509 lation reaction. Every 200 steps of MD simulation, we check if one of the TDP-43 phosphosites is in
510 contact with the active site of CK1 δ , the area of the enzyme that catalyzes the reaction, identified
511 with the residues Asp149, Phe150 and Gly151. The contact criterion is the following: the three

512 distances between the TDP-43 phosphosite and the residues of the CK1 δ active site must all be
 513 less than 1 nm; in case more than one phosphosite is in contact with the active site at the same
 514 time-step, only the closest one is taken into account. When a contact occurs, we try to switch
 515 the Ser in contact into pSer (or the opposite) with a Metropolis-like acceptance probability in
 516 Eq. 2. The reverse reaction, that is the exchange of pSer with Ser, can also occur with probability
 517 $A(p\text{Ser}, \text{Ser}) = \min(1, \exp(\beta\Delta U_P + \beta\Delta\mu_P))$, but it is less likely to happen when there is a chemical
 518 potential difference favouring the protein phosphorylation. ATP, ADP are modelled implicitly and
 519 are not explicitly simulated, with concentrations kept fixed and fully characterized through the
 520 choice of $\Delta\mu_P$.

521 The chemical potential difference in a reaction in units of $k_B T$ is given by the logarithm of
 522 the product to substrate concentration ratio. Considering that the ATP concentration in cells is
 523 around 1 mM, the concentration of ADP is around 10 μM and fixing a temperature $T = 300$ K
 524 for our simulations, we get a chemical potential difference for a phosphorylation reaction $\Delta\mu_P =$
 525 $\mu_{ADP} - \mu_{ATP} \simeq -11.5$ kcal/mol $\simeq -48$ kJ/mol (SI Text). Observe that the ATP concentration is
 526 two orders of magnitude larger than the ADP concentration, leading to a large negative $\Delta\mu_P$ that
 527 favors the exchange of Ser into pSer and disfavors the opposite reaction. Moreover we can mimic the
 528 ATP to ADP concentration ratio by changing the chemical potential difference in our simulation
 529 at fixed temperature. We used $\Delta\mu_P = 0, -5, -10$ kJ/mol for the validation of the thermodynamic
 530 consistency simulations and $\Delta\mu_P = -48$ kJ/mol for all the other simulations in dilute regime and
 531 condensate.

532 **Dephosphorylation step**

533 In our validation simulations, we assume the exchange between TDP-43 and phosphorylated TDP-43
 534 happens without chemical driving and with equilibrium concentrations, through another Metropolis-
 535 like step (reservoir exchange step). Every 200 MD steps, we check if the distances between the
 536 TDP-43 phosphosite and the 3 residues of the CK1 δ active site is larger than 25 nm (half box side
 537 length). In that case, we randomly swap the pSer of the phosphorylated TDP-43 with a Ser (or the
 538 opposite) with a Metropolis-like acceptance probability:

$$A_D(p\text{Ser}, \text{Ser}) = \min(1, \exp(-\beta\Delta U_D)) \quad (4)$$

539 where ΔU_D is again the difference between the potential energy of the configuration with Ser and
 540 the one of the configuration with pSer. In this case there is no chemical driving force, the reaction
 541 obeys detailed balance and thus it does not inject any additional energy into the system. This
 542 exchange step mimics a larger reservoir of TDP-43 and phosphorylated TDP-43 and thus enables
 543 us to simulate multiple phosphorylation cycles on the level of a single enzyme and single substrate
 544 protein simulation.

545 **Thermodynamic consistency validation simulations**

546 We simulated the system with one CK1 δ and one TDP-43 LCD with only one reactive residue.
 547 We repated the simulation for 6 different reactive serines along the TDP-43 LCD, i.e. Ser292,
 548 Ser 317, Ser 369, Ser 387, Ser 403, Ser 409, and $\Delta\mu_P = 0, -5, -10$ kJ/mol. Simulations were
 549 conducted for 20 μs in a cubic box of 50 nm side length with periodic boundaries using HPS
 550 model force field. In order to get better statistics, we took a 100 μs long trajectory for Ser 403 and
 551 $\Delta\mu_P = 0, -5$ kJ/mol. We used these longer trajectories for the estimates of $\Delta\mu_{\text{cycle}}$ with different

lag times and with the version of VAMPnet with more input distances. Errorbars on $\Delta\mu_{\text{cycle}}$ were obtained via bootstrapping of the total simulation trajectory collected.

VAMPnet architecture and training

For the bound state recognition, we performed a nonlinear dimension reduction using a neural network with two identical lobes, following the VAMPnet architecture and the hyper-parameter optimization used by Mardt et al. [48]. Each lobe is composed by an input layer with 154 nodes, one for each residue of TDP-43 LCD, one hidden layer with 30 nodes that employs exponential linear units (ELU) and an output layer with 2 nodes, ideally bound and unbound state, and a final Softmax classifier to obtain probabilities of bound and unbound configurations as output. As input for the neural network, we used the 154 distances between each residue of TDP-43 LCD and the active site of CK1 δ . We chose a learning rate of 0.5×10^{-2} and a batch size of 4×10^4 . The neural network was trained for 100 epochs on $90\mu\text{s}$ of equilibrium HPS model[31] simulation with one TDP-43 LCD and one CK1 δ .

The neural network returns the probability of being in one of the 2 output states (ideally bound and unbound state) for each snapshot of the trajectory. We assigned each snapshot to the state with higher probability, filtering those with a probability between 30% and 70% using transition-based state assignment[49]. In other words, these configurations were assigned based on the state of the previous and following snapshots, in order to filter out spurious transitions.

In order to test the generality of our method, we repeated the bound state recognition with VAMPnet, but using more input nodes. In particular, we used the distances between all the residues of TDP-43 LCD and 30 equally spaced residues of CK1 δ , resulting in an input layer of 4620 nodes. This time we used 2 hidden layers with 154 and 30 nodes each, and an output layer with 2 nodes. We reduced the learning rate to 0.5×10^{-3} and the batch size to 10^4 .

Implied timescales and Chapman-Kolmogorov test

The choice of the lag time was done by looking at the implied timescales. We can estimate the implied timescales of the Markov model from the eigenvalues of the transition matrix as:

$$t_i(\tau) = -\frac{\tau}{\ln |\lambda_i(\tau)|} \quad (5)$$

with $\lambda_i(\tau)$ the eigenvalues of $T_{ij}(\tau)$. We chose a lag time τ such that $t_i(\hat{\tau})$ is approximately constant for every $\hat{\tau} \geq \tau$. We estimated $\Delta\mu_{\text{cycle}}$ for different lag times (1,10,20 Markov chain steps) for the case of reactive Ser 403 and $\Delta\mu_{\text{P}} = 0, -5 \text{ kJ/mol}$. For $\tau \geq 10$ Markov chain steps, the estimated $\Delta\mu_{\text{cycle}}$ is in agreement with $\Delta\mu_{\text{P}}$ (SI Table S2).

We estimated the goodness of the MSM by looking at the Chapman-Kolmogorov test (CK test) [52] [53]. In a Markovian process, the transition matrix satisfies the relation

$$T_{ij}(n\tau) = [T_{ij}(\tau)]^n \quad (6)$$

with $n \geq 1$. In other words, the transition matrix of the model estimated at lag time $n\tau$ must be equal to the transition matrix to the power n of the model estimated at lag time τ . The CK test compares $T_{ij}(n\tau)$ (the estimated transition matrix) and $T_{ij}^n(\tau)$ (the predicted transition matrix) for every possible transition $i \rightleftharpoons j$ and different lag times $n\tau$.

588 **Estimate of phosphorylation rates and fit of phosphorylation processes**

589 In all the single TDP-43 LCD chain simulations, we estimated the phosphorylation rates r_P assuming
 590 the phosphorylation process is without memory and thus follows single-exponential kinetics. In
 591 all the collected 100 simulations, we had at most one phosphorylation event for each Ser residue,
 592 happening at time $t_P^i < t_{\text{tot}}$ for simulation i , with t_{tot} the total time of the simulation. In this case,
 593 we can use the maximum likelihood estimator for the rate with uniform prior distribution [55]

$$r_P = \frac{n+1}{\Theta}, \quad \text{var}(r_P) = \frac{n+1}{\Theta^2}, \text{ with } \Theta = \sum_{i=1}^n t_P^i + (100-n)t_{\text{tot}} \quad (7)$$

594 where n is the counts of simulations with one phosphorylation event for the Ser took in consideration.
 595 Instead, for the simulations in condensate, in which we have multiple TDP-43 chains and
 596 thus multiple phosphorylation events for each Ser residue, we computed r_P as the total count of
 597 phosphorylation events in the simulation divided by the total simulation time. In this case, the
 598 error on the estimate of the rate is computed as the standard error of the mean from the different
 599 replicas. In the same way we also computed all the contacts rates r_c and their error.

600 However, since the phosphorylation of a Ser can happen only if TDP-43 is bound to the enzyme,
 601 it is more appropriate to take into account the conditional probability of the phosphorylation event
 602 given the binding of TDP-43 and CK1 δ already occurred. Given $p_B(t)dt = r_B \exp(-r_B t)dt$ the probability
 603 of binding between time t and $t + dt$ and $p_P(t)dt = r_P \exp(-r_P t)dt$ the probability
 604 of having a phosphorylation between time t and $t + dt$, the conditional probability of having a
 605 phosphorylation between time t and $t + dt$ given that TDP-43 is bound to CK1 δ is

$$P(t|\text{bound})dt = \int_0^t p_B(t')dt' p_P(t-t')dt = \frac{r_B r_P}{r_B - r_P} (e^{-r_P t} - e^{-r_B t}) dt \quad (8)$$

606 If we call $P_c(t < T)$ the probability of having a phosphorylation event within time T in our
 607 simulations, we can write its complementary as:

$$1 - P_c(t < T) = 1 - \int_0^T P(t)dt = \frac{r_B r_P}{r_B - r_P} \left(\frac{e^{-r_P t}}{r_P} - \frac{e^{-r_B t}}{r_B} \right) \quad (9)$$

608 Instead, if we assume that the binding process is much faster than the phosphorylation one ($r_B \gg$
 609 r_P), than we can approximate $P(t|\text{bound}) \sim p_P(t)$ and $1 - P_c(t < T) \sim \exp(-r_P t)$.

610 From the 100 simulations used to estimate the phosphorylation rates, we computed the normalized
 611 inverse cumulative histogram of the phosphorylation events time T , where each time bin gives
 612 the phosphorylation counts for that bin plus the counts of all the following bins, divided by the
 613 number of simulations. We fitted it with $1 - P_c(t < T)$ both for a single-exponential process and a
 614 conditioned single-exponential process.

615 **Condensate identification with clustering analysis**

616 In order to identify the TDP-43 LCD condensate in the trajectory file, we used the DBSCAN
 617 (Density-Based Spatial Clustering of Applications with Noise) clustering analysis algorithm[63]. It
 618 is an efficient algorithm to identify clusters based on a euclidean distance cut-off ϵ and a minimum
 619 cluster size parameter n_{\min} . Particles with at least n_{\min} neighbors within a distance ϵ are considered

620 core particles of the cluster. Instead, particles with less than n_{\min} neighbors are considered non-core
621 particles and they are assigned to a cluster only if at least one of their neighbors is a core particle.

622 For the estimate of the condensate size and of the percentage of $CK1\delta$ in contact with the
623 condensate, we used the positions of every bead as input data and we chose the parameters $\epsilon = 1 \text{ nm}$
624 and $n_{\min} = 2$. With this choice, every isolated chain is considered as a cluster and two different
625 chains belongs to the same cluster whenever at least one of their particles is in contact (within 1 nm).
626 However, varying ϵ between 0.8 nm and 3 nm and n_{\min} between 2 and 5 does not significantly change
627 the results. We accounted for the periodic boundary conditions by centering the condensate in the
628 box at every frame.

629 Acknowledgments

630 E.Z. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation),
631 Project No. 233630050 – TRR 146. L.S.S. thank M³ODEL and ReALity (Resilience, Adaptation
632 and Longevity) and Forschungsinitiative des Landes Rheinland-Pfalz for support. This project was
633 funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - SFB 1551
634 – Project No. 464588647 . Further, we gratefully acknowledge the computing time granted on
635 the supercomputers Mogon II at Johannes Gutenberg University Mainz, which is a member of the
636 AHRP (Alliance for High Performance Computing in Rhineland Palatinate) and the Gauss Alliance
637 e.V. We thank Alex Holehouse for insightful discussions

638 Author contributions

639 E.Z. ran the simulations, analysed data, interpreted results, wrote the manuscript. D.D. provided
640 important intellectual knowledge and assisted in designing the study, reviewed the manuscript. T.S.
641 conceived the study, provided important intellectual knowledge, reviewed the manuscript. L.S.S.
642 conceived and supervised the study, interpreted results, wrote and reviewed manuscript.

643 References

- 644 1. Seifert, U. From Stochastic Thermodynamics to Thermodynamic Inference. *Annual Review of
645 Condensed Matter Physics* **10**, 171–192 (2019).
- 646 2. Weber, C. A., Zwicker, D., Jülicher, F. & Lee, C. F. Physics of active emulsions. *Reports on
647 Progress in Physics* **82**, 064601 (2019).
- 648 3. Qian, H. Phosphorylation Energy Hypothesis: Open Chemical Systems and Their Biological
649 Functions. *Annual Review of Physical Chemistry* **58**. PMID: 17059360, 113–142 (2007).
- 650 4. Speck, T. Modeling of biomolecular machines in non-equilibrium steady states. *The Journal
651 of Chemical Physics* **155**, 230901 (2021).
- 652 5. Rao, R. & Esposito, M. Nonequilibrium Thermodynamics of Chemical Reaction Networks:
653 Wisdom from Stochastic Thermodynamics. *Phys. Rev. X* **6**, 041064 (4 2016).
- 654 6. Banani, S. F., Lee, H. O., Hyman, A. A. & Rosen, M. K. Biomolecular condensates: organizers
655 of cellular biochemistry. *Nature reviews Molecular cell biology* **18**, 285–298 (2017).

- 656 7. Gil-Garcia, M. *et al.* Local environment in biomolecular condensates modulates enzymatic
657 activity across length scales. *Nature Communications* **15**, 3322 (2024).
- 658 8. Alberti, S. & Dormann, D. Liquid–Liquid Phase Separation in Disease. *Annual Review of
659 Genetics* **53**, 171–194 (2019).
- 660 9. Visser, B. S., Lipiński, W. P. & Spruijt, E. The role of biomolecular condensates in protein
661 aggregation. *Nature Reviews Chemistry*, 1–15 (2024).
- 662 10. Gopal, P. P., Nirschl, J. J., Klinman, E. & Holzbaur, E. L. F. Amyotrophic lateral sclerosis-
663 linked mutations increase the viscosity of liquid-like TDP-43 RNP granules in neurons. *Proceedings of the National Academy of Sciences* **114**, E2466–E2475 (2017).
- 664 11. Wegmann, S. *et al.* Tau protein liquid-liquid phase separation can initiate tau aggregation.
eng. *EMBO J* **37** (2018).
- 665 12. Srinivasan, E. *et al.* Alpha-Synuclein Aggregation in Parkinson’s Disease. *Frontiers in Medicine*
666 **8** (2021).
- 667 13. Yan, X. *et al.* Intra-condensate demixing of TDP-43 inside stress granules generates patholog-
668 ical aggregates. *bioRxiv*, 2024–01 (2024).
- 669 14. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amy-
670 otrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
- 671 15. Josephs, K. A. *et al.* TDP-43 is a key player in the clinical features associated with Alzheimer’s
672 disease. *Acta neuropathologica* **127**, 811–824 (2014).
- 673 16. Babu, M. M. The contribution of intrinsically disordered regions to protein function, cellular
674 complexity, and human disease. *Biochemical Society Transactions* **44**, 1185–1200 (2016).
- 675 17. Martin, E. W., Holehouse, A. S., Grace, C. R., Hughes, A., Pappu, R. V. & Mittag, T. Sequence
676 Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to
677 and upon Multisite Phosphorylation. *Journal of the American Chemical Society* **138**. PMID:
678 27807972, 15323–15335 (2016).
- 679 18. Changiarath, A. *et al.* Promoter and Gene-Body RNA-Polymerase II co-exist in partial demixed
680 condensates. *bioRxiv* (2024).
- 681 19. Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E. & Zweckstetter, M. Liquid-liquid phase
682 separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nature
683 Communications* **8**, 275 (2017).
- 684 20. Monahan, Z. *et al.* Phosphorylation of the FUS low-complexity domain disrupts phase sepa-
685 ration, aggregation, and toxicity. eng. *EMBO J* **36**, 2951–2967 (2017).
- 686 21. Ranganathan, S., Dasmeh, P., Furniss, S. & Shakhnovich, E. Phosphorylation sites are evolu-
687 tionary checkpoints against liquid–solid transition in protein condensates. *Proceedings of the
688 National Academy of Sciences* **120**, e2215828120 (2023).
- 689 22. Zwicker, D., Decker, M., Jaensch, S., Hyman, A. A. & Jülicher, F. Centrosomes are autocat-
690 alytic droplets of pericentriolar material organized by centrioles. *Proceedings of the National
691 Academy of Sciences* **111**, E2636–E2645 (2014).
- 692 23. Jülicher, F. & Weber, C. A. Droplet Physics and Intracellular Phase Separation. *Annual Review
693 of Condensed Matter Physics* **15**, 237–261 (2024).

- 696 24. Gruijs da Silva, L. A. *et al.* Disease-linked TDP-43 hyperphosphorylation suppresses TDP-43
697 condensation and aggregation. *EMBO J.* **41**, e108443.
- 698 25. Sang, D., Shu, T., Pantoja, C. F., Ibáñez de Opakua, A., Zweckstetter, M. & Holt, L. J.
699 Condensed-phase signaling can expand kinase specificity and respond to macromolecular crowding.
700 *Molecular Cell* **82**, 3693–3711.e10 (2022).
- 701 26. Buratti, E. TDP-43 post-translational modifications in health and disease. *Expert Opinion on
702 Therapeutic Targets* **22**. PMID: 29431050, 279–293 (2018).
- 703 27. Nonaka, T. *et al.* Phosphorylation of TAR DNA-binding Protein of 43 kDa (TDP-43) by
704 truncated Casein Kinase 1 δ triggers mislocalization and accumulation of TDP-43. *Journal of
705 Biological Chemistry* **291**, 5473–5483 (2016).
- 706 28. Graves, P. R. & Roach, P. J. Role of COOH-terminal Phosphorylation in the Regulation of
707 Casein Kinase I δ . *Journal of Biological Chemistry* **270**, 21689–21694 (1995).
- 708 29. Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using
709 the folding funnel: The fly-casting mechanism. *Proceedings of the National Academy of Sciences*
710 **97**, 8868–8873 (2000).
- 711 30. Tejedor, A. R., Garaizar, A., Ramírez, J. & Espinosa, J. R. ‘RNA modulation of transport
712 properties and stability in phase-separated condensates. *Biophysical Journal* **120**, 5169–5186
713 (2021).
- 714 31. Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B. & Mittal, J. Sequence determinants of
715 protein phase behavior from a coarse-grained model. *PLoS Comput. Biol.* **14**, e1005941 (2018).
- 716 32. Tesei, G. & Lindorff-Larsen, K. Improved predictions of phase behaviour of intrinsically dis-
717 ordered proteins by tuning the interaction range [version 2; peer review: 2 approved]. *Open
718 Research Europe* **2** (2023).
- 719 33. Tesei, G. *et al.* Conformational ensembles of the human intrinsically disordered proteome.
720 *Nature* **626**, 897–904 (2024).
- 721 34. Joseph, J. A. *et al.* Physics-driven coarse-grained model for biomolecular phase separation
722 with near-quantitative accuracy. *Nat Comput Sci.* **1**, 732–743 (2021).
- 723 35. Rekhi, S. *et al.* Expanding the molecular language of protein liquid–liquid phase separation.
724 *Nature Chemistry* (2024).
- 725 36. Mukherjee, S. & Schäfer, L. V. Thermodynamic forces from protein and water govern con-
726 densate formation of an intrinsically disordered protein domain. *Nature Communications* **14**,
727 5892 (2023).
- 728 37. Knoch, F. & Speck, T. Cycle representatives for the coarse-graining of systems driven into a
729 non-equilibrium steady state. *New Journal of Physics* **17**, 115004 (2015).
- 730 38. Knoch, F. & Speck, T. Nonequilibrium Markov state modeling of the globule-stretch transition.
731 *Phys. Rev. E* **95**, 012503 (1 2017).
- 732 39. Albaugh, A. & Gingrich, T. R. Simulating a chemically fueled molecular motor with nonequi-
733 librium molecular dynamics. *Nature communications* **13**, 2204 (2022).
- 734 40. Vennelakanti, V., Nazemi, A., Mehmood, R., Steeves, A. H. & Kulik, H. J. Harder, better,
735 faster, stronger: Large-scale QM and QM/MM for predictive modeling in enzymes and proteins.
736 *Current Opinion in Structural Biology* **72**, 9–17 (2022).

- 737 41. Riedmiller, K. *et al.* Substituting density functional theory in reaction barrier calculations for
738 hydrogen atom transfer in proteins. *Chem. Sci.* **15**, 2518–2527 (7 2024).
- 739 42. Brasnett, C., Kiani, A., Sami, S., Otto, S. & Marrink, S. J. Capturing chemical reactions inside
740 biomolecular condensates with reactive Martini simulations. *Communications Chemistry* **7**,
741 151 (2024).
- 742 43. Girard, M. & Bereau, T. Computer simulations of lipid regulation by molecular semigrand
743 canonical ensembles. *Biophysical Journal* **120**, 2370–2373 (2021).
- 744 44. Kametani, F. *et al.* Mass spectrometric analysis of accumulated TDP-43 in amyotrophic lateral
745 sclerosis brains. *Scientific Reports* **6**, 23281 (2016).
- 746 45. Cracco, L. *et al.* Distinguishing post-translational modifications in dominantly inherited fronto-
747 temporal dementias: FTLD-TDP Type A (GRN) vs Type B (C9orf72). *Neuropathology and
748 applied neurobiology* **48**, e12836 (2022).
- 749 46. Hasegawa, M. *et al.* Phosphorylated TDP-43 in frontotemporal lobar degeneration and amy-
750 otrophic lateral sclerosis. *Annals of Neurology* **64**, 60–70 (2008).
- 751 47. Ko, V. I., Ong, K., Cleveland, D. W., Yu, H. & Ravits, J. M. CK1 δ/ϵ kinases regulate TDP-43
752 phosphorylation and are therapeutic targets for ALS-related TDP-43 hyperphosphorylation.
753 *Neurobiology of Disease* **196**, 106516 (2024).
- 754 48. Mardt, A., Pasquali, L., Wu, H. & Noé, F. VAMPnets for deep learning of molecular kinetics.
755 *Nature communications* **9**, 5 (2018).
- 756 49. Buchete, N.-V. & Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *The
757 Journal of Physical Chemistry B* **112**. PMID: 18232681, 6057–6069 (2008).
- 758 50. Trendelkamp-Schroer, B., Wu, H., Paul, F. & Noé, F. Estimation and uncertainty of reversible
759 Markov models. *The Journal of Chemical Physics* **143**, 174101 (2015).
- 760 51. Hoffmann, M. *et al.* Deeptime: a Python library for machine learning dynamical models from
761 time series data. *Machine Learning: Science and Technology* (2021).
- 762 52. Prinz, J.-H. *et al.* Markov models of molecular kinetics: Generation and validation. *The Journal
763 of Chemical Physics* **134**, 174105 (2011).
- 764 53. Wu, H. & Noé, F. Variational approach for learning Markov processes from time series data.
765 *Journal of Nonlinear Science* **30**, 23–66 (2020).
- 766 54. Wang, J. *et al.* A molecular grammar governing the driving forces for phase separation of
767 prion-like RNA binding proteins. *Cell* **174**, 688–699 (2018).
- 768 55. Ensign, D. L. & Pande, V. S. Bayesian Single-Exponential Kinetics in Single-Molecule Experiments
769 and Simulations. *The Journal of Physical Chemistry B* **113**, 12410–12423 (2009).
- 770 56. Kõivomägi, M. *et al.* Multisite phosphorylation networks as signal processors for Cdk1. *Nature
771 structural & molecular biology* **20**, 1415–1424 (2013).
- 772 57. Liu, B., Xue, M., Qiu, Y., Konovalov, K. A., O'Connor, M. S. & Huang, X. GraphVAMPnets
773 for uncovering slow collective variables of self-assembly dynamics. *The Journal of Chemical
774 Physics* **159**, 094901 (2023).
- 775 58. Ghorbani, M., Prasad, S., Klauda, J. B. & Brooks, B. R. GraphVAMPNet, using graph
776 neural networks and variational approach to Markov processes for dynamical modeling of
777 biomolecules. *The Journal of Chemical Physics* **156**, 184103 (2022).

- 778 59. Zhang, Y., Li, S., Gong, X. & Chen, J. Toward Accurate Simulation of Coupling between
779 Protein Secondary Structure and Phase Separation. *Journal of the American Chemical Society*
780 **146**. PMID: 38112495, 342–357 (2024).
- 781 60. Das, S., Lin, Y.-H., Vernon, R. M., Forman-Kay, J. D. & Chan, H. S. Comparative roles of
782 charge, π , and hydrophobic interactions in sequence-dependent phase separation of intrin-
783 sically disordered proteins. *Proceedings of the National Academy of Sciences* **117**, 28795–28805
784 (2020).
- 785 61. Welsh, T. J. *et al.* Surface Electrostatics Govern the Emulsion Stability of Biomolecular Con-
786 densates. *Nano Letters* **22**. PMID: 35001622, 612–621 (2022).
- 787 62. Krainer, G. *et al.* Reentrant liquid condensate phase of proteins is stabilized by hydrophobic
788 and non-ionic interactions. *Biophysical Journal* **120**, 28a (2021).
- 789 63. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. *A Density-Based Algorithm for Discovering*
790 *Clusters in Large Spatial Databases with Noise* in *Knowledge Discovery and Data Mining*
791 (1996).
- 792 64. Perdikari, T. M., Jovic, N., Dignon, G. L., Kim, Y. C., Fawzi, N. L. & Mittal, J. A predictive
793 coarse-grained model for position-specific effects of post-translational modifications. *Biophys.
J.* **120**, 1187–1197 (2021).

795 Supporting information

796 Residue-level coarse-grained models with implicit solvent

797 For our coarse-grained simulations we employed the hydrophobicity scale (HPS) model [31] and a
 798 modification of it [30]. The original HPS model was fitted with IDPs data and considers proteins
 799 as fully flexible chains. In order to have a more realistic representation of the enzyme CK1 δ ,
 800 we decided to employ also a modified version of it that takes into account the presence of folded
 801 domains. In both models every residue type is represented with a particle of Lennard-Jones (LJ)
 802 size σ , charge q , mass m and hydropathy scale parameter λ . For the HPS model, the pair potential
 803 has 3 contributions

$$U_{HPS}^{ij} = U_{bond}^{ij} + U_{electrostatic}^{ij} + U_{hydrophobicity}^{ij}, \quad (S1)$$

804 while the modified HPS model has one more contribution to enhance cation-pi interactions:

$$U_{HPS}^{ij} = U_{bond}^{ij} + U_{electrostatic}^{ij} + U_{hydrophobicity}^{ij} + U_{cation-\pi}^{ij}. \quad (S2)$$

805 The bonded interactions are described by an harmonic potential

$$U_{bond}^{ij} = \frac{1}{2}k(r_{ij} - r_0)^2 \quad (S3)$$

806 with r_{ij} the distance between the neighboring residues i and j , spring constant $k = 8360\text{kJ}/(\text{molnm}^2)$
 807 and equilibrium bond length $r_0 = 0.381\text{ nm}$.

808 The interactions between non-bonded residues are modeled through the Ashbaugh-Hatch po-
 809 tential

$$U_{hydrophobicity}^{ij} = \begin{cases} U_{LJ}^{ij} + \epsilon(1 - \lambda_{ij}), & \text{if } r_{ij} \leq 2^{1/6}\sigma_{ij} \\ \lambda_{ij}U_{LJ}^{ij}, & \text{if } r_{ij} > 2^{1/6}\sigma_{ij} \end{cases} \quad (S4)$$

810 where $\sigma_{ij} = (\sigma_i + \sigma_j)/2$, $\lambda_{ij} = (\lambda_i + \lambda_j)/2$ and U_{LJ}^{ij} is the standard Lennard-Jones potential

$$U_{LJ}^{ij} = 4\epsilon \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]. \quad (S5)$$

811 The interaction is truncated at a cutoff distance of 2 nm. The parameter ϵ expresses the strength of
 812 the Lennard-Jones interaction and it is fixed to $\epsilon = 0.8368\text{ kJ/mol}$, value fitted with experimental
 813 R_g from single IDP chains [31], while the hydropathy scale parameter λ_{ij} scales down the inter-
 814 action for distances larger than the minimum of U_{LJ}^{ij} and goes from 0 (fully hydrophilic case, no
 815 attraction between residues) to 1 (fully hydrophobic case, the interaction becomes a standard LJ).
 816 Phosphorylated Ser residues are modelled as described by Perdikari et al [64].

817 Charged residues experience also salt-screened electrostatic interactions, which are modeled
 818 using a Yukawa/Debye-Hückel potential

$$U_{electrostatic} = \frac{q_i q_j e^2}{4\pi\epsilon_0\epsilon_r r} \exp(-r/r_D) \quad (S6)$$

819 where we used a Debye screening length $r_D = 1.0\text{nm}$ for an ionic strength of approximately 100
 820 mM and a relative dielectric constant of the water solvent $\epsilon_r = 80$, following the ones of the original
 821 HPS model [31]. In this case the cutoff distance is 3.5 nm.

822 For the modified HPS model, we added another LJ potential only between cation- π pairs
 823 (Arg/Lys with Phe/Trp/Tyr), as proposed by Das et al. [60]:

$$U_{cation-\pi}^{ij} = U_{LJ}^{ij}, \text{ with } \epsilon_{cation-\pi} = 3.138 \frac{kJ}{mol}. \quad (S7)$$

824 Also in this case the cutoff is 2 nm.

825 In the modified HPS model, the dynamics of folded domains follows the one of a rigid body.
 826 Moreover, the parameter λ in $U_{hydrophobicity}^{ij}$ and $\epsilon_{cation-\pi}$ in $U_{cation-\pi}^{ij}$ are scaled down by 30% for
 827 pair interactions involving residues of the folded domains, as suggested by Krainer et al. [62].

828 The Ashbaugh-Hatch pair potential for the non-bonded interactions is available at https://github.com/ezippo/ashbaugh_plugin as a HOOMD-blue plugin. The code used for the simulations is available at https://github.com/ezippo/hoomd3_phosphorylation.

831 Chemical potential difference in a phosphorylation cycle

832 The chemical reaction difference in a reaction in units of $k_B T$ is given by the logarithm of the product
 833 to substrate concentration ratio: In the phosphorylation-dephosphorylation cycle, the chemical
 834 reactions involved are the two following ones:



835 The chemical potential differences for the two reactions are

$$\Delta\mu_P = \Delta\mu_P^0 + k_B T \log \left(\frac{[\text{TDP-43}^*][\text{ADP}]}{[\text{TDP-43}][\text{ATP}]} \right) = k_B T \log \left(\frac{[\text{TDP-43}^*][\text{ADP}]}{[\text{TDP-43}][\text{ATP}]} \frac{[\text{TDP-43}]_{eq}[\text{ATP}]_{eq}}{[\text{TDP-43}^*]_{eq}[\text{ADP}]_{eq}} \right) \quad (S8)$$

$$\Delta\mu_D = \Delta\mu_D^0 + k_B T \log \left(\frac{[\text{TDP-43}][\text{P}_i]}{[\text{TDP-43}^*]} \right) = k_B T \log \left(\frac{[\text{TDP-43}][\text{P}_i]}{[\text{TDP-43}^*]} \frac{[\text{TDP-43}]_{eq}}{[\text{TDP-43}^*]_{eq}[\text{P}_i]_{eq}} \right) \quad (S9)$$

837 The total amount of chemical driving will be:

$$\begin{aligned} \Delta\mu_{\circlearrowleft} &= \Delta\mu_P + \Delta\mu_D = \\ &= k_B T \log \left(\frac{[\text{TDP-43}^*][\text{ADP}]}{[\text{TDP-43}][\text{ATP}]} \frac{[\text{TDP-43}]_{eq}[\text{ATP}]_{eq}}{[\text{TDP-43}^*]_{eq}[\text{ADP}]_{eq}} \frac{[\text{TDP-43}][\text{P}_i]}{[\text{TDP-43}^*]} \frac{[\text{TDP-43}]_{eq}}{[\text{TDP-43}^*]_{eq}[\text{P}_i]_{eq}} \right) = \\ &= k_B T \log \left(\frac{[\text{ADP}][\text{P}_i]}{[\text{ATP}]} \frac{[\text{ATP}]_{eq}}{[\text{ADP}]_{eq}[\text{P}_i]_{eq}} \right) = \Delta\mu_{ATP}^0 + k_B T \log \left(\frac{[\text{ADP}][\text{P}_i]}{[\text{ATP}]} \right) \end{aligned} \quad (S10)$$

838 Note that $\Delta\mu_{\circlearrowleft}$ is equal to the chemical potential difference for the ATP hydrolysis reaction $\text{ATP} \rightleftharpoons$
 839 $\text{ADP} + \text{P}_i$, for which the equilibrium value in standard conditions is $\Delta\mu_{ATP}^0 = -7.3 \text{ kcal/mol}$.

840 Detailed balance breaking and local detailed balance

841 In our NESS simulations of a 4-state MSM, we expect to have detailed balance for the transitions
 842 $1 \rightleftharpoons 2$ and $3 \rightleftharpoons 4$, i.e. the binding/unbinding of the enzyme with TDP-43 or phosphorylated TDP-
 843 43, since they are determined by equilibrium MD simulations, but also $4 \rightleftharpoons 1$, i.e. the reservoir
 844 exchange step, that is determined by a Metropolis step without chemical fuel. The detail balance
 845 condition is:

$$p_i T_{ij} = p_j T_{ji} \quad (S11)$$

where T_{ij} is the transition probability from state i to j and p_i is the stationary probability of being in state i . Instead the phosphorylation reaction $2 \rightleftharpoons 3$ breaks detailed balance injecting into the system an amount of energy equal to $\Delta\mu_P$. Despite the Metropolis step is built in such a way to satisfy the detailed balance condition, the $\Delta\mu_P$ added in the acceptance ratio breaks detailed balance once the algorithm is coupled to equilibrium MD simulations. In order to show this, let us call ‘A’ a microstate configuration in which we have a contact between the Ser of TDP-43 and the active site of CK1 δ and ‘B’ the same microstate configuration but soon after the phosphorylation step, with pSer instead of Ser. The probability of being in the microstate ‘A’ or ‘B’ is $p_A = \exp(-\beta\mathcal{H}_A)/Z$ and $p_B = \exp(-\beta\mathcal{H}_B)/Z$, with $\beta = 1/k_B T$ and Z the canonical partition function, and they are sampled through the MD simulation. Since the velocities in ‘A’ and ‘B’ are the same, the probability ratio will be $p_B/p_A = \exp(-\beta\Delta U_P)$.

The Metropolis acceptance ratio contains also an additional $\Delta\mu_P$, leading to the following transition probability for the phosphorylation step:

$$T_{AB} = \begin{cases} 1, & \text{if } p'_B/p'_A \geq 1 \\ \frac{p'_B}{p'_A}, & \text{if } p'_B/p'_A < 1 \end{cases} \quad (\text{S12})$$

where $p'_B/p'_A = p_B/p_A \exp(-\beta\Delta\mu_P)$. Thus detailed balance is broken for $\Delta\mu_P \neq 0$ and we get:

$$\frac{T_{AB}p_A}{T_{BA}p_B} = \frac{p'_B p_A}{p'_A p_B} = e^{-\beta\Delta\mu_P} \quad (\text{S13})$$

On the other side, if our system is a NESS, i.e. T_{ij} and p_i are constant in time, for the local detailed balance we should have

$$\Delta\mu_{\text{cycle}} = -\frac{1}{\beta} \ln \left(\frac{T_{12}T_{23}T_{34}T_{41}}{T_{14}T_{43}T_{32}T_{21}} \right) \quad (\text{S14})$$

Using Eq. S11 for the couples $(i, j) = (1, 2), (3, 4), (4, 1)$, we can simplify Eq. S14 as

$$\Delta\mu_{\text{cycle}} = -\frac{1}{\beta} \ln \left(\frac{p_2 T_{23} p_4 T_{41}}{p_1 T_{14} p_3 T_{32}} \right) = -\frac{1}{\beta} \ln \left(\frac{p_2 T_{23}}{p_3 T_{32}} \right) = \Delta\mu_P \quad (\text{S15})$$

where in the last step we use the local detailed balance condition on the phosphorylation step in Eq. S13.

Transition probabilities and transition rates

In many formulations, Eq. 3 (or Eq. S14) is expressed in terms of transition rates rather than transition probabilities. The two are related and in our case give the same result for the $\Delta\mu_{\text{cycle}}$. We estimated the time-independent transition probability $T_{ij}(\tau)$, namely the probability of having the system in state j at time $t + \tau$ given that it was in state i at time t (for every t), using the non-reversible Maximum Likelihood Estimator[50, 51]:

$$T_{ij}(\tau) = \frac{c_{ij}(\tau)}{\sum_k c_{ik}(\tau)} \quad (\text{S16})$$

where $c_{ij}(\tau)$ is the count of transitions from i to j after a lag time τ .

872 In principle, it could be useful in some cases, e.g. for continuous-time systems or non-linear
 873 reaction networks, to express Eq. 3 with transition rates k_{ij} . For a Markov process, the transition
 874 probability matrix $\mathbf{T}(\tau)$ can be expressed in terms of transition rates matrix \mathbf{k} as:

$$\mathbf{T} = \exp(\mathbf{k}\tau). \quad (\text{S17})$$

875 For small τ compared to the system timescales, Eq. S17 can be approximated as

$$\mathbf{T} \simeq \mathbb{1} + \tau\mathbf{k} \Rightarrow k_{ij} \simeq \frac{T_{ij}}{\tau}, \quad \forall i, j \text{ such that } i \neq j. \quad (\text{S18})$$

876 In such case, using T_{ij} or k_{ij} to compute $\Delta\mu_{\text{cycle}}$ does not change the result, since the factor $1/\tau$
 877 would be cancelled out in the ratio in Eq. 3 (or Eq. S14).

878 As an example, for the simulations with reactive Ser 403 and $\Delta\mu_P = -5$ kJ/mol, we computed
 879 T_{ij} with lag time $\tau = 10$ Markov chain steps (or 1 ns in simulation time). Discretizing the trajectory
 880 into 3 Markov states (state 1 with unbound CK1δ and TDP-43, state 2 with bound configuration
 881 and unphosphorylated Ser 403 and state 3 with bound configuration and phosphorylated Ser 403)
 882 leads to 2 implied timescales that are much larger than the the lag time τ (Fig. 2, SI Fig. S1). The
 883 estimated transition probability matrix is

$$\mathbf{T} = \begin{bmatrix} 0.97706502 & 0.01522616 & 0.00770882 \\ 0.03748648 & 0.92563442 & 0.0368791 \\ 0.02103749 & 0.00579359 & 0.97316892 \end{bmatrix} \quad (\text{S19})$$

884 while the transition rates matrix is

$$\mathbf{k} = \begin{bmatrix} -0.02293498 \text{ ns}^{-1} & 0.01522616 \text{ ns}^{-1} & 0.00770882 \text{ ns}^{-1} \\ 0.03748648 \text{ ns}^{-1} & -0.07436558 \text{ ns}^{-1} & 0.0368791 \text{ ns}^{-1} \\ 0.02103749 \text{ ns}^{-1} & 0.00579359 \text{ ns}^{-1} & -0.02683108 \text{ ns}^{-1} \end{bmatrix} \quad (\text{S20})$$

885 where for the diagonal elements we used the property of the rate matrices $k_{ii} = -\sum_{j \neq i} k_{ij}$.

886 **Thermodynamic consistency data**

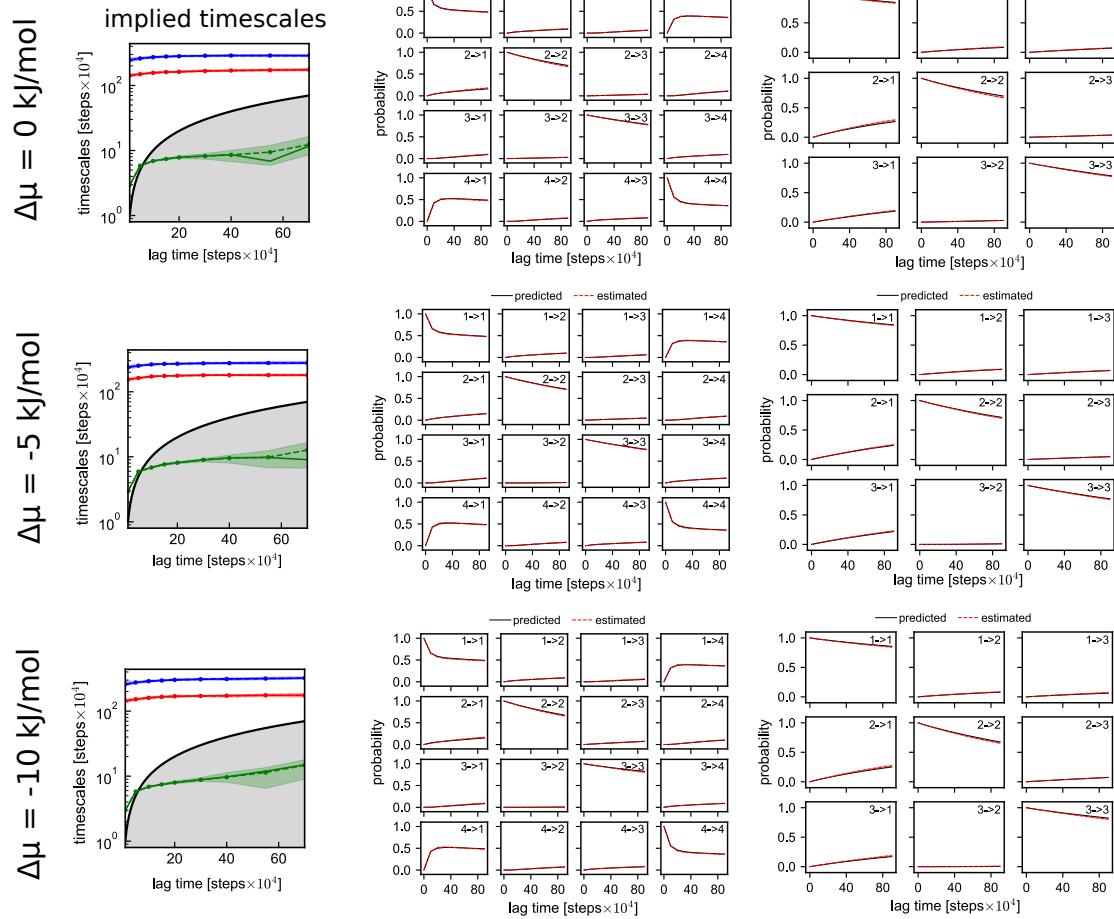
Ser	$-\Delta\mu_P$ [kJ/mol]	$-\Delta\mu_{cycle}$ [kJ/mol]	$-\Delta\mu_{cycle,3s}$ [kJ/mol]
Ser292	0	0.1 ± 0.6	0.1 ± 0.6
	5	4.6 ± 0.9	4.6 ± 0.9
	10	10.3 ± 0.8	10.3 ± 0.9
Ser317	0	-0.4 ± 0.5	-0.5 ± 0.5
	5	5.0 ± 0.7	4.9 ± 0.7
	10	10.3 ± 0.6	10.1 ± 0.8
Ser369	0	0.5 ± 0.7	0.5 ± 0.6
	5	5.0 ± 0.6	5.1 ± 0.5
	10	10.3 ± 0.7	10.2 ± 0.5
Ser387	0	0.0 ± 0.6	0.1 ± 0.4
	5	5.3 ± 0.6	5.2 ± 0.5
	10	9.3 ± 0.7	9.1 ± 0.7
Ser403	0	0.51 ± 0.26	0.52 ± 0.27
	5	4.87 ± 0.27	4.9 ± 0.3
	10	9.7 ± 1.0	9.5 ± 0.7
Ser409	0	0.0 ± 0.8	0.1 ± 0.7
	5	4.1 ± 0.7	4.1 ± 0.5
	10	10.3 ± 0.8	10.2 ± 0.8

Table S1: Estimated $\Delta\mu_{cycle}$ reported in Fig. 2E of the main text.

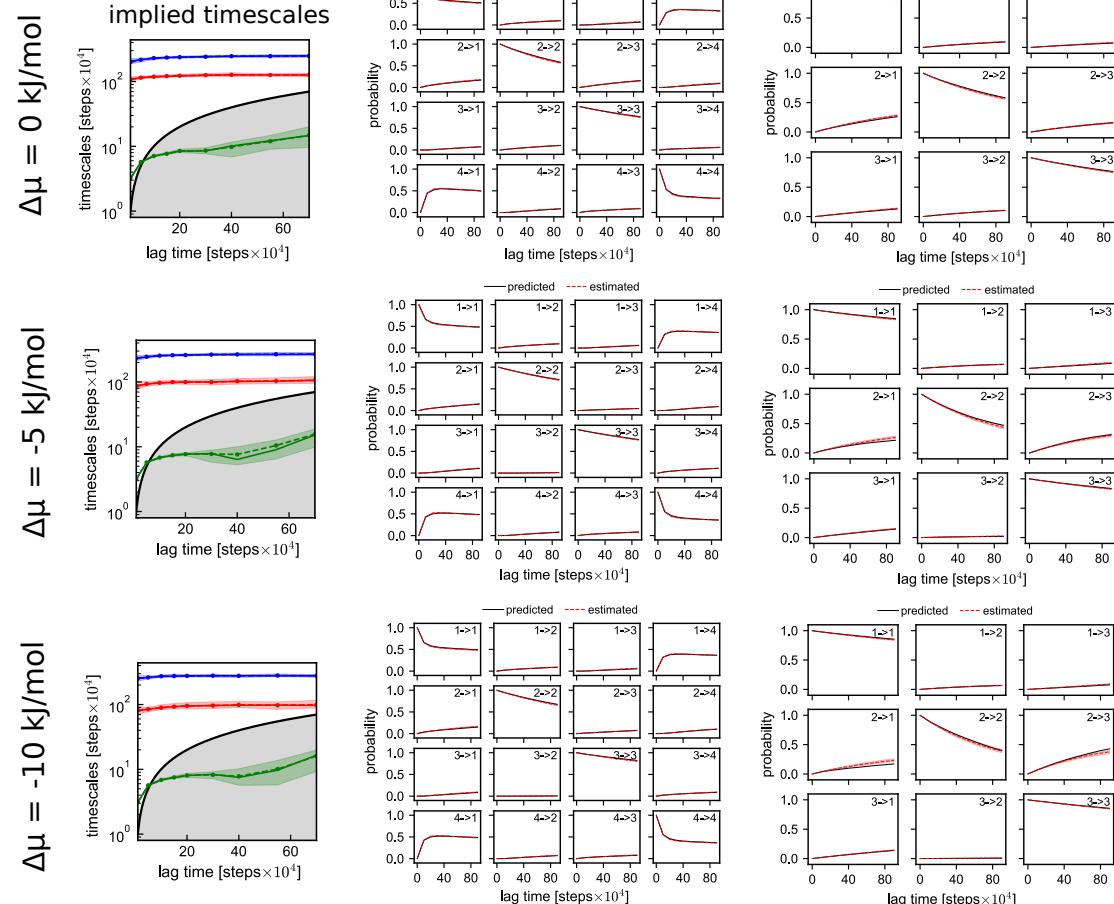
Ser	τ [steps]	$-\Delta\mu_P$ [kJ/mol]	$-\Delta\mu_{cycle}$ [kJ/mol]
Ser403	1	0	0.45 ± 0.23
	10	0	0.51 ± 0.26
	20	0	0.57 ± 0.27
Ser403	1	-5	4.29 ± 0.21
	10	-5	4.86 ± 0.27
	20	-5	4.99 ± 0.34

Table S2: Estimated $\Delta\mu_{cycle}$ for Ser 403 and $\Delta\mu_P = 0, -5$ kJ/mol for different lag times. The results for $\Delta\mu_{cycle}$ are in agreement with $\Delta\mu_P$ for $\tau \geq 10$ Markov chain steps.

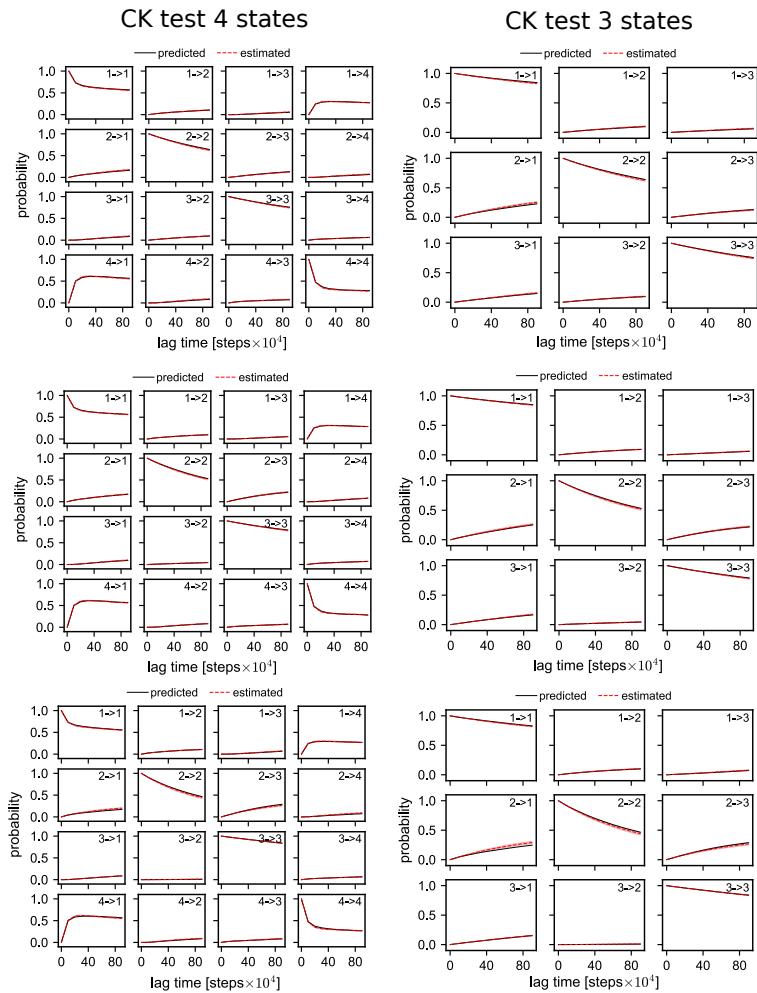
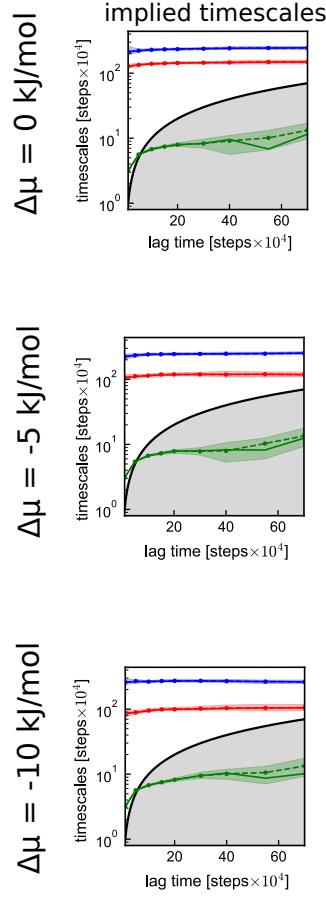
Ser292



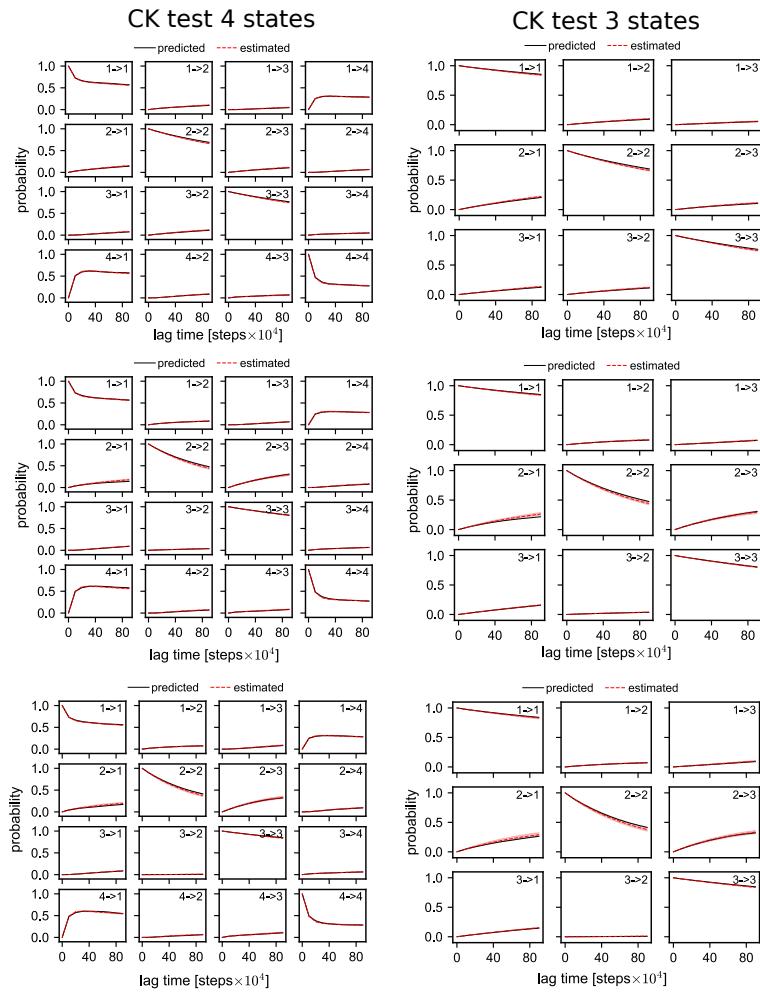
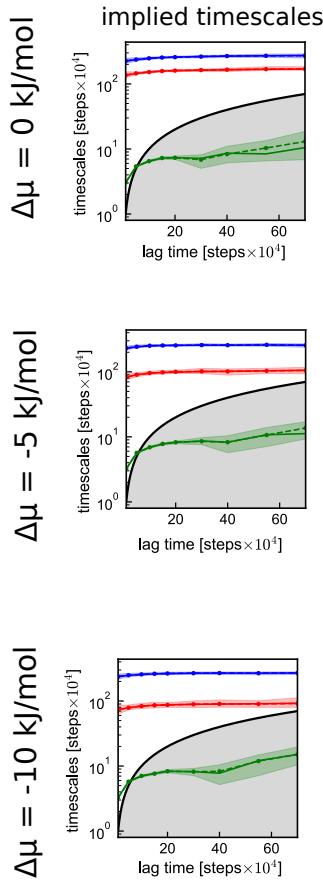
Ser317



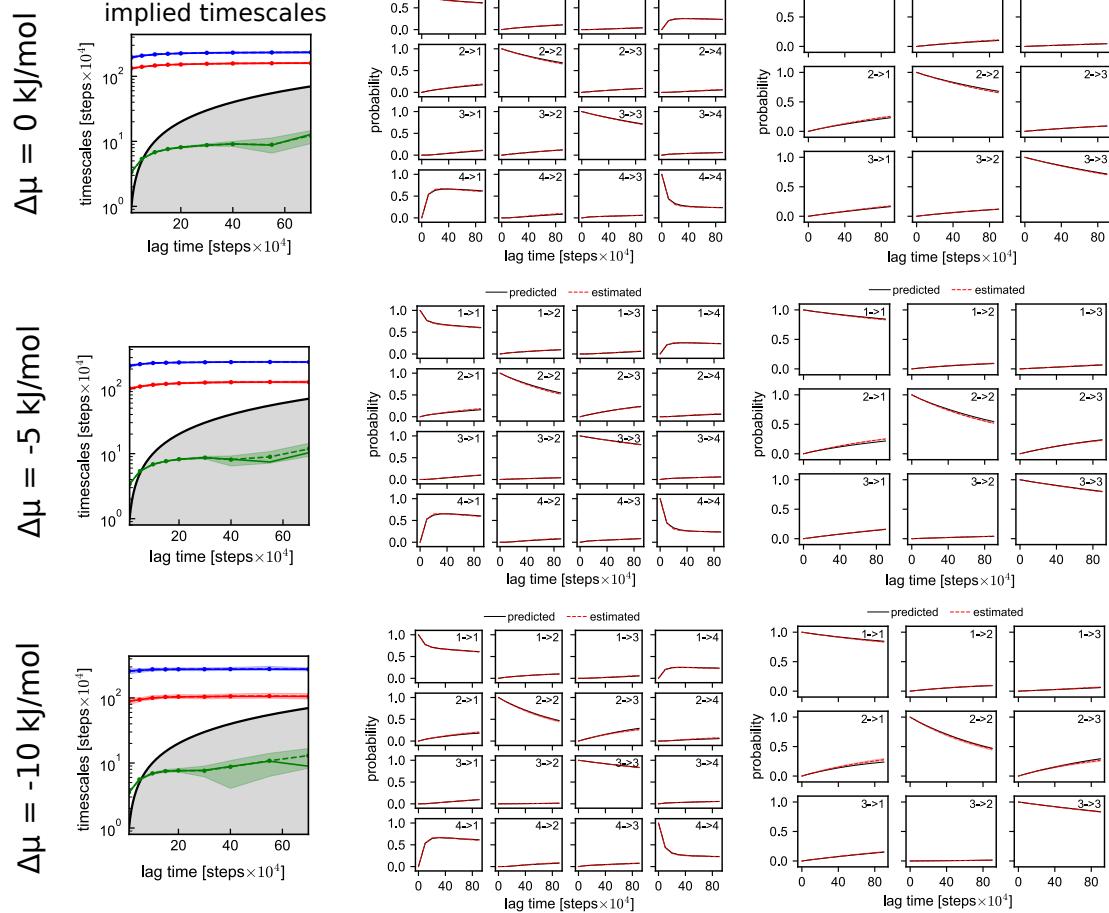
Ser369



Ser387



Ser403



Ser409

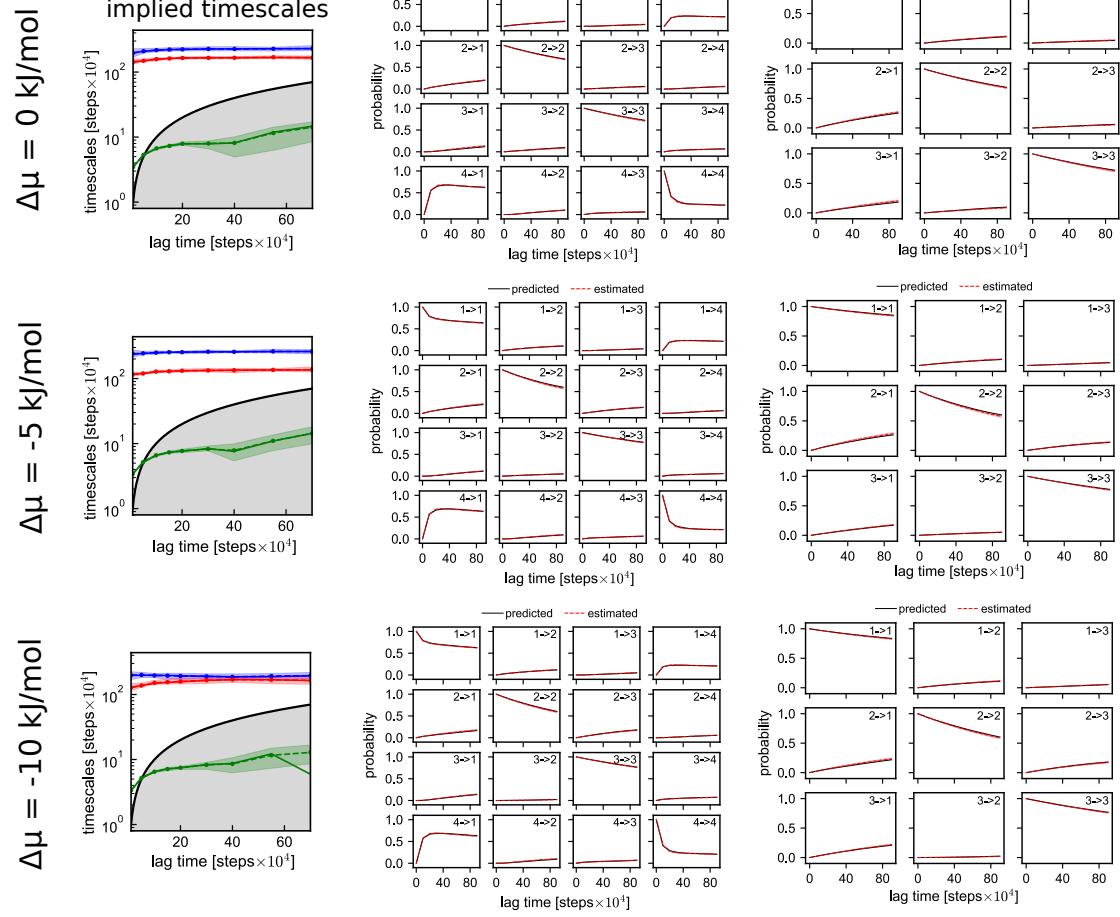


Figure S1: Implied timescales and CK test for every validation simulation. We estimated $\Delta\mu_{\text{cycle}}$ also with a 3-state MSM, merging together state 1 and 4 into the new state 1. We report the CK test also for the 3-state MSM case. The implied timescales are constant for $\tau \geq 10$ Markov chain steps. The smaller implied timescale is smaller than the lag time for $\tau \geq 10$ Markov chain steps. The CK test confirm that the MSM correctly fulfill the markovianity condition.

Ser403

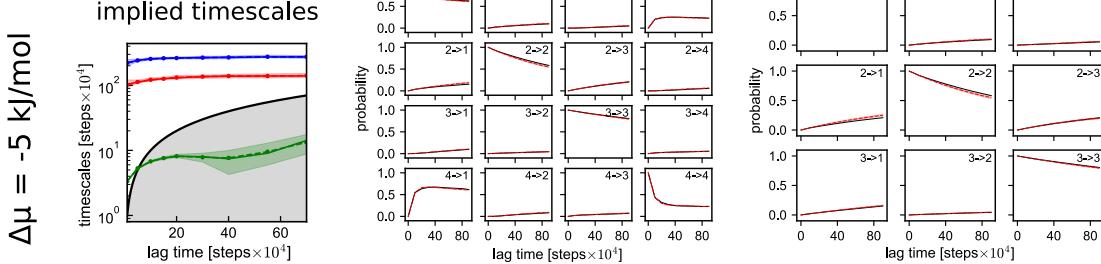


Figure S2: Implied timescales and CK test for Ser403 and $\Delta\mu_P = -5 \text{ kJ/mol}$ using a version of VAMPnet with 4260 input distances. We estimated $\Delta\mu_{\text{cycle}}$ also with a 3-state MSM, merging together state 1 and 4 into the new state 1. We report the CK test also for the 3-state MSM case.

887 Phosphorylation modifies interaction of CK1 δ with TDP-43 LCD

888 Binding free energy

889 We computed the binding free energy ΔG_{bind} between 1) CK1 δ folded-domain and wild type TDP-
 890 43 LCD, 2) full-length CK1 δ and wild type TDP-43 LCD, 3) CK1 δ folded-domain and triple phos-
 891 phorylated TDP-43 LCD (pSer 395, pSer 403, pSer 410). For the first 2 cases, we used the data
 892 from the equilibrium simulations without phosphorylation step (450 μs for case 1) and 900 μs for
 893 case 2), while for case 3) we collected 20 μs of simulation time.

894 The binding free energy is estimated as

$$\Delta G_{\text{bind}} = -\ln \left(\frac{T_{u,b}(\tau)}{T_{b,u}(\tau)} \right) \quad (\text{S21})$$

895 where $T_{u,b}(\tau)$ is the probability to have a bound state at time $t + \tau$, given an unbound state at time
 896 t , and $T_{b,u}(\tau)$ is the probability to have an unbound state at time $t + \tau$, given a bound state at time
 897 t . In order to get the transition probabilities, we discretize the simulation trajectories in bound
 898 and unbound state using VAMPnet (in the same way as in Fig. 2, as explained in Methods) and
 899 estimate $T_{u,b}$ and $T_{b,u}$ at lag time $\tau = 5 \times 10^4$ MD steps using a Maximum Likelihood estimator
 900 for MSM [50].

901 Plots

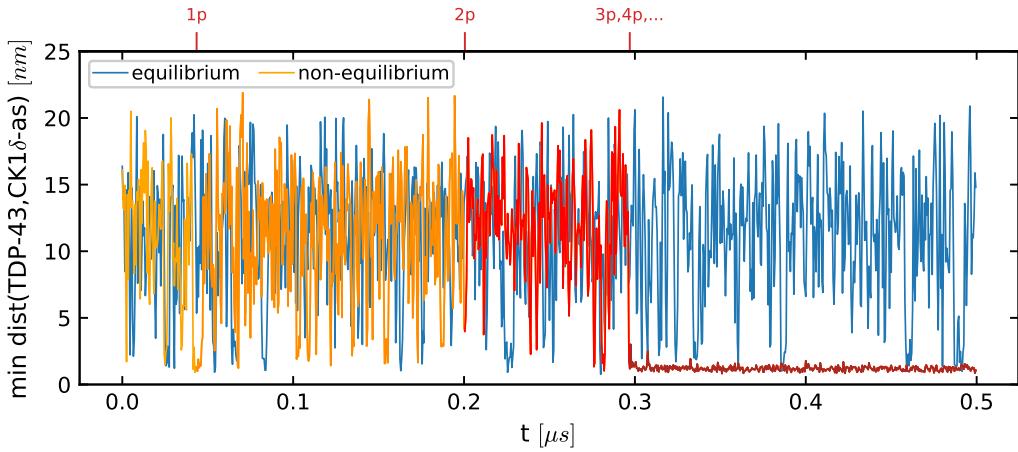


Figure S3: Example trajectory of minimum distance between residues of TDP-43 LCD and the active site of CK1 δ folded-domain in equilibrium simulation without phosphorylation (blue) and in non-equilibrium simulation (orange) in dilute concentration. The color of the non-equilibrium trajectory becomes darker after every phosphorylation event. In this example TDP-43 stays bound to the enzyme after 3 phosphorylations.

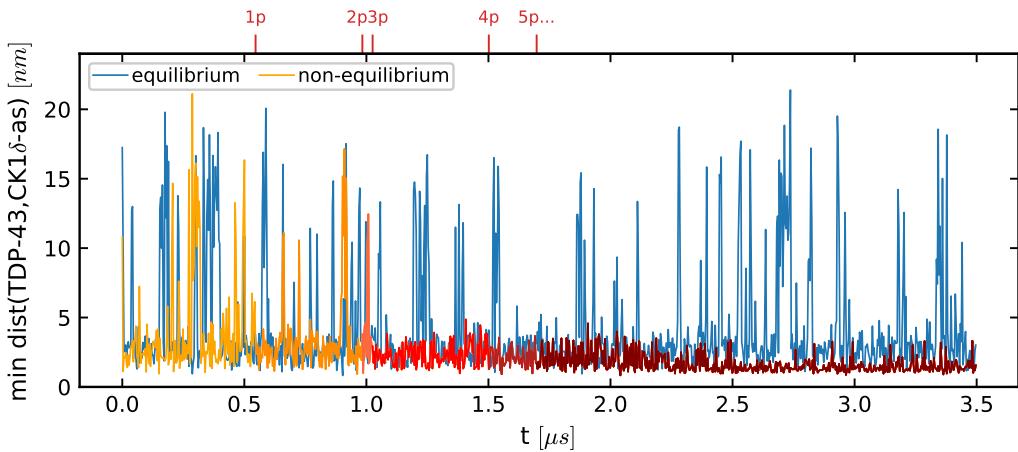


Figure S4: Example trajectory of minimum distance between residues of TDP-43 LCD and the active site of full-length CK1 δ in equilibrium simulation without phosphorylation (blue) and in non-equilibrium simulation (orange) in dilute concentration. The color of the non-equilibrium trajectory becomes darker after every phosphorylation event. In this example TDP-43 stays bound to the enzyme after 3 phosphorylations.

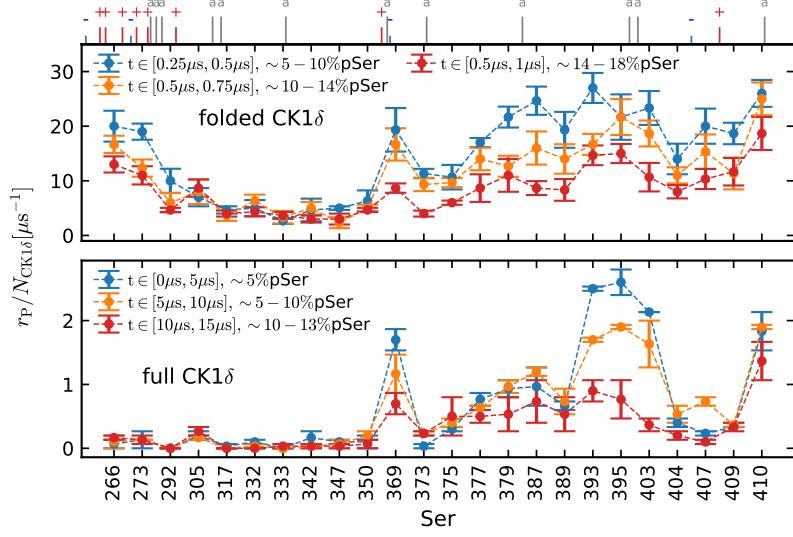


Figure S5: Phosphorylation rates r_P for every Ser of TDP-43 LCD in presence of 3 CK1 δ folded-domain (top panel) or 3 full-length CK1 δ (lower panel) in condensate for different parts of the trajectory. The ticks on top show the position of the charged and aromatic residues. Phosphorylation rates of the most phosphorylated Ser decrease with time.

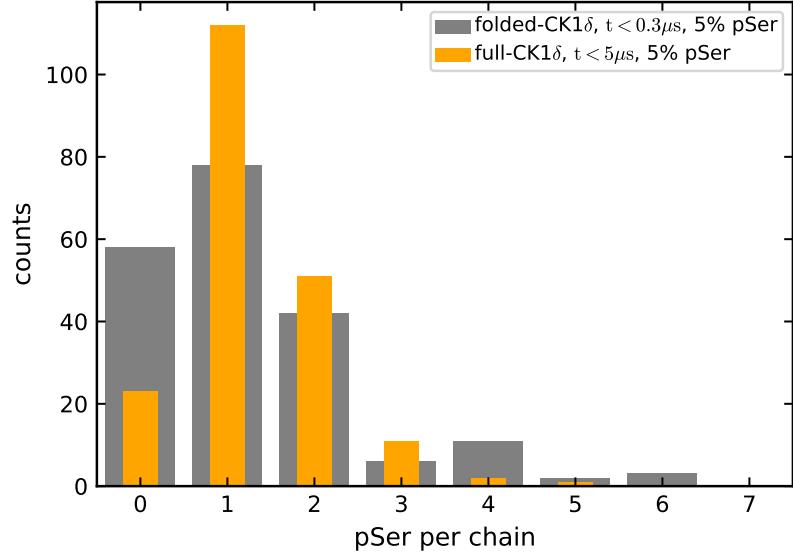


Figure S6: Histogram of the presence of phosphate per chain after $0.3 \mu s$ for the simulation of TDP-43 condensate with 3 CK1 δ folded-domain (grey) and after $5 \mu s$ for the simulation of TDP-43 condensate with 3 full-length CK1 δ (orange). In both cases about 5% of total number of Ser are phosphorylated at the time of the measurement.

₉₀₂ CK1 δ active sites and charges

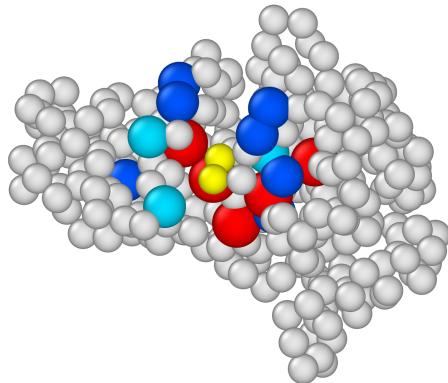


Figure S7: CK1 δ with colored charged residue close to the active site. In blue the +e charged residues, in light blue His residues (considered +0.5e in our simulations), in red -e charged residues. In yellow the active site residues.

₉₀₃ Phosphorylation acceptance

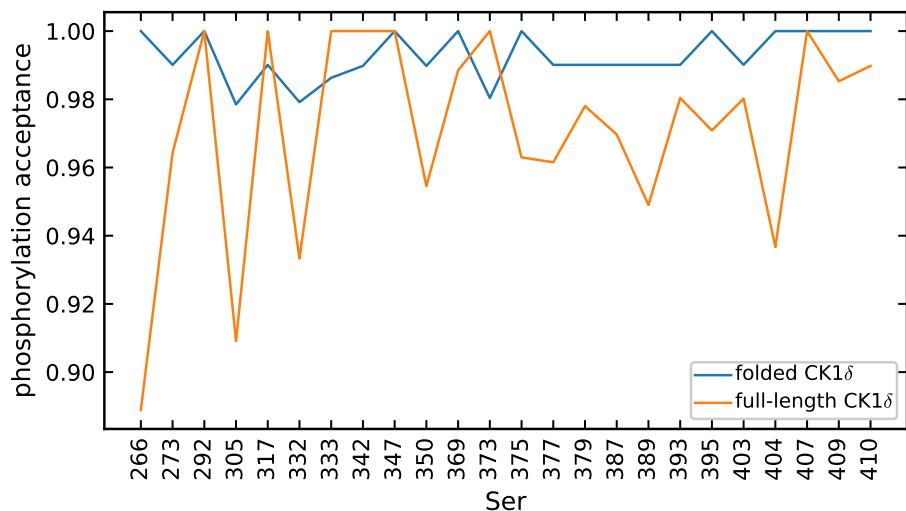
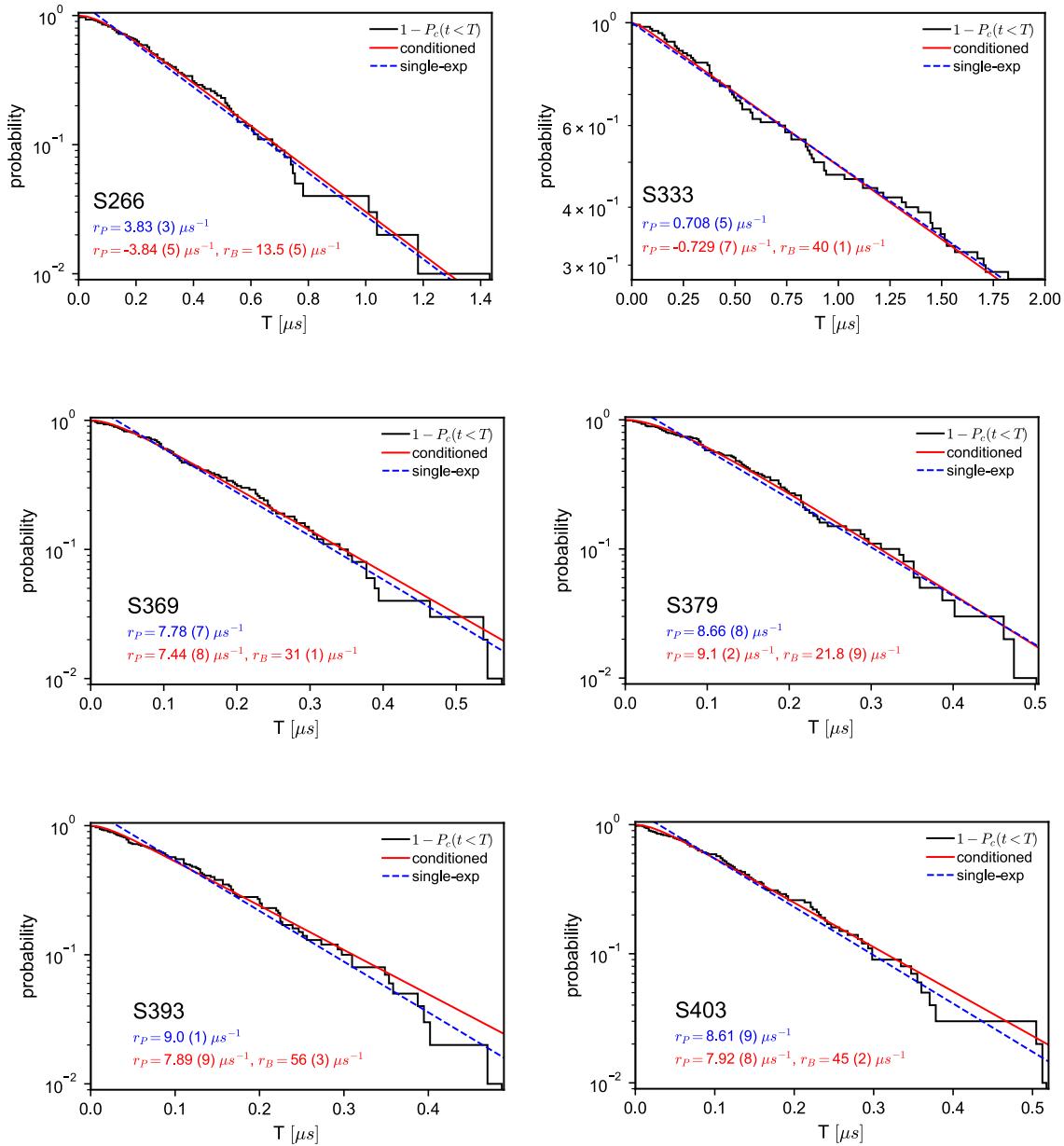


Figure S8: Acceptance ratio for phosphorylation step for simulations with CK1 δ folded-domain (blue) and full-length CK1 δ (orange). The acceptance ratio is very close to 100% for every Ser of TDP-43 LCD.

904 Phosphorylation process fit



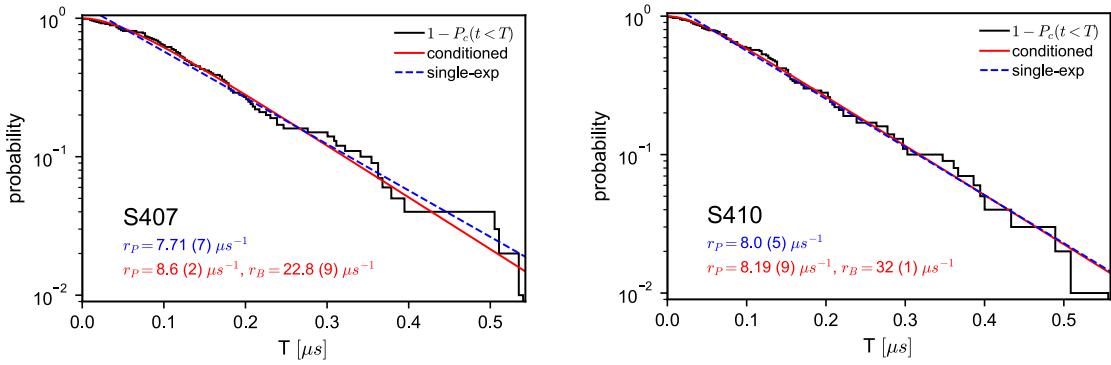


Figure S9: Normalized inverse cumulative histograms of phosphorylation times (black solid lines) and fit with simple single-exponential process (blue dotted lines, rate estimates in blue) and conditioned single-exponential process (red solid lines, rates estimates in red) for 8 different Ser residues. Most of the times the conditioned exponential process fits perfectly. The rate extrapolations from the two fits are in agreement with the bayesian estimates. The rate r_B is different for every serine.

905 Phosphorylation rank

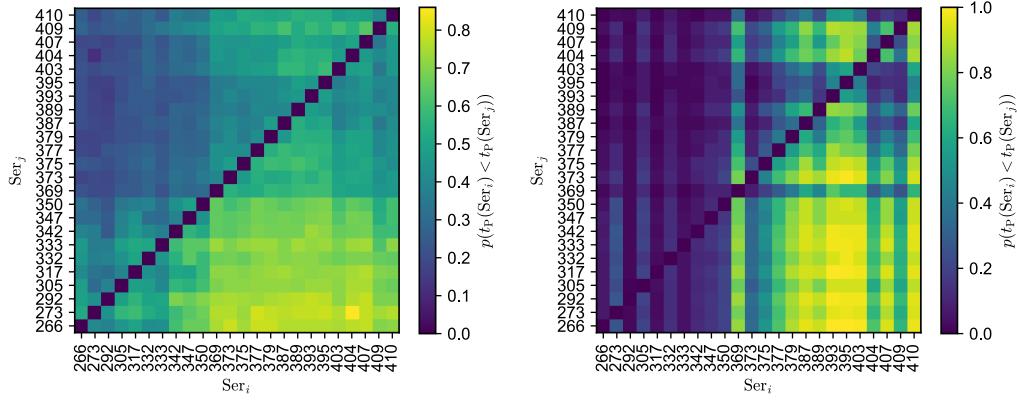
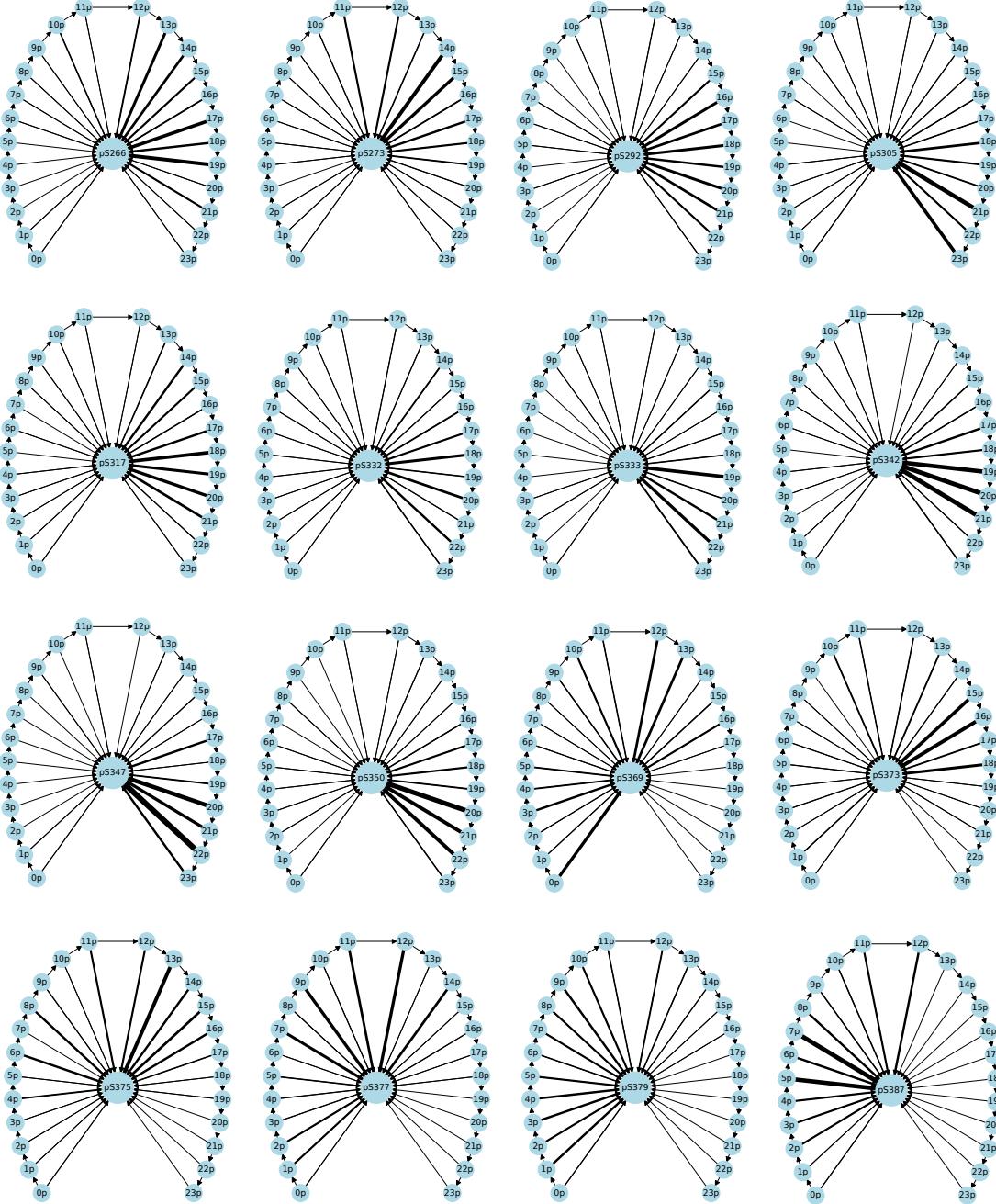


Figure S10: Probability $p(t(\text{Ser}_i) < t(\text{Ser}_j))$ of Ser_i being phosphorylated ahead of Ser_j for system with averaged-interaction chain and $\text{CK}1\delta$ folded-domain (left) and for the system with TDP-43 LCD wild type and full-length $\text{CK}1\delta$ (right), data from 100 trajectories.



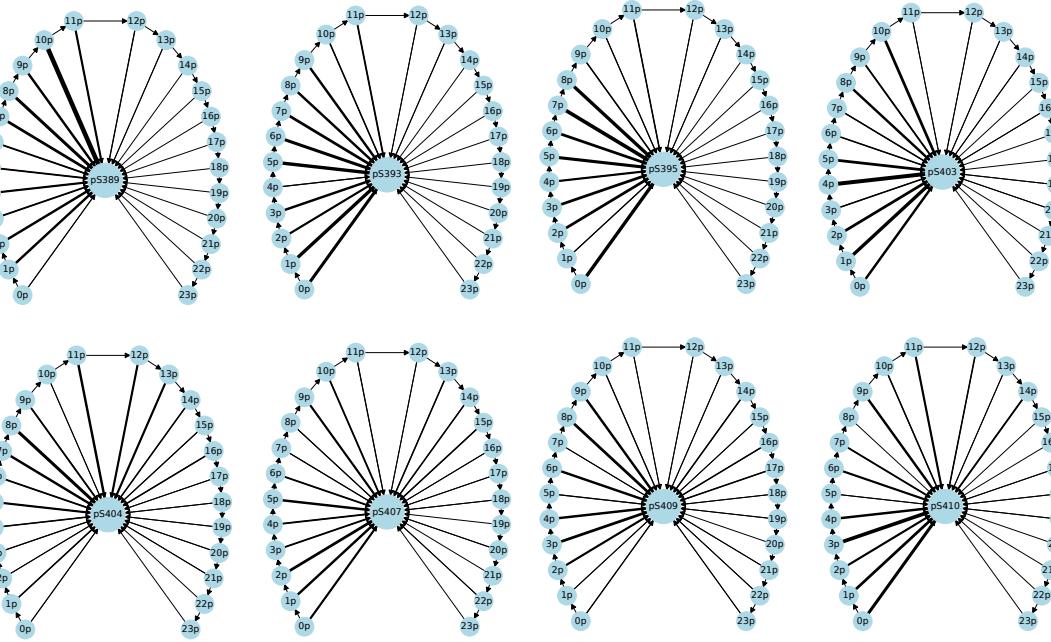


Figure S11: Phosphorylation pattern representation for every Ser of TDP-43 LCD. The thickness of the arrows represent the percentage of simulations in which the Ser in the center of the graph was phosphorylated after n other Ser residues.

906 **Correlation plots: r_P vs r_c in dilute regime for full-length CK1 δ**

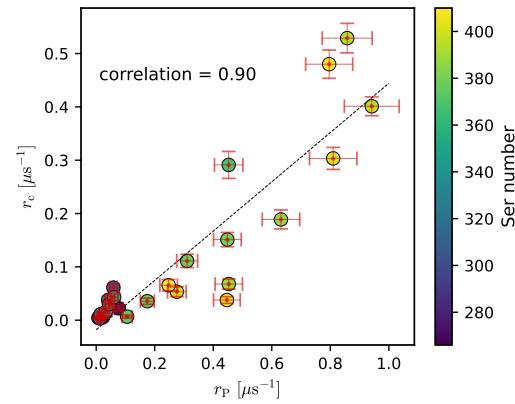


Figure S12: Correlation plots of contact frequency in equilibrium r_c and phosphorylation rates r_P in dilute regime for simulations with full-length CK1 δ .

907 Correlation plots: dilute vs dense regime phosphorylation rates

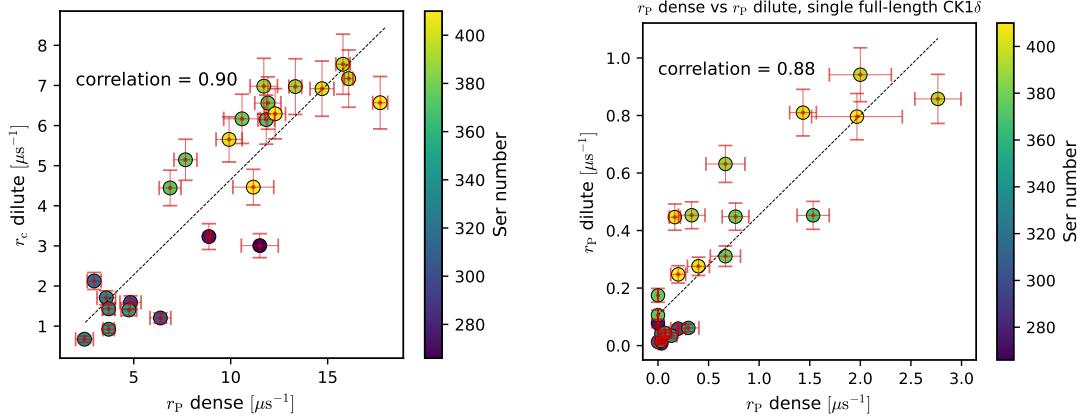


Figure S13: Correlation plots of phosphorylation rates r_P in dense (x-axis) and dilute (y-axis) regimes for simulations with one CK1 δ folded-domain (left) and one full-length CK1 δ (right).

908 Correlation plots: r_P vs r_c in condensate

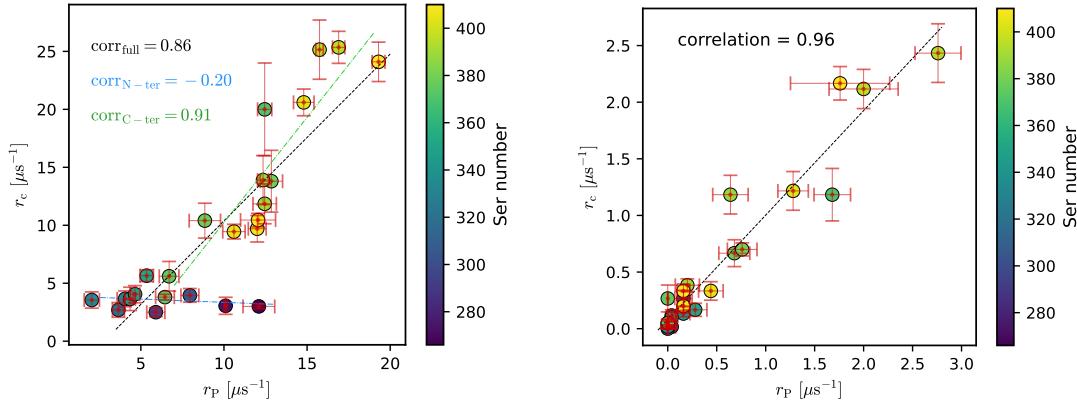


Figure S14: Correlation plots of contact frequency in equilibrium r_c and phosphorylation rates r_P in condensate for simulations with CK1 δ folded-domain (left) and full-length CK1 δ (right).

909 CK1 δ disordered domain do not cover the active site

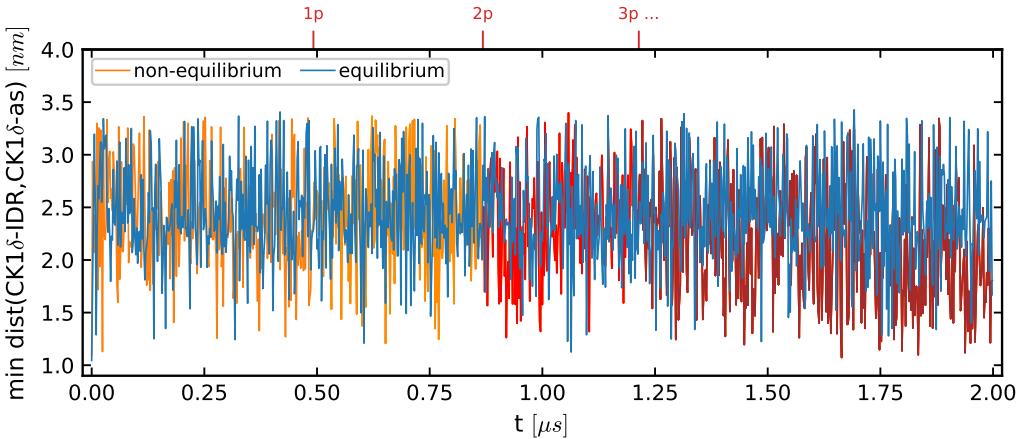


Figure S15: Example trajectory of minimum distance between residues of full-length CK1 δ IDR and its active site in equilibrium simulation without phosphorylation (blue) and in non-equilibrium simulation (orange) in dilute concentration. The color of the non-equilibrium trajectory becomes darker after every phosphorylation event. The distance never stays stable around the contact distance, but oscillates around larger values, suggesting that the CK1 δ IDR does not cover the active site for an extended time interval.

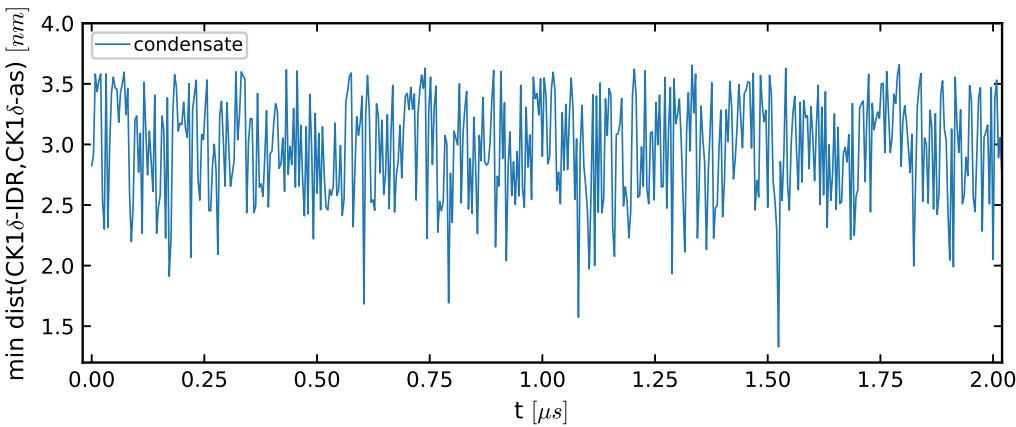


Figure S16: Example trajectory of minimum distance between residues of full-length CK1 δ IDR and its active site in condensate simulations. The distance never stays stable around the contact distance, but oscillates around larger values, suggesting that the CK1 δ IDR does not cover the active site for an extended time interval.

910 **Movies**

911 **SI Movie 1**

912 Movie from simulation of single TDP-43 LCD chain and single CK1 δ folded domain with phospho-
913 phosphorylation step (Eq. 2) and reservoir exchange step (Eq. 4) in cubic box of 50 nm side length
914 using HPS model. In this simulation, the only phosphosite is Ser403 and $\Delta\mu_P = -5$ kJ/mol. This
915 simulation was used in Fig. 2.

916 **SI Movie 2**

917 Movie from simulation of single TDP-43 LCD chain and single CK1 δ folded domain with phospho-
918 phosphorylation step in cubic box of 30 nm side length using modified HPS model. This simulation was
919 used in Fig. 3e.

920 **SI Movie 3**

921 Movie from simulation of 200 TDP-43 LCD chains and 3 CK1 δ folded-domain with phosphorylation
922 step in cubic box of 100 nm side length using modified HPS model.

923 **SI Movie 4**

924 Movie from simulation of single TDP-43 LCD chain and single full-length CK1 δ with phosphoryla-
925 tion step in cubic box of 30 nm side length using modified HPS model.

926 **SI Movie 5**

927 Movie from simulation of 200 TDP-43 LCD chains and 3 full-length CK1 δ with phosphorylation
928 step in cubic box of 100 nm side length using modified HPS model.