



# Partner-Mediated Polymorphism of an Intrinsically Disordered Protein

Christophe Bignon, Francesca Troilo, Stefano Gianni, Sonia Longhi

## ► To cite this version:

Christophe Bignon, Francesca Troilo, Stefano Gianni, Sonia Longhi. Partner-Mediated Polymorphism of an Intrinsically Disordered Protein. *Journal of Molecular Biology*, 2017, 10.1016/j.jmb.2017.11.012 . hal-01802951

**HAL Id: hal-01802951**

**<https://hal.science/hal-01802951>**

Submitted on 6 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## **Partner-mediated polymorphism of an intrinsically disordered protein**

Christophe Bignon<sup>1\*</sup>, Francesca Troilo<sup>1,2</sup>, Stefano Gianni<sup>2</sup>, and Sonia Longhi<sup>1\*</sup>

<sup>1</sup>Aix-Marseille Univ, CNRS, Architecture et Fonction des Macromolécules Biologiques (AFMB) UMR 7257, Marseille, France.

<sup>2</sup>Istituto Pasteur Italia - Fondazione Cenci Bolognetti, Istituto di Biologia e Patologia Molecolari del CNR, Dipartimento di Scienze Biochimiche “A. Rossi Fanelli”, Sapienza Università di Roma, 00185, Rome, Italy

\*To whom correspondence should be addressed.

Sonia LONGHI

CNRS and Aix-Marseille Université, AFMB, UMR 7257, Case 932

163, Avenue de Luminy, Case 932, 13288, Marseille, France

E-mail: Sonia.Longhi@afmb.univ-mrs.fr

Tel: (33) 4 91 82 55 80

Fax: (33) 4 91 26 67 20

Christophe BIGNON

CNRS and Aix-Marseille Université, AFMB, UMR 7257, Case 932

163, Avenue de Luminy, Case 932, 13288, Marseille, France

E-mail: Christophe.Bignon@afmb.univ-mrs.fr

Tel: (33) 4 91 82 55 80

Fax: (33) 4 91 26 67 20

## Abstract

Intrinsically disordered proteins (IDPs) recognize their partners through molecular recognition elements (MoREs). The MoRE of the C-terminal intrinsically disordered domain of the measles virus nucleoprotein (N<sub>TAIL</sub>) is partly pre-configured as an  $\alpha$ -helix in the free form and undergoes  $\alpha$ -helical folding upon binding to the X domain (XD) of the viral phosphoprotein. Beyond XD, N<sub>TAIL</sub> also binds the major inducible heat shock protein 70 (hsp70). So far, no structural information is available for the N<sub>TAIL</sub>/hsp70 complex. Using mutational studies combined with a protein complementation assay based on GFP reconstitution, we have investigated both N<sub>TAIL</sub>/XD and N<sub>TAIL</sub>/hsp70 interactions. Although the same N<sub>TAIL</sub> region binds the two partners, the binding mechanisms are different. Hsp70 binding is much more tolerant of MoRE substitutions than XD and the majority of substitutions lead to an increased N<sub>TAIL</sub>/hsp70 interaction strength. Furthermore, while an increased and a decreased  $\alpha$ -helicity of the MoRE lead to enhanced and reduced interaction strength with XD, respectively, the impact on hsp70 binding is negligible, suggesting that the MoRE does not adopt an  $\alpha$ -helical conformation once bound to hsp70. Here, by showing that the  $\alpha$ -helical conformation sampled by the free form of the MoRE does not systematically commit it to adopt an  $\alpha$ -helical conformation in the bound form, we provide an example of partner-mediated polymorphism of an IDP and of the relative insensitiveness of the bound structure to the pre-recognition state. The present results therefore contribute to shed light on the molecular mechanisms by which IDPs recognize different partners.

**Keywords.** Measles virus, N<sub>TAIL</sub>, XD, major inducible heat shock protein (hsp70), induced folding.

## Introduction

The measles virus (MeV) nucleoprotein consists of a structured region (N<sub>CORE</sub>, aa 1-400) [1] and a C-terminal domain (N<sub>TAIL</sub>, aa 401-525) that is intrinsically disordered [2, 3]. Intrinsically disordered proteins (IDPs) or regions (IDRs) are ubiquitous proteins/regions that lack highly populated secondary and tertiary structure in the absence of a partner or ligand [4-8].

As for many IDPs [9], including the homologous N<sub>TAIL</sub> domains from the cognate Henipaviruses [10-13], MeV N<sub>TAIL</sub> undergoes folding upon binding to the X domain (XD) of the viral phosphoprotein [14-16]. Interaction with the partner triggers  $\alpha$ -helical folding within a Molecular Recognition Element (MoRE, aa 486-502) located within one (Box2, aa 489-506) out of three conserved N<sub>TAIL</sub> regions (**Figure 1**) [3, 14, 17-24]. Binding of N<sub>TAIL</sub> to XD results in a “fuzzy” complex [25-28], *i.e.* the regions preceding and following the MoRE remain both conspicuously disordered [14, 19, 23, 29].

Although the MoRE is partly pre-structured in the free form [19, 21, 22, 24, 30, 31], computational and experimental studies revealed that MeV N<sub>TAIL</sub> folds according to a folding after binding mechanism [16, 31]. It should be reminded indeed that although conformational selection requires that the IDP populates (at least partly) the bound state in the free form, which therefore presages the final bound conformation [32, 33], the pre-existence of folded structures in an IDP does not necessarily commit to a conformational selection mechanism [31, 34, 35]. To further complicate the scenario, many IDPs display a high extent of plasticity with respect to the partner, *i.e.* they can fold into different structures upon binding to different partners (see [8] and

references therein cited). This polymorphism thus raises the question as to which is the role of preconfigured MoREs in the recognition process, and suggests that the conformation sampled in solution will not necessarily be favored upon binding to a partner [9, 36].

In line with previous reports that pointed out that structural disorder is a determinant of protein interactivity [37-39], beyond XD N<sub>TAIL</sub> binds to various viral and cellular proteins [40-48] including the major inducible heat shock protein 70 (hsp70) [49-51].

In a previous study, we have shown that progressively shortening the N<sub>TAIL</sub> fuzzy appendage from residue 401 to residue 481 results in a non-monotonic increase of its binding strength towards both XD and hsp70 [52]. This finding might reflect a similar binding mechanism in spite of the notable differences between XD and hsp70 in terms of size, fold and origin. Albeit the major hsp70-binding site has been already mapped to Box2 [50, 51], no structural information is available for the N<sub>TAIL</sub>/hsp70 complex. It therefore remains to be established whether the MoRE undergoes  $\alpha$ -helical folding also upon binding to hsp70 and whether the interaction relies on the same N<sub>TAIL</sub> residues. Deciphering the molecular mechanisms governing the N<sub>TAIL</sub>/hsp70 interaction is of broad relevance in light of its well-documented functional impact on viral transcription and replication [49, 50] and on the innate immune response [53]. In the present paper, we have addressed these questions using mutational studies combined with a protein complementation assay based on green fluorescent protein (GFP) reconstitution [54, 55].

## Results and Discussion

### *Alanine scanning mutagenesis of the MoRE and effects on XD and hsp70 binding*

In order to ascertain whether N<sub>TAIL</sub> binding to hsp70 is mediated by the same N<sub>TAIL</sub> residues as binding to XD, we targeted for alanine scanning mutagenesis the MoRE (**Figure 1**) and assessed the binding properties of the resulting variants towards both XD and hsp70 using a split-GFP reassembly assay. In this assay, wild-type (wt) N<sub>TAIL</sub> (or its variants) is fused to the N-terminal half of the green fluorescent protein (NGFP) and the binding partner under study is fused to the C-terminal half of GFP (CGFP). Upon co-expression in *E. coli*, if N<sub>TAIL</sub> interacts with the binding partner, the two GFP halves are brought to sufficiently close proximity to allow the reassembly of the two GFP halves thus leading to reconstitution of the fluorophore. The resulting fluorescence of the cell culture is measured, which is proportional to the interaction strength [54].

Since our previous studies showed that the N-terminal fuzzy region of N<sub>TAIL</sub> spanning residues 401-485 is dispensable for binding, and actually even dampens it [52], we performed alanine scanning mutagenesis of the MoRE (aa 486-504) not in the context of full-length N<sub>TAIL</sub> (aa 401-525) but in the context of the N-terminal truncation variant 471 (aa 471-525) that displays a higher fluorescence (and hence higher interaction strength) with both XD and hsp70 than full-length N<sub>TAIL</sub> [52]. Each residue of the MoRE was individually replaced with an alanine (or with a glycine when alanine was the wild-type residue) giving rise to 19 single-site variants, all derived from truncation variant 471 (**Figure 1**). Next we assessed their binding properties towards the two partners. The results obtained with XD and hsp70 are shown in **Figures 2A** and **2B**, respectively. Note that except for variant 486, which seems to be more sensitive than the others to bacterial proteolysis, the fluorescence variations reported in **Figures 2A** and **2B** cannot be accounted for by differences in the expression levels of the N<sub>TAIL</sub> variants (**Figure 2C**).

As shown in **Figure 2A**, except for variant 499 that exhibits a moderate fluorescence increase, all variants display a lower affinity for XD compared to wt471. The extent of fluorescence

reduction varies greatly from one single variant to another. Variants can be roughly divided into three groups. The first one contains single variants bearing a substitution that has either no effect or only a marginal effect on XD binding (variants 487 to 489, 493, and 497). The single substitution borne by the second group significantly reduces but does not abolish the interaction (variants 486, 490, 492, 496, 498, 500, and 502 to 504). The single substitution of the third group of variants almost completely abolishes the interaction (variants 491, 494, 495, and 501), with fluorescence values close to that provided by the negative control NGFP alone. These findings can nicely be accounted for by structural data, in that the most critical MoRE residues are those whose side chains point directly towards the partner in the XD/MoRE complex [18] (pdb code 1T6O). They also nicely confirm our previous data obtained by random mutagenesis that identified positions 491, 494, 495, 498 and 501 as the most critical [55]. A notable discrepancy however concerns residue 497 whose replacement here with an alanine has only a negligible impact on XD binding, whereas its replacement with a glycine was found to lead to a significant reduction in binding as judged from split-GFP assays [55] and isothermal titration calorimetry (ITC) [56]. There are two possible explanations for these apparently contradictory results. The first one is that alanine promotes helix formation whereas glycine has the opposite effect, as confirmed in the next part of this study. The second one is related to the fact that in our previous study [55] the R497G substitution was used in the context of full-length N<sub>TAIL</sub> (residues 401 to 525) whereas it is used here in the context of N<sub>TAIL</sub> truncation variant 471 (residues 471 to 525). Since we have shown that the N<sub>TAIL</sub> N-terminal fuzzy appendage (residues 401 to 480) dampens the N<sub>TAIL</sub>/XD interaction [52], it is conceivable that the R497G substitution has a stronger effect when used in the context of full-length N<sub>TAIL</sub> than in the context of N<sub>TAIL</sub> truncation variant 471. In addition, the finding that the R497A substitution has only a moderate effect on binding may reflect the fact that preserving the C $\beta$  of residue 497 is sufficient to maintain its stabilizing effect on complex formation, although the underlying mechanisms remain elusive. It is indeed difficult to rationalize the stabilizing effect of Ala497 in light of a previous molecular dynamics simulation study that revealed that Arg497 forms a water-mediated hydrogen bond with Tyr480 of XD [56].

The present results, which allow ranking N<sub>TAIL</sub> residues as 495>494>491>501>498 in terms of their role in binding to XD, are also in quite good agreement with recent mutational results obtained in the context of a  $\Phi$ -value analysis of the N<sub>TAIL</sub>/XD binding reaction which identified residues 498, 495 and 494 (in that order) as most critical for binding [57]. The two studies however disagree as far as the role of residue 491 is concerned, in that the latter study unexpectedly showed that the S491A substitution leads to a K<sub>D</sub> that is close to that of wtN<sub>TAIL</sub>.

The binding efficiencies of the single-site variants towards hsp70 are shown in **Figure 2B**. As previously reported [52], the interaction strength of variant 471 for hsp70 is about 7 times lower than that for XD (compare the y axis scale in **Figures 2A** and **2B**). Beyond this overall difference in interaction strength, the single-site variants behave quite differently towards hsp70. First, the substitutions of all the variants that exhibit a lower affinity compared to wt have only a marginal effect (variants 486, 491, 495, 496, 498, 499). In this respect, all these variants belong to group one described above for XD binding, with no variants falling in groups two or three. This means that, in contrast with the N<sub>TAIL</sub>/XD interaction, no single substitution on its own is able to abolish the N<sub>TAIL</sub>/hsp70 interaction. This observation suggests that no single residue is as critical for hsp70 binding as it is for XD binding, *i.e.* hsp70 is more tolerant of substitutions than XD, which implies that N<sub>TAIL</sub> interacts with XD and hsp70 using different mechanisms. The higher tolerance of hsp70 vis-à-vis N<sub>TAIL</sub> substitutions may be tightly connected with the lower affinity of the

N<sub>TAIL</sub>/hsp70 interaction. One possible scenario could be that the N<sub>TAIL</sub>/XD interaction could rely on a few highly specific interactions, while hsp70 could make use of a larger number (possibly all) of MoRE residues in a less specific and more dynamic manner. Since IDPs have been shown to exhibit also static fuzziness [26], *i.e.* the MoRE can remain highly dynamic at the surface of the partner (for examples see [12, 13, 58]), it is also conceivable that the MoRE is weakly anchored at the surface of hsp70 through numerous transient and quasi-equivalent contacts. Support of this hypothesis comes from recent studies that showed that hsp70 binds its substrates promiscuously leading to fuzzy chaperone-substrate ensembles endowed with a high conformational heterogeneity [59, 60].

The second notable difference with respect to XD binding is that the majority of variants (*i.e.*, variants 487, 489, 490, 492 to 494, 497, and 500 to 504) bind hsp70 better than wt471, indicating that the N<sub>TAIL</sub> sequence can be improved in terms of hsp70 binding. By contrast, in a previous random mutagenesis study, where substitutions were randomly introduced within N<sub>TAIL</sub> and variants were picked at random in the absence of any selection pressure, we showed that variants bearing substitutions within the MoRE tend to display a lower interaction strength towards XD, thus supporting the conclusion that the sequence of the MoRE is poorly evolvable in terms of XD binding [55]. The implications of this finding is that the sequence of N<sub>TAIL</sub> has been selected during evolution to achieve optimal binding with XD, reflecting the need for the N<sub>TAIL</sub>/XD interaction strength to be kept in a precise range so as to ensure dynamic anchoring of the polymerase complex (*i.e.*, L-P) [61] and efficient transcription re-initiation at each intergenic junction of the MeV genome [56].

Thus, it seems that in addition to relying on a different pattern of MoRE residues, the N<sub>TAIL</sub>/hsp70 and N<sub>TAIL</sub>/XD interactions do not have the same evolution potential. One possible hypothesis for the higher evolvability of the N<sub>TAIL</sub>/hsp70 interaction might reside in the fact that the two binding partners have not been subjected to an as tight co-evolution as that of the N<sub>TAIL</sub>/XD pair due to the multiple functional roles that hsp70 plays in the cell and that are not uniquely related to MeV infection. In addition, taking into account the fact that the intracellular concentration of hsp70 has been estimated at 85 nM in Vero cells following transient hyperthermic treatment [62] and that the MeV nucleoprotein is the most abundant viral protein, it is conceivable that a high affinity between the two proteins is not requested for the N<sub>TAIL</sub>/hsp70 interaction to occur and elicit the known effects on viral transcription and replication [50, 63] and on the innate immune response [53].

#### *Impact of varying the $\alpha$ -helicity of the MoRE on XD and hsp70 binding*

Since the MoRE was shown to fold into an  $\alpha$ -helix upon binding to XD [3, 14, 17-20, 22-24] and since the results of the alanine scanning mutagenesis reported in **Figures 2A** and **2B** suggest that N<sub>TAIL</sub> binds XD and hsp70 through different mechanisms, we investigated the impact of varying the helicity of the MoRE on XD and hsp70 binding. To that end, we conceived and generated a variant 471 bearing a mutated MoRE in which all the residues shown to be critical for XD binding by the alanine scanning mutagenesis (**Figure 2A**) were kept unchanged, and all the other residues were replaced with an alanine yielding a variant called “Ala471” (see **Table 1**). Concomitantly, we also generated another variant, called “Gly471” (**Table 1**), in which all the MoRE residues that had been swapped with alanine in Ala471 were replaced with a glycine. The rationale for the generation of these two variants was the following. Because of its high alanine content, Ala471 was expected to exhibit an increased  $\alpha$ -helicity, whereas Gly471 was expected to be unable to fold into an  $\alpha$ -helix even upon XD binding because of its high glycine content

and of the well-known  $\alpha$ -helix-breaking ability of this residue [64-67]. These expectations were reinforced by convergent results obtained using disorder predictions (**Figure 3A**), hydrophobic cluster analysis [68], secondary structure predictions [69] (**supplementary Figure S1**), and modeling (**Figure 3B**). Most importantly, they were experimentally confirmed by direct spectroscopic data. Indeed, far-UV circular dichroism (CD) spectroscopy analysis of AlaMoRE and GlyMoRE peptides (each encompassing residues 485-506, and bearing the above described substitutions) confirmed their markedly increased and decreased helicity, **respectively**, with respect to wtMoRE (**Figure 3C**).

The Ala471 and Gly471 variants were assessed for their ability to bind to XD and hsp70 by split-GFP reassembly. Results are reported in **Figure 3D**. In this experiment, XD interaction with full-length N<sub>TAIL</sub> (aa 401-525) was used as the reference and its fluorescence set to 100%. As already reported [52], wt471 (aa 471-525) exhibits a higher affinity for XD than full-length N<sub>TAIL</sub> (compare 401/XD and wt471/XD in **Figure 3D**). Notably, when the native MoRE sequence in wt471 was replaced with the sequence of AlaMoRE, the binding was found to increase even further (compare wt471/XD and Ala471/XD in **Figure 3D**) suggesting that a preformed  $\alpha$ -helix leads to an increased N<sub>TAIL</sub>/XD interaction strength. This finding is in agreement with our previous kinetics studies that showed that the folding of the MoRE is the rate-limiting step of the N<sub>TAIL</sub>/XD binding reaction [52]. By contrast, when the wtMoRE sequence within variant 471 was replaced with the sequence of GlyMoRE, the binding to XD drops down to the negative control level (compare Gly471/XD and Stop/XD in **Figure 3D**). This latter result definitely indicates that if the MoRE cannot fold into an  $\alpha$ -helix, then N<sub>TAIL</sub>/XD binding is impaired in spite of the presence within the MoRE of all the residues that alanine scanning mutagenesis identified as critical for the interaction. Taken together, the results obtained with the Ala471 and Gly471 variants support a scenario where the **lower** the helicity of the MoRE the **lower** the interaction strength. Importantly, results also indicate that XD is able to bind to a pre-formed  $\alpha$ -helix arguing for a mixed mechanism involving both conformational selection and induced folding.

When the same experiment was performed using hsp70 instead of XD, drastically different results were observed (**Figure 3D**). As expected, wt471 displays a much lower interaction strength towards hsp70 compared to XD (compare wt471/hsp and wt471/XD in **Figure 3D**). However, with Ala471 and Gly471, the binding to hsp70 decreases only marginally in both cases (compare wt471/hsp with Ala471/hsp or Gly471/hsp in **Figure 3D**). The possibility that different protein expression levels could be responsible for the observed differences in fluorescence was checked and ruled out (**Figure 3E**). Thus, while increasing or decreasing the  $\alpha$ -helicity of the MoRE has tremendously opposite effects on XD binding, the effects on hsp70 binding are both small and similar.

In an effort to further characterize the effect of helical propensity of N<sub>TAIL</sub> on its binding properties, in analogy to our previous studies [16, 52], we resorted to perform temperature-jump kinetic experiments on wtMoRE, AlaMoRE and GlyMoRE peptides. Unfortunately, due to the complexity of these experiments, we could not perform binding studies on hsp70. In fact, the relatively low affinity between hsp70 and N<sub>TAIL</sub> together with the small fluorescence changes induced by binding prevent an accurate determination of the (un)binding rate constants. The binding of XD to wtMoRE, AlaMoRE and GlyMoRE was measured by incubating a constant concentration of 5  $\mu$ M XD with varying concentrations of either wtMoRE or AlaMoRE or GlyMoRE (typically ranging from 2 to 50  $\mu$ M). In analogy to what we described previously, in all cases (un)binding was induced by a rapid discharge of 35 kV on a quartz cell, corresponding

to a rapid increase in temperature of 9 K. **Supplementary Figure S2** shows three typical fluorescence traces observed in these binding experiments. All curves conformed to a single exponential transition (**Supplementary Figure S2** and data not shown).

Inspection of kinetic data (**Figure 3F**) reveals that whilst the behavior of AlaMoRE is very similar to that of wtMoRE (wtMoRE:  $k_{off}^{app} = 1460 \pm 80 \text{ s}^{-1}$ ,  $k_{on}^{app} = 30 \pm 4 \text{ s}^{-1} \mu\text{M}^{-1}$ ; AlaMoRE:  $k_{off}^{app} = 1400 \pm 100 \text{ s}^{-1}$ ,  $k_{on}^{app} = 34 \pm 5 \text{ s}^{-1} \mu\text{M}^{-1}$ ), there is a detectable destabilization of the complex in the case of GlyMoRE ( $k_{off}^{app} = 1610 \pm 60 \text{ s}^{-1}$ ,  $k_{on}^{app} = 5 \pm 2 \text{ s}^{-1} \mu\text{M}^{-1}$ ). This conclusion is mirrored by a decrease of the association rate constant, which is represented by the slope of the observed rate constant as a function of reactant concentration. Thus, whilst an increase in helicity of the MoRE compared to the wt sequence does not appear to contribute to a stabilization of the complex *in vitro*, it is evident that a destabilization of the secondary structure of the MoRE corresponds to a weaker binding to XD. A possible reason for the discrepancy between split-GFP and kinetics data, as far as results with AlaMoRE are concerned, may lie in the differences in the experimental set up between the two approaches.

In conclusion, the results reported in **Figures 3D** and **F** indicate that  $\alpha$ -helical folding is a strict requirement for binding to XD, whereas favoring of disfavoring it only partially reduces the binding to hsp70. These results reinforce the idea that XD and hsp70 bind N<sub>TAIL</sub> through different molecular mechanisms.

#### *Mapping the reciprocal binding regions in the N<sub>TAIL</sub>/hsp70 interaction*

The results reported in **Figures 2** and **3** suggest that the N<sub>TAIL</sub>/hsp70 interaction does not rely on few specific MoRE residues and is also relatively insensitive to the folding state of the MoRE. These observations raised the question as to which are the minimal requirements for the N<sub>TAIL</sub>/hsp70 interaction to occur, an interaction that has been much less investigated compared to the N<sub>TAIL</sub>/XD interaction and for which molecular and structural data are still lacking. In order to fill this gap and to contribute to a better understanding of this interaction, we first used a deletion approach to map the reciprocal binding sites and then targeted for site-directed mutagenesis the hsp70-binding site of N<sub>TAIL</sub>.

Hsp70 consists of three domains: a nucleotide binding domain (NBD, 1-384), a peptide binding domain (PBD, 381-543) and a lid (543-641) (see [51] and references therein cited). To map the minimal region involved in N<sub>TAIL</sub> binding, we generated different hsp70 constructs containing one or more of these domains (**Figure 4A**, lower panel) and tested their ability to interact with wt471 or with its Ala471 and Gly471 derivatives. Results are shown in **Figure 4A**, upper panel. The three N<sub>TAIL</sub> variants provided very similar interaction profiles, with the highest interaction strength being obtained with the hsp70 construct devoid of the C-terminal lid (NBD-PBD, 1-543), and the lowest interaction strength being observed with the PBD and Lid domains.

The finding that hsp70 deletion variant devoid of the C-terminal lid (NBD-PBD, 1-543) exhibits the highest interaction strength is in agreement with our previous studies where surface plasmon resonance experiments yielded a two-fold increase in the  $K_D$  (68 *versus* 34  $\mu\text{M}$ ) of the N<sub>TAIL</sub>-binding reaction when full-length hsp70 was used instead of NBD-PBD [51]. They are also in agreement with previous reports by the group of Lila Gierasch that showed that the hsp70 C-terminal region is dispensable for binding of a peptide substrate, inter-domain allostery and co-chaperon interaction [70].



In the case of wt471, the second best binder is full-length hsp70 (1-641). PBD-Lid and NBD each show a significant binding to wt471, although lower than full-length hsp70. However, when the Ala471 and Gly471 variants were used instead of wt471, the binding of NBD proved to be slightly more efficient than full-length hsp70. Since the NBD and the PBD-lid together cover the full-length hsp70 these results suggest that wt471 might form weak contacts throughout more than one hsp70 domains. By contrast, the PBD and the Lid on their own are unable to bind to wt471 and its two derivatives. However, we cannot exclude the possibility that results obtained with the PBD and the Lid alone could arise from a low solubility/expression of these constructs, but due to the lack of affinity tag, the various hsp70 constructs fused to CGFP could not be purified as was the case of NGFP fusions. To better visualize/appreciate the binding behavior of full-length hsp70 and of its lid-free derivative, their binding data, as obtained with wt471 and with the Ala471 and Gly471 variants, were compared (**Figure 4B**). This comparison confirms the slightly reduced binding of full-length hsp70 to Ala471 and Gly471 variants with respect to wt471 as already observed in the experiment reported in **Figure 3D**. By contrast, the lid-free hsp70 binds all the three N<sub>TAIL</sub> variants in a similar manner. In conclusion, results indicate that removing the lid not only increases binding towards N<sub>TAIL</sub> but also renders it less sensitive to substitutions within the MoRE that have profound effects on XD binding.

To further dissect the N<sub>TAIL</sub>/hsp70 interaction, we generated a series of N<sub>TAIL</sub> truncation variants (**Figure 5A**) and assessed their binding properties towards hsp70 by split-GFP reassembly. The rationale for doing that was to map the minimal binding region and to assess the contribution of N<sub>TAIL</sub> Box3 (aa 517-525) to binding. Indeed, previous studies showed that although hsp70 binds with a weak affinity a Box3 peptide [49], the presence of this motif stabilizes N<sub>TAIL</sub>/hsp70 complexes [51] that are otherwise mediated by high affinity binding to Box2 [50]. As shown in **Figure 5B**, the N<sub>TAIL</sub>/hsp70 interaction strength increases with increasing N- and C-terminal shortening of N<sub>TAIL</sub>, with the wtMoRE on its own providing the highest binding. Note that, once again, the increase in fluorescence is not accounted for by an increased expression level (**Figure 5C**). These results therefore indicate that as in the case of XD, the MoRE is the only N<sub>TAIL</sub> region responsible for binding to full-length hsp70, and that Box3 behaves as a weak binding dampener.

Although the N<sub>TAIL</sub>/hsp70 interaction relies only on the MoRE, the latter is seemingly capable of binding, albeit weakly, to both NBD and PBD-lid, thus raising the question as to how such a short peptide can stretch over to cover the whole hsp70. One possibility, that remains to be assessed, is that N<sub>TAIL</sub> binds to hsp70 with a 2:1 stoichiometry. That hsp70 may form sub-stoichiometric chaperone-substrate complexes is supported by recent studies that showed that DnaK binds to hTRF-1 with a 1:3 stoichiometry [60]. Alternatively, it is conceivable that the MoRE binds to multiple hsp70 sites in a dynamic and promiscuous manner, as already observed in the case of hsp33 (Dana Reichmann, personal communication).

#### *Rational design of an N<sub>TAIL</sub> variant with enhanced interaction strength towards hsp70*

Since alanine scanning mutagenesis showed that several single substitutions lead to an increase in N<sub>TAIL</sub>/hsp70 binding, we reasoned that this could be the basis for devising an hsp70 “super binder” (hsb). The latter (see **Table 1**) was obtained by collectively introducing all the substitutions that individually increase the binding to hsp70 in the context of 471 (hsb471). Swapping the sequence of wtMoRE with that of hsb is associated with a slightly lower disorder probability (**Figure 5D**) reflecting the higher alanine content of the latter (**Table 1**). In line with expectations, this rationally designed 471 variant displays a much higher binding (2.35 times)

towards full-length hsp70 than wt471 in a split-GFP reassembly assay (compare wt471 and hsb471 in **Figure 5B**).

Since results reported in **Figure 5B** definitely identified the MoRE as the only interacting unit between N<sub>TAIL</sub> and hsp70, and since substitution of wt471 with hsb471 resulted in a significant increase in N<sub>TAIL</sub>/hsp70 binding, we reasoned that hsbMoRE should be the strongest hsp70 binder. This hypothesis was experimentally confirmed: hsbMoRE indeed binds hsp70 almost three times better than wtMoRE (**Figure 5B**). In other words, the difference between wtMoRE and hsbMoRE in terms of binding to hsp70 is higher when the MoRE is used alone than when it is used in the context of 471 (2.35 times, see above). In the results reported in **Figure 4**, hsp70 truncation variant NBD-PBD (hsp1-543) proved to be the best N<sub>TAIL</sub> binder. Results reported in **Figure 5E** show that when wtMoRE was substituted with hsbMoRE in the context of 471, the binding to hsp1-543 increases (compare wt471 and hsb471) and reaches maximal level when hsbMoRE was used alone.

The finding that replacement of as much as 70% (13 residues out of 19, **Table 1**) of the sequence of the wtMoRE with alanine or glycine residues results in an almost 3-fold increase in binding to hsp70 (**Figure 5B**), raises the question as to how can hsp70 binding to N<sub>TAIL</sub> be specific of the MoRE while being relatively independent of the sequence of the latter. It is conceivable that hsp70 recognizes not a precise amino acid sequence or motif but rather a pattern of few residues with specific chemical features with no strict positional conservation. Which could be these chemical features? The sequence of hsbMoRE is highly enriched in Ala, Gly and Leu residues (in this order) and has a null net charge (**Table 1**). Of note, these features were also found to favor binding of peptides to hsp33, an ATP-independent, redox-regulated chaperone [71]. This study also identified Asp/Glu residues as strongly disfavoring binding. Noteworthy, in hsbMoRE the two naturally occurring Asp residues were removed, thereby providing a conceptual framework to explain increased binding by hsbMoRE. Interestingly, the study by Reichmann and co-workers also unveiled that increased hydrophobicity is correlated to increased binding strength [71]. Although previous data identified exposed patches of hydrophobic residues in native proteins [72] and hydrophobic residues in peptides as targets of hsp70 binding [73], it should be emphasized that hydropathy *per se* is not sufficient to explain increased binding strength by hsbMoRE. Indeed, Ala471 is a weaker hsp70 binder than wt471 (**Figure 3D**) and wt471 is a weaker hsp70 binder than hsb471 (**Figure 5B**) although the hydropathy of AlaMoRE (1.6), as obtained using the Kyte-Doolittle scale and the expasy server (<https://www.expasy.org/>), is higher than that of wtMoRE (-0.24) and the hydropathy of hsbMoRE (0.96) is higher than that of wtMoRE. Finally, since the generation of hsbMoRE implied introducing all the Ala/Gly substitutions that individually led to increased binding strength while keeping unchanged all MoRE positions that once substituted led to diminished binding, it follows that the latter can be regarded as determinants of N<sub>TAIL</sub>/hsp70 binding. Among these residues, a stretch of three Leu (495, 496 and 498) is reminiscent of the sequence of the NRLLLTG heptapeptide co-crystallized with DnaK, the bacterial counterpart of hsp70 [74].

## Conclusions

Using alanine scanning mutagenesis, truncation variants and rational site-directed mutagenesis in combination with a protein complementation assay based on split-GFP reassembly, we have shed light on the mechanisms by which N<sub>TAIL</sub> recognizes XD and hsp70, two structured partners differing in size and fold. Results indicate that although N<sub>TAIL</sub> uses the same binding region (*i.e.*,

the MoRE) to bind the two partners, the binding mechanisms are not the same. Hsp70 is much more tolerant of N<sub>TAIL</sub> substitutions than XD, with this higher tolerance likely coming at the cost of a lower affinity compared to the N<sub>TAIL</sub>/XD pair. The N<sub>TAIL</sub>/hsp70 interaction is more evolvable compared to the N<sub>TAIL</sub>/XD interaction, *i.e.*, the majority of substitutions lead to increased interaction strength. Notably, while an increased and a decreased  $\alpha$ -helicity of the MoRE lead to enhanced and reduced interaction strength with XD, respectively, the extent of helicity of the free form of the MoRE has seemingly a negligible impact on hsp70 binding, suggesting that the MoRE does not adopt an  $\alpha$ -helical conformation once bound to hsp70. Shortening hsp70 by removing the lid results in increased binding towards N<sub>TAIL</sub>, as does the shortening of the latter to encompass only the MoRE. Finally, N<sub>TAIL</sub>/hsp70 binding can be further increased by introducing within the MoRE all the substitutions that individually result in increased interaction strength.

The enhanced interaction strength of hsbMoRE towards hsp70 NBD-PBD may be instrumental for obtaining crystals of a complex in view of structure determination by X-ray crystallography. The availability of the structure of the complex will shed light on the binding mode of this mutated MoRE to hsp70. In particular, the high-resolution structure of the complex will enable ascertaining whether the NBD and the PBD bind each a MoRE molecule, and will unveil the actual conformation that the substrate adopts once bound to hsp70. Works are in progress in our laboratory to reach this goal.

Finally, and from a more general perspective, the present results contribute to shed light on the molecular mechanisms by which IDPs recognize their partners. Indeed, in spite of the increasing interest being paid to IDPs, the molecular features that encode their binding efficiency remain elusive, and the variety of binding modes and mechanisms exhibited by IDPs (see [8] and references therein cited) suggests that there is no general mechanism for their binding-induced folding. In addition,

Here, by showing that the conformation sampled by the free form of the MoRE of N<sub>TAIL</sub> does not necessarily commit N<sub>TAIL</sub> to adopt an  $\alpha$ -helical conformation in the bound form, we bring an additional “brick in the wall”, *i.e.* we provide an additional example of partner-mediated polymorphism and of the relative insensitiveness of the bound structure to the pre-recognition state. These results are in line with recent findings that revealed that N<sub>TAIL</sub> folds as anti-parallel  $\beta$  sheet at the air/water interface [75] thereby further underscoring the extreme polymorphism of this IDP.

## Materials and Methods

### *Alanine scanning mutagenesis and generation of NGFP-N<sub>TAIL</sub> constructs*

Two complementary mutagenic primers were used in two separate PCR amplification experiments using the wtN<sub>TAIL</sub>471 construct in the pNGG vector [55, 76] as template. The first one used attB1 and the reverse mutagenic primer, and the second one used attB2 and the forward mutagenic primer. After DpnI treatment to remove methylated parental DNA, the two products of the first PCR amplification were mixed and then used as overlapping megaprimers for a second PCR amplification with primers attL1a and attL2a [76]. The second PCR product was directly used in an LR reaction (Invitrogen) with expression vector pNGG as previously described [76]. The mutagenic primers used to generate all the single-site alanine variants are listed in **supplementary Table S1**.

Full-length *wt*N<sub>TAIL</sub>401 (aa 401-525), *wt*N<sub>TAIL</sub>471 (aa 471-525) and *wt*N<sub>TAIL</sub>481 (aa 481-525) have been already described [52]. Truncation variant *wt*N<sub>TAIL</sub>401 $\Delta$ Box3 (aa 401-516) was obtained by PCR amplification using the *wt*N<sub>TAIL</sub>401 construct in pNGG as template and primers attB1 and  $\Delta$ 3-attB2. After DpnI treatment, the PCR product was used as template for a second PCR amplification using attL1a and attL2a primers. The second PCR product was directly used as template in an LR reaction with expression vector pNGG as described above.

Truncation variant *wt*MoRE (aa 485-506) coding sequence was obtained by PCR amplification using primers B1MoRE , B2MoRE and the *wt*N<sub>TAIL</sub>401 construct in pNGG as template. After DpnI treatment, the PCR product was used as template in a second PCR amplification using attL1a and attL2a primers. The second PCR product was directly used in an LR reaction with expression vector pNGG as above.

The hsb471 variant coding sequence (aa 471-525 with mutated MoRE) was obtained by PCR amplification using the *wt*N<sub>TAIL</sub>471 construct in pNGG as template and primers hsbNTailF1 and hsbNTailR1 in two separate PCR amplifications. The first one used primers attB1 and hsbNTailR1, and the second one used attB2 and hsbNTailF1. After DpnI treatment, the two products of the first PCR amplification were mixed and then used as overlapping megaprimers for a second PCR amplification with primers attL1a and attL2a. The second PCR product was directly used in an LR reaction with expression vector pNGG as described above.

The hsbMoRE variant coding sequence was obtained by PCR amplification using the *wt*471 in pNGG as template. In the first PCR amplification, the two halves of the coding sequence were amplified separately using primers 4hsb5a and 4hsb5b in one tube, and primers 4hsb5c and 4hsb5d in a second tube. After DpnI treatment, a third PCR amplification was performed using 1  $\mu$ l of each of the first two tubes and primers attL1a and attL2a. The PCR product of the third PCR amplification was then directly used in an LR reaction with expression vector pNGG as described above.

The Ala471 and Gly471 N<sub>TAIL</sub> variants coding sequences were obtained as that of hsbMoRE471 except that primer pairs AlaNTailR1 and AlaNTailF1, and GlyNTailR1 and GlyNTailF1 were respectively used in the first PCR experiment.

The sequences of all the primers used to generate the NGFP-N<sub>TAIL</sub> constructs are shown in **Supplementary Table S2**.

CaCl<sub>2</sub> competent *E. coli* TAM1 cells (Active Motif) were transformed with the LR mixtures, and selected on ampicillin-containing plates. All the sequences were checked by sequencing (GATC Biotech) and found to conform to expectations.

#### *Generation of hsp70-CGFP constructs*

Hsp70-CGFP constructs were PCR amplified using the full-length hsp70 coding sequence as template [51] and the following primers: Hsp70BspHI and 384AatII for NBD, Hsp70BspHI and Hsp70\_543 for NBD-PBD, Hsp70BspHI and Hsp70AatII for NBD-PBD-lid, Hsp70\_384 and Hsp70\_543 for PBD, Hsp70\_384 and Hsp70AatII for PBD-lid, BspHI543 and Hsp70AatII for the lid. The sequences of all the primers used to generate the hsp70-CGFP constructs are shown in **Supplementary Table S2**.

After DpnI treatment, the PCR product was purified and then digested with BspHI and AatII. After purification, digested PCR products were ligated to BspHI and AatII digested pMRBAD [54]. CaCl<sub>2</sub> competent *E. coli* TAM1 cells (Active Motif) were transformed with the ligation mixture, and selected on kanamycin plates. All the sequences were checked by sequencing (GATC Biotech) and found to conform to expectations.

The XD-CGFP construct and the split-GFP reassembly assay have been described already [52].

#### *Purification of NGFP fusions under denaturing conditions*

The expression level of the NGFP constructs was assessed as follows. After fluorescence and OD<sub>600</sub> has been measured, the whole volume of triplicate cultures (12 ml) was spun for 5' at 4000g in a single well of a 24-wells deep-well. Cell pellets were re-suspended in 1 ml of 50 mM Tris/HCl pH 8, 0.3 M NaCl, 10 mM imidazole, 0.1% Triton X100, 5 mM phenylmethylsulfonyl fluoride (PMSF) and 0.25 mg/ml Lysozyme, and frozen. After thawing, cell lysates were supplemented with 20 µg/ml DNase I and 20 mM MgSO<sub>4</sub> and incubated at 37° C for 30' under shaking. Urea (2 g per well) was added and allowed to dissolve and to denature the proteins contained in the lysate by an additional 30' incubation at 37° C under shaking. After spinning for 5' at 4000g, supernatants were supplemented with 100 µl of a 50% suspension of IMAC sepharose high performance beads (GE Healthcare). His-tagged NGFP fusions were allowed to bind to the beads for 30' at room temperature on a rotating wheel. After washing with 5 x 1 ml of 50 mM Tris/HCl pH 8, 0.3 M NaCl, 50 mM imidazole, beads were re-suspended in 50 µl of reducing SDS-PAGE loading buffer and the proteins contained in 10 µl of this suspension were resolved by SDS-PAGE.

#### *Modeling of the MoREs*

The structural models of the free form of the MoREs were obtained using the PEP-FOLD3 server (<http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD3>) [77].

#### *Synthetic peptides and far-UV circular dichroism (CD) measurements*

The synthetic peptide mimicking wtMoRE has been already described [52]. Peptides mimicking the AlaMoRE and the GlyMoRE were purchased from JPT (Berlin, Germany). All the MoRE peptides were designed to encompass residues 485-506 of N<sub>TAIL</sub>, and possess an additional non-native tyrosine at the C-terminus to allow their concentration to be inferred from the absorbance at 280 nm.

The CD spectra of the MoRE peptides were measured using a Jasco 810 dichrograph, flushed with N<sub>2</sub> and equipped with a Peltier thermoregulation system. One-mm thick quartz cuvettes were used. Peptide concentrations were 0.1 mg mL<sup>-1</sup>. Peptides were dissolved in 10 mM sodium phosphate pH 7. Spectra were measured between 190 and 260 nm at 20°C. The scanning speed was 50 nm/min, with data pitch of 0.2 nm. Each spectrum is the average of three acquisitions. The spectrum of buffer was subtracted from the protein spectrum. Spectra were smoothed using the “means-movement” smoothing procedure implemented in the Spectra Manager package. The BESTSEL website (<http://bestsel.elte.hu/>) [78] was used to analyze the experimental data in the 190–250 nm range.

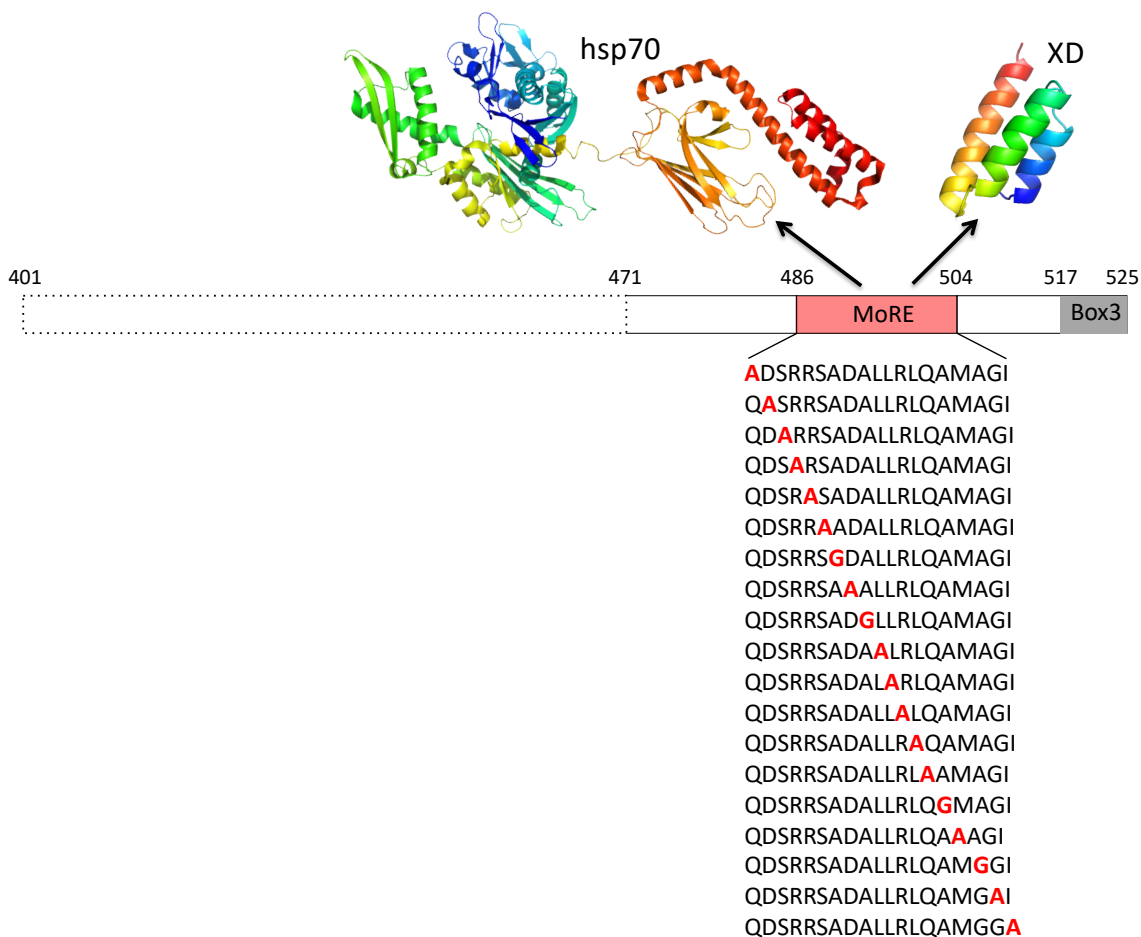
#### *Temperature-jump fluorescence spectroscopy*

Kinetic binding experiments were performed by using a Hi-Tech PTJ-64 capacitor-discharge T-jump apparatus (Hi-Tech, Salisbury, UK). Temperature was rapidly changed with a jump-size of 9 °C, from 16 °C to 25 °C. Usually 10–20 individual traces were averaged. The fluorescence change of N-acetyl-tryptophanamide (NATA) was used in control measurements. Degassed and filtered samples were slowly pumped through the 0.5 x 2 mm quartz flow cell before data acquisition. The excitation wavelength was 296 nm and the fluorescence emission was measured using a 320 nm cut-off glass filter.

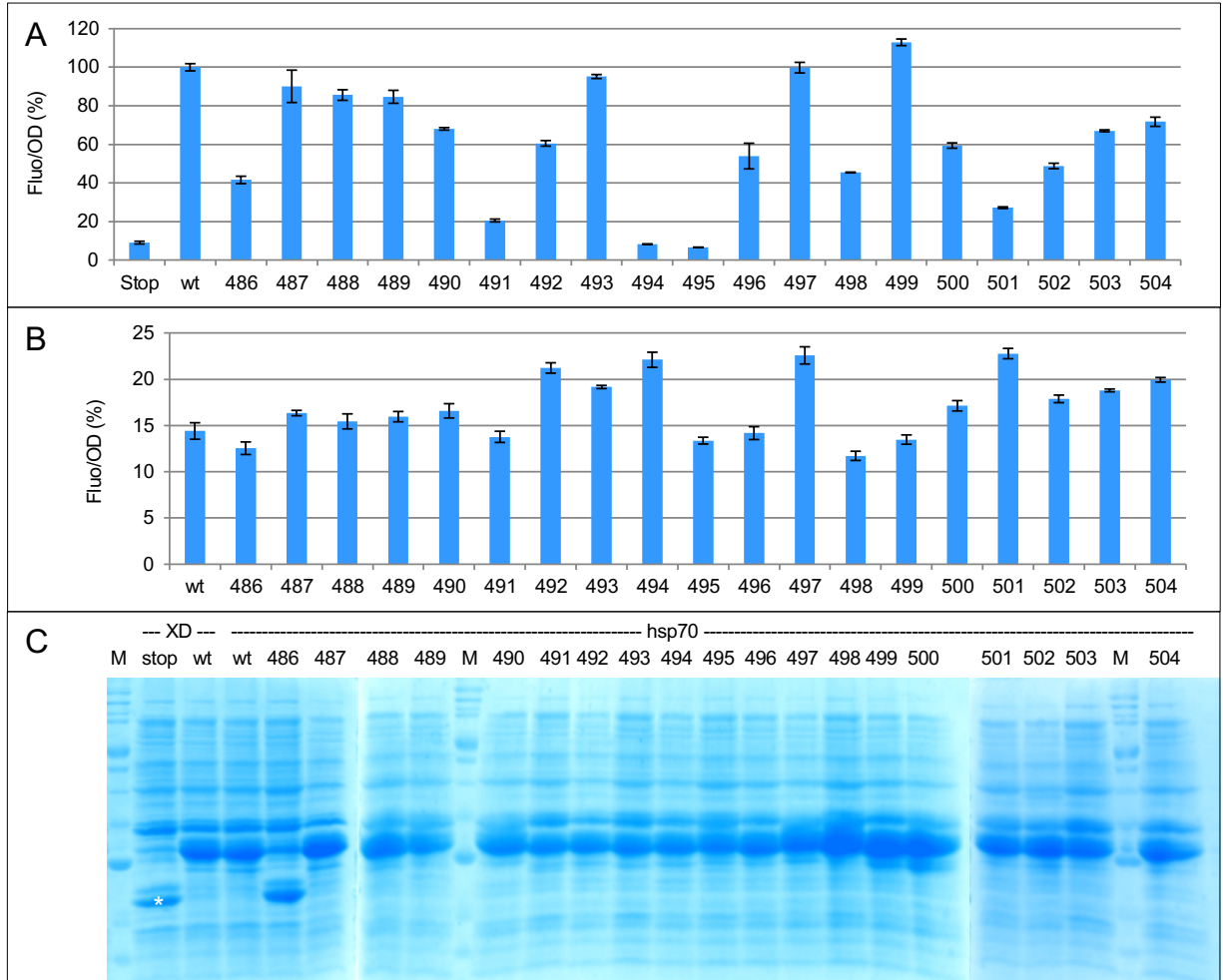
Experiments made use of a previously reported XD variant Y480W with a C-terminal hexahistidine tag [16, 52] and the three MoRE-mimicking peptides described above. The XD variant was purified as already described [16].

The experiments were carried out by mixing a constant concentration of Y480W XD (5  $\mu$ M) with excess concentrations of either wtMoRE, or AlaMoRE or GlyMoRE peptide. The buffer used was 10 mM sodium phosphate and 150 mM NaCl at pH 7.0.

## Figures

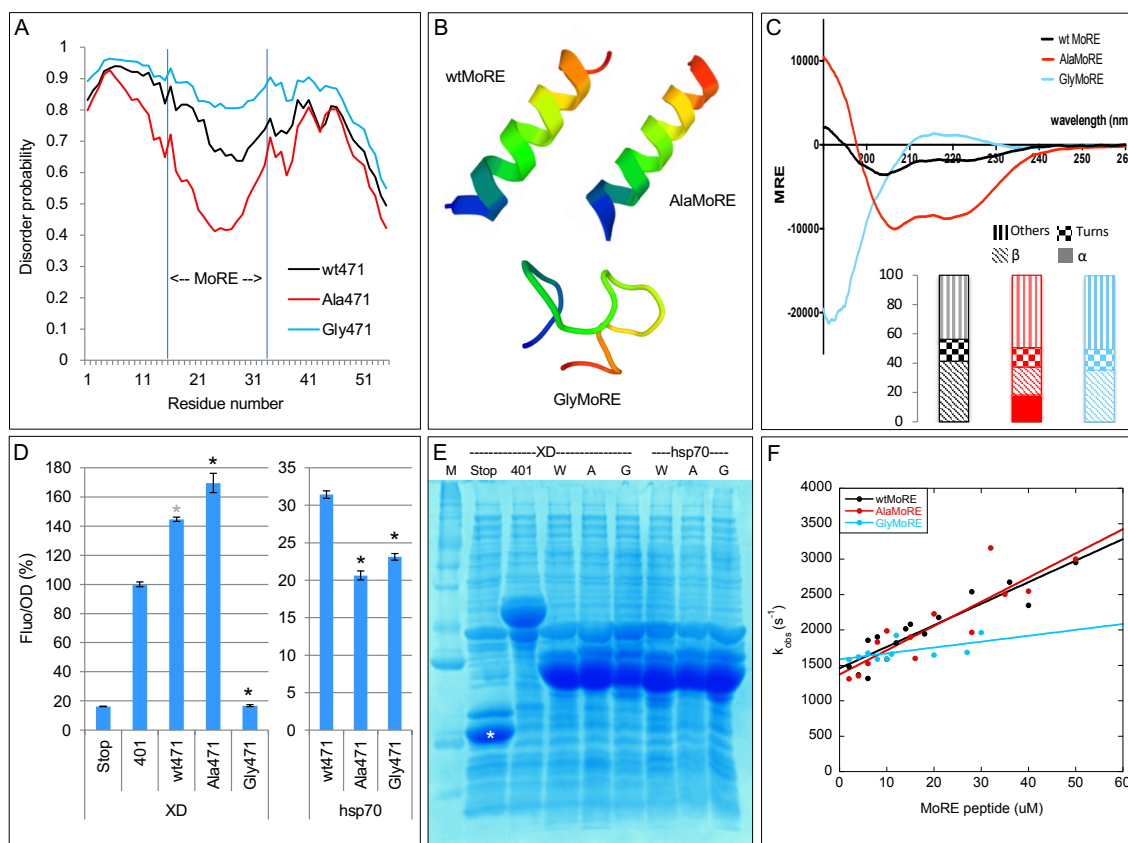


**Figure 1. Schematic representation of the variants resulting from alanine scanning mutagenesis.** Mutagenesis was performed in the context of the truncation variant 471 that spans residues 471-525 of N<sub>TAIL</sub>. Nineteen single-site variants were generated. Above the scheme are shown cartoon representations of the crystal structure of hsp70 and XD (pdb code 1OKS). The structure of hsp70 is based on pdb codes 1HJO and 4JNF. The relative orientation of the two hsp70 domains (*i.e.* amino acids 3-382 and amino acids 389-610) is based on the structure of a form encompassing residues 1-554 (pdb code 1YUW).

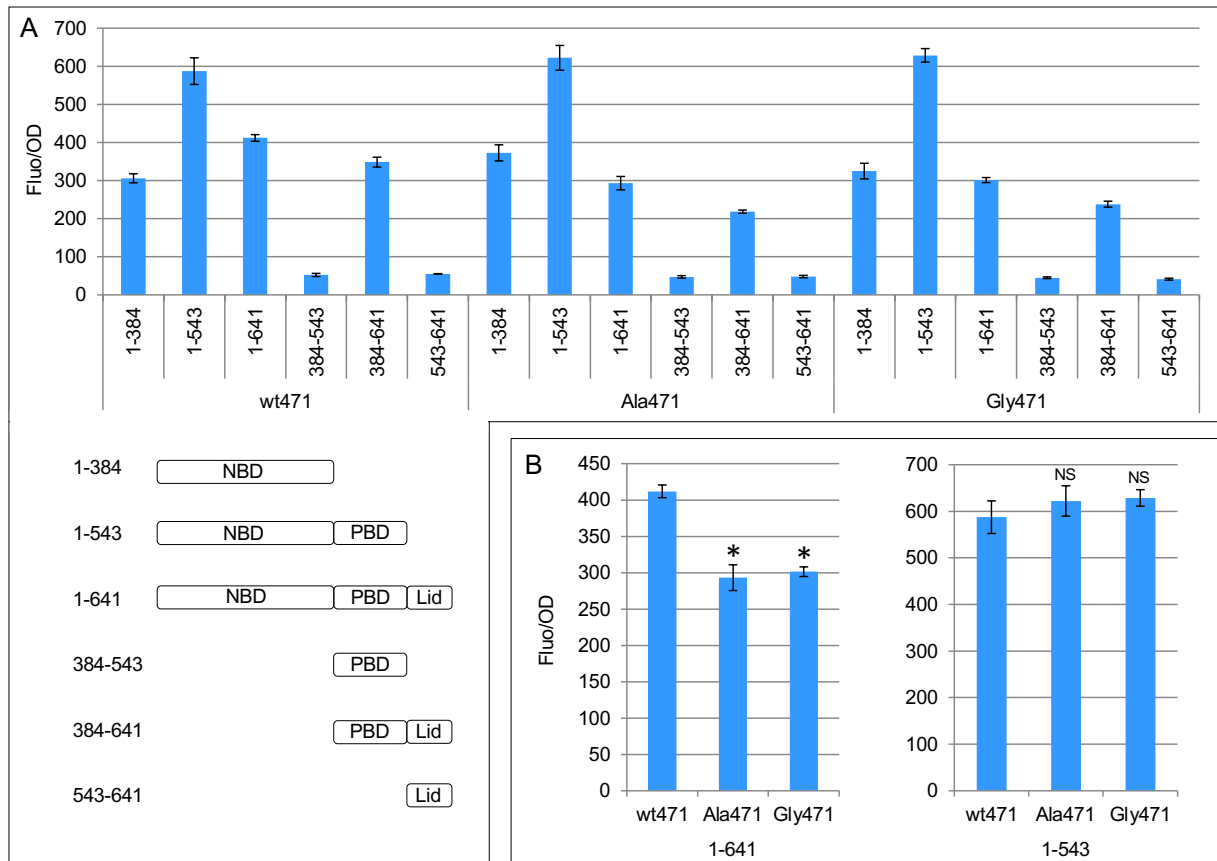


**Figure 2. Binding efficiency of alanine single-site  $N_{TAIL}$  variants towards XD and hsp70 as assessed by split-GFP reassembly.** Normalized fluorescence of NGFP- $N_{TAIL}$  variants upon co-expression with XD-CGFP (A) and hsp70-CGFP (B). Stop, NGFP alone (negative control); wt, wt471 truncation variant (aa 471-525) fused to NGFP. The numbers 486 to 504 indicate the single-site variant used (with the number corresponding to the mutated position). Note that all single-site variants were derived from wt471. Results are expressed as percentage of the positive control (wt, set to 100%). Standard deviations of triplicates are indicated for each histogram. (C) SDS-PAGE analysis of the total expression of NGFP- $N_{TAIL}$  variants. On top of the gel, XD and hsp indicate the co-expressed partner. Stop, wt, and 486 to 504 indicate the NGFP fusion used. Since only NGFP fusions were His-tagged, only  $N_{TAIL}$  variants and not CGFP-fused partners were purified under denaturing conditions. As a result, identical results were obtained for XD and hsp70. Here are shown the results obtained with hsp70. M, molecular mass markers: 212, 158, 116, 97.2, 66.4, 55.6, 42.7, 34.6, 27, 20, 14.3 kDa. Underlined values correspond to most intense bands in the gel. The white asterisk designates the NGFP fragment alone as resulting from expression of stop.

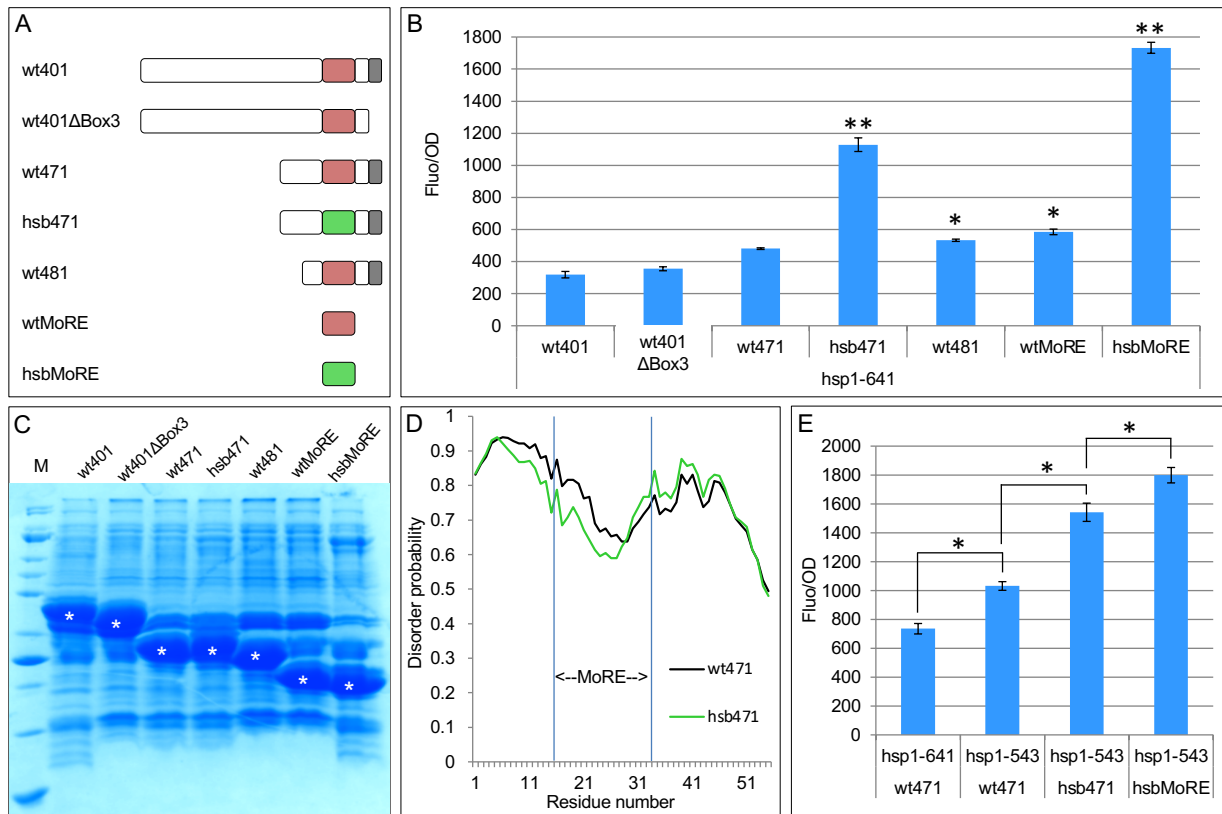




**Figure 3. Binding efficiency towards XD and hsp70 of  $N_{TAIL}$  variants mutated within the MoRE as assessed by split-GFP reassembly.** (A) Disorder probability, as obtained using IUPred [79] of wt471, Ala471 and Gly471. The x axis is residue number of 471. The position of the MoRE within the sequence is shown. (B) Structural models, as obtained using PEP-FOLD3, of wtMoRE, AlaMoRE and GlyMoRE peptides. (C) Far-UV CD spectra of the MoRE peptides. The inset shows the secondary structure content as derived using BESTSEL. (D) Normalized fluorescence of  $N_{TAIL}$  variants fused to NGFP upon co-expression with XD-CGFP (XD) and hsp70-CGFP (hsp). Stop, NGFP alone; 401, full-length wt $N_{TAIL}$  (aa 401-525); wt471, Ala471 and Gly471, 471 variants (aa 471-525) bearing the wt sequence or multiple alanine or glycine substitutions, respectively, within the MoRE (see **Table 1** for details). The black asterisk denotes a statistically significant difference (T Student's test, with  $p < 0.01$ ) with respect to wt471. The grey asterisk denotes a statistically significant difference (T Student's test, with  $p < 0.01$ ) with respect to 401. (E) SDS-PAGE analysis of the total expression of NGFP- $N_{TAIL}$  variants. XD and hsp indicate the CGFP fusions co-expressed, while labels on top of each lane indicate the expressed NGP- $N_{TAIL}$  variant (W, wt471; A, Ala471; G, Gly471). The band corresponding to NGFP alone (resulting from expression of the Stop construct) is indicated by a white star. M, molecular weight markers: 212, 158, 116, 97.2, 66.4, 55.6, 42.7, 34.6, 27, 20, 14.3 kDa. Underlined values correspond to most intense bands in the gel. Expected molecular masses of the GFP fusion proteins are: 20.4 kDa (Stop), 34 kDa (401), 25.9 kDa (wt471, Ala471, Gly471). (F) Pseudo-first-order kinetics of the binding between Y480W XD and wtMoRE, AlaMoRE and GlyMoRE peptides.



**Figure 4. Binding efficiency of hsp70 truncation variants as assessed by split-GFP reassembly.** (A) Upper panel: normalized fluorescence of hsp70 constructs schematically represented in the lower panel, fused to CGFP, as obtained upon co-expression with wt471, Ala471 and Gly471 all fused to NGFP. On the y axis are shown the fluorescence of the culture normalized as a function of culture biomass ( $OD_{600}$ ). Standard deviations of triplicate experiments are indicated for each histogram. Lower panel: Scheme of the various hsp70 constructs used in panel A. NBD, nucleotide binding domain. PBD, peptide binding domain. Numbers on the left indicate the first and last residue of the construct with respect to full-length hsp70 (aa 1-641). (B) Results obtained with full-length hsp70 (aa 1-641) or with its lid-free deletion variant (aa 1-543). The asterisk denotes a statistically significant difference (T Student's test, with  $p < 0.01$ ) with respect to 471. NS: non statistically significant with respect to wt471.



**Figure 5. Binding efficiency of N<sub>TAIL</sub> truncation and mutated variants towards hsp70 as assessed by split-GFP reassembly. (A)** Schematic representation of N<sub>TAIL</sub> variants used in panel B. The 401-485 and 505-516 regions of N<sub>TAIL</sub> are represented in white. The MoRE, encompassing residues 486-504, is represented as a red (*wild-type*) or green (mutated) box. wt401, wt full-length N<sub>TAIL</sub> (aa 401-525); wt401ΔBox3, wtN<sub>TAIL</sub> devoid of Box3 (aa 401-516); wt471, aa 471-525 with wtMoRE; hsb471, aa 471-525 with hsbMoRE; wt481, truncated wtN<sub>TAIL</sub> variant (aa 481-525); wtMoRE, truncated wtN<sub>TAIL</sub> variant (aa 485-506); hsbMoRE, mutated wtN<sub>TAIL</sub> variant (aa 485-506). **(B)** Normalized fluorescence of N<sub>TAIL</sub> constructs fused to NGFP, as obtained upon co-expression with full-length hsp70 fused to CGFP. Standard deviations of triplicate experiments are indicated for each histogram. The asterisks denote a statistically significant difference (T Student's test) with respect to wt471 (\*\* p<0.01; \* p<0.05). **(C)** SDS-PAGE analysis of the total expression of NGFP-N<sub>TAIL</sub> variants. Relevant bands are marked with a white star. M, molecular weight markers: 212, 158, 116, 97.2, 66.4, 55.6, 42.7, 34.6, 27, 20, 14.3 kDa. Underlined values correspond to most intense bands in the gel. **(D)** Disorder probability prediction, as obtained using IUPred, of wt471 and hsb471. The x axis is the amino acid sequence of 471. The position of the MoRE is shown. **(E)** Results of a split-GFP reassembly assay using full-length hsp70 (hsp1-641) or lid-free hsp70 (hsp1-543) fused to CGFP, and wt471, hsb471 or hsbMoRE fused to NGFP. The asterisk denotes a statistically significant difference (T Student's test, with p<0.05).

### Supplementary Information

Two supplementary figures (Figures S1 and S2) and two supplementary tables (Tables S1 and S2).

### **Acknowledgements**

This work was supported by the CNRS. It was also partly supported by the European program H2020 under the EVAg Research Infrastructure (grant agreement 653316). FT is a recipient of a PhD fellowship from the Italo-French University.

## References

- [1] Gutsche I, Desfosses A, Effantin G, Ling WL, Haupt M, Ruigrok RW, et al. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science*. 2015.
- [2] Longhi S, Receveur-Brechot V, Karlin D, Johansson K, Darbon H, Bhella D, et al. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *J Biol Chem*. 2003;278:18638-48.
- [3] Bourhis J, Johansson K, Receveur-Br  chot V, Oldfield CJ, Dunker AK, Canard B, et al. The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Res*. 2004;99:157-67.
- [4] Dunker AK, Garner E, Guilliot S, Romero P, Albrecht K, Hart J, et al. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pac Symp Biocomput*. 1998;3:473-84.
- [5] Wright PE, Dyson HJ. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J Mol Biol*. 1999;293:321-31.
- [6] Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem*. 2002;269:2-12.
- [7] Uversky VN. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci*. 2002;11:739-56.
- [8] Habchi J, Tompa P, Longhi S, Uversky VN. Introducing Protein Intrinsic Disorder. *Chem Rev*. 2014;114:6561-88.
- [9] Wright PE, Dyson HJ. Linking folding and binding. *Curr Opin Struct Biol*. 2009;19:31-8.
- [10] Blocquel D, Habchi J, Gruet A, Blangy S, Longhi S. Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies. *Mol Biosyst*. 2012;8:392-410.
- [11] Habchi J, Blangy S, Mamelli L, Ringkj  bing Jensen M, Blackledge M, Darbon H, et al. Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipaviruses. *J Biol Chem*. 2011;286:13583-602.
- [12] Communie G, Habchi J, Yabukarski F, Blocquel D, Schneider R, Tarbouriech N, et al. Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. *PLoS Pathog*. 2013;9:e1003631.
- [13] Baronti L, Eral  s J, Habchi J, Felli IC, Pierattelli R, Longhi S. Dynamics of the intrinsically disordered C-terminal domain of the Nipah virus nucleoprotein and interaction with the X domain of the phosphoprotein as unveiled by NMR spectroscopy. *Chem BioChem*. 2015;16:268-76.
- [14] Bourhis JM, Receveur-Br  chot V, Oglesbee M, Zhang X, Buccellato M, Darbon H, et al. The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Sci*. 2005;14:1975-92.
- [15] Blocquel D, Habchi J, Costanzo S, Doizy A, Oglesbee M, Longhi S. Interaction between the C-terminal domains of measles virus nucleoprotein and phosphoprotein: A tight complex implying one binding site. *Protein Sci*. 2012;21:1577-85.
- [16] Dosnon M, Bonetti D, Morrone A, Eral  s J, di Silvio E, Longhi S, et al. Demonstration of a folding after binding mechanism in the recognition between the measles virus NTAIL and X domains. *ACS ChemBiol*. 2015;10:795-802.

- [17] Johansson K, Bourhis JM, Campanacci V, Cambillau C, Canard B, Longhi S. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J Biol Chem*. 2003;278:44567-73.
- [18] Kingston RL, Hamel DJ, Gay LS, Dahlquist FW, Matthews BW. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc Natl Acad Sci U S A*. 2004;101:8301-6.
- [19] Belle V, Rouger S, Costanzo S, Liquiere E, Strancar J, Guigliarelli B, et al. Mapping alpha-helical induced folding within the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by site-directed spin-labeling EPR spectroscopy. *Proteins: Structure, Function and Bioinformatics*. 2008;73:973-88.
- [20] Bernard C, Gely S, Bourhis JM, Morelli X, Longhi S, Darbon H. Interaction between the C-terminal domains of N and P proteins of measles virus investigated by NMR. *FEBS Lett*. 2009;583:1084-9.
- [21] Bischak CG, Longhi S, Snead DM, Costanzo S, Terrer E, Londergan CH. Probing structural transitions in the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by vibrational spectroscopy of cyanylated cysteines. *Biophys J*. 2010;99:1676-83.
- [22] Gely S, Lowry DF, Bernard C, Ringkjøbing-Jensen M, Blackledge M, Costanzo S, et al. Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein *J Mol Recognit*. 2010;23:435-47.
- [23] Kavalenka A, Urbancic I, Belle V, Rouger S, Costanzo S, Kure S, et al. Conformational analysis of the partially disordered measles virus NTAIL-XD complex by SDSL EPR spectroscopy. *Biophys J*. 2010;98:1055-64.
- [24] Ringkjøbing Jensen M, Communie G, Ribeiro ED, Jr., Martinez N, Desfosses A, Salmon L, et al. Intrinsic disorder in measles virus nucleocapsids. *Proc Natl Acad Sci U S A*. 2011;108:9839-44.
- [25] Fuxreiter M, Tompa P. Fuzzy interactome: the limitations of models in molecular biology. *Trends Biochem Sci*. 2009;34:3.
- [26] Tompa P, Fuxreiter M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci*. 2008;33:2-8.
- [27] Fuxreiter M. Fuzziness: linking regulation to protein dynamics. *Mol Biosyst*. 2012;8:168-77.
- [28] Miskei M, Antal C, Fuxreiter M. FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Res*. 2017;45:D228-D35.
- [29] D'Urzo A, Konijnenberg A, Rossetti G, Habchi J, Li J, Carloni P, et al. Molecular Basis for Structural Heterogeneity of an Intrinsically Disordered Protein Bound to a Partner by Combined ESI-IM-MS and Modeling. *J Am Soc Mass Spectrom*. 2015;26:472-81.
- [30] Morin B, Bourhis JM, Belle V, Woudstra M, Carrière F, Guigliarelli B, et al. Assessing induced folding of an intrinsically disordered protein by site-directed spin-labeling EPR spectroscopy. *J Phys Chem B*. 2006;110:20596-608.
- [31] Wang Y, Chu X, Longhi S, Roche P, Han W, Wang E, et al. Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. *Proc Natl Acad Sci U S A*. 2013;110:E3743-E52.
- [32] Fuxreiter M, Simon I, Friedrich P, Tompa P. Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *J Mol Biol*. 2004;338:1015-26.

- [33] Lee SH, Kim DH, Han JJ, Cha EJ, Lim JE, Cho YJ, et al. Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins. *Curr Protein Pept Sci*. 2012;13:34-54.
- [34] Hammes GG, Chang YC, Oas TG. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc Natl Acad Sci USA*. 2009;106:13737-41.
- [35] Wang Y, Tang C, Wang E, Wang J. Exploration of multi-state conformational dynamics and underlying global functional landscape of maltose binding protein. *PLoS Comput Biol*. 2012;8:e1002471.
- [36] Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics*. 2008;9 Suppl 1:S1.
- [37] Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN. Flexible nets. *FEBS J*. 2005;272:5129-48.
- [38] Uversky VN, Oldfield CJ, Dunker AK. Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling. *J Mol Recognit*. 2005;18:343-84.
- [39] Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, et al. Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes. *PLoS Comput Biol*. 2006;2:e100.
- [40] De BP, Banerjee AK. Involvement of actin microfilaments in the transcription/replication of human parainfluenza virus type 3: possible role of actin in other viruses. *Microsc Res Tech*. 1999;47:114-23.
- [41] Moyer SA, Baker SC, Horikami SM. Host cell proteins required for measles virus reproduction. *J Gen Virol*. 1990;71:775-83.
- [42] Laine D, Trescol-Biémont M, Longhi S, Libeau G, Marie J, Vidalain P, et al. Measles virus nucleoprotein binds to a novel cell surface receptor distinct from FcγRII via its C-terminal domain: role in MV-induced immunosuppression. *J Virol*. 2003;77:11332-46.
- [43] Laine D, Bourhis J, Longhi S, Flacher M, Cassard L, Canard B, et al. Measles virus nucleoprotein induces cell proliferation arrest and apoptosis through NTA1/NR and NCORE/FcγRIIB1 interactions, respectively. *J Gen Virol*. 2005;86:1771-84.
- [44] Iwasaki M, Takeda M, Shirogane Y, Nakatsu Y, Nakamura T, Yanagi Y. The matrix protein of measles virus regulates viral RNA synthesis and assembly by interacting with the nucleocapsid protein. *J Virol*. 2009;83:10374-83.
- [45] Sato H, Masuda M, Miura R, Yoneda M, Kai C. Morbillivirus nucleoprotein possesses a novel nuclear localization signal and a CRM1-independent nuclear export signal. *Virology*. 2006;352:121-30.
- [46] tenOever BR, Servant MJ, Grandvaux N, Lin R, Hiscott J. Recognition of the Measles Virus Nucleocapsid as a Mechanism of IRF-3 Activation. *J Virol*. 2002;76:3659-69.
- [47] Colombo M, Bourhis JM, Chamontin C, Soriano C, Villet S, Costanzo S, et al. The interaction between the measles virus nucleoprotein and the Interferon Regulator Factor 3 relies on a specific cellular environment. *Virol J*. 2009;6:59.
- [48] Watanabe A, Yoneda M, Ikeda F, Sugai A, Sato H, Kai C. Peroxiredoxin 1 is required for efficient transcription and replication of measles virus. *J Virol*. 2011;85:2247-53.
- [49] Zhang X, Glendening C, Linke H, Parks CL, Brooks C, Udem SA, et al. Identification and characterization of a regulatory domain on the carboxyl terminus of the measles virus nucleocapsid protein. *J Virol*. 2002;76:8737-46.

- [50] Zhang X, Bourhis JM, Longhi S, Carsillo T, Buccellato M, Morin B, et al. Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus. *Virology*. 2005;337:162-74.
- [51] Couturier M, Buccellato M, Costanzo S, Bourhis JM, Shu Y, Nicaise M, et al. High Affinity Binding between Hsp70 and the C-Terminal Domain of the Measles Virus Nucleoprotein Requires an Hsp40 Co-Chaperone. *J Mol Recognit*. 2010;23:301-15.
- [52] Gruet A, Dosnon M, Blocquel D, Brunel J, Gerlier D, Das RK, et al. Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. *FEBS J*. 2016;283:576-94.
- [53] Kim MY, Shu Y, Carsillo T, Zhang J, Yu L, Peterson C, et al. hsp70 and a novel axis of type I interferon-dependent antiviral immunity in the measles virus-infected brain. *J Virol*. 2013;87:998-1009.
- [54] Wilson CG, Magliery TJ, Regan L. Detecting protein-protein interactions with GFP-fragment reassembly. *Nat Methods*. 2004;1:255-62.
- [55] Gruet A, Dosnon M, Vassena A, Lombard V, Gerlier D, Bignon C, et al. Dissecting partner recognition by an intrinsically disordered protein using descriptive random mutagenesis. *J Mol Biol*. 2013;425:3495-509.
- [56] Bloyet L, Brunel J, Dosnon M, Hamon V, Erales J, Gruet A, et al. Modulation of re-initiation of measles virus transcription at intergenic regions by PXD to NTAIL binding strength. *PLoS Path*. 2016;12:e1006058.
- [57] Bonetti D, Troilo F, Toto A, Brunori M, Longhi S, Gianni S. Analyzing the folding and binding steps of an intrinsically disordered protein by protein engineering. *Biochemistry*. 2017;in press.
- [58] Mittag T, Orlicky S, Choy WY, Tang X, Lin H, Sicheri F, et al. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proc Natl Acad Sci U S A*. 2008;105:17772-7.
- [59] Rosenzweig R, Sekhar A, Nagesh J, Kay LE. Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles. *Elife*. 2017;6.
- [60] Sekhar A, Nagesh J, Rosenzweig R, Kay LE. Conformational heterogeneity in the Hsp70 chaperone-substrate ensemble identified from analysis of NMR-detected titration data. *Protein Sci*. 2017.
- [61] Brunel J, Choppy D, Dosnon M, Bloyet LM, Devaux P, Urzua E, et al. Sequence of events in measles virus replication: role of phosphoprotein-nucleocapsid interactions. *J Virol*. 2014;88:10851-63.
- [62] Oglesbee MJ, Liu Z, Kenney H, Brooks CL. The highly inducible member of the 70 kDa family of heat shock proteins increases canine distemper virus polymerase activity. *J Gen Virol*. 1996;77:2125-35.
- [63] Shu Y, Habchi J, Costanzo S, Padilla A, Brunel J, Gerlier D, et al. Plasticity in structural and functional interactions between the phosphoprotein and nucleoprotein of measles virus. *J Biol Chem*. 2012;287:11951-67.
- [64] Scott KA, Alonso DO, Sato S, Fersht AR, Daggett V. Conformational entropy of alanine versus glycine in protein denatured states. *Proc Natl Acad Sci U S A*. 2007;104:2661-6.
- [65] Borchers W, Theillet FX, Katzer A, Finzel A, Mishall KM, Powell AT, et al. Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nat Chem Biol*. 2014;10:1000-2.
- [66] Otieno S, Kriwacki R. Probing the role of nascent helicity in p27 function as a cell cycle regulator. *PLoS One*. 2012;7:e47177.



- [67] Rogers JM, Wong CT, Clarke J. Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *J Am Chem Soc.* 2014;136:5197-200.
- [68] Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, et al. Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci.* 1997;53:621-45.
- [69] Chandonia JM, Karplus M. New methods for accurate prediction of protein secondary structure. *Proteins.* 1999;35:293-306.
- [70] Smock RG, Blackburn ME, Gierasch LM. Conserved, disordered C terminus of DnaK enhances cellular survival upon stress and DnaK in vitro chaperone activity. *J Biol Chem.* 2011;286:31821-9.
- [71] Reichmann D, Xu Y, Cremers CM, Ilbert M, Mittelman R, Fitzgerald MC, et al. Order out of disorder: working cycle of an intrinsically unfolded chaperone. *Cell.* 2012;148:947-57.
- [72] de Crouy-Chanel A, Kohiyama M, Richarme G. Interaction of DnaK with native proteins and membrane proteins correlates with their accessible hydrophobicity. *Gene.* 1999;230:163-70.
- [73] Clerico EM, Tilitsky JM, Meng W, Gierasch LM. How hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. *J Mol Biol.* 2015;427:1575-88.
- [74] Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME, et al. Structural analysis of substrate binding by the molecular chaperone DnaK. *Science.* 1996;272:1606-14.
- [75] Bénarouche A, Habchi J, Cagna A, Maniti O, Girard-Egrot A, Cavalier J, et al. Interfacial properties of NTAIL, an intrinsically disordered protein. *Biophys J.* 2017;in press.
- [76] Gruet A, Longhi S, Bignon C. One-step generation of error-prone PCR libraries using Gateway(R) technology. *Microb Cell Fact.* 2012;11:14.
- [77] Lamiable A, Thevenet P, Rey J, Vavrusa M, Derreumaux P, Tuffery P. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* 2016;44:W449-54.
- [78] Micsonai A, Wien F, Kernya L, Lee YH, Goto Y, Refregiers M, et al. Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci U S A.* 2015;112:E3095-103.
- [79] Dosztanyi Z, Csizmok V, Tompa P, Simon I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics.* 2005;21:3433-4.