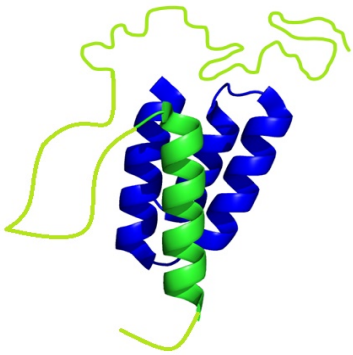
PhD Course in Biochemistry
XXXI Cycle (Academic Years 2015-2018)

Molecular mechanisms of folding of Intrinsically Disordered Proteins

PhD student
Francesca Troilo



Tutors

Prof. Stefano Gianni
Dott.ssa Sonia Longhi

PhD coordinator

Prof. Stefano Gianni

December 2018

PhD Course in Biochemistry
XXXI Cycle (Academic Years 2015-2018)

Molecular mechanisms of folding of Intrinsically Disordered Proteins

PhD student
Francesca Troilo

Tutors

Prof. Stefano Gianni
Dott.ssa Sonia Longhi

PhD coordinator

Prof. Stefano Gianni

My PhD was a joint-supervision between the University of Rome “La Sapienza” and the University of Aix-Marseille (Programma Vinci 2015).

SCIENTIFIC COMMUNICATIONS

Publications:

1. **Troilo F**, Bonetti D, Longhi S, Gianni S**. The fuzzy appendage of measles virus N_{TAIL} hampers interaction with XD through a combination of entropy and enthalpy**. IN PREPARATION.

2. Troilo F, Bignon C, Gianni S, Fuxreiter M, Longhi S. **Experimental characterization of fuzzy protein assemblies: interactions of paramyxoviral N_{TAIL} domains with their functional partners.** Methos in Enzymology, (2018) in press.

3. **Troilo F**, Bonetti D, Camilloni C, Toto A, Longhi S, Brunori M, Gianni S. **Folding Mechanism of the SH3 Domain from Grb2.** J Phys Chem B. (2018). doi: 10.1021/acs.jpcb.8b06320. PMID: 30091591

4. Bonetti D, **Troilo F**, Toto A, Travaglini-Allocatelli C, Brunori M, Gianni S. **Mechanism of Folding and Binding of the N-Terminal SH2 Domain from SHP2.** J Phys Chem B. (2018). doi: 10.1021/acs.jpcb.8b05651. PMID: 30047735

5. Bonetti D, **Troilo F**, Brunori M, Longhi S, Gianni S. **How Robust Is the Mechanism of Folding-Upon-Binding for an Intrinsically Disordered Protein?** Biophys J. (2018) 114(8):1889-1894. doi: 10.1016/j.bpj.2018.03.017. PMID: 29694866

6. Bignon C, **Troilo F**, Gianni S, Longhi S. **Partner-Mediated Polymorphism of an Intrinsically Disordered Protein.** J Mol Biol. (2018) 430(16):2493-2507. doi: 10.1016/j.jmb.2017.11.012. PMID: 29197511

7. 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) 56(29):3780-3786. doi: 10.1021/acs.biochem.7b00350. PMID: 28661120

8. **Troilo F**, Bonetti D, Toto A, Visconti L, Brunori M, Longhi S, Gianni S. **The Folding Pathway of the KIX Domain.** ACS Chem Biol. (2017) 12(6):1683-1690. doi: 10.1021/acschembio.7b00289. PMID: 28459531

Selected communications by the candidate at scientific conferences:

1. **Francesca Troilo**, Christophe BIGNON, Daniela BONETTI, Stefano GIANNI and Sonia LONGHI. "**Towards a better understanding of the molecular mechanisms by which fuzzy regions affect the folding rate of adjacent molecular recognition elements"** 2nd NGP-Net Symposium Non-globular Proteins in Molecular Physiopathology (2016) – Belgrade, Serbia (poster presentation).

2. **Francesca Troilo, "Molecular mechanisms of folding of an intrinsically disordered protein"** workshop "Physics of Biomolecules: Structure, Dynamics and Function"(2018)- Bressanone, Italy (short talk presentation)

3. **Francesca TROILO**, Christophe BIGNON, Daniela BONETTI, Stefano GIANNI and Sonia LONGHI**. "The fuzzy appendage of measles virus N_{TAIL} hampers interaction with XD through a combination of entropy and enthalpy".** BeMM Symposium (2018) – Rome, Italy (poster presentation)

INDEX

INTRODUCTION

1. Intrinsically Disordered Proteins

During the last decades, the well-known and accepted dogma that posits that a protein needs to have a well-defined three-dimensional structure to carry out its function was broken by the discovery of intrinsically disordered proteins.

Intrinsically disordered proteins (IDPs) are ubiquitous proteins lacking a well-defined three-dimensional structure and existing as an ensemble of conformations under physiological conditions of pH and salinity. Despite (or thanks to) the lack of structure, these proteins are fully functional and are involved in many biological functions. The intrinsic disorder may concern the whole protein or just some regions in the protein, named intrinsically disordered regions (IDRs) (Uversky, 2000)(Dunker et al., 2001)(Uversky, 2002) (Tompa, 2011)(Uversky & Dunker, 2010)(Wright & Dyson, 1999)(Dunker et al., 2013)(Habchi, Tompa, Longhi, & Uversky, 2014)(Dyson, 2011)(Tompa, 2012)(Uversky, 2013). IDPs/IDRs interact with their physiological partners, *via* particular regions called Molecular Recognition Elements (MoREs). MoREs are regions with an inherent propensity to fold that undergo a disorder-to-order transition upon binding (Dyson & Wright, 2005)(Dyson & Wright, 2002) and that can fold into α-helices (α-MoREs), β-strands (β-MoREs), or can adopt an irregular structure (i-MoREs) or a combination of different secondary structures (complex-MoREs).

Protein disorder is very common in nature: the presence of IDPs or IDRs is reported in all living (bacteria, archaea, eukaryota) and semi-living (i.e. viruses) organisms.

IDPs/IDRs presents some features that make them different with respect to globular proteins:

- High stability to heat and chemical denaturation

- Low resistance to proteolysis
- Low content of hydrophobic (Isoleucine, Leucine, Valine) and aromatic (Tryptophan, Tyrosine, Phenylalanine) residues, normally involved in the formation of the hydrophobic core of globular proteins, as well as of Cysteine residues, that are involved in protein conformation stability via disulphide bonds formation and coordination of prosthetic groups
- High content of Alanine and Glycine residues, of polar residues (Arginine, Glycine, Glutamine, Serine, Glutamate, Lysine) and of Proline that, despite being hydrophobic, is a structure-breaking amino acid.
- High Stokes radius
- Unusual mobility in SDS-PAGE

These peculiar features have allowed the development of bioinformatics tools for the predictions of intrinsic disorder. Bioinformatics analyses indicate that more than 50% of eukaryotic proteins are predicted to contain long (more than 30 contiguous amino acids) disordered regions. Besides the function as linkers or spacers, disorder confers many advantages to proteins: it confers flexibility, plasticity and capability to interact with multiple partners (proteins, small ligands, DNA/RNA). The generally high entropic penalty associated with the disorder-to-order transition lead to low-affinity, though specific, interactions that are ideally suited in regulatory pathways. Thanks also to the ability to undergo disorder-to-order transitions upon binding, IDPs/IDRs are indeed involved in many biological functions. Significant role of IDPs/IDRs is reported in regulation, recognition, signaling and control pathways (i.e. hub and scaffolding proteins use the disorder to bind multiple partners to activate different pathways). IDPs/IDRs may act as inhibitors or

activators, regulating the function of the partner. Besides, given the open and extended structure, IDPs/IDRs are also able to interact with small ligands (e.g. metals or ions) neutralizing or storing them and releasing them when needed by the cell. The high flexibility and exposure to the solvent make disordered regions good substrates for regulatory, proteolytic attack and post-translational modifications (acetylation, hydroxylation, ubiquitination, methylation, phosphorylation). Furthermore, disorder is often observed in proteins constituting channels, that take advantage of the flexibility and capability to recognize specifically different partners and undergoes conformational changes upon the interaction with ligands. IDPs/IDRs are also involved in regulation of transcription and translation and cell cycle control. Therefore, these proteins are fundamental for many biological functions and the defect in their functions and/or in their abundance may lead to diseases.

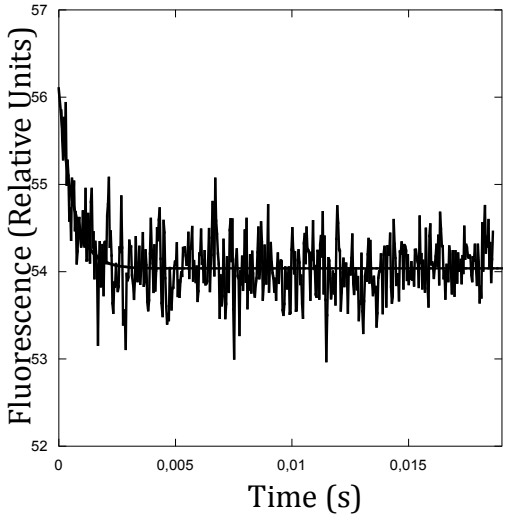
2. Folding of IDPs: kinetic studies

Following interaction with physiological partners, IDPs/IDRs undergo a (partial) disorder-to-order transition that takes place in MoREs. Therefore, studying the folding of IDPs/IDRs implies studying their interaction with their partner(s). The folding-upon-binding reaction of IDPs/IDRs involves at least two steps: the formation of the interacting complex and the conformational change of the IDP/IDR (Kiefhaber, Bachmann, & Jensen, 2012).

A powerful experimental technique to obtain information about the induced folding of IDPs/IDRs is the study of the kinetics of the interaction with their partners. A common way to study the kinetics is to induce the folding of the IDP/IDR by mixing the disordered protein with its ligand and following the reaction by optical spectroscopy. Instruments commonly used

in these mixing techniques are stopped-flow (milliseconds time scale) or continuous flow (microseconds time scale). These techniques, despite powerful to follow the kinetics of reactions, suffer from the limitation of the so-called “dead time” of the instrument. The latter is defined as the time elapsing between the end of the mixing of the solutions and the beginning of the kinetic measurement. For the stopped-flow the dead time is about 2-3 milliseconds (ms), while for the continuous flow it is about 50 microseconds (μs). Therefore, the mixing techniques cannot be used to study ultra-fast reactions that are lost in the instrument dead time.

To study ultra-fast reactions relaxation techniques are used. In these methods a system at equilibrium (in this case the IDP bound to the partner and hence folded) is perturbed by a rapid change in the experimental conditions, such as a rapid increase in temperature (temperature-jump) or pressure (pressure-jump). The return to the new equilibrium is then measured through spectroscopic techniques. In this way, reactions occurring in nano- or microseconds time scale may be studied too. The relaxation process is usually estimated following the variation of an optical probe, such as the fluorescence of an aromatic residue (tryptophan, tyrosine), as a function of the acquisition time and the fit of the resulting curve is then used to calculate the observed rate constant of the reaction (k_{obs})(Figure 1).



$$F_{(t)} = \Delta F \cdot e^{(-k_{obs} \cdot t)} + F_{\infty}$$

Figure 1. Example of a single exponential variation of tryptophan fluorescence as a function of the acquisition time in a temperature-jump binding experiment. The single exponential trace is fitted by a single exponential equation that allows calculating the k_{obs} of the reaction. ΔF is the amplitude of the trace, t is the time of acquisition and F_{∞} is the final value of the fluorescence.

As already mentioned, the reaction between IDPs/IDRs and their partners involves at least two steps: formation of the complex (regulated by the association and dissociation rate constants, k_{on} and k_{off}) and the induced folding of the IDP/IDR (regulated by the folding and unfolding rate constants, k_f and k_u), which tend to be cooperative. Therefore, in most of the cases, it is hard distinguishing between the two steps since the reaction occurs in an all-or-none way and only one rate constant, representing the overall process, can be determined. The more the system is complex (long

IDPs/IDRs) the more performing a quantitative analysis becomes hard. For short IDPs/IDRs, usually, merely two states are experimentally visible: the unbound state, with the isolate free species, and the complex, formed by the folded IDP/IDR bound to the partner. In this case a possible accumulation of intermediate(s) specie(s) is not visible. When a reaction occurs following a two-state mechanism, it is represented by a linear dependence of the observed rate constant (k_{obs}) as a function of protein concentrations. In many cases, the linear dependence of the k_{obs} with protein concentration represents an apparent two-state mechanism "hiding" intermediate(s) step(s), which are not experimentally detectable. The presence of intermediate(s) steps occurring in the reaction is usually manifested by a hyperbolic dependence of the k_{obs} as a function of protein concentration.

2.1. Induced fit and conformational selection mechanisms

Two extreme mechanisms are used to describe the folding upon binding process of IDPs/IDRs: conformational selection (Monod, Wyman, & Changeux, 1965) and induced-fit (Kiefhaber et al., 2012). In the conformational selection mechanism, the disordered protein folds before the binding with the partner, while, in the induced fit, folding occurs after binding. However, it was observed that in many cases the interaction is described by a combination of the two pathways. In some cases indeed it is reported that part of the protein folds before the interaction (conformational selection) and the rest after the binding (induced fit) (Csermely, Palotai, & Nussinov 2010), while in other cases a flux through the two pathways occurring simultaneously is detected (Hammes, Chang, & Oas, 2009).

To simplify the kinetic studies of these complicated reactions, a common approach is performing the experiments in pseudo-first order conditions. Pseudo first-order conditions are reached when one of the

reactants is used in high excess with respect to the other one, such that the reaction rate will depend just on the concentration of the reactant present at low concentration. Under these conditions, the system approaches a first-order scenario as showed in the following scheme:

$$A + B \underset{k_2}{\overset{k_1}{\rightleftarrows}} AB$$

$$[A] \gg [B]$$

$$B \underset{k_2}{\overset{k_1(A)}{\rightleftarrows}} AB$$

Kinetic binding experiments, performed under pseudo-first order conditions, may be used to distinguish between an induced fit mechanism and a conformational selection mechanism (Gianni, Dogan, & Jemth, 2014a). In the induced fit mechanism, folding occurs after binding. If this mechanism is working, kinetic experiments, conducted under pseudo-first order conditions with respect to each of the two reactants, will return the same hyperbolic behavior of k_{obs} as a function of the concentration of the protein in excess irrespective of which of the two interacting species is used as the excess ligand (Figure 2, panel A). When the conformational selection mechanism is working, folding occurs before binding and the k_{obs} will display a hyperbolic dependence as a function of protein concentration only when the specie at low, constant concentration is the unfolded protein, while, when it is the IDP/IDR that is in excess the k_{obs} will return a linear behavior (Figure 2, panel B).

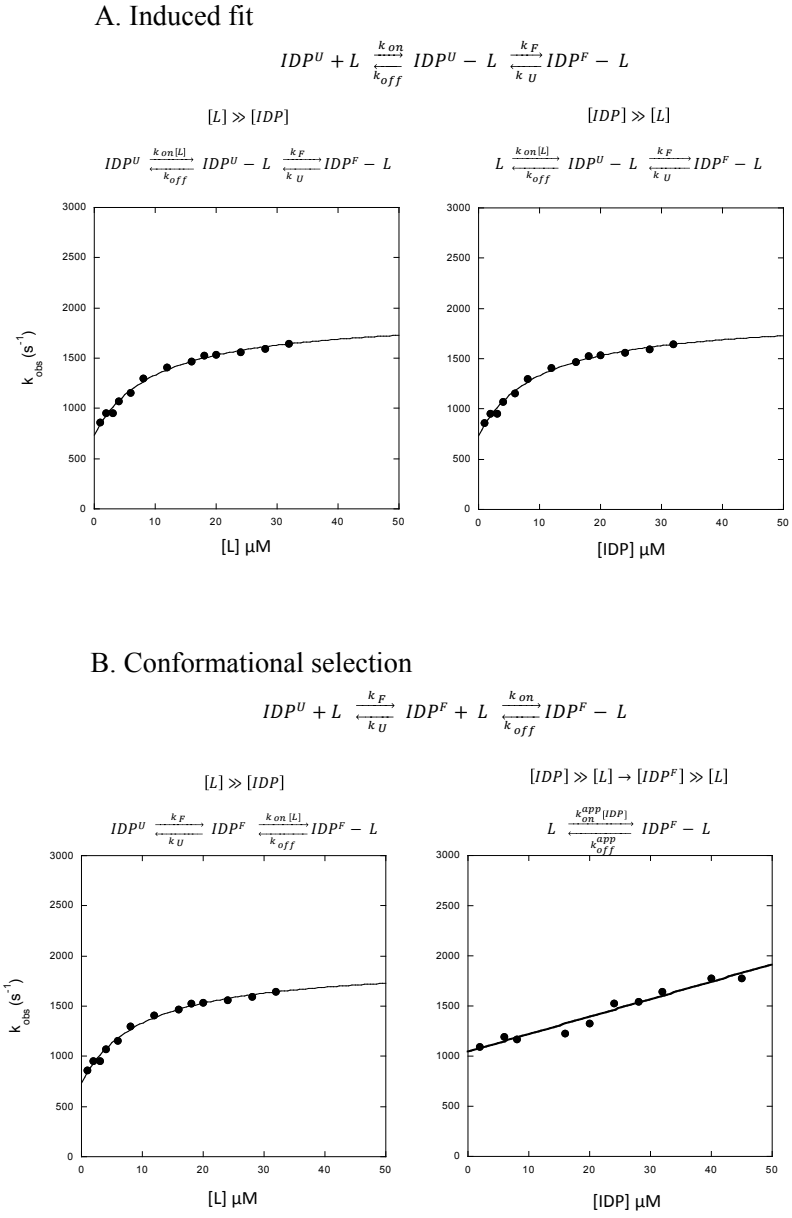


Figure 2. Schematic representation of induced fit (panel A) and conformational selection (panel B) mechanism under pseudo-first order conditions. IDP^U is the intrinsically disordered protein unfolded, IDP^F is the intrinsically disordered protein folded, L is the partner; k_f and k_u are the

folding and unfolding rate constant, k_{on} and k_{off} are the association and dissociation rate constant.

2.2. Φ-value analysis

Kinetic experiments are also used to obtain structural information about the binding and folding steps of IDPs/IDRs upon interaction with their partner(s), through the study of the transition state of the reaction.

According to the "transition state theory", introduced by Eyring in 1935, the transition state is the high free energy barrier through which the reactants are converted in products in a chemical reaction. The transition state may be described as an ensemble of conformations having the same probability to go back to the reactants or to move on to the products. By definition, the transition states never accumulate, so they can just be studied indirectly (Eyring, 1935). A powerful technique to obtain information about transition states of reactions is the Φ-value analysis. The Φ-value analysis, developed by Alan Fersht and co-workers in the late 1980s, is a method that, combining protein engineering and kinetic studies, allows inferring the structure of the transition state(s) and intermediate(s) in reactions such as protein folding, binding or catalysis, at the level of single residues (Fersht, 1995)(Fersht, Leatherbarrow & Wells, 1986)(Fersht, Leatherbarrow, & Wells, 1987)(Matouschek et al, 1989). This technique, which was initially introduced to describe the folding of barnase (Fersht, Matouchek & Serrano, 1992), is now widely used to study the folding mechanisms of small proteins. In particular, this strategy consists in the introduction of single amino acidic substitutions in the protein of interest and in the evaluation of the effect of the variation on the thermodynamics and kinetics of the system allowing interaction patterns in the transition state to be mapped out. The Φ-value is calculated as the ratio between the change in the free energy variation

between the denatured (D) (e.g. unfolded) state and the transition state (TS) and that between the unfolded state and the native state (N) upon mutation (see Figure 3):

$$\Phi = \frac{\Delta\Delta G_{D-TS}}{\Delta\Delta G_{D-N}}$$

Where the change in free energy variation between D and TS ($\Delta\Delta G_{D-TS}$) is defined as:

$$\Delta\Delta G_{D-TS} = \Delta G^{wt}_{D-TS} - \Delta G^{mut}_{D-TS}$$

$$\Delta\Delta G_{D-TS} = -RTln\frac{k^{wt}_F}{k^{mut}_F}$$

and the change in equilibrium free energy variation ($\Delta\Delta G_{D-N}$), is defined as follow:

$$\Delta\Delta G_{D-N} = \Delta G^{wt}_{D-N} - \Delta G^{mut}_{D-N}$$

$$\Delta\Delta G_{D-N} = RTlnK^{mut}_{eq} - RTlnK^{wt}_{eq}$$

$$\Delta\Delta G_{D-N} = RTln\frac{k^{mut}_F}{k^{mut}_U}\frac{k^{wt}_U}{k^{wt}_F}$$

Keq is the equilibrium constant defined as the ratio between k_F and k_U, k_F and k_U are respectively the rate constant of folding and unfolding, experimentally obtained for the *wt* and the mutants through kinetic experiments.

The Φ-values range from 0 to 1: a Φ-value equal to 0 means that the variation perturbs the transition state in the same way as the denatured state, while a Φ-value equal to 1 indicates that the effect of the substitution is the same in the transition state and in the native state. In other words, the Φ-value is a measure of the degree of nativeness of the TS. If the introduced substitution concerns a residue that is involved in the TS (i.e. a residue in a region that starts to fold in the TS), this will give rise to a Φ-value close to 1. Conversely, if the substitution pertains a residue that does not participate to folding (i.e. a residue that is located far from the region that starts to fold in the TS) then this will lead to a Φ-value close to 0 (see also below).

Usually conservative substitutions are introduced in order to induce small perturbations: buried hydrophobic side chains are truncated in order to avoid altering the stereochemistry (i.e., Ile/Val; Ala/Gly; Leu/Ala; Thr/Ser; Phe/Ala) (Fersht & Sato, 2004) and to rule out the possibility that the observed effect is imputable to the variation *per se*.

In the case of folding, where the reactant is the denatured state (D) and the product is the native state (N), the transition state (TS) represents an ensemble of conformations with the same probability to fold or unfold (Fersht, 2000) and the Φ-value indicates whether a residue is structured in the transition state in the same way as in the native state (Φ-value equal or close to 1) or whether it is as unstructured as in the denatured state (Φ-value equal or close to 0) (Figure 3).

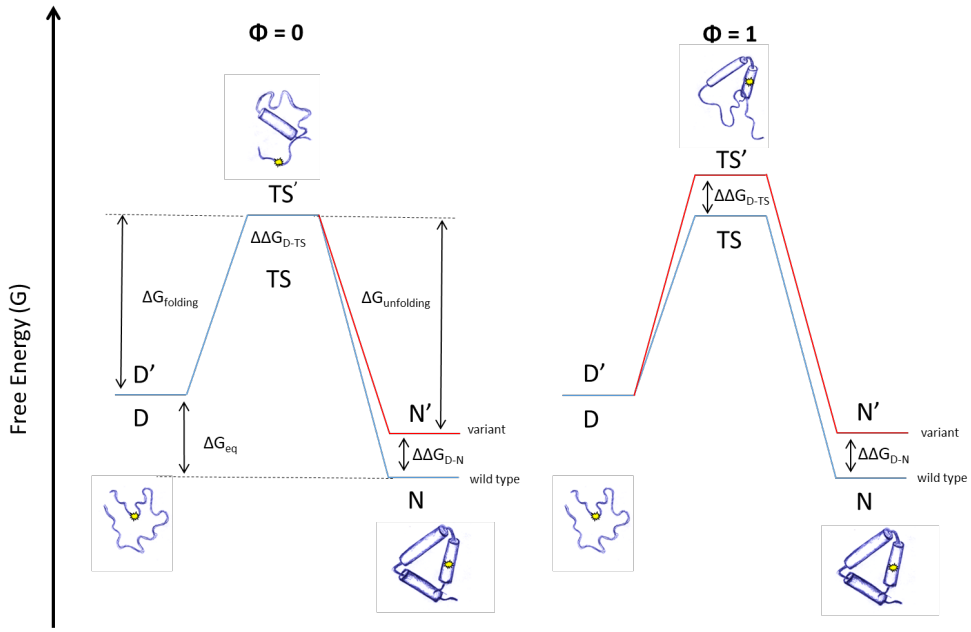


Figure 3. Schematic representation of the Φ-value analysis of a folding reaction. D and D' are the denatured state of the *wt* and the variant; N and N' are the native state of the *wt* and the variant; TS and TS' are the transition state of the *wt* and the variant.

Calculating the Φ-values for each variant allows drawing the structure of the transition state, mapping the residues with high, low or intermediate Φ-values on the structure of the protein to assess the degree of similarity with the native state and hypothesize the mechanism through which folding occurs.

The experimentally measured Φ-values can be also used as restrains to perform molecular dynamics simulations.

2.2.1. Φ-value analysis: IDPs interactions

The Φ-value analysis is also used to characterize the folding upon binding of IDPs/IDRs (Figure 4).

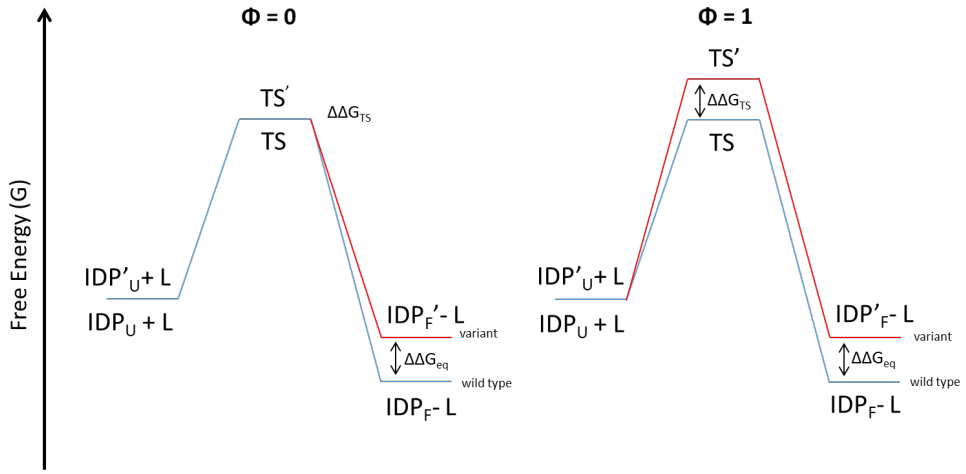


Figure 4. Schematic (simplified) representation of the Φ-value analysis of a folding upon binding reaction. IDP_U and IDP_F are the disordered protein unfolded and folded respectively. IDP'_U is the IDP, bearing the amminoacidic substitution, unfolded; TS' is the TS of the variant; IDP'_F is the IDP, bearing the amminoacidic substitution, folded.

In the case of IDPs/IDRs, the reactants are the disordered protein and its partner and the product is the complex with the folded IDP/IDR. The scheme reported in Figure 4 is an extreme simplification. In fact, in this case, the analysis is harder because the significance of the Φ-values changes depending on which step (i.e. the folding or the binding), is the rate-limiting step (representing the highest free energy barrier of the reaction that defines the transition state) of the reaction. To increase the complexity, most of the times is not possible distinguishing, experimentally, which is the limiting step of the reaction since, usually, a simple (apparent) two-state mechanism describing the overall process is observed. Thus, it is useful, before

performing a Φ-value analysis, assessing whether the mechanism is induced fit or conformational selection. In the conformational selection mechanism, the IDP/IDR folds before the binding with the partner so, if the folding is the rate-limiting step, the Φ-value analysis acquires the same significance of a classical folding analysis and does not give information about the stability of the complex. When the rate-limiting step is the binding, high Φ-values (equal or close to 1) indicate that the residue is folded in the transition state and fractional Φ-values (comprised between 0 and 1) indicate that the residue, in the transition state, is involved in the interaction with the partner.
In the induced fit mechanism, the folding occurs after the binding. In this case, when the binding is the rate-limiting step, in the transition state the IDP/IDR will be mostly unfolded, because the folding can't occur before the limiting step, and it will be represented by low Φ-values (equal or close to 0). When the folding is the rate-limiting step, the Φ-value has the same meaning of a classical folding analysis (Gianni, Dogan, & Jemth, 2014b).

Therefore, the Φ-value analysis allows obtaining information on the folding upon binding mechanisms of IDPs/IDRs since it consents hypothesizing the molecular mechanism that govern the process and allows identifying, for example, the residues that are important for the initial interaction or that fold before others.

2.2.2. Linear Free Energy Relationship

An additional classical procedure to assess the structural features of the transition state, based on the comparison of kinetic data obtained from the *wt* protein and its variants, consists in performing the Linear Free Energy Relationship analysis (LFER plot). The LFER plot is obtained by plotting the change in free energy in the transition state as a function of the change in free energy in the equilibrium state. LFERs were originally introduced to assess

the position of the transition state along the reaction coordinate during the formation of a covalent bond (Leffler, 1953). Upon altering the structure and thus the reactivity of a substrate, the dependence of activation free energy on ground-state free energy generally yields a linear profile. The slope of the observed correlation, called α, reflects the position of the transition state along the reaction coordinate. A linear LFER plot is a typical index that the folding occurs following the so-called nucleation-condensation mechanism whereby the whole protein self-assembles around a weakly formed nucleus and the transition state resembles a distorted version of the native state. The α value reflects the degree of similarity between the native and transition state. A high α value indicates that the transition state resemble a similar, although distorted, version of the native state (in the case of folding upon binding reaction the native state is the bound, folded state) while, a low α value indicates that the transition state is closer to the denatured state (in the case of a folding upon binding reaction the denatured state is the unbound, unfolded state).

3. Templated folding

A recent finding is that IDPs/IDRs could adapt their folding process depending on the structure of the partner. Φ-value analysis and molecular dynamic simulations studies of the folding after binding interaction between the intrinsically disordered c-Myb and the KIX domain revealed a structural malleability of the transition state dictated by the topology of KIX (Toto et al., 2016). This behavior suggested that IDPs/IDRs could fold via a templated mechanism involving a heterogeneous nucleation process induced by the structure of the partner.

This mechanism is in contrast with the so-called homogeneous nucleation typically observed for globular proteins, where the transition state is similar to the native state presenting an elevated fraction of native-like contacts.

This property increases the specificity of the molecular recognition process by which IDPs/IDRs interact with their partners avoiding miss-interactions.

4. Fuzziness in IDP complexes

The regions in IDPs/IDRs that undergo a disorder-to-order transition upon binding to a partner (i.e. MoREs) are usually short regions, while a considerable part of the sequence remains disordered also in the bound complex. These regions are called "fuzzy" and this type of complexes are called "fuzzy complexes". Fuzzy complexes can be classified in four structural categories: in the polymorphic complexes the IDP/IDR assumes different conformations on the partner surface (Renault et al., 2008) in the flanking complexes fuzzy regions close to the bound motifs provide additional contacts to the interaction increasing the binding affinity (Billeter et al. 1993); in the clamps complexes the fuzzy regions don't bind the partner and connect globular binding domains or motifs interacting with the partners (Clerici et al., 2009). Lastly, in the random complexes, the IDP/IDR remains completely disordered also upon the binding (Pometun, Chekmenev, & Wittebort, 2004). These four general categories can also be present in a same protein in different combinations.

Fuzzy regions also influence the interaction between IDP/IDR and partners enhancing or decreasing the binding through various mechanisms. In the conformational selection, fuzzy regions modulate the conformational equilibrium, stabilizing interaction-prone structural elements and moving the conformational equilibrium prior to binding (Adams, McBryant, Wade,

Woodcock, & Hansen, 2007)(Ghosh et al., 2010). In the flexibility/entropy modulation, fuzzy regions influence the flexibility and energy landscape of the binding interface performing transient contacts leading either to a decrease of the entropic cost of the binding and to an increase in the affinity (Pufall et al., 2005)(Lee et al., 2008). In the competitive binding, fuzzy regions compete with the binding partner, blocking the access to the binding surface, through electrostatic interactions or steric hindrance, resulting in a sort of auto-inhibition of the interaction (Stott, Watson, Howe, Grossmann, & Thomas, 2010). Finally, in the tethering mechanism, fuzzy regions enhance the binding affinity increasing the local concentration of globular, low-affinity binding domains (Daughdrill, Narayanaswami, Gilmore, Belczyk, & Brown, 2007)(Vise, Baral, Latos, & Daughdrill, 2005)(Borcherds, Becker, Chen, Chen, & Daughdrill, 2017). Therefore, the residual structural disorder in the bound form plays a role in the biological functionality of IDPs allowing a fine regulation of their activity and of that of the bound partner.

5. Frustration

Frustration in a system arises from the impossibility to achieve a specific goal. In folding, a system is considered frustrated when energetic contributions cannot be simultaneously minimized by a single conformation (Oliveberg & Wolynes, 2005). In general, proteins have the ability to find spontaneously a specific conformation, the native state, which is at the bottom of the funnel, according to the energy landscape theory. The energy landscape is the mathematical function that describes the intramolecular and solvation free energy of a given protein as a function of the microscopic degrees of freedom (Dill, Ozkan, Shell, & Weikl, 2008). The energy landscape theory was developed using tools from the statistical mechanics of disordered system and polymers (Onuchic & Wolynes, 2004). A model

simplified this theory leading to the conclusion that proteins have funnel-shaped energy landscapes since they present many high-energy states and few low-energy states. Energy landscapes are multidimensional surfaces where the vertical axis of the funnel represents the free energy of a given chain configuration, the many lateral axis represent the conformational coordinates and each conformation is represented by a point on the energy surface (Dill & Chan, 1997). According to this model, the native state represents the bottom of the funnel (i.e. the free energy minimum) and the denatured state is a set of different configurations represented by all the points on the landscape surface (Figure 5). This funnel structure is only possible for those selected sequences that are chosen so that energetic conflicts are for the most part avoided and the native structure is more stable than expected for random associating residues: this hypothesis is known as the principle of minimal frustration (Ferreiro et al., 2010). Thus, proteins are usually minimally frustrated and this feature makes protein structure robust to mutations. However, sometimes, local frustration could be a functionally useful adaptation arising from random evolution (Ferreiro, Hegler, Komives, & Wolynes, 2007). Local frustration could be useful because it may sculpt protein dynamics for specific functions, for example a specific site frustrated in a monomeric protein may become less frustrated after association in a larger assembly, driving specific association (Wang & Verkhivker, 2003)(Papoian, Ulander, & Wolynes, 2003). Also enzymes show a certain degree of frustration in their catalytic sites (Ellerby et al., 1990)(Meiering, Serrano & Fersht, 1992); therefore, looking at frustrated sites can give a hint about functional constraints on the evolution of protein energy landscapes. Considering that proteins have a certain degree of frustration, their free energy landscape will be characterized by some local minima, which are likely to contain misfolded elements. It is possible to indicate frustrated sites,

experimentally, performing, for example the Φ-value analysis: negative Φ-values or Φ-values greater than 1 are usually a signature of frustrated sites.

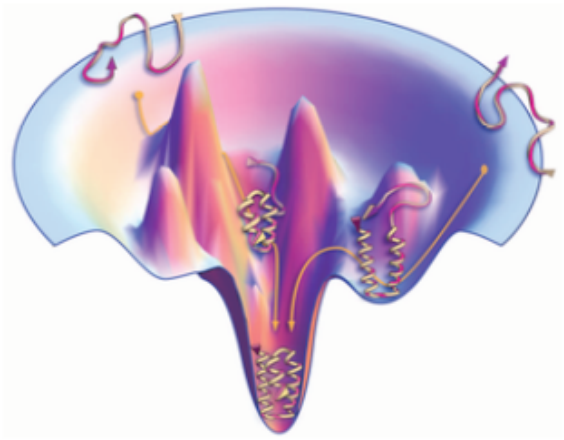


Figure 5. Schematic tridimensional representation of the energy landscape. The bottom of the funnel represents the low-energy native state and the other points of the surface are high-energy unfolded structures. Folding can occur via alternative microscopic trajectories (adapted from Dill & MacCallum, 2012).

6. Experimental system: Measles Virus N_{TAIL}

Protein disorder is highly abundant in the proteome of viruses. In addition to the biological functions carried out in eukaryotic organisms (e.g. protein-protein recognition and interaction, regulation and signal transduction, regulation of transcription and translation) the presence of disorder in viral proteins allows viruses to increase their adaptability to hostile environment and, give capability of IDPs/IDRs to bind to many different partners, confers them the ability to interact with multiple cellular factors and to highjack cellular processes at their own advantage (Tokuriki, Oldfield, Uversky, Berezovsky, & Tawfik, 2009). Moreover it was suggested

that protein disorder in viruses could be connected to the elevated mutational frequency typically occurring in viral genomes: spontaneous mutations are indeed expected to have a lower effect in a disordered protein than in a well-structured protein (Xue et al., 2010)(Uversky & Longhi, 2011).

An interesting and widely studied disordered domain is the C-terminal domain of the Nucleoprotein (N_{TAIL}) of Measles Virus. Measles Virus (MeV), a member of the *Paramyxoviridae* family, is a severe human pathogen responsible for the measles disease. The genome of MeV consists of a negative, single-stranded RNA, which is enveloped by the nucleoprotein (N) in a helical nucleocapsid. The nucleocapsid, and not naked RNA, is the substrate used by the RNA-dependent RNA polymerase (RdRp) for transcription of viral genes and replication of the genome. The RdRp is composed of the large (L) protein and the phosphoprotein (P). The P protein possesses a long N-terminal, intrinsically disordered domain (PNT amino acids 1-230) (Karlin, Longhi, & Canard, 2002a), and a C-terminal region (PCT, amino acids 231-507) organized in alternating structured and disordered modules (Karlin, Canard, & Longhi, 2003). The N protein consists of a structured N-terminal domain (N_{CORE}, amino acids 1-400) that is responsible for binding to RNA and self-assembly of N to yield the nucleocapsid (Karlin, Longhi, & Canard, 2002b)(Gutsche, Desfosses, Effantin, & Ling, 2015), and a C-terminal disordered domain (N_{TAIL}, amino acids 451-525) (Longhi et al., 2003) exposed at nucleocapsid surface. N_{TAIL} contains three regions conserved in *Morbillivirus* members named Box 1 (amino acids 401-420), Box 2 (amino acids 489-506), and Box 3 (amino acids 517-525) (Figure 6). These regions present a higher content in hydrophobic residues, which facilitates protein-protein interactions. Being exposed at the surface of the nucleocapsid, N_{TAIL} is able to interact with different partners: it binds the P viral protein but also some cellular factors

including the major inducible 70 kDa heat shock protein (hsp70) (Zhang et al., 2005), the interferon regulatory factor 3 (IRF3) (Colombo et al., 2009), peroxiredoxin 1 (Watanabe et al., 2011), the cell protein responsible for the nuclear export of N (Sato et al., 2006), casein kinase II, components of the cell cytoskeleton and also immune system factors.

During my PhD I have studied in particular the interaction between MeV N_{TAIL} and the P protein. N_{TAIL} interacts with the C-terminal X domain (XD, amino acids 459-507) of the P protein. This interaction is vital for the virus, leading to the recruitment of the polymerase complex onto the nucleocapsid template to allow transcription and replication to take place (Figure 7).

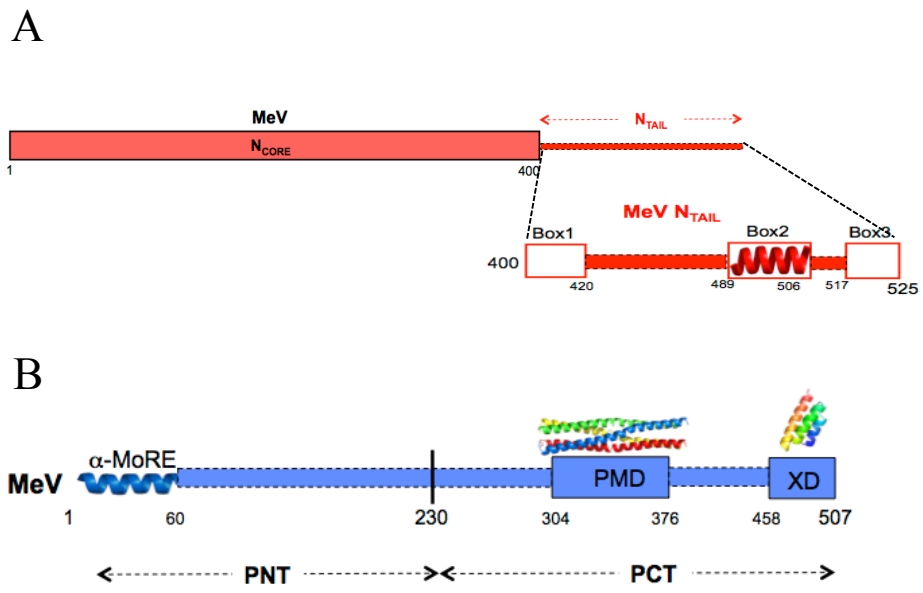


Figure 6. Schematic representation of the Nucleoprotein (A) and Phosphoprotein (B) of MeV. Figure adapted from (Longhi et al., 2017)

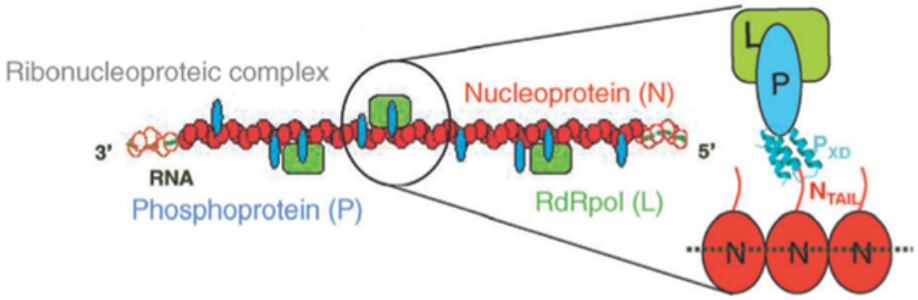


Figure 7. Schematic representation of the MeV helical nucleocapsid and of its interaction with the RNA-dependent RNA polymerase *via* the N_{TAIL}-XD interaction. Figure adapted from Habchi & Longhi 2012.

The N_{TAIL} domain interacts with XD *via* a short, order-prone molecular recognition element (α-MoRE, amino acids 486-502), located within Box 2 that, upon binding, undergoes an α-helical folding. The structure of the complex between MeV XD and the α-MoRE of N_{TAIL} is available (Protein Data Bank: 1T6O). XD is composed of three α helices; after the binding (and folding) of the α-MoRE, a four-helix complex is formed (Figure 8).

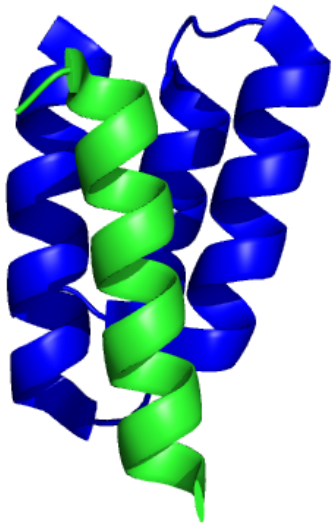


Figure 8. Representation of the three-dimensional structure of the four-helix complex between XD (in blue) and the α-MoRE (in green) of N_{TAIL} (aa 486-504). Protein Data Bank code: 1T6O

6.1. MeV N_{TAIL}-XD interaction: kinetic studies

Kinetic binding experiments were performed to study the interaction between MeV N_{TAIL} and XD. Since the reaction is very fast (milliseconds time scale), to avoid losing the reaction in the stopped-flow dead time, the temperature-jump apparatus was used. As mentioned before, and better explained in the Materials and Methods section, the temperature-jump (t-jump) is a relaxation technique that allows measuring ultra fast reactions. The t-jump, *via* a rapid increase in temperature, perturbs an equilibrium system (in this case the complex between two interacting proteins), and then, taking advantage of the change in the tryptophan fluorescence upon (un)binding, it

measures the relaxation to the new equilibrium, allowing the observed rate constant of the binding reaction to be calculated.

The kinetics of interaction between MeV N_{TAIL} and XD, performed under pseudo-first order conditions, showed a hyperbolic dependence of the k_{obs} irrespective of whether N_{TAIL} or XD was used as the excess ligand, a behavior that constitutes a signature of folding after binding (Figure 9) (Dosnon et al., 2015).

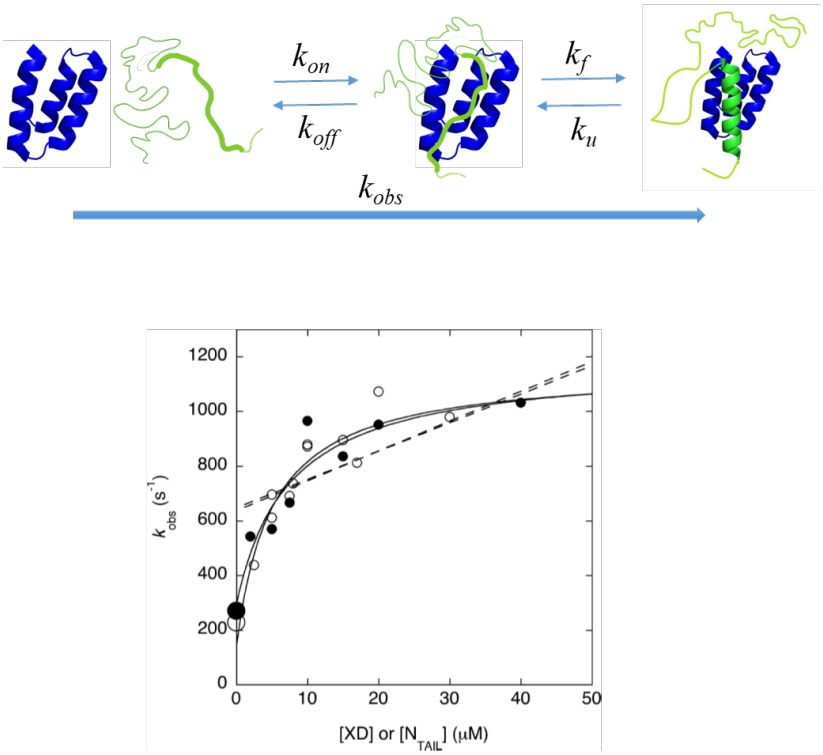


Figure 9. Top: schematic representation of the folding after binding process of the XD-N_{TAIL} interaction. k_{on} and k_{off} are the association and dissociation rate constants and k_f and k_u are the folding and unfolding rate constants. Bottom: hyperbolic dependence of the k_{obs} as a function of XD (open circles) or N_{TAIL} concentrations (fully circle). Figure from: Dosnon et al. 2015

The hyperbolic dependence of k_{obs} as a function of N_{TAIL} concentration allowed the binding, bimolecular step to be separated from the folding, monomolecular step (Dosnon et al., 2015).

An experiment performed in the presence of TFE (2,2,2-trifluoroethanol), a secondary structure stabilizer able to induce folding in peptides and proteins (Jasanoff & Fersht, 1994)(Myers, Pace, & Scholtz, 1998), showed that at high protein concentrations the reaction is limited by the folding rate of N_{TAIL}. In fact, upon increasing the TFE concentration (i.e. upon increasing the stabilization of the α-MoRE helix) a transition from a hyperbolic to a linear dependence of the k_{obs} was observed (Figure 10) (Dosnon et al., 2015).

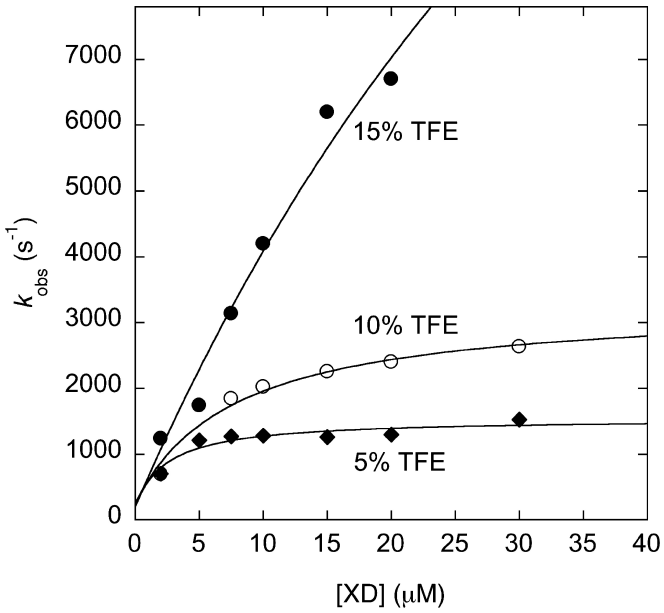


Figure 10. Kinetic binding experiment between N_{TAIL} and XD, under pseudo-first order conditions, in the presence of increasing concentrations of TFE (Figure from: Dosnon *et al.* 2015).

6.2. MeV N_{TAIL}-hsp70 interaction

MeV N_{TAIL} also interacts with the 70 kDa heat shock protein (hsp70) of the host cells. Hsp70 is an ubiquitously expressed protein involved in protein folding (acting as chaperone) and in the protection of the cell under stress conditions. Hsp70 is a large protein composed of three major functional domain: the N-terminal ATPase domain responsible for the interaction and hydrolysis of ATP; the substrate binding domain composed of a 15 kDa β-sheet subdomain and a 10 kDa helical subdomain; and a C-terminal domain rich in α-helices acting as a 'lid' for the substrate binding domain.

N_{TAIL} interacts with hsp70 through Box 2 and also, although with lower affinity, through Box 3 (Zhang et al., 2005)(Carsillo, Traylor, Choi, Niewiesk, & Oglesbee, 2006). Despite the affinity of N_{TAIL} towards hsp70 is lower (70 μM) (Couturier et al., 2010) than that towards XD (3 μM) (Dosnon et al., 2015), hsp70 competes with XD for the binding with N_{TAIL} thus reducing the stability of the N_{TAIL}-XD complex (Zhang et al., 2005).

6.3. Fuzziness in MeV N_{TAIL} complexes

After binding to XD, the N_{TAIL} regions close to the α-MoRE remain conspicuously disordered (Longhi, 2012). In a work published in 2016, Gruet and co-workers showed that gradual removal of the long N-terminal fuzzy appendage leads to an increase in the binding affinity between N_{TAIL} and XD (Gruet et al., 2016). In those studies, the authors made use of a protein

complementation assay based on split-GFP re-assembly (Magliery & Regan, 2006)(Wilson, Magliery & Regan, 2004). In this assay, XD, fused to the C-terminal domain of the GFP (Green Fluorescent Protein), is co-expressed in *E. coli* with N_{TAIL}, fused to the N-terminal domain of GFP. Upon interaction between N_{TAIL} and XD, and only in this case, the two GFP fragments can assemble in an irreversible manner, leading to reconstitution of the fluorophore and hence to a fluorescence signal that can be measured. The fluorescence is proportional to the binding strength between the two interacting proteins. Those studies demonstrated that the fuzzy region of N_{TAIL} decreases the binding strength between N_{TAIL} and XD.

The same experiments were also performed with hsp70: despite the lower affinity, and the likely different mechanisms of interaction, the truncation of the fuzzy region was found to lead to an increase in the binding strength with Hsp70 too.

The same results were also obtained in split luciferase reassembly assays. In this case, the reporter is luciferase, the assay is performed in eukaryotic cells, and the reassembly of the two halves of luciferase, leading to luminescence, is reversible. In addition, pull-down experiments that made use of XD fused to GFP and of purified, histidine-tagged N_{TAIL} either in its full-length or in its truncated form, confirmed that truncated N_{TAIL} has a higher affinity towards XD.

Thus, the fuzzy region in N_{TAIL} seems to compete with the partner, through mechanisms that are still to be unveiled, decreasing the binding affinity.

In the same work the authors showed that the residual disorder in N_{TAIL} also affects the folding rate of the α-MoRE. They performed kinetic binding experiments (under pseudo-first order conditions) between XD and either N_{TAIL} full-length or a synthetic peptide mimicking just the α-MoRE region of N_{TAIL}. In the case of full-length N_{TAIL}, a hyperbolic dependence of k_{obs} on ligand concentration was observed (the folding of N_{TAIL} becoming rate-limiting at high reactant concentrations). Conversely, when the MoRE peptide was used, a linear dependence was observed (Figure 11), meaning that the folding rate of the α-MoRE is increased, and the overall reaction can be described as schematized in Figure 12. This behavior suggests that the fuzzy appendage of N_{TAIL} influences the folding of the MoRE, lowering its rate constant.

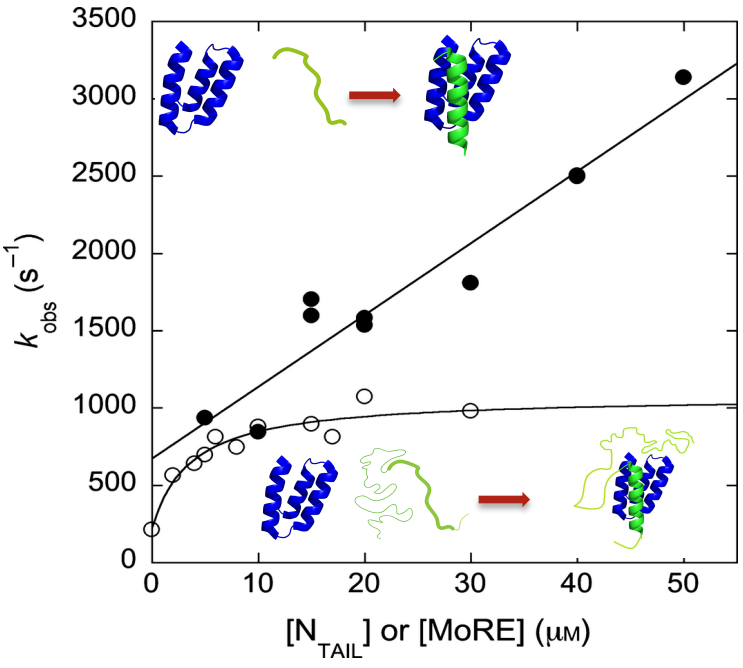


Figure 11. Kinetic binding experiments between XD and full-length N_{TAIL} (empty circles) and the α-MoRE peptide (full circles)**.** Figure modified from Gruet *et al.* 2016.

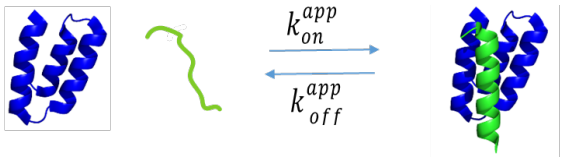


Figure 12. Schematic representation of the binding reaction between the α-MoRE and XD. The overall reaction is described by a simple two-state mechanism with an apparent association rate constant (k_{on}^{app}) and an apparent dissociation rate constant (k_{off}^{app}).

OBJECTIVES

The mechanisms regulating the interaction of intrinsically disordered proteins with their partners and the resulting disorder-to-order transitions are still far to be clearly understood.

In this thesis, using the C-terminal intrinsically disordered domain of the Nucleoprotein of Measles virus (N_{TAIL}) and its interaction with its partners as a model system, I will attempt at better understanding the mechanisms regulating the interaction between IDPs and partners, the folding pathway of Molecular Recognition Elements and the role of the regions retaining intrinsic disorder also upon binding.

The first goal of my thesis is the study of the molecular mechanism of the interaction between MeV N_{TAIL} and its partner XD (the C-terminal domain of the MeV phosphoprotein) at single residues level performing the Φ-value analysis. To this end, I will first design and produce a series of single site variants of MeV N_{TAIL} in the α-MoRE region (amino acids 488-505) and then will perform kinetics binding experiments between each variant of N_{TAIL} and its partner XD.

A work published in 2016 by Bonetti and co-workers showed that XD, in native conditions, populates an alternative state similar to an on-pathway intermediate retaining a native-like secondary structure but displaying some differences in the overall topological structure. It was observed that the substitution of the Isoleucine with Alanine in position 504 of XD leads to a variant that populates only the native state, i.e. devoid of the intermediate (Bonetti et al. 2016). Therefore, I will perform kinetic binding experiments using the XD I504A variant and the same single site N_{TAIL} variants used for the Φ-value analysis with *wt* XD, so as to assess whether the folding after binding mechanism of N_{TAIL} depends or not on the structure of the partner.

Subsequently, I will also produce a series of truncation variants of $N_{TAIL,}$ through gradual removal, by blocks of ten amino acids, of the region of N_{TAIL} that remains disordered in the complex with XD, so as to study the effect of the residual disorder on the rate of folding of the α-MoRE. Concomitantly, I will also assess which is the minimal region responsible for the shift from a hyperbolic to a linear dependence of the k_{obs} as a function of N_{TAIL} concentrations (see Figure 11). Once established whether and which part of the fuzzy region plays a role in the regulation of the folding rate of the α-MoRE, I will perform single site variation in this region and t-jump experiment with XD to assess the nature of the inhibitory effect of the fuzzy appendage. In particular, I will ascertain if the fuzzy region inhibits the folding of the MoRE through an entropic effect, merely due to the length of the sequence, or whether on the contrary it exerts an enthalpic effect due to specific aminoacidic interactions. Moreover, to assess the possible enthalpic effect of the fuzzy region on the rate of folding of the α-MoRE, I will perform kinetic binding experiment using an artificial variant of $N_{TAIL,}$ described in Gruet *et al.,* 2016, where all the amino acids in the region comprised between residues 401 and 480 are replaced with "disordered prone" residues, in order to obtain a different (i.e. highly dissimilar) and more disordered variant of N_{TAIL}.

Finally, I will also study two variants of the isolate α-MoRE region: AlaMoRE, a variant obtained by substituting with Alanine all the amino acids not essential for the interaction with XD, expected to be more ordered; and GlyMoRE, a variant obtained by substituting with Glycine all the amino acids not essential for the interaction with XD in order to obtain a more disordered sequence. Using far-UV CD spectroscopy I will assess whether the content in secondary structure in AlaMoRE and in GlyMoRE is increased and decreased, respectively, with respect to the *wt* α-MoRE. Moreover, I will

perform kinetics binding experiments with *wt* XD to assess whether a different content in (dis)order of the MoRE region affects its folding rate.

MATERIALS AND METHODS

1. Constructs generation

1.1. N_{TAIL} site-directed variants

The constructs encoding the site directed variants of N_{TAIL} were obtained using the gene encoding *wild-type* (*wt*) N_{TAIL}, inserted in the pET28 expression vector, as template to perform site-directed mutagenesis using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent technologies) according to the manufacturer's instructions. All substitutions were conservative and were confirmed by DNA sequencing.

1.2. N_{TAIL} truncation variants

The constructs encoding N-terminally heaxahistidine tagged N_{TAIL} truncation variants (from 401 to 481) with a tobacco etch virus (TEV) protease cleavage site (amino acidic sequence: ENLYFQGS) immediately upstream of the codons encoding amino acids 401, 411, 421, 431, 441, 451, 461, 471 and 481 of N_{TAIL} were already available (Gruet *et al*. 2016).

The N_{TAIL} truncation variants were also cloned into the pETG20A expression vector that drives the expression of the protein of interest fused to a Thioredoxin (Trx) solubility tag. To this end, the already available pDEST14 constructs encoding N-terminally hexahistidine tagged $N_{TAIL,}$ with a TEV protease cleavage site at different N_{TAIL} positions, were used as a templates in a first PCR amplification that used *att*B2 as reverse primer and as forward primer an oligonucleotide encompassing the *att*B1 site followed by the sequence encoding the TEV protease cleavage site, and by a sequence annealing with the 17 N_{TAIL} nucleotides downstream of the TEV cleavage site. After digestion with DpnI (1 hour at 37°C), 1 µl of the first PCR reaction was used as template for a second PCR amplification performed using primers *att*L1a and *att*L2a. These primers encode approximately half of the *att*L1 and *att*L2 recombination sites, respectively, and have been shown

to be good substrates for LR clonase (Fu et al., 2008). As the *att*L1a and *att*L2a sequences encompass the *att*B1 and *att*B2 sequences, respectively, amplicons flanked by *att*B sequences as obtained from the first PCR product provide the necessary homology for *att*L1a and *att*L2a primer hybridization, respectively. 1.5 µl of the second PCR product were used in a 5 µl LR reaction to insert the truncation variant sequences into the pETG20A plasmid by Gateway cloning. These steps were performed to produce all the truncation variant constructs bearing the Trx tag.

1.3. Artificial N_{TAIL}

An already available construct encoding the artificial N_{TAIL} sequence cloned into the pNGG plasmid (Gruet et al. 2016) was used as a template in a first PCR amplification aimed at inserting the TEV protease cleavage site upstream residue 401. The forward primer contained the *att*B1 site, followed by the sequence encoding for the TEV protease cleavage site and by 24 nucleotides annealing with the 5' region of artN_{TAIL}. The reverse primer was *att*B2. After digestion with DpnI, the artificial amplicon bearing the TEV cleavage site was used as a template for a second PCR amplification with primers attL1a and attL2a. The resulting amplicon was then inserted in the pETG20A plasmid through an LR reaction, as described for the truncation variants.

1.4. X Domain: Y480W and I504A variants

The construct encoding XD with a C-terminal hexahistidine tag and bearing the Y480W substitution (referred to as pseudo-*wt*), has been already described (Dosnon et al. 2015), as was the construct encoding pseudo-*wt* bearing the additional I504A substitution (Bonetti et al. 2016).

2. Protein expression and purification

2.1. N_{TAIL} site-directed variants

All variants were expressed in the *Escherichia coli* Rosetta T7 pLysS (Novagen) strain. Cultures were grown in Luria Bertani (LB) medium containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol at 37°C until the OD_{600} reached 0.6-0.8 and then protein expression was induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside). After induction, cells were grown at 25°C over night and then collected by centrifugation. Cells were resuspended in 5ml/gr of pellet of buffer A (50 mM sodium phosphate pH 7.2, 300 mM NaCl) supplemented with 0.1 mg/mL lysozyme, 10 µg/mL DNase I and protease inhibitor cocktail (Sigma) and disrupted by sonication. The lysate was clarified by centrifugation at 12,000 rpm for 40 minutes at 4 °C. The soluble fraction bacterial lysate was loaded onto a 5 ml His trap FF column (GE, Healthcare) preloaded with Ni^{2+} ions, previously equilibrated in buffer A. After a washing step with buffer A supplemented with 1 M NaCl, the proteins were eluted with buffer A containing 250 mM imidazole. The proteins were then concentrated until 1 – 2 ml and loaded onto a Superdex 75 16/60 column (GE, Healthcare). Isocratic elution was carried out in SEC buffer (10 mM sodium phosphate pH 7.0, 150 mM NaCl). The purity of the proteins was confirmed by SDS-PAGE and the concentration was calculated measuring the absorbance at 280 nm, using the theoretical extinction coefficient, as provided from ExPASy (https://www.expasy.org/), and applying the Lambert-Beer equation.

2.2. N_{TAIL} truncation variants and artN_{TAIL}

The N_{TAIL} truncation variants and artN_{TAIL} were expressed in Terrific Broth (TB) medium containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol. The *E. coli* strain Rosetta T7 pLysS (Novagen) was used. Cells were grown in the same conditions described for the single-site N_{TAIL} variants except that cells were collected by centrifugation after 4 hours at 37°C.

Cell lysis was carried out as described above and the soluble fraction obtained after centrifugation was loaded onto a 5 ml His trap FF column (GE, Healthcare) as described for the single site variants. In this case, the buffer used to equilibrate the column was 50 mM Tris/HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM phenyl-methyl-sulphonyl-fluoride (PMSF). After sample injection, the column was washed with 50 mM Tris/HCl pH 8.0, 1 M NaCl, 20 mM Imidazole. Elution was carried out with 50 mM Tris/HCl pH 8.0, 300 mM NaCl, 250 mM Imidazole. Eluents from IMAC were desalted using a desalting column (GE, Healthcare) previously equilibrated with desalting buffer (20 mM Tris/HCl pH 8.0, 300 mM NaCl). Cleavage with TEV protease was carried out overnight at 4°C in desalting buffer using 1 mg of TEV protease per 20 mg of target protein. Samples were then loaded onto 2.5 ml Ni^{2+} IDA Agarose resin, previously equilibrated in desalting buffer, and incubated 15 minutes at 4°C. The non-retained fraction was recovered, concentrated up to 1 – 2 ml, and then loaded onto a Superdex 75 16/60 column (GE Healthcare) and eluted in SEC buffer. The purity of the protein was confirmed by SDS-PAGE.

2.3. XD variants

The pseudo-*wt* Y480W XD variant and the I504A XD variant were expressed in *E. coli* BL21 (DE3) cells. Cultures were grown in Luria Bertani

(LB) medium containing 100 µg/ml ampicillin at 37°C until the OD_{600} reached 0.6 and then protein expression was induced with 1 mM IPTG. After induction, cells were grown at 25°C for 2 days and then collected by centrifugation. Cells were resuspended in buffer A (50 mM sodium phosphate, 300 mM NaCl) supplemented with protease inhibitor cocktail (Sigma). The lysate was clarified by centrifugation at 12,000 rpm for 40 minutes at 4 °C.

The soluble fraction of the bacterial lysate was loaded onto a 5 ml His trap FF column (GE, Healthcare). After a washing step with buffer A supplemented with 1 M NaCl, the proteins were eluted with buffer A containing 1 M imidazole. The eluted protein was then desalted with a desalting column (GE, Healthcare) using buffer A. The purity of the protein was verified by SDS-PAGE.

2.4. α-MoREs peptides

Peptides mimicking the *wt* MoRE and its variants enriched in Alanine (AlaMoRE, supposed to be more ordered) and in glycine (GlyMoRE, supposed to be more disordered) were purchased from JPT (Berlin, Germany). The MoRE peptides were designed to encompass residues 486-504 of N_{TAIL}, and possessed an additional non-native tyrosine at the C-terminus to allow their concentration to be inferred from the absorbance at 280 nm using the Lambert-Beer equation.

3. Modeling

The structural models of the free form of the three MoRE peptides were obtained using the PEP-FOLD3 server (http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD3, Lamiable et al. 2016).

4. Circular dichroism measurements

The far-UV CD spectra of the MoRE peptides were measured using a Jasco 810 dichrograph, flushed with N_2 and equipped with a Peltier thermoregulation system. Peptides were dissolved in 10 mM sodium phosphate pH 7.0 at a concentration of 0.1 mg/mL in a final volume of 300 µl. One-mm thick quartz cuvettes were used. 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. Spectra were smoothed using the ‘‘means-movement’’ smoothing procedure implemented in the Spectra Manager package. The BESTSEL website (http://bestsel.elte.hu/, 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) was used to analyze the experimental data in the 190–250 nm range.

5. Temperature-jump measurements

The interaction between MeV N_{TAIL} and XD is an extremely rapid reaction. Thus, the classical stopped-flow mixing methods cannot be use because of the dead-time of about 1-2 ms occurring in this technique. Therefore, to study the kinetics of interaction between N_{TAIL} and XD, the temperature jump (t-jump) method, a kinetic technique that allows measuring ultra rapid reactions, was used. The t-jump methodology is a relaxation method whereby the solution containing a mixture of the interacting proteins at equilibrium is perturbed by a rapid increase in temperature. After trigger, the relaxation to the new equilibrium is observed. The rapid heating is due to the electrical discharge in the solution of high voltage of tens of kV, which is capable of inducing an increase in temperature of typically 5-9°C within few

microseconds. The experiments were performed under pseudo-first order conditions, i.e. with one of the two interacting proteins in large excess. In these conditions, the reaction rate will depend solely on the concentration of the reactant present at low concentration, which simplifies quantitative analysis of the data. In practice a constant concentration (5 μM) of pseudo-*wt* Y480W XD or of pseudo-*wt* XD bearing the additional I504A substitution (referred to as I504A XD) was mixed with different N_{TAIL} (*wt* or variants) concentrations usually ranging from 2 μM to 50 μM. The buffer used was 10 mM sodium phosphate pH 7.0, 150 mM NaCl. The presence of salt is required to ensure conductivity. Each sample was degassed and slowly pumped through the 0.5x2 mm quartz flow cell of a Hi-Tech PTJ-64 capacitor-discharge Temperature-jump apparatus (Hi-Tech, Salisbury, UK) showed in Figure 1. In the cell, a rapid discharge of 12 kV leads to a temperature jump of 9°C (from 16°C to 25°C) perturbing the equilibrium as schematizes in Figure 2. The change in tryptophan fluorescence, caused by the perturbation of the equilibrium and the consequent relaxation to the initial temperature, is followed as a function of time. Usually 10-20 traces were averaged for each sample. The excitation wavelength was 296 nm and the fluorescence emission of the tryptophan was measured using a 320 nm cut-off glass filter. The fluorescence change of *N*-acetyl-tryptophanamide was used in control measurements to monitor the timescale of the rapid heating. In fact, the quantum yield of this amino acid is highly dependent on temperature and represents a classical and simple tool to evaluate the dead-time of t-jump devices. Under the condition explored, heating was completed after 20 ms. The data below 20 ms (reporting on the pre-trigger and the heating phase) were excluded from the analysis of the kinetic data.

T-jump experiments were performed between pseudo-*wt* Y480W XD and truncation variants, site-directed N_{TAIL} variants mutated in the 451-482

region and *wt* MoRE, AlaMoRE and GlyMoRE peptides from one hand, and between the I504A XD variant and the site directed N_{TAIL} variants mutated in the region of the α-MoRE (aa 486-502).

Figure 1. Temperature-jump apparatus.

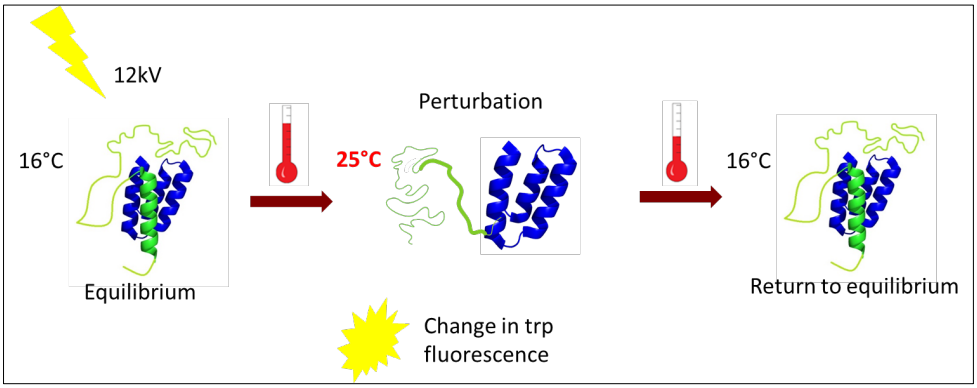


Figure 2. Schematic representation of temperature-jump experiment.

6. Data analysis

The fluorescence time courses obtained for N_{TAIL} truncation variants, N_{TAIL} site-directed variants and α-MoREs peptides were fitted by using the following single exponential equation:

$$F(t) = \Delta F \cdot e^{(k_{obs} \cdot t)} + F_{\infty}$$

where k_{obs} is the observed rate constant, ΔF is the amplitude of the trace, t is the time of acquisition of the trace and F_{∞} is the final value of the fluorescence. Observed rate constants were plotted as a function of ligand concentration and fitted with the following hyperbolic function, arising from an induced-fit model:

$$k_{obs} = \frac{[N_{TAIL}] \cdot k_F}{[N_{TAIL}] + K'_D} + k_U$$

where k_F is the folding rate constant and k_U is the unfolding rate constants of N_{TAIL} and K'_D is the dissociation constant of the initial complex between unfolded N_{TAIL} (or MoRE peptide) and XD (Bonetti et al., 2017).

Observed rate constants showing a linear dependence as a function of protein concentrations were fitted with the following linear equation:

$$k_{obs} = k_{on}^{app} \cdot [N_{TAIL}] + k_{off}^{app}$$

where k_{on}^{app} is the apparent association rate constant and k_{off}^{app} is the apparent dissociation rate constant.

RESULTS AND DISCUSSION

Paper 1.

Biochemistry

Article

pubs.acs.org/biochemistry

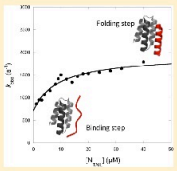
Analyzing the Folding and Binding Steps of an Intrinsically Disordered Protein by Protein Engineering

Daniela Bonetti,[†] Francesca Troilo,[†,‡] Angelo Toto,[‡] Maurizio Brunori,[†] Sonia Longhi,[‡] and Stefano Gianni*[,†]

[†]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

[‡]Aix-Marseille Univ, CNRS, Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, 13288 Marseille, France

ABSTRACT: Intrinsically disordered proteins (IDPs) are functionally active despite lacking a well-defined three-dimensional structure. Such proteins often undergo a disorder-to-order transition, or induced folding, when binding to their specific physiological partner. Because of cooperativity, the folding and binding steps typically appear as a single event, and therefore, induced folding is extremely difficult to characterize experimentally. In this perspective, the interaction between the disordered C-terminal domain of the measles virus nucleoprotein N_{TAIL} and the folded X domain of the viral phosphoprotein (XD) is particularly interesting because the inherent complexity of the observed kinetics allows characterization of the binding and folding steps individually. Here we present a detailed structural description of the folding and binding events occurring in the recognition between N_{TAIL} and XD. This result was achieved by measuring the effect of single-amino acid substitutions in N_{TAIL} on the reaction mechanism. Analysis of the experimental data allowed us (i) to identify the key residues involved in the initial recognition between the two molecules and (ii) to depict the general features of the folding pathway of N_{TAIL}. Furthermore, an analysis of the changes in stability obtained for the whole set of variants highlights how the sequence of this IDP has not been selected during evolution to fold efficiently. This feature might be a consequence of the weakly funneled nature of the energy landscape of IDPs in their unbound state and represents a plausible explanation of their highly dynamic nature even in the bound state, typically defined as "fuzziness".



One of the most surprising findings of the past decade is that many functionally active proteins lack a well-defined three-dimensional structure in isolation.[1−8] These proteins, denoted intrinsically disordered proteins (IDPs), often contain a so-called molecular recognition region that undergoes a disorder-to-order transition when binding a physiological partner.[3,9] Disorder has been suggested to be somewhat advantageous for the function and interactivity of some proteins by expanding, for example, the binding repertoire of folded proteins, by implementing their capacity to bind multiple partners and to fold into alternative conformations.[1,3,8,10] In addition, protein disorder could represent a successful strategy for decoupling binding affinity and selectivity; destabilization of the native structure, leading to unfolding, lowers the affinity for the ligand because of the increased entropic penalty, without necessarily altering specificity,[11,12] a model that has been challenged, and partly disproved, by Lah and co-workers.[13] While an experimental assessment of these hypotheses implies a deep knowledge of the system, our current understanding of the mechanism of induced folding is relatively limited, and to date, only a few proteins have been characterized to a sufficiently informative level of resolution.[14,15]

A powerful strategy for inferring the mechanism of a chemical reaction is to provide a structural characterization of each reaction intermediate(s) and the intervening transition state(s). In this context, the induced folding of an IDP upon binding to its physiological partner is characterized in theory by at least two steps: formation of a complex between the two molecules and induced folding.[16] However, it has been previously reported that folding and binding tend be cooperative, such that the overall reaction typically occurs in an all-or-none fashion and a single-exponential decay is often observed by time-resolved spectroscopy.[14,17−20] Such a cooperativity is reminiscent of what is typically detected in the folding of globular proteins[21] and represents an inherent limitation for the experimentalist; while the aim lies in defining the sequence of events occurring along the reaction path, only reagents and products may be characterized.

We recently investigated the interaction between the intrinsically disordered C-terminal domain of the measles virus nucleoprotein (N_{TAIL}; consisting of 124 residues from position 401 to 525) and the X domain (XD) of the viral phosphoprotein.[22] While XD consists of a triple-α-helical bundle, N_{TAIL} is an IDP that undergoes α-helical folding upon binding to XD,[23−27] with the resulting complex corresponding to a pseudo-four-helix fold. The disorder-to-order transition of N_{TAIL} involves a region of the protein denoted as box2 (encompassing residues 489−506). From a kinetic point of view, the fortuitous hyperbolic dependence of the observed

Received: April 18, 2017
Revised: June 29, 2017
Published: June 29, 2017

DOI: 10.1021/acs.biochem.7b00350
Biochemistry 2017, 56, 3780−3786

macroscopic rate constant on reactant concentration allowed us to isolate the folding (monomolecular) and binding (bimolecular) steps.[22] This finding, together with *ad hoc* designed experiments performed in excess of either reactants,[28,29] highlighted some clear signatures of the induced folding scenario. On the basis of these experiments, we concluded N_{TAIL} to fold only after binding XD.

In this work, we present a detailed description of the structural features of the binding and folding steps in the recognition reaction between N_{TAIL} and XD, achieved by Φ value analysis. By this technique, residue-specific structural information is inferred by comparing the kinetics of the reaction (folding and/or binding) of the wild-type protein with those of a series of conservative single variants, yielding the so-called Φ value that represents an index of nativelike structure of the mutated residue in each relevant state.[30,31] This technique, which was originally introduced to describe the structure of a metastable intermediate in the folding of barnase,[30] is also perfectly suited to characterization of the encounter between biomolecules and is therefore potentially very useful for the characterization of the folding upon binding of IDPs. Mutational kinetic data allow the discrimination of the residues that are critical for the initial recognition between the two molecules from those mainly involved in the folding of N_{TAIL}. Furthermore, analysis of the changes in the free energy of the binding and folding steps highlights how the amino acidic sequence of this IDP has not been selected during evolution to fold efficiently. As discussed below, this finding might be a consequence of the weakly funnelled nature of the energy landscape of IDPs in isolation and represents a plausible explanation of their highly dynamic nature, often termed "fuzziness".[12]

MATERIALS AND METHODS

Expression and Purification. Experiments were performed on a fluorescent pseudo-wild-type XD variant Y480W. N_{TAIL} mutants were obtained using the QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies) according to the manufacturer's instructions, and mutations were confirmed by DNA sequencing. All proteins were expressed and purified as described previously.[22] All reagents were of analytical grade.

Temperature-Jump Binding Experiments. Relaxation binding experiments were performed using a Hi-Tech PTJ-64 capacitor-discharge T-jump apparatus (Hi-Tech, Salisbury, U.K.). Pseudo-wild-type XD Y480W at a constant concentration of 2 μM was mixed with wild-type N_{TAIL} and its mutants at different concentrations, ranging from 1 to 40 μM. The temperature was rapidly changed with a jump size of 9 °C, from 16 to 25 °C. The change in the fluorescence of *N*-acetyltryptophanamide (NATA) was used in control measurements. Degassed samples were slowly pumped through the 0.5 mm × 2 mm quartz flow cell before data were acquired. Usually, 10−20 individual traces were averaged. The excitation wavelength was 296 nm, and the fluorescence emission was measured using a 320 cutoff glass filter. The buffer that was used consisted of 50 mM sodium phosphate (pH 7.2) in the presence of 300 mM NaCl. The presence of NaCl in each kinetic experiment was needed to ensure the conductivity of the solution and allow the temperature jump to occur upon the rapid discharge of the capacitor. Furthermore, NaCl was also observed to increase the solubility of some variants of N_{TAIL}. Importantly, a comparison with previous experiments in the absence of NaCl did not reveal any relevant effect of the observed affinity of XD for N_{TAIL}.[22,26]

Data Analysis. The observed binding time courses of wild-type N_{TAIL} and its site-directed mutants were fitted to a single-exponential decay to obtain the rate constant at each protein concentration. The rate constants were then plotted as a function of protein concentration and fitted to the following hyperbolic function arising from an induced fit model:

$$k_{obs} = \frac{[N_{TAIL}]}{[N_{TAIL}] + K'_D} \times k_F + k_U \quad (1)$$

where k_F and k_U are the folding and unfolding rate constants of N_{TAIL}, respectively, and K'_D represents the dissociation constant of the initial encounter complex. For each variant, $\Delta\Delta G_{K'}$ and $\Delta\Delta G_{folding}$ were than calculated from the equations

$$\Delta\Delta G_{K'} = RT \times \ln \frac{K'^{wt}_D}{K'^{mut}_D} \quad (2)$$

$$\Delta\Delta G_{folding} = RT \times \ln \frac{k_F^{wt} k_U^{mut}}{k_U^{wt} k_F^{mut}} \quad (3)$$

Finally, to calculate the Φ values, the following equations were applied:

$$\Delta\Delta G_{folding\,TS} = RT \times \ln \frac{k_F^{wt}}{k_F^{mut}} \quad (4)$$

$$\phi = \frac{\Delta\Delta G_{folding\,TS}}{\Delta\Delta G_{folding}} \quad (5)$$

RESULTS

The structure of the complex between N_{TAIL} and XD is reported in Figure 1. XD binds through its hydrophobic pocket

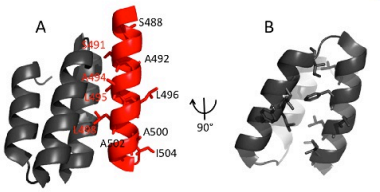


Figure 1. Structure of the complex between N_{TAIL} (red) and XD (gray) (Protein Data Bank entry 1T6O). (A) Residues of N_{TAIL} that were subjected to site-directed mutagenesis are highlighted as sticks and labeled. (B) Hydrophobic residues of the binding pocket of XD are highlighted as sticks.

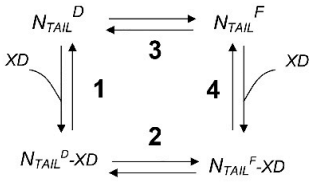
in a region of N_{TAIL} encompassing residues 489−506, which acquires α-helical folding.[23,25] To pursue our goal of providing a structural characterization of the binding and folding steps of the recognition between N_{TAIL} and XD, we performed a Φ value analysis.[30] Therefore, by following conventional rules of Φ value analysis, we designed a set of conservative substitutions that introduced a deletion of each of the targeted side chains. To avoid a perturbation of association rate constants by electrostatic effects, the five charged residues of N_{TAIL} in the region of residues 489−506 were excluded from the muta-

DOI: 10.1021/acs.biochem.7b00350
Biochemistry 2017, 56, 3780−3786

genesis. In total, 11 variants were designed, purified, and subjected to the binding-induced folding experiments described below. For the sake of clarity, the positions of N_{TAIL} that were subjected to site-directed mutagenesis are highlighted as sticks in Figure 1, where the hydrophobic residues of the binding site of XD are also depicted.

Because of the complexity of the folding upon binding, a quantitative description of the kinetics of recognition between N_{TAIL} and XD demands that some considerations about the underlying kinetic scheme be taken into account. The mechanism of binding between two molecules, involving a structural change of at least one of the reactants, should in theory be described by a squared scheme (Scheme 1).

Scheme 1



In this scenario, if the molecular flow primarily proceeds via steps 1 and 2, the overall reaction mechanism resembles an induced fit-type scenario,[32] whereby folding occurs only after binding. Alternatively, if the predominant mechanism occurs via steps 3 and 4, the folded and disordered conformations (N_{TAIL}^{F} and N_{TAIL}^{D}) preexist in solution, and therefore, the reaction is similar to a concerted or conformational selection-type model.[33] In the case of the binding between N_{TAIL} and XD via the application of a simple kinetic test, on the basis of a comparison of the experiments performed under pseudo-first-order conditions with respect to either XD or N_{TAIL},[28] we could show that N_{TAIL} folds only after binding to XD (induced fit).[22] Thus, the process is consistent with Scheme 2.

Scheme 2

$$N_{TAIL}^{D} + XD \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} N_{TAIL}^{D}\text{–}XD \underset{k_{F}}{\overset{k_{F}}{\rightleftarrows}} N_{TAIL}^{F}\text{–}XD$$

When induced fit is operative, the observed kinetics is governed by the two apparent rate constants, λ_1 and λ_2, that should be calculated from the two roots of a quadratic equation.[34] In many cases, however, including the association of N_{TAIL} and XD, only one rate constant can be determined, which jeopardizes a quantitative curve fitting. Two alternative approximations may be introduced to describe quantitatively the binding pathway of such a multistate reaction. On one hand, the association intermediate may be assumed to be in fast pre-equilibrium with the free species ($k_F \ll k_{off}$). In this case, the apparent association rate constant for induced folding at low reactant concentrations can be approximated to

$$k_{on}^{app} = \frac{1}{1 + \frac{k_{off}}{k_{on}[N_{TAIL}]}} \times k_F \quad (6)$$

whereas the apparent dissociation rate constant will approach the value of k_U.

An alternative analysis may be introduced if complex kinetics arises from a change in the rate-limiting step at different N_{TAIL} concentrations. Under these conditions, the binding transition state is postulated to have an energy higher than the barrier for the conformational change event ($k_{off} \ll k_F$). In this case, while the apparent association rate constant at low reactant concentrations will tend to k_{on}, the dissociation rate constant will be equal to

$$k_{off}^{app} = \frac{1}{1 + \frac{k_F}{k_{off}}} \times k_U \quad (7)$$

In both cases, the hyperbolic behavior of the observed rate constant arises from the accumulation of a reaction intermediate and the subsequent monomolecular folding step will become rate-limiting at an infinite reactant concentration, with an apparent rate constant of $k_F + k_U$. Thus, the apparent K_D obtained from the analysis of the apparent rate constant as a function of N_{TAIL} concentration from a hyperbolic fit returns the value of the stability of the initial complex, as expressed by its $K'_D = k_{off}/k_{on}$.

While it is generally very difficult to distinguish between these two approximations, in the case of the binding of wild-type N_{TAIL} to XD, some considerations can be put forward. In fact, it may be noted that the dissociation rate constant corresponds to a value of 720 s^{-1}. This experimentally determined value represents a lower limit for any microscopic process progressing toward unbinding, i.e., no microscopic step (including k_U) in the unbinding reaction can be slower than 720 s^{-1}. Moreover, the reciprocal of the observed dissociation rate constant is equal to the sum of the reciprocal of each individual step:

$$\frac{1}{k_{off}^{app}} = \frac{1}{k_{off}} + \frac{1}{k_U} \quad (8)$$

Thus, given that the asymptotic value of the observed rate constant is ~1700 s^{-1} (i.e., $k_F + k_U \approx 1700\ s^{-1}$), it follows that the value of k_{off}, which according to eq 6 has to be much higher than 720 s^{-1}, cannot be much lower than k_F. Accordingly, we may conclude that binding of N_{TAIL} to XD appears to approach a scenario in which the formation of the early complex is in a fast pre-equilibrium, where $k_{off}^{app} \approx k_U$.

To assess the structural features of the folding and binding steps highlighted in Scheme 2, we compared the observed kinetics of all the mutants of N_{TAIL} to that of the wild-type protein. In all cases, experiments were performed using a fluorescent pseudo-wild-type variant of XD, i.e., the Y480W single mutant, which we showed previously to bind N_{TAIL} with an affinity similar to that of wild-type XD and to display an excellent change in fluorescence upon binding.[22] In analogy to the protocol employed for wild-type N_{TAIL}, we performed temperature-jump experiments by incubating Y480W XD at a constant concentration of 2 μM with N_{TAIL} at different concentrations, typically ranging from 1 to 40 μM. The relaxation process was induced by a rapid increase in the temperature of the solution by 9 °C, from 16 to 25 °C, using a fluorescence-equipped capacitor-discharge temperature-jump instrument, the reaction being too fast for the stopped-flow apparatus. For all 11 variants, under all conditions, the observed kinetics was consistent with a single-exponential relaxation. The observed dependence of the overall macroscopic rate constant

DOI: 10.1021/acs.biochem.7b00350
Biochemistry 2017, 56, 3780−3786

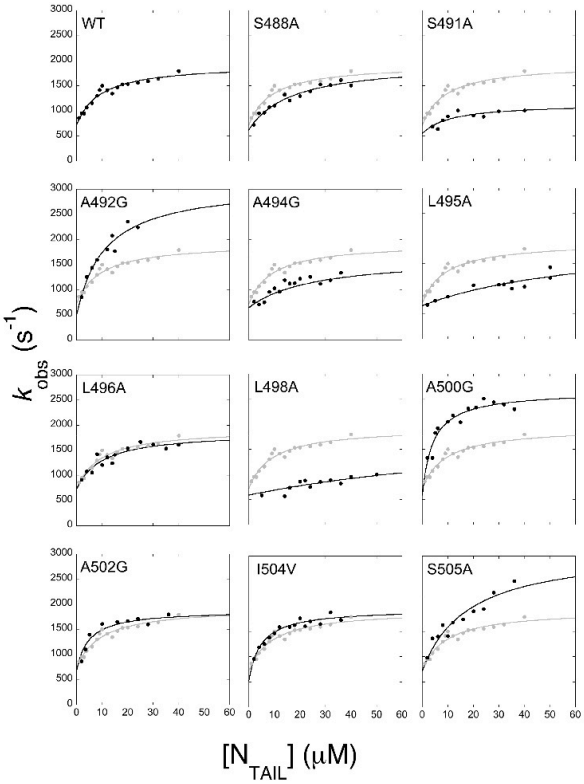


Figure 2. Kinetics of binding of N_{TAIL} and its site-directed variants to XD. Observed rate constants were measured as a function of N_{TAIL} concentration (typically ranging from 1 to 40 μM) by temperature-jump experiments, in the presence of XD Y480W at a constant concentration (2 μM): gray circles for wild-type N_{TAIL} and black circles for site-directed N_{TAIL} variants. For each variant, the overall dependence is consistent with a hyperbolic behavior.

on protein concentration is reported in Figure 2 for all variants. On the basis of the considerations discussed above, the observed kinetics of the different variants of N_{TAIL} were analyzed quantitatively to determine the folding and unfolding rate constants as well as the apparent stability of the initial encounter complex, as expressed by its K'_D, where $K'_D = k_{off}/k_{on}$. Nearly all mutants were consistent with a hyperbolic dependence, as formalized in eq 1, whereas in the case of two mutants, L495A and L498A, an apparently linear behavior was observed, indicating that these two positions are particularly important for the formation of the initial encounter complex between XD and N_{TAIL}. The calculated parameters for all the mutants together with their associated Φ values are listed in Table 1 and briefly discussed below.

■ DISCUSSION

Inferring the mechanism of folding upon binding of an IDP to its protein partner requires characterization of the structural determinants of each individual step. Given the cooperativity of these systems, however, such a goal has been to date particularly elusive,[14,15,20,35] and only in some cases, for example, Gab2,[36] could the folding of the unbound IDP be characterized. In the case of the binding of N_{TAIL} to XD, the fortuitous complexity of the observed hyperbolic kinetics, together with the mutagenesis of the residues directly involved

DOI: 10.1021/acs.biochem.7b00350
Biochemistry 2017, 56, 3780−3786

Table 1. Kinetic Folding and Binding Parameters for the Folding-after-Binding Reaction between XD Y480W and N_{TAIL} Variants

variant	k_F (s^{-1})	k_U (s^{-1})	K'_D (μM)	$\Delta\Delta G$ (kcal/mol)	$\Delta\Delta G_{IC}$ (kcal/mol)	$\Delta\Delta G^{mut-wt}_{folding}$ (kcal/mol)	$\Delta\Delta G^{mut-wt}_{folding\ TS}$ (kcal/mol)	$\Phi_{binding}$	$\Phi_{folding}$
WT	1210 ± 80	720 ± 80	9.0 ± 0.9	–	–	–	–	–	–
S488A	1400 ± 200	580 ± 60	16 ± 2	0.1 ± 0.1	0.33 ± 0.07	−0.2 ± 0.1	−0.11 ± 0.09	*a*	*b*
S491A	620 ± 60	650 ± 50	10 ± 1	0.4 ± 0.1	0.06 ± 0.08	0.34 ± 0.1	0.40 ± 0.07	0.15 ± 0.09	*b*
A492G	2600 ± 300	490 ± 50	11 ± 1	−0.6 ± 0.1	0.12 ± 0.08	−0.7 ± 0.1	−0.46 ± 0.07	−0.2 ± 0.1	0.7 ± 0.1
A494G	1000 ± 100	650 ± 60	26 ± 3	0.7 ± 0.1	0.63 ± 0.08	0.0 ± 0.1	0.10 ± 0.09	0.9 ± 0.2	*b*
L495A	1400 ± 400	600 ± 70	70 ± 20	1.1 ± 0.1	1.4 ± 0.2	−0.1 ± 0.2	−0.1 ± 0.1	1.1 ± 0.2	*b*
L496A	1100 ± 100	720 ± 70	10 ± 1	0.1 ± 0.1	0.05 ± 0.08	0.0 ± 0.1	0.04 ± 0.09	*a*	*b*
L498A	1100 ± 400	590 ± 60	100 ± 40	1.4 ± 0.1	1.4 ± 0.3	0.0 ± 0.2	0.1 ± 0.2	1.0 ± 0.3	*b*
A500G	2100 ± 200	500 ± 50	3.6 ± 0.4	−1.1 ± 0.1	−0.55 ± 0.08	−0.5 ± 0.1	−0.33 ± 0.07	0.5 ± 0.1	0.6 ± 0.2
A502G	1210 ± 80	660 ± 70	4.8 ± 0.5	−0.3 ± 0.1	−0.35 ± 0.07	−0.1 ± 0.1	0.00 ± 0.06	0.9 ± 0.3	*b*
I504V	1400 ± 100	520 ± 50	4.7 ± 0.5	−0.7 ± 0.1	−0.39 ± 0.09	−0.3 ± 0.1	−0.10 ± 0.07	0.6 ± 0.2	0.3 ± 0.3
S505A	2400 ± 400	720 ± 70	18 ± 2	0.0 ± 0.1	0.41 ± 0.08	−0.4 ± 0.1	−0.41 ± 0.10	*a*	*b*

[a]These variants show $\Delta\Delta G$ values of <0.4 kcal/mol, preventing reliable calculation of $\Phi_{binding}$. [b]These variants show $\Delta\Delta G^{mut-wt}_{folding}$ values of <0.4 kcal/mol, preventing reliable calculation of $\Phi_{folding}$.

in the folding upon binding reaction, allowed us to dissect the contribution of the different side chains of N_{TAIL} in the binding and folding steps, as briefly discussed below.

Encounter Reaction between XD and Disordered N_{TAIL}. We first discuss the identification of the residues that are critical in driving the initial encounter between XD and the disordered state of N_{TAIL}. Following a generally accepted convention,[31] the experimentally determined Φ values for the binding reaction between XD and the disordered N_{TAIL} were divided into three different groups and mapped onto the structure of the complex: small (Φ < 0.3), intermediate (0.3 < Φ < 0.7), and large (Φ > 0.7) values (Figure 3). It is evident

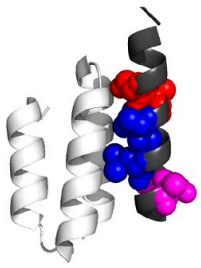


Figure 3. Cartoon representation of the complex between XD (white) and N_{TAIL} (gray) (Protein Data Bank entry 1T6O). The experimentally determined Φ values for the binding reaction are shown on the structure of N_{TAIL}. Small values (Φ < 0.3) are colored red, intermediate values (0.3 < Φ < 0.7) magenta, and large values (Φ > 0.7) blue.

that the initial encounter is essentially mediated by residues located in the central part of the helix and is predominately driven by the hydrophobic interactions involving residues A494, L495, and L498, all three showing a Φ value approaching 1. Furthermore, it may be observed that, in the high-energy complex between XD and disordered N_{TAIL}, nativelike structure gradually tapers off to the sides, with the C-terminal portion of the helix displaying Φ values higher than those observed at the N-terminus. Interestingly, substitutions A500G, A502G, and I504V, which are clustered consecutively in the C-terminal portion of the helix, all result in a stabilization of the complex (see Table 1). This finding highlights some structural frustration of wild-type N_{TAIL} in the C-terminal portion of the helix.

Folding Reaction of N_{TAIL}. A critical caveat of Φ value analysis is that substitutions should result in a change in free energy that should be large enough to be reliably measured but small enough not to abrogate folding.[31] For the majority of the variants reported in Table 1, the value of $\Delta\Delta G^{mut-wt}_{folding}$ was too low (i.e., <0.4 kcal mol^{-1}) to calculate a reliable value of Φ. In fact, fractional values of Φ could be calculated only for substitutions A492G, A500G, and I504V. Thus, relatively limited structural information about the folding transition state could be obtained.

To address the overall structural features of the transition state for folding of N_{TAIL}, we performed a linear free energy relationship analysis (LFER).[37,38] By this method, the changes in free energy of the transition state are plotted versus changes in equilibrium free energies, the slope of the observed correlation (called α value) reflecting the degree of similarity between the native and transition states and thus the position of the transition state along the reaction coordinate. As depicted in Figure 4, in the case of N_{TAIL}, the LFER yields a linear correlation with a slope of 0.82. While the linearity is consistent with what is generally observed in the folding of globular proteins,[38] indicating that N_{TAIL} folds via the so-called nucleation−condensation mechanism,[21] the α value is considerably higher than typical values of ∼0.3.[39] Thus, it appears that, after initial binding, N_{TAIL} folds via a transition state that resembles a distorted yet quite similar version of the native state (i.e., bound). In conclusion, the high degree of nativelike similarity of the transition state exceeds what is classically observed in the folding of globular proteins. Given the robustness of the α values seen in protein folding, this observation may highlight a distinctive feature of the folding of an IDP, whose generality will need to be explored by additional experimental investigations on other protein systems.

The Sequence of N_{TAIL} Is Not Evolved To Fold. A corollary of the principle of minimal frustration, which directly arises from the funneled nature of the free energy landscape with a deep minimum, is that the amino acid sequence of a

DOI: 10.1021/acs.biochem.7b00350
Biochemistry 2017, 56, 3780−3786

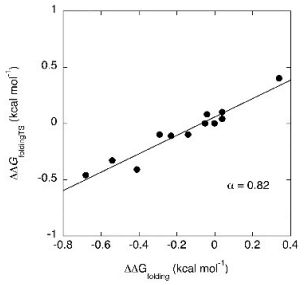


Figure 4. Linear free energy relationship (LFER) plot of the folding of N_{TAIL}. As described in the text, the changes in the free energy of the transition state ($\Delta\Delta G_{folding\ TS}$) are related to changes in equilibrium free energies ($\Delta\Delta G_{folding}$), returning a linear correlation. The α value, the slope of the observed correlation, reflects the degree of similarity between the native state and the transition state. The N_{TAIL} $\alpha = 0.82$ case indicates that folding proceeds via the nucleation−condensation mechanism, with a transition state that resembles a distorted version of the native (bound) state.

protein is evolved to maintain its shape.[40−44] Accordingly, with the relevant exception of functionally important sites, mutagenesis in a protein typically yields a destabilization of the native state.

By following these premises, we find it is of interest to comment on the effect of mutagenesis on the folding step of N_{TAIL}. From Table 1, it may be noted that of the 11 variants produced and analyzed, only one appears to destabilize the folding step of N_{TAIL}, while the others display either an increase in stability or a negligible change. This finding highlights an interesting feature of N_{TAIL} that contrasts with what is typically observed in globular proteins, indicating that the sequence of this IDP is not evolved to fold. While it would appear self-evident that the sequence of an IDP is, by definition, not designed to fold, it should nevertheless be noticed that the folding of N_{TAIL} after the initial encounter with XD is a spontaneous process driven by the overall binding, which poses the effect of mutagenesis of the folding of N_{TAIL} as an interesting conundrum.

By considering the data presented in this work together with previous experimental investigations of the physiological role of N_{TAIL} and XD, it is tempting to speculate that the lack of stability of N_{TAIL} in isolation may be a mechanism for tuning the N_{TAIL}−XD complex affinity within a functionally competent range to ensure both efficient transcription and replication. In fact, while abrogation of the N_{TAIL}−XD interaction is detrimental for virus transcription and replication, an increase in its stability appears to hamper the transcript elongation rates of the viral polymerase,[45] indicating that the K_D of the complex has to be finely regulated. This view is further supported by a recent study that identified the N_{TAIL}−XD interaction strength as a critical determinant in dictating the relative abundance of the viral transcripts, thereby ultimately determining the relative amounts of the various viral components.[46] Given the broad functional relevance of this interaction, it is conceivable that in the course of evolution the N_{TAIL} sequence has been naturally selected to finely tune its affinity for XD, and thereby its control of transcription.

From a broader perspective, it has been originally proposed that a peculiar feature of protein complexes involving IDPs is that they are highly dynamic and tend to retain a significant amount of disorder.[12] Therefore, it is reasonable to assume that the energy landscape of IDPs, even when bound to their partners, may retain a relatively high level of frustration, with a limited bias toward the main energetic minimum. On the basis of these considerations, we conclude that the peculiar effect of mutagenesis on the folding step of N_{TAIL} is a direct consequence of the fuzziness of the complex and might represent a general feature of IDPs.

AUTHOR INFORMATION

Corresponding Author

*E-mail: stefano.gianni@uniroma1.it.

ORCID

Stefano Gianni: 0000-0003-1653-1925

Funding

Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (Progetto di Interesse "Invecchiamento" to S.G.), Sapienza University of Rome (C26A155S48 and B52F16003410005 to S.G., C26N15J4A5 to A.T., and C26N15E4LB and B52F16001770005 to D.B.), and the Associazione Italiana per la Ricerca sul Cancro (Individual Grant, MFAG 2016, 18701 to S.G.). F.T. is a recipient of a Ph.D. fellowship from the Italo-French University. A.T. is a recipient of a postdoctoral fellowship from the Istituto Pasteur Italia-Cenci Bolognetti.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) Intrinsically disordered protein. *J. Mol. Graphics Modell. 19*, 26−59.

(2) Dyson, H. J. (2011) Expanding the proteome: disordered and alternatively folded proteins. *Q. Rev. Biophys. 44*, 467−518.

(3) Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol. 6*, 197−208.

(4) Habchi, J., Tompa, P., Longhi, S., and Uversky, V. N. (2014) Introducing protein intrinsic disorder. *Chem. Rev. 114*, 6561−6588.

(5) Tompa, P. (2011) Unstructural biology coming of age. *Curr. Opin. Struct. Biol. 21*, 419−425.

(6) Tompa, P., Szász, C., and Buday, L. (2005) Structural disorder throws new light on moonlighting. *Trends Biochem. Sci. 30*, 484−489.

(7) Uversky, V. N., and Dunker, A. K. (2010) Understanding protein non-folding. *Biochim. Biophys. Acta, Proteins Proteomics 1804*, 1231−1264.

(8) Wright, P. E., and Dyson, H. J. (1999) Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol. 293*, 321−331.

(9) Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol. 12*, 54−60.

(10) Uversky, V. N., Gillespie, J. R., and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins: Struct., Funct., Genet. 41*, 415−427.

(11) Spolar, R. S., and Record, M. T. J. (1994) Coupling of local folding to site-specific binding of proteins to DNA. *Science 263*, 777−784.

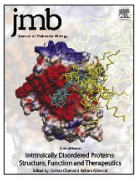
DOI: 10.1021/acs.biochem.7b00350
Biochemistry 2017, 56, 3780−3786

(12) Tompa, P., and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci. 33*, 2−8.
(13) Drobnak, I., De Jonge, N., Haesaerts, S., Vesnaver, G., Loris, R., and Lah, J. (2013) Energetic basis of uncoupling folding from binding for an intrinsically disordered protein. *J. Am. Chem. Soc. 135*, 1288−1294.
(14) Gianni, S., Dogan, J., and Jemth, P. (2016) Coupled binding and folding of intrinsically disordered proteins: what can we learn from kinetics? *Curr. Opin. Struct. Biol. 36*, 18−24.
(15) Shammas, S. L., Crabtree, M. D., Dahal, L., Wicky, B. I., and Clarke, J. (2016) Insights into Coupled Folding and Binding Mechanisms from Kinetic Studies. *J. Biol. Chem. 291*, 6689−6695.
(16) Bachmann, A., Wildemann, D., Praetorius, F., Fischer, G., and Kiefhaber, T. (2011) Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction. *Proc. Natl. Acad. Sci. U. S. A. 108*, 3952−3957.
(17) Hill, S. A., Kwa, L. G., Shammas, S. L., Lee, J. C., and Clarke, J. (2014) Mechanism of assembly of the non-covalent spectrin tetramerization domain from intrinsically disordered partners. *J. Mol. Biol. 426*, 21−35.
(18) Rogers, J. M., Oleinikovas, V., Shammas, S. L., Wong, C. T., De Sancho, D., Baker, C. M., and Clarke, J. (2014) Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *Proc. Natl. Acad. Sci. U. S. A. 111*, 15420−15425.
(19) Shammas, S. L., Travis, A. J., and Clarke, J. (2014) Allostery within a transcription coactivator is predominantly mediated through dissociation rate constants. *Proc. Natl. Acad. Sci. U. S. A. 111*, 12055−12060.
(20) Dogan, J., Gianni, S., and Jemth, P. (2014) The binding mechanisms of intrinsically disordered proteins. *Phys. Chem. Chem. Phys. 16*, 6323−6331.
(21) Fersht, A. R. (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U. S. A. 92*, 10869−10873.
(22) Dosnon, M., Bonetti, D., Morrone, A., Erales, J., di Silvio, E., Longhi, S., and Gianni, S. (2015) Demonstration of a Folding after Binding Mechanism in the Recognition between the Measles Virus NTAIL and X Domains. *ACS Chem. Biol. 10*, 795−802.
(23) Bourhis, J., Johansson, K., Receveur-Bréchot, V., Oldfield, C. J., Dunker, A. K., Canard, B., and Longhi, S. (2004) 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. 99*, 157−167.
(24) Johansson, K., Bourhis, J.-M., Campanacci, V., Cambillau, C., Canard, B., and Longhi, S. (2003) Crystal Structure of the Measles Virus Phosphoprotein Domain Responsible for the Induced Folding of the C-Terminal Domain of the Nucleoprotein. *J. Biol. Chem. 278*, 44567−44572.
(25) Kingston, R. L., Hamel, D. J., Gay, L. S., Dahlquist, F. W., and Matthews, B. W. (2004) Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc. Natl. Acad. Sci. U. S. A. 101*, 8301−8306.
(26) Bourhis, J. M., Receveur-Bréchot, V., Oglesbee, M., Zhang, X., Buccellato, M., Darbon, H., Canard, B., Finet, S., and Longhi, S. (2005) 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. 14*, 1975−1992.
(27) Gely, S., Lowry, D. F., Bernard, C., Jensen, M. R., Blackledge, M., Costanzo, S., Bourhis, J. M., Darbon, H., Daughdrill, G., and Longhi, S. (2010) Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. *J. Mol. Recognit. 23*, 435−447.
(28) Gianni, S., Dogan, J., and Jemth, P. (2014) Distinguishing induced fit from conformational selection. *Biophys. Chem. 189*, 33−39.
(29) Gianni, S., Walma, T., Arcovito, A., Calosci, N., Bellelli, A., Engstrom, A., Travaglini-Allocatelli, C., Brunori, M., Jemth, P., and Vuister, G. W. (2006) Demonstration of long-range interactions in a PDZ domain by NMR, kinetics, and protein engineering. *Structure 14*, 1801−1809.
(30) Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol. 224*, 771−782.
(31) Fersht, A. R., and Sato, S. (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U. S. A. 101*, 7976−7981.
(32) Koshland, D. E. J., Némethy, G., and Filmer, D. (1966) Comparison of experimental binding data and theoretical models in proteins containing subunits. *Biochemistry 5*, 365−385.
(33) Monod, J., Wyman, J., and Changeux, J. P. (1965) On the nature of allosteric transitions: a plausible model. *J. Mol. Biol. 12*, 88−118.
(34) Fersht, A. R. (1999) *Structure and mechanism in protein science*, Freeman, New York.
(35) Kiefhaber, T., Bachmann, A., and Jensen, K. S. (2012) Dynamics and mechanisms of coupled protein folding and binding reactions. *Curr. Opin. Struct. Biol. 22*, 21−29.
(36) Krieger, J. M., Fusco, G., Lewitzky, M., Simister, P. C., Marchant, J., Camilloni, C., Feller, S. M., and De Simone, A. (2014) Conformational recognition of an intrinsically disordered protein. *Biophys. J. 106*, 1771−1779.
(37) Leffler, J. E. (1953) Parameters for the description of transition states. *Science 117*, 340−341.
(38) Fersht, A. R. (2004) Relationship of Leffler (Bronsted) alpha values and protein folding Phi values to position of transition-state structures on reaction coordinates. *Proc. Natl. Acad. Sci. U. S. A. 101*, 14338−14342.
(39) Naganathan, A. N., and Muñoz, V. (2010) Insights into protein folding mechanisms from large scale analysis of mutational effects. *Proc. Natl. Acad. Sci. U. S. A. 107*, 8611−8616.
(40) Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct., Funct., Genet. 21*, 167−195.
(41) Onuchic, J. N., Socci, N. D., Luthey-Schulten, Z., and Wolynes, P. G. (1996) Protein folding funnels: the nature of the transition state ensemble. *Folding Des. 1*, 441−450.
(42) Wolynes, P. G. (2005) Energy landscapes and solved protein-folding problems. *Philos. Trans. R. Soc., A 363*, 453−464.
(43) Ferreiro, D. U., Hegler, J. A., Komives, E. A., and Wolynes, P. G. (2007) Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U. S. A. 104*, 19819−19824.
(44) Ferreiro, D. U., Hegler, J. A., Komives, E. A., and Wolynes, P. G. (2011) On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci. U. S. A. 108*, 3499−3503.
(45) Brunel, J., Chopy, D., Dosnon, M., Bloyet, L. M., Devaux, P., Urzua, E., Cattaneo, R., Longhi, S., and Gerlier, D. (2014) Sequence of events in measles virus replication: role of phosphoprotein-nucleocapsid interactions. *J. Virol. 88*, 10851−10863.
(46) Bloyet, L. M., Brunel, J., Dosnon, M., Hamon, V., Erales, J., Gruet, A., Lazert, C., Bignon, C., Roche, P., Longhi, S., and Gerlier, D. (2016) Modulation of Re-initiation of Measles Virus Transcription at Intergenic Regions by PXD to NTAIL Binding Strength. *PLoS Pathog. 12*, e1006058.

DOI: 10.1021/acs.biochem.7b00350
Biochemistry 2017, 56, 3780−3786

Paper 2.

Article



Partner-Mediated Polymorphism of an Intrinsically Disordered Protein

Christophe Bignon [1], Francesca Troilo [1,2], Stefano Gianni [2] and Sonia Longhi [1]

*1 - **Aix-Marseille Univ,** CNRS, Architecture et Fonction des Macromolécules Biologiques (AFMB) UMR 7257, Marseille, France*
*2 - **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*

***Correspondence to Christophe Bignon and Sonia Longhi:** CNRS and Aix-Marseille Université, AFMB, UMR 7257, Case 932, 163, Avenue de Luminy, Case 932, 13288 Marseille, France. Christophe.Bignon@afmb.univ-mrs.fr; Sonia.Longhi@afmb.univ-mrs.fr*
https://doi.org/10.1016/j.jmb.2017.11.012
Edited by Monika Fuxreiter

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_{TAIL}) is partly pre-configured as an α-helix in the free form and undergoes α-helical folding upon binding to the X domain (XD) of the viral phosphoprotein. Beyond XD, N_{TAIL} also binds the major inducible heat shock protein 70 (hsp70). So far, no structural information is available for the N_{TAIL}/hsp70 complex. Using mutational studies combined with a protein complementation assay based on green fluorescent protein reconstitution, we have investigated both N_{TAIL}/XD and N_{TAIL}/hsp70 interactions. Although the same N_{TAIL} 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_{TAIL}/hsp70 interaction strength. Furthermore, while an increased and a decreased α-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 α-helical conformation once bound to hsp70. Here, by showing that the α-helical conformation sampled by the free form of the MoRE does not systematically commit it to adopt an α-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.

Introduction

The measles virus (MeV) nucleoprotein consists of a structured region (N_{CORE}, aa 1–400) [1] and a C-terminal domain (N_{TAIL}, aa 401–525) that is intrinsically disordered [2,3]. Intrinsically disordered proteins (IDPs) or regions 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_{TAIL} domains from the cognate Henipaviruses [10–13], MeV N_{TAIL} undergoes folding upon binding to the X domain (XD) of the viral phosphoprotein [14–16]. Interaction with the partner triggers α-helical folding within a molecular recognition element (MoRE; aa 486–502) located within one (Box2, aa 489–506) out of three conserved N_{TAIL} regions (Fig. 1) [3,14,17–24]. Binding of N_{TAIL} to XD results in a "fuzzy" complex [25–28]; that is, 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_{TAIL} 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

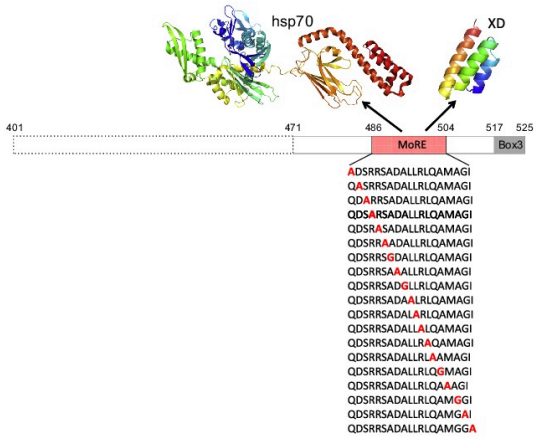


Fig. 1. Schematic representation of the variants resulting from alanine scanning mutagenesis. Mutagenesis was performed in the context of the truncation variant 471 that spans residues 471–525 of N_{TAIL}. Nineteen single-site variants were generated. Above the scheme are shown cartoon representations 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., aa 3–382 and aa 389–610) is based on the structure of a form encompassing residues 1–554 (PDB code 1YUW).

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; that is, they can fold into different structures upon binding to different partners (see ref. [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_{TAIL} 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_{TAIL} fuzzy appendage from residue 401 to residue 481 results in a non-monotonic increase of its binding strength toward 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_{TAIL}/hsp70 complex. It therefore remains to be established whether the MoRE undergoes α-helical folding also upon binding to hsp70 and whether the interaction relies on the same N_{TAIL} residues. Deciphering the molecular mechanisms governing the N_{TAIL}/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_{TAIL} binding to hsp70 is mediated by the same N_{TAIL} residues as binding to XD, we targeted for alanine scanning mutagenesis the MoRE (Fig. 1) and assessed the binding properties of the resulting variants toward both XD and hsp70 using a split-GFP reassembly assay. In this assay, wild-type (wt) N_{TAIL} (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 *Escherichia coli*, if N_{TAIL} interacts with the binding partner, the two GFP halves are brought to sufficiently close proximity to allow for 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_{TAIL} 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_{TAIL} (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_{TAIL} [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 (Fig. 1). Next we assessed their binding properties toward the two partners. The results obtained with XD and hsp70 are shown in Fig. 2A and B, respectively. Note that except for variant 486, which seems to be more sensitive than the others to bacterial proteolysis, the fluorescence variations reported in Fig. 2A and B cannot be accounted for by differences in the expression levels of the N_{TAIL} variants (Fig. 2C).

As shown in Fig. 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

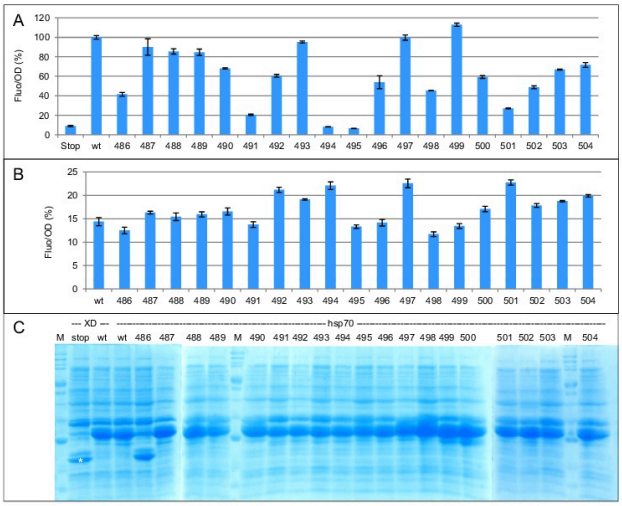


Fig. 2. Binding efficiency of alanine single-site N_{TAIL} variants toward XD and hsp70 as assessed by split-GFP reassembly. Normalized fluorescence of NGFP-N_{TAIL} variants upon co-expression with XD-CGFP (A) and hsp70-CGFP (B). Stop, NGFP alone (negative control); wt, wt471 truncation variant (aa 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.

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 toward 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 [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_{TAIL} (residues 401 to 525), whereas it is used here in the context of N_{TAIL} truncation variant 471 (residues 471 to 525). Since we have shown that the N_{TAIL} N-terminal fuzzy appendage (residues 401 to 480) dampens the N_{TAIL}/XD interaction [52], it is conceivable that the R497G substitution has a stronger effect when used in the context of full-length N_{TAIL} than in the context of N_{TAIL} 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 Cβ 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_{TAIL} 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 Φ-value analysis of the N_{TAIL}/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_D that is close to that of wtN_{TAIL}.

The binding efficiencies of the single-site variants toward hsp70 are shown in Fig. 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 Fig. 2A, B). Beyond this overall difference in interaction strength, the single-site variants behave quite differently toward 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_{TAIL}/XD interaction, no single substitution on its own is able to abolish the N_{TAIL}/hsp70 interaction. This observation suggests that no single residue is as critical for hsp70 binding as it is for XD binding; that is, hsp70 is more tolerant of substitutions than XD, which implies that N_{TAIL} interacts with XD and hsp70 using different mechanisms. The higher tolerance of hsp70 vis-à-vis N_{TAIL} substitutions may be tightly connected with the lower affinity of the N_{TAIL}/hsp70 interaction. One possible scenario could be that the N_{TAIL}/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], that is, the MoRE can remain highly dynamic at the surface of the partner (for examples, see refs. [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_{TAIL} sequence can be improved in terms of hsp70 binding. By contrast, in a previous random mutagenesis study, where substitutions were randomly introduced within N_{TAIL} 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 toward 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 are that the sequence of N_{TAIL} has been selected during evolution to achieve optimal binding with XD, reflecting the need for the N_{TAIL}/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_{TAIL}/hsp70 and N_{TAIL}/XD interactions do not have the same evolution potential. One possible hypothesis for the higher evolvability of the N_{TAIL}/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_{TAIL}/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_{TAIL}/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 α-helicity of the MoRE on XD and hsp70 binding

Since the MoRE was shown to fold into an α-helix upon binding to XD [3,14,17–20,22–24] and since the results of the alanine scanning mutagenesis reported in Fig. 2A and B suggest that N_{TAIL} 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 (Fig. 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 α-helicity, whereas Gly471 was expected to be unable to fold into an α-helix even upon XD binding because of its high glycine content and of the well-known α-helix-breaking ability of this residue [64–67]. These expectations were reinforced by convergent results obtained using disorder predictions (Fig. 3A), hydrophobic cluster analysis [69], secondary structure predictions [70] (Supplementary Fig. S1), and modeling (Fig. 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 (Fig. 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 Fig. 3D. In this experiment, XD interaction with full-length N_{TAIL} (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_{TAIL} (compare 401/XD and wt471/XD in Fig. 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 Fig. 3D), suggesting that a preformed α-helix leads to an increased N_{TAIL}/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_{TAIL}/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

Table I. Amino acid sequence of the different MoREs used in this study

	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504
wtMoRE	Q	D	S	R	R	S	A	D	A	L	L	R	L	Q	A	M	A	G	I
AlaMoRE	Q	A	A	A	A	S	A	A	A	L	A	A	L	A	A	M	A	A	A
GlyMoRE	Q	G	G	G	G	S	A	G	A	L	G	G	L	G	A	M	A	G	G
hsbMoRE	Q	A	A	A	A	S	G	A	G	L	L	A	L	Q	G	A	G	A	A

wtMoRE, wild-type sequence. AlaMoRE, the residues that were not considered as critical for XD binding based on the results of the alanine scanning mutagenesis were replaced with an alanine (shown in red). GlyMoRE, the residues that were replaced with an alanine in AlaMoRE were replaced with a glycine (shown in blue). For creating hsbMoRE, all the residues individually shown to lead to increased N_{TAIL}/hsp70 binding strength by the alanine scanning mutagenesis were collectively replaced with alanine or with glycine when the wild-type residue was alanine.

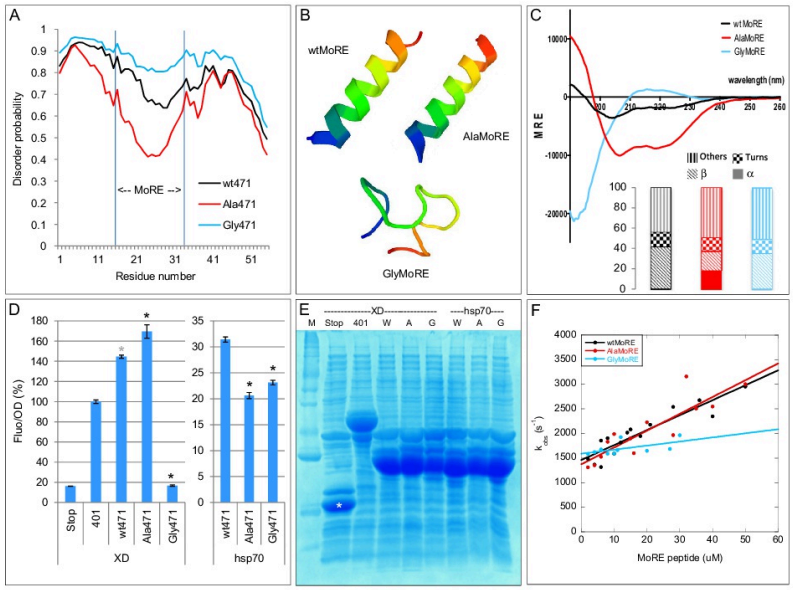


Fig. 3. Binding efficiency toward XD and hsp70 of N_{TAIL} variants mutated within the MoRE as assessed by split-GFP reassembly. (A) Disorder probability, as obtained using IUPred [68] of wt471, Ala471, and Gly471. The *x* axis is residue number of 471. The position of the MoRE within the sequence is 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 wtN_{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 gray 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 as follows: 20.4 kDa (Stop), 34 kDa (401), and 25.9 kDa (wt471, Ala471, Gly471). (F) Pseudo-first-order kinetics of the binding between Y480W XD and wtMoRE, AlaMoRE and GlyMoRE peptides.

negative control level (compare Gly471/XD and Stop/XD in Fig. 3D). This latter result definitely indicates that if the MoRE cannot fold into an α-helix, then N_{TAIL}/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 α-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 (Fig. 3D). As expected, wt471 displays a much lower interaction strength toward hsp70

compared to XD (compare wt471/hsp and wt471/XD in Fig. 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 Fig. 3D). The possibility that different protein expression levels could be responsible for the observed differences in fluorescence was checked and ruled out (Fig. 3E). Thus, while increasing or decreasing the α-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_{TAIL} 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_{TAIL} together with the small fluorescence changes induced by binding prevents 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 μM XD with varying concentrations of either wtMoRE or AlaMoRE or GlyMoRE (typically ranging from 2 to 50 μM). 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 Fig. S2 and data not shown).

Inspection of kinetic data (Fig. 3F) reveals that while the behavior of AlaMoRE is very similar to that of wtMoRE (wtMoRE: $k_{off}^{app} = 1460 \pm 80\ s^{-1}$, $k_{on}^{app} = 30 \pm 4\ s^{-1}\ \mu M^{-1}$; AlaMoRE: $k_{off}^{app} = 1400 \pm 100\ s^{-1}$, $k_{on}^{app} = 34 \pm 5\ s^{-1}\ \mu M^{-1}$), there is a detectable destabilization of the complex in the case of GlyMoRE ($k_{off}^{app} = 1610 \pm 60\ s^{-1}$, $k_{on}^{app} = 5 \pm 2\ s^{-1}\ \mu M^{-1}$). 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, while 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 Fig. 3D and F indicate that α-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_{TAIL} through different molecular mechanisms.

Mapping the reciprocal binding regions in the N_{TAIL}/hsp70 interaction

The results reported in Figs. 2 and 3 suggest that the N_{TAIL}/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_{TAIL}/hsp70 interaction to occur, an interaction that has been much less investigated compared to the N_{TAIL}/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_{TAIL}.

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 Ref. [51] and references therein cited). To map the minimal region involved in N_{TAIL} binding, we generated different hsp70 constructs containing one or more of these domains (Fig. 4A, lower panel) and tested their ability to interact with wt471 or with its Ala471 and Gly471 derivatives. Results are shown in Fig. 4A, upper panel. The three N_{TAIL} 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 μM) of the N_{TAIL}-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 Gierasch [71] that showed that the hsp70 C-terminal region is dispensable for binding of a peptide substrate, inter-domain allostery, and co-chaperon interaction.

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

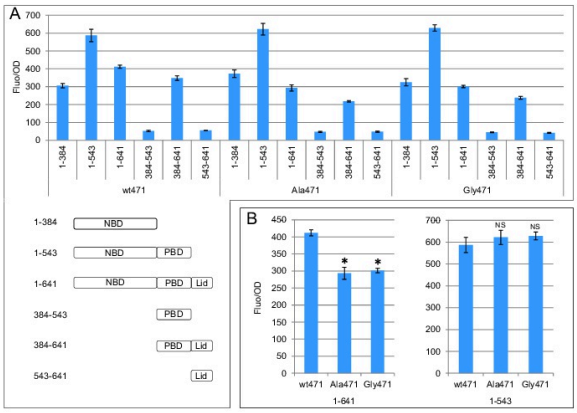


Fig. 4. Binding efficiency of hsp70 truncation variants as assessed by split-GFP reassembly. (A) Upper panel: normalized fluorescence of hsp70 constructs schematically represented in the lower panel, fused to CGFP, as obtained upon co-expression with wt471, Ala471, and Gly471 all fused to NGFP. On the *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. Numbers on the left indicate the first and last residues 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.

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 (Fig. 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 Fig. 3D. By contrast, the lid-free hsp70 binds all the three N_{TAIL} variants in a similar manner. In conclusion, results indicate that removing the lid not only increases binding toward N_{TAIL} but also renders it less sensitive to substitutions within the MoRE that have profound effects on XD binding.

To further dissect the N_{TAIL}/hsp70 interaction, we generated a series of N_{TAIL} truncation variants (Fig. 5A) and assessed their binding properties toward hsp70 by split-GFP reassembly. The rationale for doing that was to map the minimal binding region and to assess the contribution of N_{TAIL} 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_{TAIL}/hsp70 complexes [51] that are otherwise mediated by high affinity binding to Box2 [50]. As shown in Fig. 5B, the N_{TAIL}/hsp70 interaction strength increases with increasing N- and C-terminal shortening of N_{TAIL}, 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 (Fig. 5C). These results therefore indicate that as in the case of XD, the MoRE is the only N_{TAIL} region responsible for binding to full-length hsp70, and that Box3 behaves as a weak binding dampener.

Although the N_{TAIL}/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_{TAIL} 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).

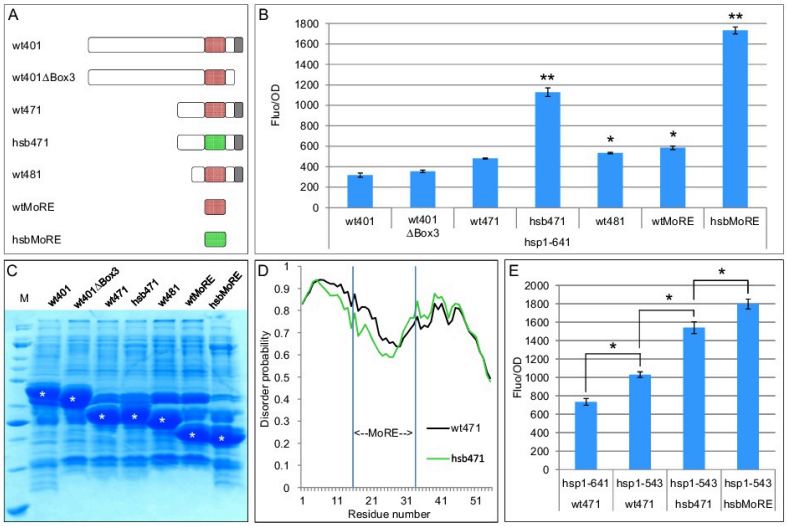


Fig. 5. Binding efficiency of N_{TAIL} truncation and mutated variants toward hsp70 as assessed by split-GFP reassembly. (A) Schematic representation of N_{TAIL} variants used in panel B. The 401–485 and 505–516 regions of N_{TAIL} are represented in white. The MoRE, encompassing residues 486–504, is represented as a red (*wild-type*) or green (mutated) box. wt401, wt full-length N_{TAIL} (aa 401–525); wt401ΔBox3, wtN_{TAIL} devoid of Box3 (aa 401–516); wt471, aa 471–525 with wtMoRE; hsb471, aa 471–525 with hsbMoRE; wt481, truncated wtN_{TAIL} variant (aa 481–525); wtMoRE, truncated wtN_{TAIL} variant (aa 485–506); hsbMoRE, mutated wtN_{TAIL} variant (aa 485–506). (B) Normalized fluorescence of N_{TAIL} 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_{TAIL} 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 $p < 0.05$).

Rational design of an N_{TAIL} variant with enhanced interaction strength toward hsp70

Since alanine scanning mutagenesis showed that several single substitutions lead to an increase in N_{TAIL}/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 (Fig. 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) toward full-length hsp70 than wt471 in a split-GFP reassembly assay (compare wt471 and hsb471 in Fig. 5B).

Since results reported in Fig. 5B definitely identified the MoRE as the only interacting unit between N_{TAIL} and hsp70, and since substitution of wt471 with hsb471 resulted in a significant increase in N_{TAIL}/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 (Fig. 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 Fig. 4, hsp70 truncation variant NBD–PBD (hsp1–543) proved to be the best N_{TAIL} binder. Results reported in Fig. 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 threefold increase in binding to hsp70 (Fig. 5B) raises the question as to how can hsp70 binding to N_{TAIL} 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 [72]. 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 [72] also unveiled that increased hydrophobicity is correlated with increased binding strength. Although previous data identified exposed patches of hydrophobic residues in native proteins [73] and hydrophobic residues in peptides as targets of hsp70 binding [74], 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 (Fig. 3D) and wt471 is a weaker hsp70 binder than hsb471 (Fig. 5B), although the hydropathy of AlaMoRE (1.6), as obtained using the Kyte–Doolittle scale and the expasy server[†], 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_{TAIL}/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 [75].

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_{TAIL} recognizes XD and hsp70, two structured partners differing in size and fold. Results indicate that although N_{TAIL} 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_{TAIL} substitutions than XD, with this higher tolerance likely coming at the cost of a lower affinity compared to the N_{TAIL}/XD pair. The N_{TAIL}/hsp70 interaction is more evolvable compared to the N_{TAIL}/XD interaction; that is, the majority of substitutions lead to increased interaction strength. Notably, while an increased and a decreased α-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 α-helical conformation once bound to hsp70. Shortening hsp70 by removing the lid results in increased binding toward N_{TAIL}, as does the shortening of the latter to encompass only the MoRE. Finally, N_{TAIL}/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 toward 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. Here, by showing that the conformation sampled by the free form of the MoRE of N_{TAIL} does not necessarily commit N_{TAIL} to adopt an α-helical conformation in the bound form, we

bring an additional "brick in the wall"; that is, 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_{TAIL} folds as anti-parallel β sheet at the air/water interface [76], thereby further underscoring the extreme polymorphism of this IDP.

Materials and Methods

Alanine scanning mutagenesis and generation of NGFP-N_{TAIL} constructs

Two complementary mutagenic primers were used in two separate PCR amplification experiments using the *wt*N_{TAIL}471 construct in the pNGG vector [55,77] 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 [77]. The second PCR product was directly used in an LR reaction (Invitrogen) with expression vector pNGG as previously described [77]. The mutagenic primers used to generate all the single-site alanine variants are listed in Supplementary Table S1.

Full-length *wt*N_{TAIL}401 (aa 401–525), *wt*N_{TAIL}471 (aa 471–525), and *wt*N_{TAIL}481 (aa 481–525) have been already described [52]. Truncation variant *wt*N_{TAIL}401ΔBox3 (aa 401–516) was obtained by PCR amplification using the *wt*N_{TAIL}401 construct in pNGG as template and primers attB1 and Δ3-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_{TAIL}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_{TAIL}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 wt471 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 μl 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_{TAIL} 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_{TAIL} constructs are shown in Supplementary Table S2.

$CaCl_2$-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, and 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]. $CaCl_2$-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_{600}

have been measured, the whole volume of triplicate cultures (12 ml) was spun for 5′ at 4000*g* 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, and 0.25 mg/ml Lysozyme, and were frozen. After thawing, cell lysates were supplemented with 20 μg/ml DNAse I and 20 mM $MgSO_4$ 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 4000*g*, 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 × 1 ml of 50 mM Tris/HCl (pH 8), 0.3 M NaCl and 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‡ [78].

Synthetic peptides and far-UV 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_{TAIL} 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_2, and equipped with a Peltier thermoregulation system. One-mm-thick quartz cuvettes were used. Peptide concentrations were 0.1 mg/ml. 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 Web site§ [79] was used to analyze the experimental data in the 190- to 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 was used in control measurements. Degassed and filtered samples were slowly pumped through the 0.5 × 2-mm quartz flow cell before data acquisition. The excitation wavelength was 296 nm and the fluorescence emission was measured using a 320-nm cutoff 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 μM) 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.

Acknowledgments

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). F.T. is a recipient of a PhD fellowship from the Italo-French University.

Appendix A. Supplementary data

Two supplementary figures (Figs. S1 and S2) and two supplementary tables (Tables S1 and S2). Supplementary data associated with this article can be found in the online version, at https://doi.org/10.1016/j.jmb.2017.11.012.

Received 3 October 2017;
Received in revised form 16 November 2017;
Accepted 19 November 2017
Available online 29 November 2017

Keywords:
measles virus;
N_{TAIL};
XD;
major inducible heat shock protein (hsp70);
induced folding

†https://www.expasy.org/

‡http://mobyle.rpbs.univ-paris-diderot.fr/cgi-bin/portal.py#forms::PEP-FOLD3

§http://bestsel.elte.hu/

Abbreviations used:
MeV, measles virus; IDPs, intrinsically disordered proteins; hsp70, heat shock protein; XD, X domain; MoRE, molecular recognition element; GFP, green fluorescent protein; CD, circular dichroism; NBD, nucleotide binding domain; PBD, peptide binding domain.

References

[1] I. Gutsche, A. Desfosses, G. Effantin, W.L. Ling, M. Haupt, R.W. Ruigrok, et al., Near-atomic cryo-EM structure of the helical measles virus nucleocapsid, Science 348 (2015) 704–707.
[2] S. Longhi, V. Receveur-Brechot, D. Karlin, K. Johansson, H. Darbon, D. Bhella, 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. 278 (2003) 18638–18648.
[3] J. Bourhis, K. Johansson, V. Receveur-Bréchot, C.J. Oldfield, A.K. Dunker, B. Canard, 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. 99 (2004) 157–167.
[4] A.K. Dunker, E. Garner, S. Guilliot, P. Romero, K. Albrecht, J. Hart, et al., Protein disorder and the evolution of molecular recognition: theory, predictions and observations, Pac. Symp. Biocomput. 3 (1998) 473–484.
[5] P.E. Wright, H.J. Dyson, Intrinsically unstructured proteins: re-assessing the protein structure–function paradigm, J. Mol. Biol. 293 (1999) 321–331.
[6] V.N. Uversky, What does it mean to be natively unfolded? Eur. J. Biochem. 269 (2002) 2–12.
[7] V.N. Uversky, Natively unfolded proteins: a point where biology waits for physics, Protein Sci. 11 (2002) 739–756.
[8] J. Habchi, P. Tompa, S. Longhi, V.N. Uversky, Introducing protein intrinsic disorder, Chem. Rev. 114 (2014) 6561–6588.
[9] P.E. Wright, H.J. Dyson, Linking folding and binding, Curr. Opin. Struct. Biol. 19 (2009) 31–38.
[10] D. Blocquel, J. Habchi, A. Gruet, S. Blangy, S. Longhi, Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies, Mol. BioSyst. 8 (2012) 392–410.
[11] J. Habchi, S. Blangy, L. Mamelli, M. Ringkjobing Jensen, M. Blackledge, H. Darbon, et al., Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipaviruses, J. Biol. Chem. 286 (2011) 13583–13602.
[12] G. Communie, J. Habchi, F. Yabukarski, D. Blocquel, R. Schneider, N. Tarbouriech, et al., Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus, PLoS Pathog. 9 (2013), e1003631.
[13] L. Baronti, J. Erales, J. Habchi, I.C. Felli, R. Pierattelli, S. Longhi, 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, ChemBioChem 16 (2015) 268–276.
[14] J.M. Bourhis, V. Receveur-Bréchot, M. Oglesbee, X. Zhang, M. Buccellato, H. Darbon, 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. 14 (2005) 1975–1992.
[15] D. Blocquel, J. Habchi, S. Costanzo, A. Doizy, M. Oglesbee, S. Longhi, Interaction between the C-terminal domains of measles virus nucleoprotein and phosphoprotein: a tight complex implying one binding site, Protein Sci. 21 (2012) 1577–1585.
[16] M. Dosnon, D. Bonetti, A. Morrone, J. Erales, E. di Silvio, S. Longhi, et al., Demonstration of a folding after binding mechanism in the recognition between the measles virus NTAIL and X domains, ACS Chem. Biol. 10 (2015) 795–802.
[17] K. Johansson, J.M. Bourhis, V. Campanacci, C. Cambillau, B. Canard, S. Longhi, Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein, J. Biol. Chem. 278 (2003) 44567–44573.
[18] R.L. Kingston, D.J. Hamel, L.S. Gay, F.W. Dahlquist, B.W. Matthews, Structural basis for the attachment of a paramyxoviral polymerase to its template, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 8301–8306.
[19] V. Belle, S. Rouger, S. Costanzo, E. Liquiere, J. Strancar, B. Guigliarelli, 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: Struct., Funct., Bioinf. 73 (2008) 973–988.
[20] C. Bernard, S. Gely, J.M. Bourhis, X. Morelli, S. Longhi, H. Darbon, Interaction between the C-terminal domains of N and P proteins of measles virus investigated by NMR, FEBS Lett. 583 (2009) 1084–1089.
[21] C.G. Bischak, S. Longhi, D.M. Snead, S. Costanzo, E. Terrer, C.H. Londergan, Probing structural transitions in the intrinsically disordered C-terminal domain of the measles virus nucleoprotein by vibrational spectroscopy of cyanylated cysteines, Biophys. J. 99 (2010) 1676–1683.
[22] S. Gely, D.F. Lowry, C. Bernard, M. Ringkjobing-Jensen, M. Blackledge, S. Costanzo, et al., Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein, J. Mol. Recognit. 23 (2010) 435–447.
[23] A. Kavalenka, I. Urbancic, V. Belle, S. Rouger, S. Costanzo, S. Kure, et al., Conformational analysis of the partially disordered measles virus NTAIL-XD complex by SDSL EPR spectroscopy, Biophys. J. 98 (2010) 1055–1064.
[24] M. Ringkjøbing Jensen, G. Communie, E.D. Ribeiro Jr., N. Martinez, A. Desfosses, L. Salmon, et al., Intrinsic disorder in measles virus nucleocapsids, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 9839–9844.
[25] M. Fuxreiter, P. Tompa, Fuzzy interactome: the limitations of models in molecular biology, Trends Biochem. Sci. 34 (2009) 3.
[26] P. Tompa, M. Fuxreiter, Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions, Trends Biochem. Sci. 33 (2008) 2–8.
[27] M. Fuxreiter, Fuzziness: linking regulation to protein dynamics, Mol. BioSyst. 8 (2012) 168–177.
[28] M. Miskei, C. Antal, M. Fuxreiter, FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies, Nucleic Acids Res. 45 (2017) D228–D235.
[29] A. D'Urzo, A. Konijnenberg, G. Rossetti, J. Habchi, J. Li, P. Carloni, 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. 26 (2015) 472–481.
[30] B. Morin, J.M. Bourhis, V. Belle, M. Woudstra, F. Carrière, B. Guigliarelli, et al., Assessing induced folding of an intrinsically

disordered protein by site-directed spin-labeling EPR spectroscopy, J. Phys. Chem. B 110 (2006) 20596–20608.
[31] Y. Wang, X. Chu, S. Longhi, P. Roche, W. Han, E. Wang, 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. 110 (2013) E3743–E3752.
[32] M. Fuxreiter, I. Simon, P. Friedrich, P. Tompa, Preformed structural elements feature in partner recognition by intrinsically unstructured proteins, J. Mol. Biol. 338 (2004) 1015–1026.
[33] S.H. Lee, D.H. Kim, J.J. Han, E.J. Cha, J.E. Lim, Y.J. Cho, et al., Understanding pre-structured motifs (PreSMos) in intrinsically unfolded proteins, Curr. Protein Pept. Sci. 13 (2012) 34–54.
[34] G.G. Hammes, Y.C. Chang, T.G. Oas, Conformational selection or induced fit: a flux description of reaction mechanism, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 13737–13741.
[35] Y. Wang, C. Tang, E. Wang, J. Wang, Exploration of multi-state conformational dynamics and underlying global functional landscape of maltose binding protein, PLoS Comput. Biol. 8 (2012), e1002471.
[36] C.J. Oldfield, J. Meng, J.Y. Yang, M.Q. Yang, V.N. Uversky, A.K. Dunker, Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners, BMC Genomics 9 (Suppl. 1) (2008) S1.
[37] A.K. Dunker, M.S. Cortese, P. Romero, L.M. Iakoucheva, V.N. Uversky, Flexible nets, FEBS J. 272 (2005) 5129–5148.
[38] V.N. Uversky, C.J. Oldfield, A.K. Dunker, Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling, J. Mol. Recognit. 18 (2005) 343–384.
[39] C. Haynes, C.J. Oldfield, F. Ji, N. Klitgord, M.E. Cusick, P. Radivojac, et al., Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes, PLoS Comput. Biol. 2 (2006), e100.
[40] B.P. De, A.K. Banerjee, Involvement of actin microfilaments in the transcription/replication of human parainfluenza virus type 3: possible role of actin in other viruses, Microsc. Res. Tech. 47 (1999) 114–123.
[41] S.A. Moyer, S.C. Baker, S.M. Horikami, Host cell proteins required for measles virus reproduction, J. Gen. Virol. 71 (1990) 775–783.
[42] D. Laine, M. Trescol-Biémont, S. Longhi, G. Libeau, J. Marie, P. Vidalain, et al., Measles virus nucleoprotein binds to a novel cell surface receptor distinct from FcgRII via its C-terminal domain: role in MV-induced immunosuppression, J. Virol. 77 (2003) 11332–11346.
[43] D. Laine, J. Bourhis, S. Longhi, M. Flacher, L. Cassard, B. Canard, et al., Measles virus nucleoprotein induces cell proliferation arrest and apoptosis through NTAIL/NR and NCORE/FcgRIIB1 interactions, respectively, J. Gen. Virol. 86 (2005) 1771–1784.
[44] M. Iwasaki, M. Takeda, Y. Shirogane, Y. Nakatsu, T. Nakamura, Y. Yanagi, The matrix protein of measles virus regulates viral RNA synthesis and assembly by interacting with the nucleocapsid protein, J. Virol. 83 (2009) 10374–10383.
[45] H. Sato, M. Masuda, R. Miura, M. Yoneda, C. Kai, Morbillivirus nucleoprotein possesses a novel nuclear localization signal and a CRM1-independent nuclear export signal, Virology 352 (2006) 121–130.
[46] B.R. tenOever, M.J. Servant, N. Grandvaux, R. Lin, J. Hiscott, Recognition of the measles virus nucleocapsid as a mechanism of IRF-3 activation, J. Virol. 76 (2002) 3659–3669.
[47] M. Colombo, J.M. Bourhis, C. Chamontin, C. Soriano, S. Villet, S. Costanzo, et al., The interaction between the measles virus nucleoprotein and the interferon regulator factor 3 relies on a specific cellular environment, Virol. J. 6 (2009) 59.
[48] A. Watanabe, M. Yoneda, F. Ikeda, A. Sugai, H. Sato, C. Kai, Peroxiredoxin 1 is required for efficient transcription and replication of measles virus, J. Virol. 85 (2011) 2247–2253.
[49] X. Zhang, C. Glendening, H. Linke, C.L. Parks, C. Brooks, S.A. Udem, et al., Identification and characterization of a regulatory domain on the carboxyl terminus of the measles virus nucleocapsid protein, J. Virol. 76 (2002) 8737–8746.
[50] X. Zhang, J.M. Bourhis, S. Longhi, T. Carsillo, M. Buccellato, B. Morin, et al., Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus, Virology 337 (2005) 162–174.
[51] M. Couturier, M. Buccellato, S. Costanzo, J.M. Bourhis, Y. Shu, M. Nicaise, 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. 23 (2010) 301–315.
[52] A. Gruet, M. Dosnon, D. Blocquel, J. Brunel, D. Gerlier, R.K. Das, et al., Fuzzy regions in an intrinsically disordered protein impair protein–protein interactions, FEBS J. 283 (2016) 576–594.
[53] M.Y. Kim, Y. Shu, T. Carsillo, J. Zhang, L. Yu, C. Peterson, et al., hsp70 and a novel axis of type I interferon-dependent antiviral immunity in the measles virus-infected brain, J. Virol. 87 (2013) 998–1009.
[54] C.G. Wilson, T.J. Magliery, L. Regan, Detecting protein–protein interactions with GFP-fragment reassembly, Nat. Methods 1 (2004) 255–262.
[55] A. Gruet, M. Dosnon, A. Vassena, V. Lombard, D. Gerlier, C. Bignon, et al., Dissecting partner recognition by an intrinsically disordered protein using descriptive random mutagenesis, J. Mol. Biol. 425 (2013) 3495–3509.
[56] L. Bloyet, J. Brunel, M. Dosnon, V. Hamon, J. Erales, A. Gruet, et al., Modulation of re-initiation of measles virus transcription at intergenic regions by PXD to NTAIL binding strength, PLoS Pathog. 12 (2016), e1006058.
[57] D. Bonetti, F. Troilo, A. Toto, M. Brunori, S. Longhi, S. Gianni, Analyzing the folding and binding steps of an intrinsically disordered protein by protein engineering, Biochemistry 56 (2017) 3780–3786.
[58] T. Mittag, S. Orlicky, W.Y. Choy, X. Tang, H. Lin, F. Sicheri, et al., Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 17772–17777.
[59] R. Rosenzweig, A. Sekhar, J. Nagesh, L.E. Kay, Promiscuous binding by Hsp70 results in conformational heterogeneity and fuzzy chaperone-substrate ensembles, Elife 6 (2017).
[60] A. Sekhar, J. Nagesh, R. Rosenzweig, L.E. Kay, Conformational heterogeneity in the Hsp70 chaperone-substrate ensemble identified from analysis of NMR-detected titration data, Protein Sci. (2017) 2207–2220.
[61] J. Brunel, D. Chopy, M. Dosnon, L.M. Bloyet, P. Devaux, E. Urzua, et al., Sequence of events in measles virus replication: role of phosphoprotein–nucleocapsid interactions, J. Virol. 88 (2014) 10851–10863.
[62] M.J. Oglesbee, Z. Liu, H. Kenney, C.L. Brooks, The highly inducible member of the 70 kDa family of heat shock proteins increases canine distemper virus polymerase activity, J. Gen. Virol. 77 (1996) 2125–2135.
[63] Y. Shu, J. Habchi, S. Costanzo, A. Padilla, J. Brunel, D. Gerlier, et al., Plasticity in structural and functional interactions between

the phosphoprotein and nucleoprotein of measles virus, J. Biol. Chem. 287 (2012) 11951–11967.
[64] K.A. Scott, D.O. Alonso, S. Sato, A.R. Fersht, V. Daggett, Conformational entropy of alanine versus glycine in protein denatured states, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 2661–2666.
[65] W. Borcherds, F.X. Theillet, A. Katzer, A. Finzel, K.M. Mishall, A.T. Powell, et al., Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells, Nat. Chem. Biol. 10 (2014) 1000–1002.
[66] S. Otieno, R. Kriwacki, Probing the role of nascent helicity in p27 function as a cell cycle regulator, PLoS One 7 (2012), e47177.
[67] J.M. Rogers, C.T. Wong, J. Clarke, Coupled folding and binding of the disordered protein PUMA does not require particular residual structure, J. Am. Chem. Soc. 136 (2014) 5197–5200.
[68] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (2005) 3433–3434.
[69] I. Callebaut, G. Labesse, P. Durand, A. Poupon, L. Canard, J. Chomilier, et al., Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives, Cell. Mol. Life Sci. 53 (1997) 621–645.
[70] J.M. Chandonia, M. Karplus, New methods for accurate prediction of protein secondary structure, Proteins 35 (1999) 293–306.
[71] R.G. Smock, M.E. Blackburn, L.M. Gierasch, Conserved, disordered C terminus of DnaK enhances cellular survival upon stress and DnaK in vitro chaperone activity, J. Biol. Chem. 286 (2011) 31821–31829.
[72] D. Reichmann, Y. Xu, C.M. Cremers, M. Ilbert, R. Mittelman, M.C. Fitzgerald, et al., Order out of disorder: working cycle of an intrinsically unfolded chaperone, Cell 148 (2012) 947–957.
[73] A. de Crouy-Chanel, M. Kohiyama, G. Richarme, Interaction of DnaK with native proteins and membrane proteins correlates with their accessible hydrophobicity, Gene 230 (1999) 163–170.
[74] E.M. Clerico, J.M. Tilitsky, W. Meng, L.M. Gierasch, How hsp70 molecular machines interact with their substrates to mediate diverse physiological functions, J. Mol. Biol. 427 (2015) 1575–1588.
[75] X. Zhu, X. Zhao, W.F. Burkholder, A. Gragerov, C.M. Ogata, M.E. Gottesman, et al., Structural analysis of substrate binding by the molecular chaperone DnaK, Science 272 (1996) 1606–1614.
[76] A. Bénarouche, J. Habchi, A. Cagna, O. Maniti, A. Girard-Egrot, J. Cavalier, et al., Interfacial properties of NTAIL, an intrinsically disordered protein, Biophys. J. (2017) (in press).
[77] A. Gruet, S. Longhi, C. Bignon, One-step generation of error-prone PCR libraries using Gateway(R) technology, Microb. Cell Factories 11 (2012) 14.
[78] A. Lamiable, P. Thevenet, J. Rey, M. Vavrusa, P. Derreumaux, P. Tuffery, PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex, Nucleic Acids Res. 44 (2016) W449–54.
[79] A. Micsonai, F. Wien, L. Kernya, Y.H. Lee, Y. Goto, M. Refregiers, et al., Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E3095–103.

Paper 3.

Biophysical Journal
Article

How Robust Is the Mechanism of Folding-Upon-Binding for an Intrinsically Disordered Protein?

Daniela Bonetti,[1] Francesca Troilo,[1] Maurizio Brunori,[1] Sonia Longhi,[2] and Stefano Gianni[1,*]
[1]Istituto Pasteur - Fondazione Cenci Bolognetti, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli," Istituto di Biologia e Patologia Molecolari del CNR, Sapienza Università di Roma, Rome, Italy and [2]Architecture et Fonction des Macromolécules Biologiques, UMR 7257, National Centre for Scientific Research, Aix-Marseille Université, Marseille, France

ABSTRACT The mechanism of interaction of an intrinsically disordered protein (IDP) with its physiological partner is characterized by a disorder-to-order transition in which a recognition and a binding step take place. Even if the mechanism is quite complex, IDPs tend to bind their partner in a cooperative manner such that it is generally possible to detect experimentally only the disordered unbound state and the structured complex. The interaction between the disordered C-terminal domain of the measles virus nucleoprotein (N_{TAIL}) and the X domain (XD) of the viral phosphoprotein allows us to detect and quantify the two distinct steps of the overall reaction. Here, we analyze the robustness of the folding of N_{TAIL} upon binding to XD by measuring the effect on both the folding and binding steps of N_{TAIL} when the structure of XD is modified. Because it has been shown that wild-type XD is structurally heterogeneous, populating an on-pathway intermediate under native conditions, we investigated the binding to 11 different site-directed variants of N_{TAIL} of one particular variant of XD (I504A XD) that populates only the native state. Data reveal that the recognition and the folding steps are both affected by the structure of XD, indicating a highly malleable pathway. The experimental results are briefly discussed in the light of previous experiments on other IDPs.

INTRODUCTION

Intrinsically disordered proteins (IDPs) represent a class of fully functional molecules that lack a well-ordered structure in isolation under physiological conditions (1–7). A typical property of IDPs lies in their capability to undergo a disorder-to-order transition upon recognition and binding to their physiological ligand, a reaction that may lead to highly dynamic complexes typically referred as "fuzzy" (8).

The folding-upon-binding reaction of IDPs is very complex because molecular recognition is intimately coupled to a structural transition. The binding-induced folding of an IDP implies in theory at least two steps: the accumulation of the complex between the two interacting partners and the induced folding (9–11). These two steps might occur in different orders such that folding may precede (conformational selection) or follow (induced fit) binding. However, it has been previously reported that folding and binding tend to be cooperative such that the overall reaction typically occurs in an all-or-none fashion, and a single exponential decay is often observed in experimentally measured time courses (12–17). Such cooperativity is reminiscent of what is typically detected in the folding of globular proteins (18). In this context, the interaction between the intrinsically disordered C-terminal domain of the measles virus nucleoprotein (N_{TAIL}) and the X domain (XD) of the viral phosphoprotein is particularly interesting. In fact, in this case, the fortuitous complexity of the dependence of the observed-rate constant upon (un)binding allows addressing quantitatively the two different steps of the overall process by analyzing quantitatively the dependence of the relaxation-rate constants (19–21). This feature represents a rare opportunity to interrogate directly the experimental system concerning the nature and structural features of the folding and binding of an IDP and possibly propose a generalization.

From a structural perspective, although N_{TAIL} is largely disordered (22), XD is a globular domain of 49 amino acids consisting of a three-helix bundle (Fig. 1) (23,24). Upon binding to XD, the disordered stretch of amino acid residues 489–506 of N_{TAIL} acquires an α-helical folding, the resulting structure of the complex corresponding to a four-helix bundle (Fig. 1) (23–26).

Previous investigations on XD in isolation have indicated that, when folded, this domain is structurally heterogeneous, populating an alternative state similar to an on-pathway folding intermediate (27). This state, although retaining a

Submitted February 7, 2018, and accepted for publication March 20, 2018.
*Correspondence: stefano.gianni@uniroma1.it
Editor: Amedeo Caflisch.
https://doi.org/10.1016/j.bpj.2018.03.017

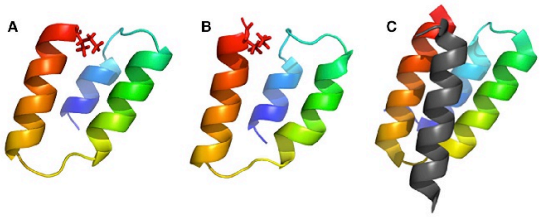


FIGURE 1 Cartoon representation of the three-dimensional structure of the free and bound forms of XD. (*A*) A representation of the native structure of XD is shown. (*B*) A representation of the structure of the on-pathway folding intermediate of XD is shown: it may be seen that although its secondary structure is native-like, its topology is somewhat different. (*C*) A representation of the three-dimensional structure of the complex between native XD (in *rainbow*) and the residues 486–504 of N_{TAIL} (in *gray*) (Protein Data Bank: 1T6O) is given, showing that the complex adopts a four-helix bundle conformation. Residue I504 is highlighted in sticks in (*A* and *B*). To see this figure in color, go online.

native-like secondary structure content, displays some significant differences in the overall topological organization (Fig. 1). Importantly, the stability of the intermediate may be tuned by site-directed mutagenesis such that variants populating solely the native state of XD were successfully obtained (27).

In this work, we resorted to analyzing the robustness of the folding of N_{TAIL} upon binding to XD by analyzing individually the effect of any given perturbation of the structure of XD upon the folding and binding steps of N_{TAIL}. In particular, our strategy was based on comparing the recognition of N_{TAIL} to wild-type XD (characterized by a mixture of native and intermediate states) with a previously characterized variant of XD (I504A) that lacks the intermediate. The results rely on an extensive mutational study wherein we carried out kinetic experiments with 11 different site-directed variants of N_{TAIL} challenged with the two forms of XD. Comparative analysis of the kinetic data shows that the binding and folding steps of N_{TAIL} are both dictated by the structure of XD, indicating that the mechanism of recognition of this IDP is very malleable and appears to be sculpted by the topology of its physiological partner.

MATERIALS AND METHODS

Expression and purification

The variant I504A XD has been already described (27). All the variants of N_{TAIL} have been produced by using the Quick-Change Lightning Site-Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA) according to the manufacturer's instructions. Mutations were all confirmed by DNA sequencing. All proteins were expressed and purified as previously described (19,20).

Temperature-jump experiments

Binding experiments between I504A XD and the variants of N_{TAIL} have been carried out by using a Hi-Tech PTJ-64 capacitor-discharge T-jump apparatus (TgK Scientific, Bradford-on-Avon, United Kingdom). A constant concentration of I504A XD (5 μM) was mixed with N_{TAIL} wild-type and its variants at different concentrations, ranging from 2 to 40 μM, in 50 mM sodium phosphate buffer (pH 7.2) with 300 mM NaCl. Carefully degassed samples were pumped through the 0.5 × 2 mm quartz flow cell before T-jump and data acquisition. Temperature was rapidly increased from 16 to 25°C, with a temperature jump of 9°C. For each experimental condition, 10 traces were usually averaged and fitted to a single exponential equation. The excitation wavelength used was 296 nm, and Trp fluorescence emission was collected using a 320 nm cutoff filter. Control experiments to calibrate the temperature jump were performed by measuring the changes in absorbance of a pH indicator (phenol red) in a buffer of known pH dependence (Tris-HCl) on temperature. Under the experimental condition explored, a discharge of 12 kV corresponded to an increase in temperature of 9°C. Additionally, the fluorescence of N-acetyltryptophanamide was used to monitor the timescale of the rapid heating. In fact, the quantum yield of this amino acid is highly dependent on temperature and represents a classical and simple tool to evaluate the dead time of T-jump devices. Under the condition explored, heating was completed after ~20 μs. Because the timescale of such an optical effect is much faster than that associated with the kinetics of binding between XD and N_{TAIL}, data below 20 μs (reporting on the pretrigger and the heating phase) were excluded from the analysis of the kinetic data.

Data analysis

The fluorescence time courses obtained for N_{TAIL} and its site-directed variants were fitted by using a single exponential equation to obtain the observed-rate constant for the relaxation time. Observed-rate constants were plotted as a function of N_{TAIL} concentration and fitted to the following hyperbolic function arising from an induced-fit model:

$$k_{obs} = \frac{[N_{TAIL}]k_F}{[N_{TAIL}] + K'_D} + k_U,$$

where k_F and k_U are the folding and unfolding rate constants of N_{TAIL}, respectively, and K'_D is the dissociation constant of the complex. $\Delta\Delta G_{K'}$ and $\Delta\Delta G_{folding}$ have been calculated for each variant in the following way:

$$\Delta\Delta G_{K'} = RT\ ln\frac{K_D^{wt}}{K_D^{mut}}$$

and

$$\Delta\Delta G_{folding} = RT\ ln\frac{k_F^{wt}k_U^{mut}}{k_U^{wt}k_F^{mut}}.$$

RESULTS AND DISCUSSION

A classical method to elucidate the mechanism of a chemical reaction is to measure the effect of a perturbation imposed on the system on each detectable kinetic step

(28–31). In the case of the binding reaction between N_{TAIL} and XD, we previously showed that these two proteins interact according to an induced-fit-type mechanism whereby the folding of N_{TAIL} is subsequent to the formation of an initial encounter complex (19,20). In fact, by comparing the reaction kinetics when performing experiments in the presence of an excess of either N_{TAIL} or XD, we observed a superimposable hyperbolic dependence of the observed-rate constants. This feature is a consequence of the symmetry of the reaction scheme of induced fit and is a signature that can be successfully used to exclude conformational selection (32,33). Furthermore, we showed that both the binding and folding steps might be quantitatively addressed by analyzing the dependence of the apparent rate constant on reactant concentration (19,20). Therefore, with the aim of analyzing the robustness of the folding and binding steps of an IDP, we used the variant I504A of XD, which populates solely the native state (27), and challenged this protein with 11 different site-directed variants of N_{TAIL}. The latter have all been recently prepared and employed to characterize the binding between N_{TAIL} and wild-type XD (20), which populates both the intermediate and native state.

Because the binding reaction is too fast to be followed by stopped-flow (19–21), we carried out the kinetic experiments using the temperature-jump methodology, incubating I504A XD at a constant concentration of 5 μM with N_{TAIL} at different concentrations, typically ranging from 2 to 40 μM. The relaxation process was triggered by a rapid increase in temperature of 9°C, from 16 to 25°C, using a fluorescence-equipped capacitor-discharge temperature-jump instrument. In all cases, the time-resolved fluorescence change, corresponding to an increase in emission, is consistent with a single exponential decay and a small perturbation of the equilibrium populations.

The dependence of the observed-rate constant on the concentration of N_{TAIL} for each of the variant is depicted in Fig. 2. Analogous to what we observed for wild-type XD (19–21), the profile of the concentration dependence is consistent for nearly all variants to a hyperbolic behavior, with the exception of A494G and L495A, for which the binding affinities were much too low to obtain reliable kinetics. Like the wild-type (19), all variants were fitted to an induced-fit mechanism as formalized in the Materials and Methods; this allowed determination of the folding and unfolding rate constants as well as the overall affinity of the initial encounter complex, as expressed by its $K'_D = k_{off}/k_{on}$. The calculated parameters for all the variants are listed in Table 1.

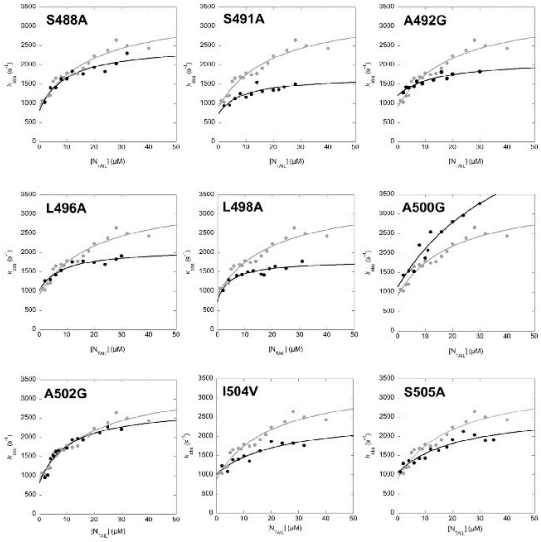


FIGURE 2 Binding kinetics of I504A XD and N_{TAIL} variants. The observed-rate constants, measured by temperature-jump kinetics experiments, have been plotted as a function of the concentration of N_{TAIL}, typically ranging from 2 to 40 μM. I504A XD was kept at a constant concentration of 5 μM. In each plot, gray circles represent the kinetic binding experiments between I504A XD and N_{TAIL} wild-type, and black circles represent the kinetic binding experiments between I504A XD and site-directed variants of N_{TAIL}. For each variant, the overall dependence is consistent with a hyperbolic behavior. Variants A494G and L495A are not shown because their binding affinity was too low to obtain reliable kinetics.

TABLE 1 Kinetic Folding and Binding Parameters for the Induced-Fit Reaction between I504A XD and N_{TAIL} Variants

Variant	k_F (s^{-1})	k_U (s^{-1})	K'_D (μM)	$\Delta\Delta G_{K'}$ (kcal/mol)	$\Delta\Delta G_{folding}$ (kcal/mol)
WT	2500 ± 200	900 ± 60	20 ± 2	NA	NA
S488A	1700 ± 200	800 ± 100	10 ± 1	negative 0.43 ± 0.09	0.2 ± 0.1
S491A	1000 ± 100	720 ± 90	9 ± 1	negative 0.49 ± 0.09	0.4 ± 0.1
A492G	900 ± 100	1210 ± 50	14 ± 1	negative 0.22 ± 0.07	0.78 ± 0.09
L496A	1100 ± 100	1020 ± 70	8 ± 1	negative 0.53 ± 0.09	0.6 ± 0.1
L498A	1000 ± 100	700 ± 100	4.5 ± 0.4	negative 0.87 ± 0.08	0.4 ± 0.1
A500G	5200 ± 500	1100 ± 100	43 ± 4	0.46 ± 0.08	negative 0.3 ± 0.1
A502G	1800 ± 200	800 ± 80	10 ± 2	negative 0.4 ± 0.1	0.12 ± 0.09
I504V	1500 ± 200	1020 ± 80	23 ± 2	0.10 ± 0.08	0.4 ± 0.1
S505A	1600 ± 200	1050 ± 70	22 ± 2	0.06 ± 0.08	0.37 ± 0.09

It is of interest to compare the effect of mutagenesis of N_{TAIL} when this protein is challenged either with wild-type XD, populating both the native and intermediate states, or the I504A mutant, populating solely the native state. Fig. 3 reports a bar chart representation of the calculated $\Delta\Delta G$ for the folding and binding steps. It is evident that the recognition step, as mirrored by the $\Delta\Delta G_{K'}$, and the coupled folding step of N_{TAIL}, expressed by the $\Delta\Delta G_{folding}$, are both strongly affected by the 3D structure of XD, with some of the variants displaying a clearly different behavior as shown by comparing data obtained with wild-type XD and I504A XD. This finding suggests that both steps are influenced by the conformation of XD. Moreover, it should be noticed that, out of all 11 N_{TAIL} variants studied, only L498A bears a substitution in a residue that is in direct contact with residue I504 of XD. Therefore, this is the sole variant for which the observed changes in $\Delta\Delta G_{K'}$ may result from direct contact between the side chains.

In an effort to gain further insight on the effect of the conformation of XD on the folding and binding of N_{TAIL}, we calculated the difference between the $\Delta\Delta G_{K'}$ obtained from kinetic experiments carried out with XD wild-type (populating two states) and that obtained in experiments performed with I504A XD. We also calculated in the same way the difference in $\Delta\Delta G_{folding}$. In this way, for each variant of N_{TAIL}, we obtained a $\Delta\Delta\Delta G_{K'}$ and a $\Delta\Delta\Delta G_{folding}$ (data shown in Table 1), which allowed us to highlight key changes in the structure of the complex (shown in Fig. 4).

In particular, a structural superposition between the intermediate, as obtained from restrained metadynamics using NMR chemical shifts (27), and the native states of XD reveal that the binding pocket for N_{TAIL}, located at the interface between helices 2 and 3, is clearly perturbed. Such a distortion appears to affect substantially the structure of the early recognition complex both in the C-terminal and in the N-terminal regions of N_{TAIL}, with residues displaying a high value of $\Delta\Delta\Delta G_{K'}$ (shown in *blue* in Fig. 4) relative to both wild-type and I504A XD. It should be recalled that the high value of $\Delta\Delta\Delta G_{K'}$ for L498A may result from a direct contact between this residue and position I504 of XD and may indicate that I504 may destabilize the binding of wild-type N_{TAIL} to the intermediate conformation. Interestingly, when folding is considered, it appears that the structural change within the partner and within the early encounter complex perturbs the folding of the whole helix, with clear changes in $\Delta\Delta\Delta G_{folding}$ that extend to the central

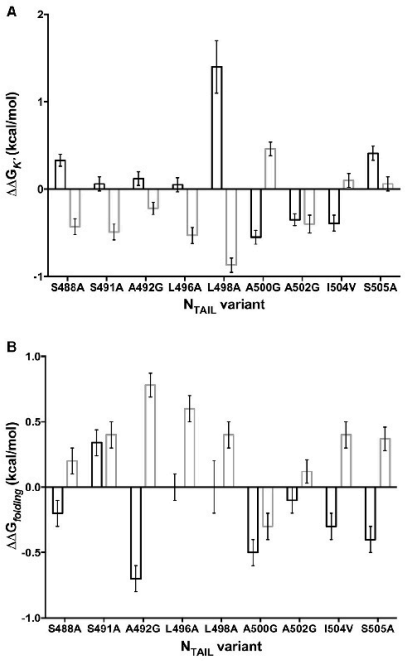


FIGURE 3 Bar chart representation of the calculated $\Delta\Delta G$. For each variant of N_{TAIL}, we show the $\Delta\Delta G$ calculated from experiments performed with XD wild-type (*black bars*) and with I504A XD (*gray bars*), each value being associated to its statistical error. (*A*) The calculated $\Delta\Delta G$ for the binding step ($\Delta\Delta G_{K'}$) is given. (*B*) The calculated $\Delta\Delta G$ for the folding step ($\Delta\Delta G_{folding}$) is given. $\Delta\Delta G$ values obtained with XD wild-type were taken from (20).

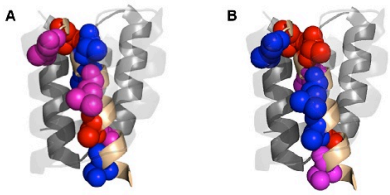


FIGURE 4 Cartoon representation of the $\Delta\Delta\Delta G$ calculated for the complex between N_{TAIL} (*gold*) and XD I504A (*gray*) that populates only the native state. In the background, the structure of the intermediate state of XD is shown. (*A*) The calculated $\Delta\Delta\Delta G_{K^*}$ for complex formation is shown. (*B*) The calculated $\Delta\Delta\Delta G_{folding}$ for N_{TAIL} folding is shown. In each panel, residues displaying $\Delta\Delta\Delta G > 0.7$ kcal/mol are represented in blue, those displaying $\Delta\Delta\Delta G$ between 0.4 and 0.7 kcal/mol are in magenta, and those with $\Delta\Delta\Delta G < 0.4$ kcal/mol are in red. To see this figure in color, go online.

structural regions of the helix. Furthermore, the drastic change in the folding behavior of the A492G mutant, when challenged to the wild-type versus I504A XD, seems to suggest that the secondary structure of N_{TAIL} is destabilized when bound to the intermediate state of XD. A rigorous description of the binding mechanism would demand a complete analysis of the Φ values, which is prevented by the relatively small changes in stability upon mutagenesis; however, we speculate that these experimental observations reflect what is expected in the case of heterogeneous nucleation, whereby the lack of a robust folding nucleus leads to a malleable mechanism, governed by the topological features of the binding partner (34,35). In this case, both folding and binding of the IDP would be highly dependent on the physiological partner.

CONCLUSIONS

The mechanism of recognition of many IDPs is a complex process involving at least two steps, the binding to the partner and the folding of the IDP (1–7,9,11). A comparison of the mechanistic features between the classical spontaneous folding of globular proteins with the binding-induced folding of IDPs revealed some critical differences. In particular, although globular proteins fold via a robust mechanism consolidated by the presence of a loosely formed yet specific nucleus (18,36), IDPs appear to fold by heterogeneous nucleation via an overall mechanism that is induced by interaction with the partner. In fact, an influence of the partner in the binding-induced folding has been proposed by studying the interaction between PUMA and Mcl1 (14,37) and later also observed for the interaction between KIX and c-Myb (35) and MLL (38).

Thanks to the possibility to distinguish between the folding and binding steps, the mechanism of the reaction between N_{TAIL} and XD unveiled in this article allows us to catch an additional glimpse of such a templated mechanism. In fact, mutational analysis of the binding and folding data reveals that both these steps are highly dependent on the fine shape of the physiological partner; this provides additional support to the notion that the energy landscape of IDPs retains a significant level of frustration in both the binding and folding steps, with a limited bias toward the main energetic minimum even once bound to the partner. Future works on different, and structurally more complex, disordered systems will hopefully allow us to test the generality of these conclusions.

AUTHOR CONTRIBUTIONS

M.B., S.L., and S.G. planned the research project. D.B. and F.T. conducted the experiments. All the authors analyzed data. S.G. wrote the main manuscript text. D.B. and F.T. prepared figures, and all the authors reviewed the manuscript.

ACKNOWLEDGMENTS

Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (Progetto di Interesse "Invecchiamento") to S.G., Sapienza University of Rome (C26A155S48, B52F16003410005, and RP11715C34AEAC9B) to S.G., the Associazione Italiana per la Ricerca sul Cancro (Individual Grant - MFAG 2016, 18701) to S.G., and the CNRS to S.L. F.T. is a recipient of a PhD fellowship from the Italo-French University.

REFERENCES

1. Dyson, H. J., and P. E. Wright. 2005. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* 6:197–208.
2. Habchi, J., P. Tompa, …, V. N. Uversky. 2014. Introducing protein intrinsic disorder. *Chem. Rev.* 114:6561–6588.
3. Tompa, P. 2011. Unstructural biology coming of age. *Curr. Opin. Struct. Biol.* 21:419–425.
4. Uversky, V. N. 2002. Natively unfolded proteins: a point where biology waits for physics. *Protein Sci.* 11:739–756.
5. Uversky, V. N., and A. K. Dunker. 2010. Understanding protein non-folding. *Biochim. Biophys. Acta.* 1804:1231–1264.
6. Wright, P. E., and H. J. Dyson. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293:321–331.
7. Wright, P. E., and H. J. Dyson. 2015. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* 16:18–29.
8. Tompa, P., and M. Fuxreiter. 2008. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem. Sci.* 33:2–8.
9. Kiefhaber, T., A. Bachmann, and K. S. Jensen. 2012. Dynamics and mechanisms of coupled protein folding and binding reactions. *Curr. Opin. Struct. Biol.* 22:21–29.
10. Bachmann, A., D. Wildemann, …, T. Kiefhaber. 2011. Mapping backbone and side-chain interactions in the transition state of a coupled protein folding and binding reaction. *Proc. Natl. Acad. Sci. USA.* 108:3952–3957.
11. Gianni, S., J. Dogan, and P. Jemth. 2016. Coupled binding and folding of intrinsically disordered proteins: what can we learn from kinetics? *Curr. Opin. Struct. Biol.* 36:18–24.
12. Ganguly, D., S. Otieno, …, J. Chen. 2012. Electrostatically accelerated coupled binding and folding of intrinsically disordered proteins. *J. Mol. Biol.* 422:674–684.

13. Hill, S. A., L. G. Kwa, …, J. Clarke. 2014. Mechanism of assembly of the non-covalent spectrin tetramerization domain from intrinsically disordered partners. *J. Mol. Biol.* 426:21–35.

14. Rogers, J. M., V. Oleinikovas, …, J. Clarke. 2014. Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *Proc. Natl. Acad. Sci. USA.* 111:15420–15425.

15. Shammas, S. L., A. J. Travis, and J. Clarke. 2014. Allostery within a transcription coactivator is predominantly mediated through dissociation rate constants. *Proc. Natl. Acad. Sci. USA.* 111:12055–12060.

16. Narayanan, R., O. K. Ganesh, …, S. J. Hagen. 2008. Kinetics of folding and binding of an intrinsically disordered protein: the inhibitor of yeast aspartic proteinase YPrA. *J. Am. Chem. Soc.* 130:11477–11485.

17. Dogan, J., T. Schmidt, …, P. Jemth. 2012. Fast association and slow transitions in the interaction between two intrinsically disordered protein domains. *J. Biol. Chem.* 287:34316–34324.

18. Fersht, A. R. 1995. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. USA.* 92:10869–10873.

19. Dosnon, M., D. Bonetti, …, S. Gianni. 2015. Demonstration of a folding after binding mechanism in the recognition between the measles virus NTAIL and X domains. *ACS Chem. Biol.* 10:795–802.

20. Bonetti, D., F. Troilo, …, S. Gianni. 2017. Analyzing the folding and binding steps of an intrinsically disordered protein by protein engineering. *Biochemistry.* 56:3780–3786.

21. Gruet, A., M. Dosnon, …, C. Bignon. 2016. Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. *FEBS J.* 283:576–594.

22. Longhi, S., V. Receveur-Bréchot, …, B. Canard. 2003. 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.* 278:18638–18648.

23. Gely, S., D. F. Lowry, …, S. Longhi. 2010. Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein. *J. Mol. Recognit.* 23:435–447.

24. Johansson, K., J. M. Bourhis, …, S. Longhi. 2003. Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *J. Biol. Chem.* 278:44567–44573.

25. Kingston, R. L., D. J. Hamel, …, B. W. Matthews. 2004. Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc. Natl. Acad. Sci. USA.* 101:8301–8306.

26. Xue, B., D. Blocquel, …, S. Longhi. 2014. Structural disorder in viral proteins. *Chem. Rev.* 114:6880–6911.

27. Bonetti, D., C. Camilloni, …, S. Gianni. 2016. Identification and structural characterization of an intermediate in the folding of the measles virus X domain. *J. Biol. Chem.* 291:10886–10892.

28. Fersht, A. R. 1999. Structure and Mechanism in Protein Science. W. H. Freeman, New York.

29. Fersht, A. R. 2004. Relationship of Leffler (Bronsted) alpha values and protein folding Phi values to position of transition-state structures on reaction coordinates. *Proc. Natl. Acad. Sci. USA.* 101:14338–14342.

30. Fersht, A. R., A. Matouschek, and L. Serrano. 1992. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* 224:771–782.

31. Leffler, J. E. 1953. Parameters for the description of transition states. *Science.* 117:340–341.

32. Gianni, S., J. Dogan, and P. Jemth. 2014. Distinguishing induced fit from conformational selection. *Biophys. Chem.* 189:33–39.

33. Olson, S. T., K. R. Srinivasan, …, J. D. Shore. 1981. Binding of high affinity heparin to antithrombin III. Stopped flow kinetic studies of the binding interaction. *J. Biol. Chem.* 256:11073–11079.

34. Giri, R., A. Morrone, …, S. Gianni. 2013. Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. *Proc. Natl. Acad. Sci. USA.* 110:14942–14947.

35. Toto, A., C. Camilloni, …, S. Gianni. 2016. Molecular recognition by templated folding of an intrinsically disordered protein. *Sci. Rep.* 6:21994–22000.

36. Itzhaki, L. S., D. E. Otzen, and A. R. Fersht. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol.* 254:260–288.

37. Rogers, J. M., C. T. Wong, and J. Clarke. 2014. Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *J. Am. Chem. Soc.* 136:5197–5200.

38. Toto, A., and S. Gianni. 2016. Mutational analysis of the binding-induced folding reaction of the mixed-lineage leukemia protein to the KIX domain. *Biochemistry.* 55:3957–3962.

Paper 4.

The fuzzy appendage of measles virus NTAIL hampers interaction with XD through a combination of entropy and enthalpy

Troilo F., Bonetti D., Longhi S., Gianni S.
IN PREPARATION.

INTRODUCTION

Intrinsically disordered proteins (IDPs) are proteins that lack a well-defined structure in the absence of a partner or ligand in physiological conditions. Despite the lack of structure, these proteins are completely functional (Wright & Dyson, 1999)(Dunker et al., 2001)(Dyson, 2011) (Habchi, Tompa, Longhi, & Uversky, 2014)(Uversky & Dunker, 2010). IDPs can be either totally or partially disordered. In the latter case, they contain both folded and intrinsically disordered regions (IDRs). Many IDPs/IDRs contain Molecular Recognition Elements (MoREs), i.e. regions that undergo a disorder-to-order transition upon interaction with the physiological partner (Dyson & Wright, 2005)(Dyson & Wright, 2002).

The folding of an IDP/IDR can occur *via* two principal mechanisms: the induced fit (the IDP/IDR folds after the binding with the partner) and the conformational selection (the protein folds before the binding).

Often, the regions flanking MoREs maintain a considerable amount of disorder also in the bound form. These regions are called fuzzy and they lead to the formation of fuzzy complexes.

N_{TAIL} is the C-terminal (ammino acids 400-525) domain of the Measles Virus nucleoprotein and it is an intrinsically disordered domain. N_{TAIL} contains a MoRE (ammino acids 486-502) that, after the binding with the C-terminal domain X (XD) of the Phosphoprotein (P), undergoes an α-helical folding. The P protein with the RNA-dependent RNA polymerase (Large protein) forms the polymerase complex. The interaction between N_{TAIL} and XD, in the viral context, is required to allow the recruitment of the polymerase complex onto the encapsidated genome (non segmented, negative-sense single stranded RNA) to allow transcription and replication to take place.

The N_{TAIL}-XD complex is an illustrative example of "fuzziness" since, upon the binding with XD, the regions flanking the α-MoRE remain disordered (Longhi, 2012).

The N_{TAIL}-XD binding reaction can be described by two steps: the binding between XD and the unfolded N_{TAIL} and the subsequent folding of the latter. Kinetics experiments of the interaction between these two domains allowed us to calculate the observed rate constant (k_{obs}) describing the overall reaction. The k_{obs} between N_{TAIL} and XD shows a hyperbolic dependence on reactant concentration that consents to separate the two steps of the reaction: the binding, bimolecular step, occurring at low reactant concentrations, and the folding, monomolecular step, occurring at high reactant concentrations. Therefore, the reaction is limited, at high reactant concentrations, by the folding rate of N_{TAIL} (Dosnon et al., 2015). By comparing the kinetics of the interaction, performed under pseudo-first order conditions (i.e. in presence of a large excess of one of the two reactants) with respect to N_{TAIL} or XD, the MoRE region was shown to fold after the binding with XD according to the so called induced fit mechanism (Dosnon et al., 2015)(Gianni, Dogan, & Jemth, 2014).

A study published by Gruet *et al* in 2016 showed that the binding between N_{TAIL} and XD is stronger in the absence of the fuzzy region, with the self-inhibitory effect being apparently not dependent on the amino acidic sequence and just only on the sequence length. In the same work the authors also performed kinetics binding experiments using either the full length N_{TAIL} (aa 401-525) or a peptide mimicking just the α-MoRE region of N_{TAIL}, and showed that the fuzzy region reduces the folding rate of the α-MoRE (Gruet et al., 2016).

In this work we studied the kinetics of the interaction between different variants of N_{TAIL} and XD in order to understand the role of the fuzzy region on the α-MoRE folding. The kinetic analysis of the interaction between different truncation variants, an artificial variant and single site variants of N_{TAIL} and XD showed that the fuzzy region, and in particular the sequence between the residues 451 and 481 hampers the folding rate of the α-MoRE with an effect dependent both on the length of the sequence (entropy) and on single amino acidic interactions (enthalpy).

MATERIALS AND METHODS

Constructs generation

The constructs encoding N-terminally heaxahistidine tagged N_{TAIL} truncation variants (from 401 to 481) with a tobacco etch virus (TEV) protease cleavage site (amino acidic sequence: ENLYFQGS) immediately upstream of the codons encoding amino acids 401, 411, 421, 431, 441, 451, 461, 471 and 481 of N_{TAIL} were already available (Gruet *et al*. 2016).

The N_{TAIL} truncation variants were also cloned into the pETG20A expression vector that drives the expression of the protein of interest fused to a Thioredoxin (Trx) solubility tag. To this end, the already available pDEST14 constructs encoding N-terminally hexahistidine tagged N_{TAIL}, with

a TEV protease cleavage site at different N_{TAIL} positions, were used as a templates in a first PCR amplification that used *att*B2 as reverse primer and as forward primer an oligonucleotide encompassing the *att*B1 site followed by the sequence encoding the TEV protease cleavage site, and by a sequence annealing with the 17 N_{TAIL} nucleotides downstream of the TEV cleavage site. After digestion with DpnI (1 hour at 37°C), 1 µl of the first PCR reaction was used as template for a second PCR amplification performed using primers *att*L1a and *att*L2a. These primers encode approximately half of the *att*L1 and *att*L2 recombination sites, respectively, and have been shown to be good substrates for LR clonase (Fu, Wehr, Edwards, & Hauge, 2008). As the *att*L1a and *att*L2a sequences encompass the *att*B1 and *att*B2 sequences, respectively, amplicons flanked by *att*B sequences as obtained from the first PCR product provide the necessary homology for *att*L1a and *att*L2a primer hybridization, respectively. 1.5 µl of the second PCR product were used in a 5 µl LR reaction to insert the truncation variant sequences into the pETG20A plasmid by Gateway cloning. These steps were performed to produce all the truncation variant constructs bearing the Trx tag.

An already available construct encoding the artificial N_{TAIL} sequence ($artN_{TAIL}$) cloned into the pNGG plasmid (Gruet et al., 2016) was used as a template in a first PCR amplification aimed at inserting the TEV protease cleavage site upstream residue 401. The forward primer contained the *att*B1 site, followed by the sequence encoding for the TEV protease cleavage site and by 24 nucleotides annealing with the 5' region of $artN_{TAIL}$. The reverse primer was *att*B2. After digestion with DpnI, the artificial amplicon bearing the TEV cleavage site was used as a template for a second PCR amplification with primers attL1a and attL2a. The resulting amplicon was then inserted in the pETG20A plasmid through an LR reaction, as described for the truncation variants.

The constructs encoding the site directed variants of N_{TAIL} were obtained using the gene encoding *wild-type* (*wt*) N_{TAIL}, inserted in the pET28 expression vector, as template to perform site-directed mutagenesis using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent technologies) according to the manufacturer's instructions. All substitutions were conservative and were confirmed by DNA sequencing.

Protein expression and purification

N_{TAIL} single site-directed variants were expressed in the *Escherichia coli* Rosetta T7 pLysS (Novagen) strain. Cultures were grown in Luria Bertani (LB) medium containing 100 μg/ml ampicillin and 34 μg/ml chloramphenicol at 37°C until the OD_{600} reached 0.6-0.8 and then protein expression was induced with 1 mM IPTG (isopropyl-β-D-thiogalactopyranoside). After induction, cells were grown at 25°C over night and then collected by centrifugation. Cells were resuspended in 5ml/gr of pellet of buffer A (50 mM sodium phosphate pH 7.2, 300 mM NaCl) supplemented with 0.1 mg/mL lysozyme, 10 μg/mL DNase I and protease inhibitor cocktail (Sigma) and disrupted by sonication. The lysate was clarified by centrifugation at 12,000 rpm for 40 minutes at 4 °C. The soluble fraction of the bacterial lysate was loaded onto a 5 ml His trap FF column (GE, Healthcare) preloaded with Ni^{2+} ions, previously equilibrated in buffer A. After a washing step with buffer A supplemented with 1 M NaCl, the proteins were eluted with buffer A containing 250 mM imidazole. The proteins were then concentrated until 1 – 2 ml and loaded onto a Superdex 75 16/60 column (GE, Healthcare). Isocratic elution was carried out in SEC buffer (10 mM sodium phosphate pH 7.0, 150 mM NaCl).

The N_{TAIL} truncation variants and artN_{TAIL} were expressed in Terrific Broth (TB) medium containing 100 μg/ml ampicillin and 34 μg/ml

chloramphenicol. The *E. coli* strain Rosetta T7 pLysS (Novagen) was used. Cells were grown in the same conditions described for the single-site N_{TAIL} variants except that cells were collected by centrifugation after 4 hours at 37°C. Cell lysis and the first step of purification through a 5 ml His trap FF column (GE, Healthcare) were carried out as described above for the single site variants. In this case, the buffer used to equilibrate the column was 50 mM Tris/HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 1 mM phenyl-methyl-sulphonyl-fluoride (PMSF). After sample injection, the column was washed with 50 mM Tris/HCl pH 8.0, 1 M NaCl, 20 mM Imidazole. Elution was carried out with 50 mM Tris/HCl pH 8.0, 300 mM NaCl, 250 mM Imidazole. Eluents from IMAC were desalted using a desalting column (GE, Healthcare) previously equilibrated with desalting buffer (20 mM Tris/HCl pH 8.0, 300 mM NaCl). Cleavage with TEV protease was carried out overnight at 4°C in desalting buffer using 1 mg of TEV protease per 20 mg of target protein. Samples were then loaded onto 2.5 ml Ni^{2+} IDA Agarose resin, previously equilibrated in desalting buffer, and incubated 15 minutes at 4°C. The non-retained fraction was recovered, concentrated up to 1 – 2 ml, and then loaded onto a Superdex 75 16/60 column (GE Healthcare) and eluted in SEC buffer.

The pseudo-*wt* XD variant Y480W was expressed and purified as previously reported in (Dosnon et al., 2015).

The purity of the all proteins was confirmed by SDS-PAGE and the concentration was calculated measuring the absorbance at 280 nm, using the theoretical extinction coefficient, as provided from ExPASy (https://www.expasy.org/), and applying the Lambert-Beer equation.

Temperature-jump measurements

Relaxation binding experiments were performed by using a Hi-Tech PTJ-64 capacitor-discharge temperature-jump apparatus (Hi-Tech, Salisbury, UK). The fluorescence change of N-acetyltryptophanamide (NATA) was used in control measurements. The experiments were carried out in pseudo first order conditions by mixing a constant concentration of pseudo *wt* Y480W XD (5 μM) with different N_{TAIL} concentrations usually ranging from 2 uM to 50 uM. The buffer used was 10 mM sodium phosphate pH 7.0, 150 mM NaCl. Degassed samples were slowly pumped through the 0.5x2 mm quartz flow cell where a rapid discharge of 12 kV leads to a temperature jump of 9°C (from 16 to 25°C). Usually 10-20 traces were averaged for each sample. The excitation wavelength was 296 nm and the change in the fluorescence emission of the tryptophan was measured as a function of the time using a 320 nm cut-off filter.

The same experiments were performed using the truncation variants, the artificial variant of N_{TAIL} and all the single site variants of N_{TAIL}.

Data analysis

The fluorescence time courses obtained for N_{TAIL} truncation variants, artN_{TAIL}, and N_{TAIL} site-directed variants were fitted by using the following single exponential equation

$$F(t) = \Delta F \cdot e^{(k_{obs} \cdot t)} + F_{\infty}$$

where k_{obs} is the observed rate constant, ΔF is the amplitude of the trace, t is the time of acquisition of the trace and F_{∞} is the final value of the fluorescence. Observed rate constants were plotted as a function of N_{TAIL} concentration and fitted to the following hyperbolic function, arising from an induced-fit model:

$$k_{obs} = \frac{[N_{TAIL}] \cdot k_F}{[N_{TAIL}] + K'_D} + k_U$$

where k_F is the folding rate constant and k_U is the unfolding rate constants of N_{TAIL} and K'_D is the dissociation constant of the initial complex between N_{TAIL} unfolded and XD (Bonetti et al., 2017).

Observed rate constants obtained for the shorter N_{TAIL} truncation variants (451, 461, 471, 481) were plotted as a function of N_{TAIL} concentration and fitted to the following linear equation:

$$k_{obs} = k_{on}^{app} \cdot [N_{TAIL}] + k_{off}^{app}$$

RESULTS AND DISCUSSION

N_{TAIL} truncation variants

The constructs encoding for truncation variants of MeV N_{TAIL}, already available in the pDEST14 plasmid (Gruet et al., 2016), produced too low quantities of proteins. Therefore, to increase the yield, the constructs were cloned in another vector, pETG20a, bearing the Thioredoxin tag (Trx) that increases protein solubility. The scheme of the constructs is shown in Figure 1.

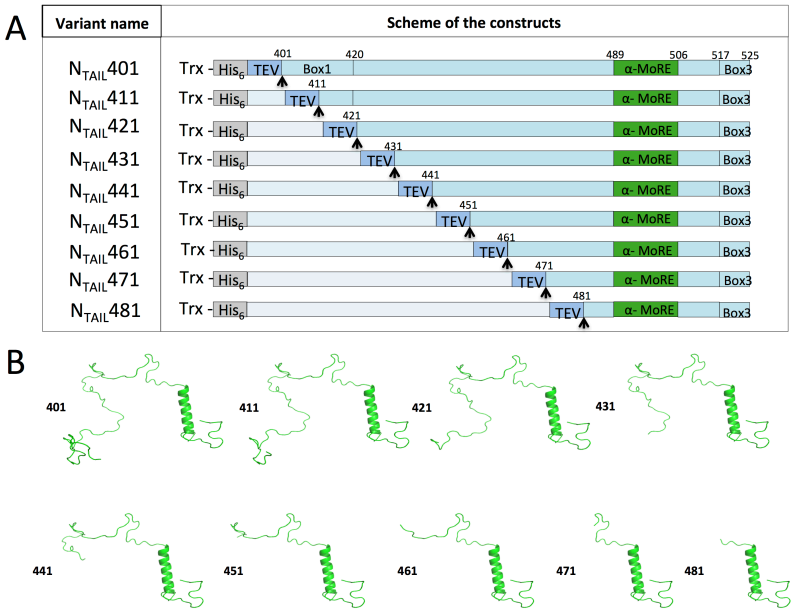


Figure 1. Panel A: Schematic representation of the truncation variant constructs cloned in the pETG20a plasmid. The Trx tag is followed by the Histidine tag and by the sequence of N_{TAIL}. The TEV site is located in such a way that, after protease cleavage, the 9 truncation variants are obtained. Each variant is 10 amino acids shorter than the previous one in the series as represented in panel B.

The temperature-jump experiments performed between each truncation variant and the pseudo-*wt* XD Y480W (that allows to follow the reaction through the change in Trp fluorescence upon the binding), under pseudo-first order conditions (i.e. XD was kept constant at low concentration of 5 μM) revealed for the longer variants (401 – 411 – 421 – 431 – 441) a hyperbolic trend of the k_{obs} similar to the *wt,* full-length, $N_{TAIL,}$ while, for the

shorter variants (451 – 461 – 471 – 481) the k_{obs} presented a linear trend as a function of N_{TAIL} concentrations as shown in Figure 2. Since it was already demonstrated that the hyperbolic trend of k_{obs} depends by the fact that at high proteins concentrations the rate of folding of the α-MoRE is the limiting step of the overall reaction (Dosnon et al., 2015), the shift from a hyperbolic to a linear behaviour of the k_{obs} means that the folding rate of the α-MoRE increases. In particular, the k_{obs} appears to increase proportionally with increasing removal of the fuzzy region from residue 441.

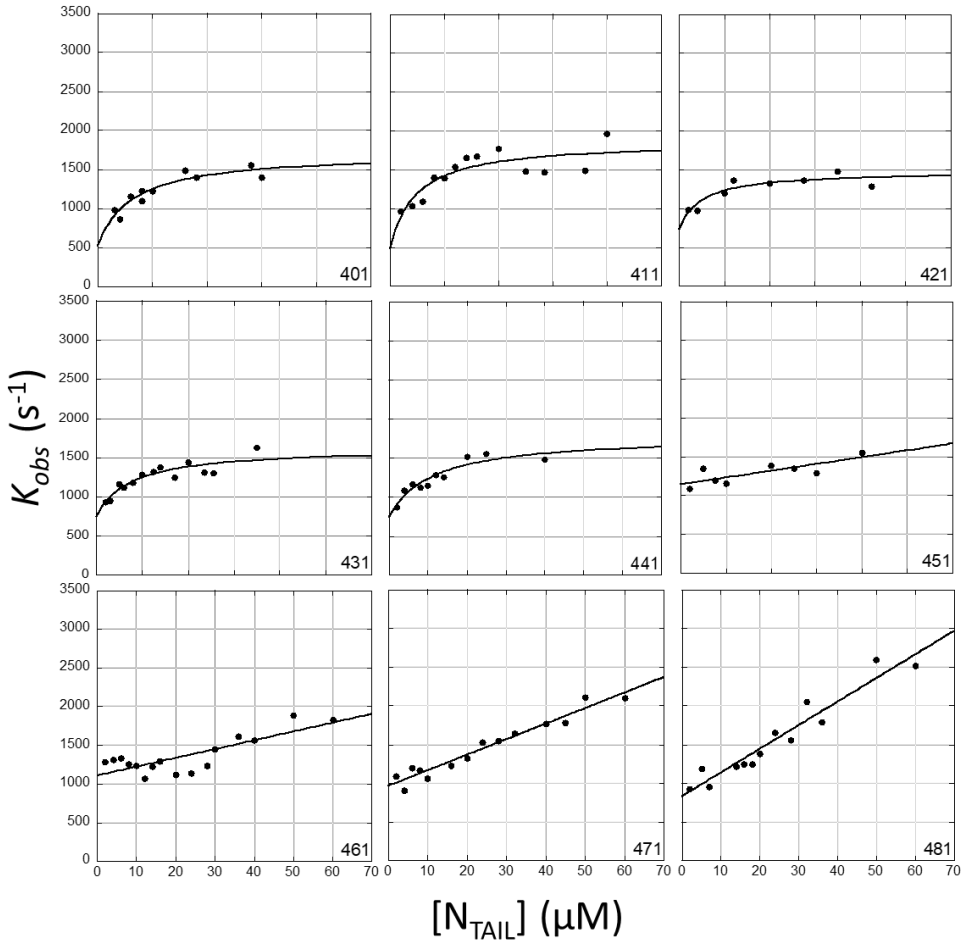


Figure 2. Kinetics of binding of N_{TAIL} truncated variants to XD. Observed rate constants were measured as a function of N_{TAIL} concentration (typically ranging from 1 to 60 μM) by temperature-jump experiments, in the presence of XD Y480W at a constant concentration (5 μM).

Artificial N_{TAIL}

To assess whether the effect of the fuzzy region on the folding rate of N_{TAIL} depended just on the length of the sequence (i.e. is due to an entropic

effect) or whether the interactions between specific residues were also involved, a temperature-jump experiment was performed, in the same conditions, between XD and an artificial variant of N_{TAIL} ($artN_{TAIL}$) already described in Gruet et al. 2016. This $artN_{TAIL}$ presents, in the region between the residues 401 and 471, an amino acidic sequence completely different with respect to the *wt*. Surprisingly, the kinetics of binding of *wt* N_{TAIL} and of $artN_{TAIL}$ to XD, showed the same hyperbolic behaviour (Figure 3) suggesting that the fuzzy region hampers the rate of folding of N_{TAIL} through steric hindrance and not through specific amino acid interactions.

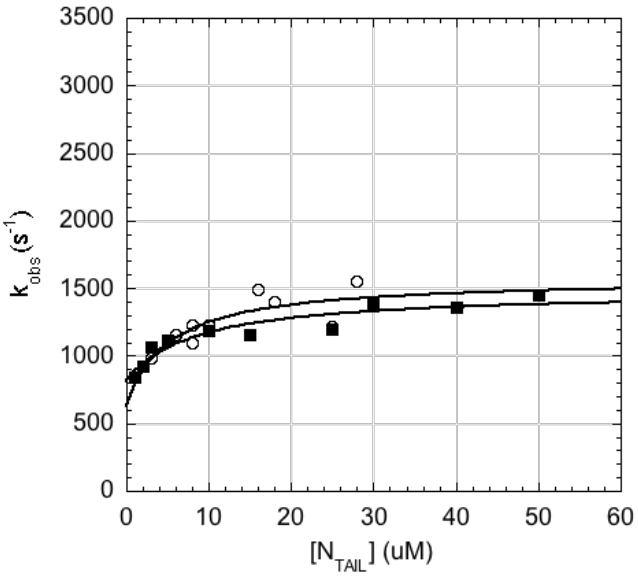


Figure 3. Kinetics of binding of *wt* N_{TAIL} (circle) and of $artN_{TAIL}$ (square) variants to XD. Both variants present the same hyperbolic dependence of k_{obs} as a function of N_{TAIL} concentrations (from 2 to 50 µM)

Site-directed N_{TAIL} variants

To confirm this result we decided to perform single site substitutions, in the context of the full-length N_{TAIL}, in the region between the residues 451 and 482. Each variant was used to perform kinetic binding experiments with XD. In case that the effect of the fuzzy region is indipendent on the amino acidics interactions, all the variants were expected to have a kinetics of interaction with XD similar to the *wt*. Surprisingly, as shown in Figure 4, in some cases the k_{obs} showed a trend dramatically different from that of *wt* N_{TAIL}. In particular, the substitutions Y451A, T454S, A459G, A465G, H466A, and L467A were found to lead to a linear dependence of k_{obs} as a function of N_{TAIL} concentrations. This finding indicated that these residues are involved in specific interactions leading to a decrease in the rate of folding of the α-MoRE.

Altogether, results lead to the conclusion that the fuzzy region of N_{TAIL} hampers the rate of folding of the α-MoRE through a combination of enthropic and enthalpic effects.

Physiologically the role of the fuzzy region in dampering the overall interaction (the binding affinity and the rate of folding of the α-MoRE) between N_{TAIL} and XD could be a strategy adopted by the virus to finely regulate the transcription and replication of the viral genome. Indeed, since the N_{TAIL} – XD interaction strength is a critical determinant in dictating the relative abundance of the viral transcripts (Bloyet et al., 2016) and hence the relative amounts of the various viral components, it needs to be tightly controlled.

From a structural point of view, it should be interesting, to perform NMR studies of the complexes between XD and the single-site N_{TAIL} variants studied, so as to assess whether the substitutions in the fuzzy region affect also the conformation of the α-MoRE.

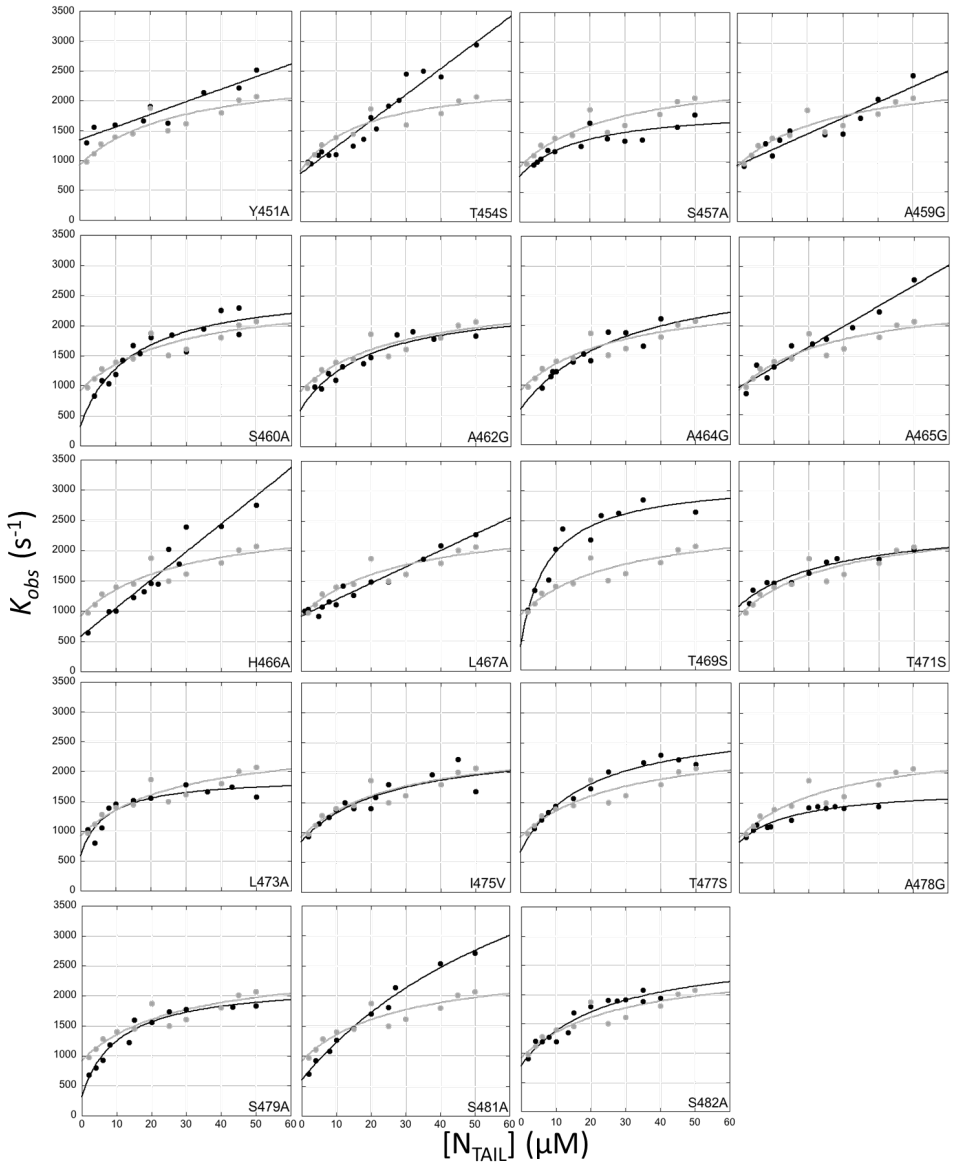


Figure 4. Kinetics of binding of N_{TAIL} and its site-directed variants to XD. Observed rate constants were measured as a function of N_{TAIL} concentration (typically ranging from 1 to 50 μM) by temperature-jump experiments, in the presence of XD Y480W at a constant concentration (5 μM). Gray circles: *wt* N_{TAIL}; black circles: site-directed N_{TAIL} variants.

REFERENCES

Bloyet, L. M., Brunel, J., Dosnon, M., Hamon, V., Erales, J., Gruet, A., Lazert, C., Bignon, C., Roche, P., Longhi, S., & Gerlier, D. 2016. RESEARCH ARTICLE Modulation of Re-initiation of Measles Virus Transcription at Intergenic Regions by PXD to NTAIL Binding Strength. ***PLOS Pathogens***, 1–39.

Bonetti, D., Troilo, F., Toto, A., Brunori, M., Longhi, S., et al. 2017. Analyzing the Folding and Binding Steps of an Intrinsically Disordered Protein by Protein Engineering. ***Biochemistry***, 56(29): 3780–3786.

Dosnon, M., Bonetti, D., Morrone, A., Erales, J., Di Silvio, E., et al. 2015. Demonstration of a folding after binding mechanism in the recognition between the measles virus NTAIL and X domains. ***ACS Chemical Biology***, 10(3): 795–802.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., et al. 2001. Intrinsically disordered protein. ***Journal of Molecular Graphics and Modelling***, 19(1): 26–59.

Dyson, H. J. 2011. Expanding the proteome: disordered and alternatively-folded proteins H. ***Q Rev Biophys***, 44(4): 467–518.

Dyson, H. J., & Wright, P. E. 2002. Coupling of folding and binding for unstructured proteins. ***Current Opinion in Structural Biology***, 12(1): 54–60.

Dyson, H. J., & Wright, P. E. 2005. Intrinsically unstructured proteins and their functions. ***Nature Reviews Molecular Cell Biology***, 6(3): 197–208.

Fu, C., Wehr, D. R., Edwards, J., & Hauge, B. 2008. Rapid one-step recombinational cloning. ***Nucleic Acids Research***, 36(9): 1–7.

Gianni, S., Dogan, J., & Jemth, P. 2014. Distinguishing induced fit from conformational selection. ***Biophysical Chemistry***, 189: 33–39.

Gruet, A., Dosnon, M., Blocquel, D., Brunel, J., Gerlier, D., et al. 2016.

Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. ***FEBS Journal***, 283(4): 576–594.

Habchi, J., Tompa, P., Longhi, S., & Uversky, V. N. 2014. Introducing protein intrinsic disorder. ***Chemical Reviews***, 114(13): 6561–6588.

Uversky, V. N., & Dunker, A. K. 2010. Understanding protein non-folding. ***Biochimica et Biophysica Acta - Proteins and Proteomics***, 1804(6): 1231–1264.

Wright, P. E., & Dyson, H. J. 1999. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. ***Journal of Molecular Biology***, 293(2): 321–331.

CONCLUSIONS AND PROSPECTS

Conclusion 1.

The mutational, kinetics and Φ-value analysis of MeV N_{TAIL} allowed investigating the contribution of single residues in the mechanism of the folding upon binding to XD.

For the binding reaction, the Φ-values analysis indicated that in the transition state the complex between N_{TAIL} and XD presents a native-like (i.e. close to the bound one) structure that gradually tapers off to the sides, showing high Φ-values of binding (> 0.7) in the C-terminal portion of the α-MoRE and lower values in the N-terminal portion. Moreover the Φ-values analysis revealed that the central part of the helix, and in particular the residues 494, 495 and 498, are involved in the interactions driving the initial encounter between N_{TAIL} and XD.

The analysis of the folding reaction resulted more complicated, because the variation in the free energy of folding generated by the substitutions was, in the most of the cases, too low to obtain reliable Φ-values, preventing a complete structural analysis of the folding transition state. However, the LFER plot analysis allowed obtaining information about the folding. The linear relationship between the changes in the free energy of the transition state and the changes in equilibrium free energy, indicates that N_{TAIL} folds via the so called nucleation-condensation mechanism. Moreover, the high α-value obtained indicates that N_{TAIL} folds *via* a transition state that resembles a distorted but similar version of the native state (i.e. bound state).

Another interesting data obtained by the mutational analysis is that only one of the 11 variants produced, appears to destabilize the folding step of N_{TAIL} while all the others, surprisingly, appear to stabilize it. This suggests that N_{TAIL} retains a high level of frustration also in the bound form and is actually not evolved to fold. Moreover, this result indicates that the folding of

N_{TAIL}, after the initial encounter with XD, is a spontaneous process driven by the overall binding. Biologically, this poor propensity to fold of N_{TAIL}, may reflect a physiological strategy adopted by the virus to finely regulate the N_{TAIL} – XD interaction in order to obtain an efficient transcription and replication. Indeed, since the N_{TAIL} – XD interaction strength is a critical determinant in dictating the relative abundance of the viral transcripts and hence the relative amounts of the various viral components, it needs to be tightly controlled.

Conclusion 2.

The analysis of the binding affinity between XD or hsp70 and variants of N_{TAIL} obtained by alanine scanning mutagenesis, truncation and rational site-directed mutagenesis, showed that although XD and hsp70 interact with the same region of N_{TAIL}, the binding mechanisms are different. Hsp70 is in fact much more tolerant of N_{TAIL} substitutions than XD (likely because of the lower affinity of the interaction between N_{TAIL} and hsp70 compared to that with XD). In addition, the majority of substitutions in N_{TAIL} lead to an increase in the interaction strength with hsp70 suggesting that the N_{TAIL}-hsp70 interaction is more evolvable compared to the N_{TAIL}-XD interaction. Moreover, binding experiments performed between hsp70 and N_{TAIL} variants presenting an increased or decreased α-helicity of the MoRE indicate that the MoRE probably doesn't fold into an α-helix upon interacting with hsp70, as judged from the fact that there are no significant differences in the binding affinity when the α-helicity of the MoRE is increased or decreased. In contrast, the α-helical folding of the MoRE was found to be necessary for the interaction with XD as showed also by the kinetic analysis of the interaction between XD and the AlaMoRE (increased α-helicity) and GlyMoRE (increased disorder) peptides. Those studies that showed that an increase in

disorder in the α-MoRE region (i.e. GlyMoRE) leads to a reduction of the folding rate of the α-MoRE upon binding to XD. The same experiment couldn't be performed with hsp70, this protein too being big and complicated to be studied by kinetic experiments.

Therefore, the intrinsically disordered domain N_{TAIL} interacts with different mechanisms and undergoes different disorder-to-order transitions depending by the interacting partner.

Conclusion 3.

The mutational and kinetic analysis (Φ-value analysis) of the interaction between N_{TAIL} and the variant of XD I504A (a variant not populating the intermediate form in the native state) compared with the Φ-value analysis performed with XD *wt* (populating an on-pathway intermediate of folding in the native state that shows a native-like secondary structure content but a different topological organization) revealed that both the binding and the folding steps of the N_{TAIL}- XD interaction are highly dependent by the shape of XD, suggesting that this IDP folds by heterogeneous nucleation *via* a mechanism induced by the shape of the partner (templated folding).

Conclusion 4.

Kinetic studies carried out with XD and artN_{TAIL} (a variant bearing a sequence highly dissimilar and slightly more disordered than *wt* N_{TAIL}) showed that artN_{TAIL} has the same hyperbolic behavior as *wt* N_{TAIL}. This indicates that the decrease in the rate of folding brought by the fuzzy appendage is sequence-independent and suggests a purely entropic mechanism.

The kinetic studies performed with XD and the truncation variants of MeV N_{TAIL} (obtained removing by blocks of ten amino acids the fuzzy region) revealed that the fuzzy region decreases the rate of folding of the α-MoRE in a non-monotonic manner. In particular, while the first variants in the set have a hyperbolic behavior, starting from variant 451, the kinetics becomes linear, indicating that the region encompassing amino acids 451 to 481 has a critical impact on the rate of folding of the α-MoRE. This behavior could either arise from a purely entropic mechanism (a steric hindrance effect brought by the fuzzy region) or, by contrast, could reflect some possible sequence-specific effects. Kinetic experiments performed using the site directed variants in the region 451-481, in the context of the full length N_{TAIL}, suggested that the effect on the rate of folding of the α-MoRE has also an enthalpic nature (i.e. due to the interactions between residues) since some single substitutions in the fuzzy region lead to an increase in the folding rate of $N_{TAIL.}$ Therefore, in conclusion, the fuzzy region of N_{TAIL} hampers the folding rate of the α-MoRE through a combination of entropic and enthalpic effects.

All together the results obtained in this thesis contributed to improve our knowledge on intrinsically disordered proteins and the mechanisms of interaction with their partners. The results obtained have the potential to be applied to IDPs in general.

First of all, we can conclude that there is no general mechanism for the binding-induced folding of IDPs, confirming the extreme polymorphism of IDPs that are able to interact with different partners *via* different mechanisms and undergo different disorder-to-order transitions depending on the interacting partners and also on the structure of the same partner. Another interesting feature of IDPs highlighted by this work is that contrary to what is

observed for globular proteins, IDPs appear to fold by heterogeneous nucleation *via* an overall mechanism induced by the interaction with the partner and depending on the structure of the partner. Moreover, as showed by the mutational and kinetics analysis of N_{TAIL}, it is reasonable to assume that the energy landscape of IDPs, also when bound to their partners, may retain a high level of frustration, with a limited bias toward the main energetic minimum, representing a plausible explanation of the high dynamic nature of IDPs (fuzziness).

Future Prospects.

The results obtained during my PhD set the basis for future studies that will lead to a better understanding of the role of structural disorder in the N_{TAIL}-XD interaction. In particular, it would be interesting to perform NMR studies of the complexes between XD and the single-site N_{TAIL} variants studied, so as to assess whether the variations in the fuzzy regions affect the conformation of the α-MoRE. Another interesting aspect to examine more in depth is the interaction between N_{TAIL} and hsp70 that, being different with respect to the N_{TAIL}-XD interaction, deserves investigation *per se* and is expected to enlarge our knowledge about the different mechanisms by which IDPs bind their partner(s) (i.e. polymorphism).

From a more general point of view, the study of the N_{TAIL}-XD interaction from Measles Virus could be very important in order to find a suitable antiviral approach to fight this re-emerging human pathogen. This interaction, in fact, is highly important for the virus being fundamental to recruit the polymerase complex on the genome and also for polymerase processivity and transcription re-initiation efficiency. Therefore, the results obtained by our mutational analysis, and in particular the discovery of specific sites in N_{TAIL} that when mutated hampers, or *a contrario* promotes

the interaction, could be exploited for the design of peptide-inspired inhibitors to be used in future antiviral approaches.

These studies could also be extended to the N_{TAIL}-XD interaction of other members of the *Paramyxoviridae* family, including the Nipah and Hendra viruses that are severe human pathogens causing respiratory and central nervous system diseases. This will allow assessing whether the role of disorder in N_{TAIL} is conserved or not in the three viruses.

BIBLIOGRAPHY

Adams, V. H., McBryant, S. J., Wade, P. A., Woodcock, C. L., & Hansen, J. C. 2007. Intrinsic disorder and autonomous domain function in the multifunctional nuclear protein, MeCP2. ***Journal of Biological Chemistry***, 282(20): 15057–15064.

Billeter, M., Qian, Y. Q., Otting, G., Muller, M., Gehring, W. & Wuthrich, K. 1993. Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain–DNA complex. ***Journal of Molecular Biology,*** 234: 1084–1093.

Bonetti, D., Camilloni, C., Visconti, L., Longhi, S., Brunori, M., Vendruscolo, M., & Gianni, S. 2016. Identification and Structural Characterization of an Intermediate in the Folding of the Measles Virus X Domain. ***The Journal of Biological Chemistry,*** 291(20): 10886-10892.

Bonetti, D., Troilo, F., Toto, A., Brunori, M., Longhi, S., & Gianni, S. 2017. Analyzing the folding and binding steps of an intrinsically disordered protein by protein engineering. ***Biochemistry,*** 56(29): 3780-3786.

Borcherds, W., Becker, A., Chen, L., Chen, J., & Daughdrill, G. W. 2017. Biophysical Letter Optimal Affinity Enhancement by a Conserved Flexible Linker Controls p53 Mimicry in MdmX. ***Biophysical Journal***, (10): 2038–2042.

Carsillo, T., Traylor, Z., Choi, C., Niewiesk, S., & Oglesbee, M. 2006. hsp72, a Host Determinant of Measles Virus Neurovirulence. ***Journal of Virology***, 80(22): 11031–11039.

Clerici, M., Gutsche, I., Gehring, N. H., Hentze, M. W., Kulozik, A., et al. 2009. Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2. ***The EMBO Journal***, 28(15): 2293–2306.

Colombo, M., Bourhis, J., Chamontin, C., Soriano, C., Villet, S., et al. 2009.

The interaction between the measles virus nucleoprotein and the Interferon Regulator Factor 3 relies on a specific cellular environment. ***Virology Journal***, 17: 1–17.

Couturier, M., Buccellato, M., Costanzo, S., Bourhis, J. M., Shu, Y., et al. 2010. High affinity binding between Hsp70 and the C-termonal domain of the measles virus nucleoprotein requires an Hsp40 co-chaperone. ***Journal of Molecular Recognition,*** 23: 300-315.

Daughdrill, G. W., Narayanaswami, P., Gilmore, S. H., Belczyk, A., & Brown, C. J. 2007. Dynamic behavior of an intrinsically unstructured linker domain is conserved in the face of negligible amino acid sequence conservation. ***Journal of Molecular Evolution***, 65(3): 277–288.

Dill K.A. & Chan H.S. 1997. From Levinthal to pathways to funnels. ***Nature Structural Biology***, 4: 10-18.

Dill K.A., & MacCallum J.L. 2012. The protein-folding problem, 50 years on. ***Science***, 338: 1042-1046.

Dill, K. A., Ozkan, S. B., Shell, M. S., & Weikl, T. R. 2008. The Protein Folding Problem. ***Annual Review of Biophysics***, 37: 289–316.

Dosnon, M., Bonetti, D., Morrone, A., Erales, J., Di Silvio, E., et al. 2015. Demonstration of a folding after binding mechanism in the recognition between the measles virus NTAIL and X domains. ***ACS Chemical Biology***, 10(3): 795–802.

Dunker, A. K., Babu, M. M., Barbar, E., Blackledge, M., Bondos, S. E., et al. 2013. What's in a name? Why these proteins are intrinsically disordered. ***Intrinsically Disordered Proteins,*** 1(1): e24157.

Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., et al. 2001. Intrinsically disordered protein. ***Journal of Molecular Graphics and Modelling***, 19(1): 26–59.

Dyson, H. J. 2011. Expanding the proteome: disordered and alternatively-

folded proteins H. ***Quarterly Reviews of Biophysics***, 44(4): 467–518.

Dyson, H. J., & Wright, P. E. 2002. Coupling of folding and binding for unstructured proteins. ***Current Opinion in Structural Biology***, 12(1): 54–60.

Dyson, H. J., & Wright, P. E. 2005. Intrinsically unstructured proteins and their functions. ***Nature Reviews Molecular Cell Biology***, 6(3): 197–208.

Ellerby, L.M., Escobar W.A., Fink A.L., Mitchinson C., and Wells J.A. 1990. The role of lysine-234 in beta-lactamase catalysis probed by site-directed mutagenesis. ***Biochemistry***, 29: 5797-5806.

Eyring, H. 1935. The activated complex in chemical reactions. ***The Journal of Chemical Physics,*** 3: 107-15

Ferreiro, D. U., Hegler, J. A., Komives, E. A., & Wolynes, P. G. 2007. Localizing frustration in native proteins and protein assemblies. ***Proceedings of the National Academy of Sciences***, 104(50): 19819–19824.

Ferreiro, D. U.,. Hegler, J. A., Komives, E. A., & Wolynes, P. G. 2010. On the role of frustration in the energy landscapes of allosteric proteins. ***Proceedings of the National Academy of Sciences***, 108: 3499-3503.

Fersht, A. R. 1995. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. ***Proceedings of the National Academy of Sciences***, 92(24): 10869–10873.

Fersht, A. R., Leatherbarrow, R. J., & Wells, T. N. 1986. Quantitative analysis of structure-activity relationship in engineered proteins by linear free-energy relationship. ***Nature,*** 322: 284-286.

Fersht, A. R., Leatherbarrow, R. J., & Wells, T. N. 1987. Structure-activity relationships in engineered proteins: analysis of use of binding energy by linear free energy relationships. ***Biochemistry,*** 26(19): 6030-8.

Fersht, A.R., Matouchek A., & Serrano L. 1992. The folding of an enzyme. I.

Theory of protein engineering analysis of stability and pathway of protein folding. ***Journal of Molecular Biology***, 224:771-82.

Fersht, A. R., & Sato, S. 2004. -Value analysis and the nature of protein-folding transition states. ***Proceedings of the National Academy of Sciences***, 101(21): 7976–7981.

Fu, C., Wehr, D. R., Edwards, J., & Hauge, B. 2008. Rapid one-step recombinational cloning. ***Nucleic Acids Research***, 36(9): 1–7.

Ghosh, R. P., Nikitina, T., Horowitz-scherer, R. A., Gierasch, L. M., Uversky, V. N., et al. 2010. Unique physical properties and interactions of the domains of methylated DNA binding protein 2 (MeCP2). ***Biochemistry***, 49(20): 4395–4410.

Gianni, S., Dogan, J., & Jemth, P. 2014a. Distinguishing induced fit from conformational selection. ***Biophysical Chemistry***, 189: 33–39.

Gianni, S., Dogan, J., & Jemth, P. 2014b. Deciphering the mechanisms of binding induced folding at nearly atomic resolution: The Φ value analysis applied to IDPs. ***Intrinsically Disordered Proteins***, 2(1): e970900.

Gruet, A., Dosnon, M., Blocquel, D., Brunel, J., Gerlier, D., et al. 2016. Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. ***FEBS Journal***, 283(4): 576–594.

Gutsche, I., Desfosses, A., Effantin, G., & Ling, W. L. 2015. Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. ***Structural Virology***, 348(6235): 704–708.

Habchi, J., Tompa, P., Longhi, S., & Uversky, V. N. 2014. Introducing protein intrinsic disorder. ***Chemical Reviews***, 114(13): 6561–6588.

Hammes, G. G., Chang, Y.-C., & Oas, T. G. 2009. Conformational selection or induced fit: A flux description of reaction mechanism. ***Proceedings of the National Academy of Sciences***, 106(33): 13737–13741.

Jasanoff, A., & Fersht, A. R. 1994. Quantitative determination of helical propensities from trifluoroethanol titration curves. ***Biochemistry,*** 33: 2129-2135.

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

Karlin, D., Longhi, S., & Canard, B. 2002a. The N-Terminal Domain of the Phosphoprotein of Morbilliviruses Belongs to the Natively Unfolded Class of Proteins. ***Virology***, 262: 251–262.

Karlin, D., Longhi, S., & Canard, B. 2002b. Substitution of Two Residues in the Measles Virus Nucleoprotein Results in an Impaired Self-Association. ***Virology***, 432: 420–432.

Kiefhaber, T., Bachmann, A., & Jensen, K. S. 2012. Dynamics and mechanisms of coupled protein folding and binding reactions. ***Current Opinion in Structural Biology***, 22(1): 21–29.

Lamiable, A., Thévenet, P., Rey, J., Vavrusa, M., Derremaux, P., & Tufféry, P. 2016. PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. ***Nucleic Acids Researh,*** 44(W1): W449459.

Lee, G. M., Pufall, M. A., Meeker, C. A., Kang, H., Graves, B. J., et al. 2008. The Affinity of Ets-1 for DNA is Modulated by Phosphorylation Through Transient Interactions of an Unstructured Region. ***Journal of Molecular Biology***, 1: 1014–1030.

Leffler, J. E. 1953. Parameters for the description of transition states. ***Science,*** 117(3039): 340-1

Longhi, S. 2012. The measles virus N(TAIL)-XD complex: an illustrative example of fuzziness. **Advances in Experimental Medicine Biology,** 725: 126-141.

Longhi, S., Bloyet, L. M., Gianni, S., & Gerlier, D. 2017. How disorder and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication. ***Cellular and Molecular Life Science,*** 74(17): 3091-3118.

Longhi, S., Receveur-Brechot, V., Karlin, D., Johansson, K., Darbon, H., et al. 2003. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moyety of the phosphoprotein. ***The Journal of Biological Chemistry,*** 278:18638-18648.

Matouschek, A., Kellis, J. T. Jr, Serrano, L., & Fersht, A. R. 1989. Mapping the transition state and pathway of protein folding by protein engineering. ***Nature.*** 340(6229): 122-6.

Meiering E.M., Serrano L., & Fersht A.R. 1992. Effect of active site residues in barnase on activity and stability. ***Journal of Molecular Biology***, 225: 585-589.

Myers, J. K., Pace, C. N., & Scholtz, J. M. 1998. Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease TI. ***Protein Science***, 383–388.

Monod, J., Wyman, J., & Changeux J. P. 1965. On the nature of allosteric transitions: a plausible model. ***Journal of Molecular Biology,*** 12: 88118

Oliveberg, M., & Wolynes, P. G. 2005. The experimental survey of protein-folding energy landscapes. ***Quarterly Reviews of Biophysics***, 3: 245–288.

Onuchic, N., & Wolynes, P. G. 2004. Theory of protein folding. ***Current Opinion in Structural Biology***, 70–75.

Papoian, G. A., Ulander, J., & Wolynes, P. G. 2003. Role of Water Mediated Interactions in Protein - Protein Recognition Landscapes. ***JACS***, (5): 9170–9178.

Pometun, M. S., Chekmenev, E. Y., & Wittebort, R. J. 2004. Quantitative Observation of Backbone Disorder in Native Elastin. ***Journal of Biological Chemistry***, 279(9): 7982–7987.

Pufall, M. A., Lee, G. M., Nelson, M. L., Kang, H. S., Velyvis, A., et al. 2005. Biochemistry: Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. ***Science***, 309(5731): 142–145.

Renault, L., Bugyi, B. & Carlier, M.F. 2008. Spire and Cordon-bleu: multifunctional regulators of actin dynamics. ***Trends in Cell Biology,*** 18: 494–504.

Sato, H., Masuda, M., Miura, R., Yoneda, M., & Kai, C. 2006. Morbillivirus nucleoprotein possesses a novel nuclear localization signal and a CRM1-independent nuclear export signal. ***Virology***, 352: 121–130.

Stott, K., Watson, M., Howe, F. S., Grossmann, J. G., & Thomas, J. O. 2010. Tail-mediated collapse of HMGB1 is dynamic and occurs via differential binding of the acidic tail to the A and B domains. ***Journal of Molecular Biology***, 403(5): 706–722.

Tokuriki, N., Oldfield, C. J., Uversky, V. N., Berezovsky, I. N., & Tawfik, D. S. 2009. Do viral proteins possess unique biophysical features ? ***Cell***,: 53–59.

Tompa, P. 2011. Unstructural biology coming of age. ***Current Opinion in Structural Biology***, 21(3): 419–425.

Tompa, P. 2012. Intrinsically disordered proteins : a 10-year recap. ***Trends in Biochemical Sciences***, 37(12): 509–516.

Toto, A., Camilloni, C., Giri, R., Brunori, M., Vendruscolo, M., et al. 2016. Molecular Recognition by Templated Folding of an Intrinsically Disordered Protein. ***Scientific Reports***, 6: 1–9.

Uversky, V. N. 2013. A decade and a half of protein intrinsic disorder: Biology still waits for physics. ***Protein Science***, 22(6): 693–724.

Uversky, V. N., & Dunker, A. K. 2010. Understanding protein non-folding. ***Biochimica et Biophysica Acta - Proteins and Proteomics***, 1804(6): 1231–1264.

Uversky, V. N., & Longhi, S. 2011. Flexible Viruses: Structural Disorder in Viral Proteins.

Vise, P. D., Baral, B., Latos, A. J., & Daughdrill, G. W. 2005. NMR chemical shift and relaxation measurements provide evidence for the coupled folding and binding of the p53 transactivation domain. ***Nucleic Acid Research***, 33(7): 2061–2077.

Wang, J., & Verkhivker, G. M. 2003. Energy Landscape Theory , Funnels , Specificity , and Optimal Criterion of Biomolecular Binding. ***Physical Review Letters***, : 1–4.

Watanabe, A., Yoneda, M., Ikeda, F., Sugai, A., Sato, H., et al. 2011. Peroxiredoxin 1 Is Required for Efficient Transcription and Replication of Measles Virus. ***Journal of Virology***, 85(5): 2247–2253.

Wilson, C. G. M., & Magliery, T. J. 2006. Reassembled GFP: detecting protein-protein interactions and protein expression patterns. ***Methods of Biochemical Analysis,*** 47: 391-405.

Wilson, C. G. M., Magliery, T. J., & Regan, L. 2004. Detecting protein-protein interactions with GFP-fragment reassembly. ***Nature Methods,*** 1: 255-262.

Wright, P. E., & Dyson, H. J. 1999. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. ***Journal of Molecular Biology***, 293(2): 321–331.

Xue, B., Williams, R. W.,Oldfield, C. J., Goh, G. K., Dunker AK, & Uversky, V. N. 2010. Viral disorder or disordered viruses: do viral proteins possess unique features? ***Protein and Peptide Letters,*** 17(8): 932-51.

Zhang, X., Bourhis, J., Longhi, S., Carsillo, T., Buccellato, M., et al. 2005. Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus. ***Virology***, 337: 162–174.

ATTACHMENTS

During my PhD I also had the opportunity to apply the kinetics techniques to study other biological systems, as detailed below.

Paper 1: The KIX domain is an 89-residues globular domain presenting two distinct binding sites and involved in mediating protein-protein interactions. The presence of two distinct binding sites in such a small domain makes it a suitable candidate to investigate the effect of the potentially divergent demands between folding and function. In the following paper we reported an extensive mutational analysis of the folding pathway of the KIX domain, based on 30 site-directed variants that have been characterized through kinetic and equilibrium denaturation experiments. This approach allowed us to assess the structures of both the transition and denatured states *via* the Φ-value analysis. Data revealed that the transition state presents mostly native-like interactions, while the denatured state is somewhat misfolded. We also performed double mutant cycles: with a second step of mutagenesis, 15 double variants were obtained allowing mapping some of the non-native contacts in the denatured state. Interestingly, such a misfolding arises from non-native interactions involving the residues critical for the function of the protein. The results described in this work appear to highlight the diverging demands between folding and function that may lead to misfolding, which may be observed in the early stages of folding.

ACS chemical biology

Articles

pubs.acs.org/acschemicalbiology

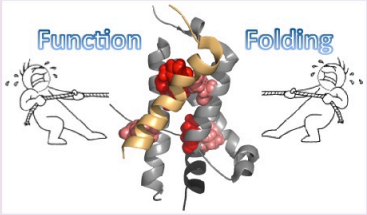
The Folding Pathway of the KIX Domain

Francesca Troilo,[†,‡] Daniela Bonetti,[†] Angelo Toto,[†] Lorenzo Visconti,[†] Maurizio Brunori,[†] Sonia Longhi,[‡] and Stefano Gianni*[,†]

[†]Istituto Pasteur - Fondazione Cenci Bolognetti, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" and Istituto di Biologia e Patologia Molecolari del CNR, Sapienza Università di Roma, 00185, Rome, Italy

[‡]Aix-Marseille Univ, CNRS, Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, 13288, Marseille, France

ABSTRACT: The KIX domain is an 89-residues globular domain with an important role in mediating protein−protein interactions. The presence of two distinct binding sites in such a small domain makes KIX a suitable candidate to investigate the effect of the potentially divergent demands between folding and function. Here, we report an extensive mutational analysis of the folding pathway of the KIX domain, based on 30 site-directed mutants, which allow us to assess the structures of both the transition and denatured states. Data reveal that, while the transition state presents mostly native-like interactions, the denatured state is somewhat misfolded. We mapped some of the non-native contacts in the denatured state using a second round of mutagenesis, based on double mutant cycles on 15 double mutants. Interestingly, such a misfolding arises from non-native interactions involving the residues critical for the function of the protein. The results described in this work appear to highlight the diverging demands between folding and function that may lead to misfolding, which may be observed in the early stages of folding.



A fundamental goal of protein folding studies is to describe the mechanism whereby a denatured protein achieves its native conformation. However, a major complexity arises from the co-operativity of the folding process—despite hundreds of weak noncovalent interactions between the residues and the solvent are being formed and broken, the native and denatured states are the only species significantly populated.[1−3] Thus, although productive folding intermediates[4−9] or off-pathway kinetic traps[10,11] may be sometimes experimentally detected, the folding reaction of small globular proteins conforms very frequently to a simple two-state reaction.

Because of the extraordinary high degree of co-operativity intrinsic to a two-state folding mechanism, structural information about the pathway may only be inferred indirectly. In fact, systematic perturbation of the structure of the protein, by conservative site-directed mutations, allows mapping the interaction pattern present in both the transition and denatured states.[12,13] This experimental characterization is critical to identifying the formation of one or more folding nuclei, stabilizing the transition state, and detecting the residual structure in the denatured state, which under physiological conditions appears to play a major role in sculpting both the folding pathway and the native topology.[14−19]

The KIX domain is a globular domain that is a part of a large coactivator protein called CBP, whose three-dimensional structure consists of three α-helices and two short 3_{10}-helices.[20,21] From a physiological perspective, KIX is capable of recognizing and binding, *via* two different binding sites located at opposite sides of the domain, several intrinsically unstructured proteins, which all fold upon binding.[22−27] One of the binding sites is called the "c-Myb site" after the transactivation domain of the oncoprotein c-Myb, whereas the other one is called the "MLL site" after the Mixed Lineage Leukemia protein, these two proteins representing the characteristic ligands of KIX. Previous site-directed mutagenesis experiments have shown that the two binding sites appear to show some functional redundancy in the activation of transcription.[28] The folding of the KIX domain was previously analyzed by equilibrium and kinetic experiments.[29−32] While earlier work suggested the presence of a reaction intermediate, based on a curvature in the dependence of the logarithm of the folding rate constant from denaturant concentration,[29] recent kinetic reanalysis of the folding of this globular domain demonstrated that this curvature was an artifact arising from time limitation of the stopped-flow apparatus.[30] Thus, we concluded that folding of KIX is consistent with a two-state scenario.

Here, we describe the experimental characterization of the folding pathway of the KIX domain, as probed by the equilibrium and kinetic analysis of 30 site directed variants, which allowed us to infer the structure of both the transition and denatured states. Data reveal that, while the folding

Received: April 4, 2017
Accepted: May 1, 2017
Published: May 1, 2017

DOI: 10.1021/acschembio.7b00289
ACS Chem. Biol. 2017, 12, 1683−1690

Table 1. Folding Parameters for KIX and Its Site-Directed Variants Measured from Equilibrium and Kinetic Experiments[a]

	k_F (s^{-1})	k_U (s^{-1})	m_F (kcal/M mol)	m_U (kcal/M mol)	m_{D-N} (kcal/M mol)	$\Delta\Delta G_{D-N}$(kcal/mol)	ϕ
WT	2400 ± 300	14.1 ± 0.9	0.76 ± 0.04	0.20 ± 0.06	0.96 ± 0.04		
T11S	2100 ± 200	21 ± 2	0.76 ± 0.08	0.2 ± 0.1	0.92 ± 0.09	0.29 ± 0.11	[b]
L14A	1700 ± 200	11 ± 1	0.73 ± 0.07	0.23 ± 0.09	0.96 ± 0.05	0.05 ± 0.11	[b]
L18A	1700 ± 300	3.0 ± 0.3	0.86 ± 0.03	0.20 ± 0.04	1.06 ± 0.03	−0.71 ± 0.13	−0.26 ± 0.17
V23A	720 ± 70	35 ± 3	0.51 ± 0.05	0.17 ± 0.09	0.68 ± 0.07	1.21 ± 0.11	0.57 ± 0.09
A25G	740 ± 70	84 ± 8	0.46 ± 0.05	0.2 ± 0.1	0.60 ± 0.09	1.72 ± 0.11	0.39 ± 0.06
I26V	2400 ± 200	23 ± 2	0.93 ± 0.02	0.18 ± 0.02	1.1 ± 0.1	0.29 ± 0.11	[b]
T29S	1400 ± 100	11 ± 1	0.65 ± 0.06	0.2 ± 0.1	0.86 ± 0.08	0.20 ± 0.11	[b]
A33G	3100 ± 500	21 ± 3	0.84 ± 0.04	0.15 ± 0.06	0.99 ± 0.04	0.07 ± 0.14	[b]
A34G	3000 ± 300	25 ± 2	0.82 ± 0.08	0.2 ± 0.1	0.99 ± 0.06	0.19 ± 0.11	[b]
L35A	2600 ± 300	25 ± 2	0.82 ± 0.08	0.2 ± 0.1	0.99 ± 0.06	0.29 ± 0.11	[b]
L43A	2100 ± 300	1.3 ± 0.2	0.90 ± 0.03	0.26 ± 0.04	1.16 ± 0.04	−1.31 ± 0.14	−0.05 ± 0.08
V44A	630 ± 70	70 ± 8	0.62 ± 0.07	0.14 ± 0.09	0.76 ± 0.06	1.70 ± 0.12	0.45 ± 0.06
A45G	1300 ± 100	27 ± 2	0.82 ± 0.03	0.17 ± 0.04	0.99 ± 0.03	0.73 ± 0.10	0.47 ± 0.13
Y46A	1800 ± 400	63 ± 9	0.76 ± 0.09	0.2 ± 0.1	0.9 ± 0.1	1.03 ± 0.17	0.15 ± 0.14
V50A	1900 ± 200	35 ± 3	0.62 ± 0.06	0.18 ± 0.08	0.80 ± 0.05	0.65 ± 0.11	0.20 ± 0.14
Y65A	1800 ± 200	26 ± 3	1.0 ± 0.2	0.2 ± 0.3	1.2 ± 0.2	0.60 ± 0.12	0.43 ± 0.17
L67A	1600 ± 300	30 ± 5	0.9 ± 0.1	0.2 ± 0.2	1.1 ± 0.1	0.50 ± 0.18	−0.03 ± 0.29
L68A	260 ± 30	7 ± 1	0.6 ± 0.1	0.26 ± 0.04	0.9 ± 0.1	0.90 ± 0.12	1.43 ± 0.22
A69G	1200 ± 100	13 ± 1	0.79 ± 0.08	0.22 ± 0.09	1.01 ± 0.04	0.35 ± 0.11	[b]
I72V	2400 ± 400	24 ± 3	0.85 ± 0.06	0.18 ± 0.08	1.03 ± 0.06	0.29 ± 0.15	[b]
Y73A	6000 ± 1000	15 ± 3	0.89 ± 0.07	0.22 ± 0.09	1.11 ± 0.05	−0.53 ± 0.17	1.00 ± 0.41
I75V	2300 ± 400	24 ± 4	0.74 ± 0.05	0.15 ± 0.06	0.89 ± 0.04	0.32 ± 0.16	[b]
L79A	2100 ± 200	8 ± 1	0.81 ± 0.03	0.21 ± 0.04	1.02 ± 0.02	−0.24 ± 0.13	[b]

[a]The mutants V10A, V19A, L22A, A47G, Y55A, A58G, and Y64A expressed poorly and could not be characterized. [b]These mutants show $\Delta\Delta G_{D-N}$ < 0.40 kcal/mol, preventing reliable calculation of the Φ values.

transition state is mainly native-like, the denatured state of the protein appears to contain non-native interactions. Furthermore, we observe that some of the mutations located in the functional sites of KIX are associated with a stabilization of the domain, a result that appears to highlight the divergent requirements of folding and function.

RESULTS

To map the structural features of the folding pathway of the KIX domain, we performed a Φ value analysis. By following this methodology,[12,13] the polypeptide chain is systematically mutated and the role of individual residues in the transition state is quantified by their respective Φ value. Quantitatively, the Φ value is expressed by normalizing the effect of the mutation on the activation free energy divided by that of the stability of the native state. Therefore, we produced 30 site directed variants of KIX, listed in Table 1, out of which seven did not express to a significant extent. The variants were designed by following the guidelines typically adopted in Φ value analysis. In particular, a conservative truncation of the side chain was introduced in each hydrophobic position of the domain, a type of mutation that, as extensively discussed previously,[12,13] represents the easiest to be interpreted, in this kind of analysis. All the purified variants were subjected to equilibrium and kinetic folding experiments.

Urea-induced equilibrium denaturation of KIX and its site directed variants was monitored by intrinsic Trp emission at 20 °C and pH 7.2 in 50 mM sodium phosphate buffer in the presence of 150 mM NaCl. In all cases, the measured transitions followed a two-state behavior (Figure 1). The thermodynamic stability and apparent m_{D-N} value for each mutant derived from two-state analysis are described below (Table 1).

Unfolding and refolding kinetics of all the proteins were measured by stopped-flow kinetics, by rapidly diluting the protein into solutions containing different urea concentrations. In all experiments, the fluorescence time course was satisfactorily fitted to a single-exponential decay. The measured chevron plots for wild-type KIX and all the variants are shown in Figure 1. In analogy to wild type KIX,[30] all site directed mutants displayed a V-shaped chevron plot, and none of them displayed any deviation from linearity in the folding and unfolding limbs. This finding reinforces our previous proposal for a two-state mechanism of this globular domain and the absence of detectable transient intermediates, a model appearing in this case also rather robust with respect to site-directed mutagenesis.

Because of its small size (89 residues),[33] KIX displays a low cooperativity, the m_{D-N} obtained from equilibrium urea-induced denaturation being 0.95 ± 0.08 kcal mol^{-1} M^{-1}; this complicates an accurate determination of the folding parameters, from equilibrium experiments alone. Thus, the equilibrium and kinetic experiments of each variant were fitted globally to the following equations:

$$\text{Equilibrium: } Y_{obs} = Y_N + Y_D \frac{e^{(m_{D-N}([urea]-[urea]_{1/2})}}{1 + e^{(m_{D-N}([urea]-[urea]_{1/2})}}$$

$$\text{Kinetics: } k_{obs} = k_F\, e^{(-m_F[urea])} + k_U\, e^{(m_U[urea])}$$

In each case, the m_{D-N} value (equivalent to the sum between m_F and m_U) was shared between the equilibrium and kinetic data sets, in agreement with the two-state formalism.[1] This fitting procedure allowed (i) testing of the robustness of the two-state mechanism for each individual mutant and (ii) decreasing of the fitting error on m_{D-N} arising from the low cooperativity observed in the equilibrium transition.

DOI: 10.1021/acschembio.7b00289
ACS Chem. Biol. 2017, 12, 1683−1690

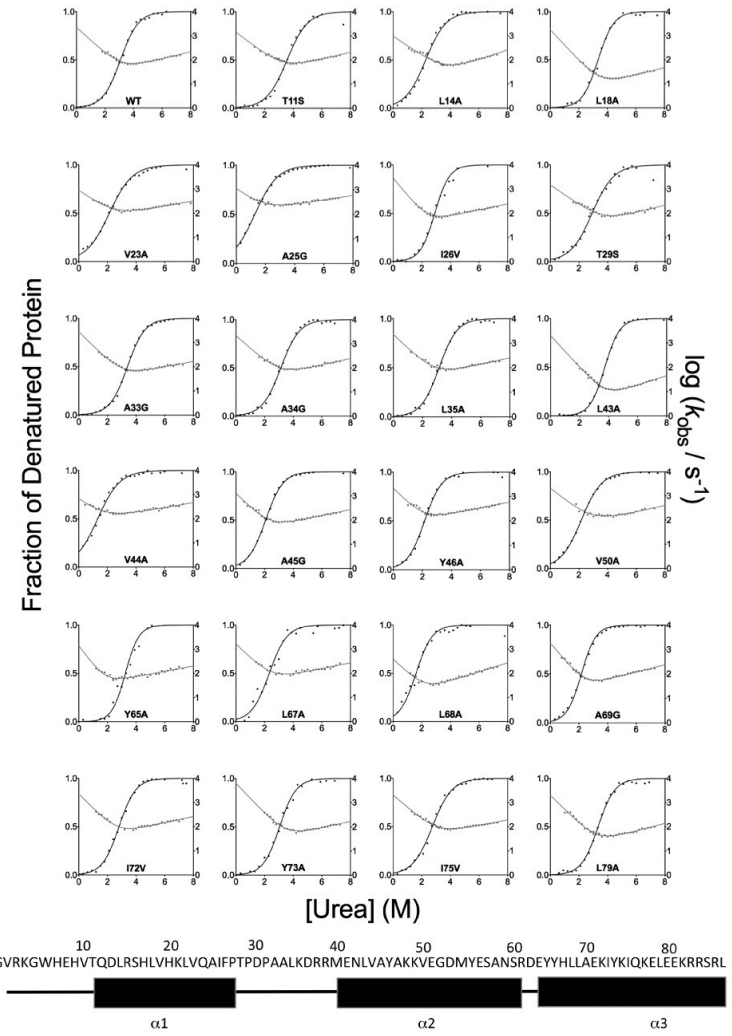


Figure 1. Equilibrium denaturations and chevron plots of KIX and its site directed mutants. All the experiments were carried out at 20 °C and pH 7.2 in 50 mM sodium phosphate buffer in the presence of 150 mM NaCl. As described in the Results section, each mutant was globally fitted to a two state mechanism by assuming the m_{D-N} value at equilibrium to be equivalent to the sum between the kinetic m_F and m_U values.[7]

DOI: 10.1021/acschembio.7b00289
ACS Chem. Biol. 2017, 12, 1683−1690

To analyze the structural distribution of Φ values, following Fersht and co-workers,[34,35] they were divided into three groups (Figure 2A): small (Φ < 0.3; red), intermediate (0.3 < Φ < 0.7; magenta), and large (Φ > 0.7; blue). The structural distribution of the Φ values indicates that the transition state has a rather diffuse native-like structure with a folding nucleus approximately located by and large in the C-terminal helix. To address the overall structural features of the transition state, we performed a linear free energy relationship analysis (LFER).[36] By following this method, the changes in free energy of the transition state are related to changes in equilibrium free energies, with the slope of the observed correlation (called α value), reflecting the degree of similarity between the native and transition states. As depicted in Figure 2B, in the case of KIX, the LFER yields a linear correlation with a slope of 0.67. This is consistent with what is observed in other globular proteins and indicates that KIX folds *via* the so-called nucleation−condensation mechanism.[37,38] This description implies that the transition state resembles a distorted version of the native state, and the whole protein self-assembles around a weakly formed nucleus.

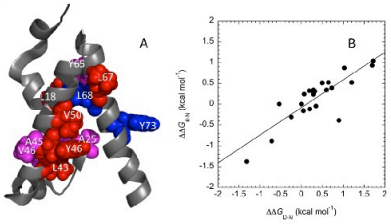


Figure 2. Folding transition state of the KIX domain. Panel A: Following a generally accepted convention,[34,35] the experimentally determined Φ values were divided into three different groups and mapped onto the native structure of KIX: small values (Φ < 0.3; red), intermediate values (0.3 < Φ < 0.7; magenta), and large values (Φ > 0.7; blue). Panel B: LFER analysis of the transition state of KIX. In agreement with the nucleation condensation mechanism,[37] we observed a linear LFER plot with an apparent α value of 0.67 ± 0.09.

A reliable test to investigate the overall structural features of folding transition and denatured states is to compare the data measured under various experimental conditions or on different site-directed mutants.[39,40] In fact, the slopes of the unfolding and refolding arms of the chevron plot (yielding the m_U and m_F values, respectively) are correlated to the changes in accessible surface area between two states; an analysis of their dependence on reaction conditions may be of diagnostic value to detect shifts of the transition and denatured states along the reaction coordinate. The correlation between the m_{D-N}, m_U, and m_F and the change in stability upon mutation is reported in Figure 3. Interestingly, it is clear that, while m_U is essentially insensitive to the thermodynamic stability of KIX, both the m_F and m_{D-N} display a similar linear decrease with decreasing protein stability. Since m_U does not change with overall thermodynamic stability, these observations indicate that the changes in the folding m values are due to movements of the denatured state toward the native state (a compaction) and not to movements of the transition state. Thus, the experiments detect a residual malleable structure in the denatured state of KIX, which is perturbed by mutagenesis. Interestingly, this compaction of the denatured state upon mutation-induced thermodynamic destabilization indicates that, contrary to what was previously observed in other proteins, the residual structure is largely due to non-native interactions.

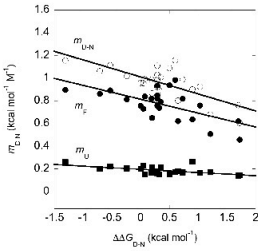


Figure 3. Dependence of m_F (closed circles), m_U (squares), and m_{D-N} (open circles) on protein stability. Lines are the best fit to a linear function. As discussed in the text, both m_F and m_{D-N} increase in parallel with increasing protein stability, strongly indicating that KIX displays residual malleable structure in its denatured state, which is primarily stabilized by non-native interactions.

Investigating Non-native Interactions in the Denatured State of KIX by Double Mutant Cycles. The changes in free energy of folding introduced by site directed mutagenesis depend on the effect of a given mutation on both the denatured and native states.[12] In particular, the experimentally measured $\Delta\Delta G_{D-N}$ equals $\Delta\Delta G_{N-N'} - \Delta\Delta G_{D-D'}$, where N′ and D′ are the native and denatured states of the variant. Because the free energy contribution arising from the interaction between residues is generally negligible in the denatured state, the observed $\Delta\Delta G_{D-N}$ resembles $\Delta\Delta G_{N-N'}$; hence, if the denatured state does not display any detectable residual structure, the measured change in protein stability caused by a mutation is mostly a measure of the perturbation induced in the native state.

Analogously, by considering a cubic thermodynamic scheme, the sum of the changes in stability introduced by two single mutations can be compared with change induced by the corresponding double mutant to assess a coupling free energy $\Delta\Delta\Delta G_{D-N} = \Delta\Delta\Delta G_{N-N'} - \Delta\Delta\Delta G_{D-D'}$, where $\Delta\Delta\Delta G_{N-N'}$ and $\Delta\Delta\Delta G_{D-D'}$ represent a measure of the strength with which the two residues interact in the native and denatured states, respectively.[40] A $\Delta\Delta\Delta G_{D-N}$ tending to 0 would be observed when the effect of the double mutant is similar to the sum of the two single mutants, whereas a $\Delta\Delta\Delta G_{D-N} \neq 0$ implies that the probed positions interact energetically.

To provide additional evidence for the presence of non-native interactions in the denatured state of KIX, we resorted to performing double mutant cycles. In particular, we focused on those positions displaying an increase in m_{D-N} value of more than 0.1 kcal mol^{-1} M^{-1} (*i.e.*, the mutations associated with a detectable increase of the accessible surface area of the denatured state). Six mutations were therefore taken into account (L18A, I26V, L43A, Y65A, L67A and Y73A), and the 15 corresponding double mutants were produced. Of these, 12 were stable enough to be expressed, purified, and characterized. In analogy to what was described above for the single mutants,

DOI: 10.1021/acschembio.7b00289
ACS Chem. Biol. 2017, 12, 1683−1690

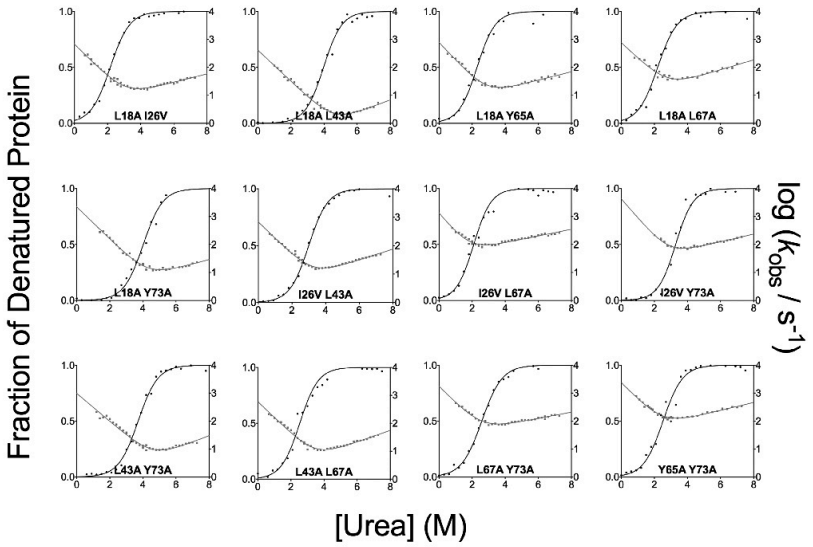


Figure 4. Double mutant cycles in the KIX domain. All the experiments were carried out at 20 °C and pH 7.2 in 50 mM sodium phosphate buffer in the presence of 150 mM NaCl. As described in the Results section, each mutant was globally fitted to a two state mechanism by assuming the m_{D-N} value at equilibrium to be equivalent to the sum between the kinetic m_F and m_U values.[7]

Table 2. Folding Parameters for the Double Mutant Cycles in the KIX Domain Measured from Equilibrium and Kinetic Experiments[a]

mutants	k_F (s^{-1})	k_U (s^{-1})	$\Delta\Delta G_{D-N}$ (kcal/mol)	$\Delta\Delta G_{D-TS}$ (kcal/mol)	$\Delta\Delta\Delta G_{D-N}$ (kcal/mol)	$\Delta\Delta\Delta G_{D-TS}$ (kcal/mol)
WT	2400 ± 300	14.1 ± 0.9				
L18A I26V	670 ± 70	3.6 ± 0.6	−0.07 ± 0.13	0.73 ± 0.09	−0.35 ± 0.22	−0.55 ± 0.17
L18A L43A	380 ± 40	0.26 ± 0.03	−1.3 ± 0.12	1.06 ± 0.09	−0.76 ± 0.23	−0.81 ± 0.34
L18A Y65A	830 ± 90	4.7 ± 0.6	−0.04 ± 0.12	0.6 ± 0.09	−0.17 ± 0.22	−0.21 ± 0.17
L18A L67A	930 ± 90	8.3 ± 0.8	0.23 ± 0.11	0.54 ± 0.09	−0.28 ± 0.25	−0.14 ± 0.21
L18A Y73A	2300 ± 400	1.7 ± 0.5	−1.20 ± 0.20	0.02 ± 0.11	−0.04 ± 0.30	−0.40 ± 0.21
I26V L43A	650 ± 60	2.3 ± 0.2	−0.29 ± 0.11	0.75 ± 0.09	0.73 ± 0.21	0.68 ± 0.16
I26V L67A	1300 ± 300	38 ± 5.0	0.93 ± 0.17	0.36 ± 0.15	0.02 ± 0.28	−0.14 ± 0.23
I26V Y73A	4300 ± 400	19.0 ± 3.0	−0.16 ± 0.12	−0.35 ± 0.09	−0.09 ± 0.24	−0.22 ± 0.18
L43A L67A	660 ± 100	1.4 ± 0.3	−0.6 ± 0.18	0.74 ± 0.11	−0.04 ± 0.29	−0.45 ± 0.21
L43A Y73A	980 ± 100	1.1 ± 0.1	−0.97 ± 0.11	0.51 ± 0.09	−0.88 ± 0.25	−1.01 ± 0.19
Y65A Y73A	1900 ± 600	31 ± 10	0.57 ± 0.29	0.11 ± 0.19	−0.61 ± 0.36	−0.53 ± 0.25
L67A Y73A	1700 ± 200	26.0 ± 3	0.54 ± 0.11	0.19 ± 0.09	−0.42 ± 0.28	−0.54 ± 0.21

[a]The mutants I26V Y65A, L43A Y65A, and Y65A L67A expressed poorly and could not be characterized.

the 12 double mutants were subjected to equilibrium and kinetic experiments (Figure 4), and the calculated folding parameters are reported in Table 2, together with their associated $\Delta\Delta\Delta G_{D-TS}$ and $\Delta\Delta\Delta G_{D-N}$. These values were therefore obtained calculating the difference in the effect observed for the double mutant minus the individual contribution of the each of the two single variants.

Inspection of Table 2 reveals that, in six cases, there was a detectable coupling free energy $\Delta\Delta\Delta G_{D-N}$ between couples of residues. Remarkably, with the exception of positions I26 and L43 that are in direct contact in the native state of KIX, all the other pairs of residues displaying a significant $\Delta\Delta\Delta G_{D-N}$ are located in different regions of the domain, suggesting the value of $\Delta\Delta\Delta G_{N-N'} \approx 0$. In those cases, the $\Delta\Delta\Delta G_{D-N}$ would tend to a value of $\Delta\Delta\Delta G_{D-D'}$ and would be therefore diagnostic of a non-native contact in the denatured state. In support of this view, it is of interest to comment on the magnitude of coupling free energies in the folding transition state, which appear to

DOI: 10.1021/acschembio.7b00289
ACS Chem. Biol. 2017, 12, 1683−1690

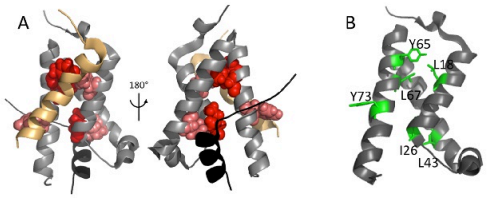


Figure 5. (A) Structural distribution of variants with increased m_{D-N} in the KIX domain. The variants of KIX displaying a detectable increase of m_{D-N} value (*i.e.*, the mutants whose *m* value is at least 0.1 kcal mol^{-1} M^{-1} higher than the wild type) are mapped onto the structure and represented in a sphere. As discussed in the text, it is evident that all the variants are located in the binding sites for either c-Myb (represented in yellow) or MLL (represented in black). The three variants resulting in a relevant stabilization of the native state (*i.e.*, L18A, L43A, and Y73A) are colored in salmon. (B) Structural distribution of the residues displaying a role in stabilizing non-native interactions in the denatured state of KIX as probed by double mutant cycles.

provide an internal control. In fact, by considering the equation $\Delta\Delta\Delta G_{D-TS} = \Delta\Delta\Delta G_{TS-TS'} - \Delta\Delta\Delta G_{D-D'}$, since both the Φ value and LFER analyses suggest the transition state to be native-like, if the coupling free energy $\Delta\Delta\Delta G_{D-N}$ is mainly represented by the $\Delta\Delta\Delta G_{D-D'}$, it follows that, in the absence of non-native like interactions in TS, $\Delta\Delta\Delta G_{D-N} \approx \Delta\Delta\Delta G_{D-TS}$. Indeed, the measured values of $\Delta\Delta\Delta G_{D-TS}$ are linearly correlated with the values of $\Delta\Delta\Delta G_{D-N}$ (Table 2), displaying a slope of 0.93 and $R^2 = 0.87$ (not shown), confirming that the coupling free energies measured by double mutant cycles report mainly on effects of the denatured state.

DISCUSSION

The early work in enzymology has shown that catalysis is improved when reaction intermediates tend to be minimally populated.[13,41] Similarly, in the case of folding, it has been proposed that optimization of folding rate constants, which might be vital for productive folding, may be achieved from the concurrent formation of secondary and tertiary structure from a highly disordered denatured state,[37] *i.e.*, by following the so-called nucleation−condensation mechanism. Consequently, it has been proposed that there is an evolutionary pressure contrasting the formation of stable structural elements in denatured states.

In this work, we have extensively characterized the folding of the KIX domain by studying the equilibrium and kinetic properties of a large number of site-directed variants, in comparison to those of the wild-type protein. While the transition state displays the classical signatures of the nucleation−condensation mechanism, with a native-like topology and a linear LFER plot (Figure 2), analysis of the denatured state properties reveals an additional complexity. In fact, analysis of the folding and unfolding *m* values of this globular domain allows detecting of the presence of a residual structure in the denatured state, which appears to be mainly stabilized by non-native interactions. We mapped some of these non-native interactions by performing double mutant cycles on a limited set of variants displaying an experimentally detectable increase of m_{D-N} value (*i.e.*, those variants which significantly perturbed D by increasing its accessible surface area).

The energy landscape theory proposes that proteins, while marginally stable, are strongly biased toward their native structure, which is therefore minimally frustrated.[42] Because of the contrasting demands of folding and function, however, it is plausible to assume that the functionally competent site(s) of a protein may contain a certain level of local frustration, such that, in those regions, the sequence is not necessarily optimized to fold.[43,44] The presence of these patterns may lead to local misfolding, which may reveal itself in the early events of folding.[45,46] By following this hypothesis, it is of interest to localize on the three-dimensional structure of native KIX the sites of mutations associated in a relevant increase of m_{D-N} value (*i.e.*, the mutants whose *m* value is at least 0.1 kcal mol^{-1} M^{-1} higher than that of the wild type; Figure 5). In fact, because variations at these positions yield to an expansion of the denatured chain, it is likely that they are directly involved in the formation of non-native interactions. Consistently, inspection of Figure 5 suggests that the positions directly involved in residual structure of the denatured state are indeed located in the binding sites of the protein. Three of these mutations, also highlighted in Figure 5, resulted in a detectable stabilization of the native state, providing further support to the hypothesis that sequence is not optimized to fold in these regions. A more direct, even if limited, investigation of the presence of non-native interaction patterns in the denatured state of KIX is represented by the presence of a detectable coupling free energy $\Delta\Delta\Delta G_{D-N}$ between residues not directly interacting in the native state. Among these residues, mapped on the structure of KIX in Figure 5B, it appears that position Y73, located in α3 in the c-Myb binding pocket, plays a particularly important role in stabilizing the residual structure of the denatured state. This finding parallels earlier NMR observations by Tollinger and co-workers,[31,32] who highlighted the capability of α3 to fold independently from the rest of the domain.

Because of its functional complexity, the KIX domain proved to be an interesting system for protein folding studies. In fact, despite its small size and simple topology, KIX interacts with different proteins *via* two distinct binding sites, which occupy a large portion of its surface. In this frame, our experimental results offer the opportunity to test the effects of the divergent demands of function and folding in proteins. Indeed, the residues involved in non-native interactions in the denatured state of KIX are those in the binding sites of the protein; an observation that appears to confirm the notion that the misfolding that may occur at the early stages of folding arises from the residues important for the function of the protein.

DOI: 10.1021/acschembio.7b00289
ACS Chem. Biol. 2017, 12, 1683−1690

MATERIALS AND METHODS

Site-Directed Mutagenesis and Protein Expression and Purification. KIX wild-type and all the site directed mutants were expressed and purified as described previously.[30,47] Mutagenesis was performed using the QuikChange mutagenesis kit (Stratagene).

Stopped-Flow Measurements. The unfolding and refolding kinetics experiments were performed on an SX-18 stopped-flow apparatus (Applied Photophysics, Leatherhead, UK). The protein samples were excited at 280 nm, and emission was measured using a 320 nm cutoff glass filter. The experiments on KIX wild-type and its variants were performed at 20 °C using an asymmetric 1 + 10 mixing in 50 mM sodium phosphate and 150 mM NaCl, at pH 7.2, into solutions containing different urea concentrations. Final protein concentration was typically 3 μM.

Equilibrium Experiments. Equilibrium unfolding experiments on KIX wild-type and its variants were carried out on a Fluoromax single photon counting spectrofluorometer (Jobin−Yvon, NJ, USA), by mixing the native protein with increasing urea concentrations at 20 °C, in 50 mM sodium phosphate and 150 mM NaCl, at pH 7.2, measuring the intrinsic tryptophan emission. Fluorescence spectra were recorded between 300 and 400 nm. The excitation wavelength was 280 nm. Protein concentration was typically 2 μM.

AUTHOR INFORMATION

Corresponding Author

*E-mail: stefano.gianni@uniroma1.it.

ORCID

Stefano Gianni: 0000-0003-1653-1925

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (Progetto di Interesse "Invecchiamento" to S.G.), Sapienza University of Rome (C26A155S48 and B52F16003410005 to S.G), and the Associazione Italiana per la Ricerca sul Cancro (Individual Grant - MFAG 2016, 18701 to S.G.). F.T. was supported by a fellowship from the Italo-French University.

REFERENCES

(1) Jackson, S. E., and Fersht, A. R. (1991) Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry 30*, 10428−10435.

(2) Jackson, W. M., and Brandts, J. F. (1970) Thermodynamics of protein denaturation. A calorimetric study of the reversible denaturation of chymotrypsinogen and conclusions regarding the accuracy of the two-state approximation. *Biochemistry 9*, 2294−2301.

(3) Tanford, C. (1970) Protein denaturation. *Adv. Protein Chem. 24*, 1−95.

(4) Roder, H., Radford, S. E., Capaldi, A. P., Shastry, M. C. R., and Kleanthous, C. (2001) Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat. Struct. Biol. 8*, 68−72.

(5) Jemth, P., Gianni, S., Day, R., Li, B., Johnson, C. M., Daggett, V., and Fersht, A. R. (2004) Demonstration of a low-energy on-pathway intermediate in a fast-folding protein by kinetics, protein engineering, and simulation. *Proc. Natl. Acad. Sci. U. S. A. 101*, 6450−6455.

(6) Parker, M. J., Spencer, J., and Clarke, A. R. (1995) An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J. Mol. Biol. 253*, 771−786.

(7) Shastry, M. C., and Roder, H. (1998) Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nat. Struct. Biol. 5*, 385−392.

(8) Ivarsson, Y., Travaglini-Allocatelli, C., Jemth, P., Malatesta, F., Brunori, M., and Gianni, S. (2007) An on-pathway intermediate in the folding of a PDZ domain. *J. Biol. Chem. 282*, 8568−8572.

(9) Travaglini-Allocatelli, C., Gianni, S., Morea, V., Tramontano, A., Soulimane, T., and Brunori, M. (2003) Exploring the cytochrome c folding mechanism: cytochrome c552 from thermus thermophilus folds through an on-pathway intermediate. *J. Biol. Chem. 278*, 41136−41140.

(10) Gianni, S., Ivarsson, Y., De Simone, A., Travaglini-Allocatelli, C., Brunori, M., and Vendruscolo, M. (2010) Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat. Struct. Mol. Biol. 17*, 1431−1437.

(11) Ivarsson, Y., Travaglini-Allocatelli, C., Brunori, M., and Gianni, S. (2008) Folding and Misfolding in a naturally occurring circularly permuted PDZ domain. *J. Biol. Chem. 283*, 8954−8960.

(12) Fersht, A. R., Matouschek, A., and Serrano, L. (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol. 224*, 771−782.

(13) Fersht, A. R. (1999) *Structure and Mechanism in Protein Science*, Freeman, New York.

(14) McCarney, E. R., Kohn, J. E., and Plaxco, K. W. (2005) Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Crit. Rev. Biochem. Mol. Biol. 40*, 181−189.

(15) Morrone, A., McCully, M. E., Bryan, P. N., Brunori, M., Daggett, V., Gianni, S., and Travaglini-Allocatelli, C. (2011) The denatured state dictates the topology of two proteins with almost identical sequence but different native structure and function. *J. Biol. Chem. 286*, 3863−3872.

(16) Religa, T. L., Markson, J. S., Mayor, U., Freund, S. M., and Fersht, A. R. (2005) Solution structure of a protein denatured state and folding intermediate. *Nature 437*, 1053−1056.

(17) van Gunsteren, W. F., Bürgi, R., Peter, C., and Daura, X. (2001) The Key to Solving the Protein-Folding Problem Lies in an Accurate Description of the Denatured State. *Angew. Chem., Int. Ed. 40*, 351−355.

(18) Wong, K. B., Clarke, J., Bond, C. J., Neira, J. L., Freund, S. M., Fersht, A. R., and Daggett, V. (2000) Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. *J. Mol. Biol. 296*, 1257−1282.

(19) Camilloni, C., Bonetti, D., Morrone, A., Giri, R., Dobson, C. M., Brunori, M., Gianni, S., and Vendruscolo, M. (2016) Towards a structural biology of the hydrophobic effect in protein folding. *Sci. Rep. 6*, 28285−28291.

(20) Radhakrishnan, I., Pérez-Alvarado, G. C., Parker, D., Dyson, H. J., Montminy, M. R., and Wright, P. E. (1997) Solution structure of the KIX domain of CBP bound to the transactivation domain of CREB: a model for activator:coactivator interactions. *Cell 91*, 741−752.

(21) Zor, T., De Guzman, R. N., Dyson, H. J., and Wright, P. E. (2004) Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol. 337*, 521−534.

(22) Brüschweiler, S., Schanda, P., Kloiber, K., Brutscher, B., Kontaxis, G., Konrat, R., and Tollinger, M. (2009) Direct observation of the dynamic process underlying allosteric signal transmission. *J. Am. Chem. Soc. 131*, 3063−3068.

(23) Law, S. M., Gagnon, J. K., Mapp, A. K., and Brooks, C. L. r. (2014) Prepaying the entropic cost for allosteric regulation in KIX. *Proc. Natl. Acad. Sci. U. S. A. 111*, 12067−12072.

(24) Lee, C. W., Arai, M., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. (2009) Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP. *Biochemistry 48*, 2115−2124.

(25) Toto, A., Camilloni, C., Giri, R., Brunori, M., Vendruscolo, M., and Gianni, S. (2016) Molecular Recognition by Templated Folding of an Intrinsically Disordered Protein. *Sci. Rep. 6*, 21994−22000.

(26) Toto, A., Giri, R., Brunori, M., and Gianni, S. (2014) The mechanism of binding of the KIX domain to the mixed lineage leukemia protein and its allosteric role in the recognition of c-Myb. *Protein Sci. 23*, 962−969.

DOI: 10.1021/acschembio.7b00289
ACS Chem. Biol. 2017, 12, 1683−1690

(27) Toto, A., and Gianni, S. (2016) Mutational Analysis of the Binding-Induced Folding Reaction of the Mixed-Lineage Leukemia Protein to the KIX Domain. *Biochemistry 55*, 3957−3962.
(28) Denis, C. M., Langelaan, D. N., Kirlin, A. C., Chitayat, S., Munro, K., Spencer, H. L., LeBrun, D. P., and Smith, S. P. (2014) Functional redundancy between the transcriptional activation domains of E2A is mediated by binding to the KIX domain of CBP/p300. *Nucleic Acids Res. 42*, 7370−7382.
(29) Horng, J. C., Tracz, S. M., Lumb, K. J., and Raleigh, D. P. (2005) Slow folding of a three-helix protein via a compact intermediate. *Biochemistry 44*, 627−634.
(30) Morrone, A., Giri, R., Brunori, M., and Gianni, S. (2012) Reassessing the folding of the KIX domain: evidence for a two-state mechanism. *Protein Sci. 21*, 1775−1779.
(31) Schanda, P., Brutscher, B., Konrat, R., and Tollinger, M. (2008) Folding of the KIX domain: characterization of the equilibrium analog of a folding intermediate using 15N/13C relaxation dispersion and fast 1H/2H amide exchange NMR spectroscopy. *J. Mol. Biol. 380*, 726−741.
(32) Tollinger, M., Kloiber, K., Agoston, B., Dorigoni, C., Lichtenecker, R., Schmid, W., and Konrat, R. (2006) An isolated helix persists in a sparsely populated form of KIX under native conditions. *Biochemistry 45*, 8885−8893.
(33) Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci. 4*, 2138−2148.
(34) Fersht, A. R., and Sato, S. (2004) Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U. S. A. 101*, 7976−7981.
(35) Fersht, A. R. (1995) Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol. 5*, 79−84.
(36) Leffler, J. E. (1953) Parameters for the description of transition states. *Science 117*, 340−341.
(37) Fersht, A. R. (1995) Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U. S. A. 92*, 10869−10873.
(38) Itzhaki, L. S., Otzen, D. E., and Fersht, A. R. (1995) The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. *J. Mol. Biol. 254*, 260−288.
(39) Sanchez, I. E., and Kiefhaber, T. (2003) Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol. 327*, 867−884.
(40) Scaloni, F., Gianni, S., Federici, L., Falini, B., and Brunori, M. (2009) Folding mechanism of the C-terminal domain of nucleophosmin: residual structure in the denatured state and its pathophysiological significance. *FASEB J. 23*, 2360−2365.
(41) Pauling, L. (1946) Molecular Architecture and Biological Reaction. *Chem. Eng. News 24*, 1375−1377.
(42) Bryngelson, J. D., Onuchic, J. N., Socci, N. D., and Wolynes, P. G. (1995) Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct., Funct., Genet. 21*, 167−195.
(43) Ferreiro, D. U., Hegler, J. A., Komives, E. A., and Wolynes, P. G. (2007) Localizing frustration in native proteins and protein assemblies. *Proc. Natl. Acad. Sci. U. S. A. 104*, 19819−19824.
(44) Ferreiro, D. U., Hegler, J. A., Komives, E. A., and Wolynes, P. G. (2011) On the role of frustration in the energy landscapes of allosteric proteins. *Proc. Natl. Acad. Sci. U. S. A. 108*, 3499−3503.
(45) Di Silvio, E., Brunori, M., and Gianni, S. (2015) Frustration Sculpts the Early Stages of Protein Folding. *Angew. Chem., Int. Ed. 54*, 10867−10869.
(46) Gianni, S., Camilloni, C., Giri, R., Toto, A., Bonetti, D., Morrone, A., Sormanni, P., Brunori, M., and Vendruscolo, M. (2014) Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc. Natl. Acad. Sci. U. S. A. 111*, 14141−14146.
(47) Giri, R., Morrone, A., Toto, A., Brunori, M., and Gianni, S. (2013) Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. *Proc. Natl. Acad. Sci. U. S. A. 110*, 14942−14947.

DOI: 10.1021/acschembio.7b00289
ACS Chem. Biol. 2017, 12, 1683−1690

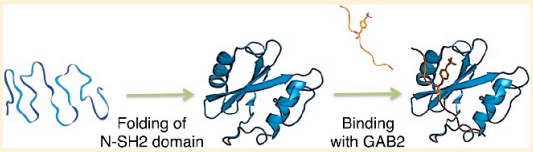
Paper 2: SHP2 is a phosphatase protein involved in many cellular pathways comprising two SH2 domains (called N-SH2 and C-SH2) and a phosphatase domain. SHP2 interacts with many partners, among which the Grb2 associated binder (Gab2), a disordered scaffolding protein. The interaction between SHP2 and Gab2 is critical in cell death and differentiation. SHP2 binds to Gab2 through its SH2 domains, which recognize specific regions of Gab2 characterized by the presence of a phosphorylated tyrosine. In order to shed light onto the dynamic and functional properties of this protein-protein interaction, we studied the mechanism of folding of N-SH2 and the binding process with Gab2 using a peptide mimicking the interacting region of Gab2. The data presented represent the first description by stopped-flow of the kinetics of binding of an SH2 domain in solution. We also could elucidate the electrostatic nature of the interaction by performing experiments at different ionic strengths, highlighting the key role of the negative charge of the phosphotyrosine in the recognition event. Furthermore, by analyzing the equilibrium and kinetics of folding of N-SH2 we demonstrate the presence of an intermediate along the folding pathway.

THE JOURNAL OF PHYSICAL CHEMISTRY B
Article
Cite This: J. Phys. Chem. B XXXX, XXX, XXX−XXX
pubs.acs.org/JPCB

Mechanism of Folding and Binding of the N-Terminal SH2 Domain from SHP2

Daniela Bonetti, Francesca Troilo, Angelo Toto, Carlo Travaglini-Allocatelli, Maurizio Brunori,* and Stefano Gianni*

Istituto Pasteur, Fondazione Cenci Bolognetti, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" and Istituto di Biologia e Patologia Molecolari del CNR, Sapienza Università di Roma, 00185, Rome, Italy



ABSTRACT: SHP2 is a phosphatase protein, involved in many cellular pathways, comprising two SH2 domains (namely N-SH2 and C-SH2) and a phosphatase domain. Among others, the interaction between SHP2 and Gab2 (Grb2 associated binder) is critical in cell death and differentiation. SHP2 binds to Gab2 through its SH2 domains, which recognize specific regions of Gab2 characterized by the presence of a phosphorylated tyrosine. In order to shed light on the dynamic and functional properties of this protein−protein interaction, we studied the mechanism of folding of N-SH2 and the binding process to a peptide mimicking a region of Gab2. The data presented represent the first description by stopped-flow of the kinetics of binding of an SH2 domain in solution. By performing experiments at different ionic strengths, we elucidate the electrostatic nature of the interaction, highlighting a key role of the negative charge of the phosphotyrosine in the recognition event of the reaction. Furthermore, by analyzing the equilibrium and kinetics of folding of N-SH2 folding we demonstrate the presence of an intermediate along the folding pathway. These results are discussed in the light of previous works on another SH2 domain.

INTRODUCTION

Src Homology 2 (SH2) domains are conserved structural motifs of about 100 amino acids that are often present in large multidomain proteins, specifically recognizing short amino acid sequences exposing phosphorylated tyrosine (phosphotyrosine) residues. The ability to recognize phosphotyrosine residues, coupled to the presence of other structural domains involved in intermolecular recognition and/or catalytic activity (e.g., PDZ, SH3, WW domains), allows these multidomain proteins to control the colocalization of a variety of proteins and effectors within a signaling cascade. Not surprisingly, many SH2 domain-containing proteins have been implicated in different diseases, such as diabetes, immune system disorders or cancer.[1−6] Despite these critical roles, the mechanism of binding and recognition between the SH2 domain and their partners has not been fully characterized, and the kinetic association and dissociation rate constants have been studied to date only by surface plasmon resonance (see for example refs 7−9).

SH2 domains belong to the $\alpha + \beta$ structural class: their conserved architecture is composed of two α-helices and a central β-sheet composed of three to five antiparallel β-strands.[10,11] The structure of different SH2 complexes shows that the bound ligand is generally orthogonal to the β-strands in an extended conformation.[12] However, different modes of ligand recognition have been described within this general framework.[13] The complex is stabilized by a network of hydrogen bonds and electrostatic interactions, mainly provided by the phosphate oxygens of the phosphotyrosine and by electrostatic or hydrophobic interactions mediated by residues C-terminal to the phosphotyrosine.[10,11]

Shp2 is a protein phosphatase involved in the Ras-ERK1/2 signaling pathway;[14] it is a multidomain protein containing two adjacent SH2 domains (N-SH2 being the N-terminal and C-SH2 the C-terminal) and a phosphatase (PHP) domain arranged in the order N-SH2, C-SH2, and PHP (from the N- to the C-termini of the full-length protein). It has been shown that N-SH2 acts as a conformational switch:[15] in the absence of its phosphotyrosine target, the N-SH2 binds to the catalytic site of the phosphatase domain of Shp2 thus inhibiting enzymatic activity, whereas in the presence of a phosphopeptide it undergoes a conformational rearrangement leading to dissociation from the phosphatase active site and activation of the enzyme. Shp2 is an allosteric enzyme, characterized by the fact that distinct molecular surfaces of the N-SH2 domain recognize the phosphotyrosine and the phosphatase domain.

Special Issue: William A. Eaton Festschrift

Received: June 13, 2018
Revised: July 25, 2018
Published: July 26, 2018

DOI: 10.1021/acs.jpcb.8b05651
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Binding of a phosphopeptide ligand at one site induces a conformational change that is transmitted to the PHP domain binding site.[15]

In order to shed light on the dynamic and functional properties of the N-SH2 domain, we here present the characterization of the mechanism of folding together with an analysis of the binding process to Gab2 (Grb2 associated binder), one of the physiological targets of Shp2.[16] Taking advantage of the fluorescence change due to Forster resonance energy transfer between the tryptophan (Trp) residues of the SH2 and a dansylated peptide, we document the first description of the kinetics of binding of a SH2 domain in solution. Furthermore, by performing experiments at different ionic strengths, we clarify the electrostatic nature of the interaction, as well as the ability of the protein to bind inorganic phosphate. From a folding perspective, we demonstrate the presence of a folding intermediate, which accumulates in the microsecond time scale. This finding contrasts what previously suggested in the folding of the Src SH2 domain.[17]

METHODS

Protein Expression and Purification. The construct encoding the N-SH2 domain of SHP2 protein (residues 3−104) was subcloned in a pET28b+ plasmid vector and then transformed in *Escherichia coli* cells BL21 (DE3). Bacterial cells were grown in LB medium, containing 30 μg/mL of kanamycin, at 37 °C until OD_{600} = 0.7−0.8, and then protein expression was induced with 1 mM IPTG. After induction, cells were grown at 37 °C overnight and then collected by centrifugation.

To purify the protein, the pellet was resuspended in buffer made of 50 mM sodium phosphate, 300 mM NaCl, pH 7.2, with the addition of antiprotease tablet (Complete EDTA-free, Roche), then sonicated and centrifuged. The soluble fraction from bacterial lysate was loaded onto a nickel-charged HisTrap Chelating HP (GE Healthcare) column equilibrated with 50 mM sodium phosphate, 300 mM NaCl, pH 7.2. Protein was then eluted with a gradient from 0 to 0.5 M imidazole by using an AKTA-prime system. Fractions containing the protein were collected and the buffer was exchanged to 50 mM Hepes pH 7.0 by using a HiTrap Desalting column (GE Healthcare). The purity of the protein was analyzed through SDS-page.

The N_{TERM} dansylated and nondansylated peptides of $Gab2_{608-620}$ (N_{TERM}-STGSVDYLALDFQ-C_{TERM}), with or without phosphorylated Tyr614, were purchased from JPT, Germany. Peptide concentrations were measured by recording the absorbance of dansyl group at the wavelength of 330 nm, and absorbance of tyrosine residue at 280 nm for non-dansylated version of the peptide.

Equilibrium Experiments. Equilibrium unfolding experiments were performed on a Fluoromax single photon counting spectrofluorometer (Jobin-Yvon, NJ, USA). N-SH2 protein, at a constant concentration of 3 μM, was excited at 280 nm and emission spectra were recorded between 300 and 400 nm, at increasing denaturant concentration. Experiments were performed at 25 °C, using a quartz cuvette with a path length of 1 cm, in buffer 50 mM TrisHCl pH 8.0. Data were fitted using eq 1:

$$Y_{obs} = (Y_N + Y_D)\frac{e^{m_{D-N}([urea]-[urea]_{1/2})}}{1 + e^{m_{D-N}([urea]-[urea]_{1/2})}} \quad (1)$$

Stopped-Flow Folding Experiments. Unfolding and refolding kinetics experiments were performed on a single-mixing SX-18 stopped-flow instrument (Applied Photophysics), monitoring the change of fluorescence emission. The excitation wavelength used was 280 nm and the fluorescence emission light was recorded by using a 320 nm cutoff glass filter. The experiments were performed at 25 °C, by using urea or guanidine hydrochloride as denaturant agent. The buffer used was 50 mM TrisHCl pH 8.0. For each denaturant concentration usually five individual traces were averaged. The final protein concentration was typically 3 μM. Semilogarithmic plot of observed rate constants versus [urea] (i.e., chevron plot) was fitted using eq 2:

$$k_{obs} = \frac{k_{I-N}\exp(-m_{I-N}[urea]/RT)}{1 + (1/k_{D-I})\exp(m_{D-I}[urea]/RT)} + k_{N-I}\exp(m_{N-I}[urea]/RT) \quad (2)$$

ANS Binding Experiments. 1-anilinonaphthalene-8-sulfonate (ANS) is a hydrophobic dye which binds hydrophobic surfaces usually found in folding intermediates.[18] Refolding experiments were performed on a single-mixing SX-18 stopped-flow instrument (Applied Photophysics) at 25 °C. N-SH2 at final concentration of 3 μM, in the presence of 50 mM Hepes, pH 7.2, and 300 μM ANS, was excited at 280 nm and ANS fluorescence was recorded using a 455 nm cutoff filter. Refolding experiments were initiated by a 11-fold dilution of the denatured protein with the appropriate buffer, progressively changing denaturant concentration.

Equilibrium binding experiments were performed on a Fluoromax single photon counting spectrofluorometer (Jobin-Yvon, NJ, USA), in absence and in the presence of 7.9 M urea. An excitation wavelength of 350 nm was used and fluorescence emission light was recorded in a quartz cuvette with 1 cm path length, between 400 and 600 nm. Final conditions were 3 μM N-SH2, 50 mM Hepes, pH 7.2, and 300 μM ANS at 25 °C.

Stopped-Flow Binding Experiments. Kinetics binding experiments were performed on a single-mixing SX-18 stopped-flow instrument (Applied Photophysics). Pseudo-first-order experiments were performed mixing a constant concentration (1 μM) of $Gab2_{608-620}$ dansylated and phosphorylated peptide with an increasing concentration of N-SH2, from 1 to 12 μM. Samples were excited at 280 nm, and the emission fluorescence light of the dansyl group was followed by using a 475 nm cutoff filter. Experiments were performed at 10 °C, and the buffer used was 50 mM Hepes pH 7.2, at different NaCl concentrations (from 0 to 1.2 M). For each acquisition, five traces were collected and averaged and then fitted to a single exponential equation. The observed rate constant was plotted as a function of N-SH2 concentration and the association rate constant (k_{on}) was calculated as the slope of a linear function fitting the observed rate constant dependence. The dissociation rate constant (k_{off}) was calculated from displacement experiments.

Observed rate constant linear dependences versus [N-SH2] were fitted using the eq 3:

$$k_{obs} = k_{on}[\text{NSH2}] + k_{off} \quad (3)$$

Stopped-Flow Displacement Experiments. The dissociation rate constant was measured by performing experiments on a single-mixing SX-18 stopped-flow instrument (Applied Photophysics). A preincubated complex of N-SH2 (1 μM) and $Gab2_{608-620}$ dansylated and phosphorylated (1 μM)

DOI: 10.1021/acs.jpcb.8b05651
J. Phys. Chem. B XXXX, XXX, XXX−XXX

was rapidly mixed with different concentrations of an excess of $Gab2_{608-620}$ phosphorylated from 10 μM to 20 μM. The excitation wavelength used was 280 nm and fluorescence emission light was collected by using a 475 nm cutoff filter. Experiments were performed at 10 °C, in the presence of 50 mM Hepes, pH 7.2, at increasing NaCl concentration ranging from 0 to 1.2 M. The observed rate constants were calculated from the average of five single traces. Observed kinetics was consistent with a single exponential decay.

RESULTS

The principal aim of this study is to provide a complete characterization of the folding and binding properties of N-SH2. In fact, while the SH2 family plays a critical role in several protein−protein interactions,[19] the biophysical work necessary to unveil the mechanism of folding and binding is still relatively limited.

The Folding Pathway of Shp2 N-SH2. In an effort to characterize the folding of N-SH2, we conducted both equilibrium and kinetic experiments. Thus, the thermodynamic stability of N-SH2 was determined by urea-induced equilibrium denaturation experiments at 25 °C, in 50 mM Tris HCl buffer at pH 8.0, monitoring the change of intrinsic fluorescence emission of the single tryptophan residue (Trp6, highlighted in Figure 1). As reported in Figure 2, the observed

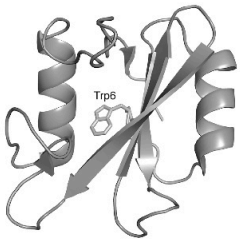


Figure 1. Structure of N-SH2 domain from SHP2 protein (PDB: 4QSY). The tryptophan residue at position 6 is highlighted as sticks.

decrease in fluorescence emission on increasing denaturant concentration is consistent with a two-state equilibrium unfolding mechanism.[20] The unfolding free energy in water (ΔG^0), calculated from a two-state model (eq 1 in Materials and Methods), is 5.1 ± 0.3 kcal/mol. The calculated $m_{D\text{-}N}$ value, which reflects the change in accessible surface area to the solvent upon unfolding, is 1.1 ± 0.1 kcal mol^{-1} M^{-1}, consistent with a protein of 108 residues.[21]

In order to obtain information on the folding mechanism of N-SH2, we performed folding and unfolding kinetic experiments at 25 °C, in 50 mM Tris HCl buffer at pH 8.0 by stopped-flow fluorimetry. At all the investigated denaturant concentrations, the (un)folding time courses were satisfactorily fitted to a single exponential decay. The logarithm of the observed rate constant as a function of denaturant concentration (chevron plot) of N-SH2 is reported in Figure 3. It is evident that at [urea] < 3 M the overall dependence of the observed rate constant displays a pronounced curvature, classically denoted as roll-over effect. This observation is generally interpreted as reflecting the population of an

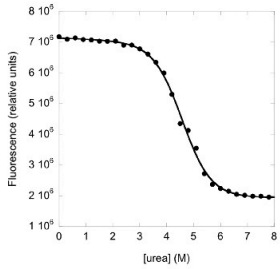


Figure 2. Equilibrium denaturation experiment of the N-SH2 domain carried out in buffer TrisHCl 50 mM pH 8.0 at 25 °C. The change of the intrinsic fluorescence of the tryptophan residue versus [urea] was fitted with a two-state equation (see text for details).

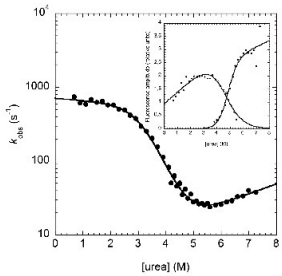


Figure 3. Semilogarithmic plot of the observed (un)folding rate constant versus [urea] measured with the stopped-flow apparatus in TrisHCl 50 mM pH 8.0 at 25 °C. The continuous line represents the best fit to a three-state folding mechanism (see text for details). Kinetic parameters obtained from data fitting are $k_{I\text{-}N}$ = 720 ± 50 s^{-1}, $m_{I\text{-}N}$ = 0.05 ± 0.01 kcal M^{-1} mol^{-1}, $k_{D\text{-}I}$ = 700 ± 200 s^{-1}, $m_{D\text{-}I}$ = 1.21 ± 0.05 kcal M^{-1} mol^{-1}, $k_{N\text{-}I}$ = 3.0 ± 0.9 s^{-1}, and $m_{N\text{-}I}$ = 0.21 ± 0.02 kcal M^{-1} mol^{-1}. Inset panel: analysis of the amplitudes of the traces obtained in unfolding (closed circles) and refolding (open circles) experiments as a function of [urea]. Lines represent the best fit to a sigmoid function.

intermediate species accumulating during the folding process and therefore demands a minimal three-state model to adequately fit the kinetic data.[22−27] The chevron plot was therefore fitted using eq 2 (see Material and Methods) and the kinetic parameters obtained from this analysis are listed in the legend to Figure 3. It should be noticed that the analysis of the (un)folding amplitudes as a function of urea concentration (inset to Figure 3) did not reveal the presence of a clear burst in fluorescence in the subms time range, with the linear extrapolation of the unfolding amplitudes at 0 M urea closely matching the observed refolding amplitudes at very low denaturant concentrations. This finding indicates that the fluorescence of the folding intermediate is similar to that of the denatured state.

DOI: 10.1021/acs.jpcb.8b05651
J. Phys. Chem. B XXXX, XXX, XXX−XXX

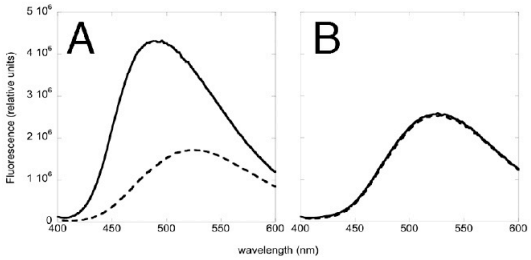


Figure 4. ANS fluorescence emission in presence (continuous line) and in absence (broken line) of N-SH2 in buffer containing 0 M urea (panel A) and 8 M urea (panel B).

Inspection of Figure 3 clearly reveals that the roll-over effect of N-SH2 occurs in a time scale that approaches the dead-time of the stopped-flow apparatus and might therefore be vitiated by the limitation of the instrument. Thus, we resorted to provide additional experiments to confirm the genuine presence of such intermediate by monitoring the binding of 1-anilinonaphthalene-8-sulfonate (ANS), a fluorescent dye that binds to hydrophobic cavities or surfaces generally found in partially folded proteins,[18] to the relevant states in the folding reaction. The binding of ANS to the native and denatured state of N-SH2 was first followed by equilibrium experiments, by measuring the fluorescence of the dye in the presence of N-SH2 at a concentration of 0 and 8 M urea (Figure 4). It is evident that, while ANS binds the native state, as probed by a detectable increase in fluorescence emission upon binding, its fluorescence spectrum was unaffected by the presence of N-SH2 under denaturing conditions, indicating that binding to the denatured state is negligible. Binding to the native state most likely occurs through a surface exposed hydrophobic patch involving residues W6, F7, F41, and L74. Consequently, we performed stopped-flow refolding experiments challenging denatured N-SH2 against a refolding buffer containing ANS (Figure 5). Interestingly, monitoring refolding, we observed a rapid decrease in the fluorescence emission of ANS, with an observed rate constant that was very similar to that observed in the absence of the dye. Since the denatured state of N-SH2 does not bind the fluorescent dye, and the fluorescence emission of ANS is higher in the presence of the native state of N-SH2 than in its absence, the observed decrease in fluorescence may be interpreted as a genuine signature of the time-dependent decrease in the population of a transiently formed hyper-fluorescent intermediate.

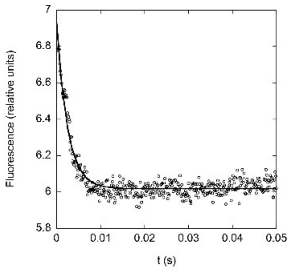


Figure 5. Stopped-flow refolding experiment trace obtained rapidly mixing denatured N-SH2 against a refolding buffer (50 mM Hepes, pH 7.2) containing 300 μM ANS, at 25 °C. The trace has been obtained as the average of 5 independently recorded traces and the line is the best fit to a single exponential function.

Ligand Binding. As outlined in the Introduction, the N-terminal SH2 domain of SHP2 plays an important role in modulating protein interactions in the assembly of signaling complexes.[1−6,10,11,15] However, although information about the specificity of binding of N-SH2 and its affinity toward different targets is available in the literature,[19] very little is known about the details of the binding mechanism of these modules. Thus, we decided to carry out kinetic binding experiments on N-SH2, using a synthetic peptide derived from the sequence of Gab2, a physiological target of N-SH2.[28]

The binding of N-SH2 to Gab2 was tested on a dansylated peptide (called $Gab2_{608-620}$; see Material and Methods section) mimicking the SH2 binding region of Gab2 and containing the phosphotyrosine Tyr614, specifically recognized by the N-SH2 domain at the interface between helix1 and strands 1 and 2.[28] The interaction between N-SH2 and the $Gab2_{608-620}$ peptide was monitored by fluorescence stopped-flow experiments following the change in the FRET signal, taking advantage of the Trp6 residue of N-SH2 as fluorescence donor and the dansyl group specifically attached to the N-terminus of $Gab2_{608-620}$ as acceptor. Since the reaction appeared to be too fast at 25 °C to be reliably followed by stopped-flow, we resorted to carrying out these experiments at 10 °C. At this temperature, and under pseudo-first order conditions, the binding time course matched single exponential at every concentrations of N-SH2. The observed rate constants show a linear dependence as a function of N-SH2 concentration (Figure 6), suggesting that binding takes place via a two-state mechanism. Fitting the data shown in Figure 6 to eq 3 (see Materials and Methods) allowed us to calculate the association rate constant $k_{on} = 167 \pm 7\ s^{-1}\ \mu M^{-1}$. Because of the large error associated with extrapolation of the linear fit to the x axis, the dissociation rate constant (k_{off}) was obtained independently from a displacement experiment in which the dansylated $Gab2_{608-620}$ peptide was preincubated with N-SH2 (1:1 stoichiometric ratio) and then challenged with different

DOI: 10.1021/acs.jpcb.8b05651
J. Phys. Chem. B XXXX, XXX, XXX−XXX

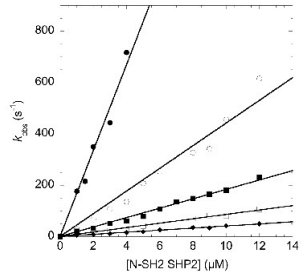


Figure 6. Pseudo-first order kinetics of the binding reaction between Dansyl-Gab2 versus different concentrations of N-SH2, in buffer Hepes 50 mM pH 7.2 with 1.2 M NaCl (closed circles), 600 mM NaCl (open circles), 300 mM NaCl (closed squares), 150 mM NaCl (open squares), and without NaCl (black diamonds), at 10 °C. Points at 0 μM N-SH2 were obtained from independent displacement experiments.

concentrations of an excess of nondansylated $Gab2_{608-620}$. The dissociation rate constant obtained by the displacement experiments (k_{off} = 1.1 ± 0.1 s^{-1}) allowed us to calculate the equilibrium dissociation constant (K_D) for the interaction between N-SH2 and the $Gab2_{608-620}$ peptide at 10 °C (K_D = 6.8 ± 0.7 nM); this result is in agreement with published equilibrium dissociation constants for this class protein,[13,19,29] and highlights a strong affinity between the two partners.

In order to explore the contribution of electrostatic interactions in controlling the affinity of complex formation, we performed kinetic binding experiments as a function of ionic strength. As shown in Table 1, the increase in ionic strength (from 0 to 1.2 M NaCl) leads to a progressive decrease of the association rate constant, but, interestingly, no variation could be detected in the dissociation rate constant. This observation suggests that the electrostatic attraction between $Gab2_{608-620}$ and N-SH2 is the driving force for fast association and formation of the encounter complex

Table 1. Kinetic Parameters of Binding of N-SH2 to a Synthetic Peptide Mimicking Gab2, Measured at Different Ionic Strengths[a]

NaCl (mM)	k_0 (s^{-1} μM^{-1})	k_{off} (s^{-1})	K_D (nM)
0	167 ± 7	1.1 ± 0.1	6.8 ± 0.7
150	44 ± 2	1.4 ± 0.1	33 ± 3
300	18.2 ± 0.3	1.8 ± 0.1	100 ± 10
600	8.5 ± 0.1	1.3 ± 0.1	150 ± 20
1200	4.0 ± 0.1	1.1 ± 0.1	270 ± 30

[a]Microscopic association (k_{on}) and dissociation (k_{off}) rate constants and equilibrium dissociation constant (K_D) calculated from pseudo-first order kinetic binding experiments and displacement experiments between Gab2 and N-SH2 in buffer Hepes 50 mM, pH 7.2, at 10 °C, at different concentrations of NaCl (from 0 to 1.2 M).

In an effort to further clarify the binding capabilities of N-SH2, we resorted to explore whether this domain could also recognize inorganic phosphate. In this case, however, since no fluorescence change could be detected when the protein was challenged with this anion, we set up an experimental test to monitor binding indirectly. In particular according to the mass action law, it may be assumed that, given that the binding of phosphate occurs only when the N-SH2 domain is folded, this anion should be capable to shift the equilibrium between denatured and native states of the protein toward the latter. Therefore, we mixed N-SH2 with solutions containing different concentrations of sodium phosphate 0−100 μM, in the presence of 5.5 M urea (i.e., at the midpoint of the chevron plot where 50% of the population of the protein is unfolded; see Figure 3). As expected from tight binding, addition of low concentrations phosphate induced refolding with an apparent K_D = 2.1 ± 1.0 μM, as obtained from the amplitude analysis of the refolding transition as a function of phosphate concentration (Figure 7).

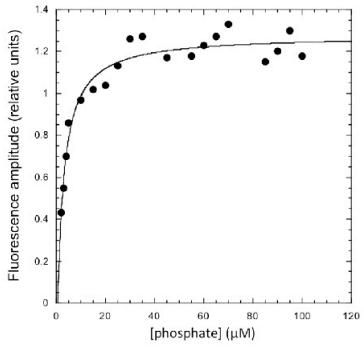


Figure 7. Dependence and analysis of the amplitudes of the observed kinetic rate constants obtained mixing N-SH2 with buffer containing different concentration of sodium phosphate at pH 7.2, in the presence of 5.5 M urea, at 25 °C.

■ DISCUSSION

The SH2 proteins represent an important class of domains, capable of mediating protein−protein interactions in several key metabolic pathways. In this context, understanding the mechanism whereby these proteins recognize their physiological partners is of critical importance to define their affinity and specificity. Moreover, folding experiments may be important to characterize their thermodynamic stability as well as to identify the presence of metastable intermediate states.

In this work we presented the characterization of the folding kinetics of N-SH2 from SHP2. Data reveal the presence of a low-energy folding intermediate, which accumulates transiently in the dead time of the stopped-flow apparatus.[23,24] Src SH2, sharing 50% sequence similarity and 35% sequence identity with N-SH2, represents the only other example in the literature where the folding kinetics of an SH2 domain was described.[17] For Src SH2, folding was observed to occur via a two-state scenario, with the denatured state containing a considerable amount of residual structure. Thus, a comparison of the folding kinetics of the two SH2 domains supports the notion that the presence of a folding intermediate is not necessary for efficient

DOI: 10.1021/acs.jpcb.8b05651
J. Phys. Chem. B XXXX, XXX, XXX−XXX

folding, which may occur productively in a two-state manner.[30] This finding, which was originally put forward from the comparison of the structurally homologous proteins Im7 and Im9,[25,31,32] indicates that intermediates may be selectively stabilized by the inherent propensity of the amino acidic sequence to fold independently, prior the formation of the native state. Furthermore, it should be noticed that, while the domain has been studied in isolation from the rest of the protein, previous work on protein folding suggested that the general features of folding of protein domains is not substantially affected by the surrounding domains.

This is the first study where the rate constants for the interaction of an SH2 domain were determined in solution. Observed kinetics indicates that N-SH2 binds its phosphorylated target in a simple bimolecular reaction, without evidence for rate-limiting conformational changes. However, it should be noticed that fast conformational changes (in the subms time regime) might not be excluded, due to the limitation of the stopped-flow apparatus. The association and dissociation rate constants were both 1 order of magnitude higher than those measured by surface plasmon resonance for the SH2 of p85, while displaying a very similar K_D in the nM concentration range.[33] It is unclear at this stage whether this discrepancy arises from a genuine difference between the two domains or, in analogy to what observed in PDZ domains,[34−36] from the different experimental techniques employed. While the dissociation rate constant is insensitive to ionic strength, we observed that the association rate constant is highly dependent on the concentration of NaCl. This finding indicates that the binding transition state is stabilized by the charge−charge interaction between the phosphorylated-peptide and N-SH2. Interestingly, while we could not detect binding when performing experiments using the nonphosphorylated peptide, it appears that inorganic phosphate at micromolar concentrations is capable of binding the protein tightly to native N-SH2, as shown by its ability to refold the protein exposed to mildly denaturing conditions. This affinity is similar to that observed for the binding of Src SH2 to random peptide containing a phosphorylated tyrosine,[19] indicating that while the specificity of the physiological interaction mediated by SH2 domains is tuned by the residues surrounding the phosphotyrosine, the phosphate group has a high affinity *per se*. Future work based on protein engineering will further clarify the molecular details of the interaction.

AUTHOR INFORMATION

Corresponding Authors

*(S.G.) E-mail: stefano.gianni@uniroma1.it.

*(M.B.) E-mail: maurizio.brunori@uniroma1.it.

ORCID

Stefano Gianni: 0000-0003-1653-1925

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Work partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (Progetto di Interesse "Invecchiamento" to S.G.), Sapienza University of Rome (C26A155S48, B52F16003410005 and RP11715C34AEAC9B to S.G), the Associazione Italiana per la Ricerca sul Cancro (Individual Grant MFAG 2016, 18701 to S.G.). F.T. is a recipient of a Ph.D. fellowship from the Italo-French University.

REFERENCES

(1) Mayer, B. J. What Have We Learned from SH2 Domains? *Methods Mol. Biol.* **2017**, *1555*, 37−43.

(2) Chia, D. J.; Subbian, E.; Buck, T. M.; Hwa, V.; Rosenfeld, R. G.; Skach, W. R.; Shinde, U.; Rotwein, P. Aberrant folding of a mutant Stat5b causes growth hormone insensitivity and proteasomal dysfunction. *J. Biol. Chem.* **2006**, *281*, 6552−6558.

(3) Zhang, M.; Jang, H.; Gaponenko, V.; Nussinov, R. Phosphorylated Calmodulin Promotes PI3K Activation by Binding to the SH2 Domains. *Biophys. J.* **2017**, *113*, 1956−1967.

(4) Mahajan, V. S.; Pillai, S. Sialic acids and autoimmune disease. *Immunol. Rev.* **2016**, *269*, 145−161.

(5) Morlacchi, P.; Robertson, F. M.; Klostergaard, J.; McMurray, J. S. Targeting SH2 domains in breast cancer. *Future Med. Chem.* **2014**, *6*, 1909−1926.

(6) Li, S.; Hsu, D. D.; Wang, H.; Feng, G. S. Dual faces of SH2-containing protein-tyrosine phosphatase Shp2/PTPN11 in tumorigenesis. *Front. Med.* **2012**, *6*, 275−279.

(7) Chen, C. H.; Chen, M. K.; Jeng, K. C.; Lung, F. D. Effects of peptidic antagonists of Grb2-SH2 on human breast cancer cells. *Protein Pept. Lett.* **2010**, *17*, 44−53.

(8) Zhang, Y.; Wavreille, A.-S.; Kunys, A. R.; Pei, D. The SH2 domains of inositol polyphosphate 5-phosphatases SHIP1 and SHIP2 have similar ligand specificity but different binding kinetics. *Biochemistry* **2009**, *48*, 11075−11083.

(9) de Mol, N. J.; Catalina, M. I.; Dekker, F. J.; Fischer, M. J.; Heck, A. J.; Liskamp, R. M. Protein flexibility and ligand rigidity: a thermodynamic and kinetic study of ITAM-based ligand binding to Syk tandem SH2. *ChemBioChem* **2005**, *6*, 2261−2270.

(10) Waksman, G.; Kominos, D.; Robertson, S. C.; Pant, N.; Baltimore, D.; Birge, R. B.; Cowburn, D.; Hanafusa, H.; Mayer, B. J.; Overduin, M.; et al. Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature* **1992**, *358*, 646−653.

(11) Waksman, G.; Shoelson, S. E.; Pant, N.; Cowburn, D.; Kuriyan, J. Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell* **1993**, *72*, 779−790.

(12) Payne, G.; Stolz, L. A.; Pei, D.; Band, H.; Shoelson, S. E.; Walsh, C. T. The phosphopeptide-binding specificity of Src family SH2 domains. *Chem. Biol.* **1994**, *1*, 99−105.

(13) Ladbury, J. E.; Arold, S. T. Energetics of Src homology domain interactions in receptor tyrosine kinase-mediated signaling. *Methods Enzymol.* **2011**, *488*, 147−183.

(14) Neel, B. G.; Gu, H.; Pao, L. The "Shp"ing news. SH2 domain-containing tyrosine phosphatases in cell signaling. *Trends Biochem. Sci.* **2003**, *28*, 284−293.

(15) Hof, P.; Pluskey, S.; Dhe-Paganon, S.; Eck, M. J.; Shoelson, S. E. Crystal structure of the tyrosine phosphatase SHP-2. *Cell* **1998**, *92*, 441−450.

(16) Cunnick, J. M.; Dorsey, J. F.; Munoz-Antonia, T.; Mei, L.; Wu, J. Requirement of SHP2 binding to Grb2-associated binder-1 for mitogen- activated protein kinase activation in response to lysophosphatidic acid and epidermal growth factor. *J. Biol. Chem.* **2000**, *275*, 13842−13848.

(17) Wildes, D.; Anderson, L. M.; Sabogal, A.; Marqusee, S. Native state energetics of the Src SH2 domain: evidence for a partially structured state in the denatured ensemble. *Protein Sci.* **2006**, *15*, 1769−1779.

(18) Stryer, L. The interaction of a naphthalene dye with apomyoglobin and apohemoglobin. A fluorescent probe of non-polar binding sites. *J. Mol. Biol.* **1965**, *13*, 482−495.

(19) Zhou, S.; Shoelson, S. E.; Chaudhuri, M.; Gish, G.; Pawson, T.; Haser, W. G.; King, F.; Roberts, T.; Ratnofsky, S.; Lechleider, R. J.; et al. SH2 domains recognize specific phosphopeptide sequences. *Cell* **1993**, *72*, 767−778.

DOI: 10.1021/acs.jpcb.8b05651
J. Phys. Chem. B XXXX, XXX, XXX−XXX

(20) Jackson, S. E.; Fersht, A. R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **1991**, *30*, 10428−10435.
(21) Myers, J. K.; Pace, C. N.; Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995**, *4*, 2138−2148.
(22) Matouschek, A.; Kellis, J. T., Jr; Serrano, L.; Bycroft, M.; Fersht, A. R. Transient folding intermediates characterized by protein engineering. *Nature* **1990**, *346*, 440−445.
(23) Khorasanizadeh, S.; Peters, I. D.; Roder, H. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat. Struct. Biol.* **1996**, *3*, 193−205.
(24) Parker, M. J.; Spencer, J.; Clarke, A. R. An integrated kinetic analysis of intermediates and transition states in protein folding reactions. *J. Mol. Biol.* **1995**, *253*, 771−786.
(25) Capaldi, A. P.; Shastry, M. C.; Kleanthous, C.; Roder, H.; Radford, S. E. Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat. Struct. Biol.* **2001**, *8*, 68−72.
(26) Gianni, S.; Ivarsson, Y.; Jemth, P.; Brunori, M.; Travaglini-Allocatelli, C. Identification and characterization of protein folding intermediates. *Biophys. Chem.* **2007**, *128*, 105−113.
(27) Travaglini-Allocatelli, C.; Gianni, S.; Morea, V.; Tramontano, A.; Soulimane, T.; Brunori, M. Exploring the cytochrome c folding mechanism: cytochrome c552 from thermus thermophilus folds through an on-pathway intermediate. *J. Biol. Chem.* **2003**, *278*, 41136−41140.
(28) Simister, P. C.; Feller, S. M. Order and disorder in large multi-site docking proteins of the Gab family–implications for signalling complex formation and inhibitor design strategies. *Mol. BioSyst.* **2012**, *8*, 33−46.
(29) Wheadon, H.; Paling, N. R.; Welham, M. J. Molecular interactions of SHP1 and SHP2 in IL-3-signalling. *Cell. Signalling* **2002**, *14*, 219−229.
(30) Daggett, V.; Fersht, A. R. Is there a unifying mechanism for protein folding? *Trends Biochem. Sci.* **2003**, *28*, 18−25.
(31) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.* **2002**, *9*, 209−216.
(32) Ferguson, N.; Capaldi, A. P.; James, R.; Kleanthous, C.; Radford, S. E. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* **1999**, *286*, 1597−1608.
(33) Felder, S.; Zhou, M.; Hu, P.; Ureña, J.; Ullrich, A.; Chaudhuri, M.; White, M.; Shoelson, S. E.; Schlessinger, J. SH2 domains exhibit high-affinity binding to tyrosine-phosphorylated peptides yet also exhibit rapid dissociation and exchange. *Mol. Cell. Biol.* **1993**, *13*, 1449−1455.
(34) Chi, C. N.; Bach, A.; Strømgaard, K.; Gianni, S.; Jemth, P. Ligand binding by PDZ domains. *Biofactors* **2012**, *38*, 338−348.
(35) Chi, C. N.; Gianni, S.; Calosci, N.; Travaglini-Allocatelli, C.; Engstrom, Å.; Jemth, P. A conserved folding mechanism for PDZ domains. *FEBS Lett.* **2007**, *581*, 1109−1113.
(36) Gianni, S.; Engström, A.; Larsson, M.; Calosci, N.; Malatesta, F.; Eklund, L.; Ngang, C. C.; Travaglini-Allocatelli, C.; Jemth, P. The kinetics of PDZ domain-ligand interactions and implications for the binding mechanism. *J. Biol. Chem.* **2005**, *280*, 34805−34812.

DOI: 10.1021/acs.jpcb.8b05651
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Paper 3: SH3 domains are small protein modules mediating protein-protein interactions recognizing motifs rich in proline in the target protein. These domains are involved in the regulation of important cellular pathways and usually present two three-stranded antiparallel β-sheets orthogonally packed onto each other, to form a single hydrophobic core.

The SH3 domain from Grb2 (Grb2-SH3) presents the typical structure of an SH3 domain and mediates the interaction between Grb2 and Gab2, a scaffolding disordered protein, triggering a number of key metabolic pathways involved in cell death and differentiation. In the following paper we performed the Φ-value analysis of the folding pathway of Grb2-SH3. This approach, combined molecular dynamic simulations, allowed us to assess the mechanism of folding of this domain and the structure of the transition state. Data suggest that Grb2-SH3 folds via a nucleation-condensation mechanism (i.e. with the concurrent formation of native-like secondary and tertiary structure) presenting a native-like, diffused transition state and without the accumulation of folding intermediates. The comparison between our data and previous folding studies on SH3 domains belonging to other proteins highlights that the proteins of this class may fold via alternative pathways, stabilized by different nuclei leading or not to accumulation of folding intermediates. This comparative analysis suggests that the alternative folding pathways for this class of SH3 domains can be selectively regulated by the specific amino acid sequences.

THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B XXXX, XXX, XXX−XXX

pubs.acs.org/JPCB

Folding Mechanism of the SH3 Domain from Grb2

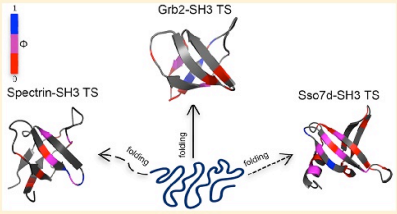
Francesca Troilo,[†] Daniela Bonetti,[†] Carlo Camilloni,[‡] Angelo Toto,[†] Sonia Longhi,[§] Maurizio Brunori,*[,†] and Stefano Gianni*[,†]

[†]Istituto Pasteur—Fondazione Cenci Bolognetti, Dipartimento di Scienze Biochimiche "A. Rossi Fanelli" and Istituto di Biologia e Patologia Molecolari del CNR, Sapienza Università di Roma, 00185 Rome, Italy

[‡]Dipartimento di Bioscienze, Università degli studi di Milano, 20133 Milan, Italy

[§]Aix-Marseille Univ, CNRS, Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR, 7257 Marseille, France

ABSTRACT: SH3 domains are small protein modules involved in the regulation of important cellular pathways. These domains mediate protein−protein interactions recognizing motifs rich in proline on the target protein. The SH3 domain from Grb2 (Grb2-SH3) presents the typical structure of an SH3 domain composed of two three-stranded antiparallel β-sheets orthogonally packed onto each other, to form a single hydrophobic core. Grb2 interacts, via SH3 domain, with Gab2, a scaffolding disordered protein, triggering some key metabolic pathways involved in cell death and differentiation. In this work we report a mutational analysis (Φ value analysis) of the folding pathway of Grb2-SH3 that, coupled with molecular dynamic simulations, allows us to assess the structure of the transition state and the mechanism of folding of this domain. Data suggest that Grb2-SH3 folds via a native-like, diffused transition state with a concurrent formation of native-like secondary and tertiary structure (nucleation−condensation mechanism) and without the accumulation of folding intermediates. The comparison between our data and previous folding studies on SH3 domains belonging to other proteins highlights that proteins of this class may fold via alternative pathways, stabilized by different nuclei leading or not to accumulation of folding intermediates. This comparative analysis suggests that the alternative folding pathways for this class of SH3 domains can be selectively regulated by the specific amino acid sequences.



INTRODUCTION

One of the most informative approaches to address the folding mechanism of globular proteins is to compare experiments performed on homologous proteins.[1−6] In fact, by describing the folding of proteins sharing the same topology while displaying a different sequence, it is theoretically possible to draw some general rules on the basic principles governing folding. Comparative folding studies have been previously reported for example on the colicin immunity proteins Im7 and Im9,[2,4,7,8] on the immunoglobulin domains,[3] on c-type cytochromes,[9−11] on homeodomain-like proteins,[5,12] on PDZ domains,[1,13−15] and on others. While all these studies suggest that the overall general features of folding are by and large defined by protein topology, it appears that a closer look at the folding pathway of the different homologues appears to highlight some features specific for each globular protein.

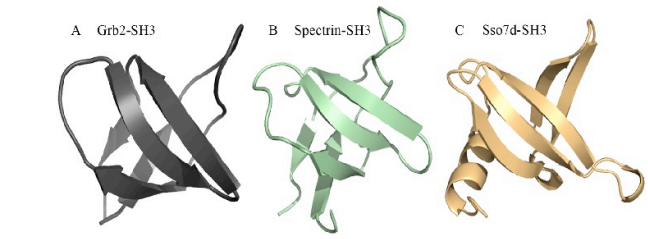
In the context of comparative folding studies, the SH3 domain represents a debated system. In fact, while an earlier comparison between the src and the spectrin SH3 domains suggested this class of proteins folds via a robust two-state mechanism characterized by a polarized and highly conserved transition state,[16−18] studies on the Sso7d domain revealed an additional complexity.[19] Indeed, while displaying less than 10% sequence identity, the Sso7d protein shares the typical topology of SH3 domains, except for the last β-strand that is a small α-helix in Sso7d. Interestingly, an experimental and computational comparison of the folding of Sso7d with SH3 domains revealed a substantial shift in the folding nucleus from the third to the second β-hairpin.[19] This finding highlighted that, even if protein topology plays a major role in the selection of the folding pathways, the specific nature of the interactions stabilizing the protein is still critical to describe folding mechanisms. Furthermore, it is of interest to note that recent studies highlighted how the folding of SH3 may also occur via a multistate scenario, with accumulation of intermediates characterized at equilibrium.[20−22] It appears therefore that, even for a deeply investigated protein system, such as the SH3 domain, folding demands a careful study to be fully understood.

The SH3 domain from Grb2 (Grb2-SH3) corresponds to the typical structure of an SH3 domain composed of two three-

Special Issue: William A. Eaton Festschrift

Received: July 3, 2018
Revised: August 8, 2018
Published: August 9, 2018

DOI: 10.1021/acs.jpcb.8b06320
J. Phys. Chem. B XXXX, XXX, XXX−XXX




```
Grb2-SH3 !---GPTYVQALFDFDPQEDGELGFRRGDFIHVMDNSDPNWWKGACHGQTGMFPRNYVTAVNIQPPPVNRNLKPDRK !
Spectrin-SH3 !MDBTGKELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGYVPAAYVKKLD--------------- !
     Identity: 19.5% !

Grb2-SH3 !PTYVQALFDF-DPQEDGELG-------PRRGDFI-HVMDNSDPNWWKGACHGQTGMFPRNYVTAVNIQPPPVNRNLKPDRK !
Sso7d-SH3 !----MATVKFKYKGEEKEVDISKIKKVWRVGKMISFTYDEGG---------GKTG---RGAVSEKDAPKELLQMLEKQKK !
     Identity: 18.8% !
```

Figure 1. Cartoon rapresentation of Grb2-SH3 (A), spectrin SH3 (B), and Sso7d-SH3 (C) structures and sequences alignments. As discussed in the text, Grb2-SH3 displays a comparable sequence identity to both Sso7d (18.8%) and spectrin SH3 (19.5%).

stranded antiparallel β-sheets orthogonally packed onto each other, to form a single hydrophobic core.[23,24] Physiologically, the domain is involved in binding a proline rich stretch of amino acids of Gab2 (encompassing residues 503−524), with this interaction triggering some key metabolic pathways involved in cell death and differentiation. From the perspective of its primary structure, it is interesting to note that Grb2-SH3 displays a similar degree of sequence identity toward Sso7d and spectrin SH3 (Figure 1), posing this system as an interesting candidate to understand further the folding mechanism of this highly studied class of proteins.

Here we present the characterization of the folding of the Grb2-SH3 domain. By carrying out kinetic experiments on 23 site-directed variants in combination with restrained molecular dynamics simulations, we present the structure of the main folding transition state. The transition state is stabilized by contacts involving both the first β-hairpin and the N- and C-termini of the protein, a finding which appears different from what was previously observed for src, spectrin, and fyn SH3.[16,18,25,26] Furthermore, we present evidence that this protein folds via a nucleation−condensation mechanism, with a diffused, rather than structurally polarized, transition state. The data are discussed in the context of previous work on other SH3 domains.

METHODS

Site-Directed Mutagenesis. Grb2-SH3 was subcloned in a pET28b+ plasmid vector. The constructs encoding the site-directed variants of SH3 were obtained using the gene encoding Grb2-SH3 WT as a template to perform site-directed mutagenesis using the QuickChange Lightning Site-Directed Mutagenesis kit (Agilent technologies) according to the manufacturer's instructions. All mutations are conservative. All mutations were confirmed by DNA sequencing.

Protein Expression and Purification. The Grb2-SH3 and all the site-directed variants were expressed in *E. coli* cells BL21 (DE3). Bacterial cells were grown in LB medium, containing 30 μg/mL of kanamycin, at 37 °C until OD_{600} = 0.7−0.8, and then protein expression was induced with 1 mM IPTG. After induction, cells were grown at 37 °C overnight and then collected by centrifugation.

To purify the protein, the bacterial pellet was resuspended in a buffer of 50 mM TrisHCl buffer, 0.5 M NaCl, pH 7.5, with the addition of antiprotease tablet (Complete EDTA-free, Roche); then sonicated; and centrifuged. The soluble fraction from bacterial lysate was loaded onto a nickel-charged HisTrap Chelating HP (GE Healthcare) column equilibrated with 50 mM TrisHCl, 0.5 M NaCl, pH 7.5. The protein was then eluted with a gradient from 0 to 1 M imidazole by using an AKTA-prime system. Fractions containing the protein were collected, and the buffer was exchanged to 25 mM Hepes pH 7.5 100 mM potassium acetate by using a HiTrap Desalting column (GE Healthcare). The purity of the protein was analyzed through SDS-page.

Protein concentration was estimated by measuring the absorbance of tryptophan residue at 280 nm and calculated through the Lambert−Beer equation.

Equilibrium Experiments. Equilibrium unfolding experiments were performed on a Fluoromax single photon counting spectrofluorometer (Jobin-Yvon, NJ). Grb2-SH3 and all the site-directed variants, at a constant concentration of 3 μM, was excited at 280 nm, and emission spectra were recorded between 300 and 400 nm, at increasing denaturant (urea) concentration. Experiments were performed at 25 °C, using a quartz cuvette with a path length of 1 cm, in 50 mM sodium phosphate buffer at pH 7.2. The fluorescence emission was then analyzed quantitatively following a two-state model. In all cases, the thermodynamic parameters calculated from curve fitting were found to be essentially insensitive to the wavelength considered, confirming the two-state nature of the reaction.

Stopped-Flow Folding Experiments. Unfolding and refolding kinetics experiments were carried out on a single-mixing SX-18 stopped-flow instrument (Applied Photophysics), monitoring the change of fluorescence emission. The experiments were performed at 25 °C in 50 mM sodium phosphate buffer at pH 7.2, by using urea as the denaturant. The excitation wavelength used was 280 nm, and the

DOI: 10.1021/acs.jpcb.8b06320
J. Phys. Chem. B XXXX, XXX, XXX−XXX

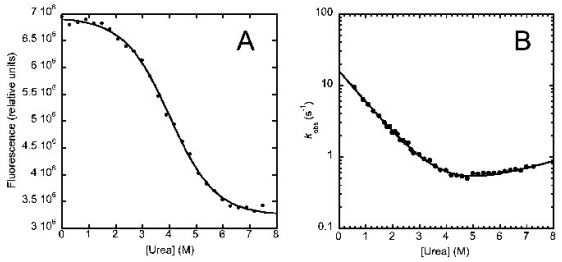


Figure 2. Folding of Grb2-SH3. (A) Equilibrium denaturation experiment of the Grb2-SH3 domain carried out in a buffer of 50 mM sodium phosphate at pH 7.2 and 25 °C. The change of the intrinsic fluorescence of the tryptophan residue versus urea concentrations was fitted with a two-state equation (see text for details). The thermodynamic parameters calculated from curve fitting were found to be essentially insensitive to the wavelength considered, confirming the two-state nature of the reaction. (B) Chevron plot of the Grb2-SH3 domain obtained in a buffer of 50 mM sodium phosphate at pH 7.2 and 25 °C. The line is the best fit to a two-state mechanism.

Table 1. Kinetic Folding Parameters of Grb2-SH3 and Its Site-Directed Variants[a]

mutant	k_F (s^{-1})	k_U (s^{-1})	m_F (kcal/M mol)	m_U (kcal/M mol)	m_{D-N} (kcal/M mol)	$[urea]_{1/2}$ (M)	ϕ
WT	16.0 ± 1.4	0.14 ± 0.02	0.59 ± 0.02	0.14 ± 0.06	0.73 ± 0.01	4.2 ± 1.0	
T1S	13.0 ± 1.3	0.21 ± 0.02	0.59 ± 0.06	0.12 ± 0.12	0.71 ± 0.10	3.1 ± 0.2	0.28 ± 0.26
Y2A	8.6 ± 0.9	0.69 ± 0.07	0.59 ± 0.10	0.07 ± 0.10	0.73 ± 0.04	1.8 ± 0.1	0.26 ± 0.07
V3A	4.6 ± 0.6	0.33 ± 0.05	0.75 ± 0.07	0.08 ± 0.09	0.83 ± 0.06	2.5 ± 0.3	0.58 ± 0.11
A5G	1.7 ± 0.7	0.36 ± 0.04	0.63 ± 0.05	0.12 ± 0.08	0.73 ± 0.06	0.2 ± 0.1	0.70 ± 0.17
L6A	8.0 ± 0.6	0.67 ± 0.05	0.68 ± 0.04	0.13 ± 0.06	0.81 ± 0.04	2.3 ± 1.1	0.30 ± 0.06
F7A	6.4 ± 0.64	0.21 ± 0.02	0.66 ± 0.07	0.11 ± 0.10	0.77 ± 0.07	3.2 ± 0.1	0.68 ± 0.15
F19A	2.6 ± 0.3	1.23 ± 0.05	0.57 ± 0.09	0.07 ± 0.13	0.73 ± 0.09	1.6 ± 0.2	0.45 ± 0.04
F24A	3.7 ± 0.3	0.70 ± 0.03	0.70 ± 0.04	0.12 ± 0.06	0.81 ± 0.04	1.8 ± 0.2	0.47 ± 0.05
I25V	6.1 ± 0.5	0.64 ± 0.03	0.70 ± 0.04	0.08 ± 0.06	0.78 ± 0.04	2.4 ± 1.7	0.38 ± 0.06
H26A	33.0 ± 8.0	0.23 ± 0.05	0.62 ± 0.05	0.15 ± 0.06	0.77 ± 0.03	3.6 ± 1.5	*b*
S31A	11.0 ± 1.0	0.42 ± 0.04	0.67 ± 0.03	0.08 ± 0.04	0.75 ± 0.03	2.8 ± 0.4	0.24 ± 0.10
A39G	16.0 ± 2.0	0.92 ± 0.07	0.71 ± 0.04	0.06 ± 0.06	0.77 ± 0.04	2.4 ± 1.3	−0.03 ± 0.07
H41A	13.0 ± 1.3	0.56 ± 0.06	0.65 ± 0.07	0.07 ± 0.07	0.72 ± 0.02	3.0 ± 0.1	0.11 ± 0.09
T44S	14.0 ± 1.0	0.53 ± 0.05	0.69 ± 0.07	0.05 ± 0.07	0.74 ± 0.02	2.8 ± 0.1	0.07 ± 0.09
Y51A	11.0 ± 0.5	0.43 ± 0.02	0.57 ± 0.02	0.14 ± 0.02	0.71 ± 0.01	2.6 ± 0.4	0.23 ± 0.07
T53S	11.0 ± 3.0	0.23 ± 0.05	0.53 ± 0.06	0.12 ± 0.09	0.65 ± 0.07	2.7 ± 0.3	0.41 ± 0.37
A54G	16.2 ± 1.6	0.62 ± 0.06	0.55 ± 0.03	0.12 ± 0.04	0.67 ± 0.02	2.4 ± 1.1	−0.05 ± 0.09
V55A	15.2 ± 1.5	0.45 ± 0.04	0.47 ± 0.05	0.23 ± 0.09	0.70 ± 0.07	3.4 ± 0.3	−0.01 ± 0.11

[a]The mutants F9A, L17A, V27A, F47A, and V52A expressed poorly and could not be characterized. [b]This mutant shows $\Delta\Delta G_{D-N}$ < 0.4 kcal mol^{-1}, preventing reliable calculation of the Φ value.[36]

fluorescence emission light was recorded by using a 320 nm cutoff glass filter. For each denaturant concentration, usually 5 individual traces were averaged. The final concentration of Grb2-SH3 and all the variants was typically 1 μM. In all cases the fluorescence time courses obtained was satisfactorily fitted by using a single exponential equation.

Molecular Dynamics Simulations. Molecular dynamics simulations of SH3 were performed using the CHARMM22* force field[27] with the TIP3P water model.[28] All the simulations were run using GROMACS[29] and PLUMED2.[30] A time step of 2 fs was used together with LINCS constraints.[31] van der Waals and Coulomb interactions were implemented with a cutoff at 0.9 nm, and long-range electrostatic effects were treated with the particle mesh Ewald method on a grid with a mesh of 0.1 nm.

A standard 200 ns molecular dynamics simulation at 300 K was performed as a reference for the native state ensemble. The starting conformation was taken from an available X-ray structure (PDB code 2VWF[23]) and solvated with 4531 water molecules and 4 sodium ions.

The transition state ensemble was determined following a standard procedure based on the interpretation of Φ value analysis in terms of fraction of native contacts. Briefly, given a set of experimental Φ values, a pseudoenergy term has been added to the force field as the squared difference between experimental and simulated Φ values in order to maximize the agreement with the experimental value while keeping the simulation stable. Given two residues that are not nearest neighbors, the native contacts between them are defined as the number of heavy side-chain atoms located within 0.65 nm in the native structure. The Φ value for a residue *i* is calculated from the fraction of native contacts that it makes in a given conformation. With this approach only Φ values between 0 and 1 can be incorporated as structural restraints.

The transition state ensemble was generated using 1000 cycles of simulated annealing. Each cycle is 200 ps long, in

DOI: 10.1021/acs.jpcb.8b06320
J. Phys. Chem. B XXXX, XXX, XXX−XXX

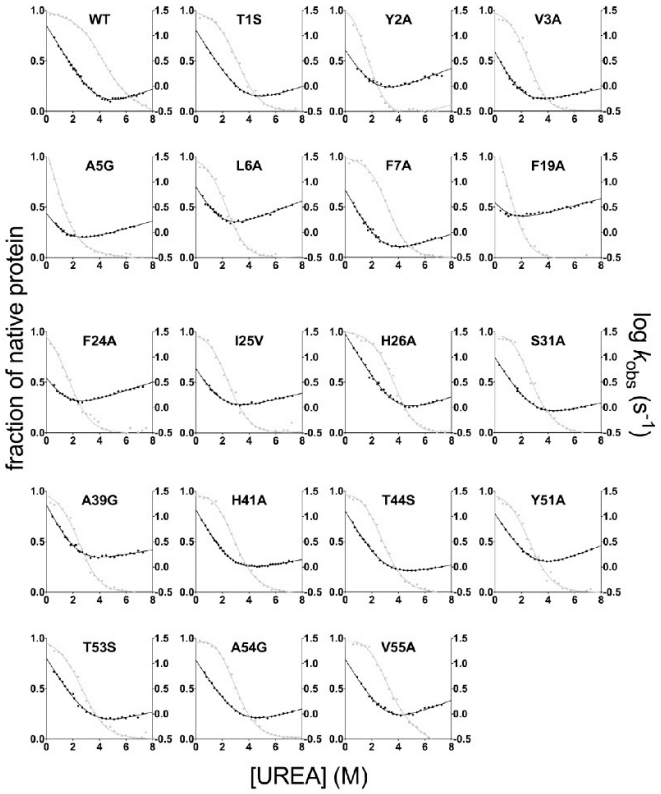


Figure 3. Equilibrium denaturations and chevron plots of Grb2-SH3 and its site-directed mutants. All experiments were carried out at 25 °C and in 50 mM sodium phosphate buffer at pH 7.2. Each mutant was globally fitted to a two-state mechanism by assuming the m_{D-N} value at equilibrium to be equivalent to the sum between the kinetic m_F and m_U values. In all cases, data were consistent with a two-state scenario,[32] indicating the absence of transient folding intermediates.

which the temperature is varied between 300 and 400 K. Only the structures sampled at the reference temperature are retained for further analysis, resulting in TSE of ∼6000 conformations.

■ RESULTS

Kinetic Folding Mechanism of Wild Type Grb2-SH3. In order to characterize the folding mechanism of Grb2-SH3 we initially conducted experiments on the wild type protein. Urea-induced equilibrium denaturation of Grb2-SH3 measured at 25 °C, pH 7.2, in 50 mM sodium phosphate buffer by a decrease in Trp emission is reported in Figure 2. The observed transition is consistent with a simple two-state behavior, suggesting the absence of stable equilibrium intermediate(s).[32]

The unfolding free energy in water derived from two-state analysis is 3.1 ± 0.4 kcal mol^{-1} displaying an m_{D-N} value of 0.73 ± 0.01 kcal mol^{-1} M^{-1}. This value, which is proportional to the change in accessible surface area upon unfolding, is consistent with what is expected from a protein of 56 amino acids.[33]

The folding and unfolding kinetics of Grb2-SH3 were measured by stopped-flow fluorimetry. As expected for a two-state folder, under all investigated conditions, folding and unfolding time courses were consistent with a single exponential decay. Furthermore, in analogy to what was previously observed on other SH3 domains, the dependence of the observed rate constant (k_{obs}) on urea concentration

DOI: 10.1021/acs.jpcb.8b06320
J. Phys. Chem. B XXXX, XXX, XXX−XXX

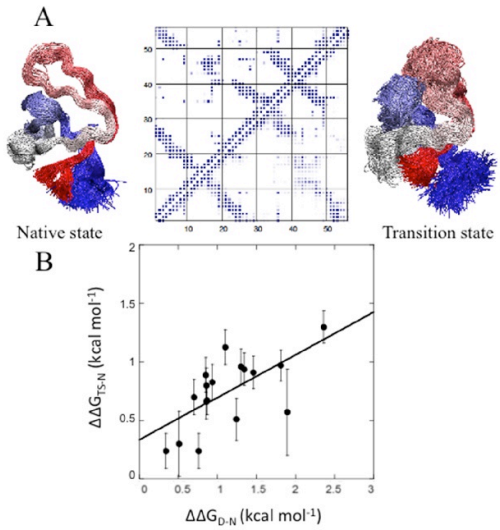


Figure 4. (A) Structure of the folding transition state of Grb2-SH3, together with the associated contact map. The top left of the contact map refers to the contacts between amino acids in the native state, whereas the bottom right refers to the contacts in the transition state. As explained in the text, the protein seems to fold via a native-like transition state characterized by the formation of the first β-hairpin, together with a consolidation of the interaction between the N- and C-termini of the protein. (B) Bronsted plot. As explained in the Results section, the linearity of the Bronsted plot suggests that this protein fold via a native-like diffused, transition state. This finding appears to be consistent with a nucleation−condensation mechanism.[45,46]

conforms to a V-shaped chevron plot (Figure 2), a typical signature of two-state folding.[32]

Since Grb2-SH3 unfolding displays a low cooperativity, with an m_{D-N} of 0.73 ± 0.01 kcal mol^{-1} M^{-1}, an accurate determination of the folding parameters from each independent experiment is complicated. Therefore, to decrease the fitting error and, at the same time, to test the robustness of two-state folding of Grb2-SH3, equilibrium and kinetic experiments were fitted globally to the following equations:

equilibrium

$$Y_{obs} = Y_N + Y_D \frac{e^{(m_{D-N}([urea]-[urea]_{1/2})}}{1 + e^{(m_{D-N}([urea]-[urea]_{1/2})}}$$

kinetics

$$K_{obs} = k_F e^{(-m_F[urea])} + k_U e^{(-m_U[urea])} \qquad m_{D-N} = m_F + m_U$$

with shared m_{D-N} values. The fitting parameters calculated from the global analysis are reported in Table 1.

Structure of the Folding Transition State of Grb2-SH3. In order to characterize the transition state of folding of Grb2-SH3, we carried out a Φ value analysis,[34,35] by producing 23 site-directed variants. The Φ value is then calculated by dividing the effect of the substitution on the activation free energy by that of the stability of the native structure. The conservative variants were designed, and the analysis was carried out using the standard rules of Φ value analysis, as formalized previously.[36]

Unfolding and folding of all the variants were measured both at equilibrium, by urea-induced denaturation, and by kinetics, using stopped-flow fluorimetry. In all cases, in analogy to what was observed for wild type Grb2-SH3, folding and unfolding kinetics were consistent with a single exponential decay. Figure 3 shows the equilibrium and kinetic experiments carried out on each site-directed variant. In all cases, data were consistent with a two-state scenario, indicating that Grb2-SH3 folds via a robust mechanism, which does not involve any transient folding intermediates.

To determine the structure of the folding transition state of Grb2-SH3, we used the experimentally measured Φ values as restrains in molecular dynamics simulations. This method, which has been previously used and validated on several different protein systems,[1,37−42] is based on the incorporation of the Φ values as biases on the fraction of formed native contacts in a molecular dynamics simulation trajectory (cf. Methods section).

The structure of the folding transition state of Grb2-SH3, together with the associated contact map, is reported in Figure 4A. It is evident that the protein seems to fold via a native-like transition state that is characterized by the formation of the first β-hairpin, together with a consolidation of the interaction between the N- and C-termini of the protein. The structure gradually tapers off, with the region encompassing the β2−β3

DOI: 10.1021/acs.jpcb.8b06320
J. Phys. Chem. B XXXX, XXX, XXX−XXX

interaction being the most disordered of the ensemble. The structural features of the transition state of folding of Grb2-SH3, in comparison to those previously depicted for other SH3 domains, are analyzed in the Discussion section.

On the basis of the Φ value analysis of src and spectrin SH3,[16,18] it has been previously suggested that the structure of the transition state of SH3 domains is highly polarized. To test this hypothesis for Grb2-SH3, we analyzed the Bronsted plot of this protein.[43] In fact, while a diffused native-like structure is expected to return linear Bronsted plots, a polarized transition state is more likely to yield a scatter in the Bronsted plot, with only some positions playing a key role in stabilizing its structure (characterized by high Φ value), with the others displaying low values of Φ.[44] As evident from Figure 4B, the transition state of Grb2-SH3 clearly displays a linear Bronsted plot, suggesting this protein folds via a native-like diffused, rather than polarized, transition state. This finding appears to be consistent with a nucleation−condensation mechanism[45,46] for this SH3, in agreement with what proposed earlier by Shakhnovich and co-workers.[26]

Robustness of Two-State Folding in Grb2-SH3. It has been proposed that some SH3 domains may retain some residual structure in their denatured state[47,48] and/or populate folding intermediates.[20−22] In order to test the robustness of the two-state folding in Grb2-SH3, we resorted to analyzing the dependence of the folding parameters as a function of protein stability. In fact, comparing the parameters measured on different site-directed variants represents an efficient test to address the overall folding characteristics of transition and denatured state.[49,50] More specifically, since the activation and ground state free energies depend on the denaturant concentration (measured by the m_U, m_F, and m_{D-N} values) and on the changes in accessible surface area between the pertinent state,[33] an analysis of their dependence may reveal signatures of shifts of the transition and denatured states along the reaction coordinate, as well as the accumulation of folding intermediates. Figure 5 depicts the correlation between the m_{D-N}, m_U, and m_F and the $\Delta\Delta G_{D-N}$ for the different site-directed variants. It is evident that, in the case of Grb2-SH3, no detectable change in m_U, m_F, and m_{D-N} values could be observed for the different variants, spanning a change in protein stability of about 3 kcal mol^{-1}. This observation suggests that, contrary to what observed in the case of fyn and PI3K SH3 domain,[20−22] the folding mechanism of this protein is robust and consistent with two-state.

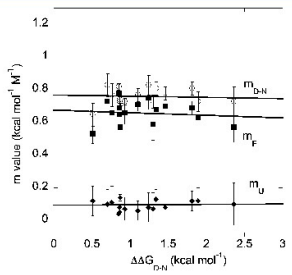


Figure 5. Correlation between the m_{D-N} (○), m_U (◆), and m_F (■) and the $\Delta\Delta G_{D-N}$ for the different site-directed variants. As discussed in the text, no detectable dependence of m_U, m_F, and m_{D-N} values can be observed for the different variants suggesting that the folding mechanism of this protein is robust and consistent with two-state.

DISCUSSION

The first comparative Φ value analysis on globular proteins was presented in two papers describing the folding of src and spectrin SH3.[16,17] These studies suggested that this protein family folds via a conserved mechanism characterized by a structurally robust transition state. Furthermore, it was pointed out that the structure of the transition state was primarily stabilized by interactions taking place in the third β-hairpin of the protein, representing a polarized folding nucleus. Subsequently, also a Φ value analysis of fyn SH3 domain was reported, further supporting the robustness of the structure of the transition state[25] showing that even drastic nonconservative mutations caused little structural rearrangements of the transition state.[51] A breakdown of such robustness could be observed in SSo7d, a protein sharing a similar topology with the other SH3 domains while displaying negligible sequence homology. In fact, in this case, a shift in the transition state nucleus from the third to the second β-hairpin was reported.[19]

In the context of previous work on SH3 domains, it is therefore interesting to note how the structure of the transition state of folding of Grb2-SH3 is different from that of src, spectrin, and fyn SH3. In fact, Grb2-SH3 displays an extended folding nucleus, which involves the β-sheet comprising the N- and C-termini of the protein together with the first β-hairpin. Since the structural architecture of the folding nucleus of Grb2-SH3 appears to be distinct from that of Sso7d, it appears that this protein family may fold through a multitude of mechanisms comprising distinct regions of the protein. Such pathways may then be selectively stabilized over others by the amino acid sequence, indicating that while the overall features of folding are defined by protein topology, the nature of the interactions stabilizing the native state is still critical to influence protein folding mechanisms. In this context, alternative pathways may emerge when the sequence is changed extensively.

A number of studies have shown that proteins may fold with or without folding intermediates, depending on solvent conditions and changes in sequence composition.[2,7,8,12,15,38] Accordingly, while the folding of SH3 domains has been classically described with a two-state mechanism, Dokholyan and co-workers predicted,[52] by analyzing different molecular dynamics simulations, that this class of protein may populate stable intermediates as a consequence of the local stabilization of individual structural elements. This finding was later supported by NMR and by pulse hydrogen exchange mass spectrometry, which revealed the presence of at least one folding intermediate in the case of Fyn[20,21] and PI3K SH3,[22] respectively. In both cases, the stabilization of the intermediate appears to arise from the stabilization of non-native hydrophobic interactions, leading to a polarized structure formation upon folding. The analysis of the Bronsted plot of Grb2-SH3 suggests this protein folds via a transition state with a diffused native-like structure. In this case, therefore, the protein seems consistent with a nucleation−condensation mechanism, characterized by a concurrent formation of native secondary and tertiary structure.[45,46] In the light of this finding, it is not surprising to observe that, contrary to the SH3 domains of Fyn

DOI: 10.1021/acs.jpcb.8b06320
J. Phys. Chem. B XXXX, XXX, XXX−XXX

and PI3K, Grb2-SH3 seems to conform to two-state folding, even when challenged with different site-directed variants, as illustrated by the robustness of the measured m_F, m_U, and m_{D-N} values, which are essentially independent of protein stability.

Taken together, our analysis of the folding pathway of Grb2-SH3 supports a view whereby this class of proteins may fold via alternative pathways, stabilized by different nuclei, that can be selectively balanced by sequence composition. In agreement with previous finding on other protein systems, local stabilization of such alternative nuclei may lead to the accumulation of intermediates, switching from two-state to multistate folding.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: maurizio.brunori@uniroma1.it.
*E-mail: stefano.gianni@uniroma1.it

ORCID

Carlo Camilloni: 0000-0002-9923-8590
Stefano Gianni: 0000-0003-1653-1925

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Work was partly supported by grants from the Italian Ministero dell'Istruzione dell'Università e della Ricerca (Progetto di Interesse "Invecchiamento" to S.G.), Sapienza University of Rome (C26A155S48, B52F16003410005 and RP11715C34AEAC9B to S.G), and the Associazione Italiana per la Ricerca sul Cancro (Individual Grant MFAG 2016, 18701 to S.G.). F.T. is a recipient of a Ph.D. fellowship from the Italo-French University.

REFERENCES

(1) Calosci, N.; Chi, C. N.; Richter, B.; Camilloni, C.; Engstrom, A.; Eklund, L.; Travaglini-Allocatelli, C.; Gianni, S.; Vendruscolo, M.; Jemth, P. Comparison of successive transition states for folding reveals alternative early folding pathways of two homologous proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 19241−19246.

(2) Capaldi, A. P.; Shastry, M. C.; Kleanthous, C.; Roder, H.; Radford, S. E. Ultrarapid mixing experiments reveal that Im7 folds via an on-pathway intermediate. *Nat. Struct. Biol.* **2001**, *8*, 68−72.

(3) Clarke, J.; Cota, E.; Fowler, S. B.; Hamill, S. J. Folding studies of Ig-like beta-sandwich proteins suggest they share a common folding pathway. *Structure* **1999**, *7*, 1145−1153.

(4) Friel, C. T.; Capaldi, A. P.; Radford, S. E. Structural analysis of the rate-limiting transition states in the folding of Im7 and Im9: similarities and differences in the folding of homologous proteins. *J. Mol. Biol.* **2003**, *326*, 293−305.

(5) Gianni, S.; Guydosh, N. R.; Khan, F.; Caldas, T. D.; Mayor, U.; White, G. W.; DeMarco, M. L.; Daggett, V.; Fersht, A. R. Unifying features in protein-folding mechanisms. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 13286−13291.

(6) Zarrine-Afsar, A.; Larson, S. M.; Davidson, A. R. The family feud: do proteins with similar structures fold via the same pathway? *Curr. Opin. Struct. Biol.* **2005**, *15*, 42−49.

(7) Capaldi, A. P.; Kleanthous, C.; Radford, S. E. Im7 folding mechanism: misfolding on a path to the native state. *Nat. Struct. Biol.* **2002**, *9*, 209−216.

(8) Ferguson, N.; Capaldi, A. P.; James, R.; Kleanthous, C.; Radford, S. E. Rapid folding with and without populated intermediates in the homologous four-helix proteins Im7 and Im9. *J. Mol. Biol.* **1999**, *286*, 1597−1608.

(9) Gianni, S.; Travaglini-Allocatelli, C.; Cutruzzola, F.; Brunori, M.; Shastry, M. C.; Roder, H. Parallel pathways in cytochrome c(551) folding. *J. Mol. Biol.* **2003**, *330*, 1145−1152.

(10) Travaglini-Allocatelli, C.; Gianni, S.; Brunori, M. A common folding mechanism in the cytochrome c family. *Trends Biochem. Sci.* **2004**, *29*, 535−541.

(11) Travaglini-Allocatelli, C.; Gianni, S.; Morea, V.; Tramontano, A.; Soulimane, T.; Brunori, M. Exploring the cytochrome c folding mechanism: cytochrome c552 from thermus thermophilus folds through an on-pathway intermediate. *J. Biol. Chem.* **2003**, *278*, 41136−41140.

(12) White, G. W.; Gianni, S.; Grossmann, J. G.; Jemth, P.; Fersht, A. R.; Daggett, V. Simulation and experiment conspire to reveal cryptic intermediates and a slide from the nucleation-condensation to framework mechanism of folding. *J. Mol. Biol.* **2005**, *350*, 757−775.

(13) Chi, C. N.; Gianni, S.; Calosci, N.; Travaglini-Allocatelli, C.; Engstrom, Å.; Jemth, P. A conserved folding mechanism for PDZ domains. *FEBS Lett.* **2007**, *581*, 1109−1113.

(14) Gianni, S.; Calosci, N.; Aelen, J. M.; Vuister, G. W.; Brunori, M.; Travaglini-Allocatelli, C. Kinetic folding mechanism of PDZ2 from PTP-BL. *Protein Eng., Des. Sel.* **2005**, *18*, 389−395.

(15) Ivarsson, Y.; Travaglini-Allocatelli, C.; Jemth, P.; Malatesta, F.; Brunori, M.; Gianni, S. An on-pathway intermediate in the folding of a PDZ domain. *J. Biol. Chem.* **2007**, *282*, 8568−8572.

(16) Grantcharova, V. P.; Riddle, D. S.; Santiago, J. V.; Baker, D. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat. Struct. Mol. Biol.* **1998**, *5*, 714−720.

(17) Martinez, J. C.; Pisabarro, M. T.; Serrano, L. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nat. Struct. Mol. Biol.* **1998**, *5*, 721−729.

(18) Martínez, J. C.; Serrano, L. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* **1999**, *6*, 1010−1016.

(19) Guerois, R.; Serrano, L. The SH3-fold family: experimental evidence and prediction of variations in the folding pathways. *J. Mol. Biol.* **2000**, *304*, 967−982.

(20) Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A. A.; Davidson, A. R.; Dobson, C. M.; Kay, L. E. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **2004**, *430*, 586−590.

(21) Ollerenshaw, J. E.; Kaya, H.; Chan, H. S.; Kay, L. E. Sparsely populated folding intermediates of the Fyn SH3 domain: matching native-centric essential dynamics and experiment. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14748−14753.

(22) Dasgupta, A.; Udgaonkar, J. B. Four-state folding of a SH3 domain: salt-induced modulation of the stabilities of the intermediates and native state. *Biochemistry* **2012**, *51*, 4723−4734.

(23) Harkiolaki, M.; Tsirka, T.; Lewitzky, M.; Simister, P. C.; Joshi, D.; Bird, L. E.; Jones, E. Y.; O'Reilly, N.; Feller, S. M. Distinct Binding Modes of Two Epitopes in Gab2 that Interact with the Sh3C Domain of Grb2. *Structure* **2009**, *17*, 809−822.

(24) Toto, A.; Bonetti, D.; De Simone, A.; Gianni, S. Understanding the mechanism of binding between Gab2 and the C terminal SH3 domain from Grb2. *Oncotarget* **2017**, *8*, 82344−82351.

(25) Northey, J. G.; Di Nardo, A. A.; Davidson, A. R. Hydrophobic core packing in the SH3 domain folding transition state. *Nat. Struct. Biol.* **2002**, *9*, 126−130.

(26) Hubner, I. A.; Edmonds, K. A.; Shakhnovich, E. I. Nucleation and the transition state of the SH3 domain. *J. Mol. Biol.* **2005**, *349*, 424−434.

(27) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, 47−49.

(28) Jorgensen, W. L. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **1981**, *103*, 335−340.

(29) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular

DOI: 10.1021/acs.jpcb.8b06320
J. Phys. Chem. B XXXX, XXX, XXX−XXX

simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*−*2*, 19−25.
(30) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(31) Hess, B. P-lincs: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116−122.
(32) Jackson, S. E.; Fersht, A. R. Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry* **1991**, *30*, 10428−10435.
(33) Myers, J. K.; Pace, C. N.; Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995**, *4*, 2138−2148.
(34) Fersht, A. R.; Matouschek, A.; Serrano, L. The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J. Mol. Biol.* **1992**, *224*, 771−782.
(35) Matouschek, A.; Kellis, J. T., Jr.; Serrano, L.; Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **1989**, *340*, 122−126.
(36) Fersht, A. R.; Sato, S. Phi-value analysis and the nature of protein-folding transition states. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 7976−7981.
(37) Geierhaas, C. D.; Salvatella, X.; Clarke, J.; Vendruscolo, M. Characterisation of transition state structures for protein folding using 'high', 'medium' and 'low' {Phi}-values. *Protein Eng., Des. Sel.* **2008**, *21*, 215−222.
(38) Gianni, S.; Ivarsson, Y.; De Simone, A.; Travaglini-Allocatelli, C.; Brunori, M.; Vendruscolo, M. Structural characterization of a misfolded intermediate populated during the folding process of a PDZ domain. *Nat. Struct. Mol. Biol.* **2010**, *17*, 1431−1437.
(39) Vendruscolo, M.; Paci, E.; Dobson, C. M.; Karplus, M. Three key residues form a critical contact network in a protein folding transition state. *Nature* **2001**, *409*, 641−645.
(40) Camilloni, C.; Bonetti, D.; Morrone, A.; Giri, R.; Dobson, C. M.; Brunori, M.; Gianni, S.; Vendruscolo, M. Towards a structural biology of the hydrophobic effect in protein folding. *Sci. Rep.* **2016**, *6*, 28285.
(41) Gianni, S.; Camilloni, C.; Giri, R.; Toto, A.; Bonetti, D.; Morrone, A.; Sormanni, P.; Brunori, M.; Vendruscolo, M. Understanding the frustration arising from the competition between function, misfolding, and aggregation in a globular protein. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 14141−14146.
(42) Gsponer, J.; Hopearuoho, H.; Whittaker, S. B.; Spence, G. R.; Moore, G. R.; Paci, E.; Radford, S. E.; Vendruscolo, M. Determination of an ensemble of structures representing the intermediate state of the bacterial immunity protein Im7. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 99−104.
(43) Leffler, J. E. Parameters for the description of transition states. *Science* **1953**, *117*, 340−341.
(44) Fersht, A. R. Relationship of Leffler (Bronsted) alpha values and protein folding Phi values to position of transition-state structures on reaction coordinates. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14338−14342.
(45) Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. Specific nucleus as the transition state for protein folding: evidence from the lattice model. *Biochemistry* **1994**, *33*, 10026−10036.
(46) Fersht, A. R. Optimization of rates of protein folding: the nucleation-condensation mechanism and its implications. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 10869−10873.
(47) Crowhurst, K. A.; Tollinger, M.; Forman-Kay, J. D. Cooperative interactions and a non-native buried Trp in the unfolded state of an SH3 domain. *J. Mol. Biol.* **2002**, *322*, 163−178.
(48) Kortemme, T.; Kelly, M. J.; Kay, L. E.; Forman-Kay, J.; Serrano, L. Similarities between the spectrin SH3 domain denatured state and its folding transition state. *J. Mol. Biol.* **2000**, *297*, 1217−1229.
(49) Sanchez, I. E.; Kiefhaber, T. Hammond behavior versus ground state effects in protein folding: evidence for narrow free energy barriers and residual structure in unfolded states. *J. Mol. Biol.* **2003**, *327*, 867−884.
(50) Scaloni, F.; Gianni, S.; Federici, L.; Falini, B.; Brunori, M. Folding mechanism of the C-terminal domain of nucleophosmin: residual structure in the denatured state and its pathophysiological significance. *FASEB J.* **2009**, *23*, 2360−2365.
(51) Northey, J. G.; Maxwell, K. L.; Davidson, A. R. Protein folding kinetics beyond the phi value: using multiple amino acid substitutions to investigate the structure of the SH3 domain folding transition state. *J. Mol. Biol.* **2002**, *320*, 389−402.
(52) Borreguero, J. M.; Ding, F.; Buldyrev, S. V.; Stanley, H. E.; Dokholyan, N. V. Multiple folding pathways of the SH3 domain. *Biophys. J.* **2004**, *87*, 521.

DOI: 10.1021/acs.jpcb.8b06320
J. Phys. Chem. B XXXX, XXX, XXX−XXX

Review (IN PRESS).

Running title. Fuzzy N_{TAIL} complexes

Experimental characterization of fuzzy protein assemblies: interactions of paramyxoviral N_{TAIL} domains with their functional partners

Francesca Troilo[1,2], Christophe Bignon[1], Stefano Gianni[2], Monika Fuxreiter[3] and Sonia Longhi[1]*

[1]CNRS and Aix-Marseille Univ, Laboratoire Architecture et Fonction des Macromolecules Biologiques (AFMB), UMR 7257, Marseille, France
[2]Istituto Pasteur – Fondazione Cenci Bolognetti, Dipartimento di Scienze Biochimiche 'A. Rossi Fanelli' and Istituto di Biologia e Patologia Molecolari del Consiglio Nazionale delle Ricerche, Sapienza Università di Roma, Rome, Italy
[3]MTA-DE Laboratory of Protein Dynamics, Department of Biochemistry and Molecular Biology, University of Debrecen, Hungary

*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

In preparation for the MIE special issue on IDPs. Deadline : June the 1st 2018

ABSTRACT

In this chapter we detail various experimental approaches to characterize the fuzziness of complexes made of the C-terminal domain of the nucleoprotein (N_{TAIL}) from three representative paramyxoviruses and of the C-terminal X domain (XD) of the homologous phosphoprotein. We discuss the advantages, the limitations as well as the caveats of the various methods. We describe experimental data showing that paramyxoviral N_{TAIL}-XD complexes are characterized by a considerable amount of conformational heterogeneity. We also detail recent data that revealed that N_{TAIL} is highly malleable, i.e. it displays a partner-mediated polymorphism. All the results suggest that N_{TAIL} plasticity and fuzziness plays a role in the coordination and regulation of the N_{TAIL} interaction network so as to ensure efficient transcription and replication.

Key words. Fuzzy interactions, experimental assessment of fuzziness, impact of fuzziness on binding, SEC, SAXS, NMR, site-directed spin-labeling EPR spectroscopy, ESI-MS and IM-MS, protein complementation assays, split-GFP reassembly, mutagenesis, kinetics.

Abbreviations. MeV, measles virus; NiV, Nipah virus; HeV, Hendra virus; N, nucleoprotein; P, phosphoprotein, RdRp, RNA-dependent RNA polymerase; L, large protein; IGR, intergenic region, N_{CORE}, globular moiety of N; N_{TAIL}, C-terminal disordered region of N; EM, electron microscopy; NMR, nuclear magnetic resonance; PNT, P N-terminal region; PCT, P C-terminal region; XD, X domain of P; PMD, P multimerization domain; ESI-MS, electrospray ionization mass spectrometry; K_D, equilibrium dissociation constant; MoRE, molecular recognition element; IDR, intrinsically disordered region; IDP, intrinsically disordered protein; RDC, residual dipolar couplings; SDSL, site-directed spin labeling; EPR, electron paramagnetic resonance; ITC, isothermal titration calorimetry; SAXS, small-angle X-ray scattering; SEC, size-exclusion chromatography; R_S, Stokes radius; Da, Dalton; MM, molecular mass; SDS-PAGE, sodium dodecyl sulfate polyacrylamide electrophoresis; SEC-MALLS, SEC coupled to

multiple angle laser light scattering; FPLC, fast protein liquid chromatography; HPLC, high-pressure liquid chromatography; Rg, radius of gyration; HSQC, Heterogeneous Single Quantum Correlation; MTSL, methanethiosulfonate derivative; τ_C, correlation time; CD, circular dichroism; CSD, charge state distribution; SASA, solvent-accessible surface area; IM, ion mobility; CCS, collision cross-section; hsp70; 70 kDa heat shock protein; GFP, green fluorescent protein; NGFP, N-terminal moiety of GFP; CGFP, C-terminal moiety of GFP; artN_{TAIL}, artificial N_{TAIL} sequence; PCR, polymerase chain reaction; OD, optical density; IPTG, isopropyl β-D-1-thiogalactopyranoside; k_{obs}, observed rate constant; k_{on}^{app}, apparent association rate constant; k_{off}^{app}, apparent dissociation rate constant.

1. INTRODUCTION

1.1. Fuzziness in molecular interactions

In contrast to the long-standing view of coupled folding to binding, protein complexes might retain considerable amount of structural and interaction multiplicity or exhibit a rapid conformational exchange even in the presence of a functional partner (Tompa & Fuxreiter, 2008). This phenomenon, termed as fuzziness, is quite common not only in intrinsically disordered (ID) proteins, but also in structured proteins and has important biological outcomes (Fuxreiter, 2018; Miskei, Gregus, et al., 2017; Sharma, Raduly, Miskei, & Fuxreiter, 2015). Retaining a high degree of conformational flexibility upon assembly decreases the entropic penalty of binding as compared to a disorder-to-order transition, leading to enhanced affinity or specificity. Fuzziness confers many further functional advantages, including the plasticity of recognition modes both with the same or alternative partners (Fuxreiter & Tompa, 2012). Fuzzy regions can mediate both specific as well as non-specific contacts, the latter having the potential to increase the local concentration nearby the partner. The long-range nature of fuzzy interactions manifests in allostery. As a result, amino acid substitutions in fuzzy regions located far away from the binding site and/or post-translational modifications have the potential to impact affinity/specificity of interactions or may exert other regulatory effects (for recent reviews on fuzziness (Duro, Miskei, & Fuxreiter, 2015; Fuxreiter, 2018; Miskei, Gregus, et al., 2017; Sharma et al., 2015)).

Fuzzy complexes are defined by two criteria (Fuxreiter, 2012; Tompa & Fuxreiter, 2008). First, the structural heterogeneity in the bound state of the protein must be characterized by structural techniques. Second, the functional impacts of perturbations (mutations, truncations, deletions) in conformationally heterogeneous regions must be demonstrated. In other words, structural multiplicity, or dynamical disorder directly contributes to the function of the complex. Here we use paramyxovirus N_{TAIL} complexes as an example to overview biophysical and structural approaches enabling the characterization of the fuzziness of protein complexes.

1.2. The *Paramyxoviridae* replicative complex, the N and P proteins and the N_{TAIL}-XD complex

The measles, Nipah and Hendra viruses are members of the *Paramyxoviridae* family within the *Mononegavirales* order. Measles virus (MeV) is one of the most infectious viruses ever known (Lamb & Parks, 2013). The Nipah and Hendra (NiV and HeV) viruses are newly emerged BSL-4 pathogens (B.T. Eaton, Mackenzie, & Wang, 2007) that have been gathered within the *Henipavirus* genus. They are the only currently known zoonotic paramyxoviruses and are responsible for severe encephalitis in human beings, with case fatality rates exceeding 75% (B. T. Eaton, Broder, Middleton, & Wang, 2006; L. F. Wang et al., 2000).

Paramyxoviruses are enveloped viruses. Their envelope consists of a lipid bilayer derived from the plasma membrane of the host cell. Beneath the envelope, the viral matrix protein (M) bridges the viral envelop glycoproteins with the nucleocapsid. As in all *Mononegavirales* members, the genome of paramyxoviruses is encapsidated by the nucleoprotein (N) giving rise to a helical nucleocapsid (N-RNA) with a characteristic herringbone-like structure (Bhella, 2007; Desfosses, Goret, Farias Estrozi, Ruigrok, & Gutsche, 2011; Gutsche et al., 2015; Karlin, Longhi, & Canard, 2002; Longhi et al., 2003; Schoehn et al., 2004). Following fusion between the viral envelope and the host cell membrane, the nucleocapsid is released in the cytoplasm where primary transcription takes place. A peculiar feature of *Mononegavirales* members is that the nucleocapsid, and *not* naked RNA, is the template used by the RNA-dependent RNA polymerase (RdRp) during both transcription and replication. The RdRp is made of the large protein (L) and the phosphoprotein (P). During RNA synthesis, P tethers L onto the nucleocapsid template through the N-P interaction. The P protein is therefore an essential polymerase cofactor since it allows the L protein to be recruited onto the nucleocapsid template. The complex formed by the N, P and L proteins constitutes the viral replicative unit (**Fig. 1A**). Once the viral ribonucleoprotein complexes are released in the cytoplasm of infected cells, transcription of viral genes starts. The polymerase enters the nucleocapsid at the 3' end of the genome where are located the promoters for transcription and replication. Transcription of the genes occurs sequentially. At each intergenic region (IGR), the polymerase ends and re-initiates at the downstream gene. Following polyadenylation, which serves as gene end

signal, the polymerase proceeds over three nucleotides without transcribing them and then restarts transcription upon recognition of a downstream gene start signal. The efficiency with which the polymerase reinitiates transcription decreases with increasing distance from the 3' end thereby leading to a gradient in the transcripts, with the most distal genes being the least expressed (Lamb & Parks, 2013) (**Fig. 1A**).

Following primary transcription, and presumably due to an increase in the intracellular concentration of N, the polymerase switches to a replication mode, i.e. it ignores the gene junctions and synthesizes a full, complementary strand of the same length as that of the genome. This positive-stranded RNA (termed antigenome) is not used as a template for transcription, and serves only as an intermediate in genome replication (see (Blocquel, Bourhis, et al., 2012; Lamb & Parks, 2013) for reviews). During replication, the nascent RNA chain is concomitantly encapsidated by N.

Although *in vitro* L can synthesize short RNA transcripts using naked RNA as substrate in the absence of P (Jordan et al., 2018; Morin, Liang, Gardner, Ross, & Whelan, 2016; Morin, Rahmeh, & Whelan, 2012), in infected cells P is required to allow recognition of the nucleocapsid template. P also serves as a chaperon towards L, acting in conjunction with hsp90 (L. M. Bloyet et al., 2016; Katoh et al., 2017).

The N protein is the most abundant viral protein. It has a major structural role as it wraps the genomic RNA within a nuclease-resistant nucleocapsid. In paramyxoviruses, N proteins are more than 500 residues in length. They consist of a well-conserved globular N-terminal moiety, N_{CORE}, which contains all the regions necessary for self-assembly and RNA-binding, and of a C-terminal domain, N_{TAIL} (**Fig. 1B**) (for reviews see (Erales, Blocquel, et al., 2015; Habchi & Longhi, 2012, 2015; Habchi, Mamelli, & Longhi, 2012; Jensen et al., 2010; Longhi, 2015; Longhi, Bloyet, Gianni, & Gerlier, 2017; Longhi & Oglesbee, 2010)). N_{TAIL} protrudes from N_{CORE} and is exposed at the surface of the viral nucleocapsid (Gutsche et al., 2015; Heggeness, Scheid, & Choppin, 1980, 1981; Karlin, Longhi, & Canard, 2002; Ringkjøbing Jensen et al., 2011) thereby allowing recruitment of P (Bankamp et al., 1996; Houben, Marion, Tarbouriech, Ruigrok, & Blanchard, 2007; Jensen et al., 2010; Kingston, Hamel, Gay, Dahlquist, & Matthews, 2004; Kingston, Walter, & Gay, 2004; Liston, Batal, DiFlumeri, & Briedis, 1997; Longhi et al., 2003).

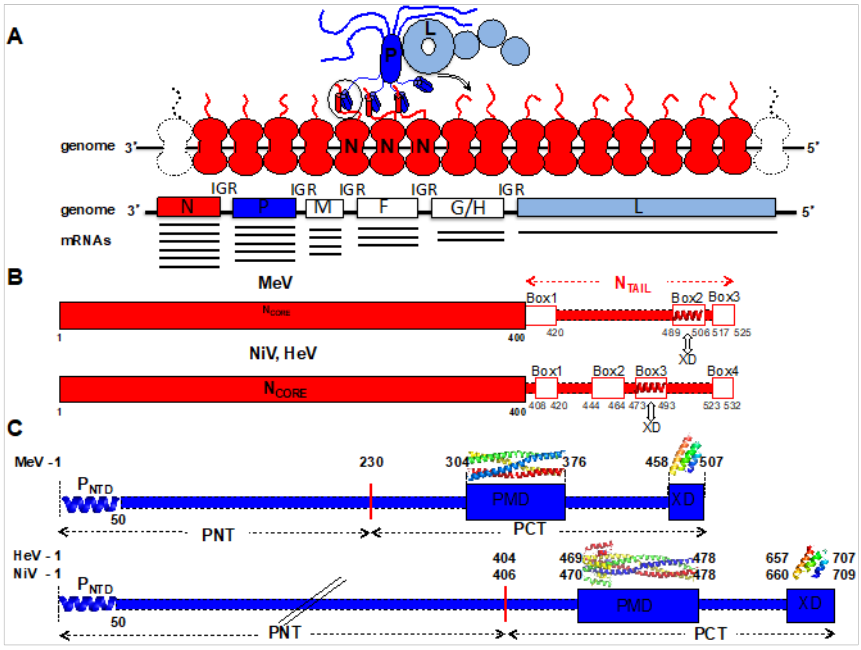


Figure 1. (A) Schematic illustration of the *Paramyxoviridae* replicative complex and of their genome. The N protein (red) is drawn with a bilobal shape according to available structural data (Gutsche et al., 2015). P (blue) is represented as a tetramer to reflect the prevalence of this oligomeric state in paramyxoviral P proteins. The L protein (light blue) is represented as a donut decorated with a tail made of three globules according to available structural data (Liang et al., 2015). The intrinsically disordered protein regions of N and P are symbolized by lines and α helixes by cylinders. The P-L complex forms the RNA-dependent RNA polymerase (RdRp) complex, which cartwheels onto the nucleocapsid complex via the X domain of P (XD). The negative-sense genomic RNA (blue dashed line) is presented in the 3' to 5' orientation. Below the genome, is shown a schematic representation of the expression gradient of the mRNA as a result of inefficient transcription re-initiation by the polymerase during transcription (Lamb & Parks, 2013). (B) Modular organization of N showing that it consists of a folded domain, N_{CORE}, and a C-terminal disordered region, N_{TAIL}. The various boxes corresponding to putative or experimentally proven MoREs are shown. The

α-MoRE involved in binding to XD is shown as a red helix and its ability to interact with XD is symbolized by a two-headed arrow. (C) Modular organization of P. Domain organization of P showing two moieties, PNT and PCT. The P editing site (see text) is shown. Structured and disordered regions are represented as large and narrow boxes respectively. PNT: N-terminal region of P; PCT: C-terminal region of P. PMD: P multimerization domain; XD: X domain consisting of a triple α-helical bundle. The α-MoRE at the N-terminal region of P (P_{NTD}), which is partly preconfigured in solution and shown to adopt a stable α-helical conformation upon binding to the monomeric form of N (Guryanov, Liljeroos, Kasaragod, Kajander, & Butcher, 2016; Yabukarski et al., 2014), is shown as a blue helix. The crystal structures of MeV PMD (PDB code 4BHV) (Blocquel et al., 2014), NiV PMD (PDB code 4N5B) (Bruhn-Johannsen et al., 2014), MeV XD (PDB code 1OKS) (Johansson et al., 2003) and HeV XD (PDB code 4HEO) (Communie, Habchi, et al., 2013) are shown. All structures were drawn using Pymol (DeLano, 2002).

The amino acid sequence of N_{TAIL} domains in *Paramyxoviridae* members is highly variable and features a compositional bias (i.e. it is enriched in polar and charged residues and depleted in hydrophobic residues) (J. Bourhis et al., 2004; Habchi, Mamelli, Darbon, & Longhi, 2010; Karlin, Ferron, Canard, & Longhi, 2003; Longhi et al., 2003). MeV N_{TAIL} is hypersensitive to proteolysis (Karlin, Longhi, & Canard, 2002) and cannot be visualized in (cryo)-electron microscopy (EM) reconstructions of nucleocapsids (Bhella, Ralph, & Yeo, 2004; Desfosses et al., 2011; Longhi et al., 2003). In a seminal study carried out in the early 2000s, these features were regarded as possible hints of intrinsic structural disorder, and the disordered nature of MeV N_{TAIL} was indeed experimentally proven (J. Bourhis et al., 2004; Longhi et al., 2003). This seminal study seeded subsequent studies focused on other paramyxoviral N proteins (e.g. Sendai virus and henipaviruses) (Habchi et al., 2010; Houben et al., 2007) that unveiled that the presence of a long disordered domain within paramyxoviral N proteins is a conserved feature.

EM studies failed to provide direct structural data on the disordered N_{TAIL} domain because this region was either unresolved (Desfosses et al., 2011) or removed by limited proteolysis (Gutsche et al., 2015). Thanks to the

combination of nuclear magnetic resonance (NMR) and small-angle scattering, a model of the RNA-bound form of full-length MeV N was obtained (Ringkjøbing Jensen et al., 2011). In this model, N_{TAIL} points towards the inner channel of the nucleocapsid and then escapes back to the solvent through the confined interstitial space between successive turns of the nucleocapsid (Ringkjøbing Jensen et al., 2011). This model provides a mechanistic basis to explain the increased rigidity of MeV nucleocapsids in which the flexible N_{TAIL} region has been proteolytically removed (Gutsche et al., 2015; Longhi et al., 2003; Schoehn et al., 2004). Subsequent studies carried out on Henipavirus nucleocapsid-like particles unveiled a very similar structural organization and showed that NiV and HeV N_{TAIL} domains too remain disordered when appended to the N_{CORE} domain (Baronti et al., 2015; Communie, Habchi, et al., 2013).

Being at least partly exposed at the surface of the nucleocapsid, N_{TAIL} can bind to various partners including the P protein.

The P protein is encoded by the P gene. Paramyxoviral P genes encode multiple proteins through either overlapping reading frames (C protein) or messenger editing (V and W proteins). The latter is a process whereby one or more non-templated G are inserted co-transcriptionally at a specific site. This insertion causes a shift in the open reading frame leading to a new protein product. Depending on the number of nucleotides inserted at the edition site, the V or W proteins are the end products. The V and W proteins share therefore with P their N-terminal domain and possess each a unique C-terminal domain. The P N-terminal (PNT) and the P C-terminal (PCT) regions, as well as the C-terminal region of V and W are therefore *bona fide* domains. Bioinformatics analyses showed that paramyxoviral P proteins share a common structural organization with PNT predicted to be intrinsically disordered and PCT consisting of alternating structured and disordered regions (Habchi et al., 2010; Karlin et al., 2003) (**Fig. 1C**). The PNT domain of henipaviruses is spectacularly large, consisting of 404 (HeV) or 406 (NiV) residues, and accounts for the extra length of *Henipavirus* P proteins. The disordered state of MeV, NiV and HeV PNT was experimentally confirmed using a wide range of biochemical and biophysical approaches (Habchi et al., 2010; Karlin, Longhi, Receveur, & Canard, 2002). Interestingly, we have recently shown that *Henipavirus* PNT domains retain

their disordered nature also in the context of the V protein (Salladini, Delauzun, & Longhi, 2017).

Within PCT, there are two structured regions: the P multimerization domain (PMD) and the C-terminal X domain (XD), which are responsible, respectively, for P oligomerization and for binding to N_{TAIL} (J. Bourhis et al., 2004; Houben et al., 2007; Johansson et al., 2003; Kingston, Hamel, et al., 2004; Kingston, Walter, et al., 2004) (**Fig. 1C**). While PMD is a coiled-coil (Blocquel et al., 2014; Bruhn-Johannsen et al., 2014; Communie, Crepin, et al., 2013), the structure of XD consists of a bundle of three antiparallel α-helices (Blanchard et al., 2004; Communie, Habchi, et al., 2013; Gely et al., 2010; Johansson et al., 2003; Kingston, Hamel, et al., 2004) (**Fig. 1C**). Its NiV counterpart is predicted to adopt a similar folding (Habchi et al., 2011). Although MeX XD is folded, electron spray ionization mass spectrometry (ESI-MS) studies unveiled that it is structurally heterogeneous, populating at least two alternative conformations under native conditions (D'Urzo et al., 2015). This conclusion is further strengthened by a recent study where the structure of a folding intermediate of MeV XD, which accounts for as much as 30% of the protein population under native conditions, was identified (Bonetti et al., 2016).

N_{TAIL} and XD interact with each other forming a 1:1 stoichiometric complex with an equilibrium dissociation constant (K_D) in the μM range (Dosnon et al., 2015; Habchi et al., 2011). Binding to XD triggers α-helical folding of a short N_{TAIL} region (Box2, aa 486-504 of MeV N, and Box3, aa 473-493 of *Henipavirus* N), referred to as a Molecular Recognition Element or α-MoRE (**Fig. 1B**) (J. Bourhis et al., 2004; Habchi et al., 2011; Johansson et al., 2003; Kingston, Hamel, et al., 2004). MoREs are short, transiently populated secondary structures within ID regions that are biased towards their bound state (Fuxreiter, Simon, Friedrich, & Tompa, 2004).

Both random and site-directed mutagenesis allowed identifying the MeV N_{TAIL} residues most critical for binding to XD (L. Bloyet et al., 2016; Gruet et al., 2013). Further investigation of single-site N_{TAIL} variants bearing substitutions within the α-MoRE revealed that the affinity of the N_{TAIL}-XD pair is directly related to the efficiency with which the viral polymerase reinitiates transcription at each IGR (L. Bloyet et al., 2016).

The Cα chemical shifts of both MeV and *Henipavirus* N_{TAIL} and the mobility of paramagnetic spin labels grafted at various N_{TAIL} positions

indicate that the α-MoRE is partly preconfigured as an α-helix in the absence of XD (Baronti et al., 2015; Belle et al., 2008; Communie, Habchi, et al., 2013; Gely et al., 2010; Martinho et al., 2013; Morin et al., 2006). The Box3 region of NiV N_{TAIL} exhibits a higher degree of pre-configuration (Baronti et al., 2015). It also populates longer-lived interconverting α-helical segments than in its HeV N_{TAIL} counterpart, as detected by site-directed spin labeling (SDSL) electron paramagnetic resonance (EPR) measurements (Martinho et al., 2013). A subsequent study that combined residual dipolar couplings (RDCs) measurements and ensemble optimization methods (Bernado et al., 2005) showed that the α-MoRE of MeV N_{TAIL} exists in a conformational equilibrium between a completely unfolded form and four partially helical conformers (Ringkjøbing Jensen et al., 2011). In the closely related Sendai virus, the α-MoRE has a similar conformational behavior, although it samples only three helical conformers (Jensen et al., 2010; Jensen et al., 2008).

Partial preconfiguration of the XD-binding region of N_{TAIL} is therefore a conserved feature shared by various members of the *Paramyxoviridae* family, implying a functional significance. The partial preconfiguration of MoREs facilitates the folding-upon-binding process: the residual structure restrains the conformational space sampled by the ID protein, thereby reducing the number of interconverting conformers in solution and rendering the structural transition to the (partially) folded conformation energetically less demanding (Fuxreiter et al., 2004). Therefore, although disorder-to-order transitions are entropically disfavored, ID proteins can finely tune their affinity towards partners by varying the extent of preconfiguration of their partially structured elements (Fuxreiter et al., 2004).

In spite of this pre-configuration, molecular dynamic simulations and kinetics studies showed that MeV N_{TAIL} folds according to a folding after binding mechanism (Dosnon et al., 2015; Y. Wang et al., 2013). Subsequent mutational studies coupled to Φ-value analysis led to a detailed structural description of the folding and binding events occurring in the recognition between N_{TAIL} and XD (Bonetti et al., 2017). By Φ-value analysis, residue-specific structural information is obtained by comparing the kinetics of the reaction (folding and/or binding) of the *wild-type* protein with that of a series of conservative single-site variants, yielding the so-called Φ value. The latter

represents an index of native-like structure of the mutated residue in the transition state (Fersht, Matouschek, & Serrano, 1992). In the case of MeV N_{TAIL}, we measured the effect of single-amino acid substitutions in N_{TAIL} on the reaction mechanism. Analysis of the experimental data allowed the identification of key residues involved in the initial recognition between the two molecules and enabled the general features of the folding pathway of N_{TAIL} to be unraveled. In addition, analysis of the changes in stability for all the variants highlighted the inherent poor folding efficiency of N_{TAIL}. We have proposed that this feature might be a consequence of the weakly funneled nature of the energy landscape of ID proteins in their unbound state that might bias for a highly heterogeneous bound state (Bonetti et al., 2017).

A folding upon binding mechanism is not a unique feature of the MeV N_{TAIL}-XD interaction. Indeed, quantitative analysis of peak intensities in HSQC spectra of HeV N_{TAIL} at various XD titration points showed a differential broadening consistent with a mechanism where XD would first bind to a short, central helix within the α-MoRE, and then this helix would be gradually extended via helical folding of the adjacent residues (Communie, Habchi, et al., 2013). A folding upon binding mechanism is also shared by the Sendai virus N_{TAIL}-XD pair, where a detailed atomic description of the molecular recognition trajectory could be obtained from relaxation dispersion studies (Schneider et al., 2015).

The crystal structure of a MeV chimeric construct in which XD is covalently attached to the α-MoRE of N_{TAIL} (aa 486-504) was solved at 1.8 Å (Kingston, Hamel, et al., 2004). The structure consists of a pseudo-four helix complex in which the α-MoRE of N_{TAIL} adopts a parallel orientation with respect to XD and is embedded in a large hydrophobic cleft delimited by XD helices α2 and α3 (Kingston, Hamel, et al., 2004). The structure of the MeV XD/α-MoRE complex was used as a template to model the structure of the homologous *Henipavirus* complexes (Habchi et al., 2011). These models were successively validated by SDSL EPR spectroscopy studies (Martinho et al., 2013).

In spite of the lack of direct structural data on *Henipavirus* N_{TAIL}-XD complexes, NMR studies provided insights into the structure of the HeV N_{TAIL}-XD complex and allowed further model refinement using constraints based on chemical shifts (Communie, Habchi, et al., 2013). Analysis of chemical shift perturbations in reciprocal titration studies, combined with the

availability of the crystal structure of XD, allowed the residues involved in the interaction to be identified (Communie, Habchi, et al., 2013). These studies revealed that albeit the binding interface is prevalently hydrophobic, the binding pocket of XD is surrounded by acidic residues that may establish electrostatic interactions with basic residues of Box3. Subsequent isothermal titration calorimetry (ITC) studies carried out at different pH values did in fact confirm the role of electrostatics in complex formation, a conclusion further strengthened by mutational studies that targeted charged residues of both N_{TAIL} and XD (Erales, Beltrandi, Roche, Maté, & Longhi, 2015). Accordingly, we proposed that the formation of the HeV N_{TAIL}-XD complex may rely on the so-called "electrostatic steering mechanism" (Schreiber, Haran, & Zhou, 2009). This model also implies a folding after binding mechanism, in agreement with NMR titration data (Communie, Habchi, et al., 2013). Neither chemical shifts nor electrostatic interactions enable drawing definite conclusions on the orientation of the MoRE at the XD surface. Through a combination of mutational and small-angle X-ray scattering (SAXS) studies, the α-MoRE was shown to adopt a parallel orientation at the surface of HeV XD (Erales, Beltrandi, et al., 2015), as in the case of the MeV complex (Kingston, Hamel, et al., 2004). The conserved parallel orientation of the MoRE at the XD surface in both MeV and HeV might reflect the need for a precise relative orientation of the whole P protein with respect to the N_{TAIL} region protruding from the nucleocapsid. In its turn, this might be related to optimal positioning of the polymerase onto the nucleocapsid template and might impart directionality to the polymerase movement along the nucleocapsid.

In both MeV and henipaviruses, following binding to XD, the majority of N_{TAIL} exhibits a conformational exchange and does not establish stable contacts with XD (Baronti et al., 2015; Belle et al., 2008; Belle, Rouger, Costanzo, Longhi, & Fournel, 2010; Blocquel, Habchi, Gruet, Blangy, & Longhi, 2012; J. M. Bourhis et al., 2005; Communie, Habchi, et al., 2013; Gely et al., 2010; Habchi et al., 2011; Martinho et al., 2013; Morin et al., 2006). These N_{TAIL}-XD complexes are therefore illustrative examples of fuzziness (Tompa & Fuxreiter, 2008)Fuxreiter, 2012 #3094;(Fuxreiter, 2018).

Below, we summarize the available experimental data supporting the fuzziness of paramyxoviral N_{TAIL}-XD complexes and describe experiments

that allowed unraveling its functional impact. We focus onto the experimental design, and for each of the various experimental approaches detailed below we emphasize the advantages and limitations and provide a few caveats.

2. EXPERIMENTAL APPROACHES TO PROBE FUZZINESS IN PARAMYXOVIRAL N_{TAIL}-XD COMPLEXES

As detailed below, various experimental approaches can be used to document fuzziness in IDP complexes. They broadly vary in terms of the molecular information provided, of protein requirements and of ease of implementation. **Fig. 2** illustrates a general suggested procedure for the experimental assessment of fuzziness with increasing level of complexity.

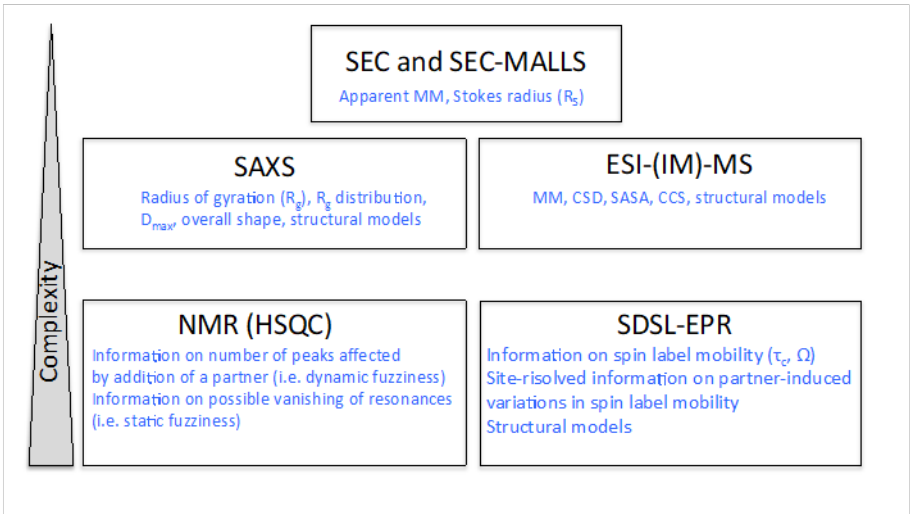


Figure 2. General suggested procedure for the experimental assessment of fuzziness in IDP complexes. The molecular information provided by each approach is indicated, as it the degree of complexity of the various approaches

2.1. SEC and SEC-MALLS

One of the easiest approaches to document lack of significant compaction upon complex formation is size-exclusion chromatography (SEC). The latter allows separating proteins according to their hydrodynamic

volume. In practice, upon injection of a mixture containing the two interacting proteins at a concentration that has to be well above the K_D so as to ensure complex formation, the Stokes radius (R_S) of the complex can be determined from the elution volume using a calibration curve obtained using globular proteins of known molecular mass (MM, in Daltons). The R_S (in Å) of the molecular mass standards is calculated according to (Uversky, 1993):

$$\log (R_S^{Obs}) = 0.369 * (\log MM) - 0.254 \quad \text{(Eq. 1)}$$

From this calibration curve one can thus infer the R_S of the complex ($R_S^{Obs\text{-}Complex}$). This experimentally observed value is then compared to the value expected for a fully folded complex ($R_s^{FF\text{-}Complex}$) made of protein A and protein B. The latter is calculated according to (Uversky, 2002):

$$\log (R_S^{FF\text{-}Complex}) = 0.357 * [\log (MM^A+MM^B)] - 0.204 \quad \text{(Eq. 2)}$$

where MM^A and MM^B are the molecular masses (in Daltons) of protein A and protein B. It is recommended to check that the elution profile is not ionic strength– and buffer-dependent. It also goes without saying that one has to check by sodium dodecyl sulfate polyacrylamide electrophoresis (SDS-PAGE) the effective presence of the two interacting proteins in the fractions of the peak supposed to correspond to the complex.

Alternatively, the Stokes radius can be determined by SEC coupled to multiple angle laser light scattering (SEC-MALLS). In this case, the Stokes radius and the molecular mass are directly measured and not inferred from a calibration curve. While SEC relies on fast protein liquid chromatography (FPLC), SEC-MALLS requires the more sophisticated and hence more expensive high-pressure liquid chromatography (HPLC) system.

In both cases, the major technical limitation resides in the ability to document complex formation. Indeed, upon injection of 500 µL of a mixture containing both MeV N_{TAIL} and XD at protein concentrations up to 5 mg/mL each onto a SEC Superdex 75 HR 10/30 column (GE, Healthcare) no peak corresponding to the complex could ever be detected. This is likely due to complex dissociation during elution in the column induced by dilution. This technical limitation can be sometimes overcome using SEC-MALLS where both the volume of the injected sample (typically, 30 µL of a protein solution at 0.3 - 1.5 mM) and that of the column are reduced. Typically, silica-based columns (Shodex) have a smaller volume (15 mL) compared to Superdex

columns routinely used in SEC. Using this approach we could indeed successfully document NiV N_{TAIL}-XD complex formation (Habchi et al., 2011): upon addition of stoichiometric amounts of XD, the N_{TAIL} peak shifted towards a lower elution volume indicating formation of the complex, as also judged from the estimated mass (25,610 ± 80 Da), a value consistent with a 1:1 complex (expected mass = ~ 22 kDa). Upon adding a two-fold molar excess of XD, no further significant shift in the elution volume of N_{TAIL} was observed, indicating that saturation was achieved for a 1:1 complex. The measured R_S of the complex is 35.4 ± 3.1 Å, a value much larger than that expected (22.3 Å) for a fully folded complex with a molecular mass equal to the sum of the masses of N_{TAIL} and XD (~ 22 kDa) (Uversky, 2002). This large value of the R_S reflects a complex that is not fully compact and that rather retains considerable conformational entropy (Habchi et al., 2011). Notably, no shift in the elution profile of HeV N_{TAIL} was observed upon addition of as much as a four-fold molar excess of XD, indicating lack of complex formation and probably reflecting a less tight complex. That the NiV N_{TAIL}-XD complex is tighter than its HeV counterpart is also supported by SDSL EPR and NMR spectroscopy data that unveiled the contribution of an additional region (i.e. Box2, aa 444-464) to binding to XD (Baronti et al., 2015; Martinho et al., 2013) (see below).

2.2. Small-angle X-ray scattering (SAXS)

SAXS is a powerful method to obtain low-resolution (20-10 Å) structural information on flexible molecules in solutions. The scattering curve, as recorded by irradiating a protein solution with X-rays and using a detector located at hundreds of cm from the sample (typically from 1 to 3 m), depends on the size and the shape of the protein. The higher the distance of the detector, the lower the scattering angle (2θ), and hence the lower the resolution. Although SAXS studies can also be performed using laboratory equipment, the use of synchrotron radiation dramatically speeds up the process of data collection and yields data of better quality. Typically 30-50 μL of protein samples are used in a concentration range of 1 to 10 mg/mL. It is recommended to collect data at various protein concentrations so as to rule out any possible attractive and/or repulsive inter-particle interactions. It is also recommended to use *strictly* the same buffer in which the protein is prepared when recording background scattering.

From the scattering curve one can derive the radius of gyration, R_g, the R_g distribution (for IDPs), the maximal intramolecular distance and the molecular mass. The R_g and the R_S correspond to two distinct parameters. The R_g is the average root-mean- square distance to the center of density in the molecule weighted by the scattering length density, whereas the R_S is the radius of a spherical particle of the same molecular mass as that of the particle under study. From the scattering data, Guinier plots (where ln(I) is plotted as a function of q^2) can be built. The slope of the plots in the Guinier region ($qR_g < 1.0$) gives the value of the R_g, while the intercept of the straight line gives the I(0), which is proportional to the molecular mass of the scatterer.

Not only SAXS is very well suited to study IDPs (Bernado & Svergun, 2010, 2012), but can also be used to study complexes made of IDPs bound to their partner(s). We successfully used SAXS to investigate the MeV N_{TAIL}-XD complex (J. M. Bourhis et al., 2005).

We first ascertain that XD possesses the same structure in solution as in the crystal. SAXS experiments performed on XD showed that it has a globular shape with an R_g of 12.1 ± 0.8 Å and a maximum diameter D_{max} of 41 ± 1 Å. The scattering profile back-calculated from the crystal structure (pdb code: 1OKS) using the program CRYSOL (Svergun, Baraberato, & Koch, 1995) is close to the experimental one (χ^2 = 1.4) and yields a theoretical R_g of 12.3 Å, thereby indicating that the overall conformation of XD in solution is similar to that observed in the crystal.

The N_{TAIL}-XD complex was prepared by mixing N_{TAIL} with a two-fold molar excess of XD at a final protein concentration of 10 mg/ml. SAXS data were also collected on N_{TAIL} at various protein concentrations ranging from 1.8 to 10 mg/ml. Background scattering was measured before or after each protein sample using the buffer solution and then subtracted from the protein scattering patterns. The experimental scattering curves at the various protein concentrations were merged to yield a single merged curve.

The scattering profile of the N_{TAIL}-XD complex was obtained as follows. To avoid any possible bias on the absolute intensities due to the concentration, the experimental merged scattering curves obtained as described above were normalized by their theoretical I(0). The scattering pattern of XD was then subtracted from the scattering pattern of the mixture containing N_{TAIL} and XD at a 1:2 molar ratio. Since the K_D between N_{TAIL} and

XD is in the µM range, the concentration of uncomplexed N_{TAIL} in solution can be assumed to be negligible at the concentration used in the SAXS experiments. The scattering pattern of the complex obtained as described above was then used for further calculations. Guinier analaysis of the curve yielded an R_g of 32.7 ± 0.7 Å. The molecular mass calculated from the forward scattering intensity I(0) is 23.5 ± 2 kDa, in agreement with the value expected for a 1:1 stoichiometric complex (21.3 kDa), thus confirming complex formation. The high value of R_g indicated that the overall structure of the N_{TAIL}-XD complex is not compact. The distance distribution function inferred from the scattering curve of the N_{TAIL}-XD complex is asymmetrical with a maximum at 20 Å, a shoulder at about 30 Å and a maximum value at 146 Å, which is typical of an elongated object (see (J. M. Bourhis et al., 2005)). The bump most probably corresponds to the intramolecular distance within the globular portion of the complex (see below), while the tail indicates that N_{TAIL} possesses regions with an extended conformation.

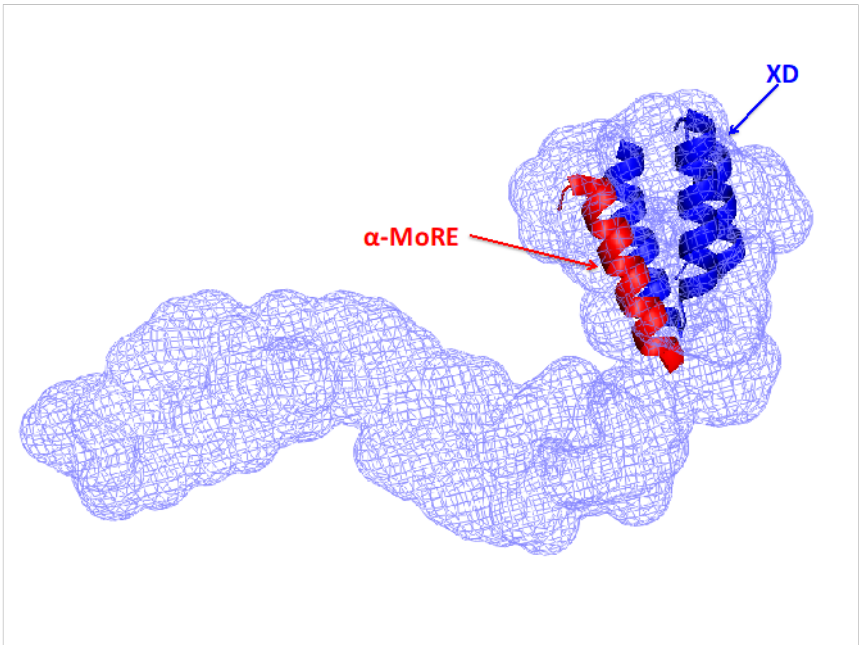


Figure 3. SAXS model of the MeV N_{TAIL}-XD complex. The structure of the chimeric construct made of XD and of the α-MoRE is shown within the globular part of the envelope obtained by *ab initio* shape reconstruction. Data are taken from (J. M. Bourhis et al., 2005).

The overall envelope of the complex was calculated *ab initio* from its scattering profile using the program GASBOR (Svergun, Petoukhov, & Koch, 2001). Several independent runs yielded different shapes with recurrent features: they were all elongated, with a globular cluster of always the same size at one extremity and an elongated protuberance with different curvatures and cross-sections. The quality of the fit to the experimental data was similar in all cases, with χ^2 values ranging from 1.3 to 1.6. The globular part was assumed to correspond to XD packed against the folded region of N_{TAIL} (i.e. the α-MoRE) while the appendage corresponds to unfolded regions of the latter. The crystal structure of the chimera made of XD and of the 486-505 region of N_{TAIL} (pdb code: 1T6O), was inserted in the shape using the program SUPCOMB (Kozin & Svergun, 2001) (**Fig. 3**). In order to get insights into the conformation that the regions of N_{TAIL} preceding and following the α-MoRE (aa 401-485 and 506-525, respectively) adopt in the complex with XD, the program package CREDO was used. CREDO is an extension of GASBOR and calculates the position of missing loops in crystal structures. The results obtained with several independent runs using the crystal structure of the chimera as template, yielded multiple models all consisting of at one side a peptide chain in an extended conformation and at the other side a chain packed against the bulky portion of the XD/α-MoRE complex. The extended conformation, protruding from the globular cluster of the complex and pointing to the solvent, corresponds to the long N-terminal region of N_{TAIL}. In contrast, the C-terminal region of N_{TAIL} (aa 506-525), always packs against the globular moiety of the envelope but at varying distances and positions. The back-calculated scattering curves from these models were found all to well (and equally) fit the experimental data (χ^2 of 0.8-1.0).

In conclusion SAXS studies clearly unveiled the presence of a long flexible appendage accommodating the N-terminal region of N_{TAIL}.

2.3. Nuclear Magnetic Resonance (NMR)

NMR spectroscopy is among the most (if not the most) powerful techniques to investigate IDPs and the structural transitions they undergo upon binding to their partner(s) (Felli & Pierattelli, 2015). Thanks to recent improvements in NMR instrumentation, pulse sequence design, and sample

preparation, new NMR tools have become available for the characterization of IDPs at atomic resolution. NMR spectroscopy provides a unique tool to access dynamics information at atomic resolution, from the picosecond timescale to slow exchange processes on the second (or even slower) timescale. As such, NMR is ideally well suited to study IDPs that are often characterized by heterogeneous dynamic properties along the polypeptide chain. Peak positions in NMR spectra depend on the resonance frequencies of the observed nuclei that in turn depend on their chemical environment. Resonance positions in NMR spectroscopy are expressed as "chemical shifts" relative to a standard and are expressed in parts *per* million (ppm). In IDPs, the lack of a stable structure results in averaging of a large part of the contributions to the chemical shift coming from the local chemical environment, thereby leading to the typical low chemical shift dispersion observed in the NMR spectra (**Fig. 4**). The chemical shifts of backbone nuclei are sensitive to the local backbone geometry, and thus provide useful information on the nature and level of secondary structural elements within the protein (Spera & Bax, 1991). The chemical shifts make NMR an atomic resolution technique in that once the resonances are assigned to their respective nuclei their response to perturbations (such as addition of partners/ligands) can be monitored.

Heterogeneous Single Quantum Correlation (HSQC) spectra are very useful to highlight structural transitions in IDPs. The approach relies on the use of a ^{15}N uniformly labeled IDP and on the monitoring of possible chemical shift changes in the backbone amide and proton resonances upon addition of an unlabeled partner. In this way, one can estimate the number of affected residues without the need of a prior full assignment of the spectrum.

^{1}H-^{15}N HSQC studies of ^{15}N uniformly labeled MeV N_{TAIL} XD either alone or in the presence of a two-fold molar excess of XD (J. M. Bourhis et al., 2005) allowed a quantitative estimation of the number of residues involved in the interaction with XD (**Fig. 4A**). 11 peaks undergo an upfield shift, which is indicative of a random coil to α-helix transition, and 5 additional ones undergo a less dramatic displacement reflecting only a change in the chemical environment without concomitant gain of regular secondary structure. The change in the chemical environment of these latter residues can result either by direct interaction with the partner or by local magnetic perturbations of residues spatially close to the newly formed α-

helix. Although this approach allows drawing conclusions about the number of residues affected by the partner and the nature of the structural transition, it does not provide information on the involved residues. Identifying the latter requires assignment of the spectrum.

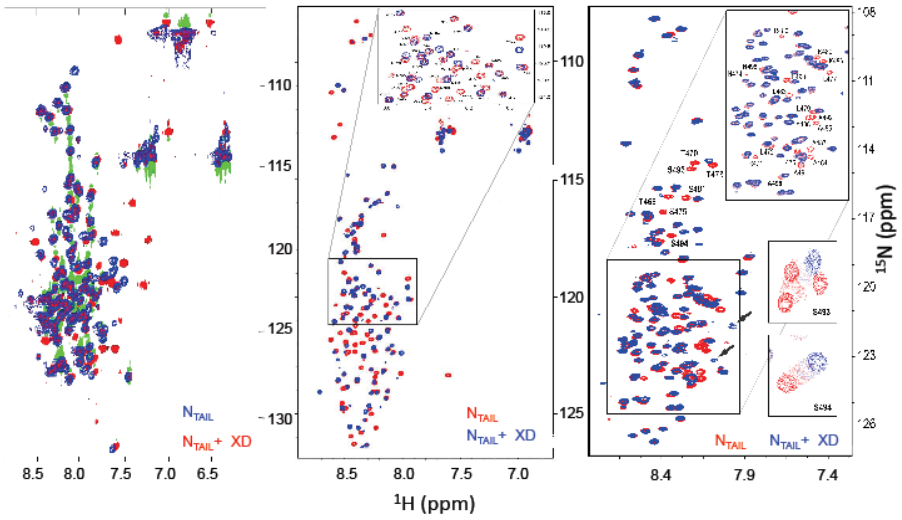


Figure 4. 2D-HSQC NMR spectra of NTAIL domains. MeV (A), HeV (B) and NiV (C) NTAIL. (**A**) HSQC spectrum of a 0.5 mM solution of MeV 15N-NTAIL alone (blue) or in the presence of a two-fold molar excess of XD (red) in 10 mM sodium phosphate at pH 7. The green spots correspond to peaks of negative intensity. The spectra were recorded at 283 K on a 500 MHz DRX Bruker spectrometer. ppm quotes for resonance shifts in parts per million of the spectrophotometer frequency (J. M. Bourhis et al., 2005). (**B**) Superposition of the 1H-15N HSQC spectra of HeV NTAIL alone at 40 µM (red) and in the presence of a 4.4-fold molar excess of XD (blue) (Communie, Habchi, et al., 2013). The spectra were obtained at a 1H frequency of 800 MHz and 293 K. The expanded regions shows the spectral assignment of a number of resonances in the spectrum of the isolated NTAIL domain. (**C**) Superimposition of the 1H-15N HSQC spectra of isolated NTAIL at 180 µM (red) and in the presence of a four-fold molar excesses of

XD (blue) (Baronti et al., 2015). The spectra were obtained at 500 MHz and 285.5 K. The expanded region shows the spectral assignment of a number of resonances in the spectrum of the isolated NTAIL domain. Two additional peaks that appear in the presence of saturating amounts of XD are indicated by an arrow.

Subsequently, the HSQC spectrum of MeV N_{TAIL} could be assigned using a double-labeled ($^{13}C/^{15}N$) N_{TAIL} sample and a series of BEST-type triple resonance experiments (Gely et al., 2010). NMR titration experiments with XD were then performed. In these experiments, ^{1}H-^{15}N HSQC spectra were collected on a 0.3 mM sample of ^{15}N-labeled N_{TAIL} in the presence of different concentrations of unlabeled XD. At low XD concentrations (0.037, 0.075, and 0.150 mM), most resonances in the HSQC experience no chemical shift changes (residues 401–474), some resonances undergo fast exchange and experience chemical shift changes with minimal line broadening (residues 475–482 and residues 507–525), and some resonances undergo intermediate exchange and experience significant line broadening (residues 483–506). For most resonances in the latter category, line broadening was so severe that prevented signal detection. At high XD concentrations (0.3 and 0.6 mM), resonances for residues 492–501, and 503–506 reappeared at new positions, suggesting the formation of a stable complex. The most pronounced amide ^{1}H and ^{15}N chemical shift changes upon binding XD concern N_{TAIL} residues 483–506. In addition, several residues in the Box3 region have chemical shift differences in the range of 20–30 Hz (Gely et al., 2010). From the chemical shifts of the free and bound form of N_{TAIL}, the secondary chemical shifts (Δδ) were computed and allowed estimating the amount of helical structure along the polypetide chain. Δδ values greater than one indicate a fractional population of helical structure in the range of 50%. This is the case of residues 491–499 in free N_{TAIL}. The ^{13}Ca and ^{13}CO resonances of these residues undergo positive shifts in their Δδ values in the presence of XD, indicating that the transient helical structure in free N_{TAIL} is stabilized in the N_{TAIL}-XD complex.

Altogether, NMR studies revealed that the majority of N_{TAIL} residues are not affected by binding to XD (i.e. aa 401-474) and therefore remain fuzzy. Among affected residues, two types of interaction are discernible: one, mediated by residues belonging to Box2, involves a significant gain of α-

helicity, while the other, attributable to Box3 residues, is not accompanied by a significant structural transition.

Similar NMR experiments were successively carried out on HeV and NiV N_{TAIL} (Habchi et al., 2011). Their spectra could be assigned using standard triple resonance experiments (Baronti et al., 2015; Communie, Habchi, et al., 2013). Experimental chemical shifts and nuclear relaxation data unveiled the presence of a transiently populated α-helix encompassing residues 470-490 in HeV N_{TAIL} and residues 470-495 in NiV N_{TAIL}, a region largely overlapping with Box3 (**Fig. 1B**) and previously shown to be the XD binding region (Blocquel, Habchi, et al., 2012; Habchi et al., 2011). Like in the case of MeV N_{TAIL}, titration studies with unlabeled XD revealed that most resonances in the HSQC experienced no chemical shift changes (**Fig. 4B** and **C**). For both HeV and NiV N_{TAIL} samples, the largest chemical shift perturbations were observed within the 470-490 region (Baronti et al., 2015; Communie, Habchi, et al., 2013). In the case of HeV, the addition of an increasing amount of unlabeled XD to ^{15}N-labeled N_{TAIL} results in severe line broadening of the NMR resonances for residues 470–490 even at sub-stoichiometric amounts of XD. In the case of NiV, peaks corresponding to this region disappear at the first titration point. Resonance vanishing without concomitant appearance of new peaks in the α-helical region, is consistent with an intermediate exchange regime occurring on the micro- to millisecond time scale (i.e. the exchange rate between free and bound N_{TAIL} is comparable to the chemical shift difference between the free and the bound form) and reflects a highly dynamic binding of Box3 at the surface of XD.

In the case of NiV, 2D ^{13}C–^{15}N CON experiments revealed the additional involvement in the interaction of residues 452–456 that undergo slow to intermediate exchange upon addition of XD.

Collectively, titration studies revealed that, in contrast with MeV, for both NiV and HeV N_{TAIL} domains, the resonances of the α-MoRE vanish upon addition of the homologous XD, a finding consistent with a highly dynamic complex in which the α-MoRE undergoes α-helical fraying at the surface of XD (Baronti et al., 2015; Communie, Habchi, et al., 2013; Habchi et al., 2011). The NiV N_{TAIL}-XD complex is slightly tighter than that of HeV (Baronti et al., 2015; Blocquel, Habchi, et al., 2012; Communie, Habchi, et al., 2013; Habchi et al., 2011), a property that is reflected in our ability to document by SEC complex formation in the case of NiV but not in the case

of HeV (Habchi et al., 2011). In further support of a tighter complex, SDSL EPR experiments showed that in the case of NiV, an additional N_{TAIL} region (i.e. Box2) is involved in the interaction with XD (Martinho et al., 2013) (see below).

Therefore, the *Henipavirus* N_{TAIL}-XD complexes are even fuzzier than the MeV one, as the fuzziness also pertains the binding region itself. This likely provides an additional way of tuning the strength of the interaction, rendering it more prone to be competed out by other partners.

2.4. Side-directed spin-labeling electron paramagnetic resonance (SDSL-EPR) spectroscopy

EPR spectroscopy is a technique that specifically detects unpaired electrons. It is based on the observation of the energy absorbed by a paramagnetic system in a magnetic field. SDSL combined with EPR is a powerful technique for detecting structural changes in proteins that are devoid of paramagnetic centers. The basic strategy of SDSL involves the insertion of a paramagnetic label at a selected site of a protein and its observation by EPR spectroscopy (for reviews see (Belle et al., 2010; Hubbell, Gross, Langen, & Lietzow, 1998; Longhi, Belle, Fournel, Guigliarelli, & Carrière, 2011). This is generally achieved through cysteine-substitution mutagenesis, followed by covalent modification of the sulfydryl group with a selective nitroxide reagent, such as the methanethiosulfonate derivative (MTSL). For a detailed procedure for the generation of variants (from site-directed mutagenesis to spin-labeling and EPR spectra recording) the interested reader may refer to (Habchi, Martinho, et al., 2012).

The EPR spectrum of the labeled protein is then recorded and variations in spin label mobility under various conditions can be monitored. The main advantages of this technique reside in the fact that (*i*) spectra can be recorded at room temperature, (*ii*) paramagnetic probes introduce a minimal perturbation of the system, and (*iii*) they probe the very local environment of the spin label. Any change in this environment resulting from either a physical or chemical event can then be identified from the change in the EPR signature. Indeed, the mobility of the spin label is reflected in the spectral shape of the spin-labeled protein. When paramagnetic labels are allowed to tumble rapidly in an isotropic way, magnetic interactions are completely averaged and the EPR spectrum displays three narrow lines. By contrast, with

slower motions the magnetic anisotropy is no longer totally averaged, thus giving rise to spectral variations. In the so-called *fast motional regime* (i.e. correlation time τ_c up to $10^{-9}s$), the spectrum is characterized by a differential broadening of lines although line positions remain constant. For $\tau_r > 10^{-9}s$, the averaging of the magnetic interactions becomes progressively less efficient, leading to shape distortions of the EPR spectrum.

Spin labels grafted within IDPs typically display a fast motional regime. To derive quantitative information on the spin label mobility, the most suited parameter is the ratio between the peak-to-peak amplitude of the low-field and central lines, h(+1)/h(0) ratio (Morin et al., 2006). This ratio decreases with decreasing spin label mobility (Morin et al., 2006).

SDSL EPR spectroscopy has emerged in the last decade as a powerful method to investigate structural transitions in IDPs (for reviews see (Casey et al., 2014; Drescher, 2012).

We have successfully used it to investigate the structural transitions that N_{TAIL} domains undergo in the presence of the homologous X domain (Belle et al., 2008; Martinho et al., 2013; Morin et al., 2006). As detailed below, the approach allowed us to precisely map the N_{TAIL} region undergoing α-helical folding.

In the case of MeV N_{TAIL}, we targeted for spin-labeling 14 sites, 12 of which are scattered in the 488 – 525 region, while two (positions 407 and 460) are located outside the reported region of interaction with XD (**Fig. 5A**). Fourteen single-site cysteine variants were designed and purified from the soluble fraction of *E. coli* and labeled as described in (Belle et al., 2008). Labeling yields ranged from 50 to 80%. The concentration of labeled proteins was evaluated by double integration of the EPR signal recorded under non-saturating conditions and comparison with that given by a 3-carboxy-proxyl standard sample. The labeling yields were estimated by calculating the ratio between the concentration of labeled proteins and the total protein concentration.

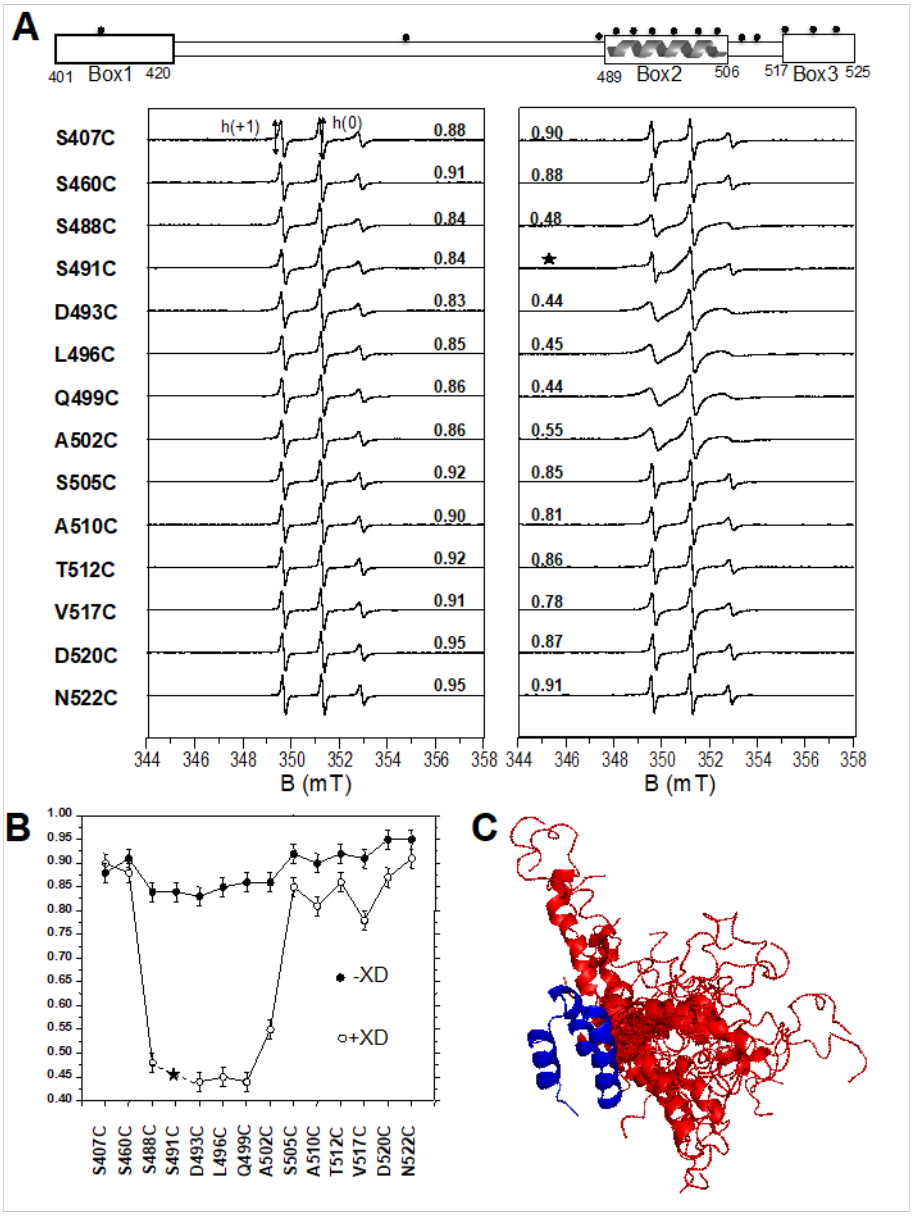


Figure 5. (**A**) Scheme of N_{TAIL} and of the positions targeted for cysteine substitution and spin labeling (diamonds) and amplitude normalized room

temperature EPR spectra of the spin-labeled MeV N_{TAIL} proteins (100 μM) in the absence (left panel) or presence of a molar excess of XD (right panel). EPR spectra were recorded at 296 K on an ESP 300E Bruker spectrometer equipped with an ELEXSYS Super High Sensitivity resonator operating at 9.9 GHz. Samples were injected into 20-μL quartz capillaries. The microwave power was 10 mW and the magnetic field modulation frequency and amplitude were 100 kHz and 0.1 mT, respectively. (**B**) h(+1)/h(0) ratios of the spin-labeled N_{TAIL} proteins either alone or in the presence of saturating amounts of XD as a function of spin-label position. The h(+1)/h(0) ratio of the spin-labeled S491C variant was not indicated, as it is not a reliable indicator of the mobility of this spin label (see text). Data of panels A and B were taken from (Belle et al., 2008). (**C**) Population of the 20 best-fit structures of the MeV complex made of XD (blue) and the 488-525 region of N_{TAIL} (red) as obtained after structural modeling at 281 K. Data were taken from (Kavalenka et al., 2010).

Like for all experimental approaches that relies on site-directed mutagenesis and/or on the covalent modification of the protein, it is recommended to check that the substitution and/or the chemical modification does not affect the overall secondary structure content and the global folding (or lack of) of the protein. In agreement, the possibility that the cysteine substitution and the introduction of the spin label might affect the overall structure and the folding propensities of the N_{TAIL} variants was checked and ruled out by far-UV circular dichroism (CD) (Belle et al., 2008).

The EPR spectra of all the spin-labeled N_{TAIL} variants exhibit a relatively narrow, single-component shape and are indicative of a high radical mobility, as judged on the basis of (h(+1)/h(0) ratio) (**Fig. 5A**, left panel). In spite of the overall high mobility, the spin labels grafted within Box2 exhibit a slightly reduced mobility, a behavior likely reflecting the partial preconfiguration of the α-MoRE. EPR spectra were also recorded in the presence of a molar excess of XD (both either in the absence or in the presence of 30% sucrose) (**Fig. 5B**, right panel). For each EPR spectrum, with the only exception of variant S491C that constitutes a special case of highly restricted motional regime, we measured the h(+1)/h(0). For position 491, which is the only one whose side chain is located at the XD-binding

interface, the mobility of the spin label was so dramatically reduced that the h(+1)/h(0) parameter is no longer appropriate (for further details see (Belle et al., 2008)). The effects of XD on the mobility of the spin label range from null (positions 407 and 460) to either dramatic (positions 488 – 502) or significant though less pronounced (positions 505 – 522). EPR experiments in the presence of 30% sucrose (a condition where the contribution of protein rotation to the EPR spectral line shape is reduced) allowed us to precisely map to residues 488–502, the N_{TAIL} region undergoing α-helical folding (Belle et al., 2008), in agreement with NMR data (J. M. Bourhis et al., 2005; Gely et al., 2010).

The possibility that lack of impact of XD on the spin label mobility of spin labels grafted at positions 407 and 460 reflects the inability of these spin-labeled variants to interact with XD was checked and ruled out by means of equilibrium displacement experiments. These results therefore clearly indicate the binding to XD has no impact on positions 407 and 460, consistent with a scenario where they do not establish contacts with XD and retain their disordered nature.

In a subsequent study, we used the information on spin-label mobility gathered in the aforementioned study as constraints to obtain an atomistic ensemble description of XD in complex with the N_{TAIL} region encompassing residues 488-525 (Kavalenka et al., 2010).

By comparing measured and modeled temperature-dependent restrictions of the side-chain conformational spaces of 12 spin-labeled N_{TAIL} variants, we showed that the 490–500 region of N_{TAIL} is prestructured in the absence of the partner, and could quantitatively estimate the extent of the α-helical sampling of the free form. By combining motional pattern analysis and experimental SDSL EPR data we could infer global structural information and could propose an ensemble model of MeV XD bound to the 488-525 region of N_{TAIL} (**Fig. 5C**). The obtained population of structures of the complex shows that while the Box2 region adopts a similar conformation in all the structures, the C-terminal region of N_{TAIL}, including Box3, samples a much wider conformational space. It therefore conserves a significant degree of freedom even in the bound form. The latter finding indicates that the fuzziness of N_{TAIL} is not restricted to the region preceding the α-MoRE and rather pertains also the C-terminal region. Therefore, in the MeV N_{TAIL}-XD complex N_{TAIL} is bound to XD through its α-MoRE, the latter being

flanked by two fuzzy appendages, of which the N-terminal one is much longer.

SDSL EPR spectroscopy studies were also carried out to investigate the partner-induced folding of *Henipavirus* N_{TAIL} domains (Martinho et al., 2013). In this case 6 positions were targeted for spin labeling, one within Box1, one within Box2, three within Box3 and one within Box4 (see Fig. 1 in (Martinho et al., 2013)). Spectra of spin-labeled variants either alone or in the presence of a molar excess of XD were recorded in the presence of 30% sucrose. Spectra were then simulated using the EPR-SIM-C software program (Strancar et al., 2005). Spectral simulation allows extracting quantitative information on spin label mobility from EPR spectra (Budil, Lee, Saxena, & Freed, 1996). In particular, it gives access to the rotational correlation time τ_c, a dynamic parameter, as well as to Ω, a normalized geometric parameter that represents the free rotational space, which varies from 0 (totally restricted movement) to 1 (totally unrestricted movement). The simulation also allows an EPR spectrum to be decomposed into possibly different components and to estimate their relative contribution, thereby allowing possible structural heterogeneity (i.e. polymorphism) to be appreciated. While in the case of HeV, XD has no effect on the mobility of the spin label grafted within Box2, in the case of NiV the spectrum observed in the presence of XD can be decomposed in three spectral components, two of which reflect a reduced mobility and one corresponds to a fraction of unbound N_{TAIL}. In the case of N_{TAIL} variants bearing a spin label grafted within Box3, the conformational heterogeneity of the bound form is even higher, being particularly pronounced in the case of NiV. This polymorphism reflects the occurrence of multiple conformations of the side chain onto which is grafted the spin label. Notably, for both HeV and NiV, no XD impact was detected on the mobility of spin labels grafted within Box1 and Box4, arguing for lack of direct interaction of these sites with XD and supporting their disordered nature in the bound form. Therefore, as in the case of MeV, these studies unveiled that the N- and C-terminal regions of N_{TAIL} remain fuzzy in the complex.

2.5. Electrospray ionization mass spectrometry (ESI-MS)

This techniques relies on the generation of gas-phase ions through non-denaturing electrospray ionization (ESI) and on analysis of the charge

state distribution (CSD) by mass spectrometry (MS). Protein CSDs can deliver important structural information, thanks to the fact that protein compactness in the original solution has a strong influence on the extent of protein ionization under electrospray conditions. The higher the structural compactness, the lower the average net charge that will be observed for any given protein (Kaltashov, Bobst, & Abzalimov, 2013). The average charge state of each component yields an estimate of solvent-accessible surface area (SASA) for the structure in solution, at the moment of transfer to the gas phase.

MS is a powerful method for the analysis of complex mixtures, as it allows detection not only of distinct masses, but also of different conformers endowed with variable degrees of compactness in the molecular ensemble (Kaltashov et al., 2013). In addition, it also allows poorly populated states to be detected. Because of this exquisite sensitivity and of its ability to directly assess species distributions, without averaging over the molecular population, ESI-MS is a valuable tool for IDP analysis (Beveridge, Chappuis, Macphee, & Barran, 2013). Ion mobility (IM) spectrometry is a separation technique that is increasingly used in combination with mass spectrometry (IM-MS) to measure gas-phase conformation of desolvated molecules (Stuchfield & Barran, 2018). The mobility of an ion is a measure of its velocity in a neutral buffer gas under the influence of a weak electric field. The number of collisions with the gas, together with the charge and shape of the ion, lead to an elution profile as an arrival time distribution, which can be used to determine the rotationally averaged collision cross-section (CCS) of the selected ion. The CCS serves as an indicator of the shape of the molecule in the gas phase (Stuchfield & Barran, 2018). ESI-IM-MS therefore allows separating ions based on their compactness, thereby unveiling possible conformational heterogeneity for a given CSD. One technical limitation of ESI-MS is that samples have to be dissolved in volatile buffers (such as ammonium acetate or formate) without salts, a condition that is not suitable for all proteins and can lead to precipitation. One of the advantages of this approach is that it requires relatively small amounts of protein (i.e. in the µg range).

We have used ESI-MS combined with IM techniques to investigate the MeV N_{TAIL}-XD complex (D'Urzo et al., 2015). Prior to infusion, equimolar amounts of the two proteins were mixed and incubated for 10

minutes at room temperature without agitation. The ESI-MS spectrum, as obtained upon infusing a mixture at 20 μM of each protein, features peaks corresponding to the free proteins along with peaks specific to the N_{TAIL}-XD complex (**Fig. 6A**). These results indicate that the complex, which relies on hydrophobic interactions, is preserved, at least partially, during the electrospray process and gas-phase ion separation. The measured mass of the complex perfectly matches the value expected for a 1:1 complex. The complex disappears upon applying denaturing solvent conditions or high declustering potentials (D'Urzo et al., 2015).

The CSD of the complex-specific peaks is quite broad and bimodal, indicating the coexistence of at least two distinct conformational states. The high-charge component of the complex (18+) was assumed to correspond to an "open" conformation, in which the disordered arms of N_{TAIL}, upstream and downstream to the α-MoRE, fluctuate maintaining high solvent accessibility. *A contrario*, the low-charge component (11+), probably represents a compact or "closed" conformation of the complex, in which the N_{TAIL} arms are packed on the surface of the partner, as suggested by the low extent of ionization, which approaches the expected value for folded globular proteins of the same mass (9.46+) (Testa, Brocca, & Grandori, 2011). The average SASA values of the "open" and "closed" states, as calculated from the empiric relation with the average charge state (Testa et al., 2011), are 16,100 $Å^2$ and 9500 $Å^2$, respectively. Notably, the closed state escaped detection in previous studies, thus raising the question as to whether this conformation reflects a real conformational heterogeneity in solution, captured only by the exquisite sensitivity of native MS, or an altered conformational ensemble induced by the electrospray conditions. The fact that the CSDs of IDPs, as detected by nano-ESI-MS, and their response to solvent conditions are protein specific (Testa et al., 2011), advocate in favor of the former hypothesis.

In order to get further insights onto the compact state of the complex, we carried out IM-MS experiments using a Waters Synapt G2 HDMS instrument (D'Urzo et al., 2015). The IM-MS spectrum obtained under the same conditions described above, is shown in **Fig. 6B**. The CSDs are similar to those observed with the QSTAR Elite instrument, although with higher relative amounts of the compact form. The additional separation of ions afforded by IM allowed detection of distinct structural species even at the

same charge state (**Fig. 6C**). Two slightly different conformations are found to populate the compact state of the N_{TAIL}–XD complex (charge states 8+ to 10+). Average CCS of 1326 and 1422 $Å^2$ can be derived for the two main peaks detectable for the 8+ charge state (**Fig. 6C**). Thus the compact state of the complex displays further structural heterogeneity: although its overall ionization suggests a collapsed structure, its arrival-time distribution reveals distinct peaks rather than the single peak typically observed for folded proteins. The presence of multiple conformers, although with different CCS values, is also observed for the 9+ and 10+ ions (**Fig. 6C**). The predominant conformer displays progressively larger CCS values (1466, 1607 $Å^2$) with increasing charge state (**Fig. 6C**). This trend is consistent with the lowest charge state (8+) corresponding to the most compact species in the original ensemble and to the species with lowest interference of Coulomb repulsions upon transfer to the gas phase. Altogether, these results suggest that both the open and closed states of the complex consist of conformational ensembles.

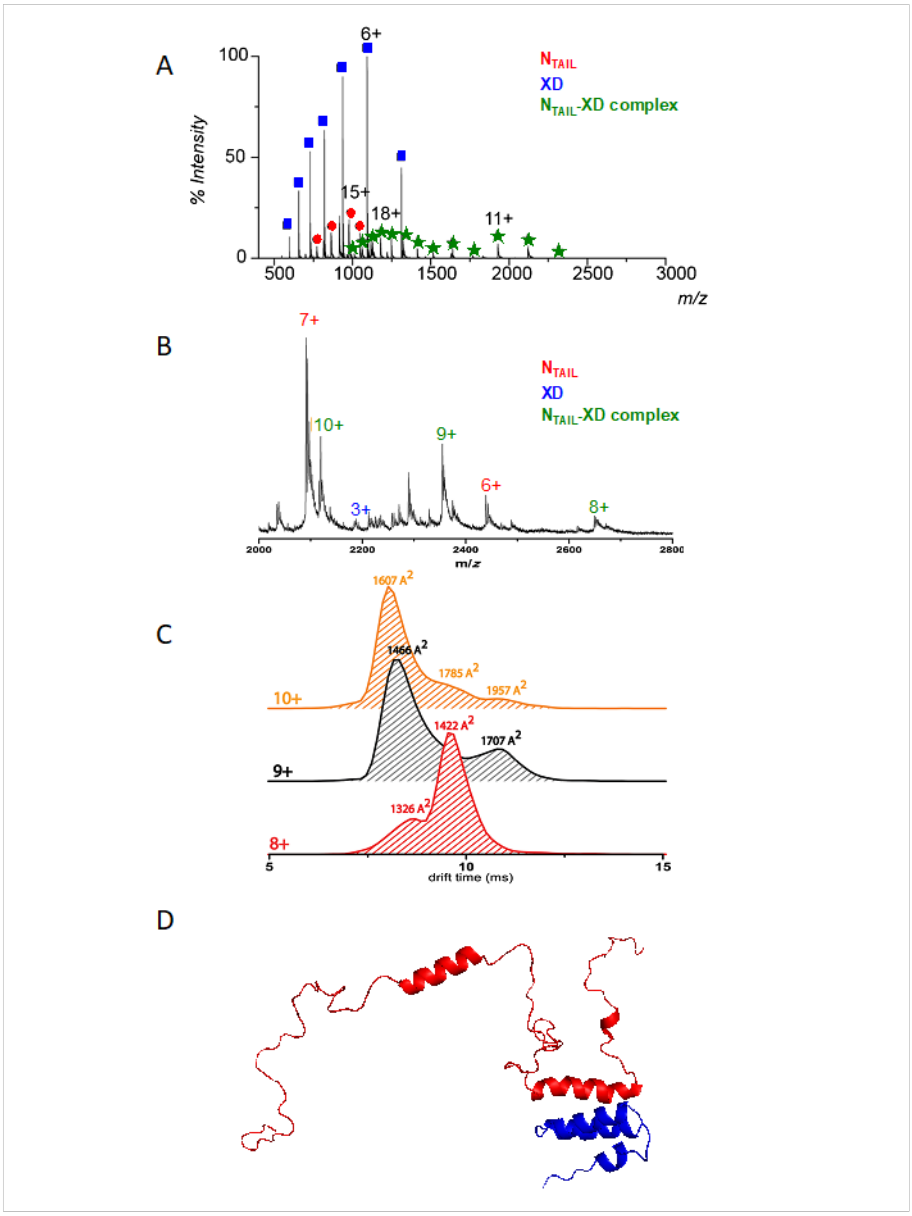


Figure 6. (**A**) Nano-ESI-MS spectrum of an equimolar mixture of MeV

N_{TAIL} and XD at 20 μM each in 10 mM ammonium acetate, pH 6.5. The spectrum was recorded using an AB Sciex QSTAR Elite instrument. (**B**) Excerpt of the higher m/z region of the nano-ESI-IM-MS spectrum of a MeV N_{TAIL} and XD mixture (same conditions as in panel A) recorded using a Waters Synapt G2 HDMS instrument. (**C**) Arrival times of the low charge (8–10+) states of the N_{TAIL}-XD complex. The 8–10+ charge states of complex show at least two different conformations, with corresponding CCSs values that vary over a wide range. (**D**) Cartoon representation of a representative structural model of the N_{TAIL}-XD complex selected from the conformational ensemble based on best agreement with the chemical shifts from NMR and the calculated SASA from ESI-MS. XD is shown in blue and N_{TAIL} in red. Data were taken from (D'Urzo et al., 2015).

Next we generated structural models of the N_{TAIL}–XD complex by combining homology modeling and PROFASI (D'Urzo et al., 2015). The latter is an all-atom Monte Carlo-based simulation tool (Jonsson, Mohanty, & Irback, 2012) that allows exploring the conformational ensemble of the complex in implicit solvent. The quality of the conformational sampling was assessed based on the agreement between calculated and experimental chemical shifts data of N_{TAIL} bound to XD (Gely et al., 2010). A quite good agreement was also found between the calculated CD spectra of the conformational ensemble and the experimental ones (Shu et al., 2012). Finally, the SASA mean value calculated from the models is 17,920±690 $Å^2$, a value in good agreement with that of the open state of the complex detected by ESI-MS. **Fig. 6D** shows one representative model selected from the conformational ensemble based on best agreement with the experimental SASA. Analysis of the contacts established between the fuzzy region with XD in this model and other models equally well fitting the experimental SASAs, unveiled their prevalently hydrophobic nature (D'Urzo et al., 2015).

While previous atomic models only concerned the α-MoRE region of N_{TAIL}, this study represents the first attempt at modeling the entire N_{TAIL}–XD complex at atomic resolution. By including the disordered arms in the molecular modeling, it was possible to extend interaction analysis to the dynamic regions of the complex, thereby setting the basis for targeting specific conformations of the complex by small-molecule inhibitors.

In conclusion, IM-MS data indicate that the compact state of the complex is characterized by high structural heterogeneity in the gas phase. The broad arrival time distributions observed for the low charge-state component could be a distinct feature of collapsed states of IDPs, as opposed to globular structures. Additional studies are however required to rule out possible artifacts induced by the electrospray process.

3. ASSESSING THE IMPACT OF THE LONG, N-TERMINAL N_{TAIL} FUZZY APPENDAGE ON BINDING

3.1. Deletion studies coupled to protein complementation assays

In section 2, we have described various methods allowing assessment of the presence of residual disorder in paramyxoviral N_{TAIL}-XD complexes. These studies allowed us to divide MeV N_{TAIL} (aa 401-525) into three regions: a fuzzy N-terminal region (aa 401-485), the α-MoRE (aa 486-504) and a fuzzy C-terminal region (aa 505-525).

Here we describe how we have assessed the impact of the N-terminal fuzzy region of N_{TAIL} on the interaction of the latter with two known binding partners, namely XD and the major inducible heat shock protein 70 (hsp70). To that end, we have serially truncated the N-terminal fuzzy region of N_{TAIL} from residue 401 to residue 481 by 10 residues intervals (see **Fig. 7**).

Among the various possible approaches to quantitatively assess protein-protein interactions (i.e. ITC, microscale thermophoresis, surface plasmon resonance, and biointerferometry), we have chosen to evaluate the impact of these truncations on the binding with XD and with hsp70 using a protein complementation assay based on the use of the green fluorescent protein (GFP) (Gruet et al., 2013; Wilson, Magliery, & Regan, 2004). GFP is a β-barrel made of 11 β strands mainly linked by loops. By cutting the loop linking β-strands 7 and 8, GFP can be split into two uneven halves made of the first 7 (NGFP) and the last 4 (CGFP) β-strands, respectively. Separately, NGFP or CGFP are unable to fluoresce. After splitting GFP into NGFP and CGFP, two proteins X and Y, known to interact with each other, can be fused to the C-terminal end of NGFP and to the N-terminal end of CGFP, respectively. When NGFP-X and Y-CGFP fusion proteins are co-expressed in *E. coli*, X and Y bind to each other within the cell, which also brings their

respective fusion partners NGFP and CGFP in close proximity. This allows NGFP and CGFP moieties to reconstitute the full-length GFP, which is fluorescent in contrast with its isolated NGFP and CGFP constituents. Thus, the interaction between two proteins X and Y (in both their *wild-type* or mutated form) can be visualized by the fluorescence of bacteria in which NGFP-X and Y-CGFP are co-expressed. Since the fluorescence is proportional to the affinity between X and Y, the interaction between different combinations of NGFP-X and Y-CGFP can be compared by simply measuring the fluorescence of the bacteria co-expressing NGFP-X and Y-CGFP. Of course, one has to rule out the possibility that observed fluorescence variations may reflect differences in expression between the variants under study. It is therefore mandatory to check expression by SDS-PAGE analysis.

In our case, X was N_{TAIL} or its truncation variants (**Fig. 7**) and Y was either XD (**Fig. 7A**) or hsp70 (**Fig. 7B**). For practical reasons, we have modified the pET11a-link-NGFP plasmid originally created by Magliery *et al.* (Magliery et al., 2005) by replacing the cloning site made of restriction sites with a Gateway® recombination cassette (Invitrogen). This allows an easier sub-cloning of the different X coding sequences under study. A detailed description of this new vector (pNGG) can be found in (Gruet, Longhi, & Bignon, 2012). For the expression of Y-CGFP, the original plasmid (pMRBAD-link-CGFP) was used, in which XD and hsp70 coding sequences were inserted by restriction cloning. The two plasmids can be co-maintained in the same bacterial cell because they have compatible origins of replications (ColE1 for pNGG and p15A for pMRBAD-link-CGFP), and because they confer a resistance to different antibiotics (ampicillin for pNGG and kanamycin for pMRBAD-link-CGFP). In these constructs, expression of NGFP-X is under the control of a T7 promoter inducible by IPTG and expression of Y-CGFP is under the control of the arabinose promoter inducible by arabinose.

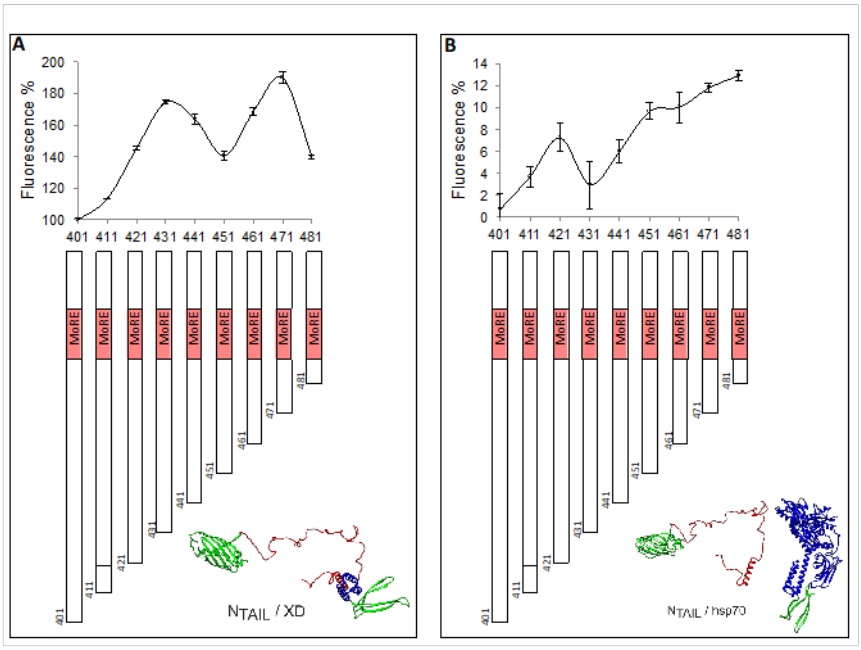


Figure 7. Split-GFP complementation assay using MeV N_{TAIL} truncation variants and MeV XD (A) or hsp70 (B). *E. coli* T7-pRos (*i.e.*, T7 cells (New England Biolabs) bearing the pLysS plasmid from Rosetta(DE3) pLysS cell (Novagen)) were co-transformed with plasmid pNGG bearing the coding sequence of either full-length N_{TAIL} (401) or its truncation variants (411 to 481), and plasmid pMRBAD-link-CGFP bearing the coding sequence of either XD (**A**) or hsp70 (**B**), and then plated on ampicillin- and kanamycin-containing agar plates. The next day, several colonies were scraped off the plate and used to seed 4 ml of Luria Bertani medium containing ampicillin, kanamycin and chloramphenicol in 24-wells deep-wells. Chloramphenicol is used to maintain the pRos plasmid. After one night at 37°C under 200 rpm shaking, 400 µl of pre-culture were used to seed each of 3 x 4 ml of terrific broth containing the same antibiotics. Cells were grown for an additional 2.5 hours at 37°C. IPTG (0.5 mM) and arabinose (2%) were added to each 4 ml culture and the deep-well was incubated over night at 17°C under shaking. The next day, deep-wells were spun for 5' at 4000 g in a swinging rotor and the supernatant was discarded. Cell pellets were resuspended in 1 ml of PBS and 10 µl of each cell suspension were diluted in 100 µl PBS in each well of a 96-wells clear bottom black plate. The optical density at 600 nm (i.e., the

number of cells) (OD_{600}) and GFP fluorescence (Fluo) of each well were measured using a TECAN GENios Plus spectrofluorimeter. The fluorescence value of each well was divided by the OD_{600} of the same well ($Fluo/OD_{600}$), and the mean value of each triplicate and standard deviation were calculated. Results are expressed in percentage (with 100% being the value provided by full-length N_{TAIL}). Below the X-axis is represented a scheme of each variant. At the bottom right is a cartoon of each interacting pair. Data were taken from (Gruet et al., 2016).

In practice, the N_{TAIL} coding sequence or that of its truncation variants (**Fig. 7**) was sub-cloned in pNGG and co-expressed in the *E. coli* strain T7-pROS with either the coding sequence of XD (**Fig. 7A**) or hsp70 (**Fig. 7B**) that had been sub-cloned in pMRBAD-link-CGFP (Gruet et al., 2016). Expression was induced by addition of 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 2% arabinose, and the cell fluorescence was measured after an overnight incubation at 17°C in a shaking incubator. In order to compensate possible differences in cell growth between cultures, the fluorescence is divided by the number of cells in the considered culture (i.e., by dividing the fluorescence by the optical density (OD) at 600 nm (OD_{600})). Each experimental point was performed in triplicate. Results are reported in **Fig. 7**. As one can see, the fluorescence increases with the truncation in a non-monotonic manner, with this being observed with both XD (**Fig. 7A**) and hsp70 (**Fig. 7B**). However, the overall fluorescence was higher for XD than for hsp70 (see the different Y-axis scales in **Fig. 7**), a finding in line with the known higher affinity of N_{TAIL} for XD (3 μM) (Dosnon et al., 2015) than for hsp70 (70 μM) (Couturier et al., 2010). In conclusion, the fuzzy N-terminal region of N_{TAIL} (aa 401-485) has a negative effect on the binding of N_{TAIL} with two binding partners that differ in both size and affinity. Interestingly, we obtained identical results when N_{TAIL} and XD from NiV and HeV were used or when another protein complementation assay based on split-luciferase (Cassonnet et al., 2011) and not on split-GFP was used. Thus, the negative effect of the fuzzy N-terminal region of N_{TAIL} (aa 401-485) on the binding seems to be a feature shared by at least three

para yxoviruses and does not depend on the assay used (Gruet et al., 2016).

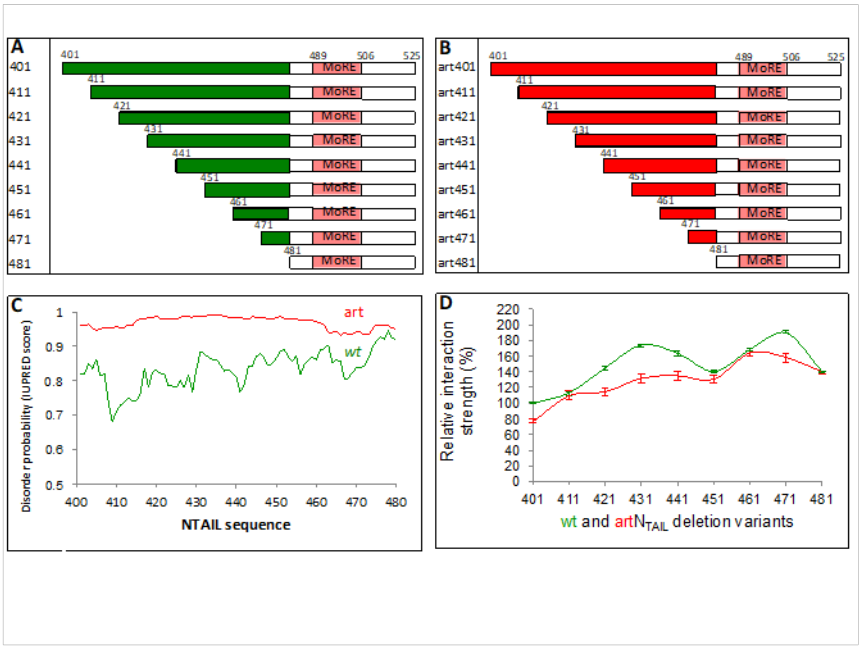


Figure 8. Split-GFP complementation assay using *wild-type* and artificial MeV N_{TAIL} truncation variants and MeV XD. *Wild-type* (wt) (**A**) and artificial (art) (**B**) MeV N_{TAIL} truncation variants. In A and B, numbers above the schemes indicate the first and last residue of the variant. (**C**) IUPred disorder prediction of wt (green line) and art (red line) N_{TAIL} region encompassing residues 400-480. (**D**) Binding values obtained by split-GFP complementation assays using wt (green line) and art (red line) MeV N_{TAIL} truncation variants. See legend of Fig. 7 for experimental details. Data were taken from (Gruet et al., 2016).

We next investigated the possible reasons of this negative effect. As we have seen in sections 2.2 to 2.4, the N-terminal region of N_{TAIL} (aa 401-485) is fuzzy, i.e. it remains disordered after binding. The first possible reason for the observed negative effect of this region on binding could be its sole disordered nature. If this were the case, then swapping the *wild-type* sequence of the fuzzy appendage of N_{TAIL} with another unrelated sequence is expected to lead to the same dampening effect provided that it is fuzzy. To

test this hypothesis, we replaced the *wild-type* fuzzy N-terminal region of N_{TAIL} (aa 401-480) with another non-natural sequence (compare **Fig. 8A** and **8B**). Compared to its *wild-type* counterpart, this artificial sequence (i) has the same number of residues, (ii) is predicted to be slightly more disordered (**Fig. 8C**), (iii) shares only 6% identity. This artificial sequence was fused with the remaining part (481-525) of the sequence of *wild-type* N_{TAIL} to reconstitute an artificial full-length N_{TAIL} (aa 401-525) (artN_{TAIL}, **Fig. 8B**). We then generated the same truncation variants as those we had generated from the *wild-type* sequence (**Figure 8B**) and compared their effect on the binding to XD. As shown in **Fig. 8D**, wtN_{TAIL} and artN_{TAIL} truncation variants feature similar binding patterns, with the binding strength increasing with the truncation. However, two differences can be noted. First, the artificial variants curve is more linear than that of the *wild-type* variants. Secondly, for the same truncation each artificial variant systematically provides a lower binding than its *wild-type* counterpart. One possible explanation for this behavior could be the higher disorder probability of the artificial sequence compared to that of the *wild-type* sequence (**Fig. 8C**). In conclusion, the negative effect of the fuzzy N-terminal region of N_{TAIL} (aa 401-485) on the binding to XD is not related to its specific sequence but to its fuzzy nature.

3.2. Kinetics studies

We further investigated the molecular mechanisms by which the fuzzy appendage of MeV N_{TAIL} influences the interaction with XD by characterizing binding kinetics. The binding between MeV N_{TAIL} and XD is an extremely rapid reaction that occurs in the μs time scale and, therefore, escapes classical stopped-flow mixing methods (dead time of about 1-2 ms). Therefore, to study the kinetics of interaction between N_{TAIL} and XD we used the temperature jump method, a kinetic technique that allows measuring ultra rapid reactions. The temperature jump methodology is a relaxation method whereby the equilibrium of a solution containing a mixture of the interacting proteins is perturbed by a rapid increase in temperature. After trigger, the relaxation to the new equilibrium is then observed. Rapid heating is due to the electrical discharge in the solution of high voltage of tens of kV, which is capable of inducing an increase in temperature of typically 5-9°C within few microseconds.

In the case of N_{TAIL} and XD we measured the change in the tryptophan fluorescence upon binding in one of the two proteins. In particular, a pseudo-*wt* tryptophan variant of XD, Y480W, was produced. The tryptophan substitution was tested and it was demonstrated that it does not affect the thermodynamic stability and structure of XD nor the stability of the complex with N_{TAIL} (Dosnon et al., 2015).

The experiments were performed under pseudo-first order conditions, with one of the two interacting proteins in large excess. In these conditions the reaction rate will depend solely on the concentration of the reactant present at low concentration, which simplifies quantitative analysis of the data. In practice a constant concentration of pseudo-*wt* Y480W XD (2 μM) was mixed with different N_{TAIL} concentrations usually ranging from 2 μM to 50 μM. The buffer used was 10 mM sodium phosphate pH 7.0, 150 mM NaCl. The presence of salt is required to ensure conductivity. Each sample was degassed and slowly pumped through the 0.5x2 mm quartz flow cell of a Hi-Tech PTJ-64 capacitor-discharge Temperature-jump apparatus (Hi-Tech, Salisbury, UK). In the cell, a rapid discharge of 12kV leads to a temperature jump of 9°C (from 11 to 20°C). The change in tryptophan fluorescence, caused by the perturbation of the equilibrium and the consequent relaxation to the initial temperature, is followed as a function of the time. Usually 10-20 traces were averaged for each sample. The excitation wavelength was 296 nm and the fluorescence emission of the tryptophan was measured using a 320 nm cut-off glass filter. The fluorescence change of *N*-acetyl-tryptophanamide was used in control measurements.

The experiment was performed using the full-length N_{TAIL} and a synthetic peptide designed to mimic just the MoRE region of N_{TAIL} (residues 485-506), in order to observe whether removal of the fuzzy region influences the rate of folding of N_{TAIL}. The full-length N_{TAIL} and the tryptophan variant of XD were expressed and purified as described in (Dosnon et al., 2015), the peptide mimicking the MoRE region was purchased from JPT (Berlin, Germany).

In all cases, the reaction followed a single exponential decay of the fluorescence and all the traces were satisfactorily fitted by the following single-exponential equation:

$$F(t) = \Delta F \cdot e^{(k_{obs} \cdot t)} + F_{\infty}$$

(Eq. 3)

where ΔF is the amplitude of the trace, k_{obs} is the observed rate constant of the reaction, t is the time of acquisition and F_{∞} is the final value of the fluorescence. We collected the observed rate constants (k_{obs}) for each N_{TAIL} or MoRE concentrations and we plotted them as a function of N_{TAIL} or MoRE concentrations.

In the case of full-length N_{TAIL}, a hyperbolic dependence of the k_{obs} on ligand concentration was observed and ascribed to a change in the rate-limiting step, with the folding of N_{TAIL} becoming rate-limiting at high reactant concentrations (Dosnon et al., 2015). Conversely, when a truncated synthetic peptide mimicking the isolated MoRE was used, a linear dependence was observed (**Fig. 9**). This behavior suggests that the fuzzy appendage of N_{TAIL} has a direct effect on the folding of the MoRE, possibly lowering its rate constant of folding (Gruet et al., 2016). The precise underlying molecular mechanisms remain however to be established.

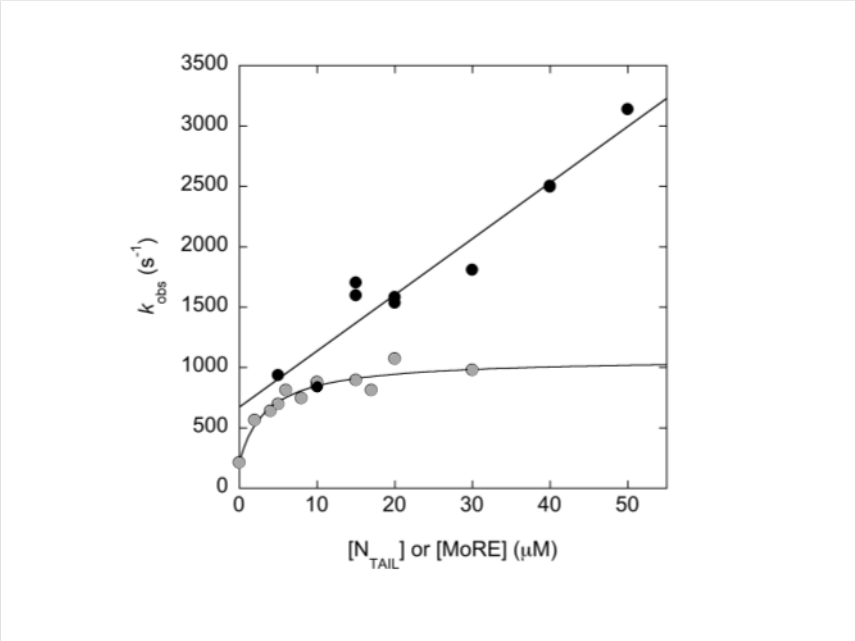


Figure 9. Pseudo-first order kinetics of the binding between Y480W XD (at a constant concentration of 2 μM) with excess concentrations of either

wtN_{TAIL} (grey circles) or a peptide mimicking the MoRE (black circles) in 10 mM sodium phosphate buffer and 150 mM NaCl at pH 7.0. Under all conditions, there was an at least fivefold difference in concentration between the two proteins to ensure pseudo-first order conditions. Experiments were carried out using a PTJ- 64 capacitor-discharge T-jump apparatus (Hi-Tech, Salisbury, UK). The temperature was rapidly changed with a jump size of 9 °C, from 11 °C to 20 °C. For both N_{TAIL} and MoRE peptide, the observed rate constant increases with increasing concentration of reactants. The data obtained with the MoRE were fitted to a linear function, while those obtained with N_{TAIL} were fitted to a hyperbolic function. The binding reaction with N_{TAIL}, but not that with the MoRE, appears to be limited, at high reactant concentrations, by a monomolecular step of ~ 10^3 s^{-1}. Data were taken from (Gruet et al., 2016).

4. ASSESSING PARTNER-MEDIATED POLYMORPHISM OF MEASLES VIRUS N_{TAIL}

4.1. Alanine-scanning and rational mutagenesis coupled to protein complementation assay

4.1.1. Alanine-scanning

In section 3.1, we have seen that the fuzzy N-terminal region of MeV N_{TAIL} (401-485) has the same negative effect on the binding to two partners of N_{TAIL} in spite of their profound differences. Indeed, XD is a small viral protein (Johansson et al., 2003) that binds the MoRE with a relatively high affinity (3 μM) (Dosnon et al., 2015), whereas hsp70 is a large cellular protein that binds the MoRE with a much lower affinity than XD (70 μM) (Couturier et al., 2010). Albeit the major hsp70-binding site had been already mapped to the MoRE (aa 486-504) (Couturier et al., 2010; Zhang et al., 2005), no structural information was available for the N_{TAIL}-hsp70 complex. In particular, it remained to be established whether the MoRE undergoes α-helical folding also upon binding to hsp70 and whether the interaction relies on the same N_{TAIL} residues. The relevance of investigating the molecular mechanisms governing the N_{TAIL}-hsp70 interaction is related to its well-documented functional impact on viral transcription and replication (Zhang et al., 2005; Zhang et al., 2002) and on the innate immune response (Kim et al., 2013).

We therefore have undertaken a study aimed at analyzing the contribution of each residue of the MoRE to binding to both XD and hsp70. To that end, we have used the alanine scanning technique. Alanine scanning consists of systematically and individually replacing each residue of a considered sequence with an alanine residue (or a glycine residue if the *wild-type* residue is already an alanine), and then assessing the effect of these substitutions on the properties of the protein. We have thus alanine scanned each residue of the MoRE from residue 486 to residue 504, and tested the effect of these substitutions on the binding of each individual variant using the split-GFP complementation assay described in section 3.1. Note that we used N_{TAIL} truncation variant 471 (aa 471-525) as backbone to derive single-site variants because it binds XD better than full-length N_{TAIL} (**Fig. 7A**). Results are reported in **Fig. 10**. In the case of XD binding (**Fig. 10**, upper panel), most of the single alanine variants exhibit a decreased binding compared to that of the *wild-type* sequence. In a few cases, and in perfect agreement with the known 3D structure of the MeV XD/α-MoRE complex (Kingston, Hamel, et al., 2004), the single alanine substitution partially or totally abrogated binding (residues 491, 494, 495, 501). These residues can therefore be considered as critical for XD binding. A quite different profile was obtained with hsp70 (**Fig. 10**, lower panel). First, many substitutions resulted in an increased binding compared to the wild-type sequence. Secondly, no single residue proved to be mandatory for binding hsp70. In conclusion, although XD and hsp70 exhibit the same sensitivity to the negative effect of the fuzzy N-terminal region of N_{TAIL} (aa 401-485) (**Fig. 7**, and section 3.1), they bind the MoRE of N_{TAIL} using different mechanisms.

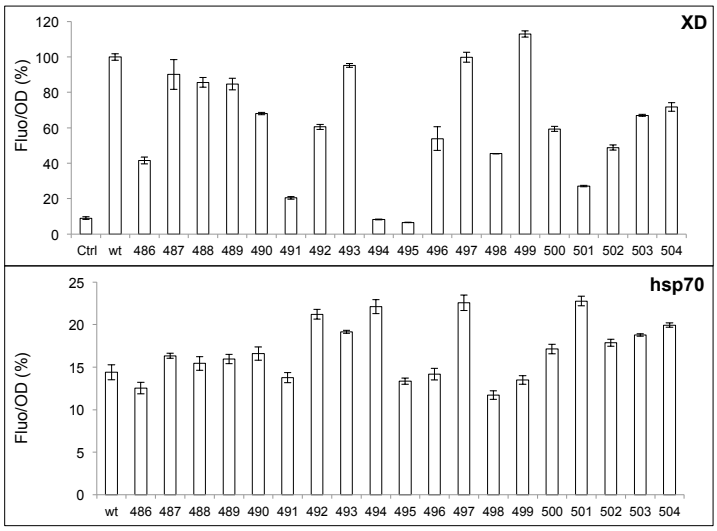


Figure 10. Alanine scanning mutagenesis of MeV N_{TAIL} MoRE. The MoRE residues (aa 486-504) of MeV N_{TAIL} truncation variant 471 (see Fig. 7) were individually mutated into an alanine residue (or a glycine residue when the *wild-type* residue was already an alanine). The binding ability of each single variant was then compared to that of the *wild-type* sequence by split-GFP complementation assay using either XD (**A**) or hsp70 (**B**). Ctrl, negative control (fluorescence background as obtained using empty pNGG vector without N_{TAIL}); wt, *wild-type* truncation variant 471. See legend of Fig. 7 for experimental details. Data were taken from (Bignon et al., 2017).

These results also suggest that the MoRE of N_{TAIL} likely adopts a different conformation at the surface of XD. The latter hypothesis was further investigated using rational mutagenesis as described in the next paragraph.

4.1.2. Rational mutagenesis

We know that the MoRE of N_{TAIL} folds as an α-helix upon XD binding (see sections 2.3 and 2.4) but nothing is known about the conformation that it adopts upon binding to hsp70. Based on the results

discussed in section 4.1.1. it is plausible that the MoRE does not fold into an α-helix once bound to hsp70.

To test this hypothesis, we conceived and generated two MoRE variants with opposite folding capabilities. In the first one, called AlaMoRE, all the residues that proved non critical for XD binding on the basis of the alanine scanning, were replaced with an alanine. In the second one, called GlyMoRE, all the residues that proved non critical for XD binding were replaced with a glycine. Because alanine is known to favor α-helix formation whereas glycine is known to have the opposite effect (Borcherds et al., 2014; Otieno & Kriwacki, 2012; Rogers, Wong, & Clarke, 2014; Scott, Alonso, Sato, Fersht, & Daggett, 2007), AlaMoRE and GlyMoRE were expected to be more and less α-helical than wtMoRE, respectively. This assumption, reinforced by disorder prediction (**Fig. 11A**) and modeling (**Fig. 11B**), was experimentally confirmed by CD analysis (**Fig. 11C**).

WtMoRE, AlaMoRE and GlyMoRE were then tested for their ability to bind XD or hsp70 by split-GFP complementation assay as described in section 3.1. This was performed using truncation variant 471 as backbone because of its higher binding ability compared to that of full-length N_{TAIL} (see 4.1.1). Results, reported in **Fig. 11D**, unambiguously indicated that increasing the α-helicity of the MoRE (AlaMoRE) slightly increases its binding to XD compared to that of wtMoRE whereas the lack of α-helicity (GlyMoRE) results in a complete loss of binding in spite of the presence of the residues revealed to be critical for XD binding by the alanine scanning.

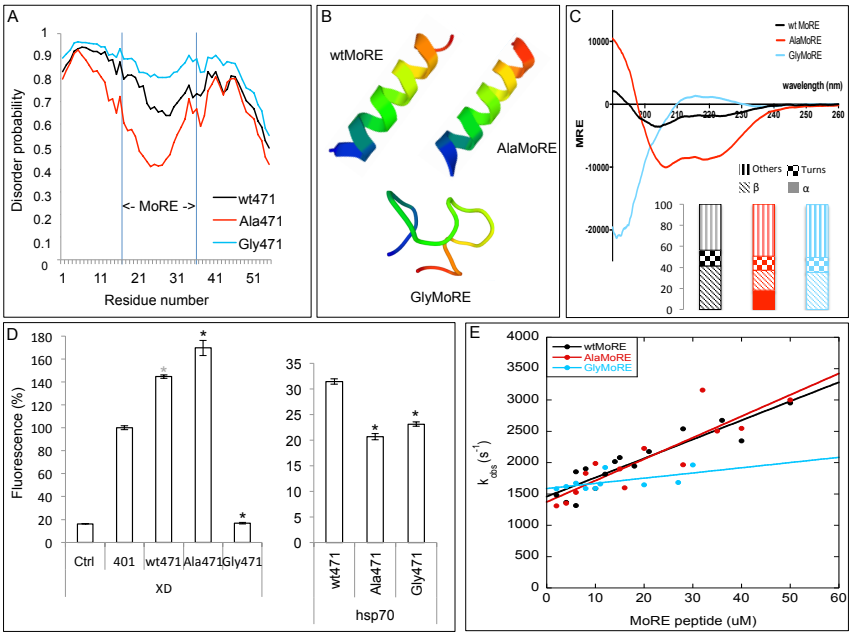


Figure 11. Binding ability of N_{TAIL} MoRE variants with different degrees of α-helicities. (**A**) IUPred disorder prediction (Dosztanyi, Csizmok, Tompa, & Simon, 2005). (**B**) Structural models as obtained using PEP-FOLD3 (Lamiable et al., 2016). (**C**) Far-UV CD spectra of MoRE peptides. The inset shows the secondary structure content as derived using BESTSEL (Micsonai et al., 2015). (**D**) Split-GFP complementation assay. On the X-axis are shown the NGFP fusions. The co-expressed CGFP fusion (XD or hsp70) used is indicated below the X-axis. Ctrl, negative control using empty pNGG vector; 401, full-length wild-type N_{TAIL}; wt471, 471 truncated variant with a wtMoRE; Ala471, 471 truncated variant with AlaMoRE; Gly471, 471 truncated variant with GlyMoRE. The black asterisk indicates a statistically significant difference (T Student's test, with $p<0.01$) with respect to wt471. The grey asterisk indicates a statistically significant difference (T Student's test, with $p<0.01$) with respect to 401. See legend of Fig. 6 for experimental details. (**E**) Pseudo-first-order kinetics of binding of MoRE peptides to XD Y480W variant. See legend of Fig. 7 for experimental details. Data were taken from (Bignon et al., 2017).

The lower XD binding ability of GlyMoRE compared to that of wtMoRE and AlaMoRE was also confirmed by kinetics experiments. In those experiments, the binding of XD to wtMoRE, AlaMoRE and GlyMoRE was measured by incubating a constant concentration (5 μM) of XD Y480W with varying concentrations of either wtMoRE or AlaMoRE or GlyMoRE (typically ranging from 2 to 50 μM). In analogy to what we described previously, in all cases (un)binding was induced by a rapid discharge of 12 kV on a quartz cell, corresponding to a rapid increase in temperature of 9°C. The fluorescence traces conformed all to a single exponential transition (see Supplementary Fig. S2 in (Bignon, Troilo, Gianni, & Longhi, 2017)). Analysis of kinetic data (**Fig. 11E**) revealed that while the behavior of AlaMoRE is close to that of wtMoRE (wtMoRE: $k_{off}^{app} = 1460 \pm 80\, s^{-1}$, $k_{on}^{app} = 30 \pm 4\, s^{-1}\mu M^{-1}$; AlaMoRE: $k_{off}^{app} = 1400 \pm 100\, s^{-1}$, $k_{on}^{app} = 34 \pm 5\, s^{-1}\, \mu M^{-1}$), there is a detectable destabilization of the complex in the case of GlyMoRE ($k_{off}^{app} = 1610 \pm 60\, s^{-1}$, $k_{on}^{app} = 5 \pm 2\, s^{-1}\mu M^{-1}$) as judged from the lower association rate constant, which is represented by the slope of the observed rate constant as a function of reactant concentration. These results indicate that, while an increase in helicity of the MoRE compared to the wt sequence does not seemingly contribute to a stabilization of the complex *in vitro*, a destabilization of the secondary structure of the MoRE is paralleled by 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 particular, it is conceivable that the differences in the concentrations used in split-GFP reassembly assays and in kinetic studies may be responsible for the observed discrepancies.

By contrast with what was observed with XD, AlaMoRE and GlyMoRE behaved similarly when assessed for their binding to hsp70: they both exhibited a partial decreased binding compared to that of wtMoRE. These results definitely indicate that XD and hsp70 do not rely on the same structural requirements to bind to the α-MoRE of N_{TAIL}. More specifically, increasing the α-helicity of the α-MoRE increases XD binding but decreases hsp70 binding suggesting that hsp70 binding does not induce α-helical folding of the α-MoRE. This conclusion is strengthened by the relative insensitiveness of hsp70 to a MoRE displaying a reduced α-helicity.

In conclusion, in addition to use a different set of N_{TAIL} residues (see section 4.1.1), XD and hsp70 do not induce and/or require the same folding within the MoRE of N_{TAIL} and are therefore likely to interact with the latter through completely different mechanisms. The discovery that the conformation sampled by the free form of the MoRE of N_{TAIL} does not necessarily commit N_{TAIL} to adopt an α-helical conformation in the bound form, provides 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 the closely related HeV N_{TAIL} folds as anti-parallel β sheet at the air/water interface (Bénarouche et al., 2017) thereby further underscoring the extreme polymorphism of this IDP.

5. CONCLUSIONS

Conformational heterogeneity in protein complexes is quite common leading to multiplicity of interaction modes (Fuxreiter, 2018; Miskei, Antal, & Fuxreiter, 2017). The experimental characterization of fuzzy assemblies is however challenging owing to the inherent multiplicity of the species, which entails resolution, spectral broadening and averaging problems.

In this chapter, we have reviewed a variety of techniques enabling the characterization of different aspects of fuzziness. Conformational heterogeneity in the bound state could be detected by various techniques such as NMR, EPR and fluorescence spectroscopy in case of highly dynamical complexes, whereas X-ray crystallography can provide information on stable, polymorphic complexes. The shape distribution of the ensemble could be obtained by SEC, SEC-MALLS and SAXS. The high-resolution ensemble can be described by molecular simulation methods that make use of experimental data as constraints (for a recent example based on SAXS data see (Cordeiro et al., 2017)). Notably, computational methods can also provide insights into the mechanisms of molecular recognition. Some issues, such as the averaging over collective motions with different timescales, have still to be solved. Single molecule approaches are ideally suited to study fuzzy complexes as they circumvent the inherent averaging problem from which all the other approaches suffer (Borgia et al., 2018). The combination of different techniques, such as NMR and SAXS for instance (Sterckx et al.,

2014), is of course an additional step forward as it allows limiting data over fitting. Future developments will focus on achieving detection of poorly populated states that might play crucial roles in function.

All these techniques together provided a comprehensive view of paramyxoviral N_{TAIL} interactions and the role of fuzziness in N_{TAIL} function.

Amongst all the interactions established by N_{TAIL}, the interaction with XD is critical as it allows the L-P complex to be recruited onto the nucleocapsid, triggering transcription and replication (Longhi, 2007, 2009, 2011; Longhi et al., 2017). The so-called "cartwheeling" mechanism proposed for *Paramyxoviridae* posits that the polymerase complex cartwheels from one N monomer to another within the nucleocapsid in order to allow transcription and replication to take place (Kolakofsky, Le Mercier, Iseni, & Garcin, 2004) (**Fig. 12A**). Hence, the N_{TAIL}-XD interaction needs to be dynamically established and broken to ensure RNA synthesis.

The affinity of the N_{TAIL}-XD interaction was shown to be a critical parameter governing both viral messenger accumulation rates (Brunel et al., 2014) and the efficiency with which the polymerase complex reinitiates transcription at each IGR (L. Bloyet et al., 2016; Bloyet, Roche, Gerlier, & Longhi, 2017). The N_{TAIL}-XD interaction thus tightly controls viral polymerase progression along the nucleocapsid. Its interaction strength has consequently to be kept into a precise window to ensure efficient transcription and replication and to keep a balanced relative amount of the various viral proteins. The fuzziness of paramyxoviral N_{TAIL}-XD complexes provides an exquisite means to modulate the binding affinity. Dynamic binding of the binding motif and the resulting polymorphism at the surface of the partner, as observed in *Henipavirus* N_{TAIL}-XD complexes, brings an additional means for regulating the interaction strength. By tuning the extent of pre-configuration of the binding motifs and/or the length of neighboring fuzzy appendages the virus can achieve an optimal binding strength.

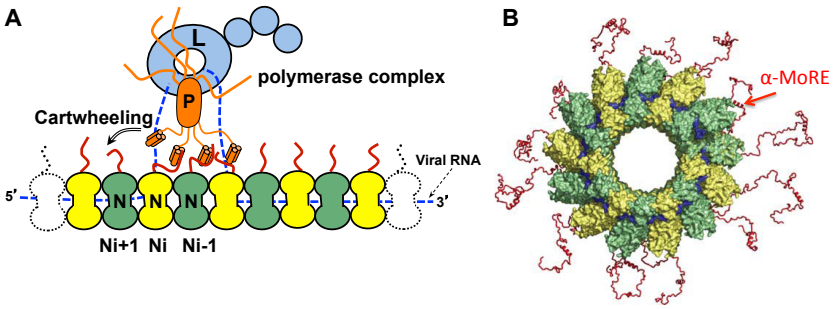


Figure 12. (**A**) Schematic representation of the P protein (orange) bound to the nucleocapsid and the L protein (light blue). IDRs are shown as lines. N_{TAIL} is shown in red, while successive N_{CORE} monomers are colored in yellow and green. The negative-sense genomic RNA (blue dashed line) is presented in the 5' to 3' orientation. Because of this, the polymerase complex is shown to cartwheel onto the nucleocapsid template from right to left. (**B**) Structural model of a single turn of the nucleocapsid made of 13 N monomers. N_{TAIL}, N_{CORE} and the RNA are colored as in panel A. For the sake of clarity, only one α-MoRE is highlighted by a red arrow. Data were taken from (Ringkjøbing Jensen et al., 2011).

The role of the fuzzy region preceding the α-MoRE in dampening the N_{TAIL}-XD interaction (Gruet et al., 2016) sheds light on why paramyxoviral N_{TAIL} domains require such a long (i.e. > 80 residues) disordered arm (**Fig. 12B**). While such a long arm might in principle be thought as affording spatial flexibility to the MoRE thereby facilitating recruitment of the L-P complex, it was recently shown that relocation of the MoRE on the structured moiety of N (i.e. N_{CORE}) is tolerated by MeV (Cox, Krumm, Thakkar, Sohn, & Plemper, 2017). In light of the observed impact of shortening the fuzzy

appendage on XD binding, it is rather tempting to speculate that in the course of evolution, the length of this region has been under selective pressure so as to ensure a "balanced" affinity towards XD. In other words, we propose that the arm has been elongated until an optimal dampening level of the α-MoRE-XD interaction was achieved. Recent data that documented enhanced transcription and concurrent reduced viral growth in a mutated MeV in which the central region of N_{TAIL} preceding the MoRE was removed (Thakkar et al., 2018), afford additional support in favor of a critical role of this region in virus replication and pathogenesis.

Fuzziness also plays an important role in tuning the interaction strength towards additional partners. The fuzzy Box3 region of MeV N_{TAIL} was shown to serve as a low-affinity binding site for hsp70 (Couturier et al., 2010; Zhang et al., 2005), a cellular protein known to stimulate viral transcription and replication (Carsillo, Zhang, Vasconcelos, Niewiesk, & Oglesbee, 2006; Oglesbee, 2007; Zhang et al., 2002), with the primary binding site having being mapped to Box2 (Carsillo, Traylor, Choi, Niewiesk, & Oglesbee, 2006; Zhang et al., 2002). Since hsp70 competitively inhibits XD binding to N_{TAIL} (Zhang et al., 2005), hsp70 could enhance transcription and genome replication by reducing the stability of N_{TAIL}-XD complexes and hence promoting P-L cartwheeling on the nucleocapsid (J. M. Bourhis et al., 2005; Zhang et al., 2005). The enhanced interaction strength with XD displayed by MeV N_{TAIL} variants devoid of Box3 may result not only from an entropic effect (like for the N-terminal fuzzy region) but also from abrogation of a destabilizing interaction between Box3 and DnaK, the *E. coli* counterpart of hsp70 (Gruet et al., 2013).

Finally, fuzzy regions flanking MoREs can also be allosterically coupled to the binding region. In line with this, a random mutagenesis study of MeV N_{TAIL} led to the identification of five regulatory regions (i.e. eBoxes) that are located in the N-terminal fuzzy region of N_{TAIL} and act as natural dampeners of the interaction (Gruet et al., 2013). The underlying molecular mechanisms remain to be elucidated.

Taken together all these results highlight that conformational heterogeneity or fuzziness in N_{TAIL} complexes, either as static or dynamic, importantly contributes to the key functions of this domain in the replicative complex of *Paramyxoviridae* members and has been optimized by evolution.

Although characterization of fuzziness is complex, it enables a more holistic view of viral protein functions and mechanisms, than the one that can be achieved through the study of shorter, more structured, constructs.

ACKNOWLEDGMENTS

S.L. wishes to thank former members of her lab who contributed to these studies: J.M. Bourhis (IBS, Grenoble, FR), J. Habchi (Dept. of Chemistry, University of Cambridge, UK) D. Blocquel (The Scripps Institute, La Jolla, CA, USA), A. Gruet (Memorial Sloan Kettering Cancer Center, New York, USA) and M. Dosnon. The authors also thank D. Gerlier (CIRI, INSERM U758, Lyon, FR), R. Pierattelli & Isabella Felli (CERM, Florence, IT), M. Blackledge & M. Ringkjøbing-Jensen (IBS, Grenoble, FR), Gary Daughdrill (University of South Florida, Tampa, FL, USA), V. Receveur-Bréchot (BIP, Marseille, FR), V. Belle & B. Guigliarelli (BIP, Marseille, FR), J. Strancar (EPR Center, Institute Josez Stephan, Ljubljana, SI), R. Grandori (University of Milano Bicocca, IT), Paolo Carloni (German Research School for Simulation Sciences and Computational Biomedicine, Jülich, D), R. Das (Dept. of Biomedical Engineering and Center for Biological Systems Engineering, Washington University in St. Louis, St. Louis, MO, USA). The studies herein reviewed were carried out with the financial support of the Agence Nationale de la Recherche, specific programs "Physico-Chimie du Vivant", ANR-08-PCVI-0020-01, and “ASTRID”, ANR-11-ASTR-003-01 to S.L. They were also supported by the Italian Ministero dell’Istruzione dell’Università e della Ricerca (Progetto di Interesse ‘Invecchiamento’ to S.G.) and by Sapienza University of Rome (C26A155S48 to S.G). They also benefited from support from the CNRS, the Direction Générale de l’Armement (DGA) and the Fondation pour la Recherche Médicale (FRM). F.T. is a recipient of a Ph.D. fellowship from the Italo-French University. This work was also financially supported by the Hungarian Academy of Sciences and the GINOP-2.3.2-15-2016-00044 project (M.F.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

Bankamp, B., Horikami, S. M., Thompson, P. D., Huber, M., Billeter, M., & Moyer, S. A. (1996). Domains of the measles virus N protein required for binding to P protein and self-assembly. *Virology, 216*(1), 272-277.

Baronti, L., Erales, J., Habchi, J., Felli, I. C., Pierattelli, R., & Longhi, S. (2015). 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, 16*(2), 268-276. doi:10.1002/cbic.201402534

Belle, V., Rouger, S., Costanzo, S., Liquiere, E., Strancar, J., Guigliarelli, B., . . . Longhi, S. (2008). 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, 73*(4), 973-988. doi:10.1002/prot.22125

Belle, V., Rouger, S., Costanzo, S., Longhi, S., & Fournel, A. (2010). Site-directed spin labeling EPR spectroscopy. In V. N. Uversky & S. Longhi (Eds.), *Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation*. Hoboken, New Jersey John Wiley and Sons.

Bénarouche, A., Habchi, J., Cagna, A., Maniti, O., Girard-Egrot, A., Cavalier, J., . . . Carrière, F. (2017). Interfacial properties of NTAIL, an intrinsically disordered protein. *Biophysical Journal, in press*.

Bernado, P., Blanchard, L., Timmins, P., Marion, D., Ruigrok, R. W., & Blackledge, M. (2005). A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc Natl Acad Sci U S A, 102*(47), 17002-17007.

Bernado, P., & Svergun, D. I. (2010). Structural Insights into Intrinsically Disordered Proteins by Small-Angle X-ray Scattering. In V. N. Uversky & S. Longhi (Eds.), *Instrumental analysis of intrinsically disordered proteins: assessing structure and conformation* (pp. 451-476). Hoboken, New Jersey John Wiley and Sons.

Bernado, P., & Svergun, D. I. (2012). Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol Biosyst, 8*(1), 151-167. doi:10.1039/c1mb05275f

Beveridge, R., Chappuis, Q., Macphee, C., & Barran, P. (2013). Mass spectrometry methods for intrinsically disordered proteins. *Analyst, 138*(1), 32-42. doi:10.1039/c2an35665a

Bhella, D. (2007). Measles virus nucleocapsid structure, conformational flexibility and the rule of six. In S. Longhi (Ed.), *Measles virus nucleoprotein*. Hauppage, NY: Nova Publishers Inc.

Bhella, D., Ralph, A., & Yeo, R. P. (2004). Conformational flexibility in recombinant measles virus nucleocapsids visualised by cryo-negative stain electron microscopy and real-space helical reconstruction. *Journal of Molecular Biology, 340*(2), 319-331.

Bignon, C., Troilo, F., Gianni, S., & Longhi, S. (2017). Partner-mediated polymorphism of an intrinsically disordered protein

. *Journal of Molecular Biology, in press*.

Blanchard, L., Tarbouriech, N., Blackledge, M., Timmins, P., Burmeister, W. P., Ruigrok, R. W., & Marion, D. (2004). Structure and dynamics of the nucleocapsid-binding domain of the Sendai virus phosphoprotein in solution. *Virology, 319*(2), 201-211.

Blocquel, D., Bourhis, J. M., Eléouët, J. F., Gerlier, D., Habchi, J., Jamin, M., . . . Yabukarski, F. (2012). Transcription et réplication des Mononégavirales: une machine moléculaire originale. *Virologie, 16*(4), 225-257.

Blocquel, D., Habchi, J., Durand, E., Sevajol, M., Ferron, F., Erales, J., . . . Longhi, S. (2014). Coiled-coil deformations in crystal structures: the measles virus phosphoprotein multimerization domain as an illustrative example. *Acta Cryst D, 70*(6), 1589-1603.

Blocquel, D., Habchi, J., Gruet, A., Blangy, S., & Longhi, S. (2012). Compaction and binding properties of the intrinsically disordered C-terminal domain of Henipavirus nucleoprotein as unveiled by deletion studies. *Mol Biosyst, 8*(1), 392-410. doi:10.1039/c1mb05401e

Bloyet, L., Brunel, J., Dosnon, M., Hamon, V., Erales, J., Gruet, A., . . . Gerlier, D. (2016). Modulation of re-initiation of measles virus transcription at

intergenic regions by PXD to NTAIL binding strength. *PLoS Pathogens, 12*(12), e1006058.

Bloyet, L. M., Roche, P., Gerlier, D., & Longhi, S. (2017). [How is transcription reinitiation governed in measles virus?]. *Med Sci (Paris), 33*(10), 843-845. doi:10.1051/medsci/20173310010

Bloyet, L. M., Welsch, J., Enchery, F., Mathieu, C., de Breyne, S., Horvat, B., . . . Gerlier, D. (2016). HSP90 Chaperoning in Addition to Phosphoprotein Required for Folding but Not for Supporting Enzymatic Activities of Measles and Nipah Virus L Polymerases. *Journal of Virology, 90*(15), 6642-6656. doi:10.1128/JVI.00602-16

Bonetti, D., Camilloni, C., Visconti, L., Longhi, S., Brunori, M., Vendruscolo, M., & Gianni, S. (2016). Identification and Structural Characterization of an Intermediate in the Folding of the Measles Virus X Domain. *Journal of Biological Chemistry, 291*(20), 10886-10892. doi:10.1074/jbc.M116.721126

Bonetti, D., Troilo, F., Toto, A., Brunori, M., Longhi, S., & Gianni, S. (2017). Analyzing the folding and binding steps of an intrinsically disordered protein by protein engineering. *Biochemistry, 56*(29), 3780-3786.

Borcherds, W., Theillet, F. X., Katzer, A., Finzel, A., Mishall, K. M., Powell, A. T., . . . Daughdrill, G. W. (2014). Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells. *Nature Chemical Biology, 10*(12), 1000-1002. doi:10.1038/nchembio.1668

Borgia, A., Borgia, M. B., Bugge, K., Kissling, V. M., Heidarsson, P. O., Fernandes, C. B., . . . Schuler, B. (2018). Extreme disorder in an ultrahigh-affinity protein complex. *Nature, 555*(7694), 61-66. doi:10.1038/nature25762

Bourhis, J., Johansson, K., Receveur-Bréchot, V., Oldfield, C. J., Dunker, A. K., Canard, B., & Longhi, S. (2004). 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 Research, 99*, 157-167.

Bourhis, J. M., Receveur-Bréchot, V., Oglesbee, M., Zhang, X., Buccellato, M., Darbon, H., . . . Longhi, S. (2005). The intrinsically disordered C-terminal domain of the measles virus nucleoprotein interacts with the C-terminal

domain of the phosphoprotein via two distinct sites and remains predominantly unfolded. *Protein Science, 14*, 1975-1992.

Bruhn-Johannsen, J. F., Barnett, K., Bibby, J., Thomas, J., Keegan, R., Rigden, D., . . . Saphire, E. O. (2014). Crystal structure of the Nipah virus phosphoprotein tetramerization domain. *Journal of Virology, 88*(1), 758-762. doi:JVI.02294-13 [pii]

10.1128/JVI.02294-13

Brunel, J., Chopy, D., Dosnon, M., Bloyet, L. M., Devaux, P., Urzua, E., . . . Gerlier, D. (2014). Sequence of events in measles virus replication: role of phosphoprotein-nucleocapsid interactions. *Journal of Virology, 88*(18), 10851-10863. doi:JVI.00664-14 [pii]

10.1128/JVI.00664-14

Budil, D. E., Lee, S., Saxena, S., & Freed, J. H. (1996). Non linear least-squares analysis of slow motion EPR spectra in one and two dimensions using a modified Levenberg-Marquardt algorithm *Journal of Magnetic Resonance Series A, 120*(2), 155-189.

Carsillo, T., Traylor, Z., Choi, C., Niewiesk, S., & Oglesbee, M. (2006). hsp72, a host determinant of measles virus neurovirulence. *Journal of Virology, 80*(22), 11031-11039. doi:JVI.01438-06 [pii]

10.1128/JVI.01438-06

Carsillo, T., Zhang, X., Vasconcelos, D., Niewiesk, S., & Oglesbee, M. (2006). A single codon in the nucleocapsid protein C terminus contributes to in vitro and in vivo fitness of Edmonston measles virus. *Journal of Virology, 80*(6), 2904-2912. doi:80/6/2904 [pii]

10.1128/JVI.80.6.2904-2912.2006

Casey, T. M., Liu, Z., Esquiaqui, J. M., Pirman, N. L., Milshteyn, E., & Fanucci, G. E. (2014). Continuous wave W- and D-band EPR spectroscopy offer "sweet-spots" for characterizing conformational changes and dynamics in intrinsically disordered proteins. *Biochemical and Biophysical Research Communications, 450*(1), 723-728. doi:10.1016/j.bbrc.2014.06.045

Cassonnet, P., Rolloy, C., Neveu, G., Vidalain, P. O., Chantier, T., Pellet, J., . . . Jacob, Y. (2011). Benchmarking a luciferase complementation assay for

detecting protein complexes. *Nature Methods, 8*(12), 990-992. doi:10.1038/nmeth.1773

nmeth.1773 [pii]

Communie, G., Crepin, T., Maurin, D., Jensen, M. R., Blackledge, M., & Ruigrok, R. W. (2013). Structure of the tetramerization domain of measles virus phosphoprotein. *Journal of Virology, 87*(12), 7166-7169. doi:10.1128/JVI.00487-13

JVI.00487-13 [pii]

Communie, G., Habchi, J., Yabukarski, F., Blocquel, D., Schneider, R., Tarbouriech, N., . . . Blackledge, M. (2013). Atomic resolution description of the interaction between the nucleoprotein and phosphoprotein of Hendra virus. *PLoS Pathog, 9*(9), e1003631. doi:e1003631

Cordeiro, T. N., Chen, P. C., De Biasio, A., Sibille, N., Blanco, F. J., Hub, J. S., . . . Bernado, P. (2017). Disentangling polydispersity in the PCNA-p15PAF complex, a disordered, transient and multivalent macromolecular assembly. *Nucleic Acids Research, 45*(3), 1501-1515. doi:10.1093/nar/gkw1183

Couturier, M., Buccellato, M., Costanzo, S., Bourhis, J. M., Shu, Y., Nicaise, M., . . . Oglesbee, M. (2010). High Affinity Binding between Hsp70 and the C-Terminal Domain of the Measles Virus Nucleoprotein Requires an Hsp40 Co-Chaperone. *J Mol Recognit, 23*(3), 301-315.

Cox, R. M., Krumm, S. A., Thakkar, V. D., Sohn, M., & Plemper, R. K. (2017). The structurally disordered paramyxovirus nucleocapsid protein tail domain is a regulator of the mRNA transcription gradient. *Sci Adv, 3*(2), e1602350. doi:10.1126/sciadv.1602350

D'Urzo, A., Konijnenberg, A., Rossetti, G., Habchi, J., Li, J., Carloni, P., . . . Grandori, R. (2015). Molecular Basis for Structural Heterogeneity of an Intrinsically Disordered Protein Bound to a Partner by Combined ESI-IM-MS and Modeling. *Journal of the American Society for Mass Spectrometry, 26*(3), 472-481. doi:10.1007/s13361-014-1048-z

DeLano, W. L. (2002). The PyMOL molecular graphics system *Proteins: Structure, Function and Bioinformatics, 30*, 442-454.

Desfosses, A., Goret, G., Farias Estrozi, L., Ruigrok, R. W., & Gutsche, I. (2011). Nucleoprotein-RNA orientation in the measles virus nucleocapsid by three-

dimensional electron microscopy. *Journal of Virology, 85*(3), 1391-1395. doi:JVI.01459-10 [pii]

10.1128/JVI.01459-10

Dosnon, M., Bonetti, D., Morrone, A., Erales, J., di Silvio, E., Longhi, S., & Gianni, S. (2015). Demonstration of a folding after binding mechanism in the recognition between the measles virus NTAIL and X domains. *ACS ChemBiol, 10*(3), 795-802.

Dosztanyi, Z., Csizmok, V., Tompa, P., & Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics, 21*(16), 3433-3434.

Drescher, M. (2012). EPR in protein science : intrinsically disordered proteins. *Top Curr Chem, 321*, 91-119. doi:10.1007/128_2011_235

Duro, N., Miskei, M., & Fuxreiter, M. (2015). Fuzziness endows viral motif-mimicry. *Mol Biosyst, 11*(10), 2821-2829. doi:10.1039/c5mb00301f

Eaton, B. T., Broder, C. C., Middleton, D., & Wang, L. F. (2006). Hendra and Nipah viruses: different and dangerous. *Nature Reviews: Microbiology, 4*(1), 23-35.

Eaton, B. T., Mackenzie, J. S., & Wang, L. F. (2007). Henipaviruses. In B. N. Fields, D. M. Knipe, & P. M. Howley (Eds.), *Fields Virology* (5th ed., pp. 1587-1600). Philadelphia: Lippincott-Raven.

Erales, J., Beltrandi, M., Roche, J., Maté, M., & Longhi, S. (2015). Insights into the Hendra virus NTAIL-XD complex: evidence for a parallel organization of the helical MoRE at the XD surface stabilized by a combination of hydrophobic and polar interactions *Biochimica and Biopysical Acta 1854*(8), 1038-1053.

Erales, J., Blocquel, D., Habchi, J., Beltrandi, M., Gruet, A., Dosnon, M., . . . Longhi, S. (2015). Order and disorder in the replicative complex of paramyxoviruses. *Advances in Experimental Medicine and Biology, 870*, 351-381.

Felli, I. C., & Pierattelli, R. (2015). *Intrinsically Disordered Proteins studied by NMR spectroscopy*. Printforce, The Netherlands: Springer.

Fersht, A. R., Matouschek, A., & Serrano, L. (1992). The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *Journal of Molecular Biology, 224*(3), 771-782. doi:0022-2836(92)90561-W [pii]

Fuxreiter, M. (2012). Fuzziness: linking regulation to protein dynamics. *Mol Biosyst, 8*(1), 168-177. doi:10.1039/c1mb05234a

Fuxreiter, M. (2018). Fuzziness in protein interactions - a historical perspective. *Journal of Molecular Biology, in press*. doi:10.1016/j.jmb.2018.02.015

Fuxreiter, M., Simon, I., Friedrich, P., & Tompa, P. (2004). Preformed structural elements feature in partner recognition by intrinsically unstructured proteins. *Journal of Molecular Biology, 338*(5), 1015-1026.

Fuxreiter, M., & Tompa, P. (2012). Fuzzy complexes: a more stochastic view of protein function. *Advances in Experimental Medicine and Biology, 725*, 1-14. doi:10.1007/978-1-4614-0659-4_1

Gely, S., Lowry, D. F., Bernard, C., Ringkjobing-Jensen, M., Blackledge, M., Costanzo, S., . . . Longhi, S. (2010). Solution structure of the C-terminal X domain of the measles virus phosphoprotein and interaction with the intrinsically disordered C-terminal domain of the nucleoprotein *J Mol Recognit, 23*, 435-447.

Gruet, A., Dosnon, M., Blocquel, D., Brunel, J., Gerlier, D., Das, R. K., . . . Bignon, C. (2016). Fuzzy regions in an intrinsically disordered protein impair protein-protein interactions. *FEBS Journal, 283*, 576-594. doi:10.1111/febs.13631

Gruet, A., Dosnon, M., Vassena, A., Lombard, V., Gerlier, D., Bignon, C., & Longhi, S. (2013). Dissecting partner recognition by an intrinsically disordered protein using descriptive random mutagenesis. *Journal of Molecular Biology, 425*(18), 3495-3509. doi:10.1016/j.jmb.2013.06.025

S0022-2836(13)00405-1 [pii]

Gruet, A., Longhi, S., & Bignon, C. (2012). One-step generation of error-prone PCR libraries using Gateway(R) technology. *Microb Cell Fact, 11*(1), 14. doi:1475-2859-11-14 [pii]

10.1186/1475-2859-11-14

Guryanov, S. G., Liljeroos, L., Kasaragod, P., Kajander, T., & Butcher, S. J. (2016). Crystal Structure of the Measles Virus Nucleoprotein Core in complex with an N-terminal Region of Phosphoprotein. *Journal of Virology, 90*(6), 2849 – 2857. doi:10.1128/JVI.02865-15

Gutsche, I., Desfosses, A., Effantin, G., Ling, W. L., Haupt, M., Ruigrok, R. W., . . . Schoehn, G. (2015). Near-atomic cryo-EM structure of the helical measles virus nucleocapsid. *Science, 348*(704-707). doi:10.1126/science.aaa5137

Habchi, J., Blangy, S., Mamelli, L., Ringkjobing Jensen, M., Blackledge, M., Darbon, H., . . . Longhi, S. (2011). Characterization of the interactions between the nucleoprotein and the phosphoprotein of Henipaviruses. *Journal of Biological Chemistry, 286*(15), 13583-13602.

Habchi, J., & Longhi, S. (2012). Structural disorder within paramyxovirus nucleoproteins and phosphoproteins. *Mol Biosyst, 8*(1), 69-81. doi:10.1039/c1mb05204g

Habchi, J., & Longhi, S. (2015). Structural Disorder within Paramyxoviral Nucleoproteins and Phosphoproteins in Their Free and Bound Forms: From Predictions to Experimental Assessment. *Int J Mol Sci, 16*(7), 15688-15726. doi:10.3390/ijms160715688

Habchi, J., Mamelli, L., Darbon, H., & Longhi, S. (2010). Structural Disorder within Henipavirus Nucleoprotein and Phosphoprotein: From Predictions to Experimental Assessment. *PLoS ONE, 5*(7), e11684. doi:10.1371/journal.pone.0011684

Habchi, J., Mamelli, L., & Longhi, S. (2012). Structural disorder within the nucleoprotein and phosphoprotein from measles, Nipah and Hendra viruses. In V. N. Uversky & S. Longhi (Eds.), *Flexible viruses: structural disorder in viral proteins* (pp. 47-94). Hoboken, New Yersey: John Wiley and Sons.

Habchi, J., Martinho, M., Gruet, A., Guigliarelli, B., Longhi, S., & Belle, V. (2012). Monitoring structural transitions in IDPs by site-directed spin labeling EPR spectroscopy. *Methods Mol Biol, 895*, 361-386. doi:10.1007/978-1-61779-927-3_21

Heggeness, M. H., Scheid, A., & Choppin, P. W. (1980). Conformation of the helical nucleocapsids of paramyxoviruses and vesicular stomatitis virus: reversible coiling and uncoiling induced by changes in salt concentration. *Proc Natl Acad Sci U S A, 77*(5), 2631-2635.

Heggeness, M. H., Scheid, A., & Choppin, P. W. (1981). The relationship of conformational changes in the Sendai virus nucleocapsid to proteolytic cleavage of the NP polypeptide. *Virology, 114*(2), 555-562.

Houben, K., Marion, D., Tarbouriech, N., Ruigrok, R. W., & Blanchard, L. (2007). Interaction of the C-terminal domains of sendai virus N and P proteins: comparison of polymerase-nucleocapsid interactions within the paramyxovirus family. *Journal of Virology, 81*(13), 6807-6816.

Hubbell, W. L., Gross, A., Langen, R., & Lietzow, M. A. (1998). Recent advances in site-directed spin labeling of proteins. *Current Opinion in Structural Biology, 8*(5), 649-656.

Jensen, M. R., Bernado, P., Houben, K., Blanchard, L., Marion, D., Ruigrok, R. W., & Blackledge, M. (2010). Structural disorder within sendai virus nucleoprotein and phosphoprotein: insight into the structural basis of molecular recognition. *Protein Pept Lett, 17*(8), 952-960. doi:BSP/ PPL/ E pub/0155 [pii]

Jensen, M. R., Houben, K., Lescop, E., Blanchard, L., Ruigrok, R. W., & Blackledge, M. (2008). Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *Journal of the American Chemical Society, 130*(25), 8055-8061. doi:10.1021/ja801332d

Johansson, K., Bourhis, J. M., Campanacci, V., Cambillau, C., Canard, B., & Longhi, S. (2003). Crystal structure of the measles virus phosphoprotein domain responsible for the induced folding of the C-terminal domain of the nucleoprotein. *Journal of Biological Chemistry, 278*(45), 44567-44573.

Jonsson, S. A., Mohanty, S., & Irback, A. (2012). Distinct phases of free alpha-synuclein--a Monte Carlo study. *Proteins, 80*(9), 2169-2177. doi:10.1002/prot.24107

Jordan, P. C., Liu, C., Raynaud, P., Lo, M. K., Spiropoulou, C. F., Symons, J. A., . . . Deval, J. (2018). Initiation, extension, and termination of RNA synthesis by a paramyxovirus polymerase. *PLoS Pathog, 14*(2), e1006889. doi:10.1371/journal.ppat.1006889

Kaltashov, I. A., Bobst, C. E., & Abzalimov, R. R. (2013). Mass spectrometry-based methods to study protein architecture and dynamics. *Protein Science, 22*(5), 530-544. doi:10.1002/pro.2238

Karlin, D., Ferron, F., Canard, B., & Longhi, S. (2003). Structural disorder and modular organization in Paramyxovirinae N and P. *Journal of General Virology, 84*(Pt 12), 3239-3252.

Karlin, D., Longhi, S., & Canard, B. (2002). Substitution of two residues in the measles virus nucleoprotein results in an impaired self-association. *Virology, 302*(2), 420-432.

Karlin, D., Longhi, S., Receveur, V., & Canard, B. (2002). The N-terminal domain of the phosphoprotein of morbilliviruses belongs to the natively unfolded class of proteins. *Virology, 296*(2), 251-262.

Katoh, H., Kubota, T., Nakatsu, Y., Tahara, M., Kidokoro, M., & Takeda, M. (2017). Heat shock protein 90 ensures efficient mumps virus replication by assisting with viral polymerase complex formation. *Journal of Virology*. doi:10.1128/JVI.02220-16

Kavalenka, A., Urbancic, I., Belle, V., Rouger, S., Costanzo, S., Kure, S., . . . Strancar, J. (2010). Conformational analysis of the partially disordered measles virus NTAIL-XD complex by SDSL EPR spectroscopy. *Biophysical Journal, 98*(6), 1055-1064.

Kim, M. Y., Shu, Y., Carsillo, T., Zhang, J., Yu, L., Peterson, C., . . . Oglesbee, M. (2013). hsp70 and a novel axis of type I interferon-dependent antiviral immunity in the measles virus-infected brain. *Journal of Virology, 87*(2), 998-1009. doi:10.1128/JVI.02710-12

Kingston, R. L., Hamel, D. J., Gay, L. S., Dahlquist, F. W., & Matthews, B. W. (2004). Structural basis for the attachment of a paramyxoviral polymerase to its template. *Proc Natl Acad Sci U S A, 101*(22), 8301-8306.

Kingston, R. L., Walter, A. B., & Gay, L. S. (2004). Characterization of nucleocapsid binding by the measles and the mumps virus phosphoprotein. *Journal of Virology, 78*(16), 8630-8640.

Kolakofsky, D., Le Mercier, P., Iseni, F., & Garcin, D. (2004). Viral DNA polymerase scanning and the gymnastics of Sendai virus RNA synthesis. *Virology, 318*(2), 463-473.

Kozin, M. B., & Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *J Appl Cryst, 34*, 33-41.

Lamb, R. A., & Parks, G. D. (2013). Paramyxoviridae. In D. M. Knipe & P. M. Howley (Eds.), *Fields Virology* (pp. 957-995). Philadelphia: Lippincott Williams & Wilkins.

Lamiable, A., Thevenet, P., Rey, J., Vavrusa, M., Derreumaux, P., & Tuffery, P. (2016). PEP-FOLD3: faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Research, 44*(W1), W449-454. doi:10.1093/nar/gkw329

Liang, B., Li, Z., Jenni, S., Rahmeh, A. A., Morin, B. M., Grant, T., . . . Whelan, S. P. (2015). Structure of the L Protein of Vesicular Stomatitis Virus from Electron Cryomicroscopy. *Cell, 162*(2), 314-327. doi:10.1016/j.cell.2015.06.018

Liston, P., Batal, R., DiFlumeri, C., & Briedis, D. J. (1997). Protein interaction domains of the measles virus nucleocapsid protein (NP). *Archives of Virology, 142*(2), 305-321.

Longhi, S. (2007). *Measles virus nucleoprotein*. Hauppage, NY: Nova Publishers Inc.

Longhi, S. (2009). Nucleocapsid structure and function. *Current Topics in Microbiology and Immunology, 329*, 103-128.

Longhi, S. (2011). Structural disorder within the measles virus nucleoprotein and phosphoprotein: functional implications for transcription and replication. In M. Luo (Ed.), *Negative strand RNA virus* (pp. 95-125). Singapore: World Scientific Publishing.

Longhi, S. (2015). Structural disorder within paramyxoviral nucleoproteins. *FEBS Letters*. doi:10.1016/j.febslet.2015.05.055

Longhi, S., Belle, V., Fournel, A., Guigliarelli, B., & Carrière, F. (2011). Probing structural transitions in both structured and disordered proteins by site-directed spin-labeling EPR spectroscopy. *J Pept Sci, 17*(5), 315-328.

Longhi, S., Bloyet, L. M., Gianni, S., & Gerlier, D. (2017). How order and disorder within paramyxoviral nucleoproteins and phosphoproteins orchestrate the molecular interplay of transcription and replication. *Cellular and Molecular Life Sciences, 74*(17), 3091-3118. doi:10.1007/s00018-017-2556-3

Longhi, S., & Oglesbee, M. (2010). Structural disorder within the measles virus nucleoprotein and phosphoprotein. *Protein and Peptide Letters, 17*(8), 961-978.

Longhi, S., Receveur-Brechot, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D., . . . Canard, B. (2003). The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. *Journal of Biological Chemistry, 278*(20), 18638-18648.

Magliery, T. J., Wilson, C. G., Pan, W., Mishler, D., Ghosh, I., Hamilton, A. D., & Regan, L. (2005). Detecting protein-protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *Journal of the American Chemical Society, 127*(1), 146-157.

Martinho, M., Habchi, J., El Habre, Z., Nesme, L., Guigliarelli, B., Belle, V., & Longhi, S. (2013). Assessing induced folding within the intrinsically disordered C-terminal domain of the Henipavirus nucleoproteins by site directed spin labeling EPR spectroscopy. *Journal of Biomolecular Structure and Dynamics, 31*(5), 453-471.

Micsonai, A., Wien, F., Kernya, L., Lee, Y. H., Goto, Y., Refregiers, M., & Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc Natl Acad Sci U S A, 112*(24), E3095-3103. doi:10.1073/pnas.1500851112

Miskei, M., Antal, C., & Fuxreiter, M. (2017). FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. *Nucleic Acids Research, 45*(D1), D228-D235. doi:10.1093/nar/gkw1019

Miskei, M., Gregus, A., Sharma, R., Duro, N., Zsolyomi, F., & Fuxreiter, M. (2017). Fuzziness enables context dependence of protein interactions. *FEBS Letters, 591*(17), 2682-2695. doi:10.1002/1873-3468.12762

Morin, B., Bourhis, J. M., Belle, V., Woudstra, M., Carrière, F., Guigliarelli, B., . . . Longhi, S. (2006). Assessing induced folding of an intrinsically disordered protein by site-directed spin-labeling EPR spectroscopy. *J. Phys. Chem. B, 110*(41), 20596-20608.

Morin, B., Liang, B., Gardner, E., Ross, R. A., & Whelan, S. P. (2016). An in vitro RNA synthesis assay for rabies virus defines critical ribonucleoprotein

interactions for polymerase activity. *Journal of Virology*. doi:10.1128/JVI.01508-16

Morin, B., Rahmeh, A. A., & Whelan, S. P. (2012). Mechanism of RNA synthesis initiation by the vesicular stomatitis virus polymerase. *EMBO Journal, 31*(5), 1320-1329. doi:10.1038/emboj.2011.483

emboj2011483 [pii]

Oglesbee, M. (2007). Nucleocapsid protein interactions with the major inducible heat shock protein. In S. Longhi (Ed.), *Measles virus nucleoprotein* (pp. 53-98). Hauppage, NY: Nova Publishers Inc.

Otieno, S., & Kriwacki, R. (2012). Probing the role of nascent helicity in p27 function as a cell cycle regulator. *PLoS One, 7*(10), e47177. doi:10.1371/journal.pone.0047177

Ringkjøbing Jensen, M., Communie, G., Ribeiro, E. D., Jr.,, Martinez, N., Desfosses, A., Salmon, L., . . . Blackledge, M. (2011). Intrinsic disorder in measles virus nucleocapsids. *Proc Natl Acad Sci U S A, 108*(24), 9839-9844.

Rogers, J. M., Wong, C. T., & Clarke, J. (2014). Coupled folding and binding of the disordered protein PUMA does not require particular residual structure. *Journal of the American Chemical Society, 136*(14), 5197-5200. doi:10.1021/ja4125065

Salladini, E., Delauzun, V., & Longhi, S. (2017). The Henipavirus V protein is a prevalently unfolded protein with a zinc-finger domain involved in binding to DDB1. *Mol Biosyst, 13*(11), 2254-2267. doi:10.1039/c7mb00488e

Schneider, R., Maurin, D., Communie, G., Kragelj, J., Hansen, D. F., Ruigrok, R. W., . . . Blackledge, M. (2015). Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR. *Journal of the American Chemical Society, 137*(3), 1220-1229. doi:10.1021/ja511066q

Schoehn, G., Mavrakis, M., Albertini, A., Wade, R., Hoenger, A., & Ruigrok, R. W. (2004). The 12 A structure of trypsin-treated measles virus N-RNA. *Journal of Molecular Biology, 339*(2), 301-312.

Schreiber, G., Haran, G., & Zhou, H. X. (2009). Fundamental aspects of protein-protein association kinetics. *Chem Rev, 109*(3), 839-860. doi:10.1021/cr800373w

Scott, K. A., Alonso, D. O., Sato, S., Fersht, A. R., & Daggett, V. (2007). Conformational entropy of alanine versus glycine in protein denatured states. *Proc Natl Acad Sci U S A, 104*(8), 2661-2666. doi:10.1073/pnas.0611182104

Sharma, R., Raduly, Z., Miskei, M., & Fuxreiter, M. (2015). Fuzzy complexes: Specific binding without complete folding. *FEBS Letters, 589*(19 Pt A), 2533-2542. doi:10.1016/j.febslet.2015.07.022

Shu, Y., Habchi, J., Costanzo, S., Padilla, A., Brunel, J., Gerlier, D., . . . Longhi, S. (2012). Plasticity in structural and functional interactions between the phosphoprotein and nucleoprotein of measles virus. *Journal of Biological Chemistry, 287*, 11951-11967.

Spera, S., & Bax, A. (1991). Empirical correlation between protein backbone conformation and Ca and

Cb 13C nuclear magnetic resonance chemical shifts. *Journal of the American Chemical Society, 113*, 5490–5492.

Sterckx, Y. G., Volkov, A. N., Vranken, W. F., Kragelj, J., Jensen, M. R., Buts, L., . . . Loris, R. (2014). Small-angle X-ray scattering- and nuclear magnetic resonance-derived conformational ensemble of the highly flexible antitoxin PaaA2. *Structure, 22*(6), 854-865. doi:10.1016/j.str.2014.03.012

Strancar, J., Koklic, T., Arsov, Z., Filipic, B., Stopar, D., & Hemminga, M. A. (2005). Spin label EPR-based characterization of biosystem complexity. *J Chem Inf Model, 45*(2), 394-406. doi:10.1021/ci049748h

Stuchfield, D., & Barran, P. (2018). Unique insights to intrinsically disordered proteins provided by ion mobility mass spectrometry. *Current Opinion in Chemical Biology, 42*, 177-185. doi:10.1016/j.cbpa.2018.01.007

Svergun, D. I., Baraberato, C., & Koch, M. H. (1995). CRYSOL - a Program to evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J Appl Cryst, 28*, 768-773.

Svergun, D. I., Petoukhov, M. V., & Koch, M. H. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal, 80*(6), 2946-2953.

Testa, L., Brocca, S., & Grandori, R. (2011). Charge-surface correlation in electrospray ionization of folded and unfolded proteins. *Anal Chem, 83*(17), 6459-6463. doi:10.1021/ac201740z

Thakkar, V. D., Cox, R. M., Sawatsky, B., da Fontoura Budaszewski, R., Sourimant, J., Wabbel, K., . . . Plemper, R. K. (2018). The Unstructured Paramyxovirus Nucleocapsid Protein Tail Domain Modulates Viral Pathogenesis through Regulation of Transcriptase Activity. *Journal of Virology, 92*(8). doi:10.1128/JVI.02064-17

Tompa, P., & Fuxreiter, M. (2008). Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends in Biochemical Sciences, 33*(1), 2-8.

Uversky, V. N. (1993). Use of fast protein size-exclusion liquid chromatography to study the unfolding of proteins which denature through the molten globule. *Biochemistry, 32*(48), 13288-13298.

Uversky, V. N. (2002). What does it mean to be natively unfolded? *European Journal of Biochemistry, 269*(1), 2-12.

Wang, L. F., Yu, M., Hansson, E., Pritchard, L. I., Shiell, B., Michalski, W. P., & Eaton, B. T. (2000). The exceptionally large genome of Hendra virus: support for creation of a new genus within the family Paramyxoviridae. *Journal of Virology, 74*(21), 9972-9979.

Wang, Y., Chu, X., Longhi, S., Roche, P., Han, W., Wang, E., & Wang, J. (2013). Multiscaled exploration of coupled folding and binding of an intrinsically disordered molecular recognition element in measles virus nucleoprotein. *Proc Natl Acad Sci U S A, 110*(40), E3743-E3752. doi:1308381110 [pii]10.1073/pnas.1308381110

Wilson, C. G., Magliery, T. J., & Regan, L. (2004). Detecting protein-protein interactions with GFP-fragment reassembly. *Nature Methods, 1*(3), 255-262. doi:10.1038/nmeth1204-255

Yabukarski, F., Lawrence, P., Tarbouriech, N., Bourhis, J. M., Delaforge, E., Jensen, M. R., . . . Jamin, M. (2014). Structure of Nipah virus unassembled

nucleoprotein in complex with its viral chaperone. *Nature Structural & Molecular Biology, 21*(9), 754-759. doi:10.1038/nsmb.2868

nsmb.2868 [pii]

Zhang, X., Bourhis, J. M., Longhi, S., Carsillo, T., Buccellato, M., Morin, B., . . . Oglesbee, M. (2005). Hsp72 recognizes a P binding motif in the measles virus N protein C-terminus. *Virology, 337*(1), 162-174.

Zhang, X., Glendening, C., Linke, H., Parks, C. L., Brooks, C., Udem, S. A., & Oglesbee, M. (2002). Identification and characterization of a regulatory domain on the carboxyl terminus of the measles virus nucleocapsid protein. *Journal of Virology, 76*(17), 8737-8746.

.