

OPEN

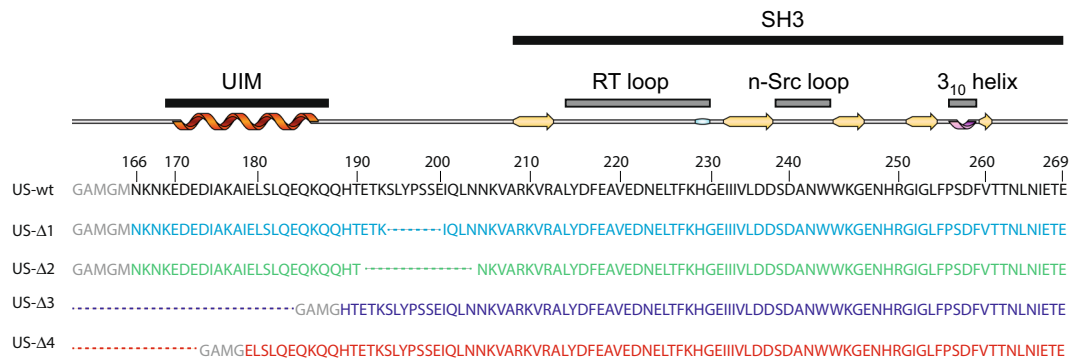
# Molecular recognition of ubiquitin and Lys63-linked diubiquitin by STAM2 UIM-SH3 dual domain: the effect of its linker length and flexibility

Minh-Ha Nguyen<sup>1</sup>, Marie Martin<sup>1</sup>, Henry Kim<sup>1</sup>, Frank Gabel<sup>2</sup>, Olivier Walker<sup>1</sup> & Maggy Hologne<sup>1</sup>

Multidomain proteins represent a broad spectrum of the protein landscape and are involved in various interactions. They could be considered as modular building blocks assembled in distinct fashion and connected by linkers of varying lengths and sequences. Due to their intrinsic flexibility, these linkers provide proteins a subtle way to modulate interactions and explore a wide range of conformational space. In the present study, we are seeking to understand the effect of the flexibility and dynamics of the linker involved in the STAM2 UIM-SH3 dual domain protein with respect to molecular recognition. We have engineered several constructs of UIM-SH3 with different length linkers or domain deletion. By means of SAXS and NMR experiments, we have shown that the modification of the linker modifies the flexibility and the dynamics of UIM-SH3. Indeed, the global tumbling of both the UIM and SH3 domain is different but not independent from each other while the length of the linker has an impact on the ps-ns time scale dynamics of the respective domains. Finally, the modification of the flexibility and dynamics of the linker has a drastic effect on the interaction of UIM-SH3 with Lys63-linked diubiquitin with a roughly eight-time weaker dissociation constant.

A wide ensemble of protein-protein interaction networks mediate communication and information exchange between cells<sup>1</sup>. Alteration or interruption in these networks usually leads to severe damage or disease<sup>2,3</sup> and has led to a tremendous exploration of protein-protein inhibition<sup>4,5</sup>. These interactions are mainly mediated by protein domains, which could be considered as modular building blocks, assembled in different fashions. They decode specific signals emerging from post-translational modifications involved in receptor signaling, endocytosis or DNA damage for instance<sup>6</sup>. Like words and grammar that rule a language, protein domains may assemble differently and adopt different architectures<sup>7,8</sup> to form multidomain proteins (MDPs) that represent more than 70% of the eukaryote proteome<sup>9,10</sup>. These identical or different domains are linked together by disordered segments characterized by variable lengths and sequences, also termed linkers. Due to their inherent variable flexibility, linkers can confer large conformational rearrangement to proteins to induce intra- or inter-domain interactions<sup>11,12</sup>. Therefore, MDPs possess the characteristics of both intrinsically disordered and well-folded proteins and their linkers can undergo post-translational modifications<sup>13</sup> rendering their study more intricate<sup>14</sup>. A wide range of approaches has reshaped our understanding of multidomain proteins, including FRET<sup>15</sup>, SAXS<sup>16</sup>, SANS<sup>17</sup> or NMR<sup>18</sup>, either as an individual technique or in combination<sup>19</sup>. From these methods, NMR spin relaxation measurements are probably the most important source of information when dealing with multidomain proteins as the collected data carry dynamical information for different spin sites measured at different magnetic fields at various experimental conditions<sup>20,21</sup>. Boosted by hardware and force field improvement, multidomain protein dynamics has also been studied by all-atoms molecular dynamics or metadynamics that allow the exploration of longer trajectories and wider conformational space<sup>22-24</sup>. Despite many studies in the field,

<sup>1</sup>Institut des Sciences Analytiques (ISA), Univ Lyon, CNRS, UMR5280, Université Claude Bernard Lyon 1, Lyon, France. <sup>2</sup>Université Grenoble Alpes, CEA, CNRS, IBS, 38042, Grenoble, France. Correspondence and requests for materials should be addressed to M.H. (email: [maggy.hologne@univ-lyon1.fr](mailto:maggy.hologne@univ-lyon1.fr))



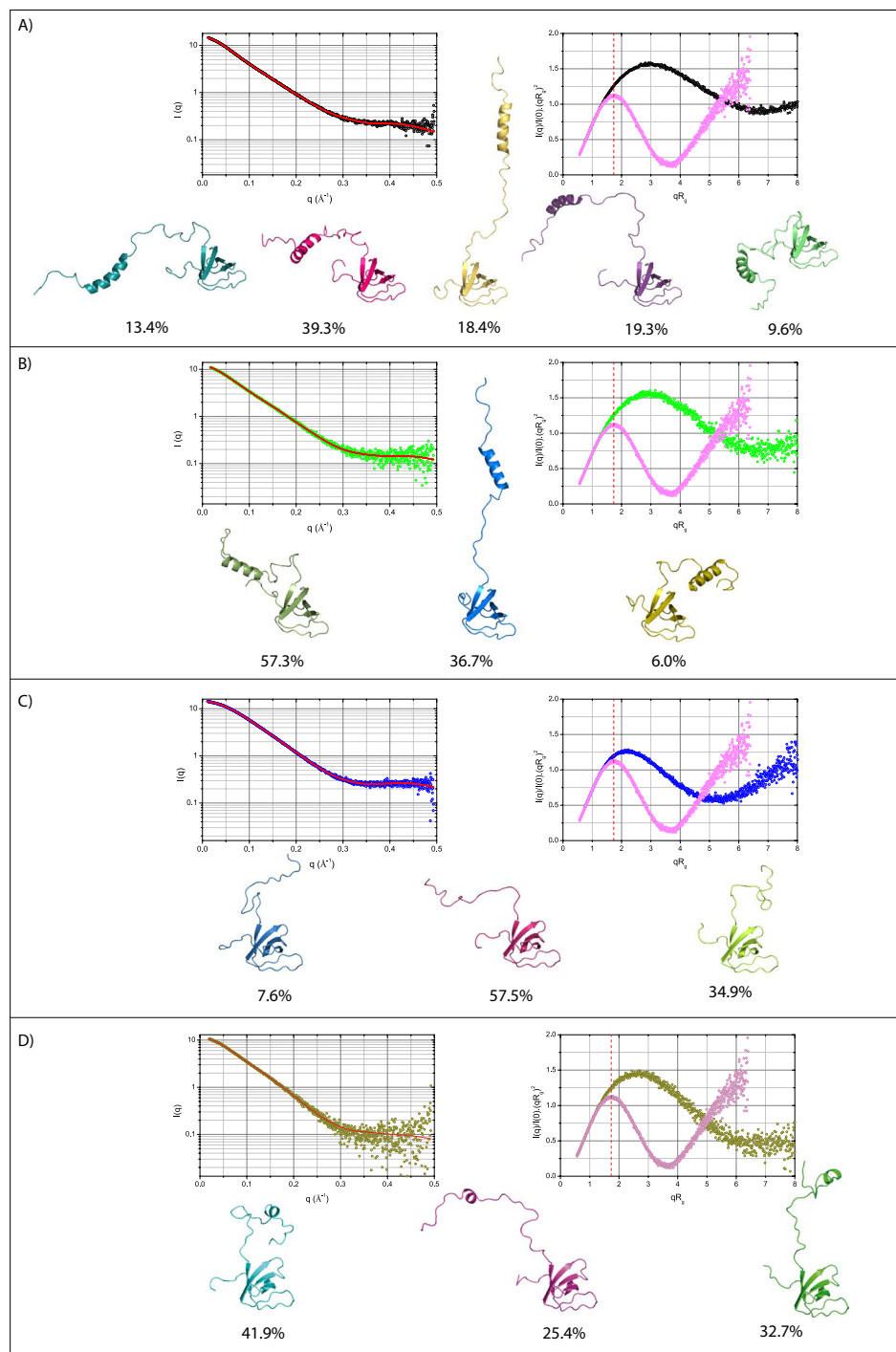
**Figure 1.** Name of the different constructs used in the present study along with their respective sequence and numbering. The schematic representation of UIM-SH3 secondary structure has been obtained by means of dssp<sup>66,67</sup> embedded into cartoon representation program SSS-drawer (<https://github.com/zharmad/SSS-Drawer>).

some questions are still pending and remain unanswered, or at least need further inquiry. For instance, what is the function of the linker flexibility and dynamics with respect to molecular recognition processes like allostery, cooperativity or avidity? To shed light on this dangling question we have chosen to explore the interaction of various STAM2 UIM-SH3 constructs with different binding partners. Indeed, STAM2 is part of the ESCRT-0 complex, which is the most upstream component of the ESCRT machinery involved in lysosomal degradation<sup>25</sup>. To do so, the proteins that are directed for degradation are first tagged by Lys63-linked polyubiquitin chains<sup>26</sup>. Here, STAM2 harbors three ubiquitin-binding domains (UBDs), namely the VHS, UIM and SH3 domains that recognize and interact with ubiquitin (Ub) moieties. In a previous work, we have delineated how these three domains interact with Ub and Lys63-linked diubiquitin chains (Lys63-Ub<sub>2</sub>). While the VHS-UIM construct cooperatively engages Lys63-Ub<sub>2</sub><sup>27</sup>, the UIM-SH3 has demonstrated an exquisite capability to bind Ub, Lys63-Ub<sub>2</sub>, UBPY and AMSH through the UIM and SH3 domains<sup>28,29</sup>. In the current article, we have selected and engineered five different constructs of the UIM-SH3 (further denoted US-WT) dual domain protein with different linker lengths or domain deletion. By means of SAXS, we show that the truncation of the linker results in a decrease in the US-WT flexibility and the conformational space sampled by each of the constructs. We have also investigated the dynamical properties at the ps-ns time scale through NMR spin relaxation measurements and we show that the correlation times related to the UIM and SH3 domains are significantly different. Moreover, the gap between the correlation times associated with each domain decreases along with the linker length. Finally, these observations were complemented by titration experiments where we have compared the interaction of Ub and Lys63-Ub<sub>2</sub> with all the newly designed constructs. Our results indicate that the binding affinity between Ub and the various US constructs is rather similar while the affinity between Lys63-Ub<sub>2</sub> and the various US constructs experiences a drastic change when the length of the linker is shortened and may have further implication in the binding and molecular recognition between multidomain proteins involved in lysosomal degradation.

### Structural Modifications Induced by UIM-SH3 Mutations

To study the influence of the linker length on the flexibility and dynamical properties, we have selected and engineered US-WT derived variants where the tether that bridges UIM and SH3 has been modified in several ways. US-Δ1 and US-Δ2 have their linker shortened for 7 and 14 amino acids, respectively. While UIM was completely removed in US-Δ3 that leaves the SH3 domain only with the full-length linker, the C-terminal half of UIM remains in US-Δ4 (Fig. 1). As we have previously shown that the UIM and SH3 domains of the US-WT construct did not interact with each other, we have checked whether any mutation would have modified this structural hallmark. For each of the mutants, a <sup>1</sup>H, <sup>15</sup>N-HSQC NMR spectrum was recorded and compared with the one obtained for US-WT. For US-Δ1, US-Δ2 and US-Δ3, the <sup>1</sup>H, <sup>15</sup>N-HSQC spectra nicely overlap with the US-WT spectrum except for missing residues or residues located in the N-terminus of US-Δ3 (see Fig. S1). This result indicates that both UIM and SH3 in US-Δ1 and US-Δ2 keep a similar fold as UIM and SH3 in US-WT and that the linker shortening does not affect the UIM helix motif. Furthermore, the SH3 domain structure in US-Δ3 is not affected by the complete UIM deletion. Like US-Δ3, the US-Δ4 <sup>1</sup>H, <sup>15</sup>N-HSQC spectrum corresponding to the SH3 domain nicely overlaps with the US-WT spectrum while some of the signals associated with the remaining UIM residues are shifted. This is likely due to a partial unwinding of the UIM helix after the cleavage of the US-WT N-terminus. Moreover, our circular dichroism results on US-Δ4 shows about the same helicity percentage than US-Δ3 while having a longer sequence (see Fig. S2 and Table ST1), which corresponds to a greater average number of helix residues (~5.7 for US-Δ4 over ~3.4 for US-Δ3). Furthermore, we can also exclude any transient interaction of the remaining UIM with SH3. To allow for a further analysis, backbone resonance assignment of US-Δ4 has been carried out by using a combination of the HNCO, HN(CA)CO, HNCACB and HN(CO)CACB experiments (see methods).

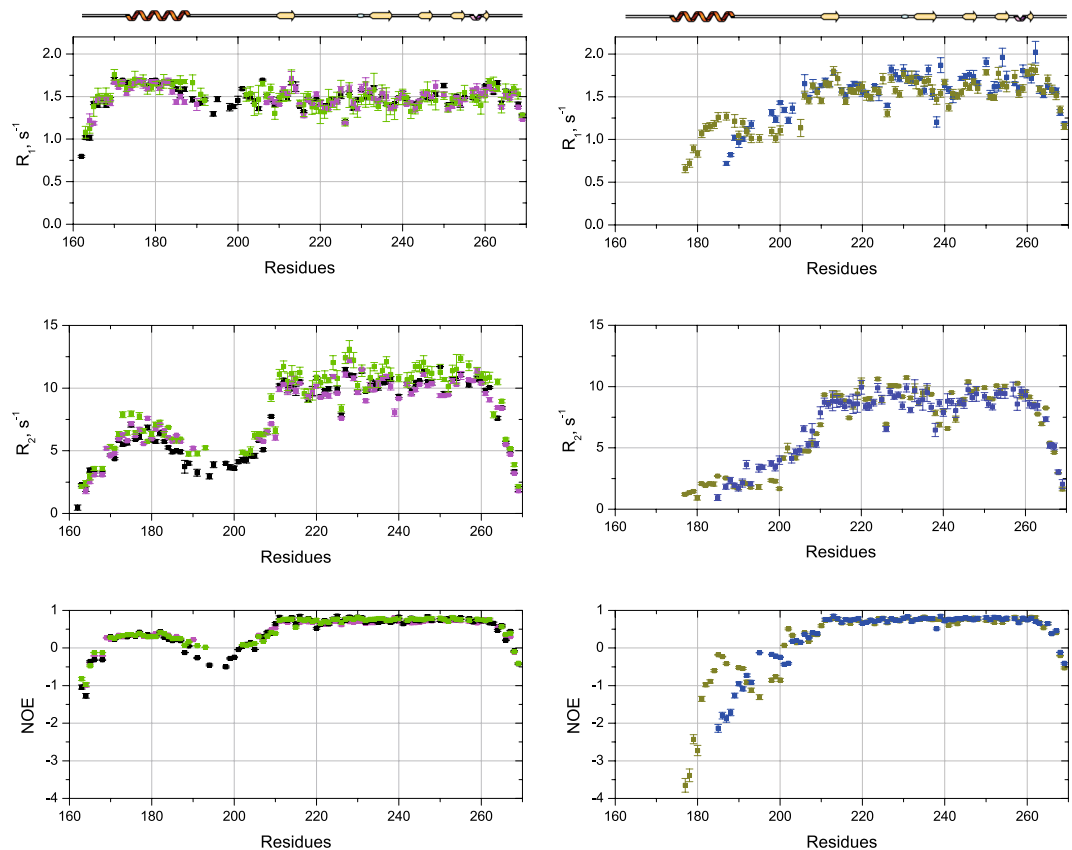
**Characterization of the flexibility of the UIM-SH3 constructs.** Thus, we were seeking to understand if any modification of the US-WT sequence would induce a drastic change of its flexibility and spatial organization. To this end, we have carried out SAXS experiments on the US-WT, Δ1, Δ3 and Δ4 constructs only. For all



**Figure 2.** SAXS data (left) and dimensionless Kratky plots (right) for **(A)** US-WT, **(B)** US- $\Delta 1$ , **(C)** US- $\Delta 3$  and **(D)** US- $\Delta 4$ . A comparison with ubiquitin is given for each Kratky plot (pink). A dashed red line is plotted for a  $qR_g$  value of  $\sqrt{3}$ . It represents the typical peak maximum for globular proteins and is consistent with ubiquitin's Kratky plot curve. The representation of the top-scoring N-ensemble of structure and their respective weight that best fit the SAXS data is presented at the bottom of each panel and are determined by means of the Multifox server<sup>64</sup>. The back-calculated SAXS data are represented by a continuous red line.

of them, data have been recorded at four different concentrations to ensure good quality data at both low and high angles (see methods). The SAXS data are presented in Fig. 2. Overall, the scattering profiles are smooth compared to a multidomain protein with fixed domain distances and reflect interdomain motion or a dynamic averaging<sup>16</sup>. Further Guinier analysis provides the radius of gyration ( $R_g$ ) values and information on the average size of the different constructs in solution. On average, US- $\Delta 3$  displays the shortest  $R_g$  while other constructs have  $R_g$  close to each other (see Table ST2). To characterize protein flexibility, one frequently resorts to a so-called Kratky plot,

which highlights the difference between a well folded, compact protein with a pronounced curve maximum and a random, unstructured chain that exhibits a plateau at the same  $qR_g$  value<sup>30</sup>. However, this kind of graphical representation is difficult to compare between objects of different size or different multidomain proteins. Therefore, we have represented dimensionless Kratky plots (Figs 2 and S3B) where  $I(q)$  is normalized to the forward scattering intensity  $I(0)$  and  $q$  is normalized to  $R_g$ . Indeed, a typical fully folded globular protein (here exemplified by ubiquitin on Fig. 2) shows a curve maximum of  $\sim 1.104$  at a  $qR_g$  value equals to  $\sqrt{3}$  ( $\sim 1.73$ ) and a shape that rises and falls symmetrically<sup>30,31</sup>. Opposite to this situation, a random chain would exhibit a nearly hyperbolic curve followed by a further increase at higher  $q$  values. In the case of all the US variants, the dimensionless Kratky plots display a significant deviation from both the 'ideal' folded case and the classical disordered case with a non-symmetrical curve, a shift of the curve maximum and a slow decrease at higher  $q$  values (Fig. 2). This behavior confirms that each US construct contains at the same time well-folded and intrinsically-disordered region. Moreover, US- $\Delta 3$  shows a curve maximum for  $qR_g$  closer to 1.73 which reflects a lower composition of unstructured flexible regions compared to the other three constructs (Figs 2C and S3B). According to the dimensionless Kratky plot, it is also noticeable that the curve depicting US-WT falls with a less pronounced slope at higher  $q$  values. Thus, it is likely that the US-WT construct is the one affected by the highest flexibility among other variants and that the shortening of the linker has removed part of the contribution of the flexible disordered region. The  $P(r)$  functions present an asymmetric shape with a maximum around 17.5 Å (Fig. S3A). For the four constructs, this peak at low  $r$  value corresponds to intradomain distances that mostly arise from SH3, the biggest among the two domains. After reaching a maximum, the  $P(r)$  curves decrease in a bumpy manner with slightly pronounced humps. This behavior firstly corresponds to the presence of the second domain UIM, that is a small helix (for US-WT, US- $\Delta 1$  and US- $\Delta 4$ ) and secondly reflects the possibility for the different constructs to adopt various spatial conformations due to their flexible linker. US-WT, US- $\Delta 1$  and US- $\Delta 4$  have a similar pattern and differ only by the extent of their curve at high  $r$  values. US- $\Delta 3$  adopts a more symmetric curve that reflects a more compact structural organization in solution. Further analysis of the  $P(r)$  function allows the determination of  $D_{max}$  that corresponds to the maximum diameter of the protein. The US-WT possesses the highest  $D_{max}$  and samples the widest volume among all the different constructs (Table ST2). A shortening of the linker region by seven amino acids does not display a significant effect on  $D_{max}$  and  $R_g$ . This result indicates that: (i) Despite the presence of seven more amino acids, the maximum diameter of US-WT is close to US- $\Delta 1$ , therefore its linker does not behave as a fully elongated tether but rather shows turns and loops. (ii) Regardless of its shorter sequence length, US- $\Delta 1$  still occupies a significant volume and displays an average size comparable to US-WT. Further deletion of the US-WT N-terminus by 14 amino acids (US- $\Delta 4$ ) induces the partial deletion of UIM but only leads to a marginal change of  $R_g$  while  $D_{max}$  drops by 11 Å to adopt a shorter maximum diameter compared to US-WT. Conversely, one can notice a drastic change after the complete deletion of the UIM domain (US- $\Delta 3$ ) where both  $R_g$  and  $D_{max}$  show a significant decrease of 4.6 and 25 Å respectively. Therefore, these results support the fact that US- $\Delta 3$  has a more compact organization along with a reduction of the explored conformational space. These findings support the idea that the UIM domain acts as a spring that can maintain the SH3 domain at a given distance. To characterize the conformation adopted by the different constructs, we have used the ten modeled structures generated by Modeller (see methods) as starting structure for SAXS data fitting. SAXS data were back-calculated by means of the FoXS program<sup>32,33</sup> and none of the initial models have succeeded to reproduce the SAXS curves and have given a systematic high  $\chi$  score (see Table ST3). Thus, it is unlikely that the different US constructs exist in a single conformation. Therefore, we used MultiFoXS for computing N-state ( $N = 1$  to 5) models of the US constructs and saving the 1000 top scoring after starting from 10000 initial conformations generated through the RRT algorithm (see methods). For the US-WT construct, the  $\chi$  score significantly decreased with a one state model compared to the best model issued from Modeller (from 101.00 to 1.62). A further increase of the number of conformations to five states improves the  $\chi$  score by around 50% and decreases the residuals (see Table ST3 and Fig. S4A). This result underlines the fact that different conformations co-exist in solution, from an elongated to a compact state along with different intermediate states. The number of states is also confirmed by the analysis of the  $R_g$  distribution computed for the 1000 best-scoring N-state models (Fig. S5). The  $R_g$  distribution in the initial pool of 10000 conformations is almost uniform and span a large amplitude from 16 to 35 Å. Whatever the model, a highly populated  $R_g$  is found around 20 Å while four other values of  $R_g$  appeared for two to five-state models. For the five-state model that shows a significant improvement of the  $\chi$  score, three values of  $R_g$  are approximately homogeneously populated at 33.8, 26.8 and 24.8 Å while a lowly populated value of  $R_g$  appears at 15.8 Å. These results demonstrate that US-WT is sufficiently flexible and dynamic to explore a broad conformational space. For the US- $\Delta 1$  construct, the  $\chi$  score significantly decreased with a one state model compared to the best model issued from Modeller (from 25.57 to 1.39). Moreover, the SAXS data are nicely reproduced with an optimum solution using a three-state models compared to the one-state model with an improvement of 28% (see Table ST3 and Fig. S4B). A four or five-state model did not improve the quality of the calculated SAXS data. As for US-WT, the one-, two- or three-state models share a common value of  $R_g$  that represents a more compact structure of US- $\Delta 1$  ( $R_g = 19.6$  Å for the three-state model). All these models also show two other values of  $R_g$  and reinforce the choice of a three-state model for the US- $\Delta 1$  construct ( $R_g$  equals 28.6 and 30.6 Å for the three-state model). For the US- $\Delta 3$  and  $\Delta 4$ , the lowest  $\chi$  scores obtained for the best models supplied by Modeller are significantly higher compared to the one-state model (55.60 compared to 1.29 and 4.17 compared to 1.73 respectively). Furthermore, we have retained an optimum three-state model that improves the  $\chi$  score by 33 and 39% compared to the one-state model for US- $\Delta 3$  and  $\Delta 4$  respectively (Table ST3). None of the four- or five-states model further improve the calculation of the SAXS curves (Fig. S4C,D). The  $R_g$  distribution related to the US- $\Delta 4$  is spread on a region comparable to the US- $\Delta 1$  construct and is consistent with the fact that the experimental  $R_g$  for US- $\Delta 1$  and US- $\Delta 4$  are quite similar despite a different sequence. For US- $\Delta 4$ , the three-state model shows two highly populated  $R_g$  values at 15.7 and 20.7 Å that correspond to compact structures



**Figure 3.** Relaxation parameters  $^{15}\text{N}$   $R_1$ ,  $^{15}\text{N}$   $R_2$  and  $^{15}\text{N}$ - $^1\text{H}$  heteronuclear NOE for the different US constructs according to the following color code: US-WT (black), US- $\Delta 1$  (green), US- $\Delta 2$  (magenta), US- $\Delta 3$  (blue) and US- $\Delta 4$  (dark yellow). The differences in the  $R_1$  and  $R_2$  levels for SH3 and UIM indicate that they tumble with different correlation times. Negative and close to zero heteronuclear NOE values also indicate a high degree of backbone flexibility in the intervening linker as well as in the N and C termini of the different constructs.

and two lowly populated  $R_g$  at 22.7 and 25.7 Å that correspond to a more extended conformation. As opposed to the US- $\Delta 4$ , the  $R_g$  distribution associated with US- $\Delta 3$  spans a narrow region between 13 and 21 Å with three main population of  $R_g$  at 15.2, 18.2 and a lower populated  $R_g$  at 20.2 Å showing that its conformational arrangement stays rather compact. Indeed, this result is in good agreement with the  $R_g$  derived from a Guinier analysis. Overall, the N-state model analysis provides a more accurate and informative view of the different US construct compared to the common experimental  $R_g$  value that represents an average value over all possible conformations. It shows that the conformational space sampled by the different US construct could be described by a combination of different  $R_g$  along with their respective weight and delineates the dynamics of the current US mutants. While the comparison of the experimental  $R_g$  did not show any dramatic difference between US-WT, US- $\Delta 1$  and US- $\Delta 4$ , the N-state model reveals a more contrasted depiction of the different conformations adopted by the various US constructs with maximum  $R_g$  values of 33.8, 30.6, 25.7 and 20.2 Å for US-WT, US- $\Delta 1$ , US- $\Delta 4$  and US- $\Delta 3$  respectively.

**Dynamical properties in the ps-ns time scale of the UIM-SH3 variants.** While the global flexibility of the different constructs has been investigated through the measurement of SAXS data, a more accurate description of the dynamics is necessary to characterize the different time scales (in the ps-ns range) that affect the various UIM-SH3 mutants. As we have noticed that a deletion of 7 amino acids in the linker region has a consequence on the flexibility of UIM-SH3, we have engineered a US- $\Delta 2$  construct where 14 amino acids have been deleted in the linker area (see Fig. 1). We have investigated the ps-ns time scale molecular motion of the different US constructs through the measurement of the three commonly used spin relaxation parameters  $^{15}\text{N}$ - $R_1$ ,  $^{15}\text{N}$ - $R_2$  and the  $^{15}\text{N}$ - $^1\text{H}$  cross-relaxation rates, via the steady-state  $^{15}\text{N}\{^1\text{H}\}$ NOEs (see methods). At a first glance, several observations can be drawn for the constructs that contain both a UIM and a SH3 domain. First, the UIM and the SH3 domains exhibit significant differences of their average  $R_2$  and NOEs (see Fig. 3 and Table ST5) while the profile of  $R_1$  values is rather flat. Second, if one considers that no conformational exchange affects either the SH3 or the UIM domain, their average  $R_2$  levels deviate significantly from what would be expected when taken individually (see Fig. S6). In this representation, the molecular mass dependence of  $R_2$  was utilized as a “molecular ruler” (calibrated using  $R_2$  data for UIM, Ub and Lys63-Ub<sub>2</sub>). For instance, the UIM individual domain shows an average  $R_2$  of  $5.1 \pm 0.6 \text{ s}^{-1}$  for an expected molecular mass of  $\sim 4.0 \text{ kDa}$ <sup>27</sup>. Thus, our  $^{15}\text{N}$  relaxation data clearly support the fact that the UIM and SH3 domains involved in the different US mutants, tumble differently but not independently.

Moreover, it is likely that UIM and SH3 do not share a common rotational diffusion tensor. The interdomain motions drastically affect the global shape of the protein and hence its rotational diffusion, anisotropy and global tumbling. Additionally, if one considers a 36% correlation between domains reported in a previous work of Bae *et al.*<sup>34</sup>, it corresponds approximately to the increase in  $R_2$  seen by the UIM and SH3 domain compared to the values expected for these domains taken individually. Therefore, we have chosen to derive the dynamics of the US mutants by considering a different rotational diffusion tensor for each of the domains. From the global tumbling analysis, it is noticeable that the differences between the correlation times that affect the UIM or the SH3 domains in US-WT, US- $\Delta 1$  and US- $\Delta 2$  are rather similar although US- $\Delta 2$  shows the shortest correlation time associated with the shortest linker. In the same order of idea, the gap between the average  $R_2$  for the UIM and SH3 domains significantly decreases when the length of the linker is shortened (see Table ST5) and clearly reflects the fact that the interdomain motion becomes less prominent when the linker is shortened. It is also noteworthy that the NOE values in the UIM C-terminus region for US- $\Delta 1$  and  $\Delta 2$  exhibit higher values compared to the US-WT and support the conclusion that the linker shortening introduces a higher stiffness at the UIM C-terminus. This also lends credence to the fact that the domain coupling becomes less obvious when the linker length decreases below a given value. When the UIM is completely or partially deleted, the SH3 domain in US- $\Delta 3$  and  $\Delta 4$  exhibits a significant decrease in its average  $R_2$  level with a concomitant increase in  $R_1$ , consistent with a decrease in the apparent molecular mass. These observations are also supported by a drastic reduction of the correlation time of the SH3 domain (see Table ST5). For US- $\Delta 3$  and  $\Delta 4$ , the long flexible N-terminus displays much lower  $R_2$  and NOE values (up to  $-3.6$  for NOE) compared to the same residue range in US-WT,  $\Delta 1$  and  $\Delta 2$ . This observation clearly reflects much faster local motions and, therefore, a much higher flexibility of this part in US- $\Delta 3$  and  $\Delta 4$ . A more likely explanation can be given with the use of reduced spectral density mapping that makes no assumption about the nature of the rotational diffusion.  $R_1$ ,  $R_2$  and NOE can directly be expressed as a linear combination of spectral densities operating at three different frequencies  $J(0)$ ,  $J(\omega_N)$  and  $J(0.87\omega_H)$ <sup>35</sup> (see supplementary materials). As previously discussed, these analytical expressions can also be used for multidomain proteins containing disordered segments<sup>20</sup>. The high frequency spectral density functions  $J(0.87\omega_H)$  are sensitive only to fast internal motions on a picosecond timescale while the zero frequency spectral density functions  $J(0)$  are sensitive to nanosecond internal motions that are faster than the global tumbling. As can be seen in Fig. S7, US- $\Delta 3$  and  $\Delta 4$  are characterized by restricted motions in the SH3 core and exhibit large  $J(0)$  values while they clearly show high contribution of picosecond timescales (see range 180–200 on Fig. S7), and hence a much higher flexibility of their N-terminus part compared to US-WT,  $\Delta 1$  and  $\Delta 2$ . As observed above, US- $\Delta 1$  and  $\Delta 2$  display a reduction of the high frequency motion concomitant with an increase in the slow nanosecond motion of the UIM C-terminus region and hence, reflects a decrease in flexibility for the corresponding constructs.

**How molecular recognition is modified by the linker length and flexibility.** As alluded to above, SAXS and NMR spin relaxation data have shown a difference in terms of rigidity and dynamics along with the length of the linker that bridges UIM and SH3. In a previous work, we have shown that US-WT has the capability to bind the deubiquitinating enzymes UBPY and AMSH<sup>28</sup> as well as Ub and Lys63-Ub<sub>2</sub><sup>29</sup>. Therefore, the next pending question could be “Do the length, rigidity and dynamics of the US-WT linker affect the interaction of US with its binding partners?”. To answer this question, we have monitored chemical shift perturbations (CSPs) of <sup>15</sup>N-labeled US constructs upon addition of the corresponding unlabeled Ub and Lys63-Ub<sub>2</sub> by means of <sup>1</sup>H, <sup>15</sup>N HSQC experiments (Fig. S8). For US-WT,  $\Delta 1$  and  $\Delta 2$  that possess a complete UIM domain, several residues located either on the UIM and the SH3 domain experience significant CSPs or strong signal broadening. Since the UIM and SH3 domains do not interact with each other, any perturbation would result from a direct interaction of the US constructs with Ub. Overall, the CSPs patterns on the UIM domain for the different complexes are similar within each other while the number of CSPs above the threshold (fixed here at 0.4 ppm) for the SH3 domain slightly decreases, concomitant with a decrease in the US linker length (Fig. S8). On the SH3 domain, the affected residues cover a region mainly centered on the RT loop and the 3<sub>10</sub> helix. On the other hand, residues K208, V207 and A206 located close to the RT loop exhibit CSPs increase that is likely due to the shortening of the linker. The quantitative analysis of the titration curves allows us to derive the corresponding dissociation constants. As seen in Table 1 and in Fig. S9, saturation of the UIM and SH3 domains in US- $\Delta 1$  and  $\Delta 2$  occurs at a higher concentration of Ub and the dissociation constant slightly increases approximately by the same amount for UIM on the one hand and SH3 on the other hand when US- $\Delta 1$  and then US- $\Delta 2$  is considered ( $\sim 1.5$  and  $\sim 2.4$  times respectively). These observations led us to conclude that the shortening of the linker has only a modest impact on US binding affinity when Ub is a binding partner. Additionally, half or entire deletion of the UIM domain completely abolished Ub binding on the UIM side while the SH3 is not affected and still shows a significant binding affinity (see Table 1) with a similar  $K_d$  compared to US-WT. This is consistent with a previous result that demonstrated how a US<sup>178E</sup> mutant completely abolished Ub binding on the UIM side<sup>28</sup>. The STAM2 protein, which plays an essential role in the endosomal sorting process, is known to have a preferential binding activity with Lys63-linked polyubiquitin chains<sup>26</sup>. Thus, we were seeking to understand if any modification of the flexibility and dynamics of the US linker would perturb its binding affinity with a Lys63-Ub<sub>2</sub> dual domain protein. We used a similar protocol as the one used for Ub binding and recorded CSPs along with the addition of Lys63-Ub<sub>2</sub> (Fig. S10). Overall, the CSP pattern for the US-WT/Lys63-Ub<sub>2</sub> complex is comparable to the one seen for the US-WT/Ub binding and does not present additional perturbed residues. Nevertheless, a closer look at CSP located on the UIM domain reveals an increase in the CSPs on the US- $\Delta 2$  N-terminus while residues Q185 and Q187 experience a drastic decrease in their respective CSPs compared to US-WT. The analysis of the titration curves reveals a shift of the saturation point for several residues, more pronounced for residues Glu218, Ala219 or Lys186 on both US- $\Delta 1$  and  $\Delta 2$ . While the dissociation constant seen on the UIM domain for the US-WT/Lys63-Ub<sub>2</sub> complex is comparable to the one seen for the US-WT/Ub complex (Table 1), the affinity of the SH3 domain for Lys63-Ub<sub>2</sub> strongly increases compared to Ub and suggests an avid interaction, as already underlined<sup>29</sup>. It is also noteworthy that

		$K_d$ ( $\mu$ M)	$K_d$ ( $\mu$ M)
		Mono Ub	Lys63-Ub <sub>2</sub>
US-WT	UIM	86 $\pm$ 31	78 $\pm$ 59
	SH3	267 $\pm$ 119	60 $\pm$ 28
US- $\Delta$ 1	UIM	148 $\pm$ 60	137 $\pm$ 57
	SH3	394 $\pm$ 127	253 $\pm$ 163
US- $\Delta$ 2	UIM	203 $\pm$ 91	99 $\pm$ 50
	SH3	669 $\pm$ 234	483 $\pm$ 196
US- $\Delta$ 3		266 $\pm$ 96	39 $\pm$ 22
US- $\Delta$ 4		158 $\pm$ 79	97 $\pm$ 49

**Table 1.** Summary of the different dissociation constants measured in the present study. In the case of interaction with Ub, dissociation constants were extracted using a 1:1 model for all constructs. In the case of interaction with Lys63-Ub<sub>2</sub>, dissociation constants were extracted for US-WT, US- $\Delta$ 1 and US- $\Delta$ 2 using a 2:1 model and for US- $\Delta$ 3 and US- $\Delta$ 4 using a 1:1 model (see methods section). Standard deviations are used as error estimates.

the affinity of the SH3 domain decreases drastically by a factor of  $\sim$ 8 with the shortening of the linker (US- $\Delta$ 2) while the dissociation constant associated with the binding of UIM with Lys63-Ub<sub>2</sub> is not affected. Additionally, Lys63-Ub<sub>2</sub> shows the tightest binding with SH3 when the whole UIM domain has been deleted (US- $\Delta$ 3). These results clearly support the conclusion that the UIM domain in US- $\Delta$ 2 has an inhibitory effect with respect to SH3. This effect may be due to a combination of a shorter linker, steric hindrance and an excluded volume constraint of UIM that reduces the conformational space available for the binding of SH3. Our results also demonstrate that multidomain proteins encourage avid binding with other multidomain proteins and that any significant modification in the linker region may affect the binding affinity of the complex.

## Discussion

US-WT is a dual domain protein that contains two ubiquitin binding domains (UBDs) and is a component of the STAM2 protein. This part of STAM2 is fundamental in the sorting process of lysosomal degradation and recognizes various binding partners<sup>28,29</sup>. More specifically, STAM2 itself is a component of the ESCRT-0 complex that binds ubiquitin or polyubiquitin chains with a preference for Lys63- over Lys48-linked chains<sup>26,36</sup>. Furthermore, to maintain a constant pool of ubiquitin in the cell, polyubiquitin chains require cleavage by the specific enzymes UBPY or AMSH. The AMSH catalytic activity is enhanced when STAM is present and is due to a given structural organization that promotes the association of STAM, Lys63-linked polyubiquitin and AMSH<sup>29,37,38</sup>. As can be seen, STAM2 and more specifically the US domains should accommodate different multidomain proteins in a spatially and timely fashion. Thus, it is of prime importance to understand how these different processes relate to the inherent plasticity of UIM-SH3, the latter domains being connected by a disordered linker of 20 amino-acids. One can also wonder if the motional properties of the linker play a role with respect to molecular recognition with various binders. Our results provide a first glimpse on the ability of US-WT to accommodate different binding partners due to the structural and dynamical characteristics of its linker. Here we took a bold and direct approach of progressively reducing linker regions and beyond-starting from WT, partial linker deletion, partial UIM deletion and finally a complete UIM deletion. The SAXS analysis of the variants demonstrates an evolution of changes in  $R_g$  and  $D_{max}$ , starting with subtle decrease and towards more drastic change when both domain and linkers were affected. Nevertheless, the sole consideration of the  $R_g$  derived from SAXS data is not sufficient to account for the inherent flexibility of the different US variants. Indeed,  $R_g$  represents an average over multi-state conformations and may hide a subtle structural dispersity of the different US variants. Therefore, we used a multi-state approach to account for the various possible conformations and their interconversion that describe the SAXS curves. As seen in the results section, the US-WT construct could be described by a five-state model that spans a wide range of  $R_g$  values from 15.8 to 33.8 Å. Conversely, the US- $\Delta$ 1,  $\Delta$ 3 and  $\Delta$ 4 only need a three-state model to match the SAXS data, US- $\Delta$ 3 being represented by the more compact conformations. To the light of our results, one can wonder why US-WT and US- $\Delta$ 1 or US- $\Delta$ 4 have close values of  $R_g$  and  $D_{max}$  even though US- $\Delta$ 1 shows a linker region deleted by seven amino acids compared to US-WT. One possible explanation is the formation of loop region in US-WT that allows conformations as compact as for US- $\Delta$ 1. The average C $\alpha$ -C $\alpha$  distance between residue Gln188 and Ala208 for the US-WT and US- $\Delta$ 1 models, gives a value of 37.7 and 28.3 Å respectively, thus a difference of 9.4 Å when one expects a end-to-end distance of  $\sim$ 23.8 Å for an individual peptide of the missing sequence in US- $\Delta$ 1<sup>39</sup>. This fact also lends credence to our hypothesis that the average  $R_g$  of US-WT is explained by the formation of loops present in different conformational states rather than several extended polymer chains. Additionally, the excluded volume due to the presence of the UIM and SH3 domains may also play a dominant role in the conformational space sampled by the different US constructs. Indeed, it is noteworthy that the excluded volume represented by the UIM or the SH3 domains prevents the flexible linker to adopt certain conformations (as illustrated by Fig. S11) in the case of US-WT,  $\Delta$ 1 or  $\Delta$ 4 but allowed in the case of US- $\Delta$ 3<sup>40</sup>. It will also greatly decrease the number of conformations otherwise accessible to a chain and will increase its average dimension or the apparent stiffness of the linker as reported by recent Monte Carlo simulations<sup>39</sup>. This apparent stiffness could also induce loop formation at a time scale of  $\sim$ 12–20ns<sup>41</sup> that is averaged on the SAXS time scale measurement.

A more specific quantitative analysis of the different US constructs motions reveals that the UIM and SH3 domains tumble with different correlation times. If the linker length is shortened from US-WT to US- $\Delta 2$ , the correlation times associated with UIM and SH3 both decrease while we observe a more pronounced decrease for the UIM domain probably due to the difference of the domain sizes. Moreover, the gap between the  $R_2$  of UIM and SH3 tends to decrease as if UIM would be part of SH3. This is not unexpected as  $R_2$  mainly reflects the global tumbling of the US constructs. Thus, it is likely that the latter domains do not behave independently and that the linker acts like a spring. It restricts the global motion of both domains leading to an increase in their correlation time and a coupling of their tumbling. Other studies have focused on a significant increase in the linker length and have reported a decrease in the correlation between domains when the linker length increases<sup>34</sup>. A similar situation has been observed in the case of identical repeating domains where the correlation times are largely overestimated compared to a situation where each domain is considered in its individual state<sup>42</sup>.

Then, we have investigated the consequences of such flexibility and global tumbling on the interaction with Ub and Lys63-Ub<sub>2</sub>, two proteins that have been shown to interact with US-WT<sup>28,29</sup>. Modifications of the linker length slightly affect the binding of Ub with either UIM or SH3 by a factor of  $\sim 2.5$ . Half or complete deletion of the UIM domain do not affect the Ub binding capability of SH3 while half of UIM is not sufficient to bind Ub. The situation is more dramatic when Lys63-Ub<sub>2</sub> is considered as a US binding partner. Indeed, the UIM domain does not suffer from a shortening of the linker and keeps a similar dissociation constant (see Table 1) while the SH3 domain exhibits a decrease in affinity by a factor of  $\sim 8$  in striking contrast with its interaction with Ub. Overall, it appears that US-WT needs a fine-tuned linker length to enable a suitable avid binding with Lys63-Ub<sub>2</sub> and to avoid steric clash that would reduce US-WT association or dissociation rate. Moreover, one can infer that the US-WT linker length, flexibility and dynamics have been designed to match with the end-to-end length of the Lys63-Ub<sub>2</sub> and more generally with polyubiquitin-linked chains that are the preferred recognition signal for the lysosomal degradation. The linker is uniquely designed to provide a synergy between amino-acids composition, length, flexibility, dynamics and the environment conditions (pH, pressure, temperature for instance) associated with allosteric or cooperative effects. In the case of STAM2, one can hypothesize that the two different intervening linkers between VHS, UIM and SH3 have been adequately shaped to allow at the same time a fast switching between Lys63-polyubiquitin chains and AMSH that transiently interact<sup>29</sup> for a rapid remodeling of the interaction map. This hypothesis is supported by our five-state model that shows a possible interconversion between a wide range of conformations for US-WT. Alteration of interactions with the polyubiquitin tag and deubiquitinating enzymes may cause a defect in the lysosomal degradation process and trigger further disease<sup>43</sup>. The question of the variability of the linker length with respect to molecular recognition is fundamental and previous studies on the Smurf2 systems have also demonstrated that a longer linker decreases the affinity<sup>44</sup> or that the linker length of different GH5 cellulase variants affects their kinetic behavior<sup>45</sup>. Other examples can be found with the linker that joins the tandem UIM domains of human Rap80. The latter one acquires a helical structure when bound to Lys63-Ub<sub>2</sub> and different lengths of the linker give rise to a decrease in affinity<sup>46</sup>. In this study, the tandem UIM can be regarded as a molecular ruler and highlights the importance of its length with respect to polyubiquitin-linked chains selectivity that can undergo various geometries and chain connections<sup>47,48</sup>.

Next, one can wonder if there is a relationship between flexibility, dynamics at the ps-ns time scale and binding affinity. Our results demonstrate that the length of the linker affects directly the global tumbling of both the UIM and SH3 domains and their binding capability with Lys63-Ub<sub>2</sub>. We can hypothesize that the outcome is a reorientation of the connected domains at a favorable rate to be consistent with the association/dissociation rate constant of any binding event. Furthermore, reduced spectral densities indicate a fast local motion located in the linker and its vicinity. This flexibility of the linker tends to reduce with its shortening, as evidenced by the higher values of  $R_2$  and the corresponding lower values of  $J(0.87\omega_H)$ . The interdomain motion may involve large-scale rotations and translations as demonstrated by the different structures sampled to account for the SAXS curves and may involve high energetic barriers. The fast local motion added to the rotational diffusion may encode successive states that help reducing these barriers and expose pre-existing conformations to binding partners. Finally, it is clearly demonstrated that flexibility and dynamics of the linker region have a direct impact on molecular recognition of the STAM2 protein with respect to polyubiquitin chains. It has to be recalled also that the lysosomal degradation sorting process is carried out by a homodimer assembled with STAM and Hrs<sup>49</sup>. The latter one also possesses a VHS and a double-sided UIM domain that add another possibility of cooperative binding. Such a process involves the interaction of the VHS and UIM domains of STAM with Lys63-linked triubiquitin (Lys63-Ub<sub>3</sub>) while the SH3 domain of STAM would bind the SH3 binding domain of AMSH<sup>37,38</sup>. We can anticipate that a shortening of the different linkers of STAM would prevent the correct positioning of polyubiquitin chains and/or AMSH and thus precludes the right catalytic task. Indeed, preliminary results obtained on a VHS-UIM mutant where we have substituted 14 amino-acids of the linker by 14 prolines, show a complete deletion of cooperative effect between VHS-UIM and Lys63-Ub<sub>2</sub>. To conclude, we have shown that the length, flexibility and dynamics of the linker binding the UIM and SH3 domains of STAM2 are fundamental characteristics to accommodate different binding partners where STAM2 can be perceived as a multiple-ligand binding platform that acts at a given time and position to trigger the correct signal outcome. The unique plasticity of the linkers involved in STAM2 also contributes to the possibility of an active/inactive switch through ubiquitin intra-molecular binding<sup>50</sup>. From a more general point of view, the linkers included in multidomain proteins could also be the next generation of druggable target as their modification may reduce or completely abolish interactions.

## Methods

**Protein expression and purification.** The human STAM2 constructs with truncation and mutation (Fig. 1) were designed in pETM60 plasmid with NusA and 6-His tag fused to the N-terminus under the regulation of a *lac* operon and have been purchased from Genecust. The plasmids were then transformed into *E. coli* BL21 GOLD (Milipore). Cells were grown in LB medium with 50 mg/l kanamycin or M9 medium supplemented with



1 mM MgSO<sub>4</sub>, 1 mM CaCl<sub>2</sub>, 6 mg/l Thiamine, 1% (v/v) trace element solution [5 g/l EDTA, 0.5 g/l FeCl<sub>3</sub>·6H<sub>2</sub>O, 5 mg/l ZnO, 1 mg/l CuCl<sub>2</sub>·2H<sub>2</sub>O, 1 mg/l Co(NO<sub>3</sub>)<sub>2</sub>·6H<sub>2</sub>O, and 1 mg/l (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>·4H<sub>2</sub>O], 50 mg/ml Kanamycin and 1 g/l <sup>15</sup>NH<sub>4</sub>Cl as sole nitrogen source for a uniform <sup>15</sup>N labelling. For <sup>13</sup>C-labelling, 2.5 g/l of <sup>13</sup>C<sub>6</sub>-D-Glucose were used instead of <sup>12</sup>C<sub>6</sub>-D-Glucose. The cells culture was grown at 37 °C to an A<sub>600</sub> of 0.6–0.8 and the overexpression is induced by adding 1 mM IPTG. After 5 h of induction at 30 °C, cells were lysed in 50 mM Tris buffer, 250 mM NaCl, 10 mM Imidazole, 0.04% (v/v) β-mercapto-ethanol, 5% (v/v) Glycerol and 1 tablet of Complete<sup>®</sup> protease inhibitors from ROCHE. The clarified cells lysate was loaded on a Ni-NTA Fast Flow column (GE Healthcare) equilibrated with 50 mM Tris-HCl (pH 7.8), 250 mM NaCl, 10 mM Imidazole, 1% (v/v) glycerol and 0.04% (v/v) β-mercaptoethanol. The bound protein was eluted with a 10–400 mM imidazole gradient. NusA and His<sub>6</sub> tag were cleaved by TEV protease at 4 °C O/N and discarded by a second Ni-NTA column. Proteins were then purified by a Superdex 75 gel filtration column (GE Healthcare) equilibrated in 20 mM sodium phosphate buffer (pH 6.8) and 130 mM NaCl. The elution peak was desalted and concentrated in a Microcon concentrator tube with 3 kD cut-off. Ubiquitin and Lys63-Ub<sub>2</sub> chains were prepared according to the previously published method<sup>51</sup>.

**Homology modeling of UIM-SH3 constructs.** The 3D structure of the UIM part (if any) in the UIM-SH3 constructs was obtained by homology modeling following a methodology similar to the modeling of the VHS-UIM construct<sup>27</sup>. The amino acid sequences of STAM2-UIM and Vps27-UIM1 share 55% identity and 70% similarity. We used the UIM1 domain<sup>52</sup> of Vps27 (PDB code 1Q0V) to model the structure of the UIM part of the UIM-SH3 construct, while the STAM2 SH3 domain (PDB code 1X2Q) was used to model the SH3 part of the UIM-SH3 construct. Models were generated by using the Modeller program<sup>53</sup> and by considering the linker region as flexible. After an alignment of the query and template sequences with Align2D, the UIM1 domain of Vps27 and the SH3 domain of STAM2 were used as input in Modeller. A total of ten structures were generated for the different UIM-SH3 constructs.

**NMR experiments.** NMR experiments were performed on a Bruker Avance III operating at a <sup>1</sup>H resonance frequency of 600 MHz (14.1 Teslas) equipped with a triple TCI cryoprobe. Sample temperature was set to 15 °C to be consistent with our previous studies on this system. The NMR samples have been exchanged into a buffer containing 20 mM sodium phosphate (pH 6.8), 10% D<sub>2</sub>O and 0.02% (w/v) NaN<sub>3</sub>.

For US-Δ4 construct, the backbone resonance assignment was carried out by using a combination of the following experiments: HNCO, HN(CA)CO, HNCACB and HN(CO)CACB.

Relaxation measurements including <sup>15</sup>N longitudinal (R<sub>1</sub>), transverse (R<sub>2</sub>) relaxation as well as the <sup>15</sup>N-<sup>1</sup>H heteronuclear cross-relaxation rates were performed using the previous published method<sup>54</sup>. NMR spectra were recorded with spectral widths of 2069 Hz in the <sup>15</sup>N dimension and 9615 Hz in the <sup>1</sup>H dimension. For the R<sub>1</sub> experiments, we used relaxation delays ranging from 40 to 2400 ms with a recycling delay of 2 s. In the case of R<sub>2</sub> experiments, we used relaxation delays from 8 to 224 ms with a recycling delay of 3 s. For heteronuclear NOE experiments, 2D spectra were recorded with and without presaturation of amide protons. The relaxation delay was set to 4.5 s in order to allow the bulk water magnetization to return as close as possible to the equilibrium state.

**NMR titration studies.** The affinities of the different constructs interacting with Lys63-Ub<sub>2</sub> or Ub were characterized using chemical shift perturbation (CSP). A series of <sup>1</sup>H-<sup>15</sup>N HSQC were recorded for <sup>15</sup>N-labeled UIM-SH3 constructs upon addition of non-labeled Lys63-Ub<sub>2</sub> or Ub until saturation. To derive the corresponding dissociation constant, we analyzed CSPs by calculating the combined amide CSP as  $\Delta\delta = [((\Delta\delta_H)^2 + (\Delta\delta_N/5)^2)]^{1/2}$  where  $\delta_H$  and  $\delta_N$  are the chemical change in <sup>1</sup>H and <sup>15</sup>N dimension upon Lys63-Ub<sub>2</sub> or Ub addition. Two models described by Varadan *et al.* have been used to extract dissociation constant<sup>55</sup>. In the case of US-Δ3 and US-Δ4, it was not possible to discriminate perturbations from either the distal or the proximal domain of Lys63-Ub<sub>2</sub> on the SH3 domain. Therefore, we used a 1:1 model for US-Δ3 and US-Δ4 upon Lys63-Ub<sub>2</sub> addition and for all US constructs upon Ub addition. The observed CSPs is a weighted average between the two extreme values corresponding to the free ( $\Delta\delta = 0$ ) and the bound state ( $\Delta\delta = \Delta\delta_{\text{bound}}$ ):

$$\Delta\delta = \Delta\delta_{\text{bound}}([L_0] + [P_0] + K_D - \sqrt{([L_0] + [P_0] + K_D)^2 - 4[L_0][P_0]})/2[P_0] \quad (1)$$

In the case of US-WT, US-Δ1 and US-Δ2, we used a 2:1 model for which each domain is interacting with one unit of Lys63-Ub<sub>2</sub> and the corresponding equation is:

$$\Delta\delta = \Delta\delta_{\text{bound}}(2[L_0] + [P_0] + K_D - \sqrt{(2[L_0] + [P_0] + K_D)^2 - 8[L_0][P_0]})/2[P_0] \quad (2)$$

All NMR data were processed with NMRPipe<sup>56</sup>, analyzed with Sparky<sup>57</sup> or CcpNmr<sup>58</sup> and relaxation rates were extracted using Relaxfit<sup>54</sup>. The different correlation times were derived by means of the Rotdif software<sup>59,60</sup>.

**SAXS experiments.** The measurements were performed at the ESRF BioSAXS beamline BM29 (Grenoble, France), using a 2D Pilatus detector at an X-ray wavelength  $\lambda = 1.008 \text{ \AA}$  with a standard single instrumental configuration (samples being automatically mounted to a capillary and 10 frames with 1 s exposure using the flow-through mode) at 20 °C. Data processing and reduction were performed using an automated standard ESRF beamline software (BSxCuBE)<sup>61</sup> and PRIMUS<sup>62</sup> while the overall parameters derived from SAXS data were processed with SCATTER<sup>63</sup>. Wild type UIM-SH3 and the truncated forms (US-Δ1, US-Δ3 and US-Δ4) were exchanged into a 20 mM Tris-HCl Buffer (pH 8.0) with 150 mM NaCl and 1 mM DTT. To eliminate any inter-particle effects, data were measured at 3–5 different concentrations (1, 2, 4, 8, and 10 mg/ml) and merged

where needed. The final proteins were prepared at different concentrations in the range between ~0.5 and 10 mg/ml to obtain high quality data from both the low angle range (low protein concentration to accurately extract the radii of gyration) and from the high angle range (high protein concentration for an accurate solvent subtraction).

**MultiFoXS N-state modeling.** In an effort to characterize the range of conformations consistent with the SAXS data for US-WT, US- $\Delta$ 1, US- $\Delta$ 3 and US- $\Delta$ 4, we analyzed the distribution of  $R_g$  through MultiFoXS N-state modeling<sup>64</sup>. As an input, we have provided the pdb structure calculated by Modeller (see paragraph above) that gave the best  $\chi$  score from FoXS modelling<sup>32</sup>. Flexible residues have been defined according to Table ST4 and 10000 conformers have been sampled. In the first step, MultiFoXS samples the input structures with a RRT algorithm<sup>64</sup> that significantly improves the sampling efficiency compared to random sampling. As a second step, a SAXS profile is calculated for each sampled conformation. Finally, the 1000 top N-states models are sorted according to their  $\chi$  values.

**Circular dichroism.** The far-UV CD experiments were performed on a Chirascan circular dichroism (CD) spectrometer (Applied Photophysics, Ltd) using a cuvette with a path length of 1 mm at room temperature. The instrument parameters were set to a step size of 0.2 nm, a spectral bandwidth of 0.5 nm, a time-per-point of 1 s. The different samples were in 20 mM sodium phosphate buffer, pH 6.8 and were measured at a concentration of 5–10  $\mu$ M. The CD spectrum of each sample was then subtracted to the buffer, normalized to their respective concentration and then converted to the mean residue molar ellipticity unit. The final spectra were deconvoluted using the CDSSTR algorithm<sup>65</sup> with the reference set 7 available at DICHROWEB (<http://dichroweb.cryst.bbk.ac.uk/html/home.shtml>).

## References

1. Scott, J. D. & Pawson, T. Cell signaling in space and time: where proteins come together and when they're apart. *Science* **326**, 1220–1224, <https://doi.org/10.1126/science.1175668> (2009).
2. Vidal, M., Cusick, M. E. & Barabasi, A. L. Interactome networks and human disease. *Cell* **144**, 986–998, <https://doi.org/10.1016/j.cell.2011.02.016> (2011).
3. Lage, K. Protein-protein interactions and genetic diseases: The interactome. *Biochim Biophys Acta* **1842**, 1971–1980, <https://doi.org/10.1016/j.bbadis.2014.05.028> (2014).
4. Arkin, M. R., Tang, Y. & Wells, J. A. Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem Biol* **21**, 1102–1114, <https://doi.org/10.1016/j.chembiol.2014.09.001> (2014).
5. Watkins, A. M. & Arora, P. S. Structure-based inhibition of protein-protein interactions. *Eur J Med Chem* **94**, 480–488, <https://doi.org/10.1016/j.ejmech.2014.09.047> (2015).
6. Deribe, Y. L., Pawson, T. & Dikic, I. Post-translational modifications in signal integration. *Nat Struct Mol Biol* **17**, 666–672, <https://doi.org/10.1038/nsmb.1842> (2010).
7. Yu, L. *et al.* Grammar of protein domain architectures. *Proc Natl Acad Sci USA* **116**, 3636, <https://doi.org/10.1073/pnas.1814684116> (2019).
8. Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C. & Teichmann, S. A. Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* **14**, 208–216, <https://doi.org/10.1016/j.sbi.2004.03.011> (2004).
9. Apic, G., Huber, W. & Teichmann, S. A. Multi-domain protein families and domain pairs: comparison with known structures and a random model of domain recombination. *J Struct Funct Genomics* **4**, 67–78, <https://doi.org/10.1023/A:1026113408773> (2003).
10. Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A. & Clarke, J. The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol* **8**, 319–330, <https://doi.org/10.1038/nrm2144> (2007).
11. Papaleo, E. *et al.* The Role of Protein Loops and Linkers in Conformational Dynamics and Allostery. *Chem Rev* **116**, 6391–6423, <https://doi.org/10.1021/acs.chemrev.5b00623> (2016).
12. Reddy Chichili, V. P., Kumar, V. & Sivaraman, J. Linkers in the structural biology of protein-protein interactions. *Protein Sci* **22**, 153–167, <https://doi.org/10.1002/pro.2206> (2013).
13. Ma, B., Tsai, C. J., Halliloglu, T. & Nussinov, R. Dynamic allostery: linkers are not merely flexible. *Structure* **19**, 907–917, <https://doi.org/10.1016/j.str.2011.06.002> (2011).
14. Delaforge, E. *et al.* Investigating the Role of Large-Scale Domain Dynamics in Protein-Protein Interactions. *Front Mol Biosci* **3**, 54, <https://doi.org/10.3389/fmolb.2016.00054> (2016).
15. Hellenkamp, B., Wortmann, P., Kandzia, F., Zacharias, M. & Hugel, T. Multidomain structure and correlated dynamics determined by self-consistent FRET networks. *Nat Methods* **14**, 174–180, <https://doi.org/10.1038/nmeth.4081> (2017).
16. Bernado, P. Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur Biophys J* **39**, 769–780, <https://doi.org/10.1007/s00249-009-0549-3> (2010).
17. Sonntag, M. *et al.* Segmental, Domain-Selective Perdeuteration and Small-Angle Neutron Scattering for Structural Analysis of Multi-Domain Proteins. *Angew Chem Int Ed Engl* **56**, 9322–9325, <https://doi.org/10.1002/anie.201702904> (2017).
18. Gobl, C., Madl, T., Simon, B. & Sattler, M. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog Nucl Magn Reson Spectrosc* **80**, 26–63, <https://doi.org/10.1016/j.pnmrs.2014.05.003> (2014).
19. Thompson, M. K., Ehlinger, A. C. & Chazin, W. J. Analysis of Functional Dynamics of Modular Multidomain Proteins by SAXS and NMR. *Methods Enzymol* **592**, 49–76, <https://doi.org/10.1016/bs.mie.2017.03.017> (2017).
20. Khan, S. N. *et al.* Distribution of Pico- and Nanosecond Motions in Disordered Proteins from Nuclear Spin Relaxation. *Biophys J* **109**, 988–999, <https://doi.org/10.1016/j.bpj.2015.06.069> (2015).
21. Viles, J. H. *et al.* Local structural plasticity of the prion protein. Analysis of NMR relaxation dynamics. *Biochemistry* **40**, 2743–2753, <https://doi.org/10.1021/bi002898a> (2001).
22. Matsunaga, Y. *et al.* Dimensionality of Collective Variables for Describing Conformational Changes of a Multi-Domain Protein. *J Phys Chem Lett* **7**, 1446–1451, <https://doi.org/10.1021/acs.jpcclett.6b00317> (2016).
23. Roy, A., Hua, D. P. & Post, C. B. Analysis of Multidomain Protein Dynamics. *J. Chem. Theory Comput* **12**, 274–280, <https://doi.org/10.1021/acs.jctc.5b00796> (2016).
24. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci USA* **115**, E4758–E4766, <https://doi.org/10.1073/pnas.1800690115> (2018).
25. Henne, W. M., Buchkovich, N. J. & Emr, S. D. The ESCRT Pathway. *Dev Cell* **21**, 77–91, <https://doi.org/10.1016/j.devcel.2011.05.015> (2011).
26. Erpapazoglou, Z., Walker, O. & Haguenaer-Tsapis, R. Versatile roles of k63-linked ubiquitin chains in trafficking. *Cells* **3**, 1027–1088, <https://doi.org/10.3390/cells3041027> (2014).
27. Lange, A. *et al.* Evidence for Cooperative and Domain-specific Binding of the Signal Transducing Adaptor Molecule 2 (STAM2) to Lys(63)-linked Diubiquitin. *J Biol Chem* **287**, 18687–18699, <https://doi.org/10.1074/jbc.M111.324954> (2012).

28. Lange, A. *et al.* Competitive binding of UBPY and ubiquitin to the STAM2 SH3 domain revealed by NMR. *FEBS Lett* **586**, 3379–3384, <https://doi.org/10.1016/j.febslet.2012.07.047> (2012).
29. Hologne, M. *et al.* NMR Reveals the Interplay among the AMSH SH3 Binding Motif, STAM2, and Lys63-Linked Diubiquitin. *J Mol Biol* **428**, 4544–4558, <https://doi.org/10.1016/j.jmb.2016.10.002> (2016).
30. Veronique, R.-B. & Dominique, D. How Random are Intrinsically Disordered Proteins? A Small Angle Scattering Perspective. *Curr Protein Pept Sci* **13**, 55–75, <https://doi.org/10.2174/138920312799277901> (2012).
31. Durand, D. *et al.* NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J Struct Biol* **169**, 45–53, <https://doi.org/10.1016/j.jsb.2009.08.009> (2010).
32. Schneidman-Duhovny, D., Hammel, M. & Sali, A. FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res* **38**, W540–544, <https://doi.org/10.1093/nar/gkq461> (2010).
33. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. Accurate SAXS profile computation and its assessment by contrast variation experiments. *Biophys J* **105**, 962–974, <https://doi.org/10.1016/j.bpj.2013.07.020> (2013).
34. Bae, S. H., Dyson, H. J. & Wright, P. E. Prediction of the rotational tumbling time for proteins with disordered segments. *J Am Chem Soc* **131**, 6814–6821, <https://doi.org/10.1021/ja809687r> (2009).
35. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A. & Kay, L. E. Spectral density function mapping using <sup>15</sup>N relaxation data exclusively. *J Biomol NMR* **6**, 153–162, <https://doi.org/10.1007/BF00211779> (1995).
36. Lauwers, E., Jacob, C. & Andre, B. K63-linked ubiquitin chains as a specific signal for protein sorting into the multivesicular body pathway. *J Cell Biol* **185**, 493–502, <https://doi.org/10.1083/jcb.200810114> (2009).
37. Davies, C. W., Paul, L. N. & Das, C. Mechanism of recruitment and activation of the endosome-associated deubiquitinase AMSH. *Biochemistry* **52**, 7818–7829, <https://doi.org/10.1021/bi401106b> (2013).
38. Baiady, N. *et al.* The Vps27/Hrs/STAM (VHS) Domain of the Signal-transducing Adaptor Molecule (STAM) Directs Associated Molecule with the SH3 Domain of STAM (AMSH) Specificity to Longer Ubiquitin Chains and Dictates the Position of Cleavage. *J Biol Chem* **291**, 2033–2042, <https://doi.org/10.1074/jbc.M115.689869> (2016).
39. Goldenberg, D. P. Computational simulation of the statistical properties of unfolded proteins. *J Mol Biol* **326**, 1615–1633, [https://doi.org/10.1016/S0022-2836\(03\)00033-0](https://doi.org/10.1016/S0022-2836(03)00033-0) (2003).
40. Segall, D. E., Nelson, P. C. & Phillips, R. Volume-exclusion effects in tethered-particle experiments: bead size matters. *Physical review letters* **96**, 088306, <https://doi.org/10.1103/PhysRevLett.96.088306> (2006).
41. Krieger, F., Fierz, B., Bieri, O., Drewello, M. & Kiefhaber, T. Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J Mol Biol* **332**, 265–274, [https://doi.org/10.1016/S0022-2836\(03\)00892-1](https://doi.org/10.1016/S0022-2836(03)00892-1) (2003).
42. Walsh, J. D., Meier, K., Ishima, R. & Gronenborn, A. M. NMR studies on domain diffusion and alignment in modular GB1 repeats. *Biophys J* **99**, 2636–2646, <https://doi.org/10.1016/j.bpj.2010.08.036> (2010).
43. Tomas, A., Futter, C. E. & Eden, E. R. EGF receptor trafficking: consequences for signaling and cancer. *Trends Cell Biol* **24**, 26–34, <https://doi.org/10.1016/j.tcb.2013.11.002> (2014).
44. Chong, P. A., Lin, H., Wrana, J. L. & Forman-Kay, J. D. Coupling of tandem Smad ubiquitination regulatory factor (Smurf) WW domains modulates target specificity. *Proc Natl Acad Sci USA* **107**, 18404–18409, <https://doi.org/10.1073/pnas.1003023107> (2010).
45. Ruiz, D. M., Turowski, V. R. & Murakami, M. T. Effects of the linker region on the structure and function of modular GH5 cellulases. *Sci Rep* **6**, 28504, <https://doi.org/10.1038/srep28504> (2016).
46. Sims, J. J. & Cohen, R. E. Linkage-specific avidity defines the lysine 63-linked polyubiquitin-binding preference of rap80. *Mol Cell* **33**, 775–783, <https://doi.org/10.1016/j.molcel.2009.02.011> (2009).
47. Castaneda, C. A. *et al.* Linkage-specific conformational ensembles of non-canonical polyubiquitin chains. *Phys Chem Chem Phys* **18**, 5771–5788, <https://doi.org/10.1039/c5cp04601g> (2016).
48. Nakasone, M. A., Livnat-Levanon, N., Glickman, M. H., Cohen, R. E. & Fushman, D. Mixed-linkage ubiquitin chains send mixed messages. *Structure* **21**, 727–740, <https://doi.org/10.1016/j.str.2013.02.019> (2013).
49. Takahashi, H., Mayers, J. R., Wang, L., Edwardson, J. M. & Audhya, A. Hrs and STAM function synergistically to bind ubiquitin-modified cargoes *in vitro*. *Biophys J* **108**, 76–84, <https://doi.org/10.1016/j.bpj.2014.11.004> (2015).
50. Hoeller, D. *et al.* Regulation of ubiquitin-binding proteins by monoubiquitination. *Nat Cell Biol* **8**, 163–169, <https://doi.org/10.1038/ncb1354> (2006).
51. Pickart, C. M. & Raasi, S. Controlled synthesis of polyubiquitin chains. *Methods Enzymol* **399**, 21–36, [https://doi.org/10.1016/S0076-6879\(05\)99002-2](https://doi.org/10.1016/S0076-6879(05)99002-2) (2005).
52. Swanson, K. A., Kang, R. S., Stamenova, S. D., Hicke, L. & Radhakrishnan, I. Solution structure of Vps27 UIM-ubiquitin complex important for endosomal sorting and receptor downregulation. *EMBO J* **22**, 4597–4606, <https://doi.org/10.1093/emboj/cdg471> (2003).
53. Sali, A. & Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**, 779–815, <https://doi.org/10.1006/jmbi.1993.1626> (1993).
54. Fushman, D., Cahill, S. & Cowburn, D. The main-chain dynamics of the dynamin pleckstrin homology (PH) domain in solution: analysis of <sup>15</sup>N relaxation with monomer/dimer equilibration. *J Mol Biol* **266**, 173–194, <https://doi.org/10.1006/jmbi.1996.0771> (1997).
55. Varadan, R. *et al.* Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *J Biol Chem* **279**, 7055–7063, <https://doi.org/10.1074/jbc.M309184200> (2004).
56. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293, <https://doi.org/10.1007/BF00197809> (1995).
57. SPARKY 3 (University of California, San Francisco).
58. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696, <https://doi.org/10.1002/prot.20449> (2005).
59. Walker, O., Varadan, R. & Fushman, D. Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from <sup>15</sup>N relaxation data using computer program ROTDIF. *J Magn Reson* **168**, 336–345, <https://doi.org/10.1016/j.jmr.2004.03.019> (2004).
60. Berlin, K., Longhini, A., Dayie, T. K. & Fushman, D. Deriving quantitative dynamics information for proteins and RNAs using ROTDIF with a graphical user interface. *J Biomol NMR* **57**, 333–352, <https://doi.org/10.1007/s10858-013-9791-1> (2013).
61. Pernot, P. *et al.* Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J Synchrotron Radiat* **20**, 660–664, <https://doi.org/10.1107/S0909049513010431> (2013).
62. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* **36**, 1277–1282, <https://doi.org/10.1107/S0021889803012779> (2003).
63. Förster, S., Apostol, L. & Bras, W. Scatter: software for the analysis of nano- and mesoscale small-angle scattering. *J Appl Crystallogr* **43**, 639–646, <https://doi.org/10.1107/S0021889810008289> (2010).
64. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res* **44**, W424–429, <https://doi.org/10.1093/nar/gkw389> (2016).
65. Johnson, W. C. Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* **35**, 307–312, [https://doi.org/10.1002/\(SICI\)1097-0134\(19990515\)35:3<307::AID-PROT4>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0134(19990515)35:3<307::AID-PROT4>3.0.CO;2-3) (1999).

66. Joosten, R. P. *et al.* A series of PDB related databases for everyday needs. *Nucleic Acids Res* **39**, D411–419, <https://doi.org/10.1093/nar/gkq1105> (2011).
67. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637, <https://doi.org/10.1002/bip.360221211> (1983).

## Acknowledgements

This work was supported by a grant overseen by the French National Research Agency (ANR) as part of the “OH risque” programme (Metadyn, ANR-14-OHRI-0006-01). The SAXS experiments were performed on beamline BM28 at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful to Petra Pernot and Martha Brennich at the ESRF for providing assistance in using beamline BM29. We are also grateful to Dr. Roland Montserret for assistance in setting up CD experiments. We thank the Rovaltain Foundation and ICL for providing experimental NMR time.

## Author Contributions

M.-H.N. performed the experiments, analyzed the data and revised the manuscript. M.M. performed the experiments and analyzed the data. H.K. performed the SAXS experiments and analyzed the data. F.G. supervised the SAXS experiments and revised the manuscript. O.W. conceived the work and has written the manuscript. M.H. has designed, performed the experiments and analyzed the data and revised the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-51182-0>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019