Molecular recognition of ubiquitin and Lysine 63 linked diubiquitin by STAM2 : the effect of the linkers length and flexibility

Minh-Ha Nguyen Thi

N° National de Thèse : 2019LYSE1321

THESE de DOCTORAT DE L'UNIVERSITE DE LYON

opérée au sein de

l'Université Claude Bernard Lyon 1

Ecole Doctorale 206

Ecole Doctorale de Chimie (Chimie, Procédés, Environnement)

Spécialité de doctorat : Résonance Magnétique Nucléaire

Discipline : Chimie

présentée et soutenue publiquement le 04/12/2019

par **Minh-Ha NGUYEN**

Molecular recognition of ubiquitin and Lysine 63 linked diubiquitin by STAM2
The effect of the linkers length and flexibility

La reconnaissance de l'ubiquitine et de la diubiqutine Lysine 63 par la protéine STAM2
L' effet de la longueur et de la flexibilité des linkers

Devant le jury composé de:

Christina SIZUN	Chargée de Recherche CNRS ISCN, Gif-sur-Yvette	Rapporteur
Christian GRIESINGER	Professeur University of Göttingen	Rapporteur
Adriana MIELE	Professeure Université de Lyon	Examinatrice
Bernhard BRUTSCHER	Directeur de Recherche CEA IBS, Grenoble	Examinateur
Frank GABEL	Chargé de Recherche CEA IBS, Grenoble	Examinateur
Maggy HOLOGNE	Maître de Conférences Université de Lyon	Co-directrice de thèse
Olivier WALKER	Maître de Conférences Université de Lyon	Directeur de thèse

"Great things are not done by one person. They're done by a team of people"
— Steve Jobs (1955-2011)

Acknowledgements

This thesis becomes a reality with considerate help and support of many people. With enormous gratitude, I would like to extend my sincere thanks to all of them.

Foremost, I would like to thank Dr. Christina Sizun and Prof. Christian Griesinger for their time and patience to review and evaluate this work, as well as Prof. Adriana Miele, Dr. Bernhard Brutscher and Dr. Frank Gabel for accepting to be part of my thesis defense committee.

To my thesis supervisors, Dr. Maggy Hologne and Dr. Olivier Walker, I would like to express my dearest appreciation for giving me the chance to work with them, for initiating my enthusiasm toward NMR by sharing their great knowledge and passion. To Olivier, thanks for your precious time, support and guidance through all ups and downs during my three years of thesis with your great sense of humour. To Maggy, thank you for teaching me everything about NMR and also for all those gourmet moments with your great culinary talent.

My special thanks to Marie Martin, who not only helped me tremendously throughout my time at the laboratory but also became a friend to whom I could share my everyday problems. Thanks for all your encouragement and useful critiques of this research work.

I would like to thank all other members of the BIOSYS and INTERACT group, Florence Guillière, Henry Kim, Corinne Sanglr, Moreno Marcellini and Yoan Monneau, for your valuable supports, great ideas and advice.

I would like to thank Dr. Olivier Cala and Dr. Thierry Granjon for their participation in my two thesis monitoring committees and for their great recommendations on the advancement of the thesis.

I wish to acknowledge the help provided by Dr. François-Xavier Cantrelle during my TGIR session at the NMR platform in Lille and the advice given by Dr. Roland Montserret during my circular dichroism experiments at the Institut Biologie et Chimie des Protéines.

I would like to thanks all the staffs at the Institut des sciences analytiques, especially Mai Canehan, Eric Legros, David Tatangelo, Maryline Deregnaucourt who made my time here much easier by providing me with their best support.

To all the members of my ”Tonkin” crew, Mimi, KLinh, Duy, Dao, Hoang, TLinh, Minh, thanks for tagging me along with great relaxing time outside of my stressful work.

To my family, Mom, Dad and my dear sister Mika, without whom I would not start this PhD thesis in the first place, thank you for supporting me unconditionally throughout my life with boundless love.

Finally a special thanks to my dear friend Fz, thank you for accompanying me through those three years, for offering me not only enormous helps in informatics but also a wonderful empathy for my everyday struggle during the thesis.

Abstract

Protein-protein interaction is considered as an important field of research, as it is the key to control variable cell processes and pathways. In eucaryotic cells, multidomain proteins (MDPs), which consist of more than one domain, take up over 70 % of the pool. Those identical or different domains of a MDP are connected to each other by a linker of variable length and flexibility. For long flexible linker, it allows the protein to sample a wide range of conformation and to adjust interaction in a subtle way.

Despite numerous efforts of research on the field, some issues remain unanswered or require further investigation. As part of this thesis, my work aims to define the role taken by the intrinsically disordered linker within MDPs. For that purpose, the STAM2 (Signal transducing adapter molecule 2) protein of the ESCRT (Endosomal Sorting Complexes Required for Transport) machinery was chosen to examine the effect of the flexibility and dynamics of the linker regions on the molecular recognition with ubiquitin and Lysine63-linked di-ubiquitin (K63-Ub_2). Such efforts were carried out by designing specific mutants altering the linker regions in different ways. The various truncated versions undergo half or complete deletion of a domain or have their linker either shortened, deleted or modified in the amino acid composition. With a combination of the several biophysical methods namely NMR (Nuclear Magnetic Resonance) spin relaxation, SAXS (Small Angle X-ray Scattering) and CD (Circular Dichroism), the study has demonstrated that the alteration in the linker region modifies the flexibility and the dynamics of the protein, one among them possibly introduces slight change in conformation. Furthermore, the modification of the linker has an impact on the inter-domain motion and alter binding affinities between STAM2 constructs and di-ubiquitin without affecting domains integrity or binding sites.

In brief, disordered linkers provide plasticity to the protein, which allow adaptability and specificity to molecular recognition process. As a further application, the linkers included in multidomain proteins could also be the next generation of druggable target as their modification may reduce or completely abolish interactions.

Résumé

Les interactions protéine-proteine sont considérées comme un domaine de recherche important puisqu'elles contrôlent la plupart des processus cellulaires. Chez les cellules eucaryotes, les protéines multi-domaines (MDP), constituées d'au moins deux domaines, représentent plus de 70 % des protéines. Au sein d'une MDP, ces domaines peuvent être identiques ou différents et sont reliés par un segment intrinsèquement désordonné de longueur et de flexibilité variable. Ces protéines peuvent alors adopter de multiples conformations dans l'espace et interagir de manière spécifique avec leurs partenaires biologiques.

Malgré de nombreux efforts de recherche dans le domaine, certaines questions restent encore non résolues ou nécessitent une étude approfondie. Mon projet de recherche est d'étudier et de définir le rôle des segments intrinsèquement désordonnés de la protéine STAM2 (Signal transducing adapter molecule 2) impliquée dans la machinerie ESCRT (Endosomal Sorting Complexe Required for Transport) , première étape dans le processus de dégradation lysosomale. Plus précisément, l'étude se focalise sur les effets de la flexibilité et la dynamique de ces segments dans le cas du processus de reconnaissance moléculaire entre STAM2 et l'ubiquitine ou di-ubiquitine. Différents mutants ont alors été conçus : soit avec un domaine totalement ou partiellement supprimé, soit avec un raccourcissement ou une suppression complète du segment ou soit avec de multiples mutations dans la séquence peptidique du segment. Ces différents construits ont été analysés en utilisant une combinaison de techniques biophysiques telles que la relaxation de spin par résonance magnétique nucléaire (RMN), la diffusion des rayons X aux petits angles (SAXS) et le dichroïsme circulaire (CD). Il a alors été démontré qu'une altération du segment désordonné peut entraîner un changement de la dynamique de la protéine et/ou un changement conformationnel. La modification de ce segment influe sur le mouvement inter-domaine et modifie l'affinité entre les construits de STAM2 et la di-ubiquitine sans modifier l'intégrité de chaque domaine et de leur site de liaison.

En résumé, les segments intrinsèquement désordonnés procurent une certaine plasticité à la protéine ce qui lui permet de s'adapter et de remplir sa fonction biologique. Il est alors possible d'imaginer dans un futur proche que ces segments soient la nouvelle génération de cibles thérapeutiques pouvant réduire ou supprimer certaines interactions nocives.

Contents

List of abbreviations

3D	Three dimension
6His	Hexa Histidine tag
AMSH	Associated Molecule with the SH3 domain
CD	Circular Dichroism
CSA	Chemical Shift Anisotropy
CSP	Chemical shift perturbation
DTT	Dithiotreitol
EGFR	Epidermal growth factor receptor
ESCRT	Endosomal transducing adaptor molecule
FRET	Förster resonance energy transfer
HRS	Hepatocyte growth factor-regulated tyrosine kinase substrate
IPTG	Isopropyl β-D-1 thiogalactopyranosid
K_d	Dissociation constant
LB	Luria-Bertani
MDP	Multidomain protein
MVB	Multivesicular body
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
PB	Phosphate buffer
ppm	part per million
PTM	Post-translational modification
SANS	Small angle neutron scattering
SAXS	Small angle X-ray scattering

STAM2	Signal Transducing Adaptor Molecule 2
TEV	Tobacco Etch Virus
Tris	Tris(hydroxymethyl) aminomethane
Ub	Ubiquitin
Ub_2	Di-ubiquitin
UBD	Ubiquitin binding domain
UBPY	Ubiquitin-specific protease 8
UIM	Ubiquitin Interacting Motif
v/v	volume per volume
w/v	weight per volume

List of Figures

List of Tables

1

Introduction

Cells communication is crucial for their survival. They need to speak with each other and the information is received, exchanged and processed from the external environment, mostly by protein-protein interaction [1]. Alteration of this protein-protein network usually leads to diseases [2] [3] [4], which makes it important to understand the mechanism of their interaction. In 2004, the human genome was fully sequenced for the coding part [5]. While the genome stays rather static except mutations that occur throughout its evolution, the proteome, which is the full set of proteins produced by that particular genome blueprint, is much more complex [6]. Indeed, depending on the cell type and different life stage, proteins are produced continually in response to the external as well as internal events. In theory, one gene should be coding for one protein, which is not the real case due to various factors such as alternative splicing [7] and post-translational modifications (PTMs) [8] [9]. The proteome therefore is highly dynamic and requires different approaches [10] [11] to fully decipher it,

other than sequencing as for the genome.

Moreover, proteins could be represented as a modular structure, where the blocks are named domains. In fact, more than 70 % of the eukaryotic proteins consist of more than one domain and they are called multi-domain proteins (MDPs), which also evolves continually [12] [13]. Via mutation, duplication and recombination of various domains out of an initial repertoire [14] [15], the diversity of the proteome becomes highly larger. In addition, identical or different domains in MDP are usually connected to each other by disordered segments, characterized by variable lengths and sequences [16] that are called linkers (Example in Figure 1.1). The benefits of having multiple domains over a single domain come from their flexible conformation but also from the stability and folding advantages [17]. Furthermore, the combination of different domains makes the MDPs able to execute sequential steps of interaction and function.

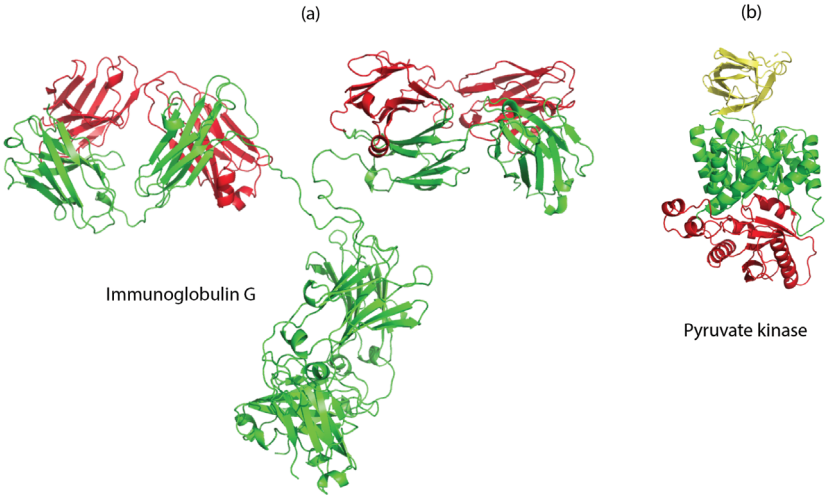


Figure 1.1 – Examples of multidomain proteins with two types of linker are shown in this figure: (a) the immunoglobulin G and its heavy chain (colored in green) which is a MDP with different domains that are linked by long flexible linkers and (b) the pyruvate kinase enzyme, also a MDP but with very short linkers in a more compact conformation.

In the case of multidomain proteins with long flexible linkers, they possess the characteristics of both intrinsically disordered and well-folded proteins that makes their study more complicated [18]. To enhance the understanding of MDPs, a large number of techniques have been practiced, namely FRET [19], SAXS [20], SANS [21], NMR [22] or a combination of the mentioned techniques [18] [23]. Regardless of the extensive research in the field, some issues remain unanswered or require further investigation, one of which is the question on the function of the linker flexibility and dynamics with respect to molecular recognition processes like allostery, cooperativity or avidity. To answer that question in the present study, a biological system was chosen that involves the STAM2 (Signal transducing adapter molecule 2) protein and its different binding partners within the ESCRT (Endosomal Sorting Complexes Required for Transport) machinery.

1. The ESCRT machinery

The ESCRT machinery is a complex system of signaling which involves variable mechanisms [24] [25] [26]. In fact the machinery consists of different cellular processes such as endocytosis of transmembrane protein and growth factor upon ligand/stimuli binding [27], formation of exosome for secretion [28] or even autophagy [29] for regulation of cell homeostasis (Figure 1.2). As part of this study, the endocytosis pathway will be solely focused. In general, upon ligand binding, the trans-membrane receptors/proteins go through either conformational change or post-translational modifications such as ubiquitination or phosphorylation, which initiates the signaling chain reaction starting by their internalization. The process of endocytosis could be done in different mechanisms: clathrin-mediated [30] [31], caveolin-mediated [32] [33] or clathrin- and cavelolin-independant [34]. Regardless of the mechanism, the internalized membrane proteins are destined to the early endosome then transferred to the late endosome through maturation [35]. During this stage the fate of the internalized proteins will be either recycling [36] back to the plasma membrane and the trans-golgi network by following the tubular system or being transported to the lysosomal compartment for degradation. This fate is indeed determined by the type of sustained post-translational modifications. For instance, LDLRs (Low-Density Lipoprotein Receptors)

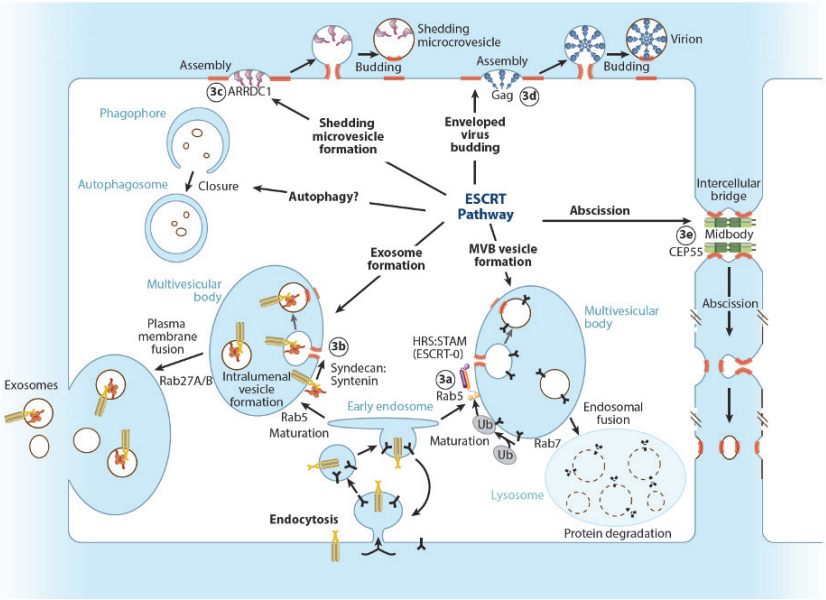


Figure 1.2 – Schematic presentation of the ESCRT pathway [24].

are recycled back to the membrane without any signal [37] while several proteins are recognized by their cytoplasmic peptide sequence [38]. In this current study, the focus will be placed on the Lysine 63-linked ubiquitin chains (K63-Ub chains) [39] [40] that tag and guide a protein toward the lysosomal degradation pathway. In fact, upon ubiquitination, the proteins will travel through different complexes of the ESCRT machinery namely: ESCRT-0, ESCRT-I, ESCRT-II, ESCRT-II, and Vsp4-Vta1 [41] [42]. It has been proved that the ESCRT-0 complex, which locates upstream of the pathway, is able to cluster ubiquitinated cargo [43] [44] and sequentially recruits the following complexes. Together with Hrs, the protein of interest STAM2 belongs to this ESCRT-0 complex [45] and executes its recruitment function by its ability to recognize K63-Ub chains (See Figure 1.3).

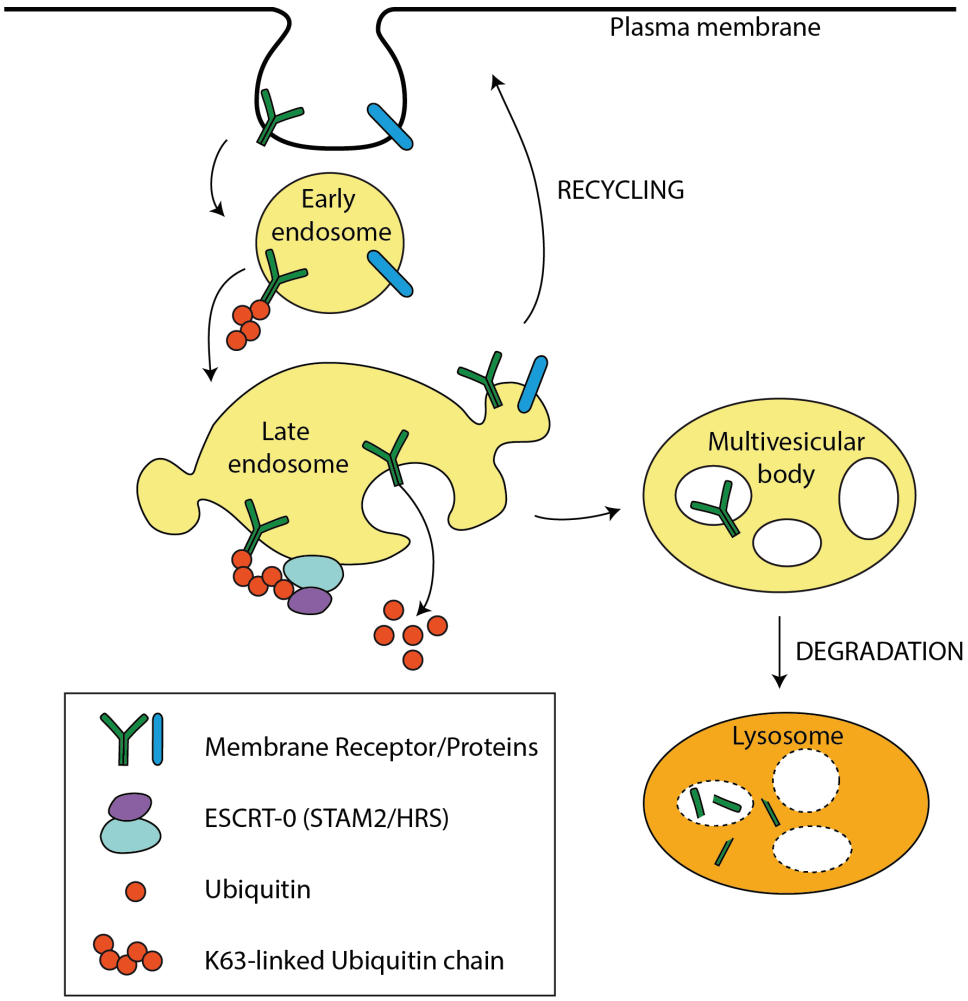


Figure 1.3 – Simplified schematic presentation of the lysosomal degradation pathway that involve STAM2/Hrs of ESCRT-0 and K63-linked ubiquitin chain as signal.

2. STAM2 and the binding partners: Ubiquitin and the Lysine 63-linked di-Ubiquitin(K63-Ub_2 chain)

STAM2, as mentioned above, is part of the ESCRT-0 complex that is located in the first step of the endocytic pathway in early endosome. STAM2 contains of 525 amino acids and harbors multiple domains such as: VHS (VPS-27, Hrs and STAM), SH3 (SRC Homology 3, UIM (Ubiquitin Binding Motif) and ITAM (Immunoreceptor tyrosine-based activation motif). The first three domains VHS, UIM and SH3 together consist of the minimum requirement for the protein STAM2 (Figure 1.4) to ensure the recognition of the protein binding partner and its proper function. VHS is a widely known domain that contains a series of α-helices which fold together to a super helix (PDB structure 1X5B) while the structure of SH3 is constituted of antiparallel β-sheets, several turns, an RT loop and short 3_{10} helices (PDB structure 1X2Q). As for the UIM domain, it is commonly known as an α-helix with a conserved sequence pattern throughout its different versions in variable protein [46].

The three domains are all ubiquitin binding domain (UBD) that can recognize and interact with Ub moieties. In a previous work, it has been demonstrated how different fragments of STAM2 interact with Ub and Lysine 63-linked diubiquitin chains (K63-Ub_2). VHS-UIM (VU-WT) and UIM-SH3 (US-WT) have shown an exquisite capability to bind Ub and Lys63-Ub_2 [47]. US-WT is additionally able to bind to the deubiquitylase enzymes UBPY and AMSH through its SH3 domain [48] [49]. The different structures that have been solved in previous studies are represented in Figure 1.5.

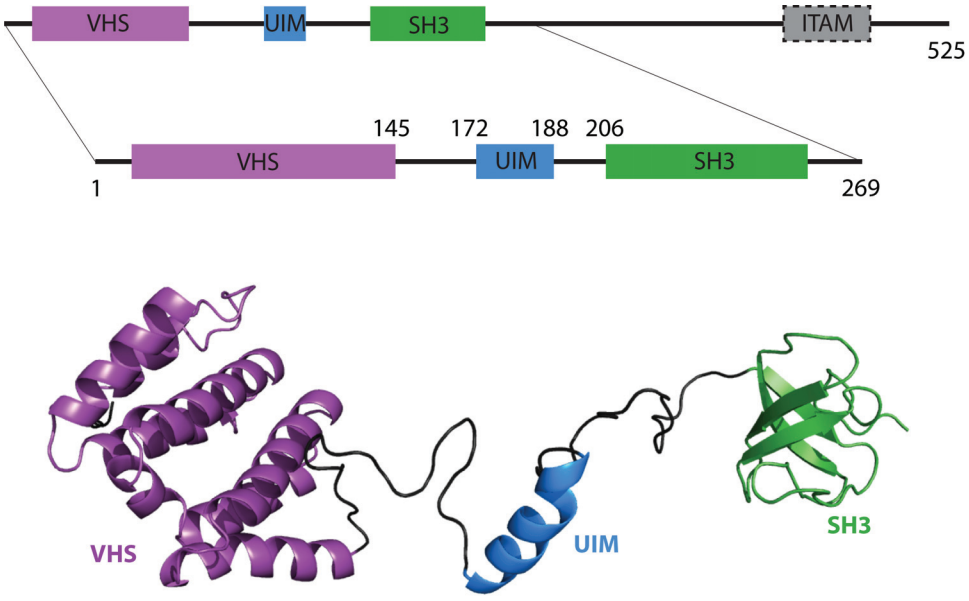


Figure 1.4 – The three domains of STAM2 VHS, UIM and SH3 in schematic and cartoon representation.

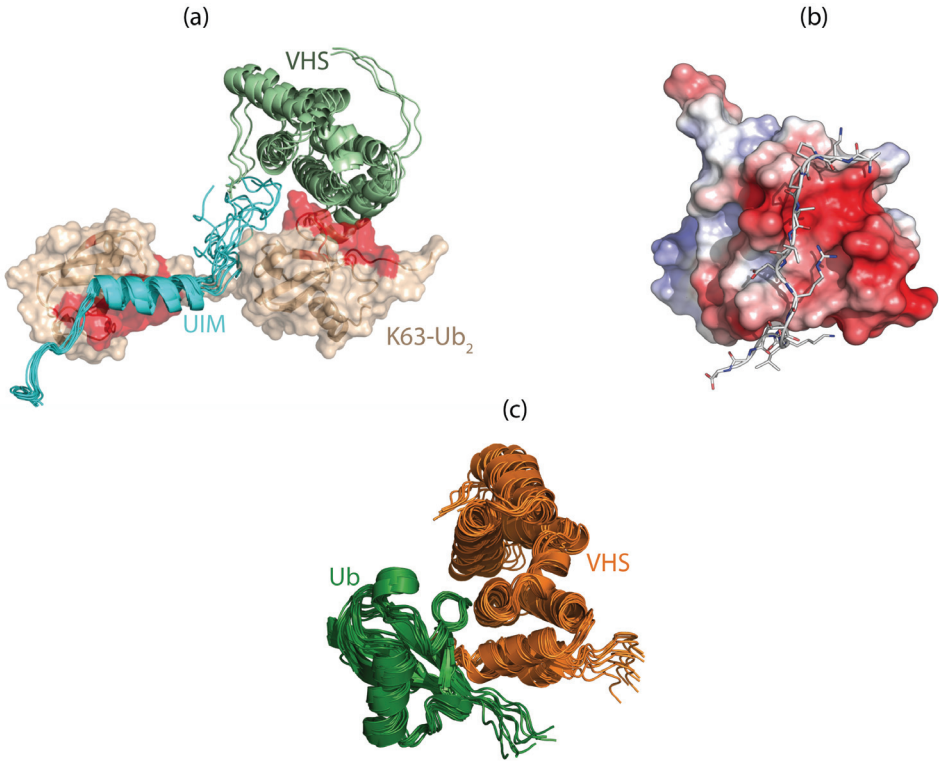


Figure 1.5 – The solved structures of the interaction between different partners: (a) NMR structure of VHS-UIM/K63-Ub_2 complex [48], (b) the NMR structure of SH3 domain (surface) and the SBM peptide issued from AMSH deubiquitylase (stick) [49] and (c) the NMR structure of VHS/Ub complex [50].

Ubiquitin is a small protein of 76 amino acids that could be found in most eukaryotic cell types. Ubiquitin or a chain of ubiquitins can be attached to other protein by covalent linkage, the process of which is called ubiquitination, and acts as a signal for various cell pathway [51]. While monomer ubiquitination is widely known to be the signal for clathrin-mediated endocytosis or chromatin remodeling [52], Lysine48-linked ubiquitin chain (K48-Ub chain) is the canonical marker for the degradation pathway by the 26S proteasome [53] [54]. As for K63-Ub chain, it has also been shown to be involved in various cellular processes such as DNA repair [55] [56] and endosomal sorting [57] [58].

3. The problematics of the current study

As mentioned before, this study aims to elucidate the role of the disordered linker in protein-protein interaction. STAM2 was chosen to be the system of interest, as its two fragments (VHS-UIM and UIM-SH3) are investigated with an emphasis on the linker region by using a combination of different approaches. The linkers are first modified in different ways: to be shortened, mutated by substitution or even removed completely. The mutants with the modified linker are further examined by means of a combination of different biophysical tools, such as (i) circular dichroism where an average contribution of different secondary structure could be extracted (ii) SAXS where the information on the conformation space sampled by the constructs could be obtained with a resolution of $\approx$ 10 Å and (iii) NMR where the dynamical properties at the ps-ns time scale could be described through spin relaxation experiments. Finally, NMR titration experiments complete the study where the interaction of Ub and K63-Ub_2 with all the newly designed constructs is visualized and the affinity is quantified. This study intends to draw a clear picture of the connection between the molecular recognition ability of multidomain and its plasticity by having long flexible linker.

2

Methods

1. Modeling

The 3D structure of the various constructs was obtained by homology modeling following a methodology similar to the modeling of the VHS-UIM construct described in the previous work of Lange *et al.* [47]. The UIM1 domain [59] of Vps27 (PDB code 1Q0V) was used to model the structure of the UIM part of all constructs (VHS-UIM, UIM-SH3 and mutants) while the STAM2's SH3 domain (PDB code 1X2Q) was used to model the SH3 part and the STAM2's VHS domain (PDB code 1X5B) was used to model the VHS part. Models were generated by using the Modeller program [60] and by considering the linker region as flexible. After an alignment of the query and template sequences with Align2D, the template sequences (PBD structures 1Q0V, 1X2Q, 1X5B) were used as input in Modeller. A total of ten structures were generated for the different constructs.

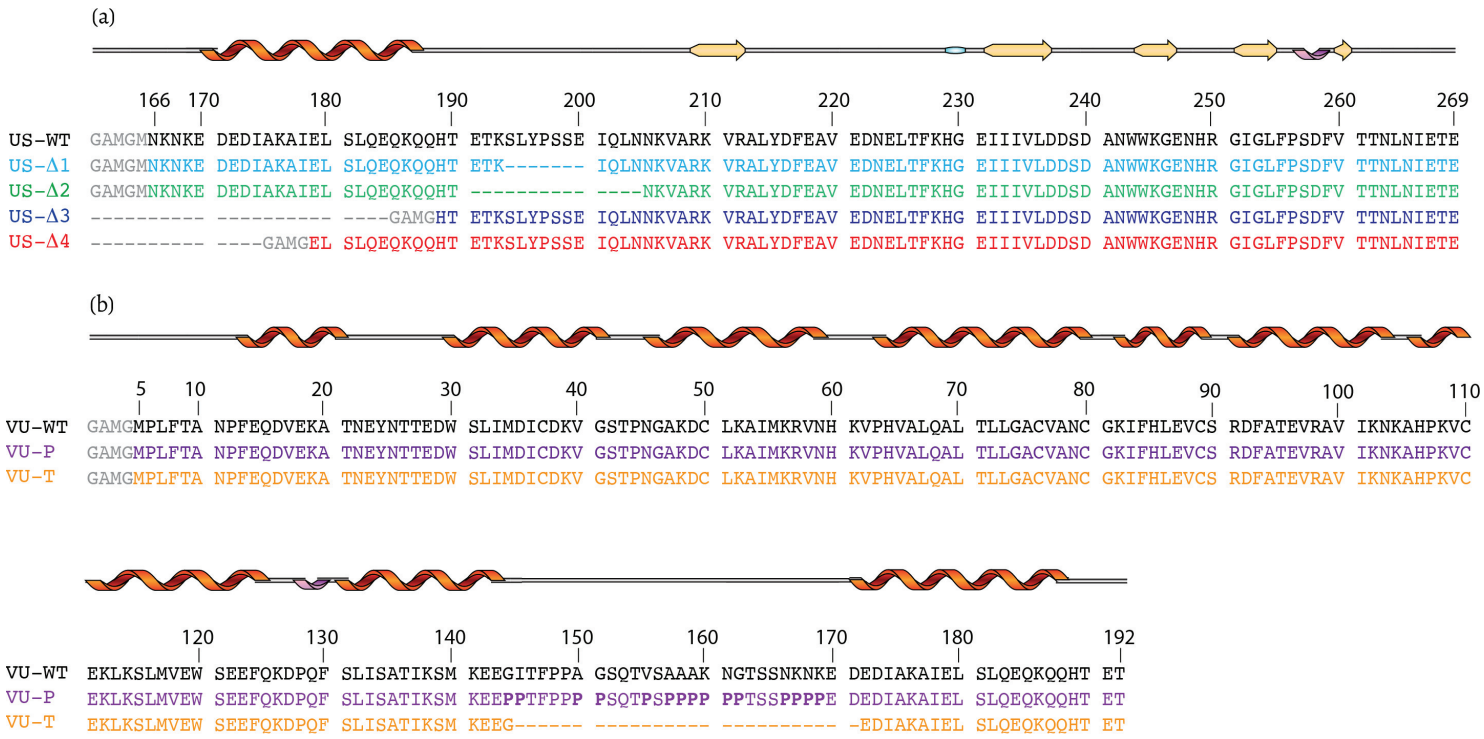


Figure 2.1 – Sequences and names of the different constructs used in the present study along with their respective sequence and numbering. The schematic representation of the secondary structure of UIM-SH3 constructs (a) and VHS-UIM (b) has been obtained by means of dssp [61] [62] embedded into cartoon representation program SSS-drawer.

2. Protein production

2.1. STAM2 constructs

The 2 human STAM2 constructs: STAM2 162-269 (US-WT) and STAM2 1-192 (VU-WT) in pETM-60 vector were provided by Dr. Daniela Höeller. The mutant constructs were engineered where the linker that bridges UIM and SH3 or VHS and UIM has been modified in different ways. For US-WT derived constructs, 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 only SH3 domain with the full length linker, half of UIM remains in US-Δ4 (see Fig 2.1a). As for VU-WT derived constructs, the linker between VHS and UIM was completely deleted in VU-T, while 13 amino acids of that linker were substituted by prolines in VU-P (see Figure 2.1b). Finally, on an attempt to solve the structure of the UIM domain, a mutant called UIM-Y was also designed with a single mutation at the position of Threonine 191 where it is substituted by a Tyrosine to make the construct visible under UV 280 nm spectroscopy and to facilitate its purification.

All constructs were expressed with their N-terminus fused to NusA and a hexa-histidine tag (6His-tag) under the regulation of a *lac* operon and have been purchased from the Genecust company. The plasmids containing the recombinant proteins were then transformed into *E. coli* BL21 GOLD (Milipore). Cells were grown in LB medium with 50 μg/ml kanamycin or M9 medium supplemented with 1 mM $MgSO_4$, 1 mM $CaCl_2$, 6 mg/l thiamine, 1 % (v/v) trace elements solution [5 g/l EDTA, 0.5 g/l $FeCl_3.6H_2O$, 5 mg/l ZnO, 1 mg/l $CuCl_2.2H_2O$, 1 mg/l $Co(NO_3)_2.6H_2O$, and 1 mg/l $(NH_4)_6Mo_7O_{24}.4H_2O$], 50 μg/ml kanamycin and 1 g/l $^{15}NH_4Cl$ as sole nitrogen source for an uniform ^{15}N labelling. For ^{13}C-labelling, 2.5 g/l of $^{13}C_6$-D-Glucose were used instead of $^{12}C_6$-D-Glucose. The cells culture was grown at 37 °C to an $OD_{600\ nm}$ of 0.6-0.8 and the overexpression is induced by adding 1 mM IPTG and the induction is performed at 30 °C for 5 hours.

For VU-WT and its derived mutants, partial deuteration ($\approx$ 80 %) is needed to reduce spin diffusion. The deuteration protocol is adapted from Dr. Christina Sizun protocol. The

cells were first grown in rich LB medium to create the biomass then transferred to 2 successive M9 media: 100 % H_2O then 100 % D_2O. The last culture in D_2O-M9 medium was grown at 37 °C until $OD_{600\ nm}$ reached 0.8. The overexpression is induced by addition of 1 mM IPTG and by an over-night incubation at 20 °C.

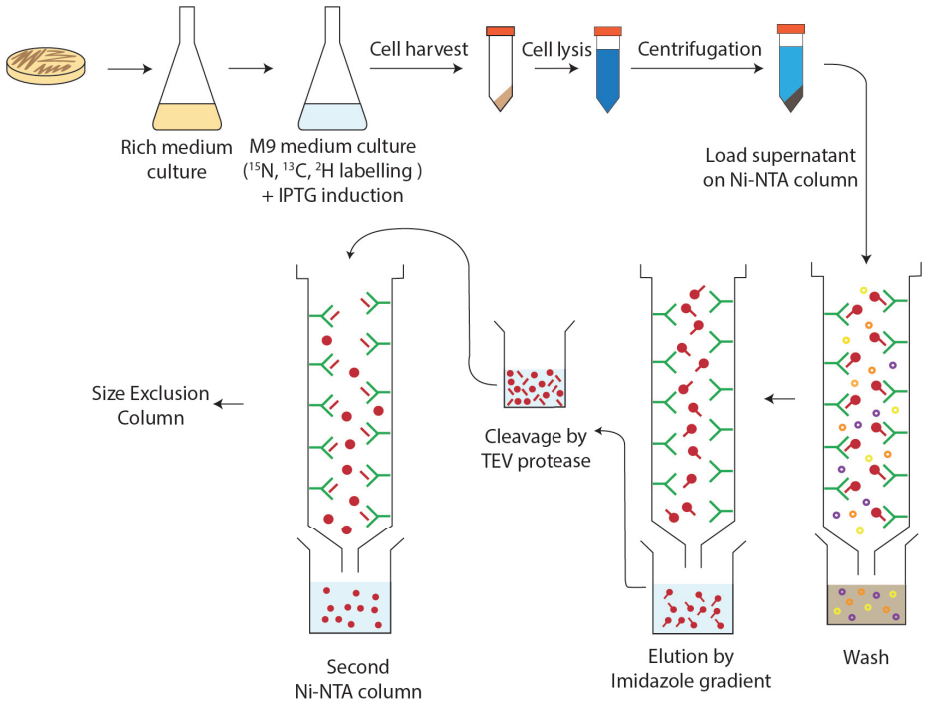


Figure 2.2 – A simplified schematic representation of the overexpression and purification of STAM2 and mutants with or without isotope labelling.

Cells were harvested by centrifugation at 3500 rpm (Beckman JA-20) and the pellets were resuspended in lysis buffer (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 the C0mplete® protease inhibitors from ROCHE) and lyzed by mild sonication. The lysate was clarified by centrifugation at 12500 rpm (Beckman JA-20) for 30 minutes. The clarified cells lysate was loaded on a His-trap Sepharose™ Fast Flow column (GE Healthcare) equilibrated with load buffer (50 mM Tris-HCl (pH 7.8), 250 mM NaCl, 10 mM imidazole, 1 %(v/v) glycerol and 0.04 %(v/v) β-mercapto-ethanol. The bound protein was eluted with a imidazole gradient and comes out of the column at around 125-150 mM imidazole . 6His-NusA were cleaved by

Tobacco Etch Virus (TEV) protease at 4°C over night while dialyzing against load buffer. The TEV site cleavage leaves 4-5 amino acids (GAMG(M)) at the N-terminus which do not belong to the protein sequence. The dialysate was then applied to the His-trap column for a second time to have 6His-NusA and TEV binding to the column. The flow-through, which contains the cleaved proteins, was concentrated to less then 1 ml, supplemented with 10 % glycerol and further purified by a Hi-load Superdex™ 75 16/60 column (GE Heathcare) equilibrated in 20 mM sodium phosphate buffer (pH 6.8) and 130 mM NaCl. The elution peak was desalted and concentrated in a Amicon® Ultra-15 centrifugal filters with 3 kD cut-off. The purity of the protein was verified by SDS-gel electrophoresis and the protein concentration was determined by measuring the absorbance at 280 nm on the Nanodrop™ using the Protein A280nm program. For each protein, the concentration was calculated using the extinction coefficient estimated by ProtParam [63] on the ExPASy server based on their own sequence.

2.2. Ubiquitin and Lysine 63-linked di-Ubiquitin chains (K63-Ub_2)

2.2.1. Ubiquitin (Ub)

Plasmids for the wild-type ubiquitin were a gift from Prof. David Fushman and the two mutants Ub-K63R and Ub-D77 were provided by Prof. Cecile Pickart (Addgene plasmid #18898 and #18896). Expression and purification of ubiquitin and mutants follow the same protocol. The ubiquitin plasmids were expressed in the Rosetta™(DE3)pLysS strains carrying the chloramphenicol-marked helper plasmid pJY2. A starter culture of 1/100 volume in rich media supplemented with 100 μg/ml ampicillin and 50 μg/ml chloramphenicol was grown for 8 hours at 37 °C then diluted in the auto-induction medium (3.3 g $(NH_4)_2SO_4$, 6.8 g KH_2PO_4, 7.1 g Na_2HPO_4, 10 g tryptone, 5 g yeast extract, 1 mM $MgCl_2$, 0.5 g glucose, 2 g lactose and 5 g glycerol per 1 liter medium) with the same concentration of antibiotics in the previous step. The auto-induced culture was grown over-night at 37 °C. Cells were then harvested and frozen at -80 °C for at least 4 hours before the next step. The frozen cell pellet was then resuspended in 50 mM Tris-HCl pH 7.6, 1 mM PMSF, 0.4 mg/ml lysozyme and 1

tablet of C0mplete® protease inhibitors (Roche). 10 mM $MgCl_2$ and 20 μg/ml DNaseI were added to the lysate and incubated on a shaker for 20 minutes at room temperature. The lysate was clarified by centrifugation at 13000 rpm (Beckman JA-20) for 30 minutes and precipitated by acidification with 0.5 %(v/v) perchloric acid under high speed stirring. The precipitate was removed by centrifugation while the supernatant was dialyzed against 50 mM ammonium acetate buffer pH 4.5 before being loaded onto a 5 ml SP Sepharose™ Fast Flow column (GE Healthcare). Ubiquitin was charged negatively at the acidic pH 4.5 and bound to the cation exchange column as all other contaminants passed through the column. For the elution step, a gradient of NaCl from 0-0.5 M was applied onto the column and ubiquitin comes out at $\approx$ 0.18-0.2 M of NaCl. The fractions were pooled and exchanged into the working buffer according to the type of experiments (See Table 2.1, page 22).

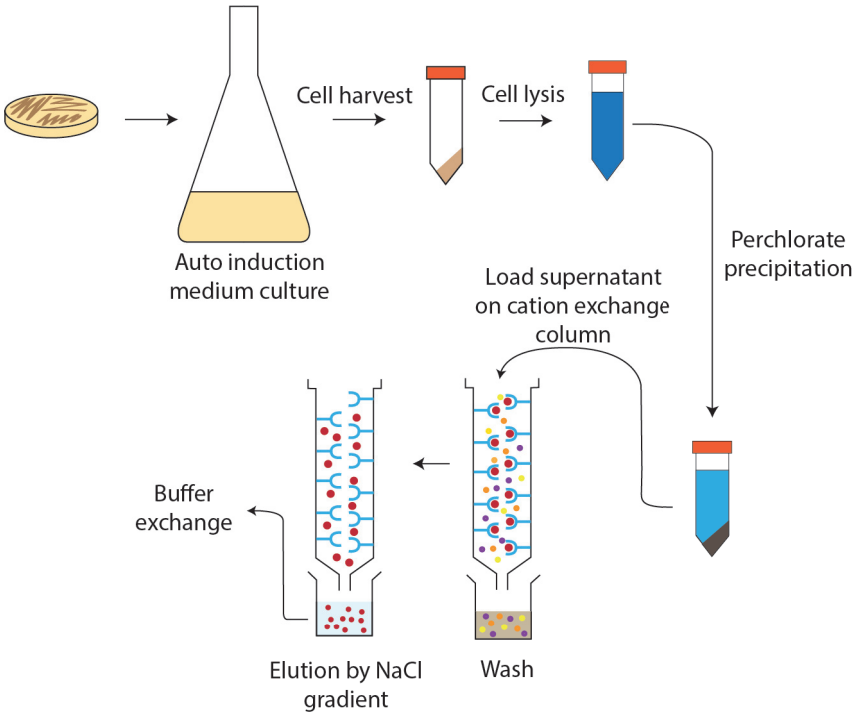


Figure 2.3 – Schematic representation of the auto-induced culture of ubiquitin and the purification using perchloric acid precipitation principle.

2.2.2. Lysine 63-linked di-Ubiquitin chains (K63-Ub_2)

The preparation of K63-Ub_2 was adapted from a previously described method by Pickart & Raasi[64]. Three enzymes were required for the ubiquitination reaction: the human E1 enzymes UBE1 and the two specific E2 enzymes UBC13 and UEV1A for K63-linked chains. Instead of using wild-type ubiquitins, two mutants K63R and D77 were used to restrain the reaction to stop at the formation of a dimer and not to form longer chains.

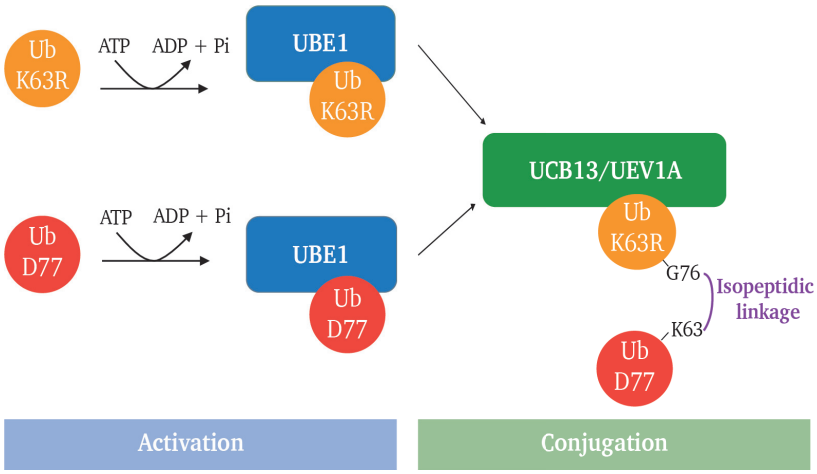


Figure 2.4 – The UbE1 enzyme binds to Ub (D77 or K63R mutant) by using 1 ATP in order to bring them to the conjugation enzyme. The latter is a complex of UBC13 and UEV1A that allow two ubiquitins to form an isopeptidic linkage between the residue G76 of Ub-K63R and K63 of Ub-D77.

UBE1

UBE1 plasmid was provided by Dr. Cynthia Wolberger (Addgene plasmid #34965). The expression of human UBE1 is done in the Rosetta™(DE3)pLysS cells. Cells were grown in rich media until $OD_{600\ nm}$ reached 0.6-0.8 then the induction is performed with 1 mM IPTG at 20 °C for 5 hours. The purifications were performed on ice throughout every step.

Cold lysis buffer (50 mM Tris-HCl pH 7.8, 250 mM NaCl, 1 mM PMSF, 1 % Triton X-100, 0.4 mg/ml lysozyme and 1 tablet of C0mplete® protease inhibitors) was added to the frozen pellet. For the lysis, the cell pellet was resuspended, then 20 μg/ml DNaseI and 10 mM $MgCl_2$

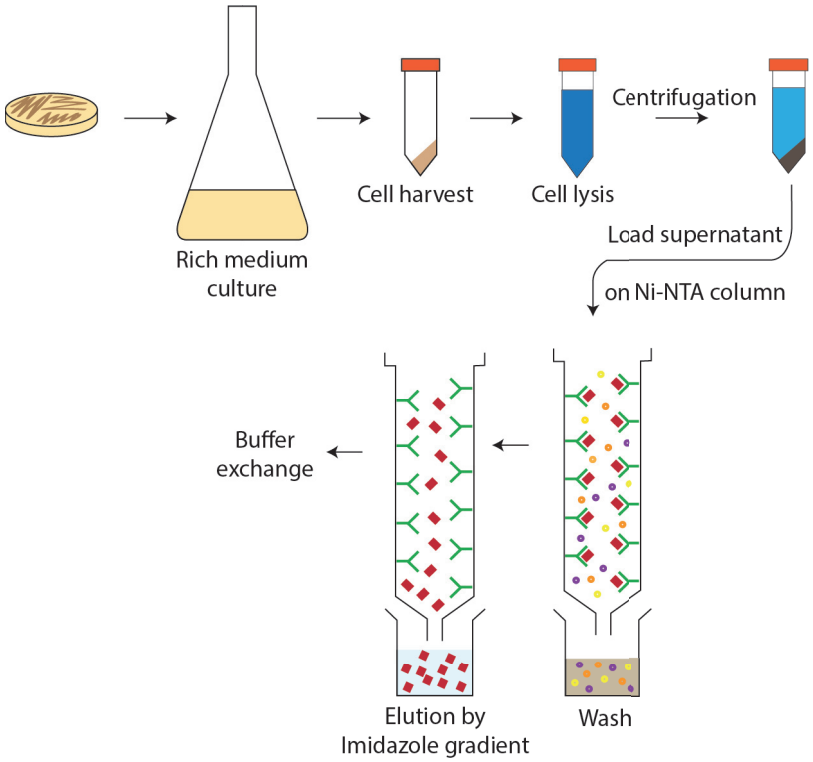


Figure 2.5 – Schematic representation of the 6His tagged E1 ubiquitin enzyme and its purification by affinity column.

were added and the lysate was incubated for 30 minutes on the shaker at 4 °C. The lysate was clarified by centrifugation at 13000 rpm for 30 minutes then 40 mM imidazole was added to the clarified lysate before loading it onto a His-trap Sepharose™ Fast Flow column (GE Healthcare). The column was equilibrated with load buffer (50 mM Tris-HCl (pH 7.8), 250 mM NaCl, 40 mM imidazole, 1 %(v/v) glycerol and 0.04 %(v/v) β-mercapto-ethanol). The bound protein was eluted with a 40-500 mM imidazole gradient. UBE1 comes out of the column at ≈ 150 mM of imidazole. The fractions were pooled together and concentrated in a Microcon® Vivaspin 20R centrifugal filters with 10 kD cut-off while exchanging to 50 mM Tris pH 7.6 for the enzymatic reaction.

E2s UBC13 and UEV1A

GST-UBC13 and GST-UEV1A plasmids were kindly provided by Dr. Michael Glickman. For expressions of the proteins, the plasmids were inserted in *E. coli* BL21 GOLD. The cells were grown in LB media until $OD_{600\ nm}$ = 0.6-0.8 and were induced by 1 mM IPTG for 4 hours at 37 °C. Cells were harvested by centrifugation then lysed by sonication in lysis buffer (50 mM Tris-HCL pH8.0, 5 % glycerol, 1 mM DTT, 0.4 mg/ml lysozyme and 1 tablet of C0mplete® protease inhibitors). The lysates were clarified by centrifugation at 12500 rpm for 30 minutes, then mixed together to form the complex GST-UBC13/GST-UEV1A before being loaded onto a GST Fast Flow column. The bound complexes of GST-UBC13 and GST-UEV1A were eluted by 10 mM reduced-glutathione and exchanged into 50 mM Tris pH 8.0 and concentrated in a Microcon® Vivaspin 20R centrifugal filters with 10 kD cut-off.

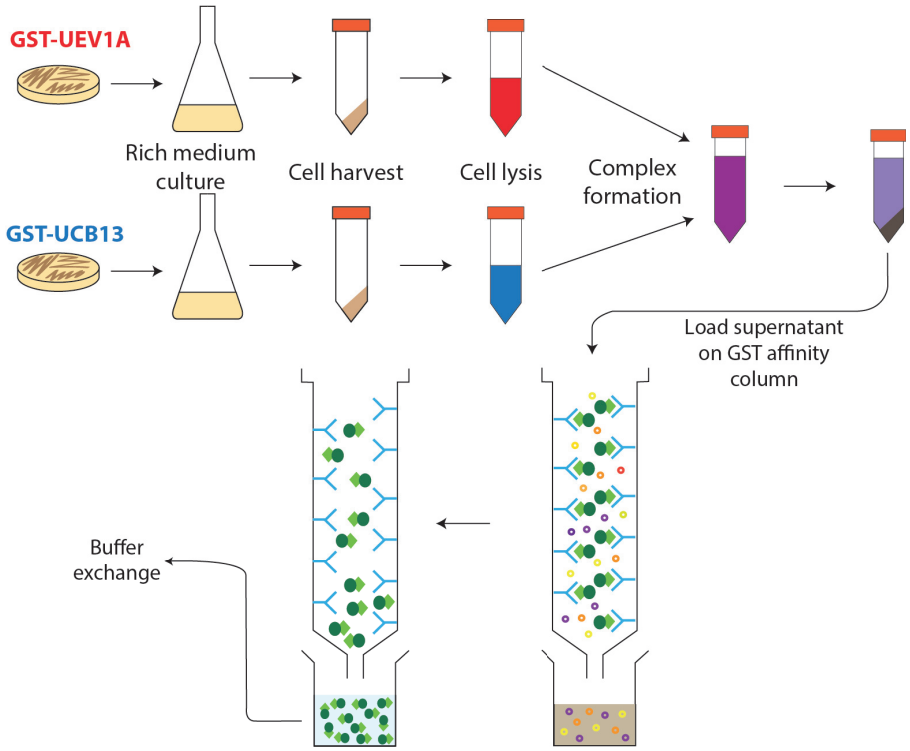


Figure 2.6 – Schematic representation of the purification of the E2s complex of GST-UVE1A/GST-UBC13.

Chain synthesis

The reaction consists of one-fifth volume of PBDM7.6 (250 mM Tris-HCl pH 7.6, 25 mM $MgCl_2$, 50 mM creatine phosphate (Sigma), 3 U/ml inorganic pyrophosphate (Sigma), 3 U/ml creatine phosphokinase (Sigma) and 10 mM ATP), 7 mg/ml of each mutant of ubiquitin K63R and D77, 0.5 mM DTT and 15 μM of the E2s complex GST-UBC13/GST-UEV1A. The reaction is activated by adding 0.1 μM UBE1 and incubated over-night at 37 °C. The reaction stopped by adding 5 mM DTT and 1 mM EDTA to reduce the disulfide links and so the dimers bound to E2 enzymes are detached. After 20 more minutes of incubation at room temperature, one-fifth volume of 2 N acetic acid was added to bring the pH down to $\approx$ 4-4.5. The reaction was then loaded onto a SP Sepharose™ Fast Flow column (GE Healthcare) equilibrated with 50 mM ammonium acetate pH 4.5. A gradient of NaCl was applied to the column and the dimer was eluted at $\approx$ 0.4 M NaCl. To further purify the dimer, a size exclusion column (Hi-load Superdex™ 75 16/60 column) was performed. The final fraction was pooled and exchanged to the working buffer according to the type of experiments (See Table 2.1).

2.3. Working buffers

The buffers were optimized for each set of proteins and for different type of experiments, which are detailed in the following tables. Sodium phosphate buffers were a mix of 55 % base (Na_2HPO_4) and 45 % acid (NaH_2PO_4). As for Tris-HCl buffer, a mix of 60% base and 40% acid was used. For di-Ubiquitin, VU-WT and constructs, due to a high molecular weight and a high structure level, DTT was added to avoid disulfide bonds, thus to prevent aggregation. NaCl was removed to avoid interfering with the recorded signal in circular dichroism (CD) and NMR experiments.

(a) Composition

Buffers	Reagents	pH
PB	20 mM sodium phosphate	6.8
PB DTT	20 mM sodium phosphate, 1 mM DTT	6.8
Tris50	50 mM Tris-HCl	7.6
Tris20 DTT	20 mM Tris-HCl, 150 mM NaCl, 1 mM DTT	8.0

(b) Usage

Proteins sets	Experiments			
	NMR	SAXS	CD	Ub_2 reaction
US-WT & mutants	PB	Tris20 DTT	PB	/
VU-WT & mutants	PB DTT	Tris20 DTT	PB	/
Ub & derived mutants	PB	Tris20 DTT	/	Tris50
K63-Ub_2 chains	PB DTT	Tris20 DTT	/	/
Ubiqutination enzymes	/	/	/	Tris20 DTT

Table 2.1 – The sodium phosphate buffer was used in NMR and CD experiments and Tris-HCl was the optimal buffer for SAXS experiments. 1 mM DTT was added for K63-linked di-ubiquitin, VU-WT and the derived mutants.

3. NMR data collection and processing

All experiments were performed at 288 K and the NMR samples were exchanged into 20 mM phosphate pH 6.8 (see Table 2.1) supplemented with 10 %(v/v) D_2O and 0.02 % (w/v) NaN_3. For US-WT and mutants, NMR experiments were all recorded on a Bruker Avance III 600 MHz spectrometer equipped with a triple TCI cryoprobe at the NMR facility of the CRMN Lyon. As for relaxation measurements, an additional set of experiments was performed on the Bruker Avance NEO 900 MHz spectrometer equipped with a triple TCI cryoprobe at the UCCS - UGSF Lille. The latter was also the spectrometer on which all NMR experiments of VU-WT and mutants were recorded. Data was processed with NMRPipe [65] and analyzed with CCPNMR [66].

3.1. US-Δ4 backbone assignment

The backbone assignment of US-Δ4 was carried out by using a combination of the following experiments: HNCO, HN(CA)CO, HNCACB and HN(CO)CACB (Figure 2.7), which were acquired on a double labeled ^{13}C,^{15}N-US-Δ4 sample. The existing assignment of US-WT was used as a starting point, mostly for the SH3 part.

To assign the carbonyl ^{13}C, we make use of the combination of two experiments HNCO and HN(CA)CO. In the HNCO experiment, magnetization is transferred from ^{1}H to ^{15}N of the residue i, then to the carbonyl ^{13}C of the preceded residue (i-1) by the ^{15}NH–^{13}CO J-coupling and then finally returns back to H_i for detection (Figure 2.7a). This experiment is used in conjunction of the HN(CA)CO where magnetization is passed from $^{1}H_i$ to $^{15}N_i$ and then via the N-Cα J-coupling to the ^{13}Cα of both residue (i) and (i-1) (Figure 2.7b). From there it is transferred to the two ^{13}CO through the ^{13}Cα - ^{13}CO J-coupling. The final signal is detected for ^{13}CO of both residues (i) and (i-1) $^{15}N_i$ and finally $^{1}H_i$.

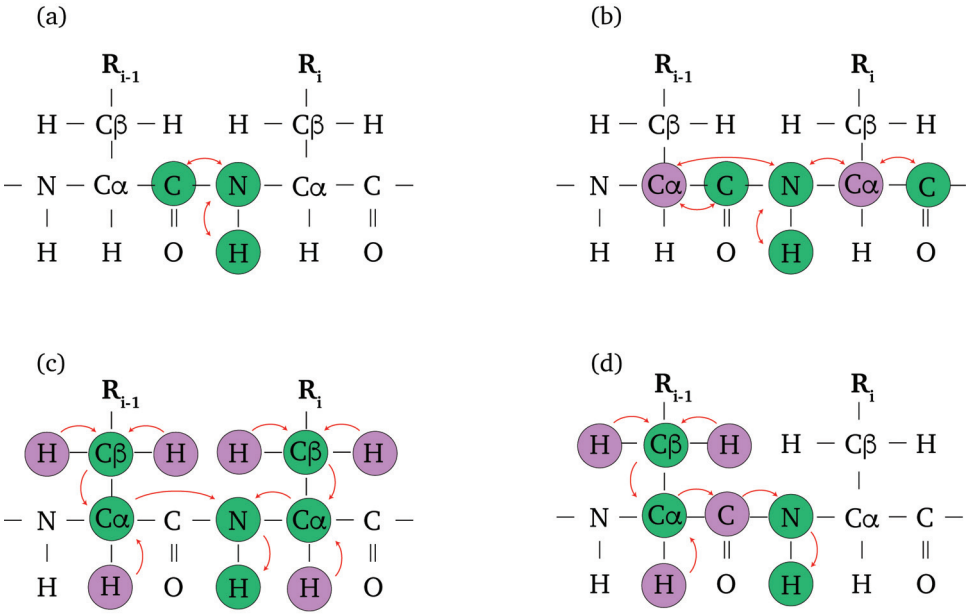


Figure 2.7 – Schematic representation of 3D NMR experiments: HNCO (a), HN(CA)CO (b), HNCACB (c) and HN(CO)CACB (d). The red arrow represents the path of the magnetization transfer. The atoms that are circled in green represents the detected atom with evolved chemical shift while the purple circled atoms stay undetected.

For the Cα and Cβ assignment, two experiments HNCACB and HN(CO)CACB were performed. In the HNCACB experiment (Figure 2.7c), the magnetization is transferred from ^{1}Hα and ^{1}Hβ to ^{13}Cα and ^{13}Cβ of both residue (i) and (i-1). The magnetization is then passed from ^{13}Cβ to ^{13}Cα to reach ^{15}N and finally the amid ^{1}H for detection. The chemical shift is therefore evolved on ^{13}Cα and ^{13}Cβ of both (i) and (i-1) residues, $^{15}N_i$ and $^{1}H_i$ but not on ^{13}CO. The complement of this HNCACB experiment is the HN(CO)CACB (Figure 2.7d) where the magnetization is designed to be transferred from ^{1}Hα and ^{1}Hβ of the residues (i-1) through$^{13}CO_{i-1}$ before reaching $^{15}N_i$ and $^{1}H_i$. The chemical shift is therefore evolved only for ^{13}Cα and ^{13}Cβ of the residue (i-1) plus $^{15}N_i$ and $^{1}H_i$.

3.2. UIM-Y backbone assignment

The backbone assignment of the UIM-Y construct was conducted with the use a set of three experiments: ^{1}H-^{15}N-HSQC, (H)N(COCA)NH (Figure 2.8) and HNCACB (Figure 2.7c). In the (H)N(COCA)NH experiments, the three atom $^{1}H_i$, $^{1}H_{(i-1)}$ and $^{1}HN_{i-1}$ is detected through the magnetization transfer path as shown in Figure 2.8.

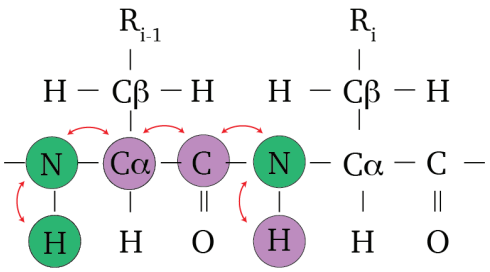


Figure 2.8 – Schematic representation of 3D NMR experiments (H)N(COCA)NH. The red arrow represents the path of the magnetization transfer. The atoms that are circled in green represents the detected atom with evolved chemical shift while the purple circled atoms stay undetected.

3.3. NMR titration experiments

^{1}H-^{15}N HSQC spectra were recorded on the ^{15}N-labeled STAM2 constructs upon addition of ^{14}N-mono-Ubiquitin or ^{15}N-K63-linked di-Ubiquitin as ligands until saturation.

The weighted amide chemical shift perturbations (CSPs) is calculated as followed

$$\Delta\delta = \sqrt{\frac{(\Delta\delta_H)^2 + (a.\Delta\delta_N)^2}{2}} \tag{2.1}$$

where $\Delta\delta_H$ and $\Delta\delta_N$ are the CSPs in the ^{1}H and ^{15}N dimensions in a successive addition of the ligands and a is the scaling factor which range from 0.1 to 0.45 [67]. In this study, a in the equation 2.1 was estimated as 0.2 based on the scaling of ^{1}H spectral width over ^{15}N spectral width. In order to extract the dissociation constant K_d, different stoechiometry models were used according to the previous work of Varadan *et al.* [68] where it is assumed that in fast exchange, the CSPs at a given intermediate state ($\Delta\delta$) can be linked to the bound state CSPs ($\Delta\delta_{bound}$) by a linear function, which varies according to different stoechiometry models.

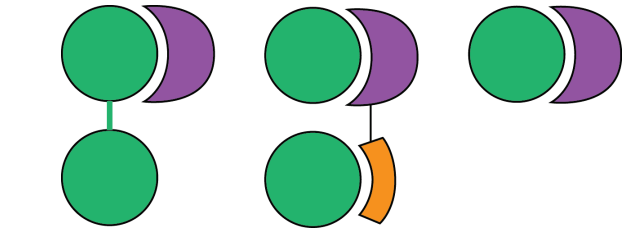
It was found in a previous work that ubiquitin has no preferential binding to SH3 or UIM domain [50]. Considering that one Ub can bind to one of the two domains UIM and SH3 independently, a 1:1 stoechiometry was applied for all US constructs upon Ub addition (Figure 2.9a). In the case of K63-Ub$_2$ titration for US-Δ4 and US-Δ3 where the SH3 domain solely remains intact, the same stoechiometry model was also applied despite the presence of 2 binding sites on the ligand:

$$\Delta\delta = \Delta\delta_{bound}.([L_0] + [P_0] + K_d - \frac{\sqrt{([L_0] + [P_0] + K_d)^2 - 4[L_0][P_0]}}{2[P_0]}) \tag{2.2}$$

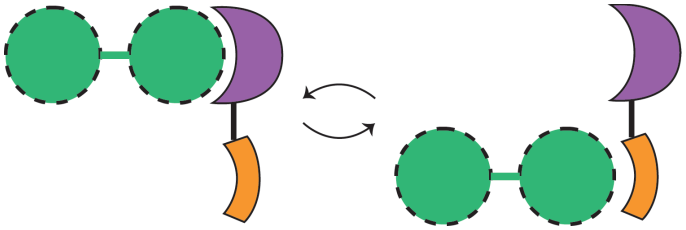
For US-WT, US-Δ1, US-Δ2, VU-WT, VU-P and VU-T upon K63-Ub$_2$ titration, a 2:1 stoechiometry was applied where each domain interacts with one unit of K63-Ub$_2$ (Figure 2.9b)

$$\Delta\delta = \Delta\delta_{bound}.(2[L_0] + [P_0] + K_d - \frac{\sqrt{(2[L_0] + [P_0] + \frac{1}{2}K_d)^2 - 8[L_0][P_0]}}{2[P_0]}) \tag{2.3}$$

The dissociation constant K_d and $\Delta\delta_{bound}$ were then fitted with non-linear regression by using an in-house Matlab (MathWorks) based program. Errors were estimated by sampling 100 initial guesses, assuming 10% error on the protein and ligand concentrations.



(a) 1:1 model



(b) 2:1 model

Figure 2.9 – Stoechiometry models used for the interaction between Ub or K63-Ub_2 and STAM2 constructs. The green round shapes represent ubiquitin, linked together or not while the purple and orange shapes represent different domains of the STAM2 constructs. The interactions in the 1:1 model (a) were considered as simultaneous while the dotted shapes mean a non-simultaneous interaction of the two K63-Ub_2 molecules in the 2:1 model (b).

3.4. NMR relaxation experiments

Spin relaxation is a powerful method which makes a major contribution to research on the dynamical properties of proteins [69] [70]. In the limitation of the present study, only basic definitions and applications of the method will be brought up [71].

In brief, at the end of the excitation by the magnetic field B_1, the total spin of the nuclei in the sample begins to return to the initial equilibrium state which is aligned with the B_0 field. This process corresponds to the disappearance of the transverse component M_{xy} and the restoration of the longitudinal component M_z. The return to equilibrium is therefore characterized by two parameters: the longitudinal relaxation time T_1, which is linked to the evolution of the M_z component and the transverse relaxation time which describes the evolution of the M_x and M_y components. In principle, T_2 is the time it takes for the signal to decay due to the dephasing between multiple nuclei presented in the molecule. In short,

T_1 and T_2 relaxation times have a tight connection to the correlation time (τ_C) of the molecule. The latter is the time it takes to rotate by 1 radian and has a tight connection to the macromolecule size. The relaxation rates R_1 and R_2, which are simply the reverse of T_1 and T_2 respectively, can be obtained by performing an interleaved pseudo-3D experiments previously described by Fushman *et al.* [72]. For the R_1, relaxation delays ranging from 40 to 2400 ms were used with a recycling delay of 2 s. In the case of R_2 experiments, the delays are ranging from 8 to 224 ms with a recycling delay of 3 s. Data were processed with NMRPipe [65] and R_1, R_2 values are extracted in CCPNMR using the "*Follow Intensity changes*" tools [66].

Another parameter used in the present study is the ^{15}N-^{1}H cross-relaxation rates via steady-state ^{15}N-^{1}H NOE (Nuclear Overhauser Effect). In short the heteronuclear NOE (hetero-NOE) corresponds to the cross relaxation effect which results in a magnetization exchange between two spins in dipolar interaction, ^{15}N and ^{1}H spins in this case. The measurement of the hetero-NOEs is done by two HSQC ^{15}N-^{1}H experiments without or with presaturation, where an irradiation is exerted on the ^{1}H nuclei to create a steady state and the perturbation on the longitudinal magnetization of ^{15}N is observed [73] [72]. 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. The hetero-NOE is finally obtained by calculating the ratio of the intensities of the ^{15}N-^{1}H correlation peaks between the two 2D experiments. In the present study, while the distance of the N-H bonds is considered to be known and constant throughout the protein, hetero-NOEs indicate the flexibility of the N-H bonds. The more rigid this linkage is, the less the proton perturbation affects the ^{15}N spin relaxation and the closer the hetero-NOE is to 1. On the contrary, as the linkage becomes more flexible, the ^{15}N spin relaxation is more affected by the proton perturbation and hetero-NOE value deviates farther away from 1. R_1, R_2 and hetero-NOEs are directly related to the motion of a NH vector and are expressed as a linear combination of spectral densities J(ω) that describe the efficiency of a given relaxation mechanism at a given frequency. These parameters are defined as:

$$R_1 = d^2[J(\omega_H - \omega_N) + 6J(\omega_H + \omega_N)] + 3(c^2 + d^2)J(\omega_N) \tag{2.4}$$

$$R_2 = \frac{d^2}{2}[J(\omega_H - \omega_N) + 6J(\omega_H) + 6J(\omega_H + \omega_N)] + (c^2 + d^2)[2J(0) + \frac{3}{2}J(\omega_N)] + R_{ex} \quad (2.5)$$

$$NOE = 1 + d^2(\frac{\gamma_H}{\gamma_N})\frac{[6J(\omega_H + \omega_N) - J(\omega_H - \omega_N)]}{R_1} \quad (2.6)$$

where

$$d = -(\frac{\mu_0}{4\pi})(\frac{\gamma_H \gamma_N h}{4\pi r_{\mathrm{HN}}^3}) \quad (2.7)$$

and

$$c = \frac{\gamma_N B_0 \Delta\sigma}{3} \quad (2.8)$$

r_{HN} is the intermolecular $^{15}N -^1 H$ distance, $\Delta\sigma$ is the anisotropy of the ^{15}N chemical shift tensor, γ_H, γ_N, ω_H, ω_N are the gyromagnetic ratios and resonance frequencies of the nuclei, h is Planck constant, and R_{ex} is the conformational exchange contribution, if any, to the measure of R_2.

A complete description of the different motions occurring at the five resonance frequencies would require the measurement of five independent relaxation parameters. A recent development of Farrow *et al.* [74] gave rise to reduced spectral densities and allows the measurement of three relaxation parameters only. The previous relaxation parameters can be recast as:

$$R_1 = 6.25d^2J(0.87\omega_H) + 3(c^2 + d^2)J(\omega_N) \quad (2.9)$$

$$R_2 = 5.4d^2J(0.87\omega_H) + (c^2 + d^2)[2J(0) + \frac{3}{2}J(\omega_N)] + R_{ex} \quad (2.10)$$

$$(1 - NOE)R_1|\frac{\gamma_N}{\gamma_H}| = 5d^2J(0.87\omega_H) \quad (2.11)$$

A complete analysis of the given relaxation parameters would require the use of given dynamical model. Another issue is to resort to a direct analysis of the different spectral

densities. NH vectors which reorient faster will have a broader distribution for J(ω). Equivalently, smaller J(0) values imply shorter τ_c values and, therefore, more rapid rotational fluctuations of the NH vector. Thus, the respective reduced spectral densities can be expressed as:

$$J(0.87\omega_H) = \frac{(1-NOE)}{5d^2} R_1 |\frac{\gamma_N}{\gamma_H}| \tag{2.12}$$

$$J(\omega_N) = \frac{R_1(1-1.25(1-NOE)|\frac{\gamma_N}{\gamma_H}|)}{3(c^2+d^2)} \tag{2.13}$$

$$J(0) = \frac{R_2 - R_1(|\frac{\gamma_N}{\gamma_H}|(0.455-0.455NOE)+0.5}{2(c^2+d^2)} \tag{2.14}$$

Starting from the spin relaxation parmeters, we can directly derive the corresponding reduced spectral densities.

With the R_1, R_2 and hetero-NOEs in hand, correlation times were analyzed and derived using the ROTDIF software [75] [76] by only considering values of the residues that belong to the secondary structure of the protein.

4. Small Angle X-Ray Scattering (SAXS) experiments

SAXS is a tool that studies the overall shape of biological macromolecules in solution at low resolution ($\approx$ 10 Å) and provides insight into their behavior in space, that makes it the choice of complementary technique. In principle, the sample is illuminated by X-rays and the scattered radiation that contains the information about the global shape of the molecule presented in the solution is captured on a detector. Measurements are made at very small angles, usually between 0.1 degrees and 5 degrees, hence the name of the technique. The experimental configuration of SAXS measurements uses a transmission geometry, in which an intense X-ray beam is first fundamental because the comparatively weak scattering signal of the sample must be measured in the immediate vicinity of the direct beam. It is also essential to use a detector that has high linearity and dynamic range as well as negligible intrinsic

noise.

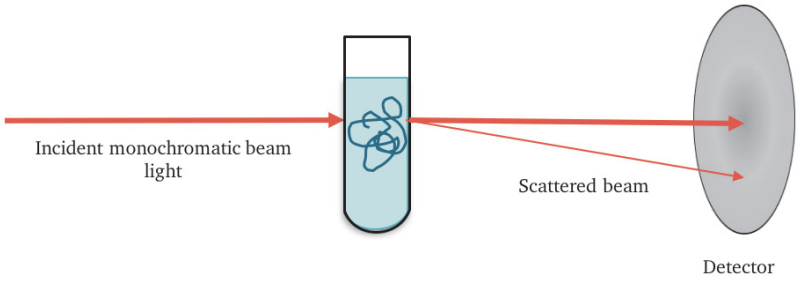


Figure 2.10 – The principle of a SAXS measurement.

The measurements were performed at the ESRF BioSAXS beamline BM29 (Grenoble, France), using a 2D Pilatus detector at an X-ray wavelength $\lambda = 1.008$ Å 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) [77] and PRIMUS [78] while the overall parameters derived from SAXS data were processed with SCÅTTER [79]. US-WT 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. Data were measured at 3-5 different concentrations ranging from 1 to 10 mg/ml to eliminate any inter-particle effects and 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), and merged where needed. In order to find the ensemble of structure and their respective weight that best fit the experimental SAXS data, back calculations were done on the Multifoxs server [80]. Among the initial model generated by Modeller (section 1.), one pdb with the best χ scores was selected as input. Flexible residues have been defined based on previous knowledges of the protein secondary structure and 10,000 conformers have been sampled. At first, MultiFoXS samples the input structures with a RRT algorithm [80] that significantly improves the sampling efficiency compared to random sampling. Afterward, a SAXS profile is calculated for each sampled conformation. Finally, the 1000 top multi-state models are sorted according to their χ values.

5. Circular Dichroism (CD) experiments

The far-UV CD experiments were performed on a Chirascan Circular Dichroism (CD) spectrometer (Applied Photophysics, Ltd) using cuvettes of 1 mm path length at room temperature. The instrument parameters were set to a step size of 0.2 nm, a spectral bandwidth of 0.5 nm and 1 point was measured per second. The different samples were exchanged into 20 mM sodium phosphate buffer, pH 6.8 and were measured at a concentration ranging from 5 to 10 μM. 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 (MRME) unit. The final spectra were deconvoluted using the CDSSTR algorithm [81] with the reference set SMP180 [82] available on the DICHROWEB server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml).

3

UIM-SH3 and the linker length

In order to study the impact of the linker length on the biological and dynamical characteristics of a multi-domain protein, US-WT derived variants were engineered where the linker between UIM and SH3 domains has been shortened by 7 and 14 amino acids (US-Δ1 and US-Δ2 respectively) and either UIM (US-Δ3) or half of UIM (US-Δ4) has been deleted (see Figure 3.1).

A previous study has demonstrated that UIM and SH3 domains do not interact with each other within the US-WT construct [48]. With this in mind, all US-WT derived mutants have been checked to confirm whether they inherit this characteristic. For each construct, a ^{1}H-^{15}N-HSQC spectrum was recorded and compared with the US-WT spectrum.

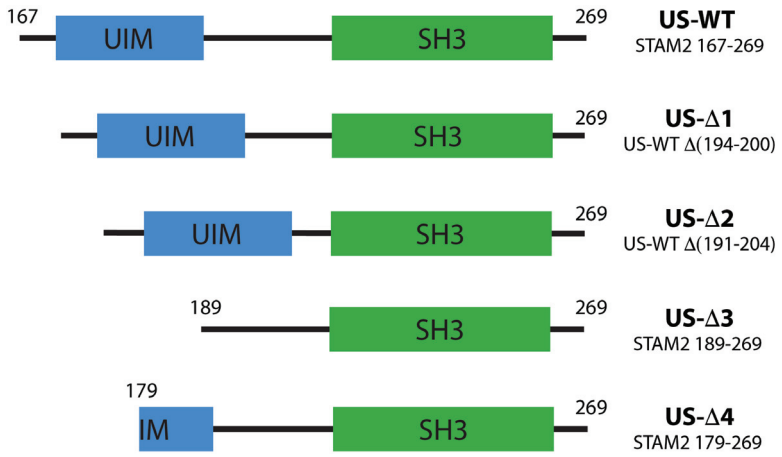


Figure 3.1 – Schematic representation of US-WT, US-Δ1, US-Δ2, US-Δ3 and US-Δ4 constructs.

For US-Δ1 and US-Δ2 a good overlapping with US-WT spectrum was observed in the aligned regions while the peaks that belong to the truncated regions disappeared (Figure 3.2). This result confirms that the linker shortening has been successfully done without affecting the structure of the remaining domains UIM and SH3 and the overlapping means that there is no interaction between the domains, a characteristic that was passed on from US-WT.

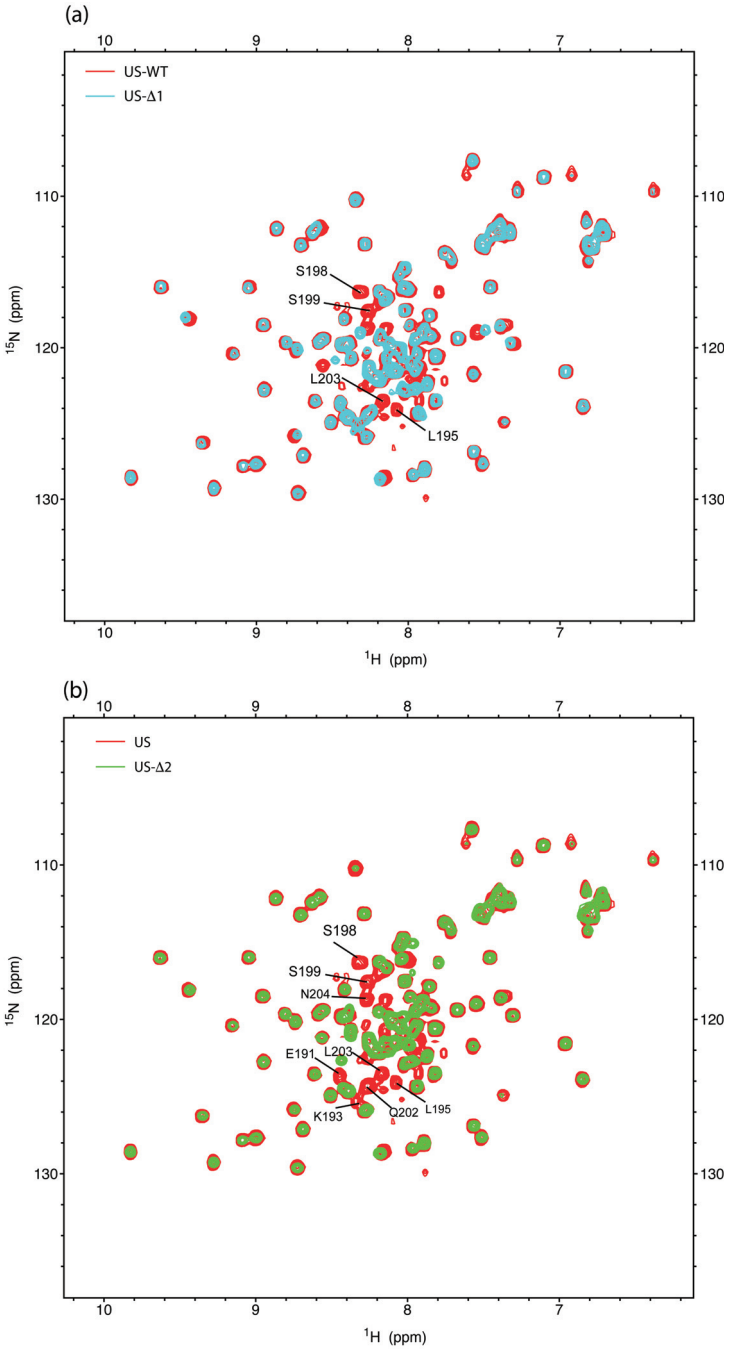


Figure 3.2 – Overlay of ^{1}H-^{15}N-HSQC spectra of US-WT (red) and US-Δ1 (blue) (a) and US-WT and US-Δ2 (green) (b). For residues present in both US-WT and US-Δ1 or US-Δ2, correlation peaks nicely overlap. The spectra were recorded at a ^{1}H frequency of 600MHz and all proteins are ^{15}N labeled.

As for US-Δ3 and US-Δ4, the peaks that belong to SH3 in ^{1}H-^{15}N-HSQC spectrum also shows a nice overlapping with US-WT, which shows the integrity of the remaining domain SH3 in spite of the removal of half or the full UIM domain. Additionally, the US-Δ4 spectrum shows a significant shift in several peaks such as E179, L180, Q183, E184, Q187, L195 (blue label in figure 3.3). Those peaks belong to the remaining of UIM domain in the C-terminus that indicates a possible change in conformation of this helical motif. For this reason, a backbone assignment of US-Δ4 has been carried out, with a focus on the N-terminus region. With that assignment, those shifted peaks were identified to facilitate further NMR experiments.

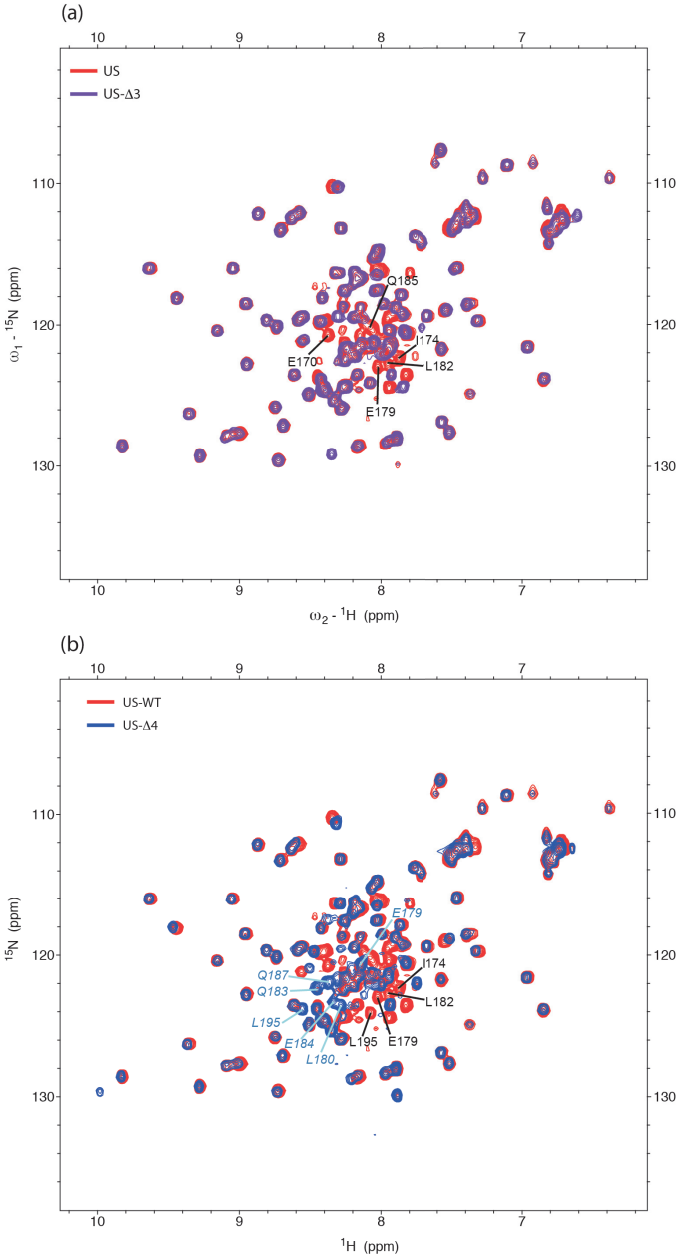


Figure 3.3 – Overlay of ^{1}H-^{15}N-HSQC spectra of US-WT (red) and US-Δ3 (purple) (a) and US-wt and US-Δ4 (dark blue) (b). For residues that belong to SH3 domain and linker, correlation peaks nicely overlap in both mutants. The cross peaks of the remaining UIM are shifted from US-WT (black label) to a new position in the spectrum (blue label) which indicates a change in conformation of the helix. The spectra were recorded at a ^{1}H frequency of 600MHz and all proteins are ^{15}N labeled.

Moreover, the unwinding of the remaining UIM helix is likely responsible for the conformational change of US-Δ4. Therefore, a series of circular dichroism experiments were performed on the US-WT and derived mutants to get information on the helicity (Figure 3.4).

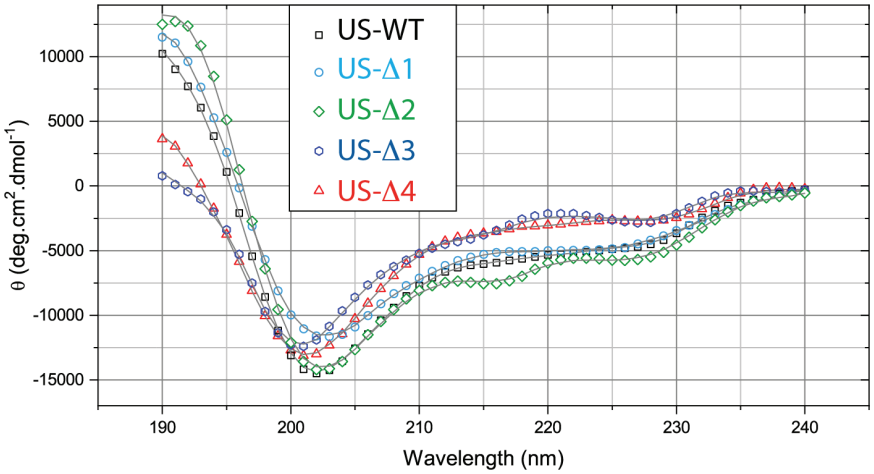


Figure 3.4 – Circular dichroism spectra obtained for US-WT (black), US-Δ1 (light blue), US-Δ2 (green), US-Δ3 (dark blue) and US-Δ4 (red). Experimental points are represented by open symbols while the fitted points are represented by gray lines.

Taking into account an error of 2 % by using the CDSSTR algorithm [83], the helicity percentage of US-Δ1 does not change comparing to US-WT while it shows a slight increase in US-Δ2 (Table 3.1). Moreover, according to the homology analysis described in section 1. of chapter 2, the α-helix of UIM has a length of 18 amino acids, which falls nicely in the range of US-WT (17.4), US-Δ1 (16.3) and US-Δ2 (19.4) average helix length. These results confirm the fact that UIM and SH3 domains keep their initial fold and structure despite a shortening of the linker. For the two mutants with alteration in the UIM motif, US-Δ4 shows about the same helicity percentage than US-Δ3 while having a longer sequence, which corresponds to a greater average number of helix residues ($\approx$ 5.7 for US-Δ4 over $\approx$ 3.4 for US-Δ3).

US-Δ4 was also purified with a 13C and 15N double labeling for a NMR backbone assignment experiment, which allowed us to have information on the chemical shifts of the backbone carbon, nitrogen and hydrogen. The backbone torsion angles has been analyzed and the prediction of the secondary structure of the protein has been done using the TALOS+ tools [87]. The results have been taken into account to reconstruct a schema of US-Δ4

Construct	Helicity	Number of amino acids	Average helix length
US-WT	16%	109	17.4
US-Δ1	16%	102	16.3
US-Δ2	20%	97	19.4
US-Δ3	4%	85	3.4
US-Δ4	6%	95	5.7

Table 3.1 – The percentage of helicity has been obtained by means of the Dichroweb server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) and the CDSSTR method [83] [84] [85] using the reference set SP175 [86]. The average helix length is calculated using the number of amino acids and the percentage of helicity.

secondary structure in 3.5. This prediction supports the previous CD results by implying a quasi-complete unwinding of the remaining UIM helix for US-Δ4. Furthermore, any transient interaction between the C-terminus half of UIM and SH3 is also excluded.

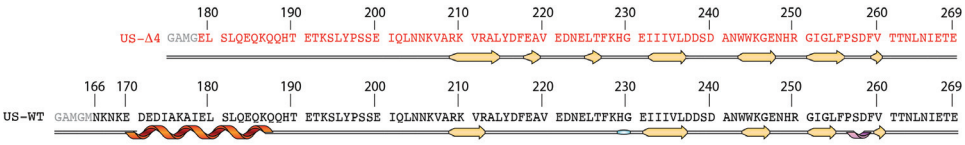


Figure 3.5 – US-WT secondary structure and the reconstruction of US-Δ4 secondary structure by using TALOS+ [87] prediction.

1. SAXS to study the global conformation and flexibility

Small Angle X-ray Scattering (SAXS) is the chosen method to study the global flexibility and spatial organization of the US-WT, US-Δ1, US-Δ3 and US-Δ4. Data have been recorded at four different concentrations (from 1 to 10 mg/ml) to ensure good quality data at both low and high angles. At first sight, the subtracted and merged scattering curves (if that was the case) are smooth lines (Figure 3.6). The bumpy form of the scattering curve, which is a signature of multidomain protein with a fixed distance between the domains or in other word, a rigid linker, was not observed in the US-WT and other mutants scattering data. This result first indicates that the constructs of interest exhibit interdomain motions, which is sign of a dynamic and flexible linker [20].

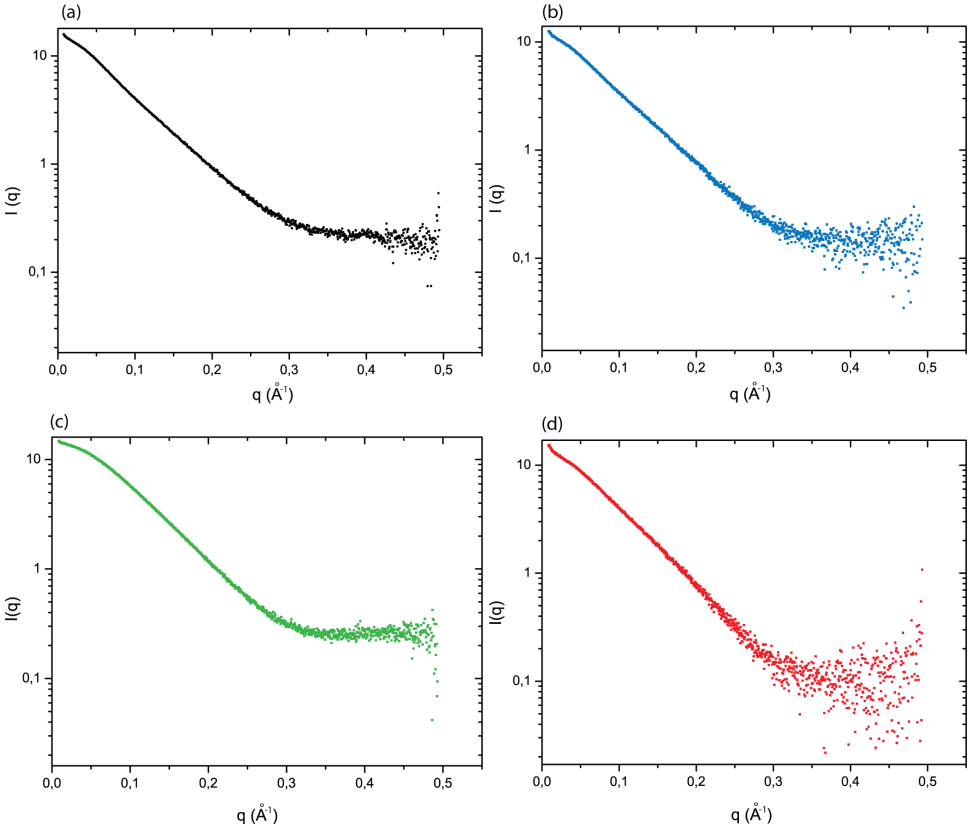


Figure 3.6 – X-ray scattering curves of US constructs after buffer subtraction and data merging: US-WT in black (a), US-Δ1 in light blue (b), US-Δ 3 in green (c) and US-Δ4 in red (d) dotted lines.

1.1. Extraction of the basic parameters and the flexibilty level

After initial processing of the raw data directly at the beamline (section 4. in Methods), the experimental curves were imported into the SCÅTTER software [79]. Few data points ($\approx$ 20) at the beginning of the experimental curve were cut to reduce non-avoidable artifact at the lowest angle. Further analysis to extract parameters such as the radius of gyration R_g, the average volume and the maximum dimension D_{max} were also performed.

A Guinier analysis was first performed where the $logI(q)$ was plotted against q^2 and the region of linearity was defined in order to extract R_g as well as the extrapolated intensity at zero scattering angle $I(0)$ [88]. The linearity of this Guinier plot at the low angle is

Construct	R_g (Å)	D_{max} (Å)	Volume (Å^3)
US-WT	22.13±0.26	95	21491
US-Δ1	21.66±0.90	93	20919
US-Δ3	17.48±0.09	70	17506
US-Δ4	21.04±0.23	84	21304

Table 3.2 – Overall parameters derived from SAXS data analysis on the SCÅTTER software [79]. The radius of gyration (R_g) was derived from a Guinier analysis while D_{max} was extracted from the distance distribution function P(r). The average volume of the different samples was derived from the analysis of the Porod-Debye plateau.

also an indication of a good quality data where there is no aggregation of the sample nor concentration dependence issue. Overall, the constructs have approximately the same R_g except US-Δ3 with the shortest R_g (see Table 3.2).

In addition to R_g and I(0), D_{max} was also determined by analyzing the P(r) function. To obtain a smooth P(r) curve with the tail reaching zero asymptotically while avoiding any negative value, D_{max} was adjusted and therefore was defined as the largest dimension of the molecule. For all constructs, the P(r) curves appear as asymmetric shape with a maximum around 17.5 Å (see Figure 3.7). This peak at low r value corresponds to intra-domain 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 corresponds to the presence of the second domain UIM, that is a small helix (for US-WT, US-Δ1 and US-Δ4) but also reflects the possibility for the different constructs to adopt various spatial conformations due to their flexible linker. US-WT, US-Δ1 and US-Δ4 have a similar pattern and differ only by the extent of their curve at high r values. US-Δ3 adopts a more symmetric curve, therefore reflects a more compact structural organization in solution.

To summarize at this point, a shortening of the linker region by seven amino acids does not affect D_{max} and R_g significantly as the maximum diameter of US-WT stays fairly close to US-Δ1 and the volume occupied by US-Δ1 and US-WT is relatively comparable. The US-WT's linker therefore does not behave as a fully elongated tether but rather shows turns and loops. Moreover, the deletion of the US-WT N-terminus by 14 amino acids (US-Δ4) induces the

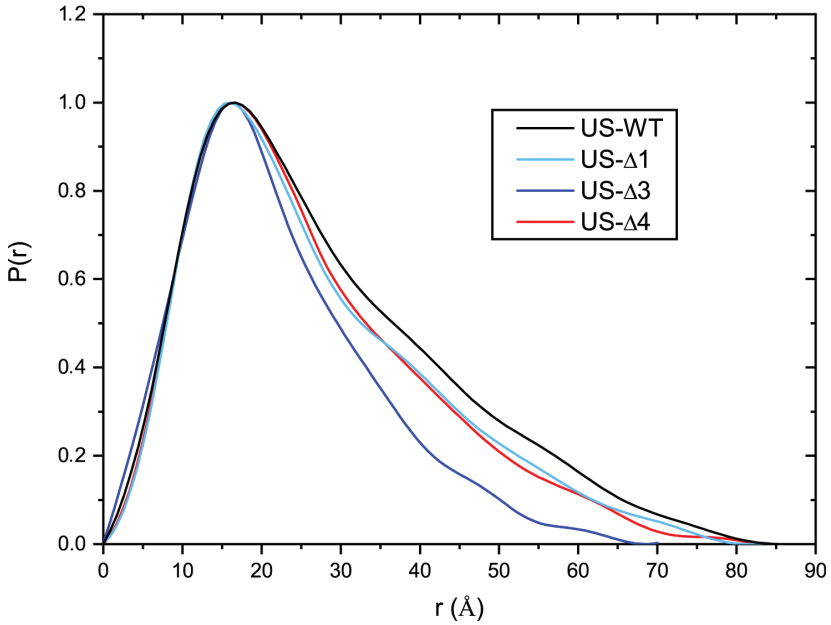


Figure 3.7 – Overlay of the normalized distance distributions P(r) for US-WT (black), US-Δ1 (light blue), US-Δ3 (dark blue) and US-Δ4 (red).

partial truncation of UIM but only a minor change of R_g was observed while D_{max} drops by 11Å compared to US-WT. On the other hand, the complete deletion of the UIM domain (US-Δ3) leads to a drastic decrease in R_g and D_{max} of 4.6 and 25 Å, respectively. Altogether, the results suggest that US-Δ3 has a reduced conformational space, hence a more compact organization and support the idea that the UIM domain acts as a spring that can maintain the SH3 domain at a given distance.

1.2. The global conformation

In the next step, Kratky plot was created to assess the flexibility of the constructs of interest. The Kratky analysis plots $q^2I(q)$ against q, highlight the differences between a well folded, compact protein with a pronounced curve maximum and a random, unstructured chain that exhibits a plateau at the same q value [89]. However, it is difficult to compare between objects of different size or different multidomain proteins with this Kratky plot. Therefore, a dimensionless Kratky was plotted (Figure 3.8) where I(q) is normalized to the

forward scattering intensity I(0) and q is normalized to R_g [90]. Ubiquitin Kratky plot was represented as a reference of a globular protein profile, which reaches a maximum of ≈1.104 at a qR_g value equals to ≈1.73 and a shape that rises and falls symmetrically [90] [91]. 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. This behavior confirms that each US construct contains at the same time well-folded and intrinsically-disordered region. Moreover, US-Δ3 shows a curve maximum at a value of qR_g closer to 1.73, which reflects a lower composition of unstructured flexible regions compared to the other three constructs (Figure 3.8). According to this normalized plot, it is also noticeable that the curve representing US-WT falls with a less pronounced slope at higher qR_g values. From this observation, we can state that the US-WT possesses the highest flexibility among other variants and that the shortening of the linker has removed part of the contribution of the flexible disordered region.

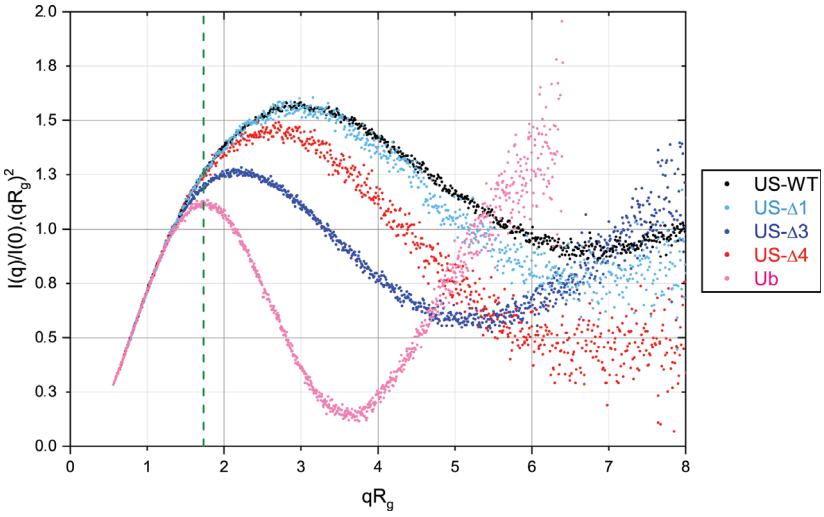


Figure 3.8 – Normalized Kratky plots of the four constructs were presented together with Ubiquitin as the reference. The green vertical dotted line represents a qRg value of $\sqrt{3}$, which crosses the typical peak maximum obtained for globular proteins.

To find the conformation adopted by the different constructs, we have used the ten

models structures generated by Modeller as starting models. For each model, a theoretical scattering curve was calculated with the FoXS server [80] [92] and compared with the experimental curve. The quality of the theoretical curve was assessed by a χ score value where a low χ score indicates a theoretical curve close to the experimental curve. For each construct, it is clear that none of the ten models is able to reproduce the experimental curve (See Table 3.4). This result suggests that the four constructs of interest US-WT and the 3 mutants US-Δ1, US-Δ3 and US-Δ4 should exist with various conformations in solution instead of a single state. To correctly fit the experimental SAXS profile, an ensemble of conformations with weighted composition must be considered as theoretical profile. Therefore, the MultiFoXs server was used in the next step, which provides a multi-state modeling. Before all analysis on MultiFoXS, flexible residues were defined, which correspond to the full-length or cut linker, the N-terminus and the C-terminus of the constructs (See Table 3.3) while the residues that belong to the SH3 and/or UIM domain are considered as rigid bodies. With the PDB model given together with the other input, the tool first computates 10000 conformations by the RRT algorithm [80] then back-calculates their scattering curves. Afterward, the number of samples is decreased to 1000 top χ scores and different multi-state ensembles are generated with weighted composition.

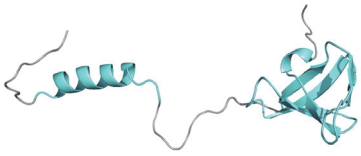
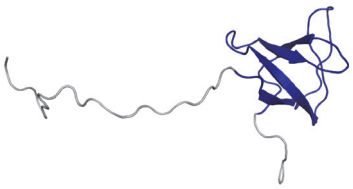
Construct	Flexible residues	Structure
US-WT	161-170 189-207 262-269	
US-Δ1	161-170 189-193 201-207 262-269	
US-Δ3	185-207 262-269	
US-Δ4	175-184 190-207 262-269	

Table 3.3 – Range of the flexible residues used for the computation of the N-state model ensemble in MultiFoXS. The structure with the best fit scoring while using FoXS was used as initial template for the MultiFoXS analysis and are represented in the far right column. The corresponding regions have been colored in grey in the different structures: US-WT(black), US-Δ1(blue), US-Δ3(dark blue) and US-Δ4(red).

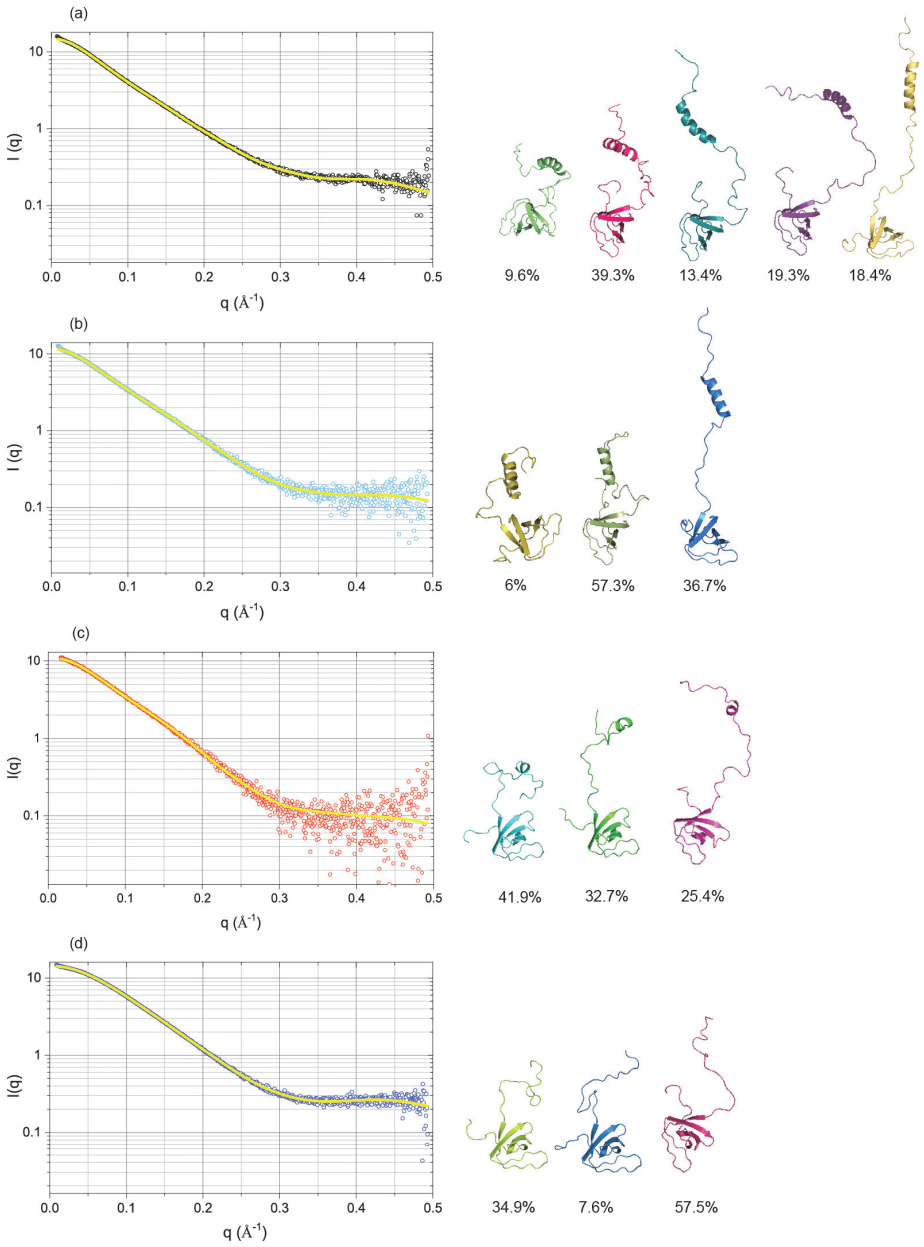


Figure 3.9 – The SAXS experimental data (dots) for US-WT (a), US-Δ1 (b), US-Δ4 (c) and US-Δ3 (d) and the back-calculated curve (continuous yellow line) from the best fit ensemble of structures are represented in the left panel. The cartoon representation of the structures and their respective weight are presented in the right panel.

For US-WT, the χ score in MultiFoXS analysis drops significantly comparing to the best

model with FoXS and improve for 50% from the one-state to the five-states ensemble (1.62 to 0.79) (see Table 3.4). The result supports the fact that US-WT is highly flexible and dynamic where different conformations co-exist in solution, from an elongated to a compact state (See Figure 3.9b).

For all other constructs, only three states are sufficient to describe the SAXS curves while increasing the number of state does not give rise to a significant decrease of the χ score. US-Δ1, US-Δ3 and US-Δ4 constructs shows a decrease of 28 %, 33 % and 39 % in χ score from the one-state to the three-states ensemble, respectively (see Table 3.4). There are accordingly less conformations needed to correctly fit the experimental data, which indicates a decrease in the flexibility of those mutants compared to US-WT.

	US-WT	US-Δ1	US-Δ3	US-Δ4
1 state	1.62	1.39	1.29	1.73
2 states	0.94	1.05	0.91	1.25
3 states	0.90	1.00	0.86	1.06
4 states	0.84	0.98	0.85	1.08
5 states	0.79	0.97	0.85	/
Modeller	101.00	25.57	55.60	4.17

Table 3.4 – Summary of the lowest χ scores for each of the different constructs and for the different multi-state models used in MultiFoXS after the computation of 1000 conformations and starting from 10000 initial conformations. For comparison, the lowest χ score calculated by FoXS for the 10 models generated by Modeller is also shown.

A further analysis of the R_g computed for the 1000 best-scoring multi-state was done. For all five models (one- to five-states) of US-WT, the R_g distribution of the initial pool of 10000 conformations spans on a large amplitude from 16 to 35 Å, with a highly populated R_g found around 20 Å, while four other values of R_g scatter toward the higher R_g (Figure 3.10). In the five-state model where the best improvement of the χ score is observed, other than the most populated R_g of $\approx$ 20 Å, three values of R_g are approximately homogeneously populated at 33.8, 26.8 and 24.8 Å while only a lowly populated value of R_g appears at 15.8 Å. The results indicate the capacity of US-WT to adopt a broad range of conformation, which makes it highly flexible in space.

For the US-Δ1 construct, the χ score of the one-state significantly improves compared to

the best scored model issued from Modeller (from 25.57 to 1.39). Compared to the one-state model, the SAXS data are nicely fitted using a three-state models and a four or five-state model did not improve the quality of the calculated SAXS data (see Table 3.4). In all US-Δ1 models, two populations of R_g are found. First is the most populated R_g around 20 Å, which is similar to US-WT. Second is a population of R_g that is located in a narrow region between 28 and 30 Å (for two- and three-states model). This result supports the fact that US-Δ1 adopts a more compact structure compared to US-WT (see Figure 3.10 b).

In the same way for US-Δ3 and Δ4, the χ score of the best fit model issued from Modeller are significantly higher compared to the one-state model (55.60 compared to 1.29 and 4.17 compared to 1.73 respectively). Similar to US-Δ1, an optimum improvement is observed with a three-state model. Indeed the χ score has improved by 33 and 39 % compared to the one-state model for US-Δ3 and Δ4 respectively (see Table 3.4) and no significant improvement is found with the four- or five-states model. The R_g distribution of US-Δ4 spans on a region of comparable length to that of US-Δ1, showing a consistency with previous observation on the similar experimental R_g of US-Δ1 and US-Δ4 despite their different sequences. On a closer look at US-Δ3, the R_g values extend between 13 and 21 Å, which shows a much narrower region compared to the other three constructs. Two main R_g populations are found at 15 and 18 Å and a lower populated R_g is located at 20.2 Å indicating a compact conformational arrangement. Indeed, the result is consistent with the experimental R_g derived from Guinier analysis.

In brief, the experimental R_g represents an average value of all possible conformations obtained by the protein in solution and is not fully informative regarding the protein flexibility. Indeed, the multi-state model analysis shows that the conformational space sampled by the different US constructs could be described by a combination of different R_g. While the comparison of the experimental R_g did not show any dramatic difference between US-WT, US-Δ1 and US-Δ4, the multi-state model reveals a deviation of the degree of flexibility of the different mutants compared to US-WT.

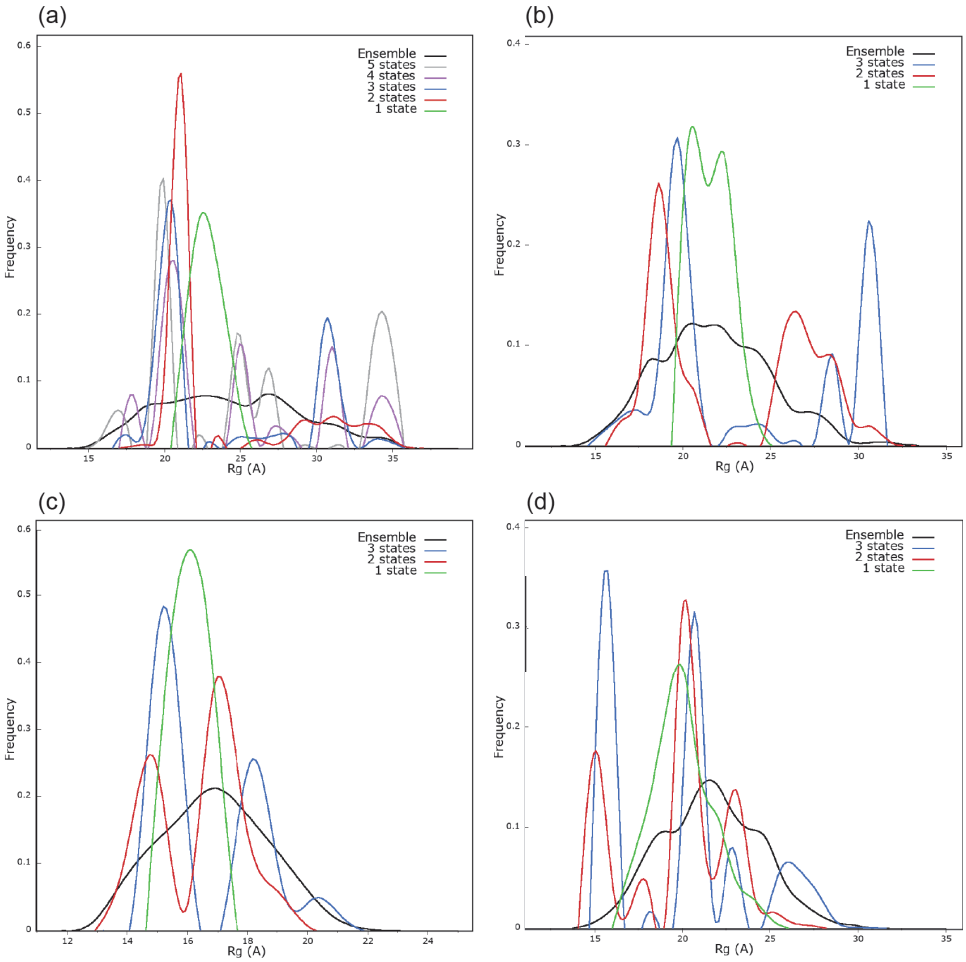


Figure 3.10 – R_g distribution for the entire ensemble of 10000 conformations (black lines) and the 1000 best-scoring N-state models (N=1 ... 5) calculated by MultiFoXS (colored lines). Up to five-state model is shown for US-WT (a) and only the three-state models are shown for US-Δ1 (b), US-Δ3 (c) and US-Δ4 (d) since increasing the number of states did not improve the χ score reported by MultiFoXS (see Table 3.4)

2. NMR spin relaxation and the dynamical characterization

By using SAXS data, the global flexibility of the constructs has been studied where the information on the shape/combination of shapes has been described. In the coming step, 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 constructs. For this study, an additional mutant US-Δ2 is introduced with 14 amino acids truncated in the linker region between UIM and SH3 domain (see Figure 2.1 in chapter 1). The measurement of the three commonly used spin relaxation parameters R_1, R_2 and the ^{15}N-^{1}H cross-relaxation rates, via the steady-state ^{15}N-^{1}H-NOEs were performed at 600 MHz.

		$\mathbf{R_1\ (s^{-1})}$	$\mathbf{R_2\ (s^{-1})}$	**NOE**
US-WT	UIM	1.65 ± 0.02	5.7 ± 0.1	0.29 ± 0.01
	SH3	1.47 ± 0.02	10.2 ± 0.2	0.74 ± 0.04
US-Δ1	UIM	1.65 ± 0.04	6.7 ± 0.2	0.32 ± 0.02
	SH3	1.45 ± 0.09	10.9 ± 0.5	0.73 ± 0.04
US-Δ2	UIM	1.60 ± 0.04	6.3 ± 0.1	0.33 ± 0.02
	SH3	1.48 ± 0.03	9.9 ± 0.1	0.69 ± 0.03
US-Δ3		1.66 ± 0.04	8.7 ± 0.4	0.76 ± 0.04
US-Δ4		1.60 ± 0.04	9.0 ± 0.2	0.73 ± 0.03

Table 3.5 – Averaged longitudinal (R_1) and transverse (R_2) relaxation rates as well as averaged steady-state ^{15}N-^{1}H-NOE values measured for the different constructs of UIM-SH3 at 15 °C and at a ^{1}H frequency of 600 MHz.

From the experimental R_1, R_2 and NOE values, a set of diffusion tensors D_{xx}, D_{yy} and D_{zz} was also deducted (see Table 3.6) in order to calculate the correlation time τ_C using the RotDif software developed by Walker *et al.* [75] with different models. For instance a fully anisotropic model was used for SH3 where 3 distinct diffusion tensors were applied. For the UIM helix, an axial model was more utilized, where the tensors $D_{parallel} = D_{xx} = D_{yy}$ and $D_{perpendicular} = D_{zz}$.

		$\mathbf{D}_{xx}$ $^{(a)}$	$\mathbf{D}_{yy}$ $^{(a)}$	$\mathbf{D}_{zz}$ $^{(a)}$	τ_C $^{(b)}$
US-WT	UIM	3.28 ± 0.49	3.28 ± 0.49	4.90 ± 0.16	4.40 ± 0.09
	SH3	1.78 ± 0.03	2.07 ± 0.04	2.47 ± 0.05	7.91 ± 0.04
US-Δ1	UIM	4.58 ± 0.26	4.58 ± 0.26	1.53 ± 0.80	4.68 ± 0.18
	SH3	1.60 ± 0.1	2.04 ± 0.09	2.57 ± 0.19	8.04 ± 0.18
US-Δ2	UIM	6.13 ± 0.59	6.13 ± 0.59	0.63 ± 0.59	3.88 ± 0.31
	SH3	1.89 ± 0.06	2.25 ± 0.05	2.32 ± 0.09	7.74 ± 0.05
US-Δ3	SH3	2.28 ± 0.07	2.54 ± 0.06	2.75 ± 0.07	6.61 ± 0.05
US-Δ4	SH3	2.03 ± 0.05	2.52 ± 0.04	2.65 ± 0.06	6.94 ± 0.05

Table 3.6 – Summary of the diffusion tensors calculated using the ROTDIF program [75] [76].
Axial and fully anisotropic models were used for UIM and SH3 domain, respectively
(a) The principal values of the diffusion tensor, in $10^7 s^{-1}$.
(b) The overall rotational correlation time, $\tau_C = 1/(6\, D_{iso})$, in nanoseconds,
with $D_{iso} = (D_{xx} + D_{yy} + D_{zz})/3$

For the constructs US-WT, US-Δ1 and US-Δ2 that contain both domain UIM and SH3, the R_1 profiles is flat and the average values of UIM and SH3 are close to each other, which are around 1.65 s^{-1} and 1.45 s^{-1} respectively. On the other hand, a significant difference of the average R_2 was observed between the two domains: around 6 s^{-1} for UIM and 10 s^{-1} for SH3 in all three constructs (see Figure 3.11 and Table 3.5).

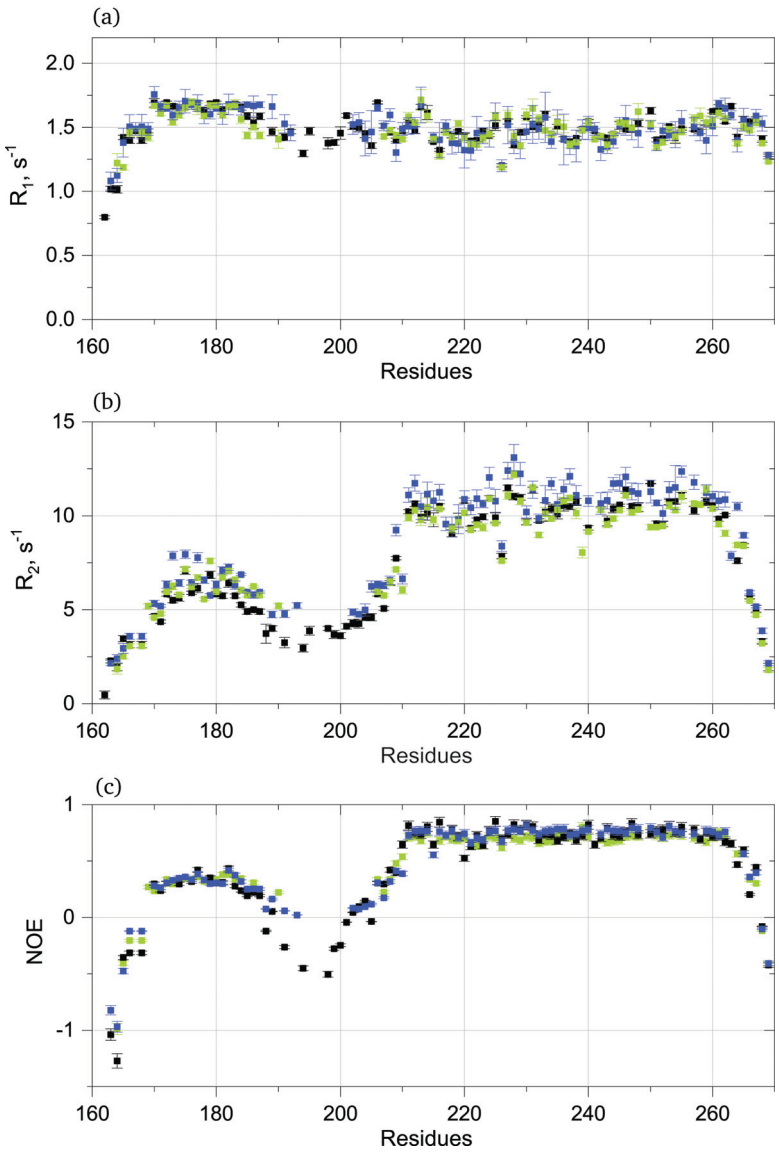


Figure 3.11 – Relaxation parameters R_1, R_2 and ^{15}N-^1H-heteronuclear NOE for the three US constructs: US-WT (black), US-Δ1 (blue) and US-Δ2 (green) obtained at a ^1H frequency of 600 MHz and at 15 °C.

The same distinction was found in the NOE level: UIM domain exhibits an average NOE value of 0.3 while the SH3 domain displays an average value of 0.72 for all constructs. The molecular mass dependance of R_2 can be used as a ”molecular mass ruler”. The calibration

curve (Figure 3.12) was obtained by considering the experimental R_2 values of UIM, Ub and K63-Ub_2 and by using the Stokes-Einstein equation [93] [47]. This curve has been obtained by assuming an isotropic overall tumbling of the protein, a squared order parameter S^2 of 0.87 and a local correlation time of 50 ps. The anisotropy on R_2 is also taken into account after averaging the R_2 values of an ensemble of 50 randomly distributed vectors with a S^2 of 0.87, a local correlation time of 50 ps and an anisotropy of 0.2 and 2.0 to plot the anisotropic curves (see dashed line on Figure 3.12). A protein specific density of 0.76 cm^3/g, a hydration shell of 3.2 Å, and a solvent viscosity of 0.00114 Pa.s at 288 K were assumed. As can be seen on the calibration curve, Ub exhibits an average R_2 of 8 ± 0.6 s^{-1} for a molecular of $\approx$ 8.0 kDa. With this in hand, R_2 value of the UIM domain and SH3 domain are plotted against their respective molecular weight of 4.0 kDa and 7.4 kDa. R_2 values of UIM and SH3 domain in US-WT, US-Δ1 and US-Δ2 both deviate considerably from the expected value of the individual domains: around 10 s^{-1} for all constructs in stead of 7.5 s^{-1}. In fact, compared to the values expected for these domains taken individually, the deviation of R_2 values falls approximately into the 36 % correlation between domains as described by Bae *et al.* [94].

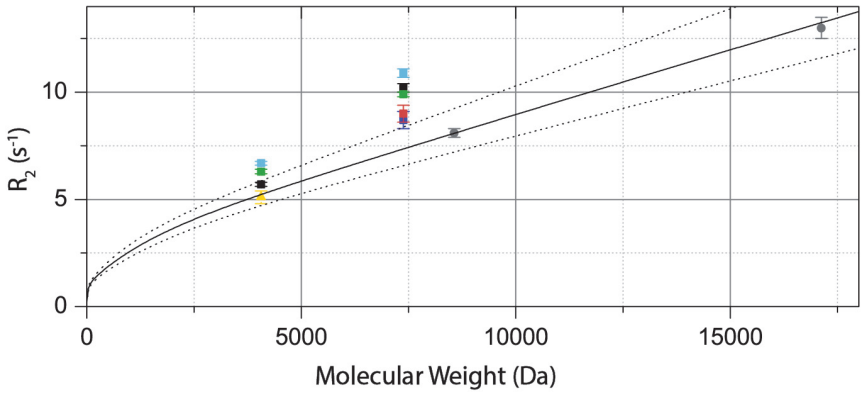


Figure 3.12 – ^{15}N transverse relaxation rate, R_2, as a function of molecular mass. The respective R_2 derived for the SH3 or the UIM domain involved in the different mutants are colored black (US-WT), blue (US-Δ1), green (US-Δ2), dark blue (US-Δ3) and red (US-Δ4). The calibration plain curve was obtained by considering the experimental R_2 values of Ub and K63-Ub_2 distal Ubiquitin unit (grey). The dotted black line represents the effect of the anisotropy on R_2 after averaging the R_2 values of an ensemble of 50 randomly distributed vectors with a S^2 of 0.87, a local correlation time of 50 ps and an anisotropy of 0.2 (lower curve) and 2.0 (upper curve). The UIM R_2 (yellow) value has been taken from a previous study [48].

Furthermore, as proven in the previous section, UIM and SH3 do not have any interdomain interaction. The observation supports the idea that in solution, the two domains UIM and SH3 of the same molecule tumble differently with distinct R_2 values and hence correlation time, but in a dependent manner. This is likely due to the presence of the long linker between them. Moreover, it is likely that UIM and SH3 do not share a common rotational diffusion tensor. The interdomain motion drastically affects the global shape of the protein and hence its rotational diffusion, anisotropy and global tumbling. When analyzing the correlation time τ_C, the two domains UIM and SH3 show approximately 3.5 ns of difference for US-WT and US-Δ1 while US-Δ2 with the shortest linker length shows a slightly higher gap (3.8 ns). Similarly, the gap between the average R_2 for the UIM and SH3 domains significantly decreases when the length of the linker is shortened. Those R_2 gap are 4.5, 4.2 and 3.6 for US-WT, US-Δ1 and US-Δ2 respectively (see Table 3.5). Therefore, linker shortening decreases significantly the interdomain motion and its influence on the global tumbling of the dual domain. Additionally, by zooming in the C-terminus region (188-195) of UIM in the two mutants US-Δ1 and Δ2, the NOE display a higher value compared to US-WT, which indicates an increase in the stiffness of the linker at UIM side, thus a decrease of the interdomain motion.

For US-Δ3 and Δ4, the average R_1 level of the SH3 domain increases while the average R_2 level as well as the correlation time significantly decreases. Indeed, UIM is partially or completely deleted for those two constructs. The interdomain motion is therefore abolished and the average R_2 values decreases to the expected values for SH3 domain alone. (See Figure 3.13 and Table 3.5). As for the rest of the molecule, the long flexible linker displays much lower R_2 and NOE values. For instance, the NOE values go down to -3.6 for the US-Δ3 construct, compared to the same residue range in US-WT, Δ1 and Δ2. This result indicates a much faster local motion in the linker region with the absence of half or the full UIM domain, therefore a much higher flexibility.

As previously shown in the Methods chapter, 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 $J(0.87\omega_H)$ [74]. These analytical expressions have been already used to study dynamics of multidomain proteins containing disordered segments [95].

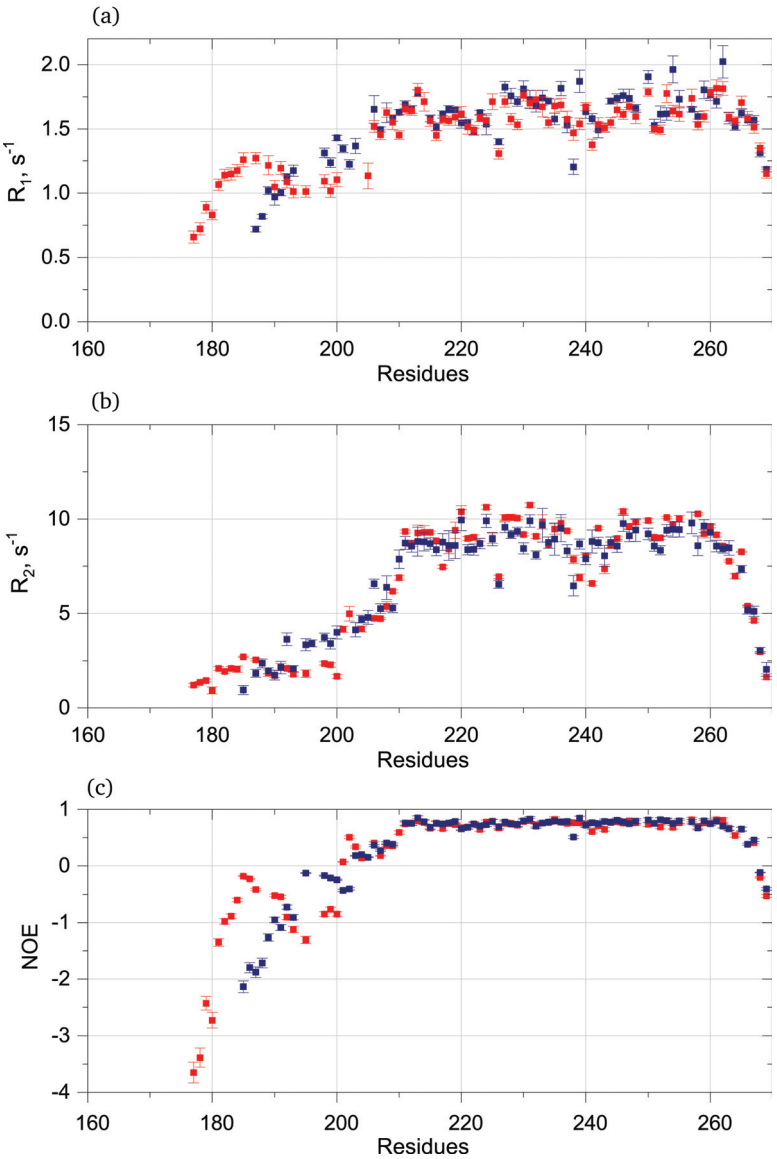


Figure 3.13 – Relaxation parameters R_1, R_2 and ^{15}N-^{1}H-heteronuclear NOE for the two constructs: US-Δ3 (dark blue) and US-Δ4 (red) obtained at a ^{1}H frequency of 600 MHz at 15 °C.

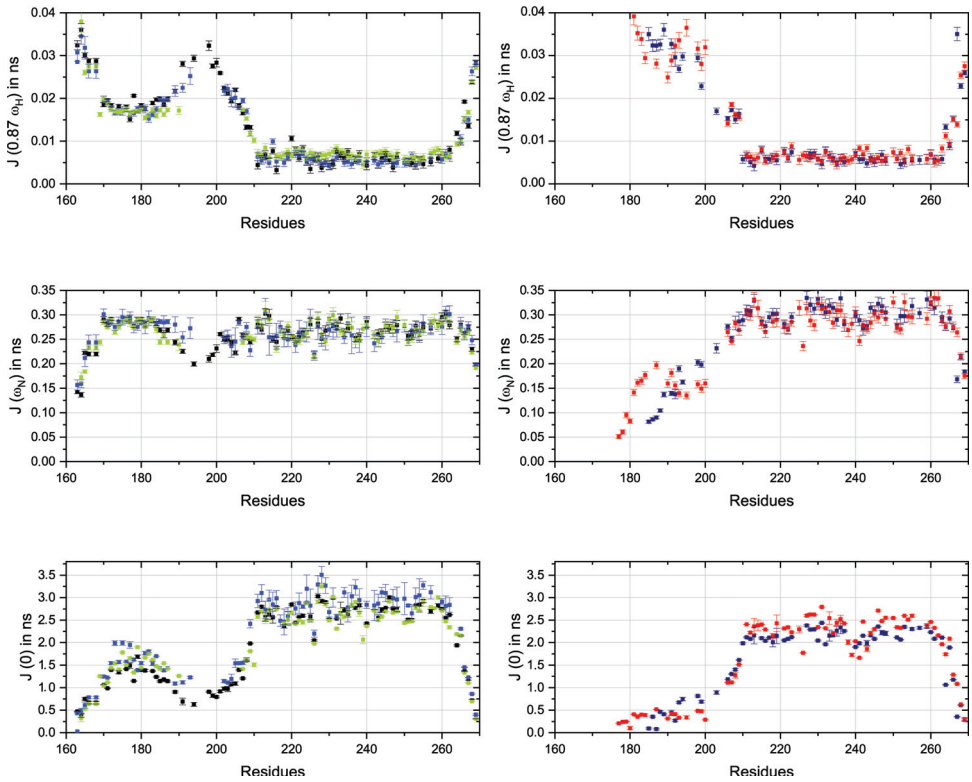


Figure 3.14 – Reduced spectral densities for the US variants US-WT (black), US-Δ1 (blue), US-Δ2 (green), US-Δ3 (dark blue) and US-Δ4 (red) using the equations derived in chapter 2. We have assumed a CSA level of -160ppm and a NH distance of 1.02 Å.

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 Figure 3.14, US-Δ3 and Δ4 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 Figure 3.14), and hence a much higher flexibility of their N-terminus part compared to US-WT, Δ1 and Δ2. As observed above, US-Δ1 and Δ2 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, reflect a decrease in flexibility for the corresponding constructs.

3. Chemical shift perturbation and the visualisation of molecular recognition

The function of a protein is reflected by various parameters such as the catalytic activity (for enzymes), the solubility, the hydrophobicity hence the cellular localization etc. In a previous study, it has been shown that US-WT has the capability to bind the deubiquitinating enzymes UBPY and AMSH [48] as well as Ub and K63-Ub_2 [49]. Thus in the present study, the capacity of binding to its molecular partner was considered as its ability to function properly. The Ub and K63-Ub_2 binding capacity is monitored through chemical shift perturbation experiments (CSP). ^{1}H-^{15}N-HSQC spectra were recorded on ^{15}N-labeled US constructs upon addition of unlabeled Ub or K63-Ub_2 until saturation.

3.1. Interaction with ubiquitin (Ub)

As illustrated previously, there is no interaction between UIM and SH3 domains on US-WT, US-Δ1 and Δ2 constructs, therefore any perturbations are resulting solely from the interaction between the constructs and Ub or K63-Ub_2. Several residues on UIM and SH3 exhibit significant shifts and signal broadening (Figure 3.15). For instance, by taking a closer look at the region shown in the magnified area of Figure 3.15, the residues I174, A175, A177, E179 and L182 signals disappear at low concentration of Ub added, which indicates a fairly tight binding and significant perturbation of this domain. On the other hand, the residues E218, A219 and S257 are an example of residues with an important CSP, where they are shifted for $\approx$ 0.2 ppm in the ^{1}H dimension for E218 and A219 and around 1.5 ppm in the ^{15}N dimension for S257. At first glance, a fairly similar situation is observed in other constructs with several modifications which is discussed in further analysis.

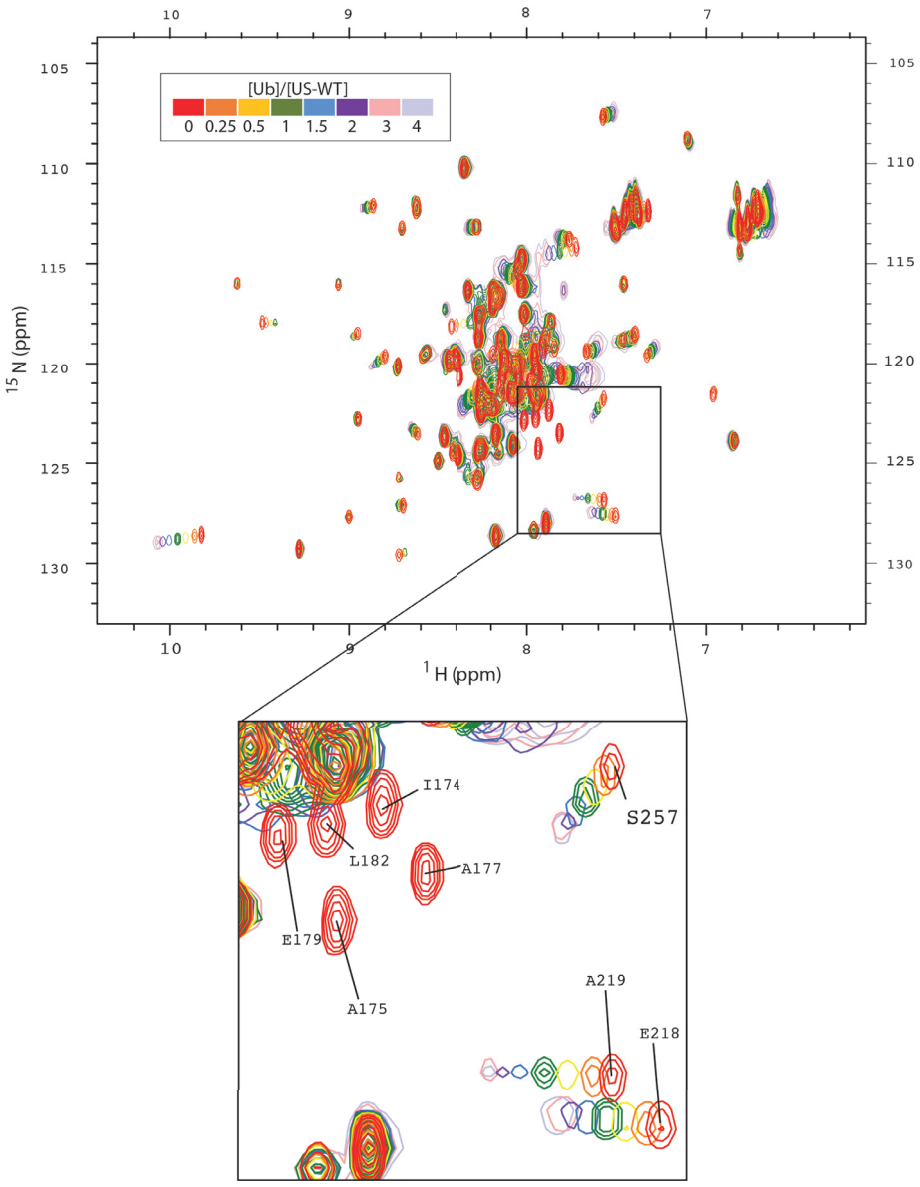


Figure 3.15 – Overlay of the ^{1}H-^{15}N-HSQC spectra of ^{15}N US-WT during titration with ^{14}NUb. The concentration of the added Ub increases according to the [Ub]/[US-WT] ratio ranging from 0.25 to 4. The spectrum of US-WT before any addition of Ub is in red where [US-WT] started at 400 μM and the last point of titration is light blue with [US-WT] = 340 μM. The spectra were recorded at a ^{1}H frequency of 600MHz.

The shift in two dimensions of every resonance was taken into account. The trajectory was calculated based on the formula mentioned in the chapter 2 and plotted to get the

chemical shift perturbations (CSPs) profile (see Figure 3.16 and 3.17). On the UIM domain the CSPs profile remains comparable between the three constructs US-WT, US-Δ1 and Δ2, which indicate the same binding site and interaction mode (see Figure 3.16). On the other hand, several differences are observed in the RT loop and the 3_{10} helix. While the number of CSPs that rises above the chosen threshold of 0.07 ppm remains the same, the overall CSPs intensity decreases with the shortening of the linker. Conversely in the vicinity of the cut linker at the SH3 side the residues K208, V207 and A206 exhibit an increase in CSPs which indicates a stronger perturbation of the surrounding environment.

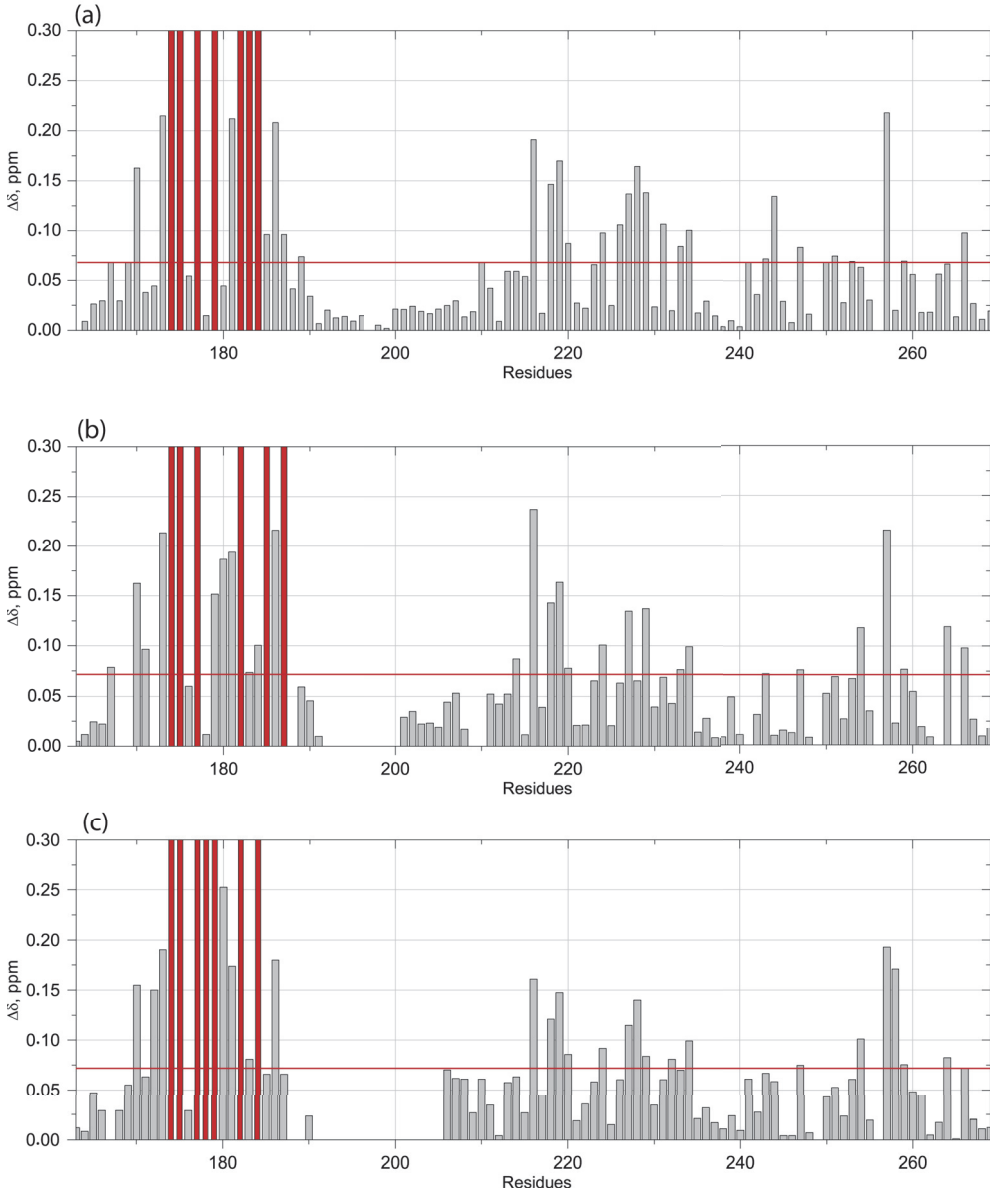


Figure 3.16 – Chemical shift pertubations were mapped on ^{15}N-labeled US-WT (a), US-Δ1 (b) and US-Δ2 (c) upon saturation with ^{14}N-Ub. Red bars indicate residues that disappear during the titration. The red line at 0.07ppm represents the threshold above which the residues are considered for K_d analysis.

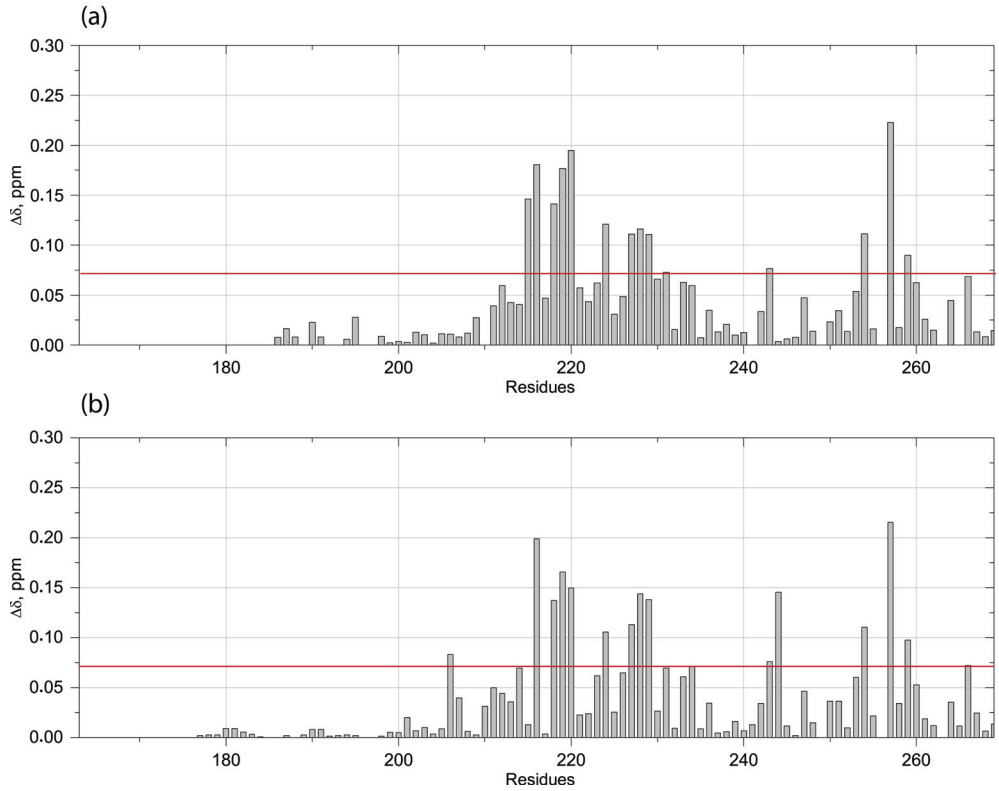


Figure 3.17 – Chemical shift perturbations were mapped on ^{15}N-labeled US-Δ3 (a) and US-Δ4 (b) upon saturation with ^{14}N-Ub. Only those curves that show final CSPs above a threshold of 0.07 ppm were considered for K_d analysis.

In the next step, the titration curves were fitted to the stoechiometry 1:1 model (see section chapter 1) and the dissociation constant K_d was derived for each residue that rises over the 0.07 ppm threshold. At first glance, the titration curves of US-Δ1 and Δ2 reach the saturation at a higher [Ub]/[Substrate] ratio ≈ 4-6 compared to US-WT that start to saturate at around a [Ub]/[Substrate] ratio of 2-3 (see Figure 3.18).

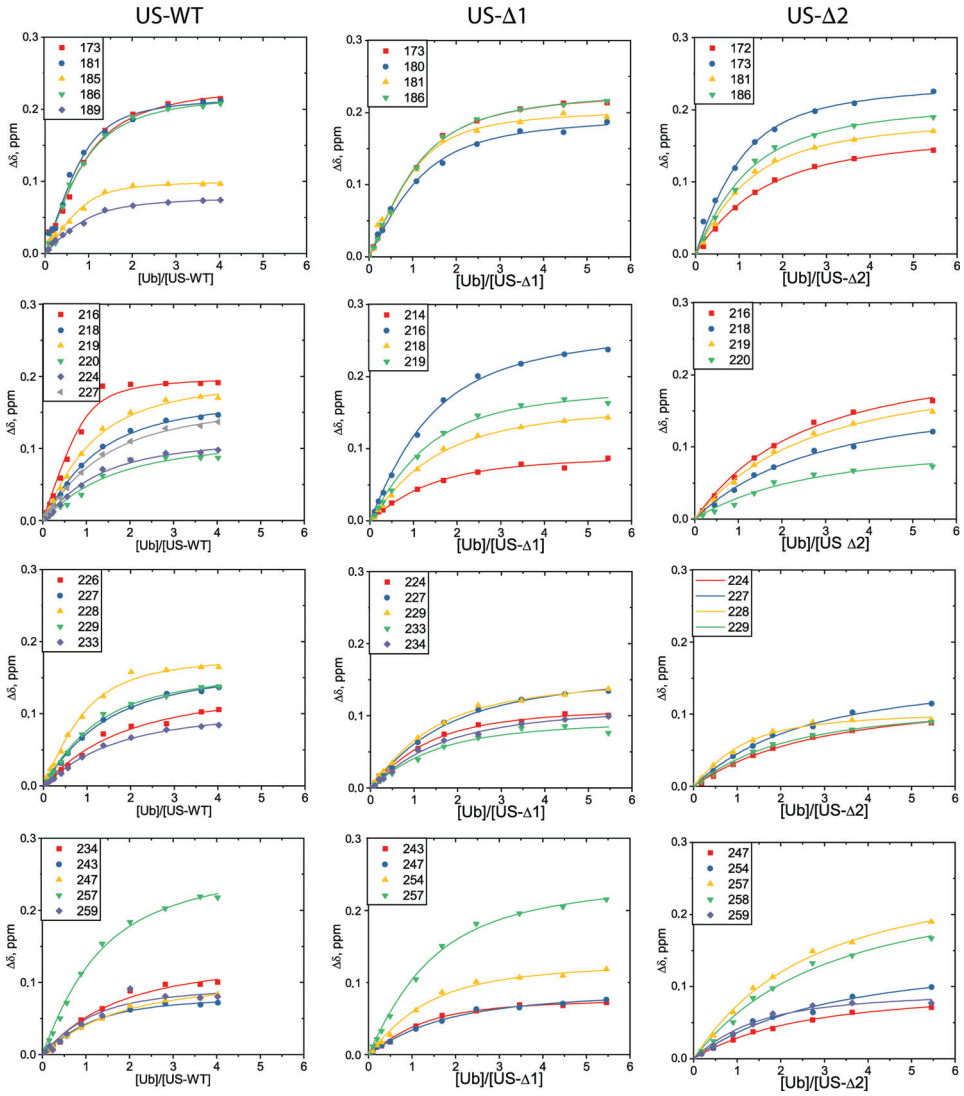


Figure 3.18 – The CSPs monitored during the titration of US-WT (left), US-Δ1 (middle), US-Δ2 (right) by Ub were plotted against the ligand to protein ratio. For each residue presented, the smooth line of the same color depicts the fitted curves obtained by the Matlab program mentioned in the previous chapter, using the 1:1 model

Quantitatively speaking, the dissociation constant of UIM and SH3 domains increases in the same manner for each of the two mutants US-Δ1 and Δ2, which raise for $\approx$ 1.5 and $\approx$ 2.4 times respectively (see Table 3.7). The linker length therefore plays a minor role in the Ub recognition of US-WT as when it is shortened the affinity does not drastically change.

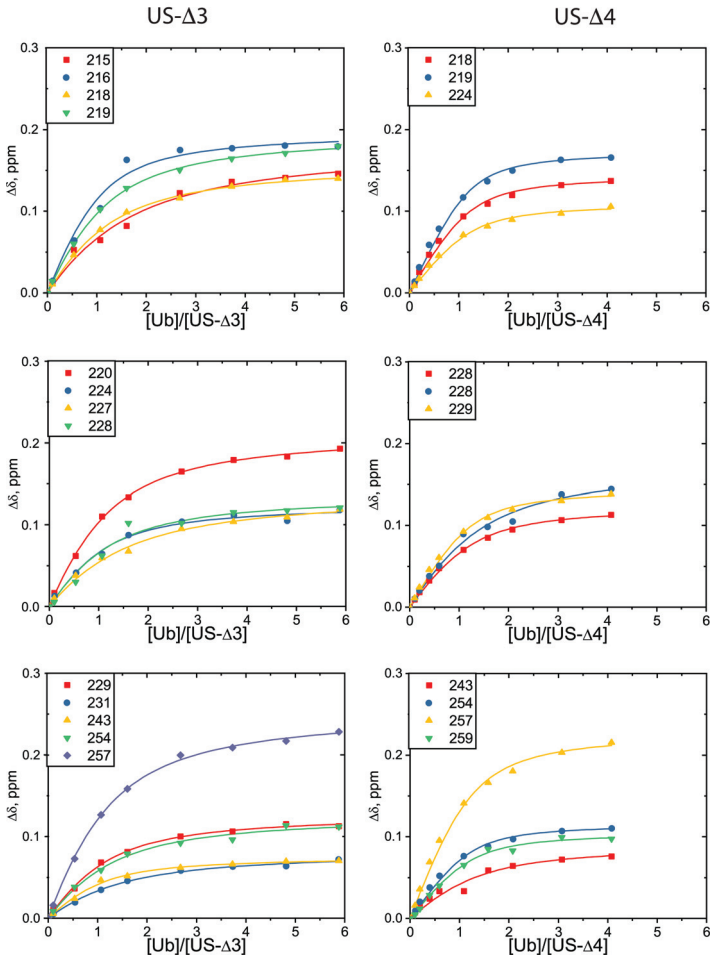


Figure 3.19 – The experimental titration curves of some representative residues of the interaction between Ub and the two constructs US-Δ3 and US-Δ4, together with the fitted curved extracted from the in-house Matlab program (see Chapter 2).

For the two US-Δ3 and Δ4 constructs, the truncation of half or the total of UIM domain completely abolish Ub binding. The two mutants are consequently considered as SH3 domain alone with a long flexible linker at the N-terminus. Indeed, the CSP pattern of the SH3 domain remains unchanged compared to that of the three constructs US-WT, US-Δ1 and Δ2. Likewise, the titration curves of SH3 residues in US-Δ3 and Δ4 show a similar look as seen in the other constructs (Figure 3.19) and the dissociation constant of SH3 in US-Δ3 remains similar to the SH3 domain of US-WT. US-Δ4 even exhibits a higher affinity where the K_d slightly reduces

for $\approx$ 1.7 times (see Table 3.7). Overall, the binding to Ub does not require any constraint in space since the monomer of Ub can easily intercalate for the interaction, as long as there is an intact domain presented.

		K_d (μM) **Ub**	**K_d (μM)** **K63-Ub_2**
US-WT	UIM	86 ± 31	78 ± 59
	SH3	267 ± 119	60 ± 28
US-Δ1	UIM	148 ± 60	137 ± 57
	SH3	394 ± 127	253 ± 163
US-Δ2	UIM	203 ± 91	99 ± 50
	SH3	669 ± 234	483 ± 196
US-Δ3	SH3	266 ± 96	39 ± 22
US-Δ4	SH3	158 ± 79	97 ± 49

Table 3.7 – 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 K63-Ub_2, dissociation constants were extracted for US-WT, US-Δ1 and US-Δ2 using a 2:1 model and for US-Δ3 and US-Δ4 using a 1:1 model (see Chapter Methods). Standard deviations are used as error estimates.

3.2. Interaction with Lysine 63-linked di-ubiquitin (K63-Ub_2)

The STAM2 protein, which plays an essential role in the endosomal sorting process, is known to have a preferential binding activity with K63-linked polyubiquin chains. As seen in the previous section, the modification of the conformation, flexibility and dynamics of the protein has negligible effect on the Ub binding. Thus raising a question if any modification of this US linker would perturb its affinity with a multi-domain binding partner, K63-Ub_2 dual domain protein in this case. The same protocol was followed as for Ub monomer. ^{1}H-^{15}N-HSQC spectra of the different ^{15}N-labeled constructs were recorded upon addition of unlabeled K63-Ub_2 and the CSPs were monitored.

As can be seen in the series of ^{1}H-^{15}N-HSQC spectra, US-WT titrated by K63-Ub_2 shows a similar appearance as the titration by Ub. Due to a strong binding, the residues I174, A175, A177, E179 and L182 resonances of the UIM domain for instance start to disappear at a ratio [Ub_2]/[US-WT] of 0.1 (Figure 3.20, upper panel). A fairly similar aspect is observed in US-Δ1 titration with practically the same resonances moving in the same direction as in US-WT titration by Ub_2 (Figure 3.20, lower panel) with comparable trajectories. However, for the five residues mentioned above, the cross peaks disappearance occurs only at a [Ub_2]/[US-WT] ratio of 0.2, which is slightly higher than that of US-WT. The CSPs patterns of K63-Ub_2 titration remain fairly similar to the one of Ub titration as no additional perturbations were observed. However in both mutants US-Δ1 and US-Δ2, several CSPs are significantly lower such as D216 and F259 in SH3 region (see Figure 3.21).

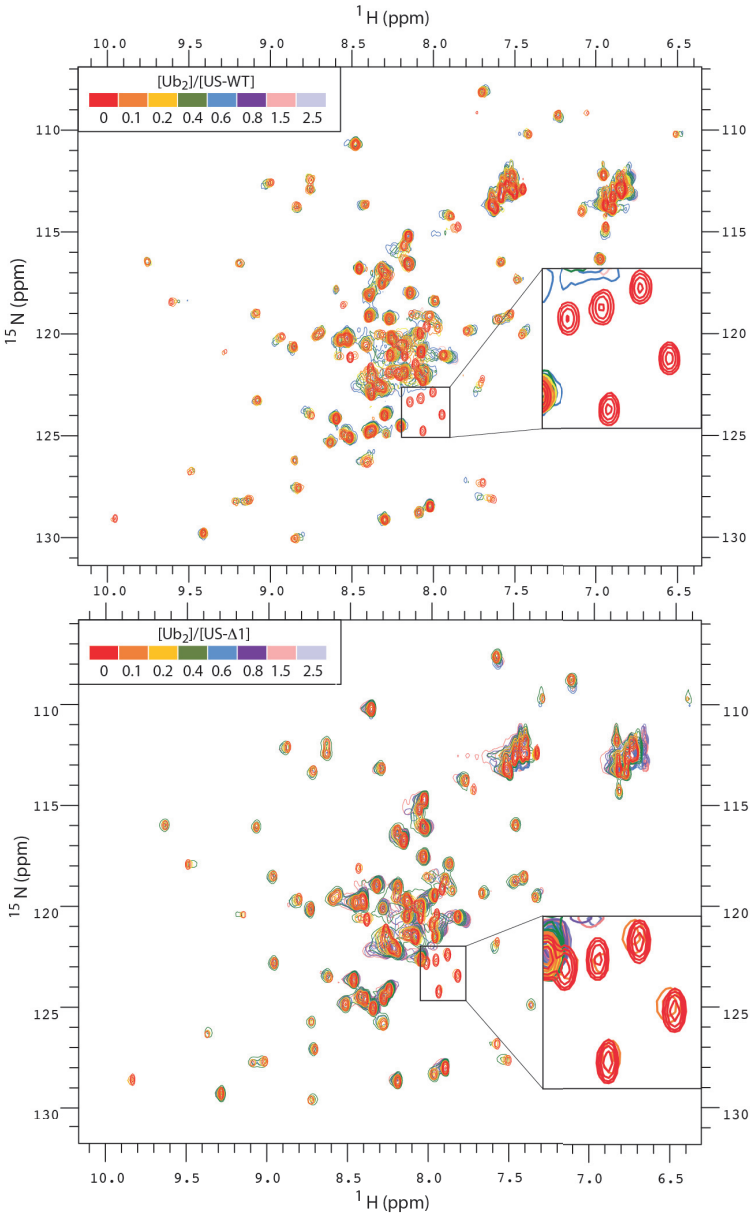


Figure 3.20 – Overlay of the ^{1}H-^{15}N-HSQC spectra of ^{15}N US-WT (top) and ^{15}N US-Δ1 (bottom) during titration with ^{14}N K63-Ub$_2$. The added K63-Ub$_2$ inscrease following the [Ub$_2$]/[US-WT] ratio from 0.1 to 2.5. The spectrum of US-WT before addition of K63-Ub$_2$ is in red where [US-WT] = 300 μM and the last point of titration is light blue with [US-WT] = 250 μM. The spectra were recorded at a ^{1}H frequency of 600MHz.

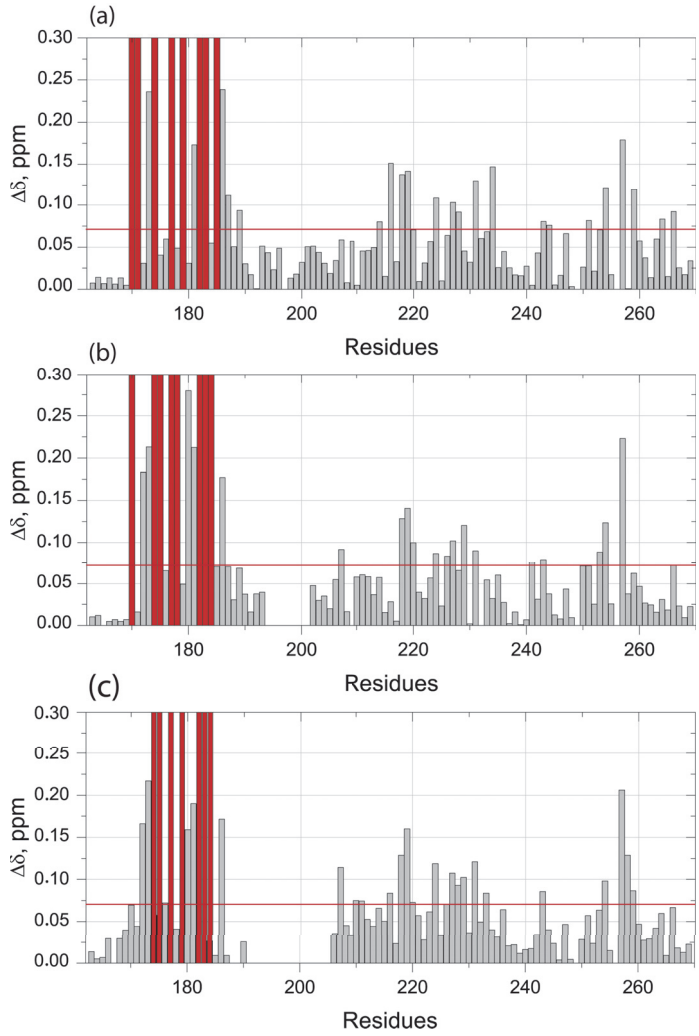


Figure 3.21 – Chemical shift perturbations were mapped on ^{15}N-labeled US-WT (a), US-Δ1 (b) and US-Δ2 (c) upon saturation with ^{14}N K63-Ub_2. Red bars indicate residues that disappear during the titration.

At the same time, certain residues of the UIM domain also show an increase in the CSPs on the US-Δ2 N-terminus while residues Q185 and Q187 experience a drastic decrease in their respective CSPs compared to US-WT. At the SH3 side of the linker of the 2 mutants US-Δ1 and US-Δ2, some residues show higher perturbations than in US-WT, for instance V207, K210 and V211 (see Figure 3.21).

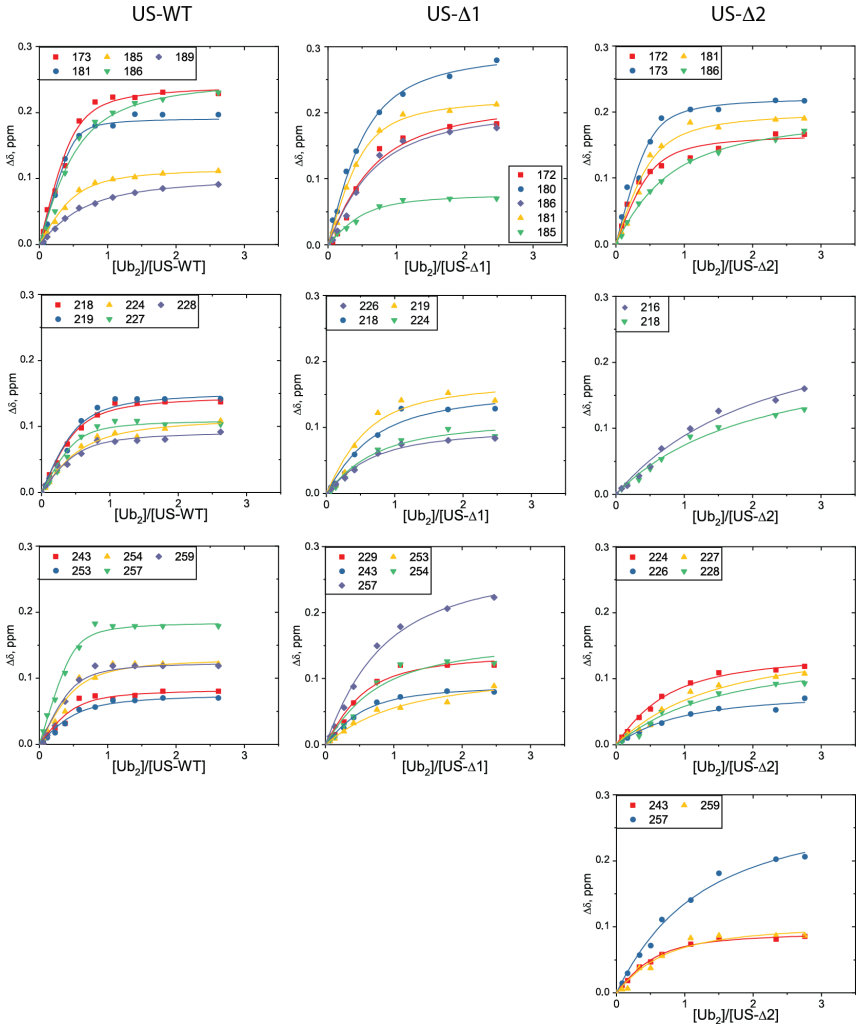


Figure 3.22 – The titration curves of some representative residues for the interaction between K63-Ub_2 and US-WT (left), US-Δ1 (middle), US-Δ2 (right) were represented with the experimental in dotted plot. For each residues presented, the smooth line of the same color depicts the fitted curves by using the 2:1 model.

In order to thoroughly quantify the modification, dissociation constants were calculated by fitting titration curves with the 2:1 stoechiometry model, which was chosen based on several observations. First the 2:1 model was the best model to correctly fit the titration curves. Second the titration induces great signal broadening on both UIM and SH3 domain at low concentration of K63-Ub_2, which indicates a significant increase in the molecular weight of the complex and suggests that there could be two molecules of K63-Ub_2 that come to bind to the titrated protein, although not simultaneously. Only residues with CSPs raising over the defined threshold of 0.07 in all mutants are taken into account for the calculation of K_d.

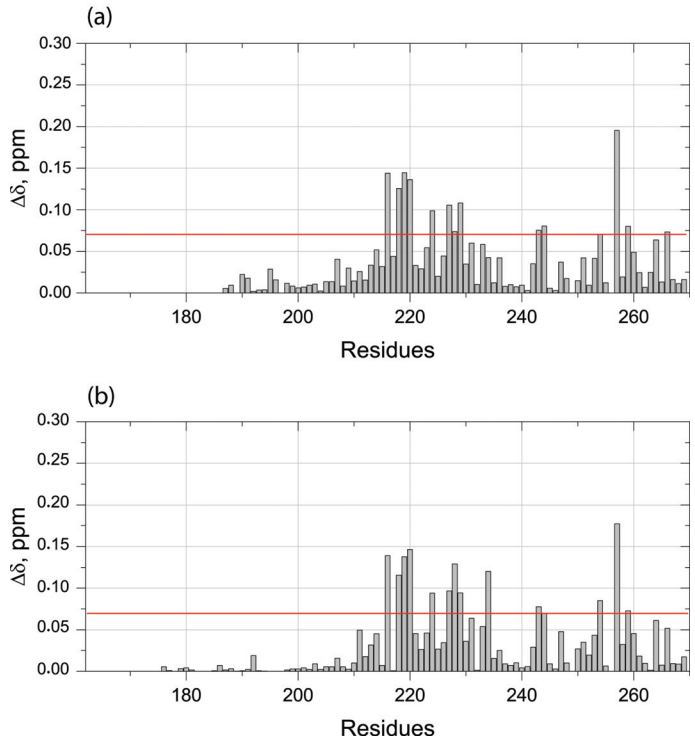


Figure 3.23 – Chemical shift perturbations were mapped on ^{15}N-labeled US-Δ3 (a) and US-Δ4 (b) upon saturation with ^{14}N K63-Ub_2.

While the dissociation constant seen on the UIM domain for the US-WT/ K63-Ub_2 complex is comparable to the one seen for the US-WT/Ub complex (Table 3.7), the affinity of the SH3 domain for K63-Ub_2 strongly increases compared to Ub and suggests an avid interaction, as already underlined in a previous work [49].

Moreover, as seen on the fitted titration curves, several residues such as E218, A219 or K186 exhibit important delay in saturation compared to US-WT (see Figure 3.22). This observation correlates with a significant increase in the dissociation constant of SH3 domain in US-Δ1 ($\approx$ 4.5 times) and US-Δ2 ($\approx$ 8 times) compared to US-WT (See Table 3.7). On the other hand, the UIM affinity remains similar or only shows a minor increase of less than 2 folds.

For the two mutants with partial or complete deletion of the UIM domain, the situation remains similar to what was seen in their titration with Ub. Indeed, the US-Δ3/K63-Ub_2 and US-Δ4/K63-Ub_2 complexes show the same CSP pattern as their complex with Ub (see Figure 3.23 and 3.16). Likewise the titration curves obtained for SH3 residues also stay fairly in the same shape (see Figure 3.24 and 3.18). When looking at the affinity of the interaction, US-Δ3 shows a slight decrease in K_d (39 $\pm$ 22 μM) compared to US-Δ4 (97 $\pm$ 49 μM) or to the other three constructs (Table 3.7). The affinity of the US-Δ3 toward K63-Ub_2 is therefore the highest among all constructs. This result gives sign of a mild inhibitory effect of the UIM domain toward SH3 as the total absence of the UIM domain in US-Δ3 has increased its affinity for K63-Ub_2. The effect accentuates with the shortening of the linker as the affinity for K63-Ub_2 progressively decreases with the number of residues truncated. Therefore, a steric hindrance and an excluded volume constraint of UIM occur which reduces the conformational space available for the binding of SH3.

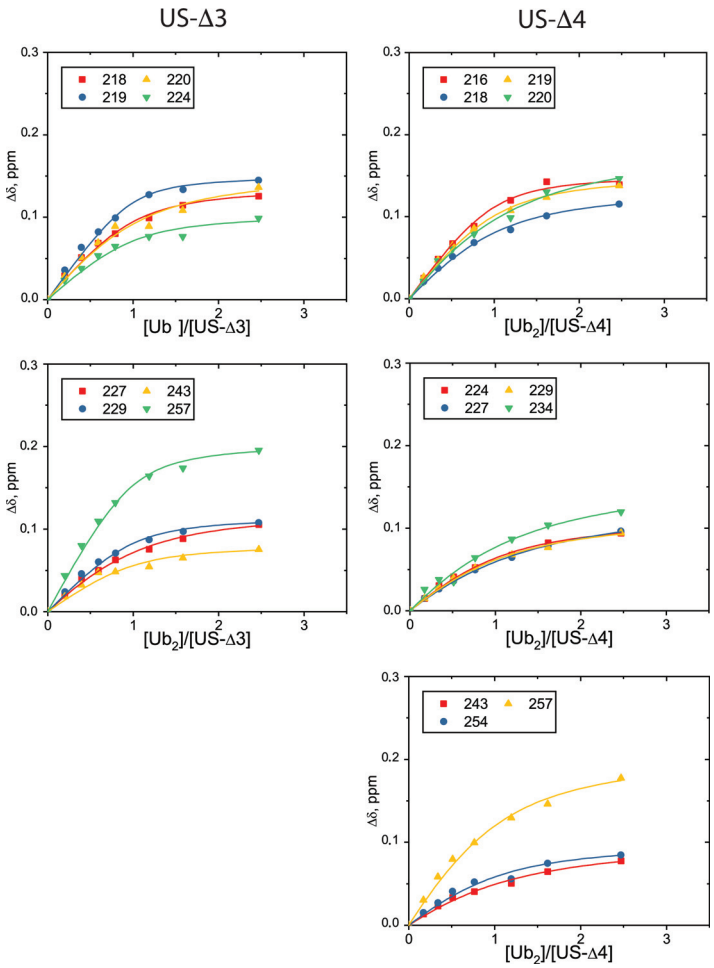


Figure 3.24 – The experimental titration curves of some representative residues for the interaction between K63-Ub_2 and US-Δ3 (left), US-Δ4 (right) were presented in dotted plot. For each residue presented, the smooth line of the same color depicts the fitted curves by using the 1:1 model.

The results elaborated in this chapter were submitted and accepted for publication in ”Scientific reports” with the title ”Molecular recognition of ubiquitin and Lys63-linked diubiquitin by STAM2 UIM-SH3 dual domain: the effect of its linker length and flexibility”

4

VHS-UIM and the important presence of its flexible linker

The VU-WT is a segment of the STAM2 protein with VHS and UIM domains tethered together by a linker of 27 amino acids. To define the importance of this linker existence on the protein dynamics and characteristics, a set of two derived mutants was designed: the VU-P mutant with a modified linker where 14 residues were substituted by prolines and the VU-T mutant with 27 amino acids of the linker completely removed (see Figure 4.1).

Similar to the US-WT construct, it has been previously proven that no interaction occurs between the two domains VHS and UIM[47]. A set of ^{1}H-^{15}N-HSQC spectra was recorded at a ^{1}H frequency of 900 MHz for the 3 constructs VU-WT, VU-P and VU-T. The choice of a high magnetic field is based on the need to acquire high resolution spectra for each of the constructs. Indeed, the VU constructs all contain a total of nine α helices, which

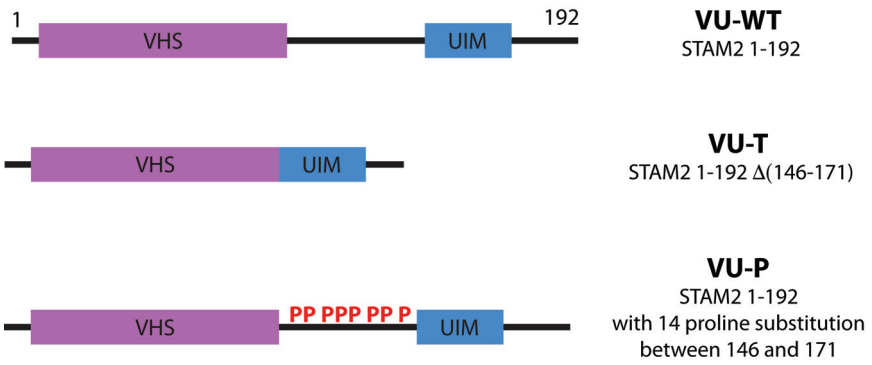


Figure 4.1 – Schematic representation of VU-WT, VU-T and VU-P constructs.

NMR cross peaks fall within a narrow region on the spectra and render them difficult to interpret. Moreover, larger proteins tend to tumble more slowly in solution which increase relaxation rates, thus a low signal-to-noise level. For this reason, all VU constructs samples were produced with additional deuterium labelling (up to 70 %, see Methods) to lessen this effect (Figure 4.2). The ^{1}H-^{15}N-HSQC spectrum of VU-WT was then compared to the previous individual UIM and VHS HSQC spectrum to check that there is no change in folding and no interaction between the two domains (Figure 4.3).

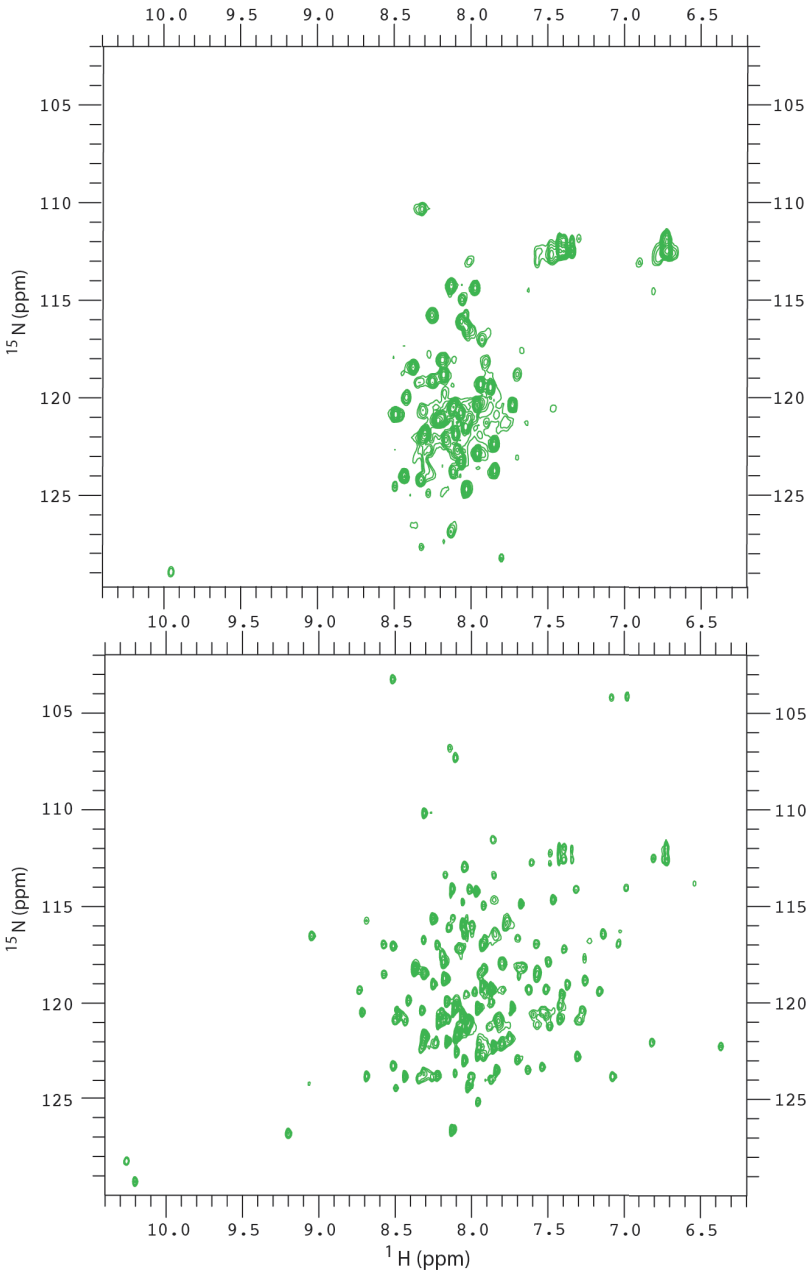


Figure 4.2 – ^{1}H-^{15}N-HSQC spectrum was acquired at a ^{1}H frequency of 600 MHz and at 15 °C for the ^{1}H-VU-P sample (upper panel) while the ^{2}H-VU-P sample spectrum was recorded a ^{1}H frequency of 900 MHz at same temperature (lower panel). The high relaxation rates due to the large size of the protein make it hard to obtain a high resolution of the spectrum. By reducing spin diffusion with the partial deuteration, the signal-to-noise ratio of the resonances located in the border region significantly increases.

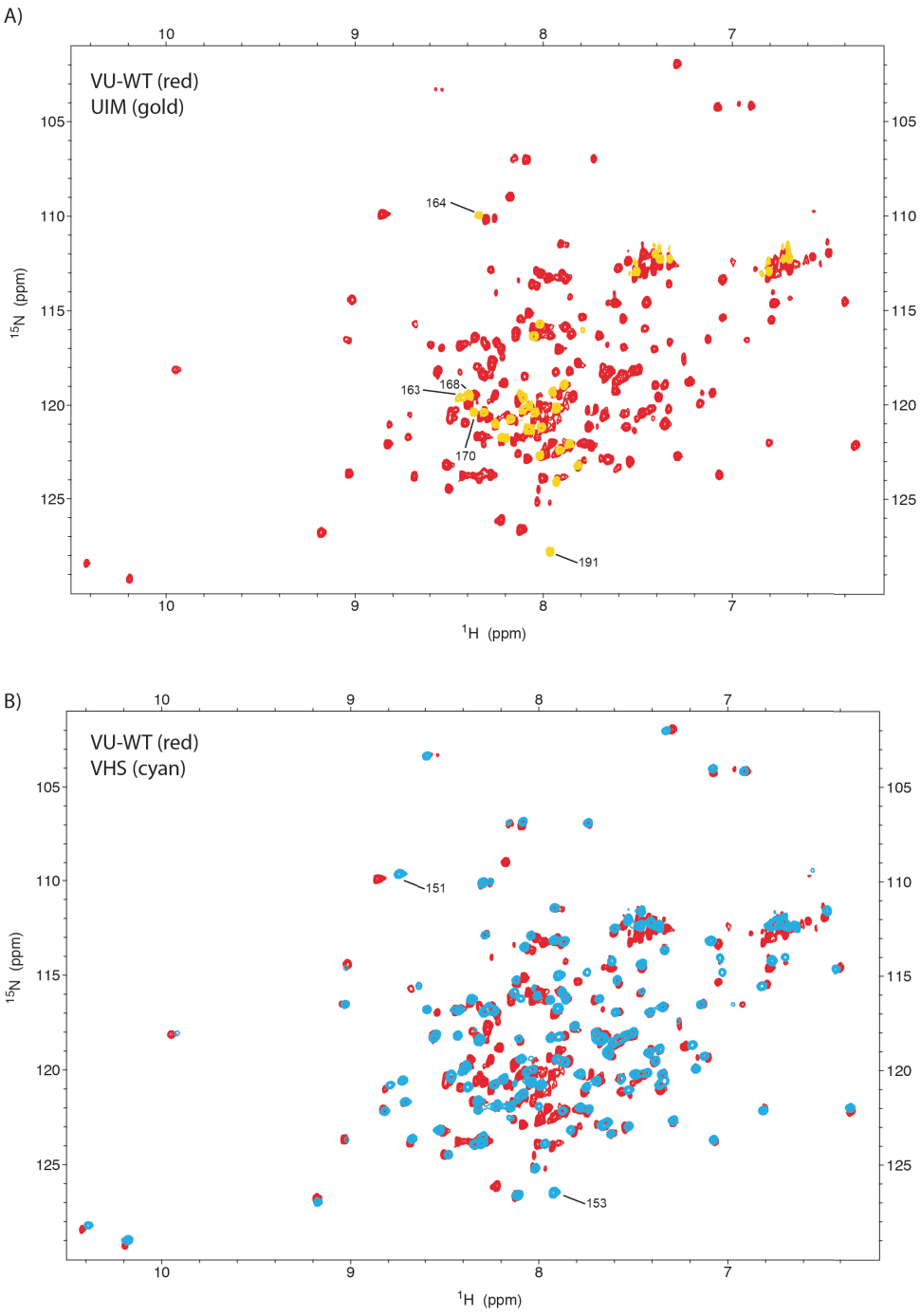


Figure 4.3 – The VU-WT ^{1}H-^{15}N-HSQC spectrum was overlayed with UIM (upper panel) and VHS spectrum (lower panel). Spectra were recorded at a ^{1}H frequency of 900 MHz and at 15 °C

1. VU-T and the missing linker

For VU-T, partial overlapping was observed in the two domain VHS and UIM (see Figure 4.4) with slight shifts, suggesting that the structure of each domain remains mostly intact. However, a possible constraint due to the absence of the linker has a significant impact on the electronic environment surrounding the two domains VHS and UIM, which explains the slight shifts in a large amount of cross peaks when compared the VU-T to the individual domain VHS and UIM (Figure 4.4).

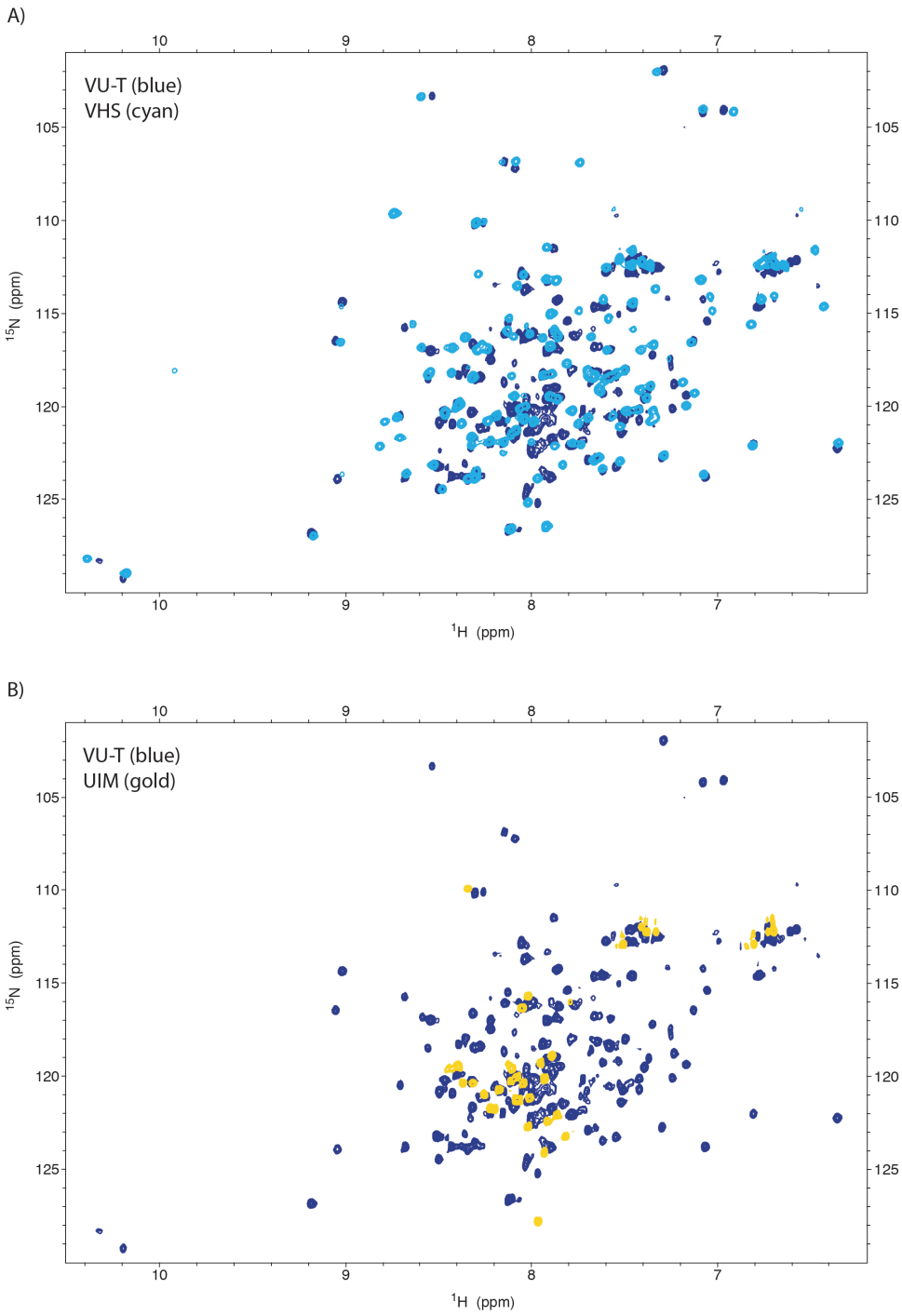


Figure 4.4 – The VU-T ^{1}H-^{15}N-HSQC spectrum was overlayed with VHS spectrum (upper panel) and UIM spectrum (lower panel). Spectra was recorded at a ^{1}H frequency of 900 MHz and at 15 °C

To obtain further information on the integrity of the mutants, circular dichrosim experiments were performed (see Methods). Similarly to the US constructs, the CDSSTR algorithm was used to seek for the contribution of different secondary structure in reference to two difference sets: SMP180 [82] and SP175 [86]. Regarding 2% of error using this algorithm, the helicity percentage of the VU-WT construct are 62 % (64 %) while the disordered contribution is 27 % (25 %), which makes the average length of the helix part approximately 119 (123) residues while roughly 50 residues cover the disordered region (See Table 4.1). The results nicely agree with the expected number of helix residues issued from the VU-WT construct model (124 residues). As for VU-T, the number of the helix residues remains similar to VU-WT, which is 130 (125), while the number of residues in unstructured region decreases drastically to 22 (28) concomitant with the 27 truncated amino acids.

Construct	Helix (%)	Disordered (%)	Helix (residues)	Disordered (residues)
VU-WT(192)	62 % (64 %)	27 % (25 %)	119 (123)	52 (48)
VU-T(165)	79 % (76 %)	13 % (17 %)	130 (125)	22 (28)

Table 4.1 – The percentage of helicity has been obtained by means of the Dichroweb server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) and the CDSSTR method [83] [84] [85] with the use of two different sets of reference. The results found with the reference set SMP180 [82] is represented in black and in blue between brackets are the results given by using the set SP175 [86] . The average length of the disordered region is calculated using the number of amino acids.

Therefore, regardless the reference set that was used in the CDSSTR method, the VU-T construct shows a significant change in the secondary structure composition. In fact, considering the average length of this unstructured region, VU-T has $\approx$ 20-30 residues less than VU-WT that fairly coincides with the 27 amino acids truncated. The removal of this linker therefore does not affect the helical structure of VHS and UIM domains considering approximately 124 residues implicated in α-helices in the structure (see Table4.1). Consequently, any modification in the protein conformational and dynamical characteristic is taken into account solely as the result of the absent linker.

In the next step, further NMR experiments were performed to examine different properties of the protein. To facilitate the analysis of those experiments, previous backbone

assignment of the VU-WT construct that covers roughly 70 % of the total residues has been used as reference to assign the ^{1}H-^{15}N-HSQC spectrum of VU-T.

1.1. Modification of the protein dynamics

In order to study the dynamical characteristics at the different time scales (in the ps-ns range), measurement of the three commonly used spin relaxation parameters R_1, R_2 and the ^{1}H-^{15}N cross-relaxation rates, via the steady-state ^{1}H-^{15}N-NOEs were performed at ^{1}H frequency of 900 MHz. In general, for the two constructs, a high flexibility is observed in the C-terminus and N-terminus extremities, which is reflected by a significant decrease of R_2 and NOEs compared to the well folded core of VHS (Figure 4.5). In addition, for VU-WT, the R_1, R_2 and NOE values of VHS and UIM exhibit significantly distinct levels indicating an independent tumbling and rotational diffusion tensor of the two domains. Effectively, the structure of VHS is mainly constituted by a bundle of 9 helices. Thus, VHS exhibits a higher level of compactness than UIM, the structure of which was determined as an α-helix by homology modelling. Moreover, within each domain VHS or UIM, the values of R_1, R_2 and NOE are fairly homogeneous with very few fluctuations showing a fairly stable structure throughout the two domains.

For VU-T where the total linker has been removed, a significant modification is observed. First looking at all three relaxation parameters, VHS does not exhibit any modifications as there is a nice overlapping with VU-WT values (Figure 4.5). By continuing throughout the sequence, the relaxation rates of UIM exhibit a drastic change, especially in the N-terminus region. In fact, the first half of UIM appears to share the same level of R_1 as the VHS domain of $\approx$ 0.5 s^{-1} then goes up again toward its original level. A similar situation occurs for the R_2 rate where the level of the N-terminus half of UIM is dragged up to $\approx$ 25-30 s^{-1}. As for NOEs, throughout the whole length of UIM, fluctuations are observed but in general it exhibits a higher level of NOEs compared to UIM in VU-WT. By examining solely the profile of the three spin relaxation parameters, the UIM helix appears to exhibit a more stable conformation at the N-terminus extremity but keeps fairly the same dynamics toward the end of its secondary structure. Effectively, by looking at the average NOE values, the original gap between the

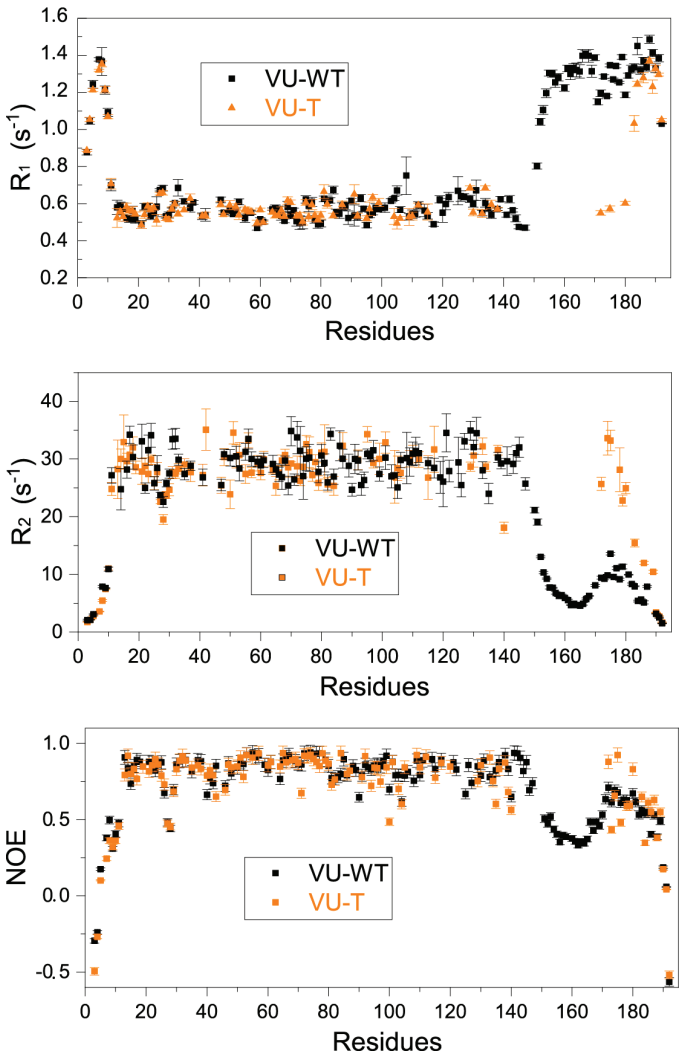


Figure 4.5 – Relaxation parameters R_1, R_2 and ^{1}H-^{15}N-heteronuclear NOE for the two constructs: VU-WT (black) and VU-T(orange).

average relaxation rates of VHS and UIM domains (of VU-WT) shows considerable decrease. The two domains are brought up close to each other in term of R_1 and R_2 levels. The interdomain motion is considered nearly abolished as the two domains behave as one entity. Therefore, the average R_1, R_2 and NOEs of VU-T were calculated by taking into account also residues of UIM. An increase in R_2 and NOEs was observed compared to the VHS domain of VU-WT which reflects the fact that VHS become more heavy with the additional UIM at its

C-terminal (see Table 4.2).

		R_1 (s^{-1})	R_2 (s^{-1})	NOE
VU-WT	VHS	0.57 ± 0.02	26.9 ± 1.9	0.84 ± 0.01
	UIM	1.30 ± 0.02	8.9 ± 0.2	0.60 ± 0.03
VU-T		0.6 ± 0.02	28.7 ± 1.7	0.80 ± 0.04

Table 4.2 – Averaged longitudinal (R_1) and transverse (R_2) relaxation rates as well as averaged steady-state ^{1}H-^{15}N-NOE values measured for the two constructs VU-WT and VU-T at 15 °C and at a ^{1}H frequency of 900 MHz.

This result is due to the space constraint occurring when the long linker is missing, which drags the two domains notably close to each other to the point where the boundary between them is abolished. Likewise the correlation time of VU-T was also calculated considering VHS and UIM as a whole. The value of τ_C increases, which is 15.7 ns compared to 14.8 ns for VHS in VU-WT. This result agrees well with the increase in R_2 and NOEs value (see Table 4.3).

		D_{xx} $^{(a)}$	D_{yy} $^{(a)}$	D_{zz} $^{(a)}$	τ_C $^{(b)}$
VU-WT	VHS	1.00 ± 0.07	1.13 ± 0.04	1.26 ± 0.05	14.76 ± 0.22
	UIM	2.81 ± 1.38	2.81 ± 1.38	7.28 ± 1.01	3.87 ± 0.22
VU-T		0.79 ± 0.06	0.91 ± 0.07	1.47 ± 0.08	15.74 ± 0.27

Table 4.3 – Summary of the diffusion tensors calculated using the ROTDIF program [75] [76].
Axial and fully anisotropic models were used for UIM and VHS domain, respectively
(a) The principal values of the diffusion tensors, in $10^7 s^{-1}$.
(b) The overall rotational correlation time, $\tau_C = 1/(6\ D_{iso})$, in nanoseconds,
with $D_{iso} = (D_{xx} + D_{yy} + D_{zz})/3$

Additionally to spin relaxation parameters R_1, R_2 and NOE, the spectral densities operating at three different frequencies J(0), J(ω_N) and J(0.87ω_H) are also extracted [74]. In fact, the three analytical expressions make no assumption about the nature of the rotational diffusion. As can be seen in Figure 4.6, the VHS domain does not exhibit significant difference between VU-WT and VU-T while UIM shows significantly higher J(0) level and lower J(0.87ω_H), that reflect a high contribution of slow motion at nanosecond timescales

while the fast internal motion is nearly nullified. The results suggest that the N-terminus part of UIM undergoes a steric hindrance as it approaches the VHS domain while its C-terminus part keeps a fast motion at the picosecond timescales, indicating the dynamic characteristic of the UIM α-helix.

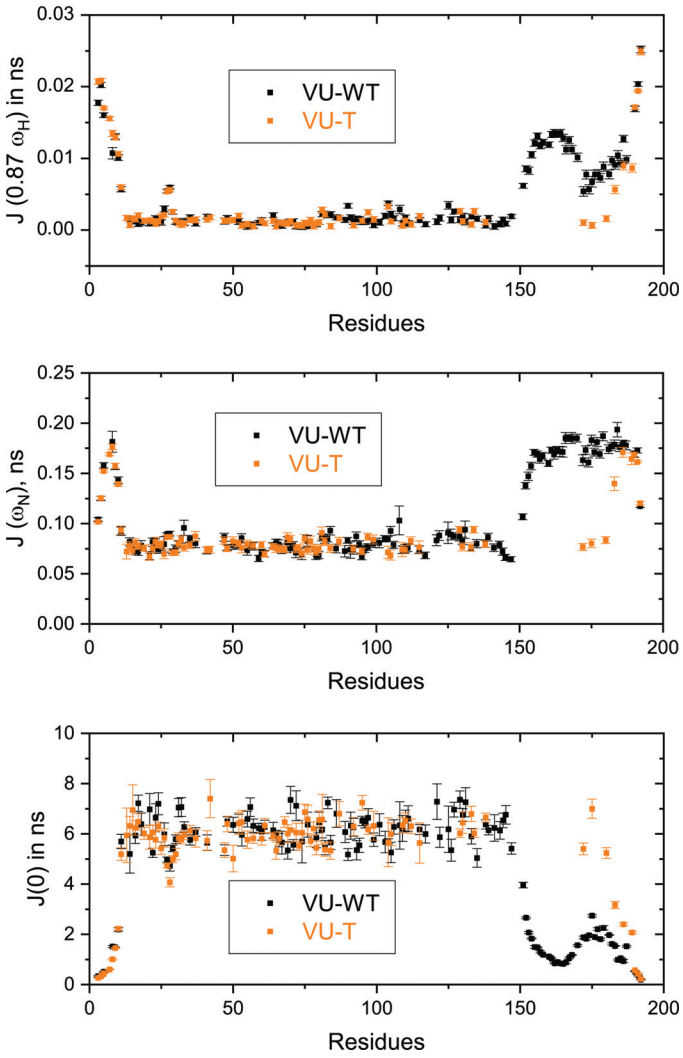


Figure 4.6 – Reduced spectral densities for the VU variants VU-WT (black) and VU-T (orange) using the equations derived in chapter 2.

1.2. Modification in the interaction with K63-Ub_2

The linker between VHS and UIM domains appears to play a significant role in maintaining the interdomain motion. With that in mind, the interaction between the VU constructs and K63-Ub_2 was investigated to perceive whether this modification of the protein dynamics has any connection to its molecular recognition capacity. Titration experiments is performed by a series of ^{1}H-^{15}N-HSQC spectra on ^{15}N-labeled VU constructs upon addition of non labeled K63-Ub_2.

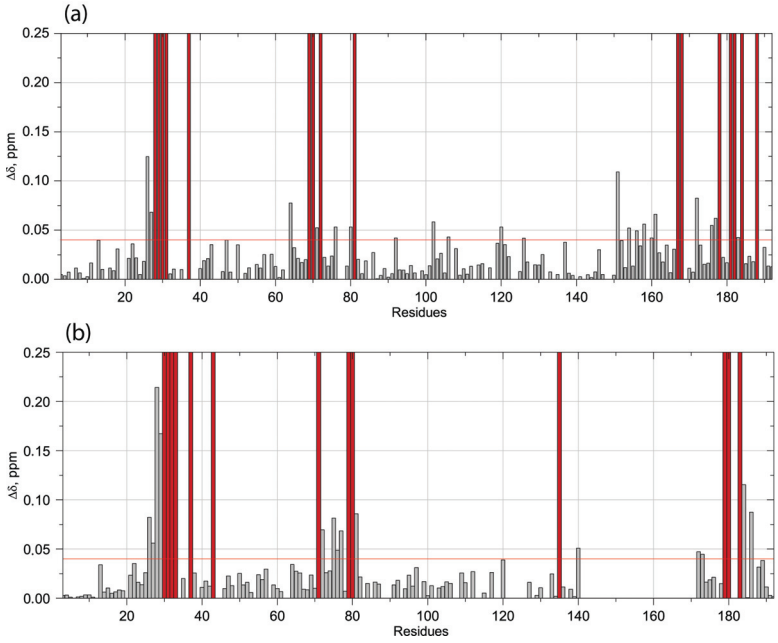


Figure 4.7 – Chemical shift perturbations were mapped on ^{15}N-labeled VU-WT (a) and VU-T (b) upon saturation with ^{14}N K63-Ub_2. Red bars indicate residues that disappear during the titration. The horizontal red line at 0.04 ppm represents the threshold beyond the residues are used for further quantitative analysis in order to find the dissociation constant.

At first glance, the CSPs patterns show several changes (Figure 4.7). On closer look at the two residues E28 and D29 in VU-WT, the signal disappears after the first addition of Ub_2 suggesting a tight binding at this location (Figure 4.8, upper panel). Conversely for VU-T, it is possible to follow the CSP of these two residues E28 and D29 (Figure 4.8, lower panel).

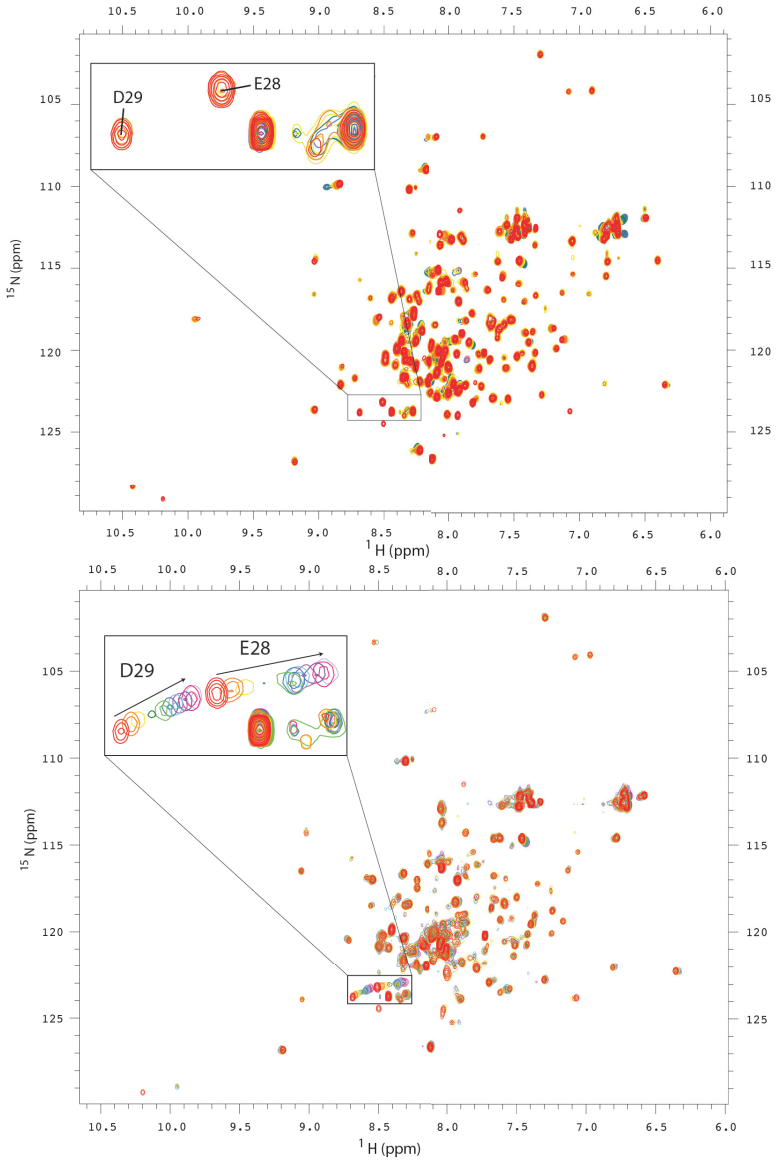


Figure 4.8 – Overlay of the ^{1}H-^{15}N-HSQC spectra of ^{15}N VU-WT (upper panel) and ^{15}N VU-T (lower panel) during titration with ^{14}N K63-Ub_2. The concentration of the added K63-Ub_2 follow a ligand to subtrate ratio ranging from 0.1 to 2.5 for VU-WT and up to 4 for VU-T. The spectrum of the 2 constructs before any addition of K63-Ub_2 are colored in red: [VU-WT] = 160μM and [VU-T] = 180μM. The last point of titration is colored in light blue where the concentration of VU-WT and VU-T is 140 μM and 150μM respectively. A zoomed box was selected to show the disappearing of of the two residues E28 and D29 upon the first addition of K63-Ub_2 in VU-WT. The same two residues conversely nicely shift in VU-T while drawing a fairly long trajectory.

On the CSPs map, the VHS domain displays the same chemical shift perturbation pattern for VU-WT and VU-T. However, except several resonances that immediately disappear after the addition of a low concentration of K63-Ub_2, most of the perturbations show a low intensity ($\approx$ 0.25 ppm). Overall, the perturbation pattern for the VU-WT and VU-T is similar as well as the direction of the shifts. This clearly indicate that the VHS domain in VU-WT or VU-T share the same mode of binding with K63-Ub_2. To obtain information regarding the binding site, a threshold of 0.04 ppm has been chosen. The residues with perturbation rising above that threshold are considered as being located at or very close to the interaction region. In all two constructs, the VHS CSPs are mainly located within two areas centered around: T26, T27, E28, D29 and T71, L72. This result consistently agrees with the previous CSPs studies on the VU-WT constructs [47], indicating a similar interaction sites between the VHS domain and K63-Ub_2. Regarding UIM domain in VU-WT, higher perturbations occur as most of the cross peaks that belong to UIM namely K167, N168, E179, L182, Q183, Q185 and Q188 that disappear upon addition of K63-Ub_2. Other residues such as I174, K176 and A177 undergo signal coalescence indicating a fairly tight binding.

		K63-Ub_2 K_d (μM)
VU-WT	VHS	53.9 $\pm$ 29.2
	UIM	72.1 $\pm$ 29.1
VU-T	VHS	76.6 $\pm$ 29.5
	UIM	2548.0 $\pm$ 162

Table 4.4 – Dissociation constants measured for the interaction between VU-WT and VU-T with K63-Ub2 using a 2:1 stocechiometry model (See Methods)

For the VU-T construct, the interaction site in UIM domain is nearly abolished as all residues in the C-terminus of UIM that originally exhibit great alteration in VU-WT construct do not show any perturbation. The remaining N-terminus residues: D173, I174, K176 and A177 display smaller CSPs compared to the corresponding residues in VU-WT. The modification of the CSPs profile nicely correlates with the further quantitative analysis.

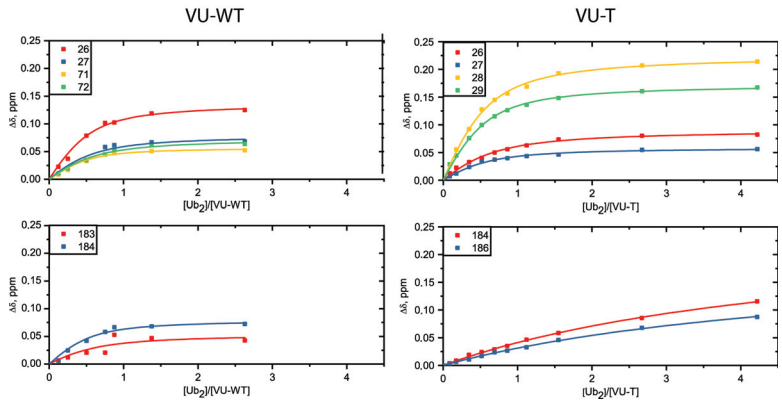


Figure 4.9 – The titration curves of some representative residues for the interaction between K63-Ub_2 and VU-WT (left) and VU-T (right) were represented with the experimental data points in dotted plot. For each residue presented, the smooth line of the same color depicts the fitted curves by using the 2:1 model.

Effectively, the dissociation constant of the VHS domain remains in the same value range, 53.9 μM for VU-WT and 76.6 μM for VU-T (Table 4.4) although a slight decrease in affinity is perceived in the mutant VU-T. The most significant change can be observed in the UIM domain. The dissociation constant considerably goes out of range in VU-T, where it increases by factor of 35 when compared to that of VU-WT construct, which makes the UIM and K63-Ub_2 interaction to switch from a low micromolar affinity to a low millimolar affinity. This outcome is reflected as well by looking at the fitted titration curves (see Figure 4.9). In fact, unlike VU-WT where the titration reaches a plateau at a Ub_2 ratio of 2, the UIM part of VU-T requires a higher concentration of Ub_2 in order to reach saturation. The result suggests that VHS domain in VU-T has an inhibitory effect toward UIM and prevents its binding with K63-Ub_2. Therefore, this absence of the linker first leads to a great restriction in space of the molecule, which makes the molecular recognition considerably harder and significantly alters the proper function of the protein.

2. VU-P and the role of the linker flexibility

The VU-P mutant was designed with the purpose to seek for the impact of the linker flexibility. In fact, proline is the most conformationally restricted amino acid residue with a cyclic side chain and so the rotation around the peptidic bond is constrained by its inclusion in the pyrrolidine ring. The substitution of 17 amino acids by prolines in the linker sequence is expected to give the molecule higher stiffness in this unstructured region.

Similar to VU-T construct, the VU-P ^{1}H-^{15}N-HSQC spectrum exhibits partial overlapping with the VHS and UIM spectra with slight shifts on several additional cross peaks comparing to VU-T (Figure 4.10). The situation is possibly due to the substituted prolines that bring change to the electronic environment surrounding the two domains VHS and UIM, suggesting that conformational modification might occur in VU-P construct.

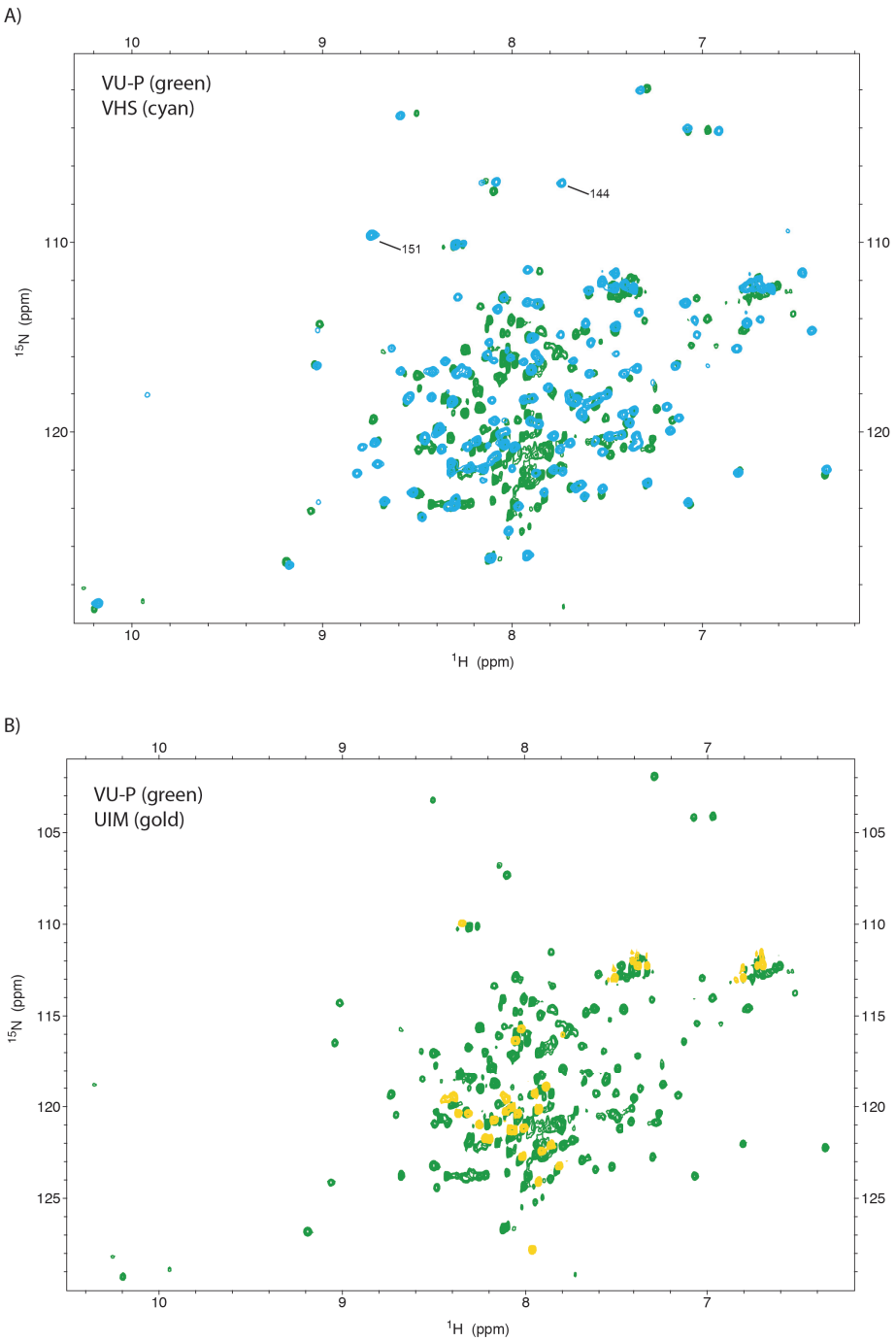


Figure 4.10 – The VU-P ^{1}H-^{15}N-HSQC spectrum was overlayed with UIM spectrum (upper panel) and VHS spectrum (lower panel). Spectra were recorded at a ^{1}H frequency of 900 MHz and at 15 °C.

2.1. Conformational and dynamical modification

Circular dichroism data are obtained for VU-P using the same experimental conditions and analysis method as the previous two constructs VU-WT and VU-T. Regardless the reference set SPM180 [82] or SP175 [86], a higher percentage of helicity is noticed which is (75 % and 74 %) concomitant with a decrease in the disordered contribution (16 % and 18%). The average length of the α-helix part therefore increases to ≈ 144 residues, which is ≈ 22 more residues than in VU-WT. On the other hand, a difference of ≈ 21 residues is observed for the disordered region in VU-P, which nicely fits with the previous observation on the α-helix as though part of the linker is rearranged into a helical structure.

Construct	Helix (%)	Disordered (%)	Helix (residues)	Disordered (residues)
VU-WT(192)	62 % (64 %)	27 % (25 %)	119 (123)	52 (48)
VU-P(192)	75 % (74 %)	16 % (18 %)	144 (142)	31 (35)

Table 4.5 – The percentage of helicity has been obtained by means of the Dichroweb server (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml) and the CDSSTR method [83] with the use of two different sets of reference. The results found with the reference set SMP180 [82] is represented in black and in blue between brackets are the results given by using the set SP175 [86]. The average length of the disordered region is calculated using the number of amino acid.

By studying the relaxation parameters of the VU-P construct, the same situation as VU-WT is observed. The gap between the level of R_1 and NOE values of the VHS and UIM domains remains the same, suggesting that the proline substitution has minor effect on the independent tumbling of the two domains (Figure 4.11 and Table 4.6). However, on a closer look at the linker region, all three relaxation parameters R_1, R_2 and NOE show different levels comparing to that of VU-WT. In VU-WT, the linker and the two domains exhibit a smooth connection in term of the R_1, R_2 and NOEs values.

As for VU-P, on all three relaxation parameters R_1, R_2 and NOEs, the linker region (146-171) appears to be rather flat throughout the linker length with an abrupt deviation from the average level of VHS and UIM domains (Figure 4.11). Likewise, the average R_2 of the VHS domain slightly increases to 32.2 s^{-1}, that draws the gap between the 2 domains

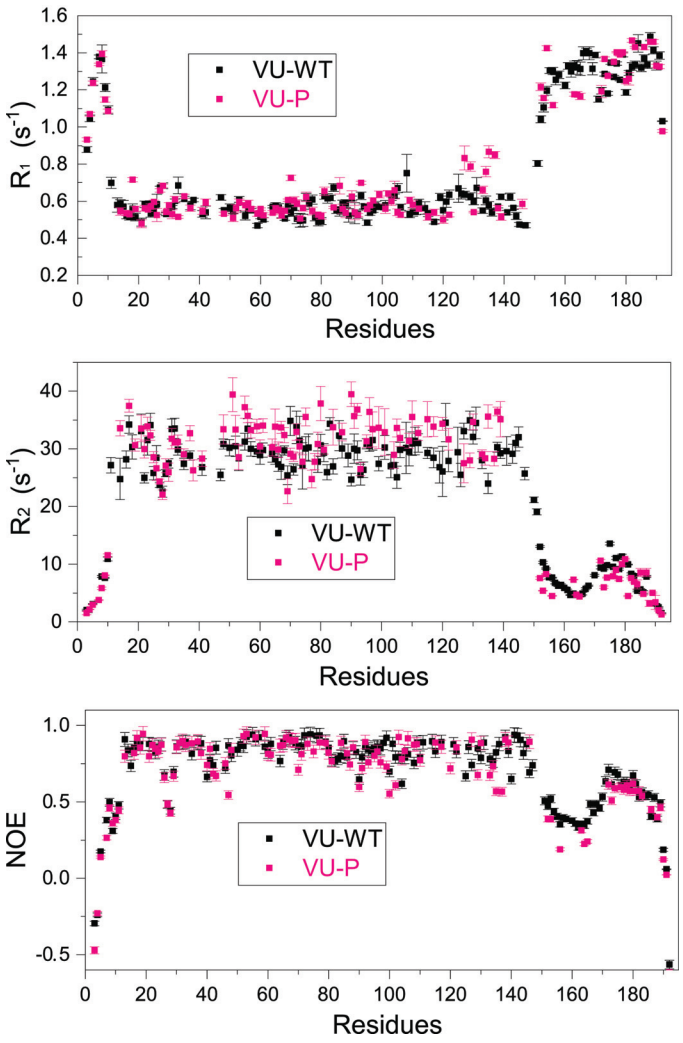


Figure 4.11 – Relaxation parameters R_1, R_2 and ^{1}H-^{15}N-heteronuclear NOE for the two constructs: VU-WT (black) and VU-T(pink).

higher, suggesting a greater impact of the linker in restricting the interdomain motion. The results imply a local stabilization, consistent with the fact that proline provides stiffness to the peptidic bond and hence a smaller degree of freedom in space.

		R_1 (s^{-1})	R_2 (s^{-1})	NOE
VU-WT	VHS	0.57 ± 0.02	26.9 ± 1.9	0.84 ± 0.01
	UIM	1.30 ± 0.02	8.9 ± 0.2	0.60 ± 0.03
VU-P	VHS	0.59 ± 0.02	32.2 ± 1.7	0.81 ± 0.04
	UIM	1.35 ± 0.01	7.8 ± 0.3	0.57 ± 0.03
		1.30 ± 0.01	7.2 ± 0.2	0.48 ± 0.02

Table 4.6 – Averaged longitudinal (R_1), transverse (R_2) relaxation rates and steady-state ^{1}H-^{15}N-NOE values measured for VU-WT and VU-P at 15 °C and at a ^{1}H frequency of 900 MHz. For the UIM domain of the VU-P constructs, the averaged relaxation rates are calculated by taking into account only original UIM residues from E172 to Q188 (in black) or with the linker residues included from S152 to Q188 (in blue).

The result is supported by examining the J(0) values (in the range of 146 to 171 see Figure 4.12), which appear to slightly decrease and reach the same level as the UIM domain. The internal motion of the proline-rich linker at nanosecond timescales is reduced suggesting a possible folding of the region.

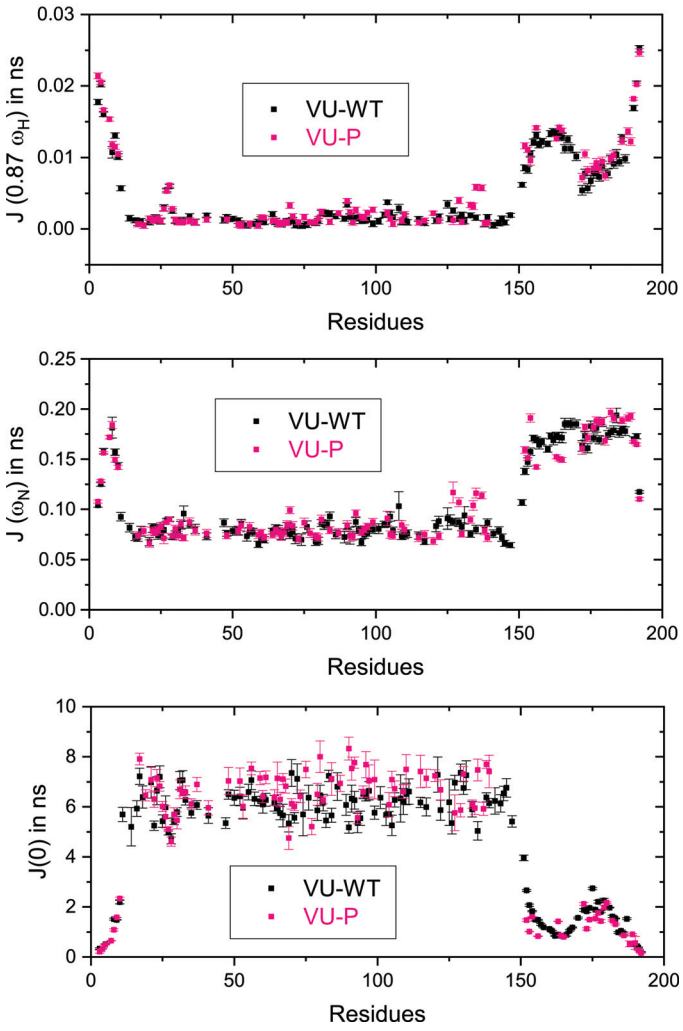


Figure 4.12 – Reduced spectral densities for the VU variants VU-WT (black) and VU-P (pink) using the equations derived in chapter 2.

Moreover, to correctly extract the correlation time of each domain from the three ^{1}H-^{15}N-HSQC relaxation parameters, the residues that participate in the secondary structure of the domain are solely examined. For the UIM domain of VU-P, while using the same range of residues (172-188) the correlation time obtained is 1.81 ns. When the linker residues (152-171) are also included in the calculation, the correlation time goes up to 4.48 ns which is fairly close to the original τ_c of the UIM helix (see Table 4.7), supporting the idea of the linker folding. In summary, the substitution of 14 residues among the 27 amino acids by proline stabilizes this linker region and potentially leads to a conformational change where the originally-disordered linker could be partially folded into an α-helix secondary structure. The modification of the amino acids composition in the linker region that connects VHS and UIM therefore should give rise to a significant change in the dynamical as well as the conformational characteristics of the protein.

		$\mathbf{D}_{xx}$ $^{(a)}$	$\mathbf{D}_{yy}$ $^{(a)}$	$\mathbf{D}_{zz}$ $^{(a)}$	τ_C $^{(b)}$
VU-WT	VHS	1.00 ± 0.07	1.13 ± 0.04	1.26 ± 0.05	14.76 ± 0.22
	UIM	2.81 ± 1.38	2.81 ± 1.38	7.28 ± 1.01	3.87 ± 0.22
VU-P	VHS	0.88 ± 0.06	0.99 ± 0.06	1.14 ± 0.08	16.61 ± 0.30
	UIM	3.05 ± 2.01	3.05 ± 2.01	21.56 ± 11.88	1.81 ± 1.07
		4.42 ± 0.97	4.42 ± 0.97	2.34 ± 0.37	4.48 ± 0.27

Table 4.7 – Summary of the diffusion tensors calculated using the ROTDIF program [75] [76]. Axial and fully anisotropic models were used for UIM and VHS domain, respectively
(a) The principal values of the diffusion tensors, in $10^7 s^{-1}$.
(b) The overall rotational correlation time, $\tau_C = 1/(6\ D_{iso})$, in nanoseconds,
with $D_{iso} = (D_{xx} + D_{yy} + D_{zz})/3$

2.2. Change in affinity toward K63-Ub_2

The proline substitution increases the linker stiffness and appears to slightly decrease the interdomain dependency. In the subsequent step, molecular recognition of the construct toward K63-Ub_2 is examined. Compared to VU-WT, similar perturbations are observed in both domains although VU-WT exhibits additional CSPs in the C-terminus region of its linker: S152, T154, A158 and A159 (Figure 4.13). Moreover, a vast majority of perturbations observed in VU-WT undergoes signal coalescence such as the residues E28, D29, A69, L70 and L72 in VHS domain and I174, K176 and A177 in UIM domain, which are not seen in VU-P (Figure 4.14).

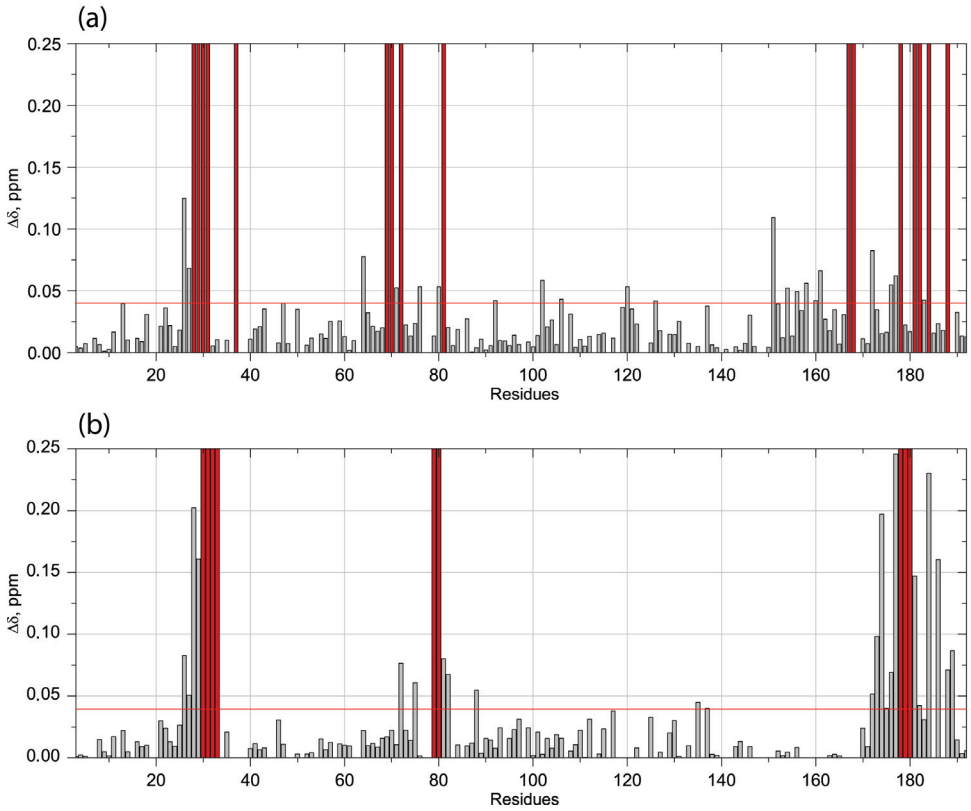


Figure 4.13 – Chemical shift pertubations were mapped on ^{15}N-labeled VU-WT (a) and VU-P (b) upon saturation with ^{14}N K63-Ub_2. Red bars indicate residues that disappear during the titration. The threshold for K_d calculation are represented as th horizontal red line at 0.04 ppm.

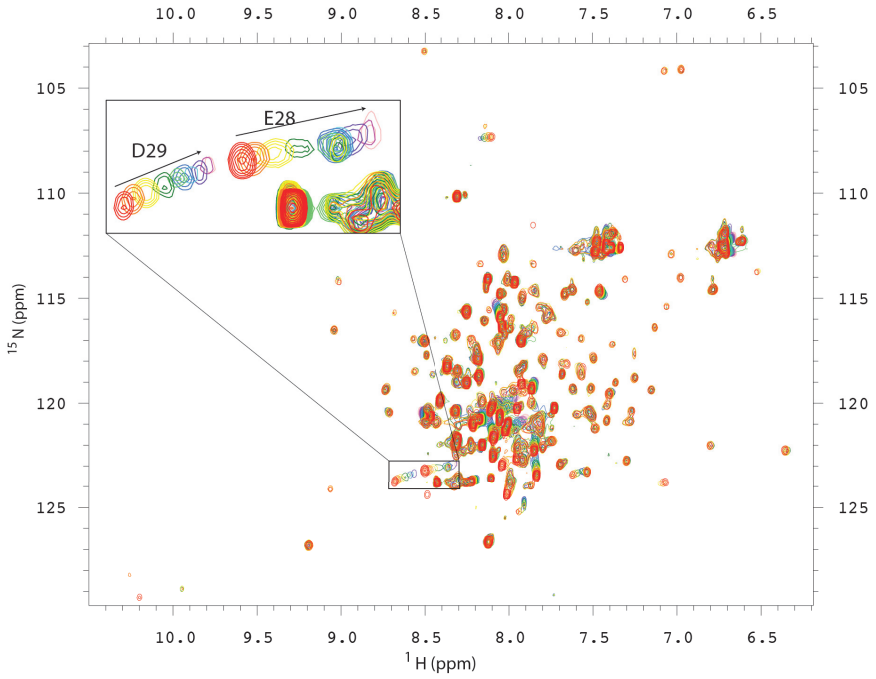


Figure 4.14 – Overlay of the ^{1}H-^{15}N-HSQC spectra of ^{15}N VU-P during titration with ^{14}N Ub_2. The concentration of the added K63-Ub_2 increases according to the K63-Ub_2 ligand concentration to subtrate ratio ranging from 0.1 to 3.5. The spectrum of the VU-P before addition of K63-Ub_2 are colored in red : [VU-WT] = 160μM and [VU-T] = 180μM. The last point of titration is colored in light blue where the concentration of VU-WT and VU-T is 140 μM and 150μM respectively In the zoomed box, two residues E28 and D29 show a great shift mostly in the ^{1}H dimension.

This result is supported by the quantitative analysis of the titration. The experimental data are first fitted as described in the Methods chapter. Comparing to VU-WT and VU-T, the residues located in VHS domain reach a plateau at a [Ub_2]/[VU-P] ratio of 2.5 approximately (see Figure 4.15). As for the UIM domain, several residues nearly reach saturation at a [K63-Ub_2]/[VU-P] ratio of 3.5.

Dissociation constant (K_d) are then calculated based on the fitted curves of the experimental data (See Methods). The VHS domain of VU-P exhibits a comparable K_d (62 μM) to that of VU-WT (53.9 μM) although similarly to VU-T, a small decrease in affinity is observed. As for UIM, a 6-fold increase in K_d is seen in VU-P (453 μM comparing to 73.9 μM in VU-WT, see Table 4.8). Thus the drop is not as drastic as seen in VU-T, as the affinity remains in the micromolar range. This result indicates a mild impact of the linker stiffness

		K63-Ub$_2$ K$_d$ (μM)
VU-WT	VHS	53.9 $\pm$ 29.2
	UIM	72.1 $\pm$ 29.1
VU-P	VHS	62.0 $\pm$ 11.3
	UIM	453.0 $\pm$ 159.8

Table 4.8 – Dissociation constants measured for the interaction between VU-WT and VU-P with K63-Ub$_2$ using a 2:1 stoechiometry model (See Methods).

on its molecular recognition capacity. In fact the interdomain motion between VHS and UIM is slightly more restricted that probably makes VU-P to have difficulties to bind its interaction partner in space. The dependency of the two domains is thereby weaken, which could have impact on the protein function, as multidomain protein generally needs a certain dependency to perform in a sequential fashion.

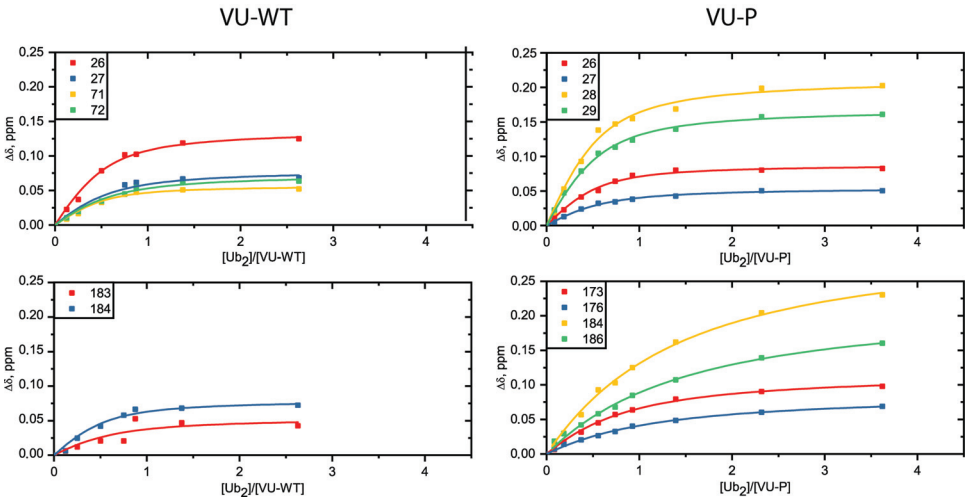


Figure 4.15 – The experimental titration curves of some representative residues for the interaction between K63-Ub$_2$ and VU-WT (left) and VU-P (right) were plotted in dot. For each residue presented, the smooth line of the same color depict the fitted curves by using the 2:1 stoechiometry model.

5

Discussion on the role of STAM2 linkers

The ESCRT machinery is responsible for various signaling processes, among which the sorting mechanism of the lysosomal degradation represents an essential pathway where K63-Ub_2 is involved as the signal instead of K48-Ub_2 which was well known as the signal of proteasomal degradation [96] [97]. STAM2, the protein of interest in this present study, is known to harbor three ubiquitin binding domains (UBDs) [49] [47] that make the protein capable to establish interactions with ubiquitin and K63 poly-ubiquitin chains. Together with HRS protein, STAM2 constitutes the ESCRT-0 complex which situates in the first chain of the lysosomal pathway mentioned in the introduction chapter. Other than ubiquitin and poly-ubiquitin chains, STAM2 can recognize additionally binding partners such as the deubiquitylase AMSH and UBPY [98] [99] [49]. In fact, the enzymes are able to cleave the particular K63 iso-peptidic bonds that links together monomers in the K63 polyubiquitin chains. Thus, this interaction is responsible for maintaining a constant pool of ubiquitin in cell

and could be considered as a recycling process. In short, the interactions established between different binding partners at different time is crucial to keep the pathway functioning properly. As mentioned before, multidomain proteins possess the advantages of having domains with different nature thus they are able to execute sequential steps of interaction and function. This advantage should have a close connection to their plasticity in space. In the present study, the N-terminus fragment of STAM2 contains 3 domains VHS, UIM and SH3 with two long disordered linkers between them, which give rise to a question whether the dynamical properties of the linker play a role in molecular recognition with various binding partners. Here the two dual domains VHS-UIM (VU) and UIM-SH3 (US) are separately studied with manipulation on the linker region in different manner: progressively reducing the linker regions starting from WT (US-Δ1, US-Δ2) or completely removing the linker (VU-T), partially and completely deleting of UIM (US-Δ3, US-Δ4) and finally proline substitution on two third of one linker length (VU-P).

On the US constructs, SAXS data show evolution of changes in R_g and D_{max} ranging from minor decrease when the linker is shortened (US-Δ1) toward drastic change when both domain and linkers were affected (US-Δ3 and US-Δ4). However, as the R_g parameter represents an average of all existing conformations of the protein in solution, it is not sufficient to conclude on the flexibility of the different US variants based on the R_g values derived from SAXS data solely. To assess the possible hidden structural dispersity, the muti-state approach MultiFoXS was performed to extract information on the conformation range obtained by the constructs of interest.

As presented in chapter 3, for US-WT, a five-state model best describes and fits the experimental data. The range of R_g values is relatively large, that goes from 15.8 to 33.8 Å. As for US-Δ1, US-Δ3 and US-Δ4 constructs, a three-state model is sufficient to describe the experimental data, as a bigger ensemble does not improve significantly χ value, indicating a reduction of plasticity of the molecule upon mutation. Moreover, looking at the three-state model of US-Δ1, the range of R_g range is slightly shorter (19 to 31 Å) than that of US-WT and all intermediate R_g are excluded, suggesting a possible restriction in conformation switching. By comparing US-WT with US-Δ1, same R_g and D_{max} value are observed even though the mutant possesses a linker with 7 amino acids shorter. It is possible that the linker of US-WT

arranges in a loop region that allows conformations with similar compactness as US-Δ1. The average Cα-Cα distance between residue Q188 and A208 for the US-WT and US-Δ1 models, gives a value of 37.7 and 28.3 Å respectively, thus a difference of 9.4 Å which makes the two distances fairly similar. However an end-to-end distance of ≈ 23.8 Å is expected for a peptide of 7 amino acids long [100]. This observation supports the previous hypothesis that US-WT forms a loop of different conformational states rather than several extended polymer chains.

Moreover, the presence of the UIM and SH3 domains may also play a dominant role in the conformational space occupied by the different US constructs due to the excluded volume. Effectively, in UIM or the SH3 domains prevent the flexible linker to adopt certain conformations [101] in the case of US-WT (Figure 5.1, left), US-Δ1 or US-Δ4 but allowed in the case of US-Δ3 (Figure 5.1, right). This excluded volumes could decrease the number of conformations accessible to a chain and could increase the stiffness of the linker as reported by recent Monte Carlo simulations [100].

NMR spin relaxation is a powerful method that allows a more specific quantitative study on the motion of different constructs. It is revealed that in either of the constructs US-WT, VU-WT and mutants, each domain tumbles with different correlation times. For the US constructs, the more drastic change is observed in US-Δ2 where the linker is shortened by 14 amino acids. Indeed, the correlation times associated with UIM and SH3 both decrease although UIM exhibits a stronger decrease due to its small size. Moreover, the gap between the R_2 level of UIM and SH3 tends to decrease as if UIM would be part of SH3. The shortened linker reduces the interdomain motion as well as restricts the global motion of both domains that leads to a decrease in their correlation time and a coupling of their tumbling. On another point of view, when the linker length increases significantly, it is reported in another study [94] that the correlation between domains decreases significantly. A similar situation occurs with identical repeating domains, where the correlation times are largely overestimated compared to a situation where each domain is considered in its individual state [102]. In short, a long linker plays important role in maintaining the independence of the two domains.

How would it affect the molecular recognition ability, thus the function of the protein? To answer this question, further investigation on the interaction with Ub and K63-Ub_2 was performed. The latter two proteins have been shown in previous works to have interaction

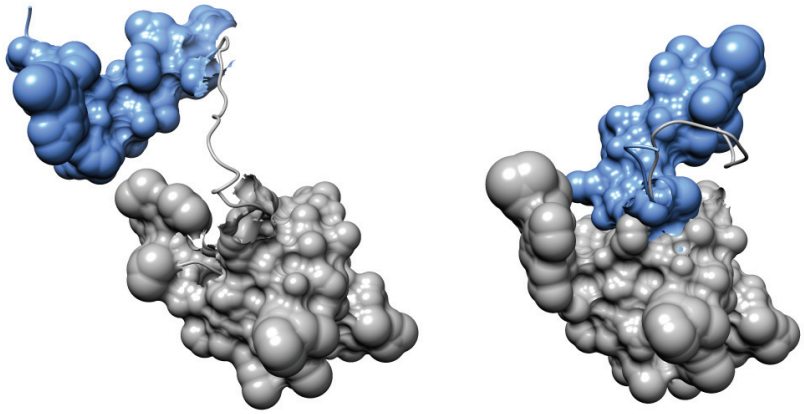


Figure 5.1 – Effect of the excluded volume on US-WT (left) and US-Δ3 (right) models. To demonstrate that the excluded volume due to UIM (blue) and SH3 (grey) restrict the conformational space available to US-WT compared to US-Δ3, we have added the UIM domain to the N-terminus of US-Δ3 by aligning the UIM C-terminus (H189-T192) on the same residues of US-Δ3. This representation clearly shows that the excluded volume due to the UIM and SH3 domains restricts the conformational space available to US-WT while the linker in the US-Δ3 model may adopt a wrap-around conformation and may result in steric clashes if the same orientation of the linker would be used for US-WT (right model). The excluded volume is represented by the solvent accessible surface area and has been calculated using the formula *residue.areaSAS* of Chimera.

with the US-WT construct [47] [49]. When the linker length is modified, the affinity of both UIM and SH3 towards Ub slightly reduces by a factor of $\approx$ 1.5 for US-Δ1 and 2.5 US-Δ4. While half of UIM is not sufficient to bind Ub, the binding affinity of SH3 in US-Δ4 and US-Δ3 toward Ub does not exhibit drastic change. When it comes to K63-Ub_2 as the binding partner, the change is much stronger. The UIM situation remains unchanged as its affinity with K63-Ub_2 keeps a fairly similar dissociation constant between US-WT, US-Δ1 and US-Δ2, see Table 3.7). As for SH3, a gradual decrease in affinity up to $\approx$ 8 folds was observed (60, 253 and 483 μM in US-WT, US-Δ1 and US-Δ2 respectively, (see Table 3.7) as the linker is progressively shortened. Comparing to Ub interaction, K63-Ub_2 brings more space constraints

to the interaction. It appears that US-WT therefore needs a refined linker length to etablish an appropriate avid binding and to avoid steric clash that would reduce US-WT association or dissociation rate with K62-Ub_2 [19]. Thus, the long flexible linker connecting those two domains must be designed to match with the interaction with K63-Ub_2 and more generally with polyubiquitin-linked chains that are the specific signal of the lysosomal degradation. Additionally, as referred in the previous chapter, the UIM domain might play a minor inhibitory effect on the SH3 domain, which was not originally known as an UBD and has only been proved to have interaction with Ub and chains in a recent work [49].

In the present study, VU-T represents another approach where the linker is completely removed. A more drastic change was observed in the dynamical properties of the UIM domain. In fact, the deletion has dragged the two domains back to back, resulting in a dependent tumbling of the N-terminus half of UIM. Indeed while the fast picosecond motion is drastically reduced in this region, slow local motion has raised to reach the VHS level as if it becomes part of the preceded domain, suggesting that the interdomain motion is nearly abolished in VU-T. This change in the dynamical characteristics indeed results in a dramatic decrease in the binding capacity to its specific ligand K63-Ub_2. The UIM domain becomes hardly accessible to its ligand, which is possibly due to the mask effect of VHS, caused by the lack of space between the two domains.

Nevertheless, it has been shown in the chapter 3 that the length of the linker plays a crucial role in keeping the independence of the domains, which is important for multidomain to be able to perform multiple interaction/function at a specific time and space in the cellular pathway. In another approach, the linker length was kept unchanged, but the amino acids composition was greatly modified with substitution by prolines. Here the linker may undergo a rearrangement of secondary structure, which increases its stiffness while reducing the dependency tumbling of the two domains VHS and UIM. This occurrence has further impact on the molecular recognition ability of the protein, where the K63-Ub_2 affinity of VU-P shows a slight decrease. Conversely, a mutant where the linker residues would be substituted by amino acids that could bring a high degree of freedom to the peptidic bond such as glycine would be interested to examine for instance. In short, for multidomain protein, the linker usually provides a synergy between amino acids composition, length, flexibility, dynamics

and the environment conditions (pH, pressure, temperature for instance) associated with allosteric or cooperative effects. Any modification of those elements of the linker without touching the functional domain would greatly affect the protein function in one way or another.

It is likely that the two linkers that tether three domains of STAM2 have been designed properly to allow multiple partner interaction during the sorting process for a rapid remodeling of the interaction map. Indeed, fast switching between K63-polyubiquitin chains and AMSH has been reported in the previous work of Hologne *et al.* [49]. Indeed, from the multi-states analysis of SAXS data, it has been shown that US-WT should exhibit in continue interconversion between a wide range of conformations, which nicely supports the switching-partner hypothesis. The relationship between flexibility, dynamics at the ps-ns time scale and binding affinity therefore is significant. The results have shown that for UIM-SH3 the length of the linker affects directly the global tumbling of both the UIM and SH3 domains and their binding capability with K63-Ub_2. Furthermore, reduced spectral densities indicate a fast local motion located in the linker and its vicinity as a lower value of J(0.87ωH) is observed comparing to the same region in US-WT. Therefore, the shortening of this linker reduces its flexibility, as well as the interdomain dependency. Its higher value of J(0) also reflects the greater contribution of slow motion, which could involve domain rotations and translations which correlates with the reduction to a three-state ensemble that best fits the SAXS experimental curves comparing to the five-state ensemble of US-WT. In brief, shortening or complete removal of the linker and alteration of the amino acids composition all lead to modification of the flexibility and dynamics of the linker region, sometime can drive to a conformational change and it is proved in the present study that those changes of the linker region have a direct impact on molecular recognition of the STAM2 protein with respect to polyubiquitin chains.

6

The helical structure of UIM

As referred in the previous chapters, the N-terminal (1-269) fragment of STAM2 is the minimum requirement for the protein to function properly. It is composed of three distinct domains VHS, UIM and SH3. VHS is a widely known domain that contains a series of α-helices which fold together to a super helix (PDB structure 1X5B) while SH3 structure is constituted of antiparallel β-sheets, several turns, loop and short helices (PDB structure 1X2Q). As for the UIM domain, it is commonly known as an α-helix with a conserved sequence pattern throughout its different versions in several proteins[46]. The study of the two segment VHS-UIM and UIM-SH3 raises a problem concerning the precise length of the UIM domain.

The conserved sequence pattern of UIM [103] was defined as followed

X - Ac - Ac - Ac - Ac - Φ - X - X - A - X - X - X - S - X - X - Ac - X - X -X -X

with Ac as acidic residues (Glutamic acid or Aspartic acid), Φ as large hydrophobic residues (typically Leucine), A as Alanine, S as Serine and X as less conserved residues. In fact, the large amount of less conserved residues of UIM makes it difficult to assess the exact location on the sequence where the α-helix starts to fold, by using solely homology analysis. Therefore a thorough study is conducted in order to define the structure of the UIM domain in STAM2 protein.

1. UIM-Y

The UIM motif in STAM2 protein is a peptide of 31 amino acids, among which no aromatic residue is present, which makes UIM invisible during a protein quantification at UV 280nm. Therefore, we have decided to design another version of UIM where the residue T190 is substituted by a tyrosine. The mutant called UIM-Y therefore facilitates the purification steps where it is visible under UV 280 nm spectrometer. The residues T190 was chosen by its location in the rear end of the domain and should not take part in the core sequence, therefore would not affect the structure. As seen in Figure 6.1, the ^{1}H,^{15}N-HSQC spectra of the original UIM and the mutant UIM-Y show perfect overlapping is observed on all resonances, except the mutated T190 which shows a big shift and the three residues Q188, H189 and E191 that are located in the vicinity. This observation suggests that the mutation T190Y does not affect the structural integrity of UIM and could be a good substitution of the UIM itself for the future study.

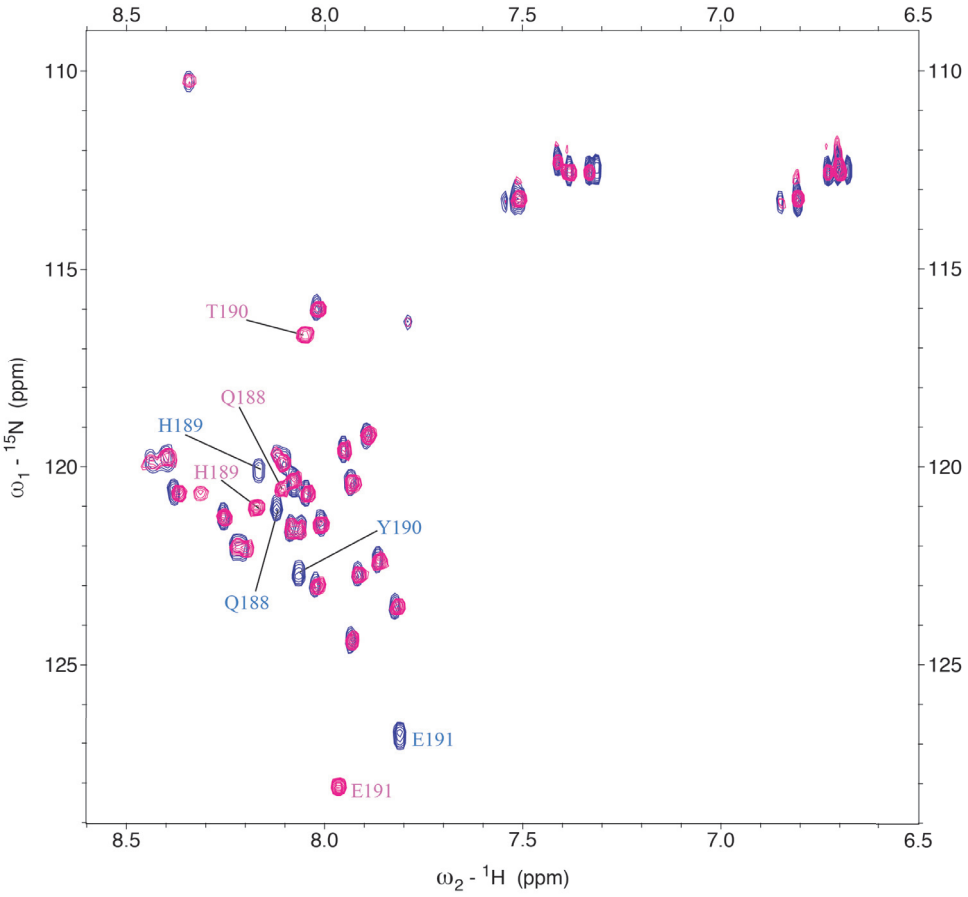


Figure 6.1 – The overlay of the UIM-Y (blue) and UIM (pink) ^{1}H-^{15}N-HSQC spectra. The residues Q188, H189, T190 or Y190 and E191 of the two spectra are shown with labels of the same color. Spectra were recorded at a ^{1}H frequency of 600 MHz and at 15 °C.

2. UIM-Y dynamics

A series of 3D NMR assignment experiments were done on the double labeled $^{15}N,^{13}C$-UIM-Y (see Methods) in order to assign the 31 resonances of UIM-Y (see Figure 6.2).

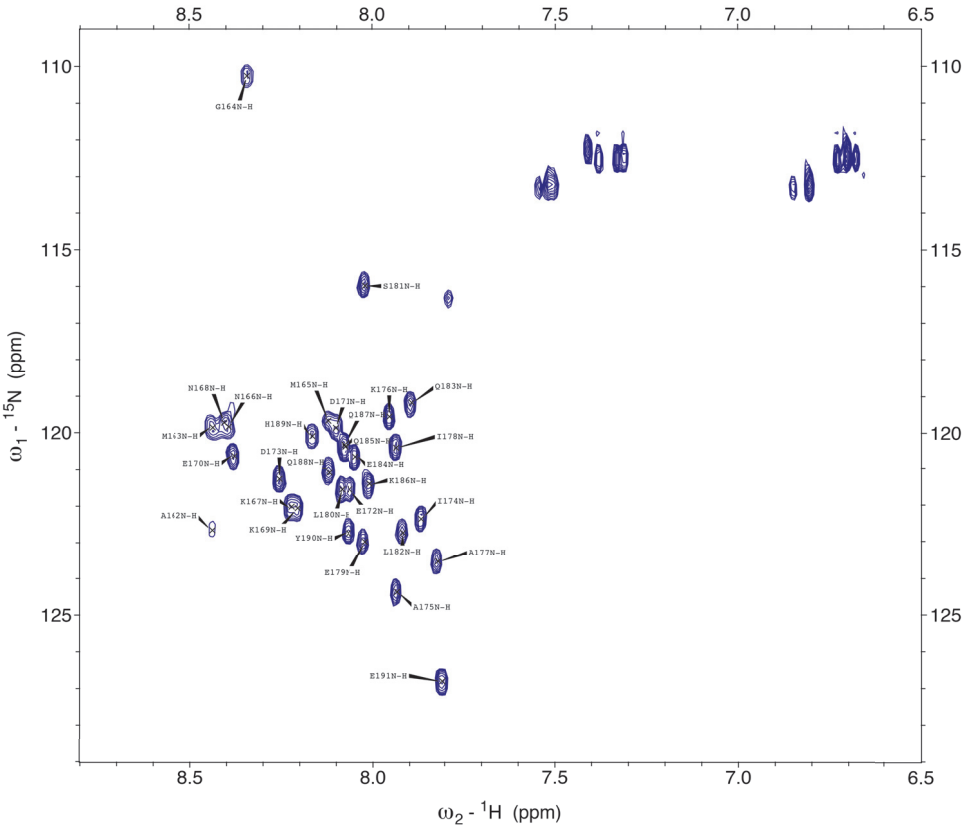


Figure 6.2 – Backbone assignment of the UIM-Y ^{1}H-^{15}N-HSQC spectra.

A first glance on the three $^{1}H,^{15}N$ spin relaxation parameters shows a similar pattern seen in either VHS-UIM or UIM-SH3 at the UIM region. On all R_1, R_2 and NOE profiles, the curves all start with a low value then raise at the region of 163-165, reach a plateau and go down again slowly from residue 185-188. The results reassure the precision of the previous studies. Indeed, the region flanking 172 and 187 are considered the total length of UIM helix when doing calculations of all parameters such as: K_d to Ub and Ub_2, average R_1, R_2, NOE

and correlation time τ_C.

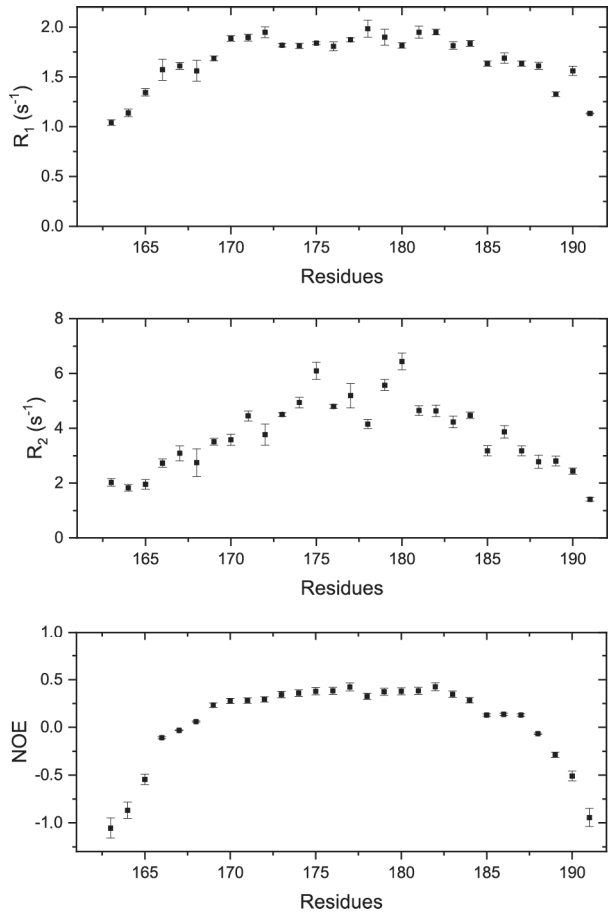


Figure 6.3 – R_1, R_2 and NOE ^{1}H-^{15}N spin relaxation rates of UIM-Y.

As shown in the Table 6.1 the values of R_1 and R_2 however change from the values found when in US-WT. Indeed, R_1 level is 1.8 s^{-1} while it was around 1.65 s^{-1} in US-WT and R_2 is 4.6 s^{-1} while it was 5.7 s^{-1} in US-WT. The correlation time also decreases in UIM-Y (3.2 ns) compared to UIM in US-WT (4.4 ns). In fact, without being linked to a larger domain such as SH3 that introduces an interdomain motion, the UIM domain relaxation rates follows the R_2 molecular ruler (Figure 3.12).

As for the NOE average, it stays stably at 0.32 which is similar to the value of 0.29 found in US-WT. However, the two extremities of UIM goes much lower in the negative values where

	R_1 (s^{-1})	R_2 (s^{-1})	NOE	τ_C (ns)
UIM-Y	1.88 ± 0.04	4.6 ± 0.2	0.32 ± 0.03	3.2 ± 0.6
UIM in US-WT	1.65 ± 0.02	5.7 ± 0.1	0.29 ± 0.01	4.4 ± 0.4

Table 6.1 – Averaged longitudinal (R_1) and transverse (R_2) relaxation rates ^{1}H-^{15}N-NOE values measured for the two constructs UIM-Y and UIM in US-WT at 15 °C and at a ^{1}H frequency of 600 MHz. The average has been obtained using the values of the residues ranging between 172 and 187. Correlation times (τc) have been calculated using the ROTDIF program [75] [76].

they reach -1.0, which was not seen when in the multidomain protein. The results thus reflect the high freedom of movement of the two ends of UIM when it is not flanked by a larger domain and also indicates a high dynamics characteristics of this α-helix.

3. Side chain assignment (Upcoming works)

In order to solve the structure of UIM, the side chain assignment is also needed. In the upcoming works, the following experiments are performed for such informations required : HN(CO)CACB, HCCH-TOCSY, ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC.

7

Conclusion

Throughout this study, the impact of the linker region on the molecular recognition of STAM2 toward ubiquitin and di-ubiquitin chain has been clarified. By modifying this disordered region in different manner, it has been shown that flexibility and dynamics have a fairly great contribution to the proper interaction between the protein and its binding partner. Indeed, for the UIM-SH3 fragments, the combination of different approaches such as SAXS, NMR and CD has brought not only a panoramic view on the wide range of conformations of the protein but also a high resolution perspective on the dynamics at ps-ns timescale. As for the N-terminus fragment VHS-UIM, NMR spin relaxation data provides unquestionably insight to the importance of the flexible linker with respect to its molecular recognition. However, several limitations remain for this part of the study. First, due to the drastic change of the linker region, the overlapping between ^{1}H-^{15}N-HSQC spectra of the mutants and the individual domain is not perfect. This problem makes the assignment of those mutants

incomplete, especially in the linker region (VU-P) or in the vicinity of the linker hinge (VU-T). Dynamics information lacking in that region could be completed with a thorough backbone assignment of the mutants using classical 3D NMR experiments. Moreover, as mentioned in the chapter 4, as a result of the mutation, conformational change might have occured, which requires further investigation. It could be a thorough study of the structure by NMR or global shape examination by SAXS or FRET.

Additionally, the K63-linked chain specific deubiquitinating enzyme AMSH has been shown to bind to the SH3 domain of STAM2 and to enhances the recycling of the EGF receptor [104] [58] [105]. In order to maintain a constant pool of ubiquitin in the cell, polyubiquitin chains require cleavage by this specific enzyme. Indeed, its catalytic activity increases when STAM2 is present, resulting from a specific structural organization that promotes the association of STAM, K63- linked polyubiquitin chains and AMSH. Therefore, it is possible that a modification of the linker flexibility would prevent the correct positioning of different components in this trimer-complex and keeps them from performing their proper function. The study of the latter must include the presence of the three domains VHS, UIM and SH3 which requires novel labelling technique such as methyl-groups labeling [106] [107].

In brief, the study demonstrates that other than the structured domains of a MDP, the disordered regions that connect them together also play an important role in the protein interaction. The presence of those linkers, as well as their flexible characteristics are essential for the protein to receive its binding partners at different time and space. As a further application, the linkers included in multidomain proteins could also be the next generation of druggable target as their modification may reduce or completely abolish interactions.

Bibliography

[1] Scott, J. D. & Pawson, T. (2009). Cell signaling in space and time: where proteins come together and when they're apart. *Science*, 326(5957), 1220–4.

[2] Altshuler, D., Daly, M. J., & Lander, E. S. (2008). Genetic mapping in human disease. *Science*, 322(5903), 881–888.

[3] Lage, K. (2014). Protein-protein interactions and genetic diseases: The interactome. *Biochim Biophys Acta*, 1842(10), 1971–1980.

[4] Vidal, M., Cusick, M. E., & Barabasi, A. L. (2011). Interactome networks and human disease. *Cell*, 144(6), 986–98.

[5] Consortium, I. H. G. S. *et al.* (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011), 931.

[6] Koonin, E. V., Wolf, Y. I., & Karev, G. P. (2002). The structure of the protein universe and genome evolution. *Nature*, 420(6912), 218–23.

[7] Lee, Y. & Rio, D. C. (2015). Mechanisms and regulation of alternative pre-mrna splicing. *Annual review of biochemistry*, 84, 291–323.

[8] Rape, M. (2018). Ubiquitylation at the crossroads of development and disease. *Nature reviews Molecular cell biology*, 19(1), 59.

[9] Verdin, E. & Ott, M. (2015). 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nature reviews Molecular cell biology*, 16(4), 258.

[10] Kim, M.-S., Pinto, S. M., Getnet, D., Nirujogi, R. S., Manda, S. S., Chaerkady, R., Madugundu, A. K., Kelkar, D. S., Isserlin, R., Jain, S., *et al.* (2014). A draft map of the human proteome. *Nature*, 509(7502), 575.

[11] Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., *et al.* (2014). Mass-spectrometry-based draft of the human proteome. *Nature*, 509(7502), 582.

[12] Apic, G., Gough, J., & Teichmann, S. A. (2001). Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *Journal of molecular biology*, 310(2), 311–325.

[13] Ekman, D., Björklund, Å. K., Frey-Skött, J., & Elofsson, A. (2005). Multi-domain proteins in the three kingdoms of life: orphan domains and other unassigned regions. *Journal of molecular biology*, 348(1), 231–243.

[14] Han, J. H., Batey, S., Nickson, A. A., Teichmann, S. A., & Clarke, J. (2007). The folding and evolution of multidomain proteins. *Nat Rev Mol Cell Biol*, 8(4), 319–30.

[15] Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., & Teichmann, S. A. (2004). Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol*, 14(2), 208–16.

[16] George, R. A. & Heringa, J. (2002). An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng*, 15(11), 871–9.

[17] Vishwanath, S., de Brevern, A. G., & Srinivasan, N. (2018). Same but not alike: Structure, flexibility and energetics of domains in multi-domain proteins are influenced by the presence of other domains. *PLoS computational biology*, 14(2), e1006008.

[18] Delaforge, E., Milles, S., Huang, J. R., Bouvier, D., Jensen, M. R., Sattler, M., Hart, D. J., & Blackledge, M. (2016). Investigating the role of large-scale domain dynamics in protein-protein interactions. *Front Mol Biosci*, 3, 54.

[19] Hellenkamp, B., Wortmann, P., Kandzia, F., Zacharias, M., & Hugel, T. (2017). Multidomain structure and correlated dynamics determined by self-consistent fret networks. *Nat Methods*, 14(2), 174–180.

[20] Bernado, P. (2010). Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering. *Eur Biophys J*, 39(5), 769–80.

[21] Sonntag, M., Jagtap, P. K. A., Simon, B., Appavou, M. S., Geerlof, A., Stehle, R., Gabel, F., Hennig, J., & Sattler, M. (2017). Segmental, domain-selective perdeuteration and small-angle neutron scattering for structural analysis of multi-domain proteins. *Angew Chem Int Ed Engl*, 56(32), 9322–9325.

[22] Gobl, C., Madl, T., Simon, B., & Sattler, M. (2014). NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog Nucl Magn Reson Spectrosc*, 80, 26–63.

[23] Zhang, Y., Wen, B., Peng, J., Zuo, X., Gong, Q., & Zhang, Z. (2015). Determining structural ensembles of flexible multi-domain proteins using small-angle X-ray scattering and molecular dynamics simulations. *Protein & cell*, 6(8), 619–623.

[24] McCullough, J., Colf, L. A., & Sundquist, W. I. (2013). Membrane fission reactions of the mammalian ESCRT pathway. *Annu Rev Biochem*, 82, 663–92.

[25] Henne, W. M., Buchkovich, N. J., & Emr, S. D. (2011). The ESCRT pathway. *Dev Cell*, 21(1), 77–91.

[26] Wegner, C. S., Rodahl, L. M., & Stenmark, H. (2011). ESCRT proteins and cell signalling. *Traffic*, 12(10), 1291–7.

[27] Sorkin, A. & von Zastrow, M. (2009). Endocytosis and signalling: intertwining molecular networks. *Nat Rev Mol Cell Biol*, 10(9), 609–22.

[28] Colombo, M., Moita, C., van Niel, G., Kowal, J., Vigneron, J., Benaroch, P., Manel, N., Moita, L. F., Théry, C., & Raposo, G. (2013). Analysis of ESCRT functions in exosome

biogenesis, composition and secretion highlights the heterogeneity of extracellular vesicles. *J Cell Sci*, 126(24), 5553–5565.

[29] Rusten, T. E. & Stenmark, H. (2009). How do ESCRT proteins control autophagy? *Journal of cell science*, 122(13), 2179–2183.

[30] Pearse, B. (1987). Clathrin and coated vesicles. *The EMBO journal*, 6(9), 2507–2512.

[31] Pishvaee, B. & Payne, G. S. (1998). Clathrin coats—threads laid bare. *Cell*, 95(4), 443–446.

[32] Glenney, J. R. & Soppet, D. (1992). Sequence and expression of caveolin, a protein component of caveolae plasma membrane domains phosphorylated on tyrosine in rous sarcoma virus-transformed fibroblasts. *Proceedings of the National Academy of Sciences*, 89(21), 10517–10521.

[33] Kurzchalia, T. V., Dupree, P., Parton, R. G., Kellner, R., Virta, H., Lehnert, M., & Simons, K. (1992). Vip21, a 21-kd membrane protein is an integral component of trans-golgi-network-derived transport vesicles. *The Journal of cell biology*, 118(5), 1003–1014.

[34] Mayor, S. & Pagano, R. E. (2007). Pathways of clathrin-independent endocytosis. *Nature reviews Molecular cell biology*, 8(8), 603.

[35] Stoorvogel, W., Strous, G. J., Geuze, H. J., Oorschot, V., & Schwartzt, A. L. (1991). Late endosomes derive from early endosomes by maturation. *Cell*, 65(3), 417–427.

[36] Hsu, V. W., Bai, M., & Li, J. (2012). Getting active: protein sorting in endocytic recycling. *Nat Rev Mol Cell Biol*, 13(5), 323–8.

[37] Maxfield, F. R. & McGraw, T. E. (2004). Endocytic recycling. *Nature reviews Molecular cell biology*, 5(2), 121.

[38] Bonifacino, J. S. & Traub, L. M. (2003). Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annual review of biochemistry*, 72(1), 395–447.

[39] Hicke, L. & Dunn, R. (2003). Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. *Annual review of cell and developmental biology*, 19(1), 141–172.

[40] Shih, S. C., Sloper-Mould, K. E., & Hicke, L. (2000). Monoubiquitin carries a novel internalization signal that is appended to activated receptors. *The EMBO journal*, 19(2), 187–198.

[41] Hurley, J. H. (2008). ESCRT complexes and the biogenesis of multivesicular bodies. *Curr Opin Cell Biol*, 20(1), 4–11.

[42] Hurley, J. H. (2010). The ESCRT complexes. *Crit Rev Biochem Mol Biol*, 45(6), 463–87.

[43] Raiborg, C., Bache, K. G., Gillooly, D. J., Madshus, I. H., Stang, E., & Stenmark, H. (2002). Hrs sorts ubiquitinated proteins into clathrin-coated microdomains of early endosomes. *Nat Cell Biol*, 4(5), 394–8.

[44] Raiborg, C. & Stenmark, H. (2009). The ESCRT machinery in endosomal sorting of ubiquitylated membrane proteins. *Nature*, 458(7237), 445–52.

[45] Mayers, J. R., Fyfe, I., Schuh, A. L., Chapman, E. R., Edwardson, J. M., & Audhya, A. (2011). ESCRT-0 assembles as a heterotetrameric complex on membranes and binds multiple ubiquitinylated cargoes simultaneously. *J Biol Chem*, 286(11), 9636–45.

[46] Young, P., Deveraux, Q., Beal, R. E., Pickart, C. M., & Rechsteiner, M. (1998). Characterization of two polyubiquitin binding sites in the 26 S protease subunit 5a. *Journal of Biological Chemistry*, 273(10), 5461–5467.

[47] Lange, A., Castaneda, C., Hoeller, D., Lancelin, J. M., Fushman, D., & Walker, O. (2012a). Evidence for cooperative and domain-specific binding of the signal transducing adaptor molecule 2 (STAM2) to Lys63-linked diubiquitin. *J Biol Chem*, 287(22), 18687–99.

[48] Lange, A., Ismail, M. B., Riviere, G., Hologne, M., Lacabanne, D., Guilliere, F., Lancelin,

J. M., Krimm, I., & Walker, O. (2012b). Competitive binding of UBPY and ubiquitin to the STAM2 SH3 domain revealed by NMR. *FEBS Lett*, 586(19), 3379–84.

[49] Hologne, M., Cantrelle, F. X., Riviere, G., Guilliere, F., Trivelli, X., & Walker, O. (2016). NMR reveals the interplay among the AMSH SH3 binding motif, STAM2, and Lys63-linked diubiquitin. *J Mol Biol*, 428(22), 4544–4558.

[50] Lange, A., Hoeller, D., Wienk, H., Marcillat, O., Lancelin, J. M., & Walker, O. (2011). NMR reveals a different mode of binding of the stam2 VHS domain to ubiquitin and diubiquitin. *Biochemistry*, 50(1), 48–62.

[51] Pickart, C. M. & Fushman, D. (2004). Polyubiquitin chains: polymeric protein signals. *Curr Opin Chem Biol*, 8(6), 610–6.

[52] Kinyamu, H., Chen, J., & Archer, T. (2005). Linking the ubiquitin–proteasome pathway to chromatin remodeling/modification by nuclear receptors. *Journal of molecular endocrinology*, 34(2), 281–297.

[53] Baumeister, W., Walz, J., Zühl, F., & Seemüller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. *Cell*, 92(3), 367–380.

[54] Thrower, J. S., Hoffman, L., Rechsteiner, M., & Pickart, C. M. (2000). Recognition of the polyubiquitin proteolytic signal. *The EMBO journal*, 19(1), 94–102.

[55] Hofmann, R. M. & Pickart, C. M. (1999). Noncanonical MMS2-encoded ubiquitin-conjugating enzyme functions in assembly of novel polyubiquitin chains for DNA repair. *Cell*, 96(5), 645–53.

[56] Hofmann, K. (2009). Ubiquitin-binding domains and their role in the DNA damage response. *DNA repair*, 8(4), 544–556.

[57] Duncan, L. M., Piper, S., Dodd, R. B., Saville, M. K., Sanderson, C. M., Luzio, J. P., & Lehner, P. J. (2006). Lysine-63-linked ubiquitination is required for endolysosomal degradation of class i molecules. *The EMBO Journal*, 25, 1635–1645.

[58] Huang, F., Zeng, X., Kim, W., Balasubramani, M., Fortian, A., Gygi, S. P., Yates, N. A., & Sorkin, A. (2013). Lysine 63-linked polyubiquitination is required for EGF receptor degradation. *Proc Natl Acad Sci U S A*, 110(39), 15722–7.

[59] Swanson, K. A., Kang, R. S., Stamenova, S. D., Hicke, L., & Radhakrishnan, I. (2003). Solution structure of vps27 UIM–ubiquitin complex important for endosomal sorting and receptor downregulation. *The EMBO journal*, 22(18), 4597–4606.

[60] Šali, A. & Blundell, T. L. (1993). Comparative protein modelling by satisfaction of spatial restraints. *Journal of molecular biology*, 234(3), 779–815.

[61] Joosten, R. P., Te Beek, T. A., Krieger, E., Hekkelman, M. L., Hooft, R. W., Schneider, R., Sander, C., & Vriend, G. (2010). A series of pdb related databases for everyday needs. *Nucleic acids research*, 39(suppl_1), D411–D419.

[62] Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers: Original Research on Biomolecules*, 22(12), 2577–2637.

[63] Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., Bairoch, A., *et al.* (2005). Protein identification and analysis tools on the expasy server. In *The proteomics protocols handbook* (pp. 571–607). Springer.

[64] Pickart, C. M. & Raasi, S. (2005). Controlled synthesis of polyubiquitin chains. *Methods Enzymol*, 399, 21–36.

[65] Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. (1995). NMRpipe: a multidimensional spectral processing system based on unix pipes. *Journal of biomolecular NMR*, 6(3), 277–293.

[66] Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., & Laue, E. D. (2005). The ccpn data model for NMR spectroscopy: development of a software pipeline. *Proteins: Structure, Function, and*

Bioinformatics, 59(4), 687–696.

[67] Williamson, M. P. (2013). Using chemical shift perturbation to characterise ligand binding. *Prog Nucl Magn Reson Spectrosc*, 73, 1–16.

[68] Varadan, R., Assfalg, M., Haririnia, A., Raasi, S., Pickart, C., & Fushman, D. (2004). Solution conformation of Lys63-linked di-ubiquitin chain provides clues to functional diversity of polyubiquitin signaling. *J Biol Chem*, 279(8), 7055–63.

[69] Palmer III, A. G. (2001). NMR probes of molecular dynamics: overview and comparison with other techniques. *Annual review of biophysics and biomolecular structure*, 30(1), 129–155.

[70] Jarymowycz, V. A. & Stone, M. J. (2006). Fast time scale dynamics of protein backbones: Nmr relaxation methods, applications, and functional consequences. *Chem Rev*, 106(5), 1624–71.

[71] Doucleff, M., Hatcher-Skeers, M., & Crane, N. J. (2011). *Pocket guide to biomolecular NMR*. Springer Science & Business Media.

[72] Fushman, D., Cahill, S., & Cowburn, D. (1997). The main-chain dynamics of the dynamin pleckstrin homology (ph) domain in solution: analysis of 15N relaxation with monomer/dimer equilibration. *J Mol Biol*, 266(1), 173–94.

[73] Kay, L. E., Torchia, D. A., & Bax, A. (1989). Backbone dynamics of proteins as studied by nitrogen-15 inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, 28(23), 8972–8979.

[74] Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., & Kay, L. E. (1995). Spectral density function mapping using 15N relaxation data exclusively. *J Biomol NMR*, 6(2), 153–62.

[75] Walker, O., Varadan, R., & Fushman, D. (2004). Efficient and accurate determination of the overall rotational diffusion tensor of a molecule from 15N relaxation data using computer program rotdif. *Journal of Magnetic Resonance*, 168(2), 336–345.

[76] Berlin, K., Longhini, A., Dayie, T. K., & Fushman, D. (2013). Deriving quantitative dynamics information for proteins and rnas using rotdif with a graphical user interface. *Journal of biomolecular NMR*, 57(4), 333–352.

[77] Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L., & McSweeney, S. (2013). Upgraded esrf bm29 beamline for SAXS on macromolecules in solution. *J Synchrotron Radiat*, 20(Pt 4), 660–4.

[78] Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H., & Svergun, D. I. (2003). Primus: a windows pc-based system for small-angle scattering data analysis. *Journal of applied crystallography*, 36(5), 1277–1282.

[79] Förster, S., Apostol, L., & Bras, W. (2010). Scatter: software for the analysis of nano-and mesoscale small-angle scattering. *Journal of Applied Crystallography*, 43(3), 639–646.

[80] Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., & Sali, A. (2016). FoXS, FoXSdock and multiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic acids research*, 44(W1), W424–W429.

[81] Johnson, W. C. (1999). Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins: Structure, Function, and Bioinformatics*, 35(3), 307–312.

[82] Abdul-Gader, A., Miles, A. J., & Wallace, B. A. (2011). A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. *Bioinformatics*, 27(12), 1630–1636.

[83] Compton, L. A. & Johnson Jr, W. C. (1986). Analysis of protein circular dichroism

spectra for secondary structure using a simple matrix multiplication. *Analytical biochemistry*, 155(1), 155–167.

[84] Manavalan, P. & Johnson Jr, W. C. (1987). Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Analytical biochemistry*, 167(1), 76–85.

[85] Sreerama, N. & Woody, R. W. (2000). Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical biochemistry*, 287(2), 252–260.

[86] Lees, J. G., Miles, A. J., Wien, F., & Wallace, B. (2006). A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics*, 22(16), 1955–1962.

[87] Shen, Y., Delaglio, F., Cornilescu, G., & Bax, A. (2009). Talos+: a hybrid method for predicting protein backbone torsion angles from nmr chemical shifts. *Journal of biomolecular NMR*, 44(4), 213–223.

[88] Feigin, L., Svergun, D. I., *et al.* (1987). *Structure analysis by small-angle X-ray and neutron scattering*, volume 1. Springer.

[89] Putnam, C. D., Hammel, M., Hura, G. L., & Tainer, J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Quarterly reviews of biophysics*, 40(3), 191–285.

[90] Receveur-Brechot, V. & Durand, D. (2012). How random are intrinsically disordered proteins? a small angle scattering perspective. *Curr Protein Pept Sci*, 13(1), 55–75.

[91] Durand, D., Vives, C., Cannella, D., Perez, J., Pebay-Peyroula, E., Vachette, P., & Fieschi, F. (2010). NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J Struct Biol*, 169(1), 45–53.

[92] Schneidman-Duhovny, D., Hammel, M., & Sali, A. (2010). FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic acids research*, 38(suppl 2), W540–W544.

[93] Cavanagh, J., Faribrother, W. J., Palmer, A. G., Rance, M., & Skelton, N. J. (2007). Chapter 5 - relaxation and dynamic processes. In J. Cavanagh, W. J. Faribrother, A. G. Palmer, M. Rance, & N. J. Skelton (Eds.), *Protein NMR Spectroscopy (Second Edition)* (pp. 333 – 404). Burlington: Academic Press, second edition edition.

[94] Bae, S.-H., Dyson, H. J., & Wright, P. E. (2009). Prediction of the rotational tumbling time for proteins with disordered segments. *Journal of the American Chemical Society*, 131(19), 6814–6821.

[95] Khan, S. N., Charlier, C., Augustyniak, R., Salvi, N., Dejean, V., Bodenhausen, G., Lequin, O., Pelupessy, P., & Ferrage, F. (2015). Distribution of pico- and nanosecond motions in disordered proteins from nuclear spin relaxation. *Biophys J*, 109(5), 988–99.

[96] Erpapazoglou, Z., Walker, O., & Haguenauer-Tsapis, R. (2014). Versatile roles of k63-linked ubiquitin chains in trafficking. *Cells*, 3(4), 1027–88.

[97] Lauwers, E., Jacob, C., & Andre, B. (2009). K63-linked ubiquitin chains as a specific signal for protein sorting into the multivesicular body pathway. *J Cell Biol*, 185(3), 493–502.

[98] Baiady, N., Padala, P., Mashahreh, B., Cohen-Kfir, E., Todd, E. A., Du Pont, K. E., Berndsen, C. E., & Wiener, R. (2016). 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(4), 2033–42.

[99] Davies, C. W., Paul, L. N., & Das, C. (2013). Mechanism of recruitment and activation

of the endosome-associated deubiquitinase AMSH. *Biochemistry*, 52(44), 7818–29.

[100] Goldenberg, D. P. (2003). Computational simulation of the statistical properties of unfolded proteins. *Journal of molecular biology*, 326(5), 1615–1633.

[101] Segall, D. E., Nelson, P. C., & Phillips, R. (2006). Volume-exclusion effects in tethered-particle experiments: bead size matters. *Physical review letters*, 96(8), 088306.

[102] Walsh, J. D., Meier, K., Ishima, R., & Gronenborn, A. M. (2010). NMR studies on domain diffusion and alignment in modular GB1 repeats. *Biophys J*, 99(8), 2636–46.

[103] Hofmann, K. & Falquet, L. (2001). A ubiquitin-interacting motif conserved in components of the proteasomal and lysosomal protein degradation systems. *Trends in biochemical sciences*, 26(6), 347–350.

[104] McCullough, J., Clague, M. J., & Urbe, S. (2004). AMSH is an endosome-associated ubiquitin isopeptidase. *J Cell Biol*, 166(4), 487–92.

[105] Clague, M. J., Coulson, J. M., & Urbe, S. (2012). Cellular functions of the DUBs. *J Cell Sci*, 125(Pt 2), 277–86.

[106] Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C., & Kay, L. E. (1999). A robust and cost-effective method for the production of val, leu, ile (δ1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *Journal of biomolecular NMR*, 13(4), 369–374.

[107] Ruschak, A. M. & Kay, L. E. (2010). Methyl groups as probes of supra-molecular structure, dynamics and function. *Journal of biomolecular NMR*, 46(1), 75.