

Mass spectrometry insights into a tandem ubiquitin-binding domain hybrid engineered for the selective recognition of unanchored polyubiquitin

Daniel Scott^{1,2}, Tom P Garner³, Jed Long^{2,4}, Jo Strachan⁵, Sharad C Mistry⁶, Andrew R Bottrill⁶, David J Tooth¹, Mark S Searle^{2,4}, Neil J Oldham², Rob Layfield^{1,*}

¹School of Life Sciences, Queen's Medical Centre, University of Nottingham, Nottingham NG7 2UH, UK

²School of Chemistry, University Park, University of Nottingham, Nottingham NG7 2RD, UK

³Department of Biochemistry, Albert Einstein College of Medicine of Yeshiva University, Bronx, NY 10461, USA

⁴Centre for Biomolecular Sciences, University Park, University of Nottingham, Nottingham NG7 2RD, UK

⁵Institute of Cell Biology, University of Edinburgh, Edinburgh EH9 3BF, UK

⁶Protein & Nucleic Acid Chemistry Laboratory, University of Leicester, Leicester, LE1 9HN, UK

*To whom correspondence should be addressed:

Dr. Robert Layfield

Tel: +44 115 8230107

robert.layfield@nottingham.ac.uk

Abbreviations: DUB: deubiquitinating enzyme, UBA: Ubiquitin associated domain, UBD: ubiquitin-binding domain, t-UBD: ubiquitin-binding domain hybrid, TUBEs: tandem ubiquitin-binding entities

Keywords: Native mass spectrometry; Protein engineering; Ubiquitination; Ubiquitin-binding domain; Unanchored polyubiquitin

Total number of words: 5813

Abstract

Unanchored polyubiquitin chains are emerging as important regulators of cellular physiology with diverse roles paralleling those of substrate-conjugated polyubiquitin. However tools able to discriminate unanchored polyubiquitin chains of different isopeptide linkages have not been described. We describe the design of a linker-optimised ubiquitin-binding domain hybrid (t-UBD) containing two UBDs, a ZnF-UBP domain in tandem with a linkage-selective UBA domain, which exploits avidity effects to afford selective recognition of unanchored Lys48-linked polyubiquitin chains. Utilising native MS to quantitatively probe binding affinities we confirm cooperative binding of the UBDs within the synthetic protein, and desired binding specificity for Lys48-linked ubiquitin dimers. Furthermore MS/MS analyses indicate that the t-UBD, when applied as an affinity enrichment reagent, can be used to favour the purification of endogenous unanchored Lys48-linked polyubiquitin chains from mammalian cell extracts. Our study indicates that strategies for the rational design and engineering of polyubiquitin chain-selective binding in non-biological polymers are possible, paving the way for the generation of reagents to probe unanchored polyubiquitin chains of different linkages and more broadly the 'ubiquitome'.

Statement of significance

Importance and functional significance of the research (max 200 words)

Unanchored (substrate-free) polyubiquitin chains are naturally-occurring covalently-linked polymers of the ubiquitin protein which are emerging as important regulators of ubiquitin-mediated processes. However little is known about their composition/structure, assembly mechanisms and precise functions in part due to the inability of current molecular tools to distinguish unanchored polyubiquitin chains from the more widely characterised substrate-anchored forms. We reasoned it should be possible to design and synthesize a non-biological protein containing tandem ubiquitin-binding domains (natural ubiquitin recognition elements) which would exploit avidity effects to afford the selective recognition and purification of specific forms of unanchored polyubiquitin, Lys48-linked chains. We describe use of a computational design approach based on combination of existing molecular structures, the application of quantitative native mass spectrometry to define binding affinities, as well as a demonstration using MS/MS that the synthetic protein exhibits desired binding specificity *in vitro* and in cell extracts. This is the first exploration of the utility of this type of selective artificial tandem ubiquitin-binding domain hybrid and indicates the rational design and engineering of more general polyubiquitin chain-selective binding in non-biological polymers is viable. Such reagents are essential for future mechanistic studies of ubiquitin-mediated signalling *in vivo*.

1 Introduction

Ubiquitination, the covalent post-translational modification of target proteins with ubiquitin, is central to the regulation of a diverse array of biological processes [1]. Signalling versatility arises from the ability of ubiquitin to form eight structurally unique polyubiquitin chain types, which vary according to linkage (isopeptide and peptide) and chain length, on selected (usually) Lys residues within the substrate sequence [2]. However, more recently physiological roles for unanchored (substrate-free) polyubiquitin chains have also begun to emerge, for example with respect to regulation of protein kinase activity [3], innate immune signalling [4, 5], and the aggresome response [6, 7]. In each case non-covalent binding of unanchored polyubiquitin chains to specific effector proteins underlies the respective biological process targeted. As for ubiquitin signalling mediated *via* conjugated polyubiquitin, unanchored chains with different linkages translate in to different signalling outcomes. However, to date molecular tools to study this unanchored fraction of the ‘ubiquitome’ are limited.

Non-covalent recognition of (poly)ubiquitin is afforded by a ‘tool box’ of ubiquitin-binding domains (UBDs) found within ubiquitin-binding proteins, including small helical structural motifs and zinc-finger domains [8], which recognise distinct surface patches on ubiquitin. The three-helix bundle ubiquitin-associated (UBA) domain typically recognises a hydrophobic patch centred on Ile44 of ubiquitin, while the zinc-finger ZnF A20 domain binds specifically to the polar Asp58 face of ubiquitin [9]. The ZnF-UBP domain specifically targets the free C-terminus (Gly75/Gly76) of ubiquitin [10], thus distinguishing between conjugated polyubiquitin chains and those that are substrate-free. A limited number of isolated UBDs can decode polyubiquitin signals in a linkage-selective manner. The archetypal example is the UBA2 domain of hHR23A, a proteasome shuttle protein which recognises Lys48-linked chains by simultaneous binding to Ile44 sites on distal and proximal ubiquitins in a diubiquitin unit, *via* two binding faces on the same UBA [11]. More recently, natural UBDs selective for Lys29/Lys33-linked diubiquitin have also been identified [12]. Alternatively, linkage selectivity of ubiquitin-binding proteins can be achieved by placing multiple non-specific UBDs in tandem, with the linker region between them positioning the domains for efficient avid binding across different ubiquitins, which in the case of Rap80 achieves selective recognition of Lys63-linked polyubiquitin chains [13]. The deubiquitinating enzyme USP5 is a good example of a natural multi-UBD protein, with binding by a ZnF-UBP domain, catalytic site and two UBA domains promoting the selective recognition of unanchored polyubiquitin chains, over monoubiquitin, albeit independent of chain linkage [14].

Insights into the biological significance of different polyubiquitin assemblies, and the ubiquitin-mediated processes they regulate, have in part been afforded by exploiting the inherent specificities of UBDs as *in vitro* tools. We previously used the isolated ZnF-UBP domain of USP5 to allow the affinity purification of unanchored polyubiquitin from a range of different species [15, 16], whilst alternative UBDs have been widely utilised to purify or detect ubiquitin-modified proteins *in vitro* [17], and also to probe ubiquitin modifications in living cells [18], often with linkage selectivity. Furthermore artificial protein sequences, termed Tandem Ubiquitin-Binding Entities (TUBEs), representing multiple repeats of the same (and very recently different [19]) UBD sequence have been developed as tools to probe polyubiquitin chains [20]. As for their natural counterparts, avidity effects account for the high affinity binding of TUBEs to polyubiquitin, although these polymers do not typically exhibit linkage selectivity [17]. In contrast, Sims and co-workers [13] have extended the Rap80 model and described the design of a Lys63 polyubiquitin chain-selective agent as a sensor for localisation and linkage-dependence of ubiquitin signalling by using multiple tandem natural ubiquitin interacting motifs (UIMs). Thus UBDs are in principle amenable to synthetic biology approaches involving assembly of tandem UBDs with engineered recognition properties.

Indeed a future challenge is the development of bio-inspired synthetic reagents for the selective detection and purification of polyubiquitin chains of defined chain linkage/length, and which distinguish unanchored from conjugated polyubiquitin. Towards this goal we demonstrate the rational design and validation using different MS approaches, of a tandem UBD (t-UBD) hybrid protein containing a ZnF-UBP domain joined to a Lys48-polyubiquitin-selective UBA domain, which exploits avidity effects to allow the recognition and purification of unanchored Lys48-linked polyubiquitin chains. This study serves as a first proof-of-concept that polyubiquitin-selective recognition can be achieved using novel assemblies of differing UBDs in tandem with optimised linkage-selective avidity effects.

2 Materials and methods

2.1 t-UBD gene construction

Molecular modelling (see supplementary information for details) was completed to determine the optimum linker length to join the UBA2 domain of hHR23A and the ZnF-UBP domain of USP16, to generate the t-UBD. Accordingly the codon optimised DNA coding sequence for the UBA2 of hHR23A (residues 315-362), a 2xAsp4xGly linker and the ZnF-UBP domain of USP16 (residues 23-143) was assembled N to C terminally,

with *Bam*HI and *Xho*I restriction sites included at the 5' and 3' ends respectively. The synthetic DNA was commercially sourced (Dundee Cell Products) and cloned in to the corresponding *Bam*HI and *Xho*I sites of pGEX-4T-1 (GE Healthcare). Point mutations (ZnF-UBP-Arg84Ala and insertions/deletions at the linker of Gly residues) to the t-UBD gene were introduced by site-directed mutagenesis (QuikChange kit, Stratagene, Agilent, Stockport, UK), construct integrity was confirmed by DNA sequencing.

2.2 Protein expression and purification

The t-UBD (see supplementary information for detail), ZnF-UBP domain of USP5 (as per [15]) and monoubiquitin (WT and a mutant lacking the C-terminal Gly75 and Gly76 (Δ GG), as per [21]) were expressed, purified and desalted into 25 mM ammonium acetate (pH 7) *in house*. Desalted diubiquitin (Lys48 and Lys63 linked, Boston Biochem) was commercially sourced.

2.3 Native mass spectrometry

Native ESI experiments were performed on a Waters (Altrincham, UK) Synapt G1 High Definition Mass Spectrometer (HDMS) - a hybrid quadrupole- time of flight instrument, equipped with standard electrospray source. The instrument was operated in positive-ion mode, over the scanning range of m/z 1000-4000. Instrument parameters were: ESI capillary voltage, 2.5 kV, sample cone voltage, 30 V, extraction cone voltage, 5 V, trap and transfer collision voltage, 8 and 5 V respectively. Pressures maintained throughout the instrument were; backing pressure, 4-4.2 mBar, trap pressure, 2.5×10^{-2} mbar and TOF region pressure, 1.8×10^{-6} mbar. Waters MassLynx 4.1 software was used to operate the instrument and subsequently process data.

Using ESI-MS, the binding affinity of the t-UBD for ubiquitin substrates was determined by titrating the ubiquitin substrate (1 μ M) against the t-UBD at varying concentrations (2, 4, 6, and 8 μ M) (three replicates at each titre value). Samples were prepared as 50:50 (v/v) solution mixtures in 25 mM ammonium acetate (pH 7) and directly infused into the mass spectrometer using a 100 μ L Hamilton syringe (Bonaduz, Switzerland) and syringe pump (Harvard 22 dual syringe pump, model 55-2222 Holliston, MA, USA). Minimum smoothing and background subtraction was applied to the obtained spectra prior to analysis. Binding affinities were determined by comparing the ratios of signal intensities attributed to unbound (poly)ubiquitin and bound t-UBD•(poly)ubiquitin ions (see [22] for details on K_d determination).

2.4 Pull-down assays

Purified t-UBD or ZnF-UBP was immobilized to cyanogen bromide-activated Sepharose 4B (GE Healthcare) at equimolar concentration (t-UBD – 1.33 mg/ml; ZnF-UBP - 1 mg/ml) as previously described [15], wherein bait proves the limiting factor in the capture of unanchored polyubiquitin chains from cell extracts. t-UBD or ZnF-UBP Sepharose was then used to capture either commercial diubiquitin or endogenous unanchored polyubiquitin from HEK293T cell lysates, as previously described [15] (see supplementary information for detail). Sepharose captures were then either western blotted, submitted to a deubiquitination assay, or digested with trypsin and analysed by LC-MS/MS.

2.5 Western blotting

50 μ L of t-UBD or ZnF-UBP Sepharose captures were eluted into gel loading buffer, separated by 5-20% polyacrylamide SDS PAGE and immunoblotted against ubiquitin (VU-1, Life Sensors, PA, USA) or the Lys48 polyubiquitin linkage (Apu-2, Millipore, Watford, UK), according to the manufacturer's guidelines. Blots were detected by ECL (Western lightning plus, Perkin Elmer).

2.6 Deubiquitination assay

50 μ L t-UBD or ZnF-UBP Sepharose captures were incubated overnight at 37°C, with 50 μ L of full-length human USP5 (10 ng/ μ L, ENZO Life Sciences, Exeter, UK) or 50 μ L of the catalytic core of the USP2 (25 ng/ μ L, ENZO Life Sciences, Exeter, UK), in DUB buffer (50 mM Tris (pH 7.5), 1 mM DTT). The reaction was quenched with the addition of gel loading buffer, and immunoblotted against ubiquitin (VU-1).

2.7 Protein digestion, analysis by tandem mass spectrometry and data processing

Proteins affinity purified on t-UBD or ZnF-UBP Sepharose were eluted into 0.1% formic acid (pH 2), and digested according to a modified FASP strategy [23]. Briefly, eluate was buffer exchanged (Amicon ultra, 0.5 ml, 10 kDa MWCO) into digestion buffer (2 M urea, 0.1 M ammonium bicarbonate (pH 8), 5% (v/v) acetonitrile), reduced (50 mM DTT), alkylated (100 mM chloroacetamide, Sigma, Dorset, UK) and digested with trypsin (sequencing grade modified porcine trypsin, Promega, Southampton, UK). After overnight digestion, peptides were extracted (70% acetonitrile, 0.1% formic acid) and vacuum centrifuged to dryness. LC-MS/MS was carried out using an RSLC nano HPLC system (Dionex, UK) and an LTQ-Orbitrap-Velos mass spectrometer (Thermo Scientific) (see supporting information for detail on sample loading and analysis).

3 Results

3.1 Designing and purifying the t-UBD

We previously showed that the ZnF-UBP domain from the deubiquitinating enzyme USP5 recognises, with relatively high affinity, the free C-terminus of unanchored ubiquitin ($K_d \approx 2.3 \mu\text{M}$ vs monoubiquitin [22]), and can be used to affinity purify endogenous unanchored polyubiquitin chains from a range of biological sources [15, 16]. Importantly the specificity of the ZnF-UBP domain excludes the purification of conjugated polyubiquitin and associated covalently modified target proteins. However subsequent quantitative electrospray ionisation mass spectrometry (ESI-MS) interaction studies confirmed that this UBD exhibits no intrinsic selectivity for unanchored ubiquitin dimers or tetramers of different isopeptide linkages [22]. A further caveat to the use of the isolated ZnF-UBP domain in the purification of unanchored polyubiquitin is that it also fails to discriminate between free ubiquitin monomers and chains, the former of which predominate *in vivo*. Thus, purifications can be dominated by abundant monoubiquitin which can impact downstream analyses [15].

Here we describe the design of a novel synthetic t-UBD which for the first time utilises two different UBDs in tandem to exploit avidity effects that afford the selective recognition and purification of unanchored polyubiquitin chains of defined linkage. We reasoned that by utilising the ZnF-UBP domain (residues 23-143) of the deubiquitinating enzyme USP16 ($K_d \approx 6.5 \mu\text{M}$ vs monoubiquitin [24]) in place of the ZnF-UBP domain of USP5 ($K_d \approx 2.8 \mu\text{M}$ vs monoubiquitin) we would further disfavour recognition of monoubiquitin. Thus our construct included the ZnF-UBP domain (residues 23-143) of USP16 in tandem with the Lys48-selective UBA2 domain of hHR23A (residues 315-362) [11]. Accordingly the ZnF-UBP domain of USP16 (pdb code 2I50) was modelled onto the existing USP5 ZnF-UBP:monoubiquitin complex (pdb code 2G45). This model guided docking of the ZnF-UBP domain on to the protruding free C-terminus of the proximal ubiquitin of the UBA2:Lys48-diubiquitin complex (pdb code 1ZO6). Molecular dynamic (MD) simulations were used (see supplementary information) to generate a sterically and energetically plausible model of the complex, which indicated a requirement for a linker length of at least 17 Å (equivalent to six residues) between the final structured C-terminal Phe360 residue of UBA2 and first structured N-terminal Val23 residue of the ZnF-UBP. We inserted a linker which accommodated the two native C-terminal aspartates from the UBA2 sequence (see supporting information Figure 1 for complete primary sequence). Thus, the final arrangement of the t-UBD placed the UBA2 domain N-terminal to the ZnF-UBP domain separated by a six amino acid (Asp₂Gly₄) flexible

linker (Figure 1). A plasmid for the expression of the t-UBD in *E. coli* was generated from the appropriate DNA sequence and used to produce the synthetic protein with an N-terminal GST affinity tag. Expressed protein was purified using glutathione-Sepharose, with t-UBD released using thrombin (with a residual N-terminal GlySer, from the thrombin cleavage site). Following purification, the construct was assessed by SDS-PAGE and mass spectrometry (supplementary Figure 2 and 3).

3.2 Probing t-UBD specificity by native mass spectrometry

To determine if the desired avidity effects were realised in the synthetic t-UBD protein, we performed native ESI-MS interaction studies in combination with protein mutagenesis. We used non-denaturing conditions to maintain non-covalent UBD:(poly)ubiquitin complexes in the gas phase [22], permitting binding affinities to be estimated from titration analyses. We have previously shown that this methodology gives affinities in excellent agreement with data derived from other biophysical techniques but with increased speed and reduced sample consumption [22]. Using ESI-MS we found that the t-UBD readily formed complexes with unanchored Lys48-diubiquitin, binding with a $K_d \approx 1 \mu\text{M}$, compared to Lys63-diubiquitin with $K_d \approx 9 \mu\text{M}$ (Figure 2, left and Table I) and monoubiquitin with $K_d \approx 11 \mu\text{M}$ (Table I). Consistent with our ESI-MS data, protein pull-down assays using t-UBD Sepharose confirmed preferential capture of unanchored Lys48-diubiquitin compared to Lys63-diubiquitin (Figure 2, right); in comparison, under the conditions used, the higher affinity (than USP16) isolated ZnF-UBP domain from USP5 was able to capture Lys63-diubiquitin ($K_d = 2.8 \pm 0.3 \mu\text{M}$).

Since the isolated UBA2 domain has a previously reported K_d vs Lys48-diubiquitin of $\approx 18 \mu\text{M}$ and vs monoubiquitin of $\approx 200 \mu\text{M}$ [22], and the ZnF-UBP (USP16) a K_d vs monoubiquitin of $\approx 6.5 \mu\text{M}$ [24], we rationalised the higher affinity ($\approx 1 \mu\text{M}$) interaction of the t-UBD with unanchored Lys48-diubiquitin results from both UBDs simultaneously recognising both ubiquitin units in the Lys48-linked dimer. Further, binding of the t-UBD to Lys63-diubiquitin with $K_d \approx 9 \mu\text{M}$ was entirely consistent with recognition of just the proximal ubiquitin by the ZnF-UBP and the functionality of that UBD. To test the contribution of both UBDs to recognition of unanchored Lys48-diubiquitin we mutated a critical Arg84 residue in the ZnF-UBP sequence of the t-UBD and assessed binding of the mutant. In this case an apparent K_d of $\approx 23 \mu\text{M}$ vs Lys48-diubiquitin (Table I), comparable to that we reported for the UBA2 alone ($K_d \approx 18 \mu\text{M}$), is consistent with abrogating the ZnF-UBP interaction with only the UBA2 contributing to binding. The data suggests that tandem UBDs are contributing to an avidity effect which underlies high affinity recognition of Lys48-diubiquitin. Likewise, use of

a monoubiquitin mutant lacking the C-terminal Gly75 and Gly76, critical for recognition by the ZnF-UBP domain [24], produced an apparent K_d with the wild-type t-UBD of $\approx 198 \mu\text{M}$ (Table I), consistent with only weak binding at the UBA2. Thus, we can conclude that our engineered avidity effect contributes to a ≈ 11 -fold preference of the t-UBD for unanchored Lys48-diubiquitin over monoubiquitin and that the t-UBD exhibits selectivity for Lys48-linked dimers (at least 7-fold over Lys63-linked dimers). It is noteworthy that selectivity of comparable magnitude is reported for the naturally occurring tandem UIMs of Rap80, which show an ≈ 24 -fold selectivity for Lys63-diubiquitin over monoubiquitin [13].

3.3 Optimising the t-UBD linker length

We next assessed if our rational design strategy had arrived at the optimum flexible linker length connecting the UBA2 and ZnF-UBP domains for the recognition of unanchored Lys48-diubiquitin. Previous studies have shown that in the natural tandem UIM construct Rap80, the linker length and its structural integrity significantly modulates the affinity for Lys63-linked polyubiquitin chains through avidity effects, by controlling UIM orientation and the relative phasing of binding sites along the helix [13]. Accordingly, we used site-directed mutagenesis to determine the effects on Lys48-diubiquitin recognition by shortening or lengthening the six residue flexible linker in the t-UBD. t-UBD sequences with 4-8 residue linkers were produced and again native ESI-MS was used to estimate binding affinities *vs* unanchored Lys48-diubiquitin (Figure 3). We found that the addition or removal of a single Gly residue from the linker increased the apparent K_d values, thus diminishing the avidity effect. The addition or removal of two residues produced apparent K_d values of $\approx 10 \mu\text{M}$, close to those reported for the isolated ZnF-UBP *vs* monoubiquitin ($K_d \approx 6.5 \mu\text{M}$ [24]) suggesting a major contribution of only this domain to the binding event. We speculate that weaker binding with shorter and longer linkers presumably reflect clashes between the two UBDs, linker and diubiquitin (shortening) and higher entropic costs associated with ordering the linker upon binding (lengthening).

3.4 Purification of endogenous unanchored polyubiquitin chains by the t-UBD

We went on to assess the binding properties of the t-UBD with respect to the purification of endogenous unanchored polyubiquitin chains from mammalian cells. Western blot analyses confirmed that in comparison to the isolated ZnF-UBP domain from USP5, under bait-limiting conditions, the t-UBD displayed preferential capture of longer unanchored polyubiquitin chains from HEK293T cell extracts (Figure 4, grey triangle), as detected with both anti-ubiquitin (Figure 4A) or Lys48-polyubiquitin specific antibodies (Figure 4B).

Importantly analysis of the blots revealed a clear advantage of the t-UBD over the original isolated ZnF-UBP domain, with almost no capture of free monoubiquitin associated with the former, reflecting the lower binding affinity of the ZnF-UBP domain of USP16 (Figure 4A). The unanchored nature of the purified chains was confirmed by treatment of samples captured on beads with the broad spectrum deubiquitinating enzyme USP2 (catalytic core, Figure 4D) and also full-length USP5 (Figure 4C), which is highly selective for substrate-free polyubiquitin [10]; in both cases chains were completely disassembled to ubiquitin monomers. There was no obvious preference of the t-UBD over the isolated ZnF-UBP domain for endogenous Lys48-diubiquitin although this may represent the relatively high abundance of this linkage in endogenous dimers [15]. The preferential binding to longer chains of the t-UBD over the isolated ZnF-UBP domain supports previous ESI-MS data in which the UBA2 domain displays a higher affinity for tetraubiquitin *vs* diubiquitin (8 μ M *vs* 18 μ M) [22]. We speculate that as with tandem UIMs [13], the compact nature of long Lys48-linked chains may lead to increased avidity effects owing to simultaneous interaction of non-adjacent ubiquitins with the t-UBD.

To confirm the identity of polyubiquitin linkages in the purified endogenous unanchored chain mixtures, we performed LC-MS/MS analyses of samples captured in parallel from HEK293T cells using the t-UBD compared to the isolated ZnF-UBP from USP5. From the ZnF-UBP captured samples we detected signature peptides, indicating a significant abundance of all seven isopeptide linkages of unanchored polyubiquitin chains (Lys6, Lys11, Lys27, Lys29, Lys33, Lys48, Lys63; Figure 5 – two biological replicas presented). The peptide sequences identified and representative MS/MS spectra can be found in supplementary Figures 4 and 5. Detection of Lys6, Lys27, Lys29 and Lys33 linkages represents the first direct demonstration that the endogenous unanchored polyubiquitin pool includes these chain types. Although we do not provide a quantitative description of relative abundance of signature peptides (and hence polyubiquitin linkage) in the endogenous unanchored polyubiquitin pool, the ~four-fold increase in the spectral count of Lys48 signature peptides in the samples produced using the t-UBD (Figure 5) strongly indicates linkage selectivity is maintained in these *in vitro* pull-down.

4 Discussion

Unanchored polyubiquitin chains are emerging as important regulators of cellular physiology, with diverse roles which parallel those of substrate-conjugated polyubiquitin. Seminal work from the Chen lab defined roles for

endogenous unanchored polyubiquitin chains with different linkages in the direct activation of some protein kinases [3]. Subsequently a variety of physiological roles have emerged [4-7, 25-27].

We previously exploited the isolated ZnF-UBP domain from USP5 to afford the first affinity purifications of endogenous unanchored polyubiquitin chains from a range of biological sources [15, 16]. Although the purification of conjugated polyubiquitin was excluded, the approach is however non-selective with respect to different polyubiquitin linkages and also is complicated by the capture of abundant free ubiquitin monomers which predominate *in vivo*. Here we show it is feasible to design synthetic reagents which exploit avidity effects to permit the recognition and purification of unanchored polyubiquitin chains of defined linkage, in this case Lys48, by placing a ZnF-UBP domain in concert with a linkage-selective UBA domain. Our design perspective built upon the structural insights that natural linkage-selectivity tandem UBD interactions have so far provided; avidity effects are a key factor in fine tuning polyubiquitin selectivity and can be used to disfavour binding to monomeric ubiquitin. Indeed utilising the lower affinity (than that of USP5) ZnF-UBP domain of USP16 in the t-UBD sequence (K_d vs monoubiquitin $\approx 6.5 \mu\text{M}$), we were able to strongly disfavour the capture of monoubiquitin in pull-downs from cell extracts.

Several lines of evidence support the notion that within the artificial t-UBD construct both UBDs are functional individually, and furthermore work cooperatively to afford higher affinity interactions with Lys48-diubiquitin. Most persuasively, ESI-MS interaction studies demonstrated binding of the t-UBD protein to unanchored Lys48-diubiquitin with an apparent K_d of $\approx 1\mu\text{M}$, which is greater than that associated with either of the individual UBDs in isolation. In contrast, binding of t-UBD to unanchored Lys63-diubiquitin was of similar affinity to the isolated ZnF-UBP:monoubiquitin interaction, suggesting in this case simple recognition at the proximal ubiquitin in the dimer by the ZnF-UBP alone. Use of mutant monoubiquitin as ligand and mutagenesis of the ZnF-UBP domain within the t-UBD provided additional confirmation the functionality of both domains.

As noted previously the magnitude of the avidity effect we generated (an ≈ 11 -fold preference of the t-UBD for unanchored Lys48-diubiquitin over monoubiquitin) may appear relatively modest, but this is of a similar magnitude to that for the naturally occurring tandem UIMs in Rap80, which display an ≈ 24 -fold preference for Lys63-diubiquitin over monoubiquitin [13]. Both our designed (t-UBD) and these natural (Rap80) UBDs do however represent the minimal tandem array, and indeed much higher affinity binding can be achieved by

placing multiple UBDs in tandem, as was achieved by Sims and co-workers using three identical UIMs in a single structured α -helix [13]. This similarly explains the high affinity (non-selective) polyubiquitin-binding properties of TUBEs (consisting of a tandem array of flexibly linked UBA domains). Thus, it should be possible to extend our design strategy to incorporate additional UBDs in order to promote further interactions which favour the recognition of longer unanchored polyubiquitin chains. Reassuringly, our modelling approach led to the optimum flexible linker length (with longer or shorter linkers associated with weaker binding), although it would be of interest to explore the effects on binding affinity of substituting rigid helical linkers to more tightly constrain the tandem UBD arrangement. Given that our MS/MS analyses using the isolated linkage-independent ZnF-UBP domain of USP5 alongside the t-UBD revealed that all of the ubiquitin isopeptide linkages participate in unanchored polyubiquitin chain formation in mammalian cells, we anticipate substituting other linkage-selective UBDs for the UBA2 to permit the recognition, detection and purification of unanchored chains of all linkage type allowing further dissection of the linkage-selective roles of unanchored polyubiquitin in regulating cell physiology.

In summary our work serves as proof-of-concept for the rational design of UBD hybrids with defined unanchored polyubiquitin selectivity, which we plan to evolve through further design iterations. We have demonstrated one application of the t-UBD, in the selective affinity purification of endogenous unanchored polyubiquitin chains, but such tools could easily be adapted beyond purification to generate probes to monitor dynamic changes in ubiquitination patterns in live cells or as transient inhibitors of ubiquitin-dependent pathways *in vivo* [17].

Acknowledgements

We wish to thank the University of Nottingham and the Birmingham-Nottingham Strategic Collaboration Fund for supporting elements of this work.

5 References

- [1] Komander, D., The emerging complexity of protein ubiquitination. *Biochemical Society Transactions* 2009, *37*, 937-953.
- [2] Komander, D., Rape, M., The ubiquitin code. *Annual review of biochemistry* 2012, *81*, 203-229.
- [3] Xia, Z. P., Sun, L. J., Chen, X., Pineda, G., *et al.*, Direct activation of protein kinases by unanchored polyubiquitin chains. *Nature* 2009, *461*, 114-119.
- [4] Zeng, W., Sun, L., Jiang, X., Chen, X., *et al.*, Reconstitution of the RIG-I pathway reveals a signaling role of unanchored polyubiquitin chains in innate immunity. *Cell* 2010, *141*, 315-330.
- [5] Pertel, T., Hausmann, S., Morger, D., Zuger, S., *et al.*, TRIM5 is an innate immune sensor for the retrovirus capsid lattice. *Nature* 2011, *472*, 361-365.
- [6] Ouyang, H., Ali, Y. O., Ravichandran, M., Dong, A., *et al.*, Protein aggregates are recruited to aggresome by histone deacetylase 6 via unanchored ubiquitin C termini. *The Journal of biological chemistry* 2012, *287*, 2317-2327.
- [7] Hao, R., Nanduri, P., Rao, Y., Panichelli, R. S., *et al.*, Proteasomes activate aggresome disassembly and clearance by producing unanchored ubiquitin chains. *Molecular cell* 2013, *51*, 819-828.
- [8] Husnjak, K., Dikic, I., Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions. *Annual review of biochemistry* 2012, *81*, 291-322.
- [9] Garner, T. P., Strachan, J., Shedden, E. C., Long, J. E., *et al.*, Independent Interactions of Ubiquitin-Binding Domains in a Ubiquitin-Mediated Ternary Complex. *Biochemistry* 2011, *50*, 9076-9087.
- [10] Reyes-Turcu, F. E., Horton, J. R., Mullally, J. E., Heroux, A., *et al.*, The ubiquitin binding domain ZnFUBP recognizes the C-terminal diglycine motif of unanchored ubiquitin. *Cell* 2006, *124*, 1197-1208.
- [11] Varadan, R., Assfalg, M., Raasi, S., Pickart, C., Fushman, D., Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain. *Molecular cell* 2005, *18*, 687-698.
- [12] Michel, M. A., Elliott, P. R., Swatek, K. N., Simicek, M., *et al.*, Assembly and specific recognition of k29- and k33-linked polyubiquitin. *Molecular cell* 2015, *58*, 95-109.
- [13] Sims, J. J., Cohen, R. E., Linkage-specific avidity defines the lysine 63-linked polyubiquitin-binding preference of rap80. *Molecular cell* 2009, *33*, 775-783.
- [14] Reyes-Turcu, F. E., Shanks, J. R., Komander, D., Wilkinson, K. D., Recognition of polyubiquitin isoforms by the multiple ubiquitin binding modules of isopeptidase T. *The Journal of biological chemistry* 2008, *283*, 19581-19592.
- [15] Strachan, J., Roach, L., Sokratous, K., Tooth, D., *et al.*, Insights into the molecular composition of endogenous unanchored polyubiquitin chains. *Journal of proteome research* 2012, *11*, 1969-1980.
- [16] Strachan, J., Shaw, B., Krishna, V., Scott, D., del Pozo, J., Broad Utility of an Affinity-enrichment Strategy for Unanchored Polyubiquitin Chains. *J Proteomics Bioinform* 2013.
- [17] Scott, D., Oldham, N. J., Strachan, J., Searle, M. S., Layfield, R., Ubiquitin-binding domains: mechanisms of ubiquitin recognition and use as tools to investigate ubiquitin-modified proteomes. *Proteomics* 2015, *15*, 844-861.
- [18] van Wijk, S. J., Fiskin, E., Putyrski, M., Pampaloni, F., *et al.*, Fluorescence-based sensors to monitor localization and functions of linear and K63-linked ubiquitin chains in cells. *Molecular cell* 2012, *47*, 797-809.
- [19] Gao, Y., Li, Y., Zhang, C., Zhao, M., *et al.*, Enhanced Purification of Ubiquitinated Proteins by Engineered Tandem Hybrid Ubiquitin-binding Domains (ThUBDs). *Molecular & cellular proteomics : MCP* 2016.
- [20] Hjerpe, R., Aillet, F., Lopitz-Otsoa, F., Lang, V., *et al.*, Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities. *EMBO reports* 2009, *10*, 1250-1258.

- [21] Long, J., Gallagher, T. R., Cavey, J. R., Sheppard, P. W., *et al.*, Ubiquitin recognition by the ubiquitin-associated domain of p62 involves a novel conformational switch. *The Journal of biological chemistry* 2008, 283, 5427-5440.
- [22] Sokratous, K., Roach, L. V., Channing, D., Strachan, J., *et al.*, Probing Affinity and Ubiquitin Linkage Selectivity of Ubiquitin-Binding Domains Using Mass Spectrometry. *Journal of the American Chemical Society* 2012, 134, 6416-6424.
- [23] Wisniewski, J. R., Zougman, A., Nagaraj, N., Mann, M., Universal sample preparation method for proteome analysis. *Nature methods* 2009, 6, 359-362.
- [24] Pai, M. T., Tzeng, S. R., Kovacs, J. J., Keaton, M. A., *et al.*, Solution structure of the Ubp-M BUZ domain, a highly specific protein module that recognizes the C-terminal tail of free ubiquitin. *Journal of molecular biology* 2007, 370, 290-302.
- [25] Nanduri, P., Hao, R., Fitzpatrick, T., Yao, T. P., Chaperone-mediated 26S Proteasome Remodeling Facilitates Free K63 Ubiquitin Chain Production and Aggresome Clearance. *The Journal of biological chemistry* 2015, 290, 9455-9464.
- [26] Jiang, X., Kinch, L. N., Brautigam, C. A., Chen, X., *et al.*, Ubiquitin-induced oligomerization of the RNA sensors RIG-I and MDA5 activates antiviral innate immune response. *Immunity* 2012, 36, 959-973.
- [27] Banerjee, I., Miyake, Y., Nobs, S. P., Schneider, C., *et al.*, Influenza A virus uses the aggresome processing machinery for host cell entry. *Science* 2014, 346, 473-477.

Figure legends

Figure 1: t-UBD structure. (Right) Model of the t-UBD hybrid protein produced by docking (guided by pdb code 2I50) the ZnF-UBP domain of USP16 (blue) on to the protruding free C-terminus of the proximal ubiquitin in a model of the hHR23A-UBA2:Lys48-diubiquitin complex (green (partly obscured), yellow; pdb code 1ZO6). The distance between C-terminal Phe360 of UBA2 and N-terminal Val6 of ZnF-UBP (both red) was estimated as 17.2 Å and indicated an absolute minimum linker length of six residues. We subsequently included a six residue flexible linker, including four glycine residues (white) along with two C-terminal aspartates donated by the UBA2 sequence. (Left) Cartoon representation of the t-UBD model.

Figure 2: Confirming the functionality of the t-UBD with unanchored ubiquitin dimers. (Left) Representative ESI-MS spectra showing the t-UBD (6 μM) in complex with Lys48 (top) or Lys63-linked (bottom) diubiquitin (1 μM). The t-UBD more readily forms complexes with unanchored Lys48-diubiquitin than with Lys63-diubiquitin, as evidenced by the stronger signal for bound complex in the Lys48-diubiquitin spectrum and for unbound t-UBD in the Lys63-diubiquitin spectrum. Titration analyses were used to estimate K_d values for the interactions (see Table I). (Right) Representative western blots (anti-ubiquitin) of protein pull-down assays with equimolar t-UBD or isolated ZnF-UBP domain from USP5 (UBP) on beads, used to capture unanchored Lys48-diubiquitin or Lys63-diubiquitin. The t-UBD captures only Lys48-diubiquitin under these conditions. The higher affinity isolated ZnF-UBP domain (USP5) ($K_d \approx 3.6 \mu\text{M}$ for Lys48-diubiquitin and $2.8\mu\text{M}$ for Lys63-diubiquitin [22]) is capable of capturing both dimers.

Figure 3: Validating the t-UBD design strategy. Effects of shortening or lengthening the six residue t-UBD linker on the recognition of unanchored Lys48-diubiquitin. Site-directed mutagenesis was used to insert or delete one or two glycine residues within the t-UBD linker, and ESI-MS applied to estimate K_d values for the interaction with Lys48-diubiquitin.

Figure 4: Application of t-UBD to the purification of endogenous unanchored polyubiquitin. Capture of endogenous unanchored Lys48-linked polyubiquitin chains using the t-UBD. Representative anti-ubiquitin western blots showing endogenous unanchored polyubiquitin chains, captured from mammalian HEK293T cell extracts, using equimolar t-UBD or isolated ZnF-UBP domain from USP5 (UBP) on beads (with protein on beads limiting). Grey circles indicate numbers of ubiquitins (#Ub) within the chains. (A) Probed with anti-

ubiquitin. (B) Probed with Lys48-linked polyubiquitin-specific antibodies. (C) Captured unanchored polyubiquitin chains as in (A) treated with full-length USP5, which catalyses complete disassembly to monoubiquitin, confirming their unanchored nature (anti-ubiquitin). (D) Captured unanchored polyubiquitin chains as in (A) treated with USP2 catalytic core, which again catalyses complete disassembly to monoubiquitin, further confirming they represent polyubiquitin (anti-ubiquitin).

Figure 5: Confirming the specificity of t-UBD *in vitro*. LC-MS/MS analyses confirm the identity of the polyubiquitin linkages in the purified endogenous unanchored polyubiquitin chains from HEK293T cells, captured in parallel using the t-UBD or linkage-independent ZnF-UBP domain (USP5) (two biological replicas presented). Purified samples on beads as in Figure 4, prepared using equimolar t-UBD or ZnF-UBP, were subject to tryptic digestion and LC-MS/MS analysis to identify polyubiquitin signature peptides. After custom database searching, only peptides corresponding to ubiquitin were identified. The data was manually interrogated for peptides modified with GlyGly (+114.042927 on Lysine) and LeuArgGlyGly (+383.228103 on Lysine). The bars represent total spectral count of each ubiquitin signature peptide during the analysis of each sample.

Table I – ESI-MS-derived dissociation constant (K_d) values for UBD: (poly)ubiquitin interactions

	monoubiquitin	Δ GlyGly- monoubiquitin	Lys48-diubiquitin	Lys63-diubiquitin
UBA2 ^a	200 ± 14	nd	18 ± 2	126 ± 18
ZnF-UBP ^b	6.5	nd	nd	nd
t-UBD	10.9 ± 0.4	197.5 ± 4.3	1.3 ± 0.3	9.9 ± 1.5
t-UBD (ZnF-UBP) mutant	nd	nd	23 ± 1.8	nd

The K_d values reported here for the t-UBD (mean ± SD, μ M) were averaged from values measured at different concentrations; nd - not determined. Values generated in this study are shaded. ^aDetermined by ESI-MS [22].

^bDetermined by ITC [24]. Note that values of the t-UBD versus Lys63-diubiquitin and monoubiquitin are consistent with only the ZnF-UBP making a major contribution to binding in these cases.

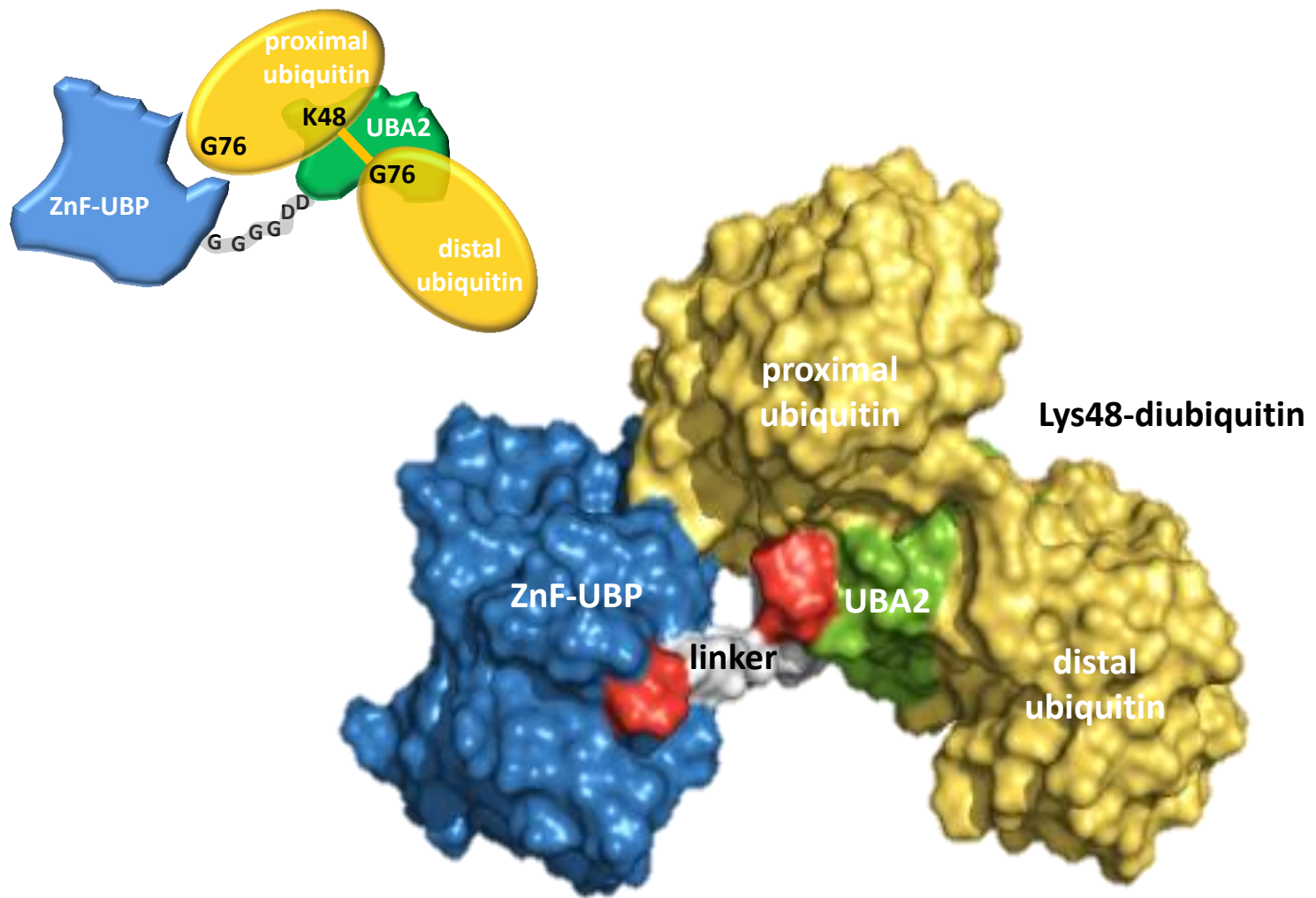


Figure 1

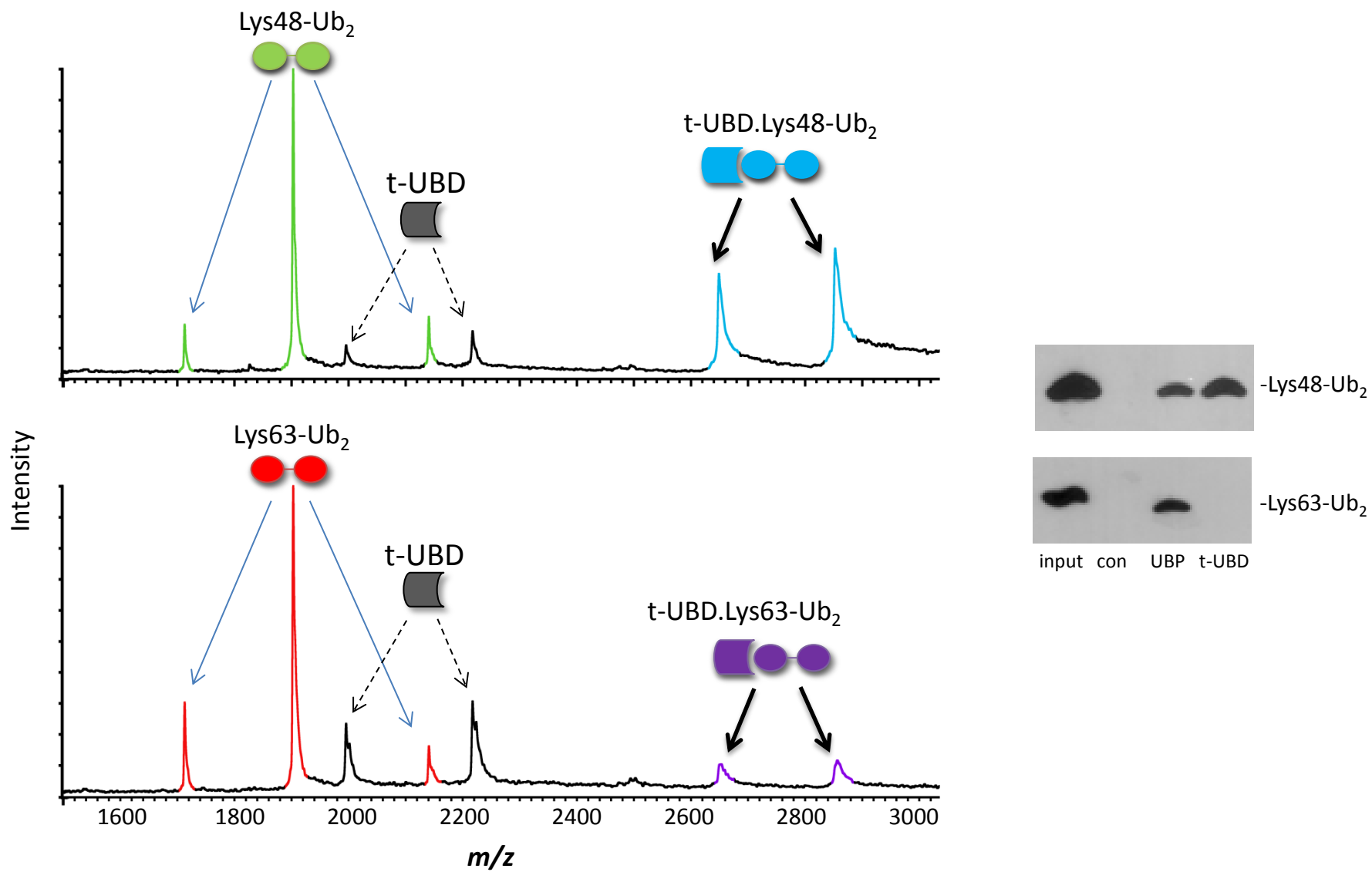


Figure 2

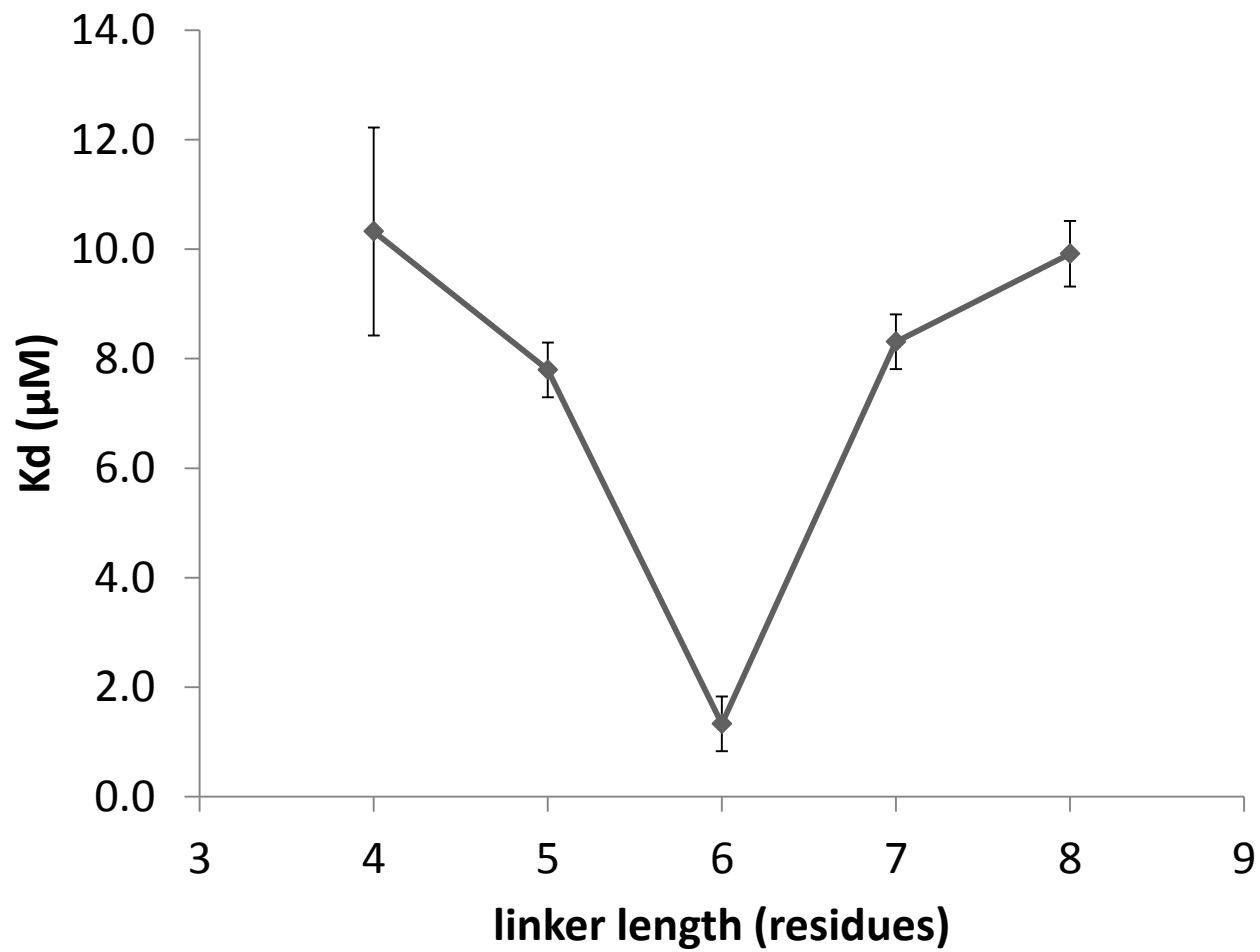


Figure 3

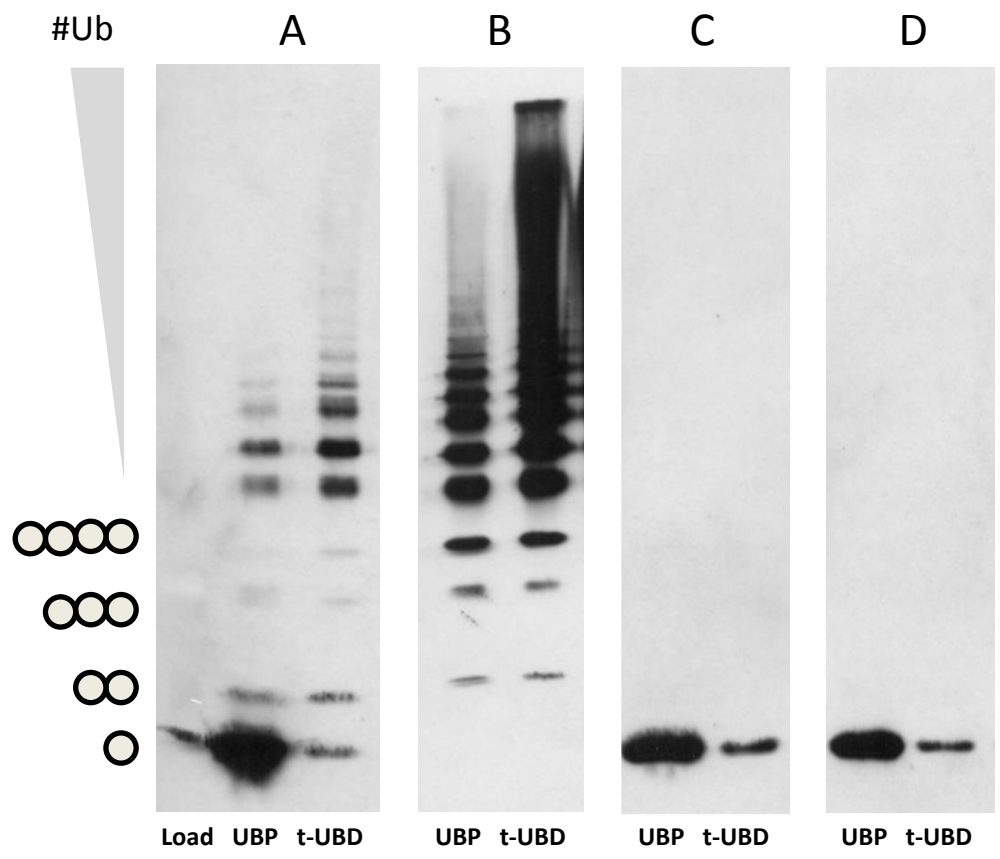


Figure 4

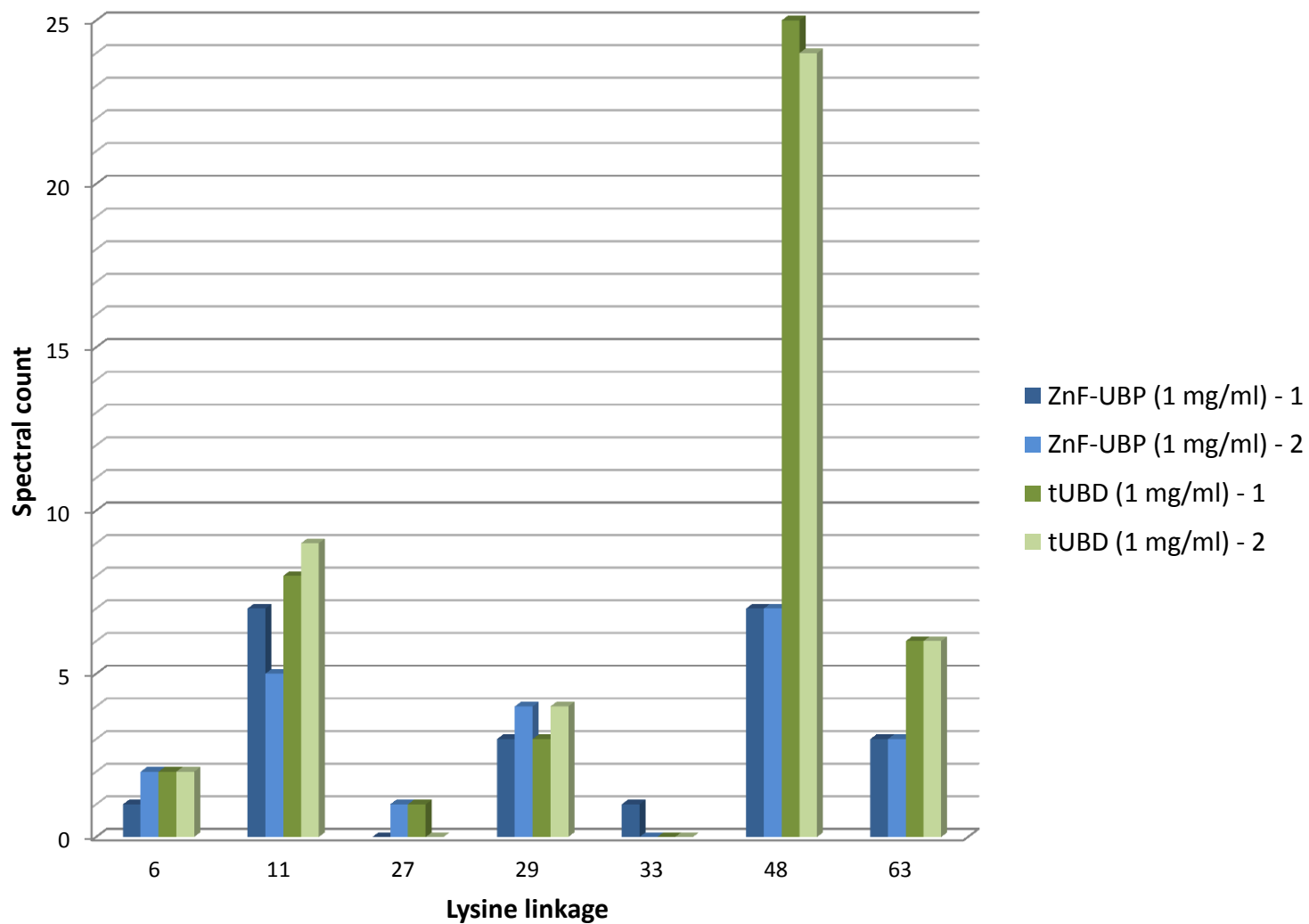


Figure 5

Supplementary information

t-UBD molecular modelling

The initial model of the ZnF-UBP domain in complex with monoubiquitin was generated using MOLMOL [1] by overlaying the NMR structure of the ZnF-UBP domain of USP16 (pdb 2I50) [2] with the related ZnF-UBP domain of USP5 in complex with monoubiquitin (pdb 2G45) [3]. The resulting model of the complex was then overlaid with the proximal ubiquitin of Lys48-linked diubiquitin in complex with the UBA2 domain of hHR23A (pdb 1ZO6) [4] to generate a ternary complex model of the two UBDs in complex with Lys48-linked diubiquitin. The ternary complex model was optimized thermodynamically *via* energy minimization using an AMBER10 force field in MOE (Molecular Operating Environment (MOE) 2014.10 Chemical Computing Group Inc., Montreal, Canada). The subsequent model was analysed to assess the optimal linkage strategy for the two UBDs based on the distance between the N and C-terminal residues of the ZnF-UBP and UBA domain respectively, and the average distance spanned by disordered loops found in the PDB database. Linkers of 5-7 residues were modelled using the general loop building procedure of the MOE Homology Modeller and minimized in MOE [5]. A 5-residue linker caused significant steric clashes while the 6 and 7 residue linkers were well tolerated.

t-UBD expression and purification

Competent BL21 cells were transformed with the GST-t-UBD expression plasmid and grown in Luria broth (LB) medium, supplemented with ampicillin, at 37°C. Cultures were incubated until $OD_{600} \approx 0.6$ and induced with 0.2 mM IPTG at 20°C for ≈ 16 hours. Bacteria were harvested by centrifugation and re-suspended in lysis buffer (10 mM Tris (pH 7.5), 150 mM NaCl, 0.1% (v/v) Triton X-100) supplemented with DNaseI (Sigma) and protease inhibitor cocktail (Sigma). The re-suspension was then lysed by sonication and clarified by centrifugation. Soluble GST-t-UBD was affinity purified on glutathione-Sepharose 4B (GE Healthcare) for 1 hour at 4°C. Following the binding step, the Sepharose was washed with thrombin cleavage buffer (20 mM Tris (pH 8.4), 150 mM NaCl, 2.5 mM $CaCl_2$) and incubated with thrombin (Sigma) for 16 hours at 4°C. Thrombin cleaved t-UBD, was further purified by anion exchange chromatography into 5 mM potassium phosphate buffer, with an increasing NaCl gradient. Fractions containing purified t-UBD were identified by SDS-PAGE and subsequent Coomassie staining, pooled and concentrated by ultrafiltration using Vivaspin 20 columns (10 kDa MWCO) (GE Healthcare) into 25 mM ammonium acetate (pH 7).

Purification of endogenous unanchored polyubiquitin

Untreated HEK293T cells (grown to $\approx 70\%$ confluence) were scraped into homogenising buffer (50 mM Tris, (pH 7.5), 150 mM NaCl, 0.5% (v/v) NP-40) supplemented with N-ethylmaleimide (NEM; 5 mM), MG-132 (20 μM), mammalian protease inhibitor cocktail and mammalian phosphatase inhibitor cocktail (both 0.1% (v/v) respectively). The resuspension was sonicated, clarified by centrifugation (13,000 rpm, 10 min, 4°C) and the supernatant passed through glass wool. The protein concentration of the lysate was determined using a Thermo Scientific Pierce BCA Protein Assay Kit, prior to quenching excess NEM (DTT, [10 mM]). The lysate was heated at 75°C for 20 minutes and centrifuged (13,000 rpm, 10 min, 4°C). The supernatant (from 15 mg unheated lysate) was incubated with 100 μl of t-UBD or ZnF-UBP Sepharose overnight, washed in homogenising buffer and then with deubiquitination (DUB) buffer (50 mM Tris (pH 7.5), 1 mM DTT).

LC-MS/MS methods

LC-MS/MS was carried out using an RSLCnano HPLC system (Dionex, UK) and an LTQ-Orbitrap-Velos mass spectrometer (Thermo Scientific). Samples were loaded at high flow rate onto a reverse-phase trap column (0.3mm i.d. x 1mm), containing 5 μm C18 300 Å Acclaim PepMap media (Dionex) maintained at a temperature of 37 °C. The loading buffer was 0.1% formic acid / 0.05% trifluoroacetic acid / 2% acetonitrile. Peptides were eluted from the trap column at a flow rate of 0.3 $\mu\text{l}/\text{min}$ and through a reverse-phase PicoFrit capillary column (75 μm i.d. x 400mm) containing Symmetry C18 100 Å media (Waters, UK) that was packed *in house* using a high pressure device (Proxeon Biosystems, Denmark). Peptides were eluted over a period of two hours, with the output of the column sprayed directly into the nanospray ion source of the LTQ-Orbitrap-Velos mass spectrometer.

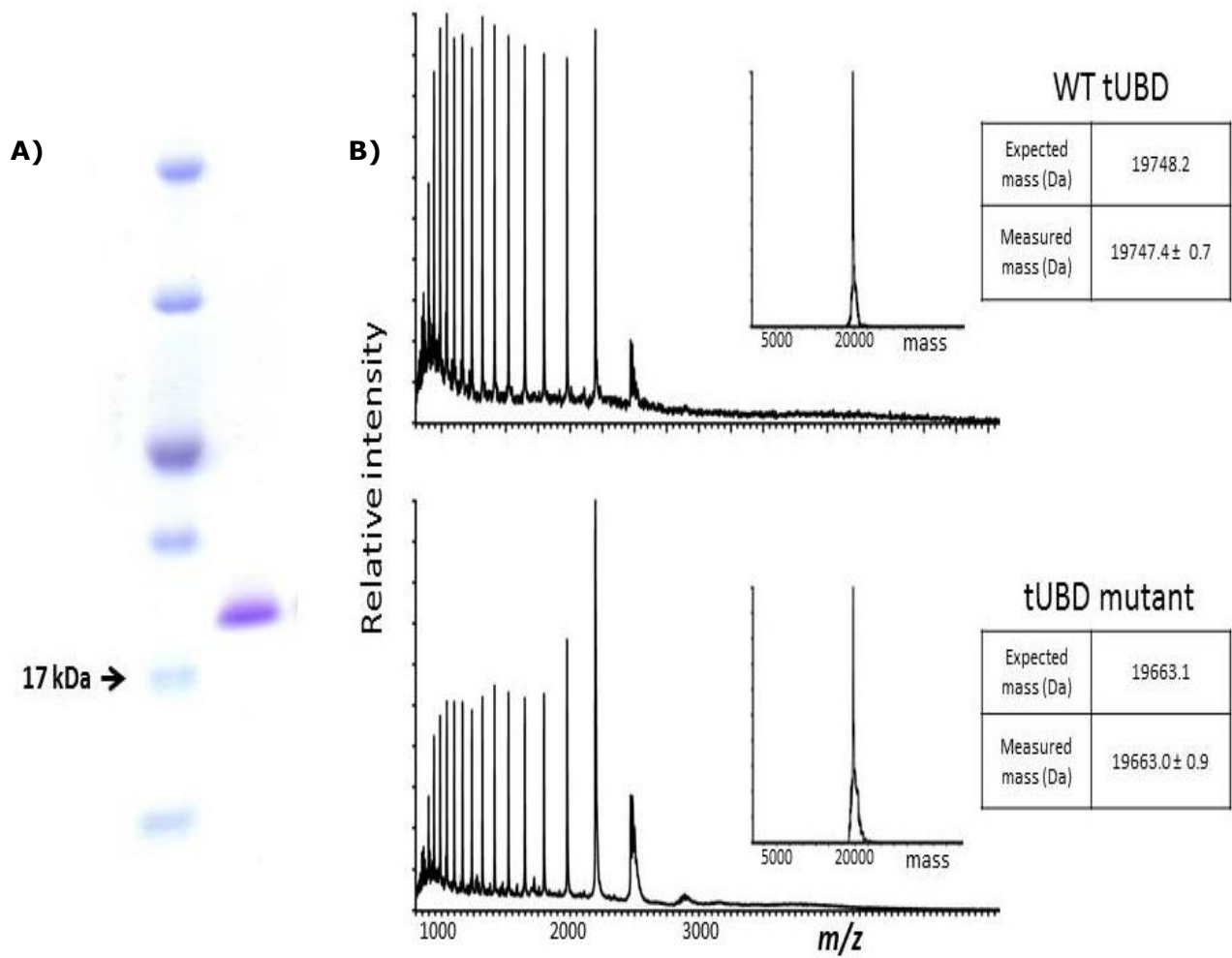
The LTQ-Orbitrap-Velos mass spectrometer was set to acquire a 1 microscan FTMS scan event at 60000 resolution over the m/z range 300-1800 Da in positive ion mode. The maximum injection time for MS was 500 ms and the AGC target setting was $1e^6$. Accurate calibration of the FTMS scan was achieved using a background ion lock mass for $\text{C}_6\text{H}_{10}\text{O}_{14}\text{S}_3$ (401.922718 Da). Subsequently up to 10 data dependent HCD MS/MS were triggered from the FTMS scan. The isolation width was 2.0 Da, normalized collision energy 42.5. Dynamic exclusion was enabled. The maximum injection time for MS/MS was 250 ms and the AGC target setting was $5e^4$.

The .raw data file obtained from each LC-MS/MS acquisition was processed using Proteome Discoverer (version 1.4.0.288, Thermo Scientific), searching each file in turn using Mascot (version 2.2.04, Matrix Science Ltd.) [6] against a custom UniprotTB_20140416 database (unknown version, 3997 entries) [7], to which the human ubiquitin sequence was added. The peptide tolerance was set to 10 ppm and the MS/MS tolerance was set to 0.02 Da. Fixed modifications were set as carbamidomethyl (+57.021464 on cysteine) and variable modifications set as oxidation (+15.994915 on methionine), GlyGly (+114.042927 on Lysine) and LeuArgGlyGly (+383.228103 on Lysine). A decoy database search was performed. The output from Proteome Discoverer was further processed using Scaffold Q+S [8] (version 4.4.1.1, Proteome Software). Upon import, the data was searched using X!Tandem [9] (The Global Proteome Machine Organization). PeptideProphet [10] and ProteinProphet [11] (Institute for Systems Biology) probability thresholds of 95% were calculated from the decoy searches and Scaffold was used to calculate an improved 95% peptide and protein probability threshold based on the data from the two different search algorithms. Peptide Probabilities from Mascot were assigned by the Scaffold Local FDR algorithm. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Proteins sharing significant peptide evidence were grouped into clusters.

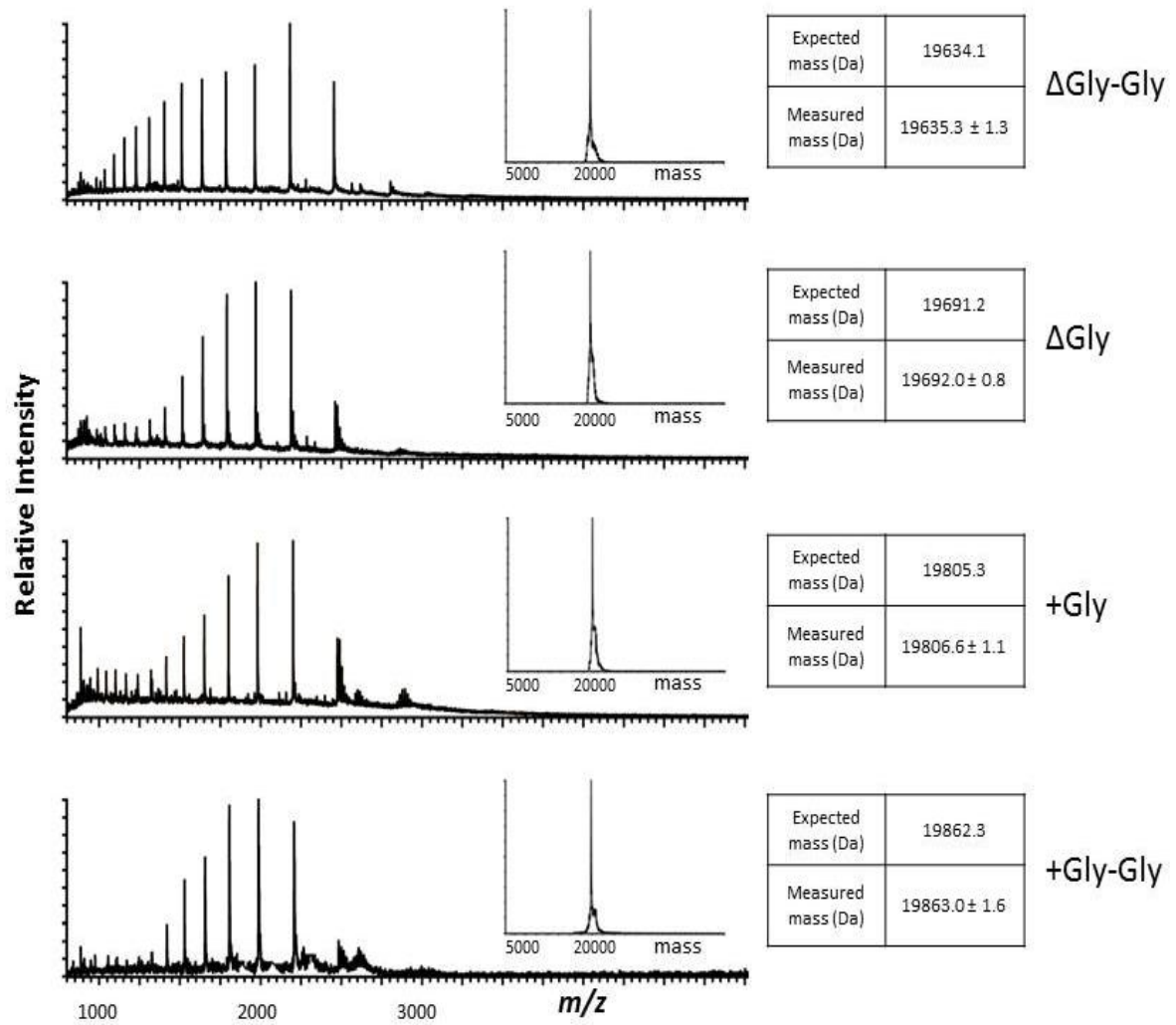
Supplementary Figure 1: Primary sequence of the t-UBD. The N-terminal Gly-Ser overhang, residual from thrombin cleavage product is indicated, followed by residues 315-362 of hHR23A UBA2 (blue), a four Gly linker (in addition to the two Asp residues C-terminal of the UBA2 domain) and residues 23-143 of the ZnF-UBP domain of USP16 (green). F and V (bold) represent the C-terminal Phe360 of UBA2 and N-terminal Val6 of ZnF-UBP, respectively. R (red) represents the Arg84 of the ZnF-UBP domain which was mutated to abrogate binding at this domain.

GSQEKEAIERLKALGFPE~~SLVIQAYFACEKNE~~
NLAANFLLSQN**F**DDGGGG**V**CRHIRKGGLEQGNL
KKALVNVEWNICQDCKTDNKVKDKAEETE~~EEK~~
PSVWLCLKCGHQGCG**R**NSQEQHALKH~~YLTPRS~~
EPHCLVLSLDNWSVWCYVCDNEVQYCSSNQLG
QVVDYVRKQAS

Supplementary Figure 2: Purification of the t-UBD reagent. A) A representative Coomassie stained SDS-PAGE gel of purified t-UBD. B) Spectra of WT and mutant t-UBD sprayed from 80 % MeCN, 0.1 % Formic acid.



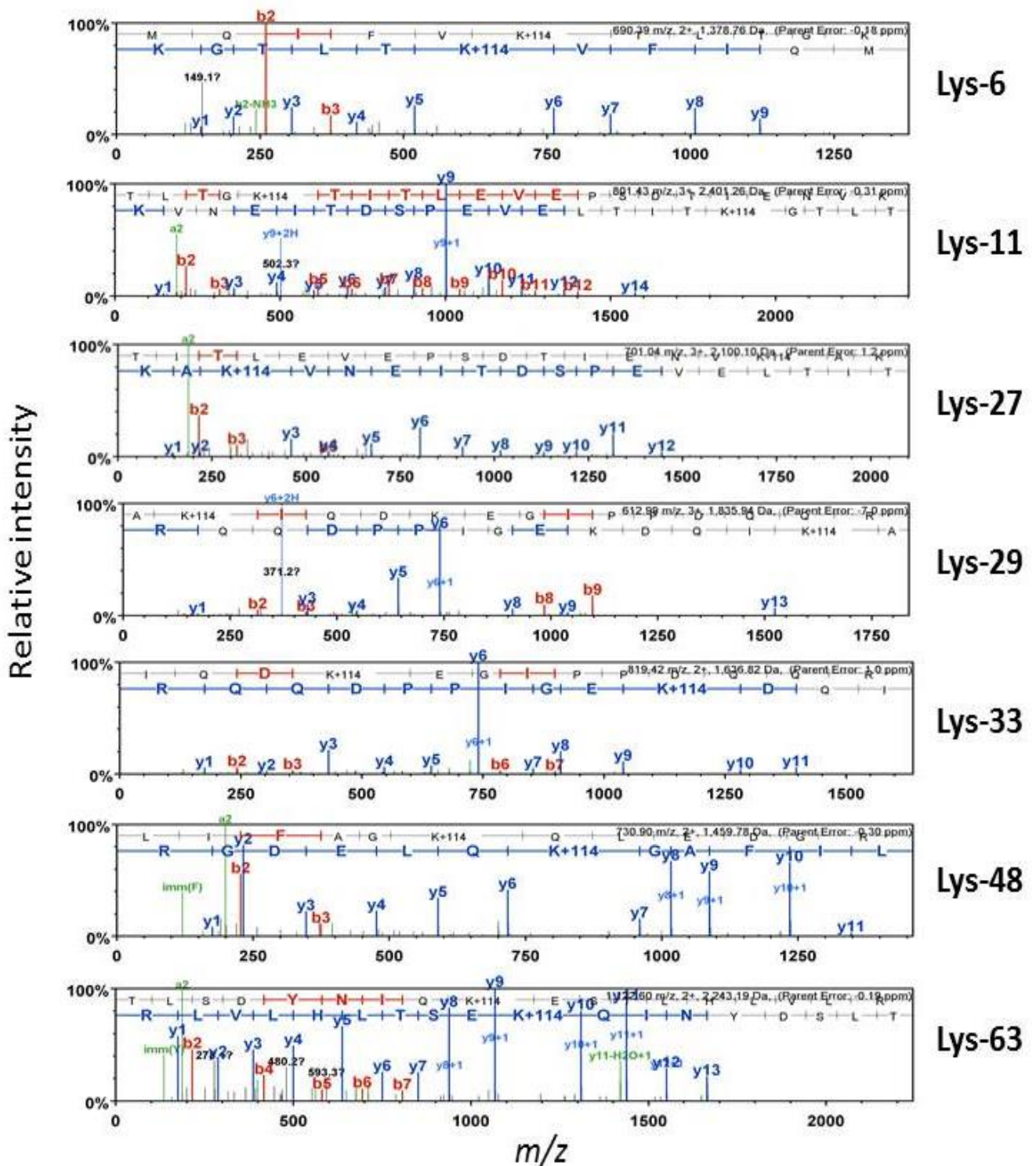
Supplementary Figure 3: Spectra of t-UBD linker mutations. Spectra of t-UBD linker mutants sprayed from 80 % MeCN, 0.1 % formic acid. The mutation is indicated to the right.



Supplementary Figure 4: The signature ubiquitin tryptic peptides identified by MS/MS analysis of purified unanchored polyubiquitin. Additionally it has been reported that a minimal population of LRGG modified peptides result from trypsin missed cleavage.

Linkage	Signature peptide
K6	MQIFVK(GG)TLTGK
K11	TLTGK(GG)TITLEVEPSDTIENVK
K27	TITLEVEPSDTIENVK(GG)AK
K29	AK(GG)IQDK
K33	IQDK(GG)EGIPPDQQR
K48	LIFAGK(GG)QLEDGR
K63	TLSDYNIQK(GG)ESTLHLVLR

Supplementary Figure 5: Representative MS/MS spectra of the purified polyubiquitin linkages. MS/MS spectra of different polyubiquitin signature peptides as indicated derived from a sample of endogenous unanchored polyubiquitin captured from HEK293T cells using the linkage-independent isolated ZnF-UBP domain of USP5.



References

1. Koradi, R., M. Billeter, and K. Wuthrich, *MOLMOL: a program for display and analysis of macromolecular structures*. J Mol Graph, 1996. **14**(1): p. 51-5, 29-32.
2. Pai, M.T., et al., *Solution structure of the Ubp-M BUZ domain, a highly specific protein module that recognizes the C-terminal tail of free ubiquitin*. J Mol Biol, 2007. **370**(2): p. 290-302.
3. Reyes-Turcu, F.E., et al., *The ubiquitin binding domain ZnF UBP recognizes the C-terminal diglycine motif of unanchored ubiquitin*. Cell, 2006. **124**(6): p. 1197-208.
4. Varadan, R., et al., *Structural determinants for selective recognition of a Lys48-linked polyubiquitin chain by a UBA domain*. Mol Cell, 2005. **18**(6): p. 687-98.
5. Labute, P., *Protonate3D: assignment of ionization states and hydrogen coordinates to macromolecular structures*. Proteins, 2009. **75**(1): p. 187-205.
6. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data*. Electrophoresis, 1999. **20**(18): p. 3551-67.
7. *The Universal Protein Resource (UniProt) in 2010*. Nucleic Acids Res, 2010. **38**(Database issue): p. D142-8.
8. Searle, B.C., *Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies*. Proteomics, 2010. **10**(6): p. 1265-9.
9. Craig, R. and R.C. Beavis, *TANDEM: matching proteins with tandem mass spectra*. Bioinformatics, 2004. **20**(9): p. 1466-7.
10. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search*. Anal Chem, 2002. **74**(20): p. 5383-92.
11. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry*. Anal Chem, 2003. **75**(17): p. 4646-58.