Structure and recognition of polyubiquitin chains of different lengths and linkage
David Fushman1 and Keith D. Wilkinson2*

Addresses: 1Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, USA; 2Department of Biochemistry, Emory University School of Medicine, Atlanta, GA 30322, USA

* Corresponding author: Keith D. Wilkinson (keith.wilkinson@emory.edu)

Abstract
The polyubiquitin signal is post-translationally attached to a large number of proteins, often directing formation of macromolecular complexes resulting in the translocation, assembly or degradation of the attached protein. Recent structural and functional studies reveal general mechanisms by which different architectures and length of the signal are distinguished.

Introduction
Post-translational modification of proteins is a widespread mechanism to fine tune the structure, function and location of proteins encoded by the surprisingly small number of genes in most organisms. Among these modifications, one of the most versatile and intriguing modifications is the covalent attachment of ubiquitin or ubiquitin-like proteins to target proteins [1-5]. In this commentary we will argue that, in contrast to modifications such as phosphorylation, methylation, acetylation, etc., the attachment of one or more ubiquitins provides a large interaction surface by which the modification can be decoded [6] and results in a vast number of potential signals by virtue of the varied architectures linking the multiple ubiquitin molecules. Ubiquitination of proteins directs the modified proteins to different cellular fates such as translocation, assembly or degradation, depending on the type of tagging. Recent studies have begun to define the roles of these different polyubiquitin signals in physiology and disease [7-11], and it is obvious that the manipulation of these signals and their recognition will be important in understanding and treating human disease. For example, defects in the pathway by which Ubiquitination tags proteins for proteolysis in the proteosome have been linked to neurodegenerative disorders such as Alzheimer’s [7] and Parkinson’s disease [8].

Diversity of polyubiquitin chains and linkages
Polyubiquitin chains consist of ubiquitin monomers linked to each other covalently by an isopeptide bond between the C-terminus of one ubiquitin and an amine from one of the seven lysines or the N-terminal methionine of the next ubiquitin. A variety of polyubiquitin chains have been detected in cellular lysates [12], varying in length and linkage (Figure 1). Moreover, the chains can also be mixed, i.e. containing other heterologous proteins from the ubiquitin family such as SUMO (small ubiquitin-like modifier) [13].

Polyubiquitin can signal in the cell as unanchored chains [14, 15], with one free C-terminus on the proximal subunit, or as a molecular tag attached through this C-terminus to a lysine in a target protein. As the latter might have several lysines, the same protein can carry several different polyubiquitin tags associated with it [16].

Assembly and disassembly of different polyubiquitin chains
Some unanchored polyubiquitin chains are synthesised by the sequential reactions of two enzymes: E1 (the ubiquitin activating enzyme) and E2 (the ubiquitin conjugating enzyme). Pioneering work from Cecile Pickart showed that E1 and E2-25K synthesize unanchored homogeneous...
Unanchored chains are shown here. These are often found attached to a target protein via an isopeptide bond between the free C-terminus of the proximal ubiquitin and an amino group on the target protein. Chains can be “linear” (no more than one amino group of each ubiquitin is linked to another ubiquitin) or “branched” (at least one ubiquitin is attached to other ubiquitins via two or more different amino groups). The ubiquitin containing the free C-terminus is the “proximal” subunit and the ubiquitin(s) lacking any amino group linkage is a “distal subunit”. The polymers are referred to as “homogeneous” if all linkages use the same amino group on each ubiquitin. They are termed “heterogeneous” if ubiquitin linkages involve different amino groups on different subunits. Chains containing other ubiquitin-like proteins are called “heterologous” or “mixed”. The type of linkage is designated by the identity of the amino group used, i.e. Met1-, Lys6-, Lys11-, Lys27-, Lys29-, Lys33-, Lys48-, Lys63-linked. Note that Met1-linked ubiquitin is identical to the proprotein product of the polyubiquitin gene and has often been referred to as linear polyubiquitin. We propose the Met1-linked nomenclature to avoid the confusion with unbranched chains described above. In the “open” conformations the hydrophobic patch of ubiquitin is exposed, while in the “closed” conformations this patch is buried by ubiquitin-ubiquitin interactions. Abbreviations: SUMO, small ubiquitin-like modifier.
of a given chain can be dynamic [23]. For instance, the promiscuous UBCH5 family of E2s combines with either of two E3 ligases, CIAP (a cellular inhibitor of apoptosis) or APC (the anaphase promoting complex), to monoubiquitinate their respective substrates [24-27]. Subsequently Lys11-linked chains are elaborated by these E3s using the E2 enzyme UBE (ubiquitin-activating enzyme) 2S. BRCA1/BARD1 autoubiquitination also shows this progression of priming (monoubiquitination) followed by elongation (polyubiquitination). Six different E2s can monoubiquitinate BRCA1, while different E2s are then used to elongate the chains with Lys63 (UBC13/MMS2), Lys48 (UBE2K), or Lys6 (UBCH5) linkages [28]. Certain pairs of E2 and E3 can also synthesize branched chains [16]. In another classic example of chain remodelling, A20 (a large enzyme with both deubiquitinating enzymes and E3 ligase activity) is thought to first trim K63-linked substrates of the NF-κB (nuclear transcription factor kappa B) pathway and then elaborate a K48-linked chain that results in the degradation of the substrate.

How does chain architecture determine the fate of polyubiquitinated proteins?
Importantly, all these polyubiquitin forms appear to serve different functions (targeting/localization, complex assembly, modulation of function or stability). The underlying structure-function hypothesis is that the chain linkage (and possibly chain length) defines the ensemble of structures/conformations that a particular chain can adopt [10, 24, 29-31]. In turn, this structure defines the chain’s ability to interact with specific receptors in a linkage specific manner [22, 32-39]. The critical question is how signalling specificity is achieved despite the vast variety of polyubiquitin chain structures/conformations.

Two features of ubiquitin chains are essential for this signalling: the hydrophobic patch on one face of the monomers [40, 41], and a highly flexible C-terminus, resulting in significant flexibility of each ubiquitin-ubiquitin linker in the chain [42]. In fact, Lys48-linked di- and tetra-ubiquitin chains under physiological conditions are in equilibrium between a predominantly “closed” conformation (where the hydrophobic patches are sequestered at the ubiquitin/ubiquitin interface) and one or more open conformations [42-44], whereas Lys63-linked chains predominantly adopt an extended structure exposing the hydrophobic patches and making them readily available for interactions with receptors [38, 45]. The emerging structural data and computer modelling indicate an even greater structural variability for chains connected through other lysines [19, 32, 46, 47]. We anticipate that chain branching might result in structures and intra-chain contacts unavailable for linear-topology chains, thus potentially further enhancing the signalling capability of polyubiquitin.

Polyubiquitin receptors
Specific recognition of polyubiquitin is accomplished by proteins containing one or more ubiquitin-binding domains. There are at least 20 families of ubiquitin-binding domains (Table 1) and many polyubiquitin-binding proteins contain multiple copies, often consisting of two or three different ubiquitin-binding domains connected by flexible linkers [48-53]. The most common of these domains, including the UBA (ubiquitin-associated) domain and UIM (ubiquitin-interacting motif), bind to a conserved hydrophobic patch on ubiquitin consisting of Leu8, Ile44, and Val70 [48-53]. The affinity of all these domains for ubiquitin is modest (µM) but tight binding is achieved by avidity; the immobilization of polyubiquitin by binding to one ubiquitin-binding domain lowers the entropic barrier for binding of an adjacent ubiquitin to another ubiquitin-binding domain, or for other protein-protein interactions between the ligand and the intact receptor [33, 48, 54].

Shuttling receptors
Moving ubiquinated proteins within the cell is facilitated by shuttling receptors [55-59]. These ubiquitin-binding proteins contain one or more ubiquitin-binding domains. One type of shuttling receptor, the UBX (ubiquitin regulatory X domain) family of p97 cofactors, contains both a ubiquitin-binding domain (of the UBA family) and a ubiquitin-like domain (of the UBX family) and acts as an adapter localizing polyubiquitinated proteins to the p97 AAA ATPase (an ATPase associated with a variety of cellular activities) [60, 61]. Subsequent delivery of the ubiquitinated proteins to the proteasome could be direct or the substrate could be handed off to other shuttling proteins [62]. Ubiquitin-like domains can act as autoinhibitory sequences by binding to the ubiquitin-binding domain, competing with ubiquitin binding [55] and/or directing the ubiquitin-binding protein to another receptor or the proteasome. Networks of ubiquitin binding domain/ubiquitin-like domain containing proteins could assemble into oligomers or a lattice that would provide a highly selective array of ubiquitin-binding sites that exhibit specificity for particular polyubiquitin chain linkages. Indeed, studies on artificial oligomeric UBA domains such as GST (glutathione S-transferase)-UBA fusions or TUBES (tandem ubiquitin-binding entities) have emphasized that the observed specificity is determined more by the oligomeric state of these domains than it is by the weak specificity inherent in the isolated domains [37, 48].

Specific recognition of chain linkage
Specificity of chain binding is a required attribute of receptors that sort proteins carrying differently linked ubiquitins to different cellular fates. For instance Lys6-, Lys11-, Lys29- and Lys48-linked substrates can be directed...
to the proteasome [9, 63], while Lys63-linked and Met1-linked chains serve non-proteolytic roles by assembling protein complexes in the DNA-damage response and NF-κB signalling pathways [15, 64, 65].

In general, it appears that proteins consisting of flexibly linked ubiquitin-binding domains show little chain specificity, in part because the polyubiquitin chain itself is very flexible [43, 44, 47]. If this conformational flexibility is restricted in a physiological complex or lattice, even these binding proteins could show modest specificity. In contrast, receptors that recognize either the isopeptide linkage or adjacent faces of ubiquitin that are juxtaposed only in certain linkages can show much higher chain linkage specificity [35, 53, 54, 66, 67] (see Figure 2).

**Importance of chain length**

Since the initial observation of Pickart’s group that K48-linked tetraubiquitin was the minimal effective length to deliver proteins to the proteasome [68], the question of how chain length affects the fate of a ubiquitinated protein has been debated. Chains must achieve a length that provides sufficient binding affinity through avidity, but long chains may be so tightly bound as to prevent dissociation of catalytic intermediates or “hand off” from one receptor to the next [69]. In part, chain length can be controlled by the degree of processiveness (repetitive rounds of catalysis without releasing the substrate) exhibited by the ubiquitin ligases [70-72] or by ubiquitin chain editing catalyzed by deubiquitinating enzymes. The modular nature of receptors containing multiple ubiquitin-binding domains and the ability of longer polyubiquitin chains to bind multiple receptors could serve to act as length sensors (Figure 3). For instance, the deubiquitinating enzyme USP5 selectively binds tetraubiquitin with sub-nanomolar affinity using an ensemble of four ubiquitin-binding sites [18]. Alternatively, a lattice consisting of ubiquitin-binding domain/ubiquitin-like domain-containing proteins that bind each other and/or compete for ubiquitin and ubiquitin-like domain binding can show a marked length dependence on the resulting complexes (as for example, Dsk2 and Rpn10 [56]). Longer polyubiquitin chains could be “handed off” from one receptor to another in such a lattice, as exemplified by the trafficking of ubiquitinated proteins through the ESCRT (endosomal sorting complex required for transport) complex [69]).

---

### Table 1. Ubiquitin binding domains (See text and [48-53] for details)

<table>
<thead>
<tr>
<th>Structure</th>
<th>Domain</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Single Helices</strong></td>
<td>UIM</td>
<td>Binds to groove on ubiquitin containing hydrophobic patch</td>
</tr>
<tr>
<td></td>
<td>DUIM</td>
<td>Double UIM</td>
</tr>
<tr>
<td></td>
<td>MIU</td>
<td>Binds to groove on ubiquitin but in opposite orientation</td>
</tr>
<tr>
<td><strong>Three helix bundles</strong></td>
<td>UBA</td>
<td>Widespread, often combined with other ubiquitin-binding domains</td>
</tr>
<tr>
<td></td>
<td>CUE</td>
<td>UBA variant involved in ERAD</td>
</tr>
<tr>
<td></td>
<td>VHS</td>
<td>Seven helices in folded domain, endosomal sorting</td>
</tr>
<tr>
<td></td>
<td>GAT</td>
<td>Found in endosomal sorting pathways, UBA+2 ZnF</td>
</tr>
<tr>
<td><strong>Zinc Fingers</strong></td>
<td>NZF</td>
<td>Found in Npl4, TAB2, and NZF2 of ESCRT II</td>
</tr>
<tr>
<td></td>
<td>ZnF UBP</td>
<td>Found in deubiquitinating enzymes, BRAP2 ligase and HDAC6</td>
</tr>
<tr>
<td></td>
<td>A20 ZnF</td>
<td>Variant of ZnF ubiquitin-specific protease domain, found in A20 deubiquitinating enzyme and Rabex 5</td>
</tr>
<tr>
<td></td>
<td>UBP</td>
<td>Found in Y-family polymerases involved in DNA repair</td>
</tr>
<tr>
<td><strong>Catalytic Domains</strong></td>
<td>UBC</td>
<td>Catalytic domain of E2 ubiquitin-conjugating enzymes</td>
</tr>
<tr>
<td></td>
<td>UEV</td>
<td>Variant UBC lacking enzymatic Cys residue, inactive</td>
</tr>
<tr>
<td></td>
<td>SENP</td>
<td>SUMO-specific protease domain</td>
</tr>
<tr>
<td></td>
<td>USP</td>
<td>Ubiquitin-specific protease deubiquitinating enzyme domain</td>
</tr>
<tr>
<td></td>
<td>UCH</td>
<td>Ubiquitin C-terminal hydrolase deubiquitinating enzyme domain</td>
</tr>
<tr>
<td></td>
<td>OTU</td>
<td>Ovarian Tumor-related deubiquitinating enzyme domain</td>
</tr>
<tr>
<td></td>
<td>Jabi1/MPN</td>
<td>Metalloprotease deubiquitinating enzymes</td>
</tr>
<tr>
<td><strong>Other Domains</strong></td>
<td>GLUE</td>
<td>GRAM-like ubiquitin-binding domain in EAP45</td>
</tr>
<tr>
<td></td>
<td>PRU</td>
<td>PH domain of Rpn13</td>
</tr>
<tr>
<td></td>
<td>SH3</td>
<td>SH3 domain in many endocytic proteins</td>
</tr>
<tr>
<td></td>
<td>UBM</td>
<td>Ubiquitin-binding domain in Y-family DNA polymerases</td>
</tr>
<tr>
<td></td>
<td>PFU</td>
<td>Ubiquitin-binding domain of PLAA family proteins</td>
</tr>
<tr>
<td></td>
<td>WD40</td>
<td>WD40 domain of UFD2/DOA1/PLAA</td>
</tr>
<tr>
<td></td>
<td>UBAR</td>
<td>Coiled Coil, binds to Met1-linked polyubiquitin</td>
</tr>
</tbody>
</table>

Abbreviations: DOA1, degradation of alpha2 1; ERAD, endoplasmic reticulum-associated protein degradation; GLUE, GRAM-like ubiquitin-binding domain in EAP45; MIU, ubiquitin interacting domain; NZF, Npl4 zinc fingers; OTU, Ovarian Tumor; PFL, PLAA family ubiquitin binding domain; PH, pleckstrin homology; PLAA, phospholipase A2 activating protein; SUMO, small ubiquitin-like modifier; UBA, ubiquitin associated; UBAR, ubiquitin binding in ABIN and NEMO proteins; UBC, ubiquitin C; UBM, ubiquitin-binding motif; UBP, ubiquitin-specific protease domain; UCH, ubiquitin C-terminal hydrolase; UEV, ubiquitin E2 variant; UFD2, ubiquitin fusion degradation 2; USP, ubiquitin-specific protease; UBP, ubiquitin-specific zinc fingers.
Localization of the polyubiquitin signal

Met1-linked chains (both unanchored and attached to NEMO – an NF-κB modifier) can activate signalling in the NF-κB pathway and directly activate some kinases [15, 29, 73, 74]. Unanchored Met1-linked ubiquitin is also the primary gene product of several ubiquitin genes and is post-translationally assembled in response to genotoxic stress and TNF-α by the E3 ligase LUBAC (linear ubiquitin chain assembly complex) [75, 76]. Levels of M1-linked ubiquitin chains in cells are normally very low [77], in part because the primary gene product is co-translationally cleaved to monomeric ubiquitin at the ribosome and in part because of the presence of large amounts of USP5 deubiquitinating enzyme, the enzyme responsible for disassembling polyubiquitin intermediates that might otherwise accumulate in the cell [18]. Thus, it is unlikely that Met1-linked chains are widely distributed in the cell, and more likely that they are locally generated at the site(s) of signalling when required. A similar argument can be made in the case of ubiquitinated proteins that accumulate during signalling. A great deal of cellular specificity in the ubiquitin pathway seems to depend on the use of adaptors and scaffolds that colocalize polyubiquitin and the enzymes that metabolize it [78]. For instance, deubiquitinating enzymes are very often found in the same protein complexes as the ubiquitin ligases that synthesize polyubiquitin [78, 79]. This suggests that if a polyubiquitin or polyubiquitinated protein is not properly channelled to its target by ubiquitin receptors it can be disassembled before it leaves the site of synthesis.

Future directions

The incredible diversity of polyubiquitin chains observed in vivo suggests that there is a similar complexity in the receptors that recognize chains. It also seems likely that additional ubiquitin-binding motifs and domains remain to be discovered. In spite of the limited specificity of isolated ubiquitin binding domains different chain architectures effectively signal different cellular fates, suggesting that a deeper understanding will be gained from studies on binding specificity in the context of the full-length receptors. As engineered and synthetic polyubiquitins become available [46, 80-87], these will aid structure determination of polyubiquitin-receptor complexes that will be vital to understand the decoding of the polyubiquitin signals. Finally, we need a more sophisticated understanding of the “hand off” of a receptor bound polyubiquitin from one receptor to the next in a pathway. For instance it is still a mystery how shuttling receptors hand off ubiquitin from the ligases and chaperone complexes to the proteasome, or how sequential ESCRT complexes can direct endocytic cargos carrying the ubiquitin signal [69]. A detailed understanding of these issues may pay dividends in treating diseases involving errors in the ubiquitination pathway.
Figure 3. Length dependence of linear chain structure and recognition

Each row represents a different length of linear ubiquitin: a) extended chains with exposed hydrophobicity such as Met1- and Lys63-linked; b) closed conformations where some of the ubiquitins contain intramolecularly buried hydrophobic patches as seen in Lys48-linked chains [67, 88, 89]; c) length selection by binding to a single receptor polypeptide (UBA [ubiquitin associated domain]2 of hHR23a and USP5 [ubiquitin-specific protease 5] are shown [18, 39]); and d) combinatorial length selection by two different ubiquitin receptors (RPN10 and DSK2 are shown [56]). Abbreviations: UBA, ubiquitin-associated domain; UBL, ubiquitin-like; UBX, ubiquitin regulatory X domain; USP, ubiquitin-specific protease; ZnF, zinc finger.

Abbreviations
AAA ATPase, ATPases associated with a variety of cellular activities; ESCRT, endosomal sorting complex required for transport; GST, glutathione S-transferase; LUBAC, linear ubiquitin chain assembly complex; NF-κB, nuclear transcription factor kappa B; SUMO, small ubiquitin-like modifier; TUBE, tandem ubiquitin-binding entity; UBA, ubiquitin associated; UBC, ubiquitin C; UBE, ubiquitin-activating enzyme; UBX, ubiquitin regulatory X domain; UIM, ubiquitin interacting motif; USP5, ubiquitin-specific protease 5.

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
We thank Robert Cohen for sharing his manuscript and helpful comments.

References


F1000 Factor 10


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 7
Evaluated by Christiane Riedinger and Jane Endicott 12 Aug 2009, Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 7
Evaluated by Zhen-Qiang Pan 23 Apr 2009, Keith Wilkinson 17 Nov 2011


F1000 Factor 9


F1000 Factor 7


F1000 Factor 9
Evaluated by Zhen-Qiang Pan 23 Apr 2009, Keith Wilkinson 17 Nov 2011


F1000 Factor 9


F1000 Factor 9
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011


F1000 Factor 6
Evaluated by Keith Wilkinson 17 Nov 2011

