**Trends in Biochemical Sciences**

**The ball and chain of polyubiquitin structural studies**

---Manuscript Draft---

<table>
<thead>
<tr>
<th>Manuscript Number:</th>
<th>TIBS-D-15-00167R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article Type:</td>
<td>Review</td>
</tr>
<tr>
<td>Corresponding Author:</td>
<td>Annalisa Pastore</td>
</tr>
<tr>
<td></td>
<td>London, UNITED KINGDOM</td>
</tr>
<tr>
<td>First Author:</td>
<td>Caterina Alfano</td>
</tr>
<tr>
<td>Order of Authors:</td>
<td>Caterina Alfano</td>
</tr>
<tr>
<td></td>
<td>Serena Faggiano</td>
</tr>
<tr>
<td></td>
<td>Annalisa Pastore</td>
</tr>
<tr>
<td>Abstract:</td>
<td>Ubiquitylation is a post-translational modification implicated in several different cellular pathways. The possibility to forming chains through covalent cross-linking between any of the seven lysines or the initial methionine to the C terminus of another moiety provides ubiquitin (Ub) with special flexibility in its function in signalling. Here, we review the knowledge accumulated over the last years about the functions and structural features of polyUb chains. This analysis reveals the need to understand further the functional role of some of the linkages and the structural code that determines recognition of polyUbs by protein partners.</td>
</tr>
</tbody>
</table>

*Powered by Editorial Manager® and ProduXion Manager® from Aries Systems Corporation*
Trend Box

- Ubiquitin (Ub), a small signalling protein, forms chains thanks to covalent linkage between any of the 7 lysines or the N-terminal methionine of one subunit and the C terminus of another Ub.

- PolyUb chains both with the same or mixed linkages are involved in several cellular functions from proteasomal targeting to protein regulation and have thus acquired an enormous importance in cellular signalling.

- PolyUbs exhibit a unique repertoire of conformational states that is dependent on the specific linkages, which have different flexibilities.

- Further complexity is added by the interaction of polyUbs with cellular partners, which modulate their structure and functions.

- Understanding the structural and functional aspects of the polyUb code continues to offer an important challenge.
Review

The ball and chain of polyubiquitin structures

Caterina Alfano¹, Serena Faggiano², Annalisa Pastore*¹

¹Department of Clinical and Basic Neuroscience, King’s College London, London, United Kingdom
²Department of Pharmacy, University of Parma, Parma, Italy

*To whom correspondence should be addressed.

Email: annalisa.pastore@crick.ac.uk

Keywords: biophysics, post-translational modifications, proteasomal pathway, signalling, structure, ubiquitin

Running title: Structural studies of polyubiquitin chains

Ubiquitylation is a post-translational modification implicated in several different cellular pathways. The possibility of forming chains through covalent cross-linking between any of the seven lysines, or the initial methionine, and the C terminus of another moiety provides ubiquitin (Ub) with special flexibility in its function in signalling. Here, we review the knowledge accumulated over the past several years about the functions and structural features of polyUb chains. This analysis reveals the need to understand further the functional role of some of the linkages and the structural code that determines recognition of polyUbs by protein partners.

Why are polyUb chains interesting?

Post-translational modifications are one of the keys to understanding the complexity of eukaryotic genomes: capable of considerably changing protein surfaces and binding properties of a protein, they act as a common mechanism for modulating and regulating protein function. Eukaryotes have a large repertoire of post-translational modifications, which include phosphorylation, acetylation,
methylation, ubiquitylation, sumoylation, neddylation, glycosylation and glycation [1]. Among these, ubiquitylation is a modification that brings particular flexibility both from structural and functional perspectives. Ub (76 amino acid residues) is a ubiquitous globular protein (Box 1) that is evolutionarily stable, being highly conserved among eukaryotes and sharing a sequence identity between human and yeast as high as 96% [2].

Ubiquitylation occurs by the covalent attachment of the C-terminal carboxyl group of Ub to the ε-amine of lysine residues of a substrate protein through an isopeptide bond. The acceptor can be either a different protein or Ub itself. When the latter occurs, this leads to formation of chains with different linkages that may involve Lys6, Lys11, Lys27, Lys29, Lys33, Lys48 or Lys63 [3]. Chains can also contain linear (via the N-terminal Met1 of Ub) or mixed linkages [4].

The seven lysine residues are present on different surfaces of the Ub monomer and have different degrees of exposures to the solvent. Lys6, Lys33 and Lys63 are fully exposed and do not form intramolecular contacts. Lys11 and Lys27 are the least exposed to the solvent, as they point towards the hydrophobic core and are involved in salt-bridges with the carboxylate oxygen atoms of Glu34 and Asp52, respectively. Finally, Lys29 and Lys48 form hydrogen bonds from their ε-amino nitrogen atoms and the carbonyl oxygen atoms of Glu16 and Ala46, respectively, and are partially exposed to the solvent. The amino acid position involved in the linkage thus dictates the structural properties of the corresponding chain and differences in the biological function. All chain types are present in both yeast and mammalian cells but the relative abundance of the individual chains depends on the specific cell state [5-7].

The richness of conformations that the different linkages confer has been the subject of intense studies over the past 1-2 decades, following advancements in producing ubiquitin chains in vitro. In this review, we summarize our knowledge of the function(s) of polyUb and describe the conformational landscape of polyUb chains both in isolation and in complexes with other proteins. Although our list
of examples will not be exhaustive given the continuous flow of publications, it is hoped that this review will provide a useful guideline to discuss the different aspects of polyUb versatility.

The functional role of polyUb

The different polyUb chains are involved in different metabolic pathways, where they seem to have an essential role in regulation. All linkages of polyUb homopolymers can target proteins for proteasomal degradation. However, Lys48 polyUb chains are most clearly involved in targeting a substrate for degradation [8-10]. PolyUbs with Lys48 and Lys63 linkages are also involved in cellular signalling [11]. Lys11, Lys63 and Met1 linkages seem to be involved in the regulation of cell cycle and nuclear factor-κB activation [12-14]. Lys11 chains also mediate protein degradation during cell division and bind to the S5 subunit of the proteasome [12]. The complexity of the polyUb code does not end here: mixed chains with linkages from different lysines also seem to target proteins for proteasomal degradation [15, 16]. Branched chains, in which more than one Ub domain is attached to the same subunit, have also been described [4]. Since these chains bind to the proteasome but with low affinity, they are thought to be involved in other pathways. Mixed, multiply branched chains with Lys6, Lys27 and Lys48 linkages are attached to the monoubiquitinating ligase RING1b. This modification is needed to stimulate the RING1b enzyme activity toward the substrate histone H2A [17, 18].

Ub function can in turn be modulated by post-translational modifications such as acetylation or phosphorylation [19-21]. Phosphorylation of eight out of the eleven possible sites (Thr7, Thr12, Thr14, Ser20, Ser57, Tyr59, Ser65, and Thr66) has been observed in mammals [20, 22-26]. Preferential phosphorylation of Ser65 is favored in Lys6 and Lys11 polyUb chains, while it is disfavored in Lys27 chains [27]. Parkin E3 ligase is activated by pSer65-Ub initiating a complex chain of events [28, 29], whereas the deubiquitylating efficiency of deubiquitinating (DUB) enzymes USP2, USP5, USP8, USP15, USP30, ataxin-3, and AMSH is inhibited by phosphorylation at this position [21, 27, 30].
Interestingly, the site of phosphorylation (distal or proximal domain; see Box 1) is relevant: USP2 is inhibited by phosphorylation of the distal Ser65, while AMSH is inhibited by phosphorylation of the proximal Ser65 [30].

Acetylation also significantly affects the ubiquitylation cascade, apparently by repressing polyUb elongation in vitro while stabilizing the monoubiquitylation state in cells [20]. The known acetylated sites are Lys6 and Lys48 [20]. Ub acetylation can occur after Ub is conjugated to partners and affects Ub interaction with proteins including DUBs, Ub binding domains (UBDs) and enzymes involved in Ub chain assembly [20, 31, 32].

The kaleidoscopic world of polyUb structure

Ubiquitylation is catalyzed in vivo by a three-step enzymatic cascade. First, an E1 activating enzyme uses ATP to link the C-terminal glycine of Ub via a thioester bond to a cysteine in the E1 active site [33-35]. The activated Ub intermediate is then transferred to the catalytic cysteine of an E2 enzyme. The E3 is finally responsible for further transfer of Ub to the substrate [36]. Target specificity and independent regulation of different linkages occurs through the E2 and E3 enzymes [37]. E3 ligases are divided into four families, HECT, RING-finger, U-box, and PHD-finger, based on the mechanism of catalysis. HECT E3 ligases temporarily accept activated Ub, whereas RING E3 ligases catalyse direct transfer of Ub from the E2 to the substrate [38].

A prerequisite to allow in vitro structural and functional studies of polyUb chains is to produce suitable quantities of pure samples. This is often not an easy task given that some of the enzymes responsible for linkage are unknown. Several strategies have thus been designed to circumvent the problem, some of which rely on enzymatic approaches (Box 2). Other methods are based on chemical synthesis (Box 3).

The first Lys48 diUb structure appeared in 1992 [39]. Since then, several crystallographic and NMR structures of polyUb chains have been solved: more than 100 structures are available in the PDB data.
repository. Most of these studies relate to Lys48 and Lys63 linkage chains, whereas the first structure of a Met1 (or linear) diUb was released only in 2009 [40] followed by the structures of Lys6 and Lys11 diUbs in 2010 [41, 42]. We had to wait until 2015 to see structures of Lys29 [43, 44] and Lys33 diUb [44, 45]. Notably, structural information on Lys27 linkage chains is still missing. Two leitmotifs common to these structures are open/closed conformations and flexibility (Figure 1). The linker geometry determines whether the chain forms a more compact ‘closed’ assembly, in which the domains pack against each other typically through various hydrophobic patches, or forms a more extended open structure. Open structures are usually flexible and cover a large number of isoenergetic states of the energy landscape. Next, we briefly discuss the most important features of these structures. The order chosen matches the order of publication for each particular linkage.

**Lys48 linked chains: open or closed?**

There are several structures of isolated Lys48 di and tetraUb [39, 46-54] (Table 1). The predominant state of Lys48 diUb chains at near-physiological conditions is a ‘closed’ conformation in which the Ile44-centred hydrophobic interfaces (Box 1) of each subunit pack against each other (pdb code: 1aar and 3m3j), rendering this closed conformation incompetent for interaction with protein partners [39, 48]. However, this is thought of as only one of the several possible conformations of the Lys48 diUb energy landscape. Solution studies demonstrated that open and closed conformations are in fast exchange with each other [54, 55] and that this equilibrium is critical for recognition of Lys48-linked polyUb chains (2pea, 2pe9, 3ns8 and 3aul). Lys48 tetraUbs form a pseudo-tetragonal structure in which pairs of hydrophobic patches interact, suggesting that diUb is the minimal structural unit. As for diUb chains, tetraUbs are thought to visit closed and open conformations in fast equilibrium and use this dynamical conformation state for partner recognition (1tbe, 1f9j, 2o6v and 3alb) [46, 47, 50, 52].
Lys63 and Met1 linked chains: it could not be more extended

In all available structures (which are from X-ray diffraction), Lys63 chains have an entirely different three-dimensional structure from Lys48-linked chains [40, 49, 56-58]. The conformation is extended and open, with no well-defined non-covalent interface between moieties. The inter-domain interaction between the Ile44-centred hydrophobic patches of the two moieties seen in Lys48 chains is not sterically possible because the Lys63 side chain is on the opposite side of the subunit compared to the Ile44-centred hydrophobic patch, while the linker is not long enough to allow inter-domain interactions between the two patches. This leaves the hydrophobic patches exposed and competent for binding. In the crystal structures (2jf5, 3h7p, 3hm3), the hydrophobic patches do not lie on the same surface of the Ub molecule but are sequentially rotated by ~90° [40, 57, 58]. In solution it is more likely that the linker is flexible, thus allowing the individual monomers to undergo free rotational movements that would change the relative orientation of the patches. What has been named the ‘fluidity’ or dynamics of these chains may be important for function [58].

The overall conformation of Met1 chains is virtually equivalent to Lys63 chains and consists of structures in which individual Ub moieties are rotationally unrestrained and highly flexible (3axc and 2w9n). This is expected because Met1 is spatially close to Lys63 in the Ub structure (Box 1). The two types of chains thus differ mainly by the chemistry of the linkage between domains [40].

Non-canonical chains: Lys6, Lys11, Lys29 and Lys33-linked chains

The only known diUb structure of the Lys6 linkage is a crystal structure (2xk5) [41]. This reveals an asymmetric compact conformation in which the proximal Ub domain commits only two of its canonical hydrophobic residues (Ile44 and Val70) to bind a slightly different but close hydrophobic patch in the distal Ub domain containing Ile36, Leu71 and Leu73 (referred to as the Ile36 patch). This asymmetric interaction leaves part of the canonical Ile44-centred patch of the distal domain exposed...
and available to binding to other partners. It is, however, possible that in solution the dynamic nature of polyUb chains allows other interfaces.

Three different structures of Lys11 chains have been published (3nob, 2xew, 2mbo), which were determined under different conditions [42, 59, 60]. Notably, the orientation of the two Ub moieties is different in all three structures and leads to a different degree exposure of the Ile44-centred hydrophobic patches. This could however be partially the consequence of crystal packing in the crystal structure.

The crystal structure of Lys29 diUb (4s22) adopts an extended conformation, distinct from that observed in Lys63 and in Met1 linked chains. The inter-domain contacts are mainly polar and involve Arg42 and Arg72 of the distal Ub domain and Glu16, Glu18 and Asp21 of the proximal domain [43, 44].

The crystal structures of Lys33 di- and triUb were solved recently (4xyz, 4y1h, 5af4 and 5af5) [44, 45]. Lys33-linked diUb adopts a compact structure where the two monomeric subunits interact extensively through two hydrophobic surfaces and involving the interdigitation of patches formed by Ile36, Leu71, Leu73, Leu8, Ile13 and Leu69. This arrangement leaves the Ile44-centred patch exposed to the solvent. This compact conformation is similar to the closed conformation adopted by one of the Lys11 diUb structures (3nob) [59] but distinct from that observed in Lys6 and Lys48 diUbs. DUB enzymes can distinguish between these linkages [61]. The crystal structure of Lys33 triUb reveals an open conformation where the monomeric subunits form no mutual interactions except for the isopeptide linkage [44, 45].

**PolyUb chains in complex with UBDs**

To understand the mechanisms by which proteins recognise different Ub linkages, it is essential to look at the structures of polyUb complexes with other partners. One of the first structures and the simplest model for polyUb interactions was that of the Ub associated 2 (UBA2) domain of hHR23A
with Lys48 polyUb. UBA2 forms asymmetric interactions with the proximal and distal Ub domains (1zo6) [62]. The proximal Ub has higher-affinity binding than the distal domain, as also seen in the S5a/Lys48 diUb complex (2kde) [63, 64]. The interacting surface of Ub is centred on I44, implying a closed-to-open conformation transition of the diUb upon binding to UBA2 that results sandwiching of UBA2 between the two Ub subunits. The authors suggested that selectivity is given by the entropic cost of opening the interface and immobilizing the UBA2 domain.

After this seminal work, several other papers have reported structures of polyUb chains with other proteins. The currently available structures (Tables 2,3) can be divided into two families: the structures of complexes with UBDs and those of complexes with DUBs. From them we can address questions such as: Are there general rules that dictate Ub interactions with other proteins? Can we explain through structural determinants chain-selectivity? It is perhaps still too soon to answer the question of whether complexes of short polyUb chains follow the same rules as longer ones.

There seems to be two distinct strategies for interaction and linkage specificity. One, preferentially adopted by UBDs, in which linkage specificity is not mediated by direct interactions with the linkage. In the second, adopted by several DUBs, the isopeptide linkage participates to recognition. Another common theme to Ub complexes is the importance of the distal Ub in recognition, which almost always establishes direct contacts with the UBD (Figure 2 and Table 2).

Representative examples of UBD/polyUb complexes are the crystal structures of the RAP80 tandem Ub interacting motifs (UIMs) in complex with Lys63 diUb (3a1q) [65], the UBAN (Ubiquitin Binding of ABIN and NEMO) motif of NEMO in complex with Met1 diUb (2zvn) [66], the NMR-based structural model of the second UBA of hHR23A in complex with Lys48 diUb (1zo6) [62] and a Fab fragment Apu2.16 in complex with Lys63 diUb (3dvn) [67].

Recently, the structure of the nitrate zinc finger 1 (NZF1) domain of TRABID in complex with Lys29 diUb (4s1z) was published [43]. In this case, there is no direct interaction between the linkage and the UBDs. Instead, the NZF1 domain of TRABID interacts both with the distal and proximal Ub domains.
and recognises the Ile44-centred patch on the distal domain and residues around Lys29 of the proximal domain. The overall conformation and relative orientation of the two Ub subunits change in the complex with respect to the arrangement on the unbound Lys29 diUb. Structural remodelling enables the NZF1 to bind simultaneously both the distal and the proximal domains. This conformation is thought to confer binding selectivity.

The accumulated structural knowledge has revealed that recognition of polyUb is often achieved through rather weak binding of different domains to more than one Ub position in the polyUb chain. This allows the formation of tight complexes through a mechanism of avidity and/or cooperativity even when interaction between individual sites and monoUb might be weak (Box 4). The complex of the individual tandem UIMs of RAP80 with monoUb is, for instance, too weak (Kd value 500 μM) to form a stable complex (3a1q) [68] but the simultaneous interactions of multiple UIMs with the multiple subunits of a Lys63 polyUb chain significantly strengthen the complex. A possible advantage of having several weak interactions is the possibility of gaining linkage specificity when the arrangement of the domains favors avid binding to one polyUb linkage over others.

**The critical role of spacers between UBDs**

The length and the flexibility of the linkers between UBDs are crucial for an avidity mechanism that could be sensitive to the orientation of the binding subunits [65, 68]. The binding sites must be optimally placed for simultaneous binding. A similar mechanism is observed in the structures of the complexes of hHR23A with Lys48 diUb, of NEMO with Met1 diUb [62, 66], and of antibodies against Lys63-linked Ub chains in complex with Lys63 diUb (3dvn) [67]. RAP80 recognises the hydrophobic patches centred on Ile44 of both Ub moieties by the tandem UIM regions: UIM1 interacts with the proximal Ub domain whereas UIM2 interacts solely with the distal domain. The authors suggested that selectivity could be ensured by the inter-UIM region which has the right geometry and length to adopt a conformation suitable for this type of recognition. Indeed, the linker forms a twelve amino acid long
alpha helix, which functions as a ‘molecular spacer’ and dictates the distance and the relative orientation between the two UIMs.

In the UBAN motif of NEMO in complex with Met1 diUb (2zvn) [66], the UBAN forms a typical parallel coiled-coil homodimer held together by hydrophobic interactions. This homodimer binds two diUb chains, one at each side of the coiled-coil. The distal Ub domain interacts mainly by its hydrophobic patch centred on Ile44, while the proximal domain interacts through polar interactions with NEMO residues. Linkage specificity is granted by the fact that the simultaneous binding of both proximal and distal domains can occur only with Met1 chains.

Linkage specificity is also dictated by the spacing between the distal and proximal Ub domains in the complex of Met1 diUb and the HOIL-1L NZF domain of LUBAC (linear Ub chain assembly complex; the enzyme responsible for the formation of Met1 chains) (3b0a) [69]. HOIL-1L NZF contains two Ub binding regions arranged to simultaneously bind both the proximal and the distal domains while having no direct contact between the HOIL-1L NZF domain and the linkage.

The NZF domains of TAB2 and TAB3 in complex with Lys63 diUb (3a9j, 3a9k and 2wwz) [70, 71] are almost identical and present the same structural features. Both proteins use two different well-conserved hydrophobic regions of the NZF domain to interact with the hydrophobic patches of the distal and proximal Ub domains and have no direct interaction with the Lys63 isopeptide bond. Like RAP80, the two Ub-binding sites of the TAB proteins are arranged to optimise binding affinity. The distance between them enables the simultaneous binding of the distal and the proximal domains, enhancing the affinity.

Interaction of polyUb with DUB enzymes

Much fewer structures are available of polyUb complexes with DUBs. The available structures are representative of only three of the linkage types (Met1, Lys11 and Lys63) and only three of the five DUB subfamilies (Table 3). Additionally, the structure of OTUB1 (an OTU DUB) in complex with
the E2s UbcH5b/UBE2D2 and Ubc13 (4dhz) indirectly suggests the arrangement of a Lys48 diUb bound to an OTU DUB protein [72]. In this structure, an unbound monoUb was observed in the crystal in addition to the bound Ub. The two Ubs are roughly positioned to allow a Lys48 linkage. Likewise, a model of the Josephin domain of ataxin-3 (Josephin subfamily) with Lys48 diUb was suggested [73] but it remains unclear whether this has physiological meaning.

DUB recognition seems much more differentiated and strongly dependent on the specific DUB subfamily, some of which do not have linkage specificity (reviewed in [38]), such as the Ub-specific protease (USP) subfamily. The OTU subfamily can have different specificities: OTUB1 and Cezanne, for instance, prefer Lys48 and Lys11 linkages, respectively, and OTULIN is specific for Met1 linkages. TRABID and CYLD are examples of dual specificity, the former cleaving Lys29 and Lys33 linkages, the latter being specific for Met1 and Lys63. Also DUBs of the OTU family exhibit dual specificity [31].

Overall, two features seem to be common (Figure 3). In all complexes the distal Ub makes contacts through the Ile44 patch or nearby residues. This is true also for the structure of USP21 in complex with linear diUb-aldehyde (Figure 3c) where the proximal Ub is covalently attached to the catalytic cysteine. There are several examples in which specificity is dictated by direct recognition of the (iso)peptide linkage. In the complex between AMSH-LP DUB (JAMM family) and Lys63 diUb (2znv) [74], for instance, Lys63 diUb keeps an extended conformation as in the unbound form. AMSH-LP is a zinc dependent DUB enzyme that specifically cleaves Lys63 polyUb chains. The binding affinity mainly relies on the interaction of AMSH-LP with two hydrophobic patches of the distal Ub domain centred on Ile44 (Leu8, Ile44, Val70) and on Ile36 (Ile36, Leu71, Leu73). While the interaction between AMSH-LP with distal Ub domain is critical for binding affinity, the interaction of AMSH-LP with proximal Ub is critical for catalysis. To allow cleavage of the isopeptide bond, Lys63 of the proximal Ub has to be kept in the right orientation for catalysis. Thus, the surface around the Lys63 of
the proximal Ub is located inside a concave surface made by two regions of AMSH-LP where the tripeptide Glu62-Lys63-Glu64 is correctly aligned for cleavage.

OTULIN (OTU family), also known as FAM105B or Gumby, seems to be the only DUB specific for Met1 linkages. The catalytic domain of OTULIN interacts with Met1 diUb (3znz and 4ksl) [75, 76] through the Ile44 centred hydrophobic patch of the distal Ub domain, while the proximal domain binds OTULIN by a different surface. The binding interface between OTULIN and the proximal Ub is quite extensive and is very important for binding specificity as it orients the proximal Ub to place the Met1 linkage in the catalytic pocket and positions it for cleavage. OTULIN is the first DUB enzyme to show a substrate-assisted catalytic mechanism: the proximal Ub participates directly in the organization of the catalytic site with Met1 displacing the catalytic His by steric hindrance from its unbound position. This movement results in activation of the otherwise inactive enzyme. Glu16 of the proximal Ub also participates in restricting the movement of the catalytic His and in introducing a negative charge.

Concluding remarks

We have seen here how complex the repertoire of structures allowed by chains of a protein as small as Ub can be. The possibility of so many different linkages and the interactions with other partners amplifies the modularity of Ub behaviour, allowing the adoption of very different conformational landscapes. Our understanding of the structural determinants of polyUb functions and interactions continues to expand; we have tried to summarize the current knowledge. We have also discussed the challenges and questions that remain to be addressed in the near future, before a more accurate understanding is available that clarifies the role of polyUb in signalling and provides new insights into its importance for all cellular functions. An important challenge for the future is to solve the structure of Lys27-linked polyUb. We will also need to understand if there is a more universal code for polyUb interactions and whether and how longer chains differ from diUbs. To answer to these questions, more structures of Ub complexes are needed. When they become available, we may also be in the position
of understanding the significance that the flexibility of different linkages plays for partner recognition and the role of each linkage. Finally, a fascinating open question is how post-translational modifications modulate the Ub code.

Acknowledgements

We thank Pierandrea Temussi and David J. Thomas for critical reading and Doriano Lamba for help with Figure 2 preparation. This review was possible thanks to the financial support of MRC.

References

2 Goldstein, G., et al. (1975) Isolation of a polypeptide that has lymphocyte-differentiating properties and is probably represented universally in living cells. Proc Natl Acad Sci USA 72, 11-15
4 Kim, H.T., et al. (2007) Certain pairs of ubiquitin-conjugating enzymes (E2s) and ubiquitin-protein ligases (E3s) synthesize nondegradable forked ubiquitin chains containing all possible isopeptide linkages. J. Biol. Chem. 282, 17375-17386


23 Shiromizu, T., et al. (2013) Identification of missing proteins in the neXtProt database and unregistered phosphopeptides in the PhosphoSitePlus database as part of the Chromosome-centric Human Proteome Project. *J Proteome Res* 12, 2414-2421


65 Sato, Y., et al. (2009) Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. EMBO J. 28, 2461-2468
66 Rahighi, S., et al. (2009) Specific recognition of linear ubiquitin chains by NEMO is important for NF-kappaB activation. Cell 136, 1098-1109
103 Tokunaga, F., et al. (2012) Specific recognition of linear polyubiquitin by A20 zinc finger 7 is involved in NF-kappaB regulation. EMBO J. 31, 3856-3870
111 Shrestha, R.K., et al. (2014) Insights into the mechanism of deubiquitination by JAMM deubiquitinases from cocrystal structures of the enzyme with the substrate and product. Biochemistry 53, 3199-3217
Box 1: The Ub identity card

Ub is a small protein (76 amino acid residues) with a globular, compact and tightly hydrogen-bonded fold.

Mammalian cells have four Ub genes, which encode polyUb or fusions with ribosomal proteins. The polyUb precursor is a head-to-tail linear polymer with the carboxyl group of the C-terminal Gly76 of one subunit being covalently bound to the α-amino group of the N-terminal Met1 of the next molecule. Cleavage of these linear chains and of Ub fusion proteins produces the mature monomeric protein in which the N terminus, no longer involved in the peptide bond, changes orientation and packs against the structural core forming a compact globular structure [77, 78]. The fold consists of a mixed β-sheet with five strands, a short 310-helix and 3.5 turns of an α-helix, which packs against the cavity of the β-sheet (Figure I) [79]. The first seven residues in the N-terminal region of the protein are tightly constrained. This renders the N terminus not easily accessible, preventing degradation. Ub is highly stable over a wide range of pH and temperature values and against protease digestion despite the presence of seven lysine residues and four arginine residues, which could easily be attacked by trypsin [80]. The only part of the protein protruding out from the structure is the C-terminal Leu-Arg-Gly-Gly motif, which does not interact with the rest of the molecule and is thus accessible to enzymes involved in the formation of both isopeptide and peptide bonds. Other small proteins with folds similar to Ub have also been identified, among which are NEDD8, SUMO-1, SUMO-2, SUMO-3, ATG8, ISG15, and FAT10. These proteins are collectively called Ubl. A key feature of Ub is the solvent-exposed hydrophobic-patch centred on Leu8, Ile44, and Val70 (Figure I, in red). This is the preferential target of most UBDs and DUBs despite their diversity in function and structure [81, 82]. Alternative hydrophobic surfaces are also possible, such as the patch centred on Ile36 (Ile36, Leu71, Leu73, in orange in Figure I). In polyUbs, the C terminus of one subunit (conventionally named the distal domain) can covalently attach to one of each of the seven lysine residues (Figure I, in blue) or to the N terminus Met1 (Figure I, in yellow) of a second subunit (conventionally named the proximal domain).
Box 2: Producing polyUbs in the laboratory in large quantities

Enzymatic production can be achieved in vitro either using wild-type or a mutated Ub as a substrate (for a recent extensive review see [83]) (Figure I). Mutants are more modular as chain elongation can be controlled step by step. This allows the possibility of differentiating between the different positions of the chain and obtaining chains in which each position can be differentially isotope labelled as required, such as in NMR studies [48, 60, 84]. However, producing mutants is more laborious than using the commercial wild-type protein. More recently, protocols for the production of Lys48, Lys63, and Lys11 linked chains using wild-type Ub have been developed [42, 85-87]. The method is based on the formation of unanchored Ub chains by E1 and a specific E2/E3 enzyme and separation of the products by cation exchange chromatography. It requires the use of an E2/E3 enzyme specific for a target chain type. A limitation is, however, that no selective isotope labelling of a particular position is possible, as the reagents are indistinguishable. Also, chains longer than tetraUb can be separated with difficulty by cation exchange chromatography. Other protocols allow the production of Lys29 and Lys33 chains exploiting DUB enzymes to cleave unwanted chain types formed as secondary products [43-45]. Finally, production of milligrams of Met1 chains can be achieved using suitable plasmids for expression of the required chain length. Several of these plasmids for the expression of di, tri and tetraUb are already available [85, 88].
Box 3: Chemical synthesis of polyUb chains

Chemical methods for the production of Ub chains have been developed to circumvent the need of identification of the specific E2/E3 system for the ligation of a selected chain type (for previous reviews see [89-91]). Another advantage is their suitability to produce Ub chains with selective isotope labelling in a desired position of the chain. The most exploited approach for the chemical synthesis of an isopeptide bond between two Ub building blocks is expressed protein ligation (EPL) (Figure I). This technique involves native chemical ligation (NCL) between a Ub subunit carrying a C-terminal thioester (α-thioester) and a second Ub with a free thiol adjacent to an amine group (mercaptolysine) [92-94]. α-thioesters are generated from thiolysis of an intein fusion protein, while the mercaptolysine is obtained through solid-phase peptide synthesis (SPPS). EPL consists in a transthioesterification between the thiol and the α-thioester, followed by a spontaneous intramolecular S-to-N acyl transfer and a desulphurization step, resulting in a native isopeptide bond. An evolution of NCL is genetically encoded orthogonal protection and activated ligation (GOPAL) [41]. As in NCL, a Ub α-thioester is obtained from intein fusion thiolysis. All lysines are blocked with a protecting group. The second ligand has a genetically encoded modified lysine, protected with a different protecting group, which is removed to expose the selected amine group to ligation. Finally, the other protecting groups are also removed. GOPAL has allowed synthesis of Lys6 and Lys29 diubiquitin [41].
An extended GOPAL strategy was applied to the preparation of longer chains [95]. Di and triUb α-thioesters were prepared using intein chemistry, and a different protecting group for lysines was used that could be removed in milder conditions. Other methods were developed for synthesis of Ub chains formed via non-native linkages. For example, Ubs were linked via a triazole bond exploiting a copper-catalyzed azide-alkyne “click” reaction [96], or using thiolene coupling chemistry [97]. Recent advances include the chemical synthesis of phosphorylated diUb [30]. The chemical synthesis of chains remains the preferred method for the production of Lys27 linked chains, since the E2/E3 necessary for enzymatic ligation are not known.

![Figure I. Expressed protein ligation (EPL)](image)

**Box 4: Avidity versus affinity**

Affinity measures the strength of binding between two molecules, each of which contain a single binding site. Avidity is instead a term originally derived from immunology to refer to the properties of antibody-antigen binding. It was meant as a measure of the overall strength of binding of an antigen with many antigenic determinants and multivalent antibodies. It is dependent on three major parameters: affinity of the antibody for the epitope, valence of both the antibody and antigen and structural rearrangements of the parts that interact. As a result we measure more than the sum of the individual affinities. For instance, IgM antibodies may have low affinity but they have high avidity due to their 10 weak binding sites contrary to the two strong binding sites of IgG antibodies. In the cited example, the authors tested whether Rap80, which contains tandem UIMs, could achieve linkage selectivity by exploiting the distinct orientation and spacing of Ub subunits of particular polyUb linkages. They hypothesised that multiple UBDs could be arrayed in space to optimize simultaneous interactions with both Ub subunits and adopt a configuration which could selectively bind only one type of Ub-Ub linkage. Binding by the first of multiple UBDs to one Ub molecule would then position the
second UBD more favorably for the interaction with a second Ub in the chain. This mechanism can be seen as a way to increase the local concentration of the interactor by restricting its conformational space thus making the binding even more favorable.
<table>
<thead>
<tr>
<th>Chain type</th>
<th>PDB</th>
<th>Comments</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met1 diUb</td>
<td>2w9n</td>
<td>X-ray, 2.25 Å</td>
<td>[40]</td>
</tr>
<tr>
<td>Met1 diUb</td>
<td>3axc</td>
<td>X-ray, 2.19 Å</td>
<td>[98]</td>
</tr>
<tr>
<td>Lys6 diUb</td>
<td>2xk5</td>
<td>X-ray, 3.00 Å</td>
<td>[41]</td>
</tr>
<tr>
<td>Lys6 triUb</td>
<td>3zIz</td>
<td>X-ray, 2.90 Å</td>
<td>[84]</td>
</tr>
<tr>
<td>Lys11 diUb</td>
<td>2mbo/2mbq</td>
<td>NMR, inconsistent with the crystal structures</td>
<td>[60]</td>
</tr>
<tr>
<td>Lys11 diUb</td>
<td>2xew</td>
<td>X-ray, 2.20 Å</td>
<td>[42]</td>
</tr>
<tr>
<td>Lys11 diUb</td>
<td>3nob</td>
<td>X-ray, 2.19 Å</td>
<td>[59]</td>
</tr>
<tr>
<td>Lys29 diUb</td>
<td>4s22</td>
<td>X-ray, 2.30 Å</td>
<td>[43]</td>
</tr>
<tr>
<td>Lys33 diUb</td>
<td>4xyz</td>
<td>X-ray, 1.65 Å</td>
<td>[45]</td>
</tr>
<tr>
<td>Lys33 diUb</td>
<td>5af4</td>
<td>X-ray, 1.85 Å</td>
<td>[44]</td>
</tr>
<tr>
<td>Lys33 triUb</td>
<td>4y1h</td>
<td>X-ray, 1.40 Å</td>
<td>[45]</td>
</tr>
<tr>
<td>Lys33 triUb</td>
<td>5af5</td>
<td>X-ray, 1.68 Å</td>
<td>[44]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>1aar</td>
<td>X-ray, 2.3 Å</td>
<td>[39]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>1gJz</td>
<td>NMR, truncated construct 1-51</td>
<td>[99]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>3aul</td>
<td>X-ray, 2.39 Å</td>
<td>[53]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>3m3j</td>
<td>X-ray, 1.60 Å</td>
<td>[51]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>3ns8</td>
<td>X-ray, 1.71 Å</td>
<td>[54]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>2bgf</td>
<td>NMR, used as a model to implement methods</td>
<td>[100]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>2pe9/2pea</td>
<td>NMR, used as a model to implement methods</td>
<td>[101]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>2jfl</td>
<td>X-ray, 1.95 Å</td>
<td>[40]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>3h7p/3h7s</td>
<td>X-ray bound to Cd or Zn, 1.90 Å and 2.30 Å</td>
<td>[57]</td>
</tr>
<tr>
<td>Lys48 tetraUb</td>
<td>1f9j</td>
<td>X-ray, 2.70 Å</td>
<td>[47]</td>
</tr>
<tr>
<td>Lys48 tetraUb</td>
<td>1tbe</td>
<td>X-ray, 2.40 Å</td>
<td>[46]</td>
</tr>
<tr>
<td>Lys48 tetraUb</td>
<td>2o6v</td>
<td>X-ray, 2.2 Å validated by NMR data</td>
<td>[50]</td>
</tr>
<tr>
<td>Lys48 tetraUb</td>
<td>3alb</td>
<td>X-ray, 1.85 Å</td>
<td>[52]</td>
</tr>
<tr>
<td>Lys63 tetraUb</td>
<td>3hm3</td>
<td>X-ray, 1.96 Å</td>
<td>[58]</td>
</tr>
</tbody>
</table>
TABLE 2: Structures of polyUb chains in complexes with UBDs.

<table>
<thead>
<tr>
<th>Chain type</th>
<th>Partner</th>
<th>PDB ID*</th>
<th>UBD</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met1 diUb</td>
<td>NEMO</td>
<td>2zvn/2zvo</td>
<td>UBAN</td>
<td>[66]</td>
</tr>
<tr>
<td>Met1 diUb</td>
<td>HOIL-1L</td>
<td>3b08/3b0a</td>
<td>NZF</td>
<td>[69]</td>
</tr>
<tr>
<td>Met1 diUb</td>
<td>Met1 linkage specific antibody</td>
<td>3u30</td>
<td>Fab</td>
<td>[102]</td>
</tr>
<tr>
<td>Met1 tetraUb</td>
<td>A20</td>
<td>3vuy</td>
<td>A20 ZnF</td>
<td>[103]</td>
</tr>
<tr>
<td>Lys29 diUb</td>
<td>TRABID</td>
<td>4s1z</td>
<td>NZF</td>
<td>[43]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>gp78CUE</td>
<td>2lvp/2lvq</td>
<td>CUE</td>
<td>[104]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>S5a</td>
<td>2kde/2kdf</td>
<td>UIM</td>
<td>[64]</td>
</tr>
<tr>
<td>Lys48 diUb</td>
<td>hHR23A</td>
<td>1zo6</td>
<td>UBA</td>
<td>[62]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>FAAP20</td>
<td>3wwq</td>
<td>NZF</td>
<td>[105]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>RAP80</td>
<td>2rr9</td>
<td>UIM</td>
<td>[106]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>RAP80</td>
<td>3a1q</td>
<td>UIM</td>
<td>[65]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>NEMO</td>
<td>3jsv</td>
<td>UBAN</td>
<td>[107]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>RIG-I</td>
<td>4nqk</td>
<td>2CARD</td>
<td>[108]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>Lys63 linkage specific antibody</td>
<td>3dvg/3dvn</td>
<td>Fab</td>
<td>[67]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>TAB2 and TAB3</td>
<td>3a9j/3a9k</td>
<td>NZF</td>
<td>[70]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>TAB2</td>
<td>2wx0/2wwz</td>
<td>NZF</td>
<td>[71]</td>
</tr>
<tr>
<td>Lys63 triUb</td>
<td>TAB2</td>
<td>2wx1</td>
<td>NZF</td>
<td>[71]</td>
</tr>
</tbody>
</table>

*They are all X-ray structures with the exception of 2kde/2kdf, 2rr9 and 1zo6.*
<table>
<thead>
<tr>
<th>Chain type</th>
<th>Partner</th>
<th>PDB</th>
<th>DUB family</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met1 diUb</td>
<td>CYLD</td>
<td>3wxe/3wxf</td>
<td>USP</td>
<td>[109]</td>
</tr>
<tr>
<td>Met1 diUb</td>
<td>OTULIN/Gumby/Fam105</td>
<td>4ksl</td>
<td>OTU</td>
<td>[76]</td>
</tr>
<tr>
<td>Met1 diUb</td>
<td>OTULIN/Gumby/Fam105</td>
<td>3znz</td>
<td>OTU</td>
<td>[75]</td>
</tr>
<tr>
<td>Met1 diUb aldehyde</td>
<td>USP21</td>
<td>2y5b</td>
<td>USP</td>
<td>[110]</td>
</tr>
<tr>
<td>Lys11 diUb</td>
<td>OTUD2</td>
<td>4boz</td>
<td>OTU</td>
<td>[31]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>CYLD</td>
<td>3wxg</td>
<td>USP</td>
<td>[109]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>Sst2</td>
<td>4nql</td>
<td>JAMM</td>
<td>[111]</td>
</tr>
<tr>
<td>Lys63 diUb</td>
<td>AMSH-LP</td>
<td>2znv</td>
<td>JAMM</td>
<td>[74]</td>
</tr>
</tbody>
</table>
Figure Legends

Figure 1. Representative structures of diUbs with different linkers. Top: The structures of distal Ub were superposed on monoUb (1ubq) and then displaced. The side chain of the proximal linker is explicitly indicated in magenta. As a reference, the Ile44 patch (Leu8, Ile44 and Val70) is indicated by warm red spheres corresponding to the Cα atom. When relevant, alternative hydrophobic patches as indicated in the text are added as orange spheres. Bottom: A schematic representation of the structures using the same colour code used above with indications of the position of the Ile44 patch (in warm red). The relative PDB codes are indicated. The structures of longer chains are not shown for sake of simplicity.

Figure 2. Structures of representative complexes of diUbs with UBD partners. Distal Ub domains (always shown on the top) are indicated with lighter blue ribbons, while the proximal Ubs are shown in dark violet. The Ile44 and Ile36 centred patches are indicated as red and orange spheres respectively. We also added His68 to the Ile44 patch as this residue is often also involved in interaction. The various partners are indicated as black ribbons. The structures correspond to PDB entries: a) 3b08, b) 4s1z, c) 3a9j, d) 3wwq, e) 2kde, f) 2rr9, g) 2zvn, h) 3jsv, i) 1zo6, j) 2lvp, k) 2lvq, l) 4nqk, m) 3u30 and n) 3dvg. The linkage and the partner name are reported. When more than one structure was reported in the same article for the same linkage only one is reported. The only exception is CUE because the two structures significantly differ for the packing of the diUb against the partner. Please note that we have reported the structures available in the PDB Databank; this does not necessary correspond to specificity of binding for a certain chain type. For example, FAAP20 also binds Met1- and Lys48-linked chains. NEMO is also not specific for Lys63-linked chains.

Figure 3. Structures of representative complexes of diUbs with DUB partners. The distal and proximal Ub domains and the hydrophobic patches are indicated as in Figure 2. The structures correspond to PDB entries: a) 4ksl, b) 4boz, c) 2y5b, d) 3wxe, e) 3wxg, f) 4nql and g) 2zvn. The Ub linkage and the
partner name are indicated. Note that in c) the proximal Ub of Met1 diUb is covalently linked to the catalytic cysteine of USP21 through its C-terminal Gly.
Outstanding Questions Box

- What is the structure of Ly27 polyUbs? This information is currently missing because we do not know the enzymes responsible for these linkages, which makes their production difficult. Answering this question would complete the ‘periodic table’ of polyUb structure. To do this, we must design a strategy for the chemical or enzymatic formation of these chains or identify the specific enzymes.

- Also missing are the structures of representative members of different DUB families and many more structures of longer chains. Also in the latter case a bottleneck remains the production of suitable quantities of these polyUb chains. This remains a challenging task.

- Is there a general code of polyUb interaction? PolyUbs are very promiscuous and interact with several very different cellular partners. Is there any general structural determinant that would allow us to predict the structure of polyUbs in complexes?

- How does the intrinsic flexibility of most linkages relate to their functional properties? What is the gain from these dynamical properties?

- **What is the exact role of each polyUb linkage in the proteasomal pathway?** It was originally thought that only Lys48 linkages are involved. More recently this view has changed with the realization that chains with all linkages interact with proteasomal components. Is there any specificity or selectivity?

- How is the Ub/polyUb code further modulated by post-translational modifications such as acetylation and phosphorylation? This is an emerging theme that adds a further layer of complexity but, at the same time, increases the repertoire of possible conformational states.
Figure 3

OTU family

a) Met1
   OTULIN

b) Lys11
   OTUD2

c) Met1 aldehyde
   USP21

d) Met1
   CYLD

e) Lys63
   CYLD

USP family

JAMM family

f) Lys63
   Sst2

g) Lys63
   AMSH-LP