Supporting Information

StUbEx PLUS – a modified Stable tagged Ubiquitin Exchange system for Peptide Level purification and in-depth mapping of Ubiquitination Sites

Vyacheslav Akimov[†,§], Louise C. B. Olsen[†,§], Sten V. F. Hansen[†], Inigo Barrio-Hernandez[†], Michele Puglia[†], Søren S. Jensen[†], Ilia A. Solov'yov[‡], Irina Kratchmarova[†], Blagoy Blagoev[†,*]

SUPPORTING FIGURES AND TABLES

Figure S1. Comparison of the primary sequences and the *in silico* 3-D structures of wild type ubiquitin and Ubi+.

Figure S2. Distribution of identified ubiquitination sites in terms of protein abundance, cellular compartment, biological process and molecular function.

Figure S3. PRM comparative analyses of ubiquitin chains in wild type and StUbEx PLUS U2OS cells.

Additional information:

Interpretation of the molecular dynamics (MD) simulations of Ubi+ and WT ubiquitin structures.

Primary PRM data from Skyline for the analysis of ubiquitin chains.

Table S1. Lysine ubiquitinated peptides and sites.

Table S2. peptides from N-terminally ubiquitinated proteins.

SUPPORTING FIGURE LEGENDS

Figure S1. Comparison of the primary sequences and the *in silico* 3-D structures of wild type ubiquitin and Ubi+. (A) Schematics of the lenti-viral plasmids used for generating stable, tetracycline-inducible, StUbEx PLUS cell lines. Insert below provides the cDNA and corresponding protein sequences of wild-type (WT) ubiquitin and the RNAi-resistant Ubi+. (B) Comparison of the structures of WT ubiquitin and Ubi+. Top panel: surface representation of the known crystal structure of ubiquitin (PDB ID: 1UBQ) after the performed MD simulations. Color indicates the standard types of the residues, while magenta highlights the six incorporated histidine residues. Bottom panel: direct overlay comparison of the secondary and tertiary structures of WT (blue) and Ubi+ (red). The positions of the lysine residues are indicated with blue (WT) and red (Ubi+) spheres; bottom right inset: arrangement of the inserted six histidine residues in Ubi+. (C) Stability of the WT ubiquitin and its Ubi+ variant. Time evolution of the root mean square deviation (RMSD) of the WT (green) and of the Ubi+ (orange) as calculated from the performed MD simulation. RMSD is computed for the backbone atoms of both proteins, following Eq. (1) (see Additional Information below), relatively to the equilibrated structures. (D) Energy distribution for WT ubiquitin and Ubi+. Distribution of the internal energy, computed for both WT (green) and Ubi+ (orange) from the performed 100 ns long MD simulation. The red lines indicate Gaussian fitting of the distribution with Eq. (2) (see Additional Information below), being indicative for a stable configuration.

Figure S2. Distribution of identified ubiquitination sites in terms of protein abundance, cellular compartment, biological process and molecular function. (A) Distribution of the identified ubiquitinated proteins over the U2OS proteome in respect to cellular abundances. Proteins were separated in bins according to their copy number in the cell (as reported in Back et al., 2011) and the bar plot show the percentage of the proteins within each bin with identified ubiquitination sites in the StUbEx PLUS dataset. (B) GO term enrichment analyses for the entire U2OS proteome (dark green bars) and for the ubiquitinated proteins from the StUbEx PLUS dataset (light green), shown as percentage of the proteins from the total following each of the most significant terms. The analysis

was performed using GOCC (cell compartment), GOBP (biological process) and GOMF (molecular function).

Figure S3. PRM comparative analyses of ubiquitin chains in wild type and StUbEx PLUS U2OS cells. Relative quantification of the tryptic peptides originating from ubiquitin chains from whole cell lysates StUbEx PLUS U2OS and Wild type U2OS cells by PRM LC-MS/MS. Quantitation is based on the top five fragments of each peptide normalized to a unique peptide of cytoplasmic actin.

Interpretation of the molecular dynamics (MD) simulations of Ubi+ and WT ubiquitin structures

The root mean square deviation (RMSD) of a protein is a useful characteristic that permits judging about its overall stability. Mathematically, it is virtually defined as:

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N} |\vec{r}_i(t) - \vec{r}_i(t_{ref})|^2}{N}} \qquad (1)$$

Here the summation runs over all atoms for which the RMSD should be computed, with *i* being an index the *i*-th atom, *N* is the total number of atoms, $\vec{r}_i(t)$ describes the position of an atom at the time instance *t*, while $\vec{r}_i(t_{ref})$ describes it at the reference time instance t_{ref}. Figure S1C shows RMSD time evolution calculated for the WT ubiquitin and for the Ubi+. RMSD for both proteins flattens out at a value of about 1.5 Å, which indicates that both proteins are stable in the simulation. The horizontal dashed lines indicate the average RMSD values obtained for the last 70 ns of the simulation. In both cases, the reference time instance t_{ref}, is taken as the last snapshot after the equilibration simulation (see methods for details). The saturation of RMSD at a value below 2 Å is strongly supporting the stability of the two computational models [1,2]. The fact that WT ubiquitin is stable is not surprising, but the

comparison of RMSD for the WT and Ubi+ shows that the Ubi+ should be as stable as the WT and have a very similar secondary structure.

RMSD flattening means that on average the atoms in both structures are not moving more than ~1.5 Å from the reference configuration, which is an indication that both structures, which were highly similar after equilibration, preserve this property. This can also be seen in Figure S1B, where 3D rendering of the proteins is presented and illustrates that all main folding motifs of the WT ubiquitin are preserved in Ubi+. Note, that the deviations of the six histidines themselves are not taken into the RMSD calculation of the Ubi+, therefore allowing to directly compare the WT ubiquitin with the identical part of the Ubi+ sequence.

To further illustrate the stability of Ubi+, we have computed and analyzed the probability distribution of the internal energy of Ubi+ and WT ubiquitin, as featured in Figure S1D. Both distributions show a clear Gaussian profile, that can be fitted by

$$p(E) = \frac{1}{\sqrt{2\pi\langle E\rangle^2}} Exp\left(-\frac{(E-E_0)^2}{2\langle E\rangle^2}\right) \qquad (2)$$

where E is the internal (potential) energy of the protein, E_0 is its average value, and $\langle E\rangle$ is the energy standard deviation. According to the foundations of statistical physics, at thermal equilibrium, the potential energy of a stable protein configuration will feature a single Gaussian distribution, as also evidenced in Figure S1D for both proteins.

References

[1] K. A. Jepsen, and I. A. Solov'yov. On binding specificity of (6-4) photolyase to a T(6-4)T DNA photoproduct. European Physical Journal D 71, 155 (2017).

[2] A. Reese, N. H. List, J. Kongsted, and I. A. Solov'yov. How Far Does a Receptor Influence Vibrational Properties of an Odorant? PLoS ONE 11, pp. 10.1371/journal.pone.0152345 (2016).

Figure S1

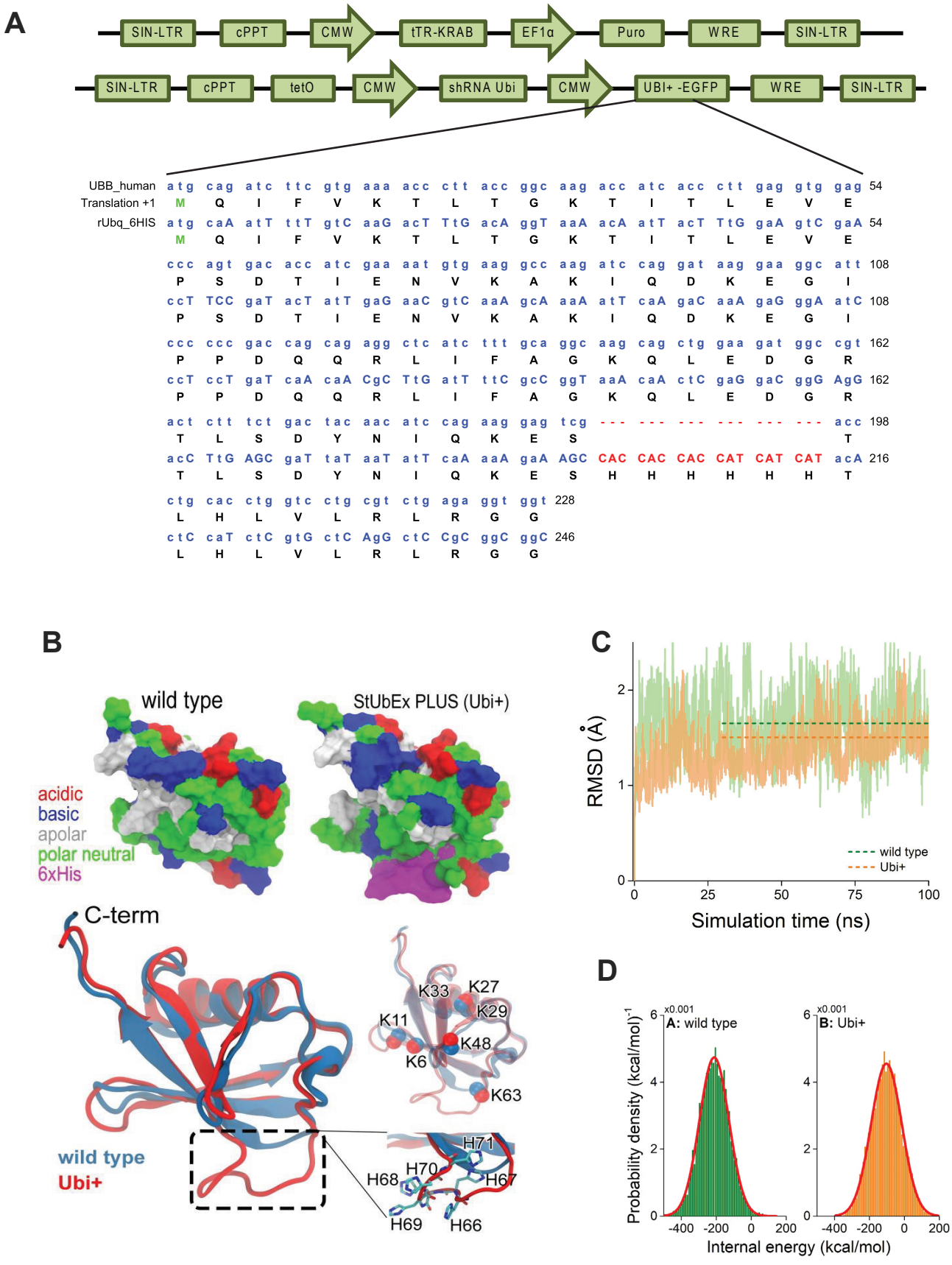

A
SIN-LTR
cPPT
CMW
tTR-KRAB
EF1α
Puro
WRE
SIN-LTR
SIN-LTR
cPPT
tetO
CMW
shRNA Ubi
CMW
UBI+ -EGFP
WRE
SIN-LTR
UBB_human atg cag atc ttc gtg aaa acc ctt acc ggc aag acc atc acc ctt gag gtg gag 54
Translation +1 M Q I F V K T L T G K T I T L E V E
rUbq_6HIS atg caA atT ttT gtC aaG acT TtG acA ggT aaA acA atT acT TtG gaA gtC gaA 54
M Q I F V K T L T G K T I T L E V E
ccc agt gac acc atc gaa aat gtg aag gcc aag atc cag gat aag gaa ggc att 108
P S D T I E N V K A K I Q D K E G I
ccT TCC gaT acT atT gaG aaC gtC aaA gcA aaA atT caA gaC aaA gaG ggA atC 108
P S D T I E N V K A K I Q D K E G I
ccc ccc gac cag cag agg ctc atc ttt gca ggc aag cag ctg gaa gat ggc cgt 162
P P D Q Q R L I F A G K Q L E D G R
ccT ccT gaT caA caA CgC TtG atT ttC gcC ggT aaA caA ctC gaG gaC ggG AgG 162
P P D Q Q R L I F A G K Q L E D G R
act ctt tct gac tac aac atc cag aag gag tcg --- --- --- --- --- --- acc 198
T L S D Y N I Q K E S T
acC TtG AGC gaT taT aaT atT caA aaA gaA AGC CAC CAC CAC CAT CAT CAT acA 216
T L S D Y N I Q K E S H H H H H H T
ctg cac ctg gtc ctg cgt ctg aga ggt ggt 228
L H L V L R L R G G
ctC caT ctC gtG ctC AgG ctC CgC ggC ggC 246
L H L V L R L R G G
B
wild type
StUbEx PLUS (Ubi+)
acidic
basic
apolar
polar neutral
6xHis
C-term
K33
K27
K29
K11
K48
K6
K63
H71
H70
H68
H67
H69
H66
wild type
Ubi+
C
RMSD (Å)
0
1
2
0
25
50
75
100
Simulation time (ns)
wild type
Ubi+
D
Probability density (kcal/mol)-1
x0.001
A: wild type
B: Ubi+
0
2
4
6
-400
-200
0
200
Internal energy (kcal/mol)

Figure S2

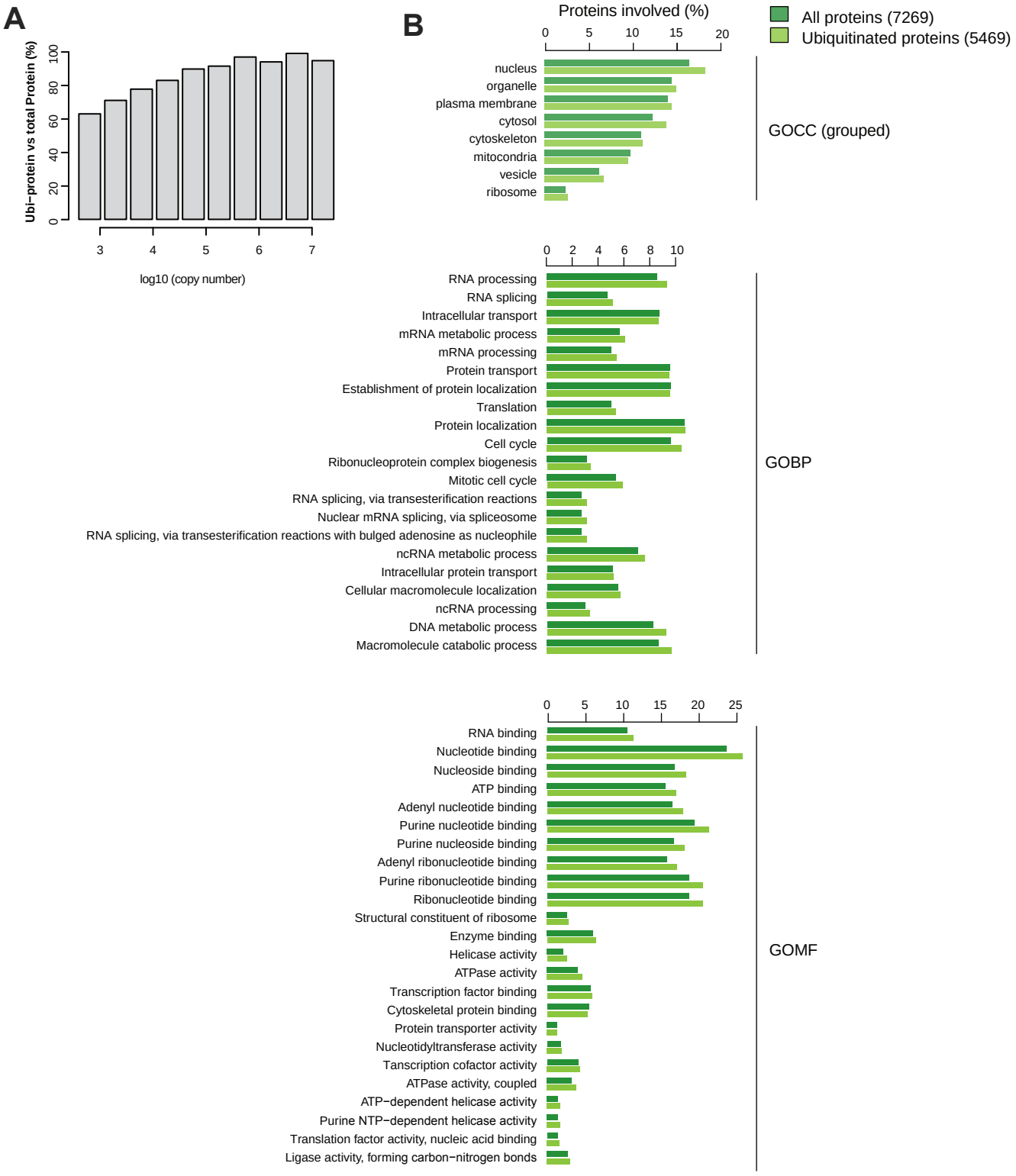
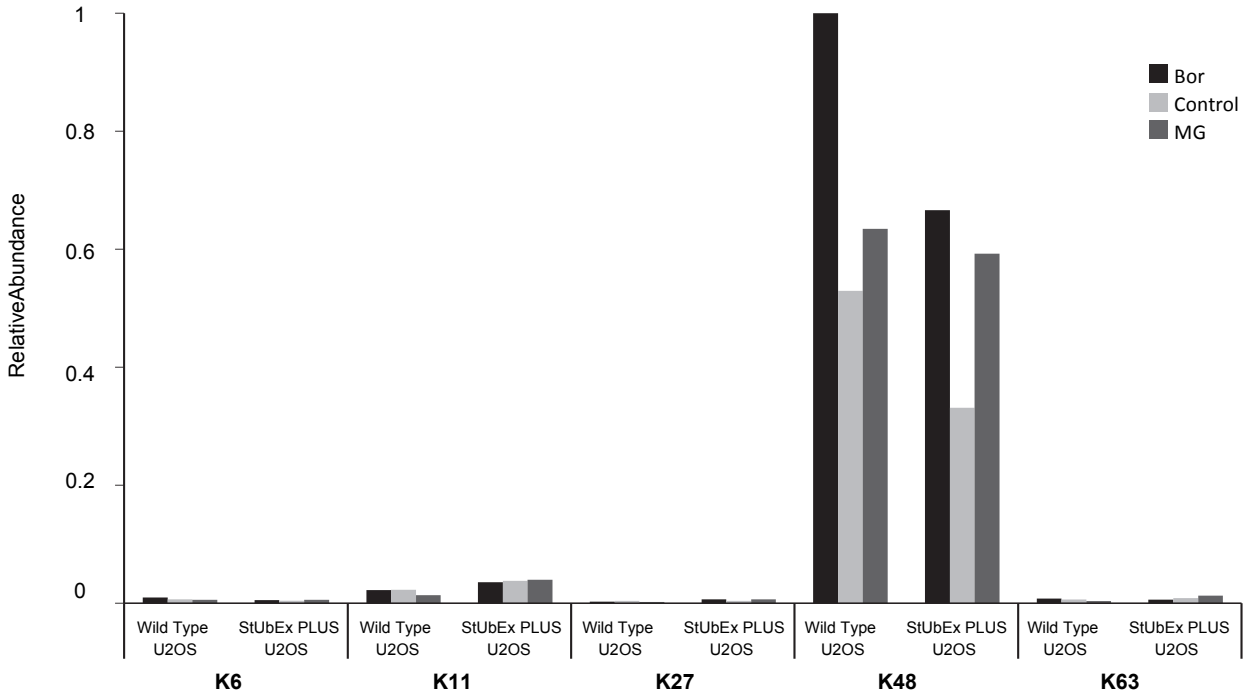

A
Ubi−protein vs total Protein (%)
0
20
40
60
80
100
3
4
5
6
7
log10 (copy number)
B
Proteins involved (%)
0
5
10
15
20
All proteins (7269)
Ubiquitinated proteins (5469)
nucleus
organelle
plasma membrane
cytosol
cytoskeleton
mitocondria
vesicle
ribosome
GOCC (grouped)
0
2
4
6
8
10
RNA processing
RNA splicing
Intracellular transport
mRNA metabolic process
mRNA processing
Protein transport
Establishment of protein localization
Translation
Protein localization
Cell cycle
Ribonucleoprotein complex biogenesis
Mitotic cell cycle
RNA splicing, via transesterification reactions
Nuclear mRNA splicing, via spliceosome
RNA splicing, via transesterification reactions with bulged adenosine as nucleophile
ncRNA metabolic process
Intracellular protein transport
Cellular macromolecule localization
ncRNA processing
DNA metabolic process
Macromolecule catabolic process
GOBP
0
5
10
15
20
25
RNA binding
Nucleotide binding
Nucleoside binding
ATP binding
Adenyl nucleotide binding
Purine nucleotide binding
Purine nucleoside binding
Adenyl ribonucleotide binding
Purine ribonucleotide binding
Ribonucleotide binding
Structural constituent of ribosome
Enzyme binding
Helicase activity
ATPase activity
Transcription factor binding
Cytoskeletal protein binding
Protein transporter activity
Nucleotidyltransferase activity
Tanscription cofactor activity
ATPase activity, coupled
ATP−dependent helicase activity
Purine NTP−dependent helicase activity
Translation factor activity, nucleic acid binding
Ligase activity, forming carbon−nitrogen bonds
GOMF

Figure S3



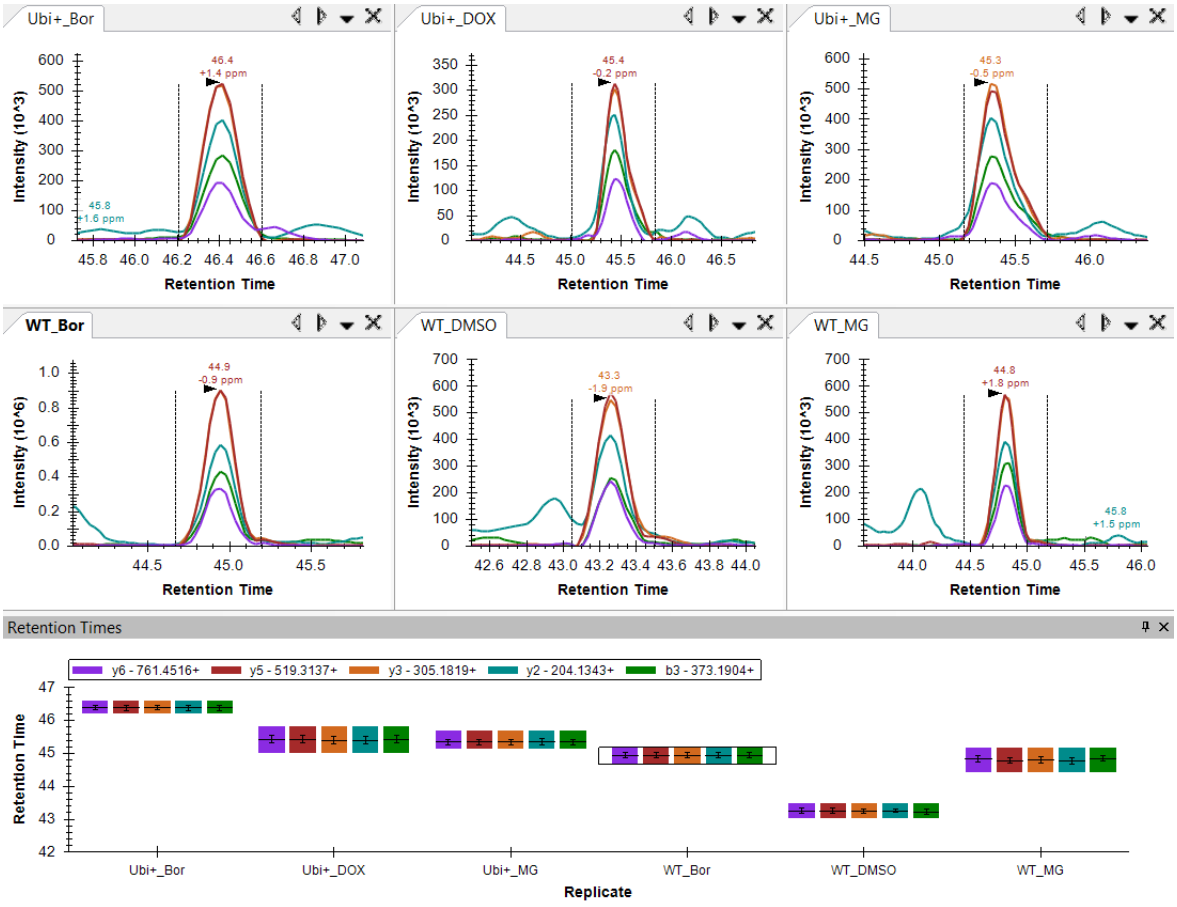
PRM Transition List

Peptide	Peptide Sequence	m/z	Fragment	Peak Rank
K6 Chain	MQIFVK(gg)TLTGK	460.6 (3+)	precursor	1
		519.31	y5	2
		305.18	y3	3
		204.13	y2	4
		373.19	b3	5
		761.45	y6	6
K11 Chain	TLTGK(gg)TITLEVEPSDTIENVK	801.43 (3+)	precursor	1
		1002.51	y9	2
		1131.55	y10	3
		1172.65	b10	4
		905.46	y8	5
		1230.62	y11	6
K27 Chain	TITLEVEPSDTIENVK(gg)AK	701.04 (3+)	precursor	1
		1315.69	y11	2
		802.44	y6	3
		460.29	y3	4
		673.40	y5	5
		915.53	y7	6
K48 Chain	LIFAGK(gg)QLEDGR	487.6 (3+)	precursor	1
		717.35	y6	2
		589.29	y5	3
		476.21	y4	4
		1016.51	y8	5
		347.17	y3	6
K63 Chain	TLSDYNIQK(gg)ESTLHLVLR	561.81 (4+)	precursor	1
		637.41	y5	2
		500.36	y4	3
		387.27	y3	4
		938.58	y8	5
		1067.62	y9	6
Actin, Cytoplasmic (for normalization)	GYSFTTTAER	566.77 (2+)	precursor	1
		678.34	y6	2
		912.44	y8	3
		577.29	y5	4
		825.41	y7	5
		476.25	y4	6

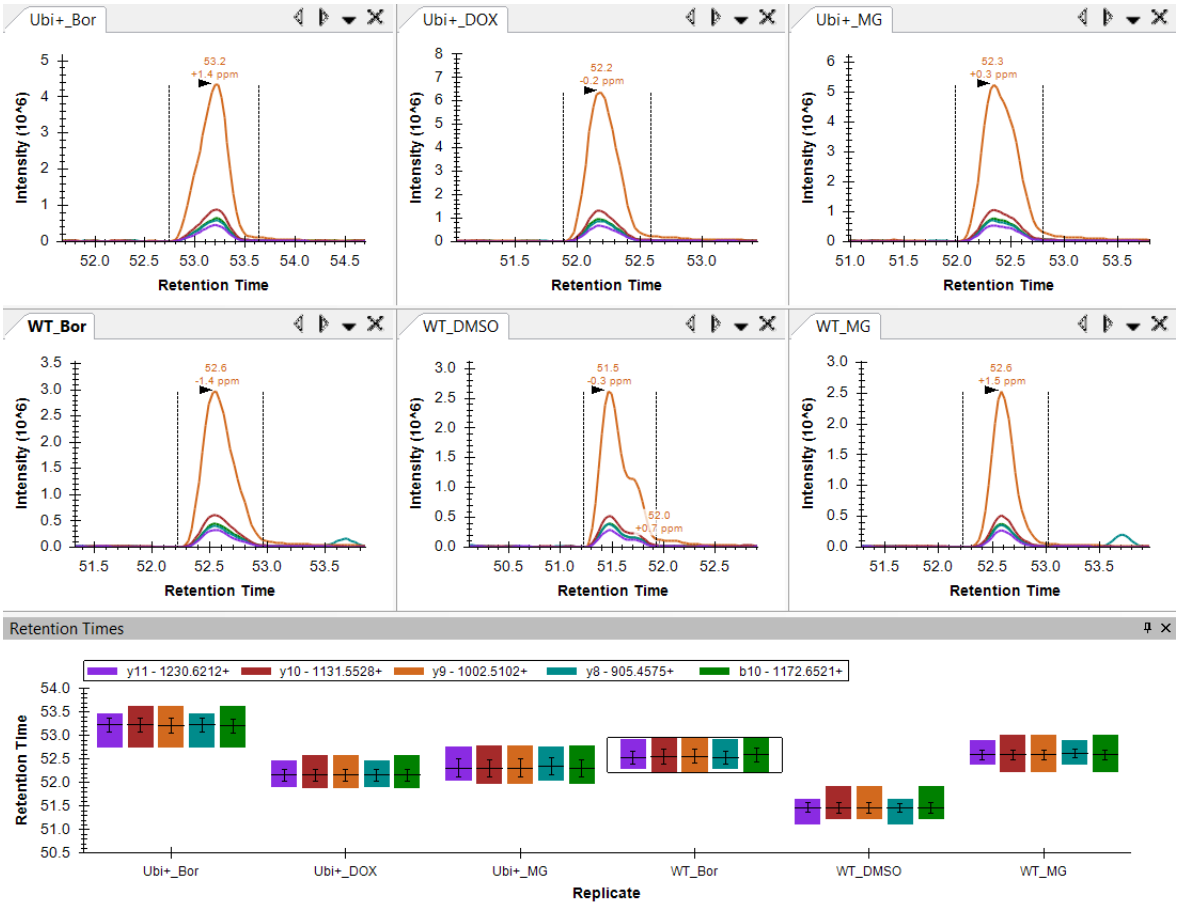
*K(gg) represents a lysine residue carrying the diGly remnant from ubiquitin modification

Primary Skyline PRM data

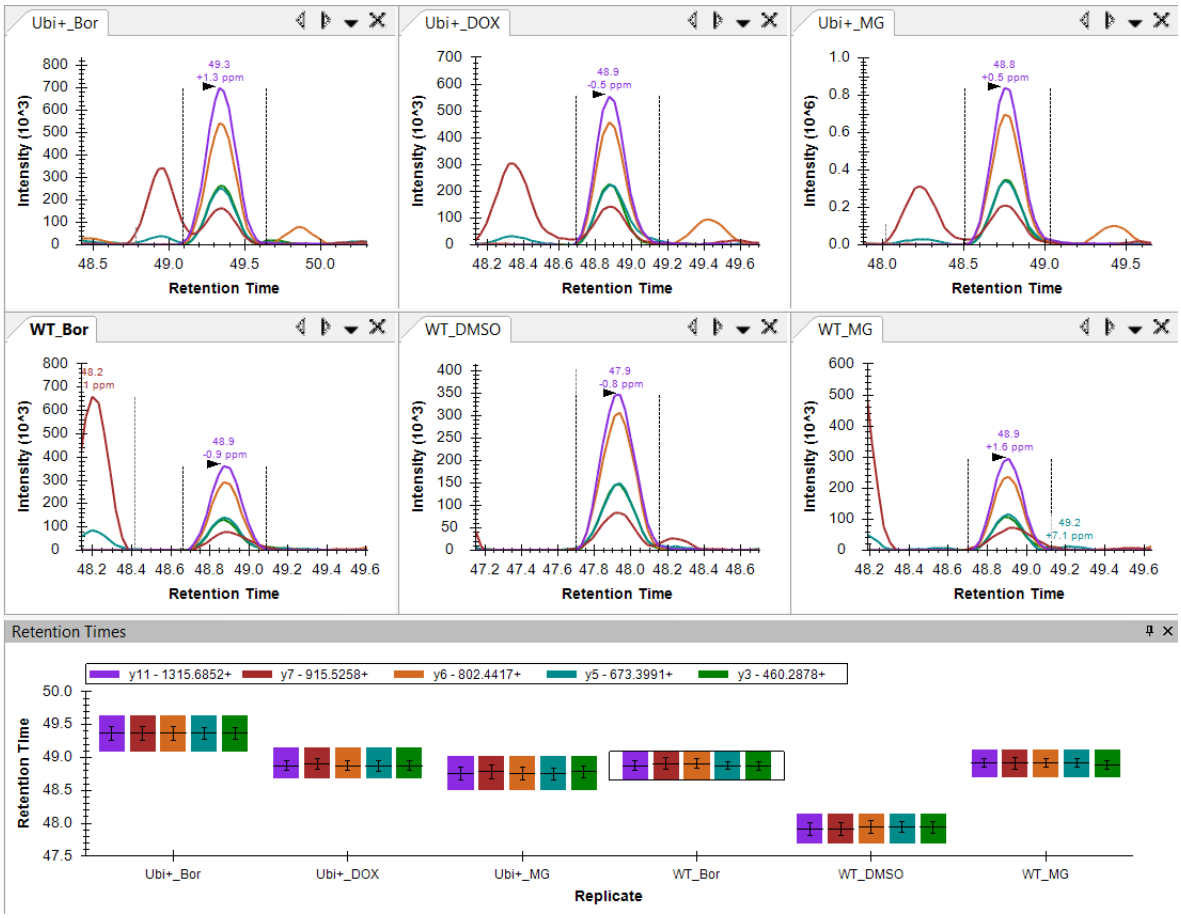
K6 chain - MQIFVK(gg)TLTGK



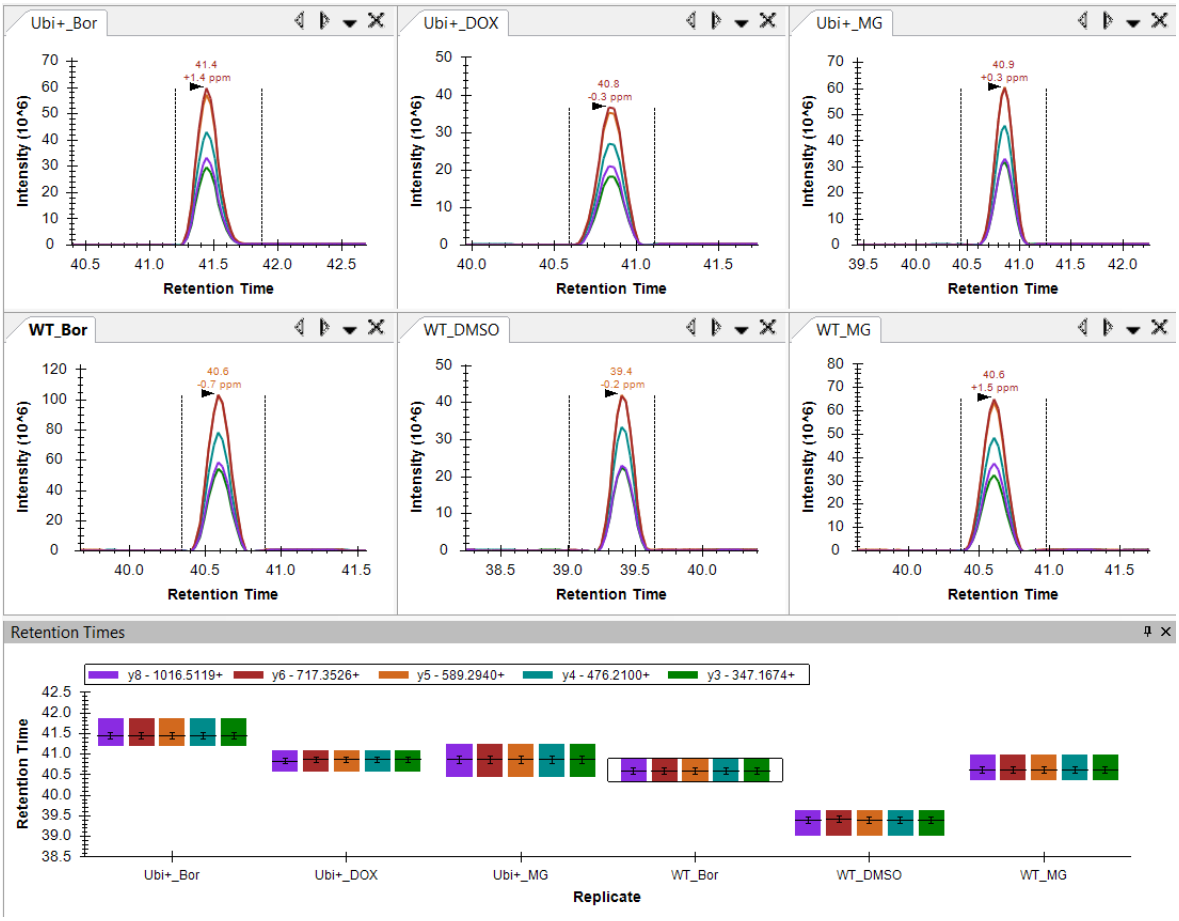
K11 chain - TLTGK(gg)TITLEVEPSDTIENVK



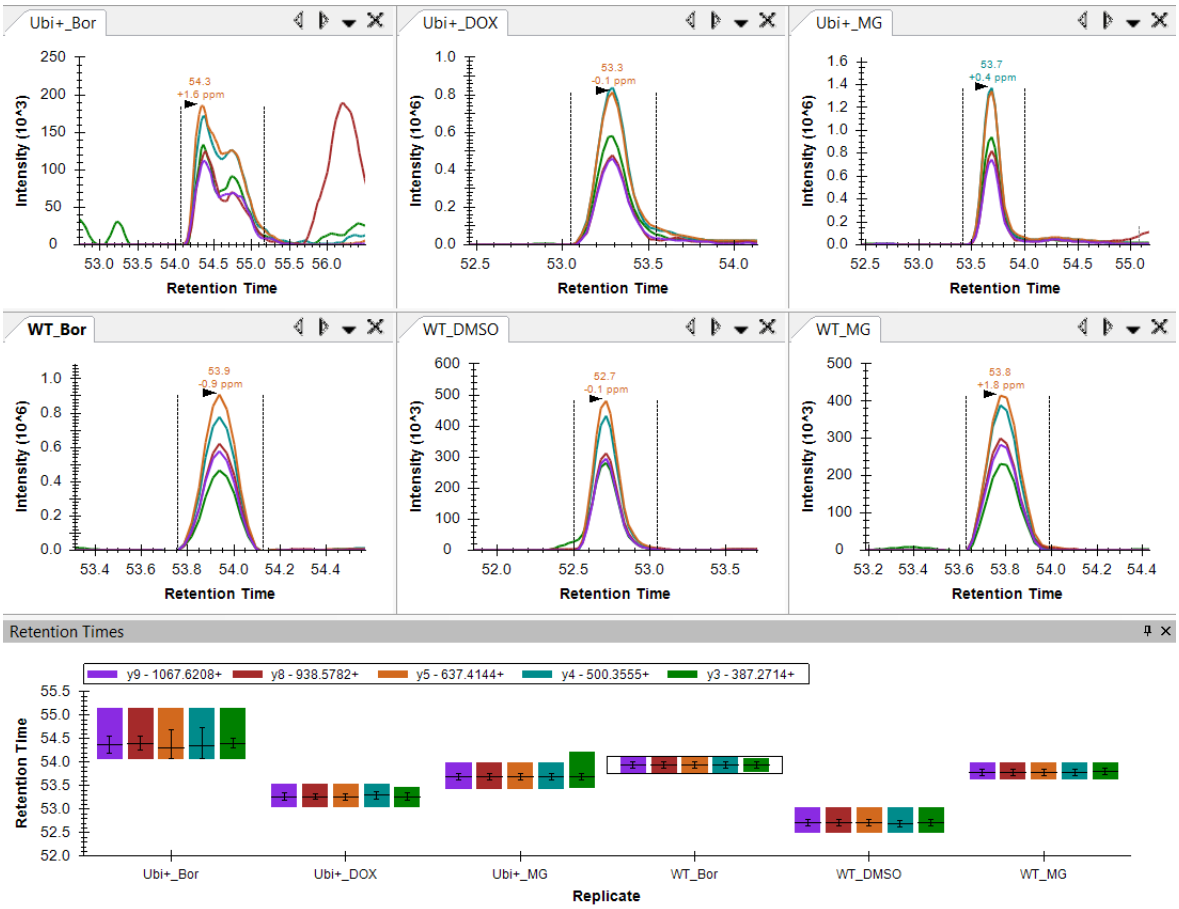
K27 chain - TITLEVEPSDTIENVK(gg)AK



K48 chain - LIFAGK(gg)QLEDGR



K63 chain - TLSDYNIQK(gg)ESTLHLVLR



Actin - GYSFTTTAER