

Global, Site Specific Identification and Quantification of Endogenous SUMO Modifications under Native Conditions

Hongbo Gu¹, Xiaoying Jia¹, Jianmin Ren¹, Elizabeth Komives², Matthew P. Stokes¹

Cell Signaling Technology, Inc., Danvers, MA 01923
Department of Chemistry & Biochemistry, UCSD, La Jolla, CA 92093

INTRODUCTION

The family of Small Ubiquitin-like Modifier (SUMO) proteins in humans includes four distinct genes: SUMO1, SUMO2/3 (which differ by only three residues) and SUMO4. SUMOylation regulates the function of various proteins by reversible, covalent, isopeptide-bond attachment between the C-terminus of SUMO and a free ε-amino group of the target protein's lysine, similar to Ubiquitin (Ub). Ub conjugation mainly targets proteins for degradation by the proteasome, but has also been implicated in DNA repair, receptor signaling and cell communication. The function of SUMO conjugation is less well-understood, but SUMOylated proteins are involved in gene expression, DNA repair, nuclear import, heat shock, cell motility, and lipid metabolism. SUMO targets are generally low-abundance proteins, and the amount of the modification at steady-state is also low. In recent years several groups have developed methods for analysis of SUMOylated proteins. However, many of these involve overexpression of mutant SUMO sequences. Therefore, methods that allow proteome-level identification of endogenous SUMOylation sites are needed.

We recently described the novel application of wild-type α-lytic protease (WaLP) to proteome digestion for shotgun proteomics. Although relatively relaxed specificity was observed, WaLP prefers to cleave after threonine residues and never cleaves after arginine (Fig. 1b). We show here that WaLP cleaves at the C-terminal TGG sequence (in SUMO 2/3 and 4) leaving a SUMO-remnant KGG at the position of SUMO attachment. The resulting KGG-containing peptides can then be identified using methods already developed for Ub-profiling (Fig. 1d). The method allows identification of SUMO attachment sites under completely native conditions using the Ub-profiling workflow by simply substituting WaLP for trypsin. The same sample can be subjected to analysis of both Ub-attachment and SUMO-attachment by digesting the sample with either trypsin or WaLP respectively.

METHODS

Hela cells were heat shocked and lysed in urea buffer. Protein digestion was performed in 2M urea by alpha-lytic protease (WaLP) that preferentially cut at the C-terminus of T, S, A and V resulting a Glyly remnant left on the lysine residue previously carrying SUMO-1 or SUMO-2/3. Digested peptides were subject to immunoprecipitation (IAP) using the K-ε-GG remnant antibody and LC-MS/MS analyzed for identification and quantification of SUMOylation sites. To validate the KGG sites identified by the method, specific SUMO-proteases were used to remove endogenous SUMO tails from modified proteins, which were then digested by WaLP and trypsin, respectively. Label-free quantification was applied to determine the quantitative changes of identified KGG peptides from WaLP and trypsin digest in parallel.

CONCLUSIONS

There were about 1,100 non-redundant KGG peptides identified from 5mg Hela lysate digested with WaLP and enriched with KGG antibody IAP including known SUMOylated proteins including TRIM28, TRIM33, SUMO1 and SUMO-2/3. By using specific SUMO-proteases (SENP 1 and 2), both SUMO-1 and SUMO-2/3 were efficiently removed *in-vitro*, as confirmed by western blots, while the total ubiquitination level was not affected by SUMO-proteases. We compared the quantitative changes of identified KGG peptides from WaLP and trypsin digest with and without SUMO-protease treatment. 612 out of 724 WaLP-digested KGG peptides (SUMOylated sites) decreased in abundance, while only 16 out of 9,031 trypsin digested peptides (Ubiquitinated sites) decreased with SUMO protease treatment. The results validated that the KGG sites identified by our method truly come from SUMOylation instead of ubiquitination.

Protein function classification placed proteins involved in 1) transcriptional regulation; 2) translation; 3) RNA processing; and 4) DNA binding and repair as top protein groups modified by SUMOylation that indicated a potential correlation between SUMOylation and epigenetics. More importantly, more than 2/3 of the KGG sites identified by our method are novel SUMOylation sites, providing a rich resource for the field of SUMOylation.

REFERENCES

Meyer, J. G. et al. Expanding proteome coverage with orthogonal-specificity α-lytic proteases. *Mol Cell Proteomics* 13, 823-835, doi:10.1074/mcp.M113.034710 (2014).
Udeshi, N. D., Mertins, P., Svinkina, T. & Carr, S. A. Large-scale identification of ubiquitination sites by mass spectrometry. *Nat Protoc* 8, 1950-1960 (2013).
Becker, J. et al. Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat Struct Mol Biol*, 525-531, doi:10.1038/nsmb.2526 (2013).

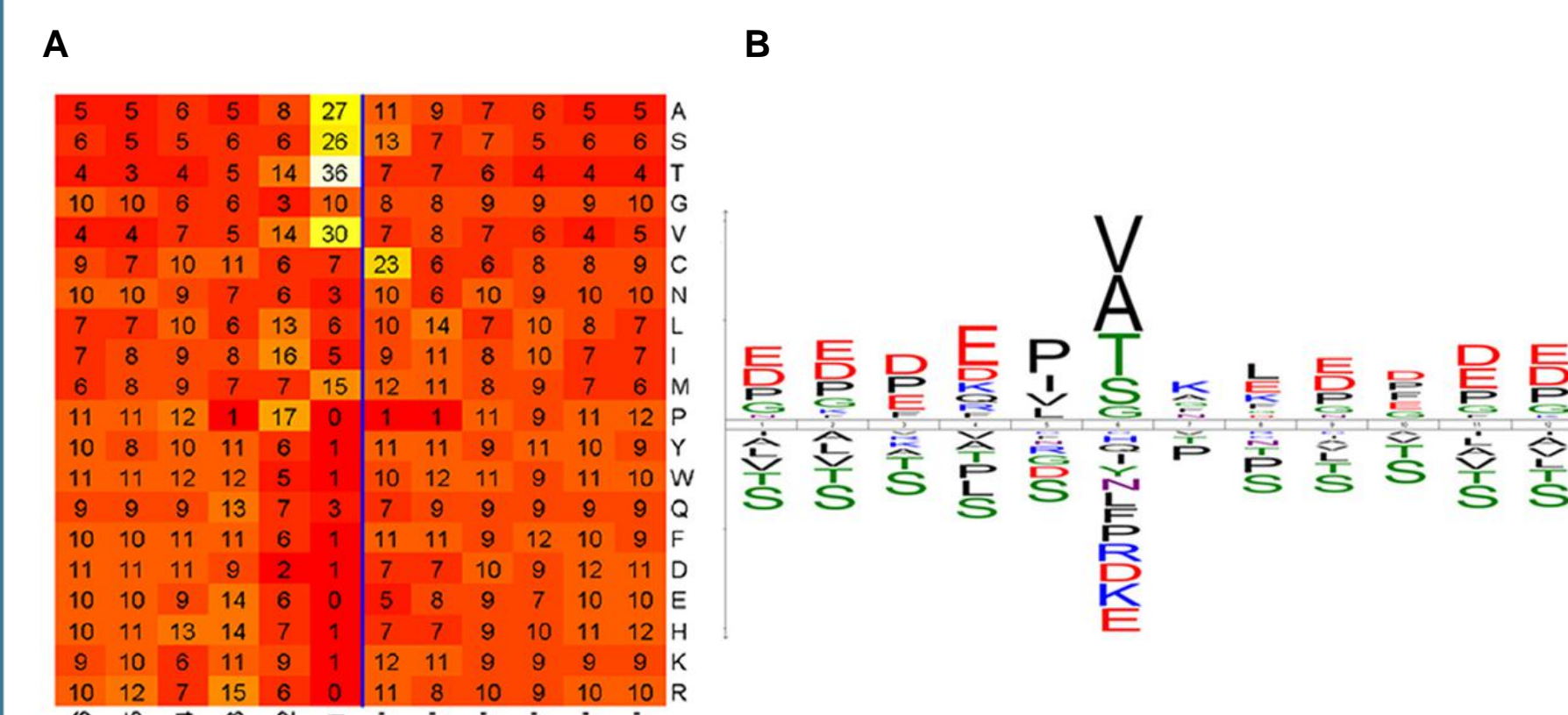


Figure 1: Heat map summarizing the analysis of the occurrence of each amino acid at each position in the identified peptide from the WaLP digest in reference 1. (A), WaLP peptides, counts normalized for the occurrence of each amino acid at each position. (B), iceLogo depicting the enrichment and depletion of specific amino acids at each position in the WaLP peptides, with residues colored according to property (acidic, red; basic, blue; hydrophobic, black; small/neutral, green). WaLP yields peptides with the following P1 (C-terminal) residues: A, 20%; V, 20%; S, 16%; T, 16%; G, 8%; L, 6%. The cleavage site is marked by a vertical line in the heat map.

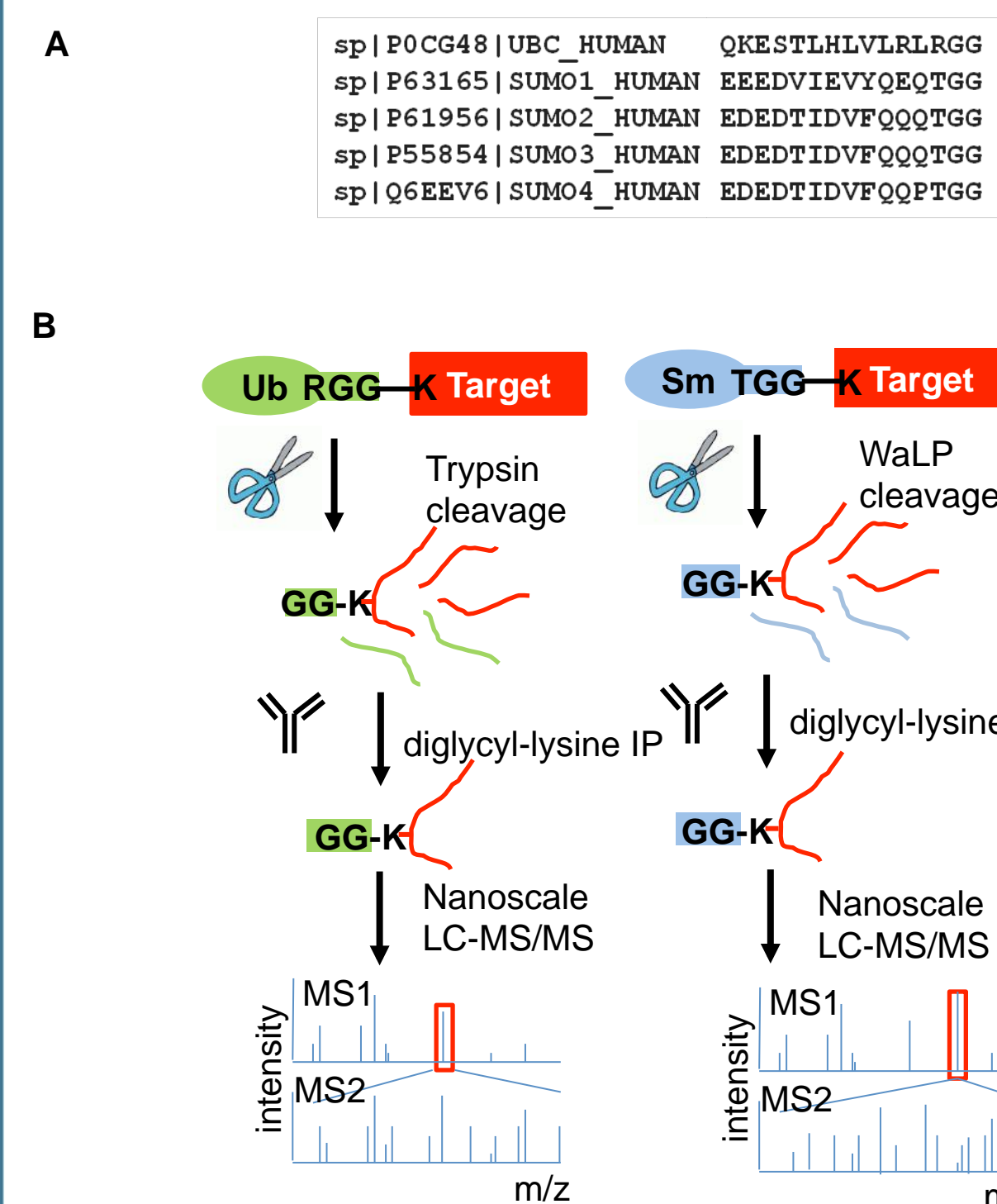


Figure 2: A Strategy for mapping endogenous SUMO1/2/3 attachment sites. (A), C-terminal sequence alignment of mature human SUMO 1-4 and Ubiquitin. (B), Cartoon showing ubiquitin-remnant and SUMO-remnant mapping strategy. Proteins modified by ubiquitin digested with trypsin and modified by SUMO digested with WaLP leave a diglycine attached to the ε-amino of the lysine where ubiquitin or SUMO was attached. An antibody specific for the KGG the remnant is used to enrich peptides from ubiquitinated sites and SUMOylated sites that are then identified by nanoLC-MS/MS.

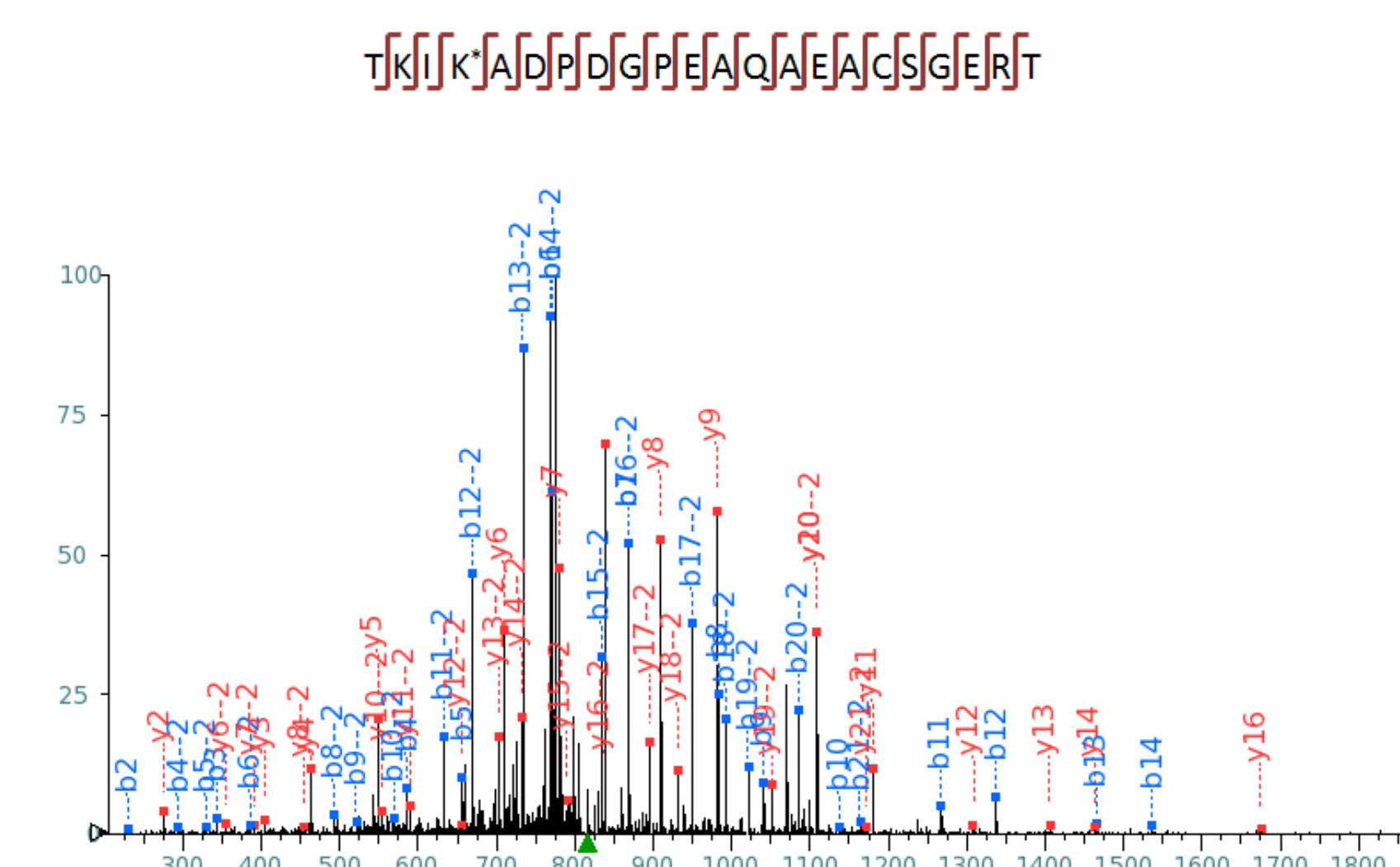


Figure 3: A known SUMOylation site (K5 of NHP2_Human, Q9NX24) was confidently identified from WaLP digest of Hela cells by KGG antibody enrichment and LC-MS/MS. Green triangle indicates the m/z of the precursor peptide.

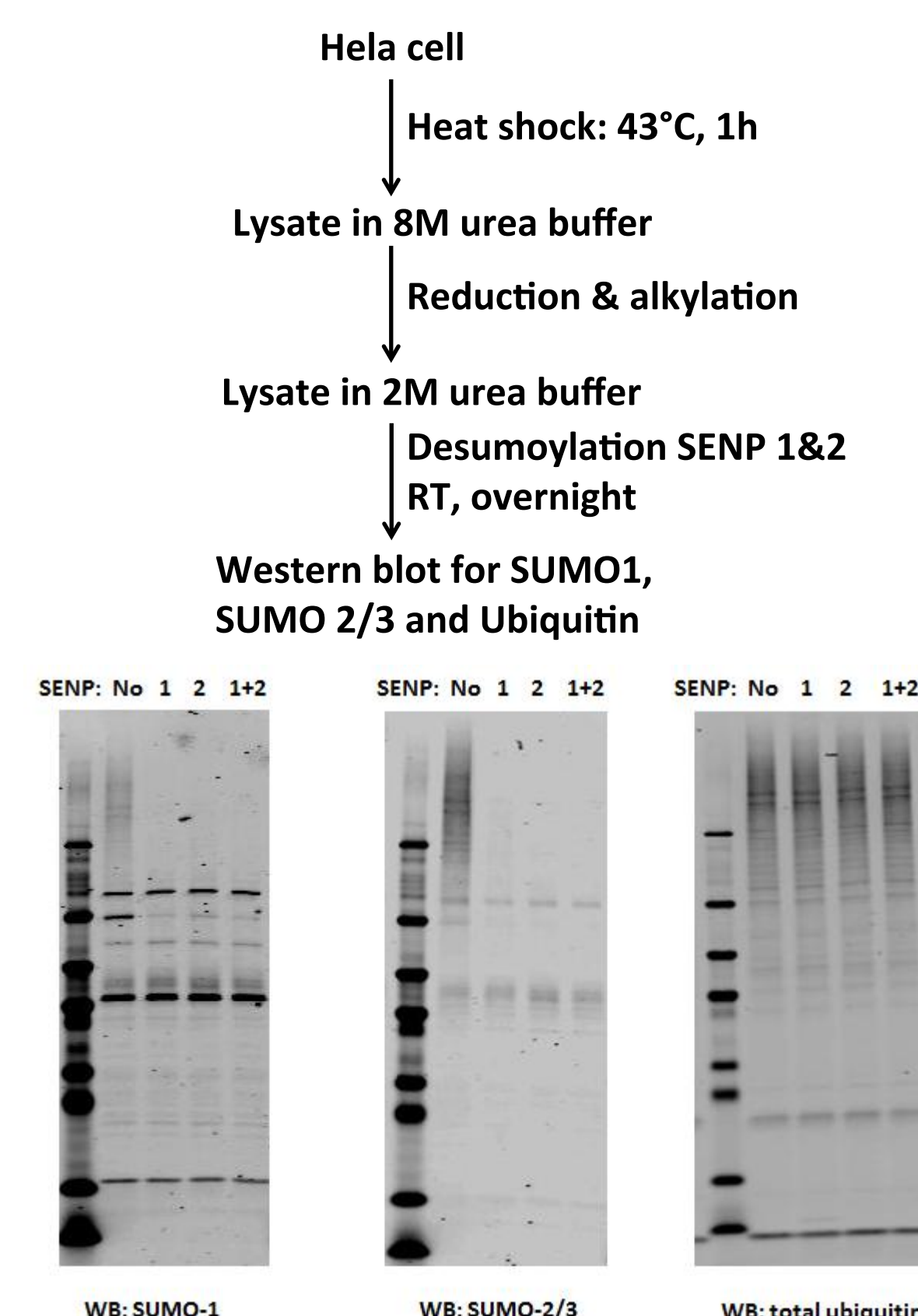


Figure 4: A strategy to validate identified SUMOylation sites. Endogenous SUMO attachment was removed by SUMO proteases SENP 1 and SENP 2. Endogenous ubiquitin levels were not affected by SENP 1 and SENP 2 treatment.

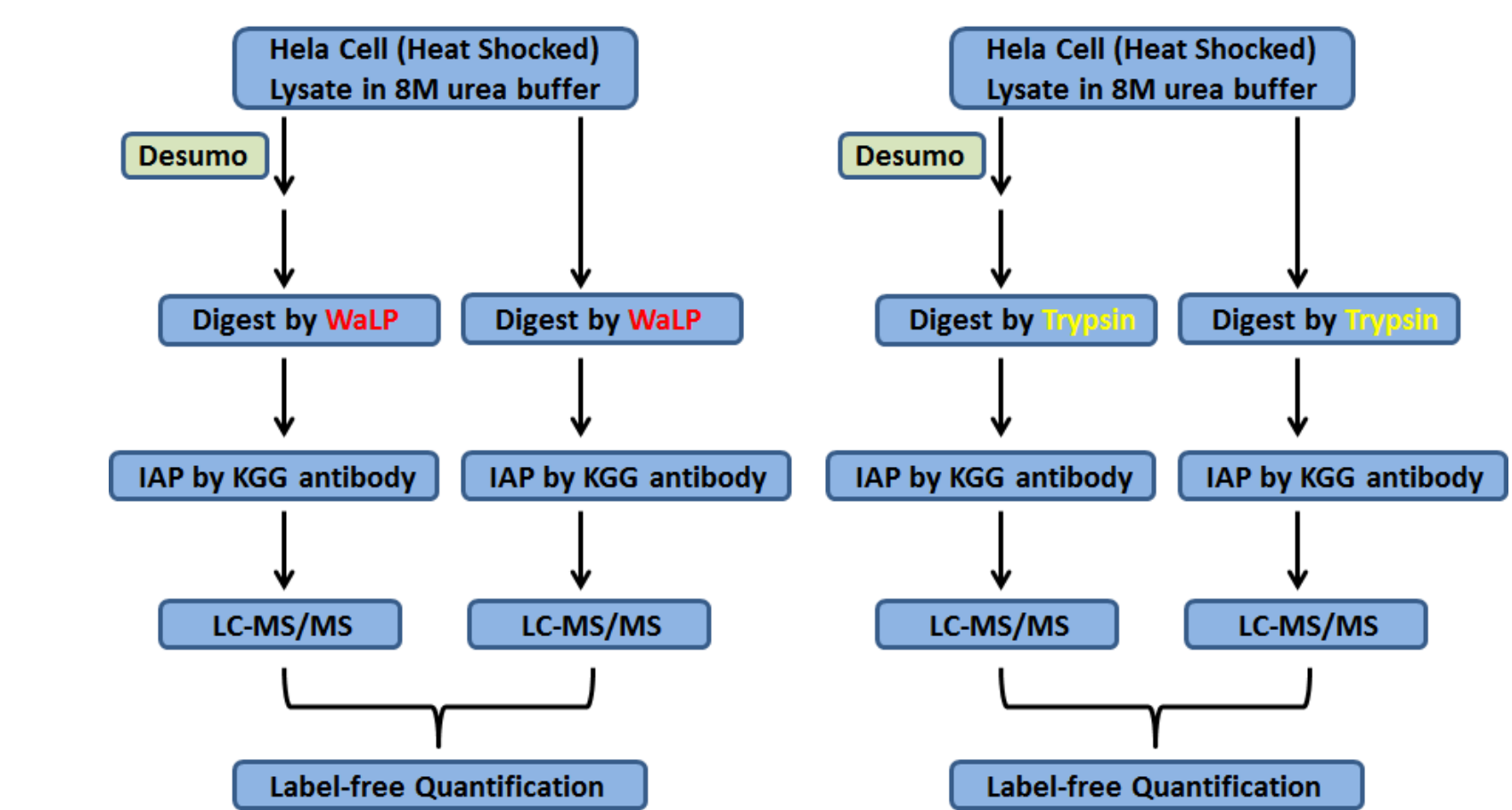


Figure 5: A parallel quantitative analysis of KGG peptides enriched from WaLP and Trypsin digest of Hela cells. Label-free quantification was applied to quantify peptides from digested treated with SENP 1+2 vs. untreated.

Sample	Protease	De-sumo	Replicate	Redundant KGG peptide	Non-redundant KGG peptide
Hela cells heat shock treated	WaLP	No	1	612	450
	WaLP	No	2	829	638
	WaLP	Yes	1	147	115
	WaLP	Yes	2	143	108
	Trypsin	No	1	7,456	5,866
	Trypsin	No	2	7,155	5,600
	Trypsin	Yes	1	7,026	5,523
	Trypsin	Yes	2	6,971	5,428

Table 1: Summary of number of identified KGG peptides from WaLP and Trypsin digest. There is a dramatic drop in the number of KGG peptides from WaLP digest treated with SUMO proteases, which is consistent with western blots showing efficient removal of endogenous SUMO.

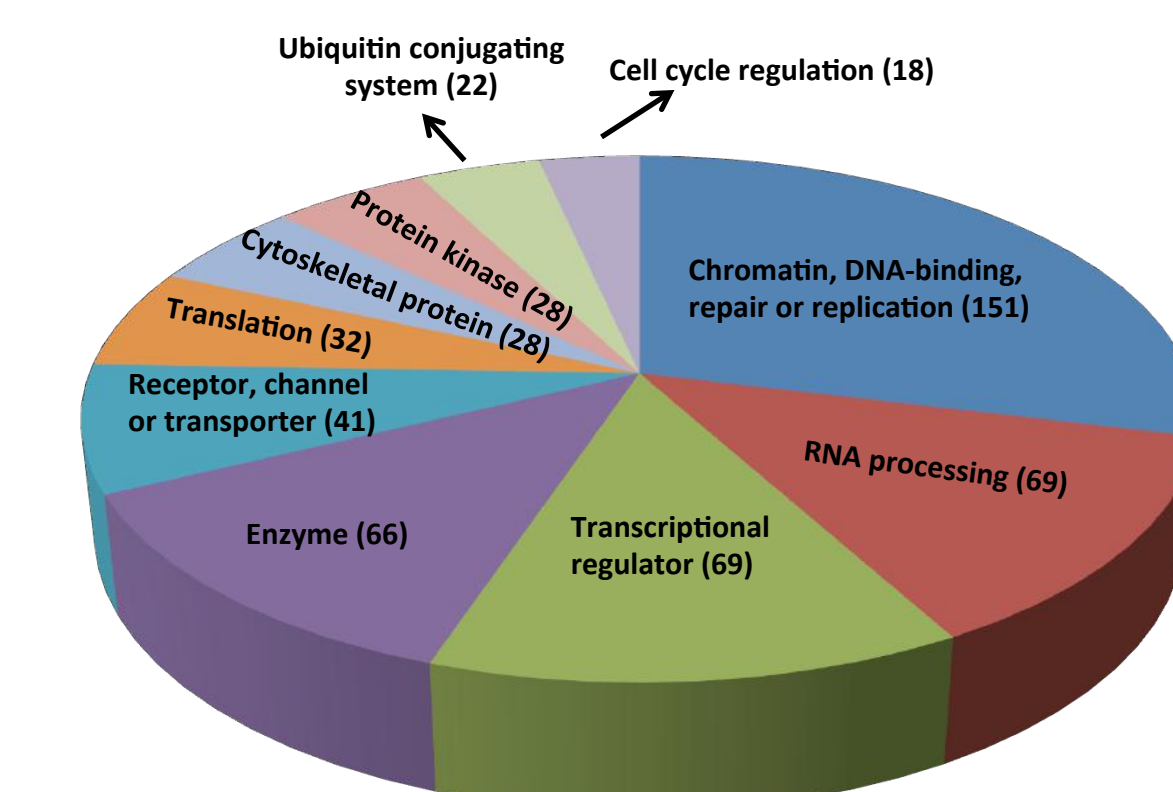


Figure 6: Classification of identified SUMOylated proteins.

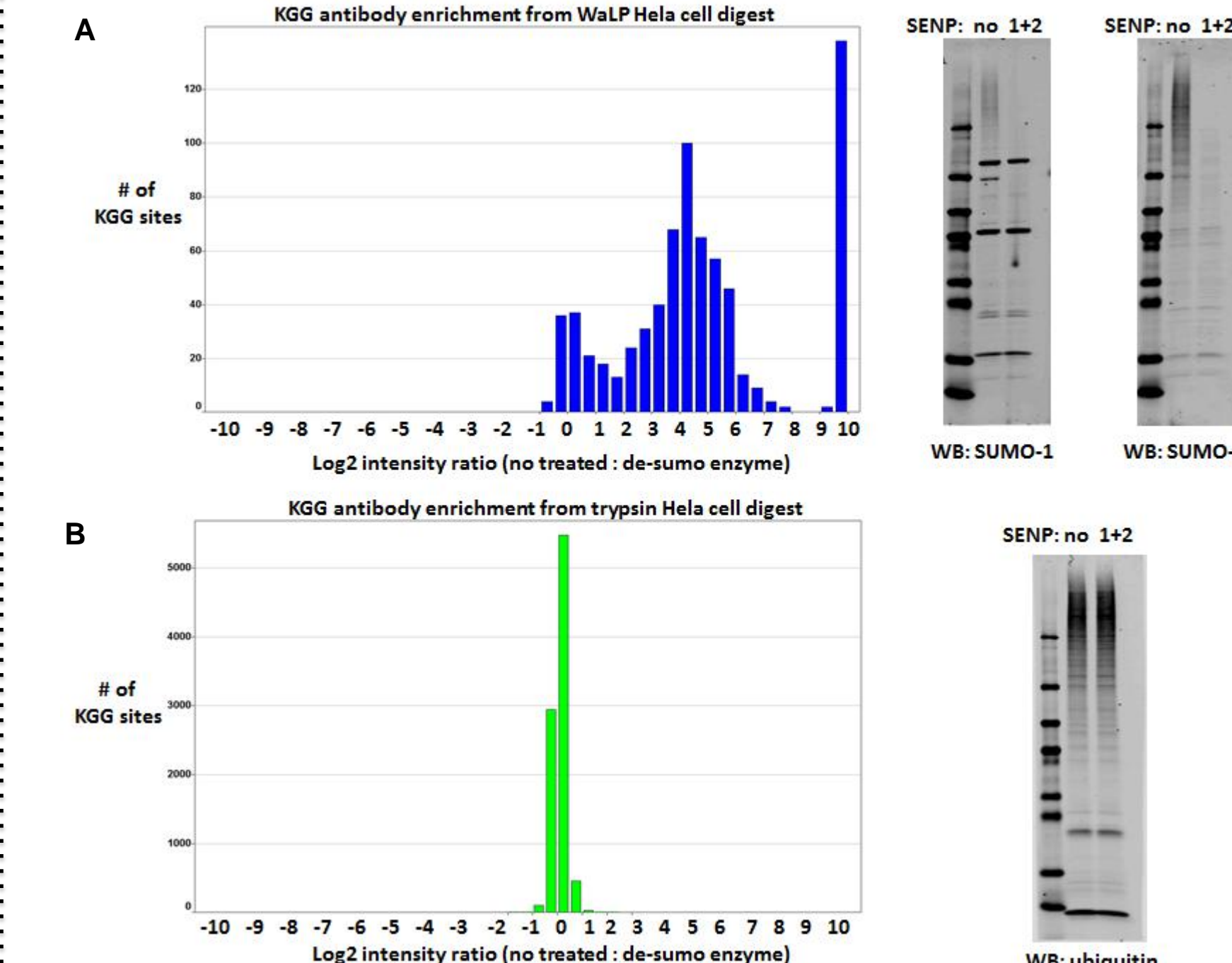


Figure 7: (A), Bar graphs of results from quantification of KGG peptides after treatment with SENP 1+2 showing the dramatic reduction of KGG-sites observed after WaLP digestion. Importantly, for over 400 sites, SENP 1+2 treatment resulted in no detectable KGG modification upon WaLP digestion. These sites were assigned as a log(2) value of 10. (B), No significant reduction in intensity of KGG sites was observed after tryptic digestion. Both quantitative results are consistent with western blots.

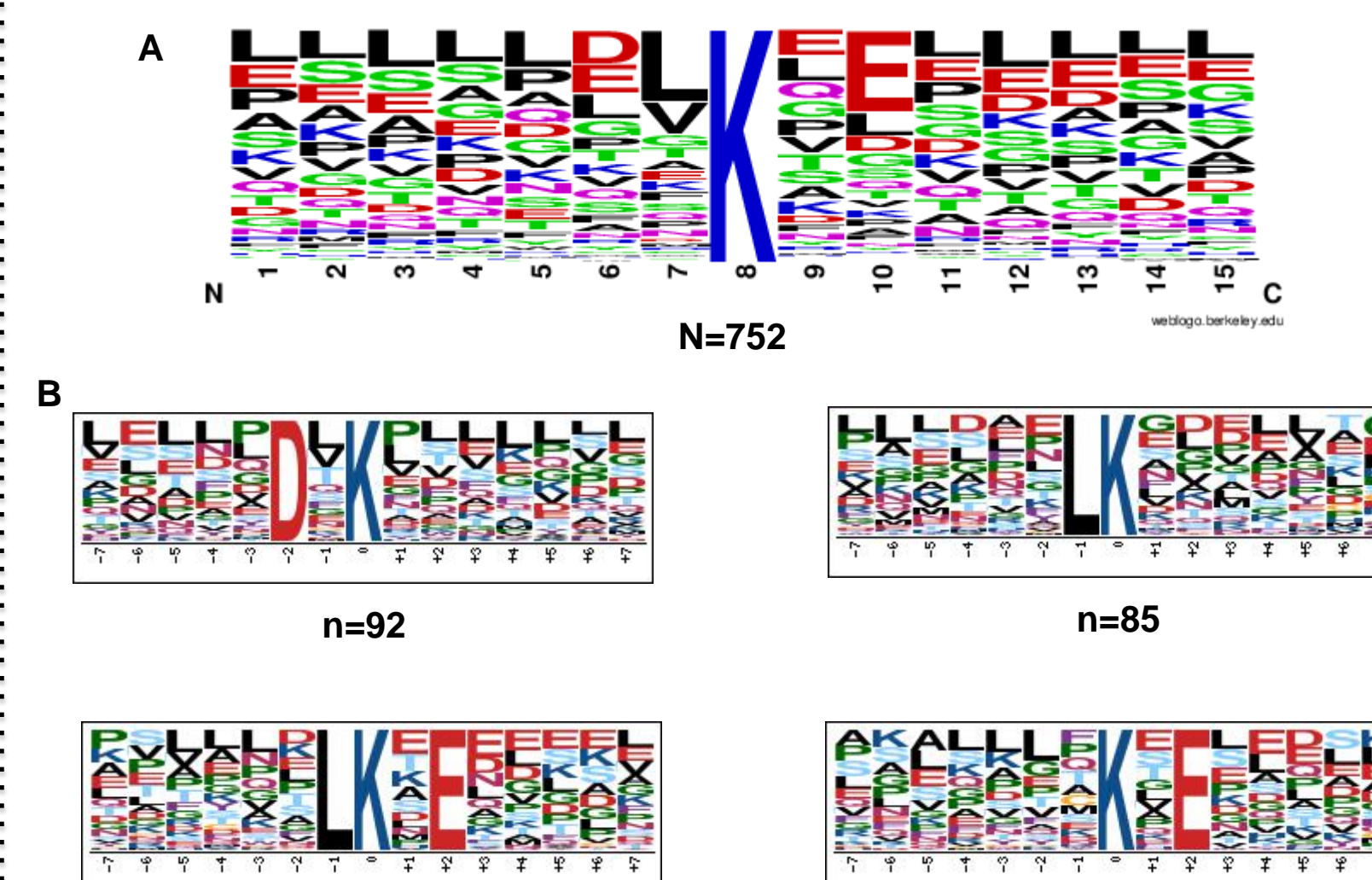


Figure 8: Motif analysis of identified KGG peptides from WaLP digest. (A), Motif logo of all identified KGG peptides (n=752). The size of the amino acid represents the relative occurrence of the amino acid at a specific position. (B), Top four consensus motif sequences.