Supporting Information

Exopeptidase assisted N- and C-terminal proteome sequencing

Andreas O. Helbig[1] & Andreas Tholey[1,*]

[1]Systematic Proteome Research & Bioanalytics, Institute for Experimental Medicine, Christian-Albrechts-Universität zu Kiel, Kiel, Germany

* to whom correspondence should be addressed:

Andreas Tholey

Systematic Proteome Research & Bioanalytics, Institute for Experimental Medicine

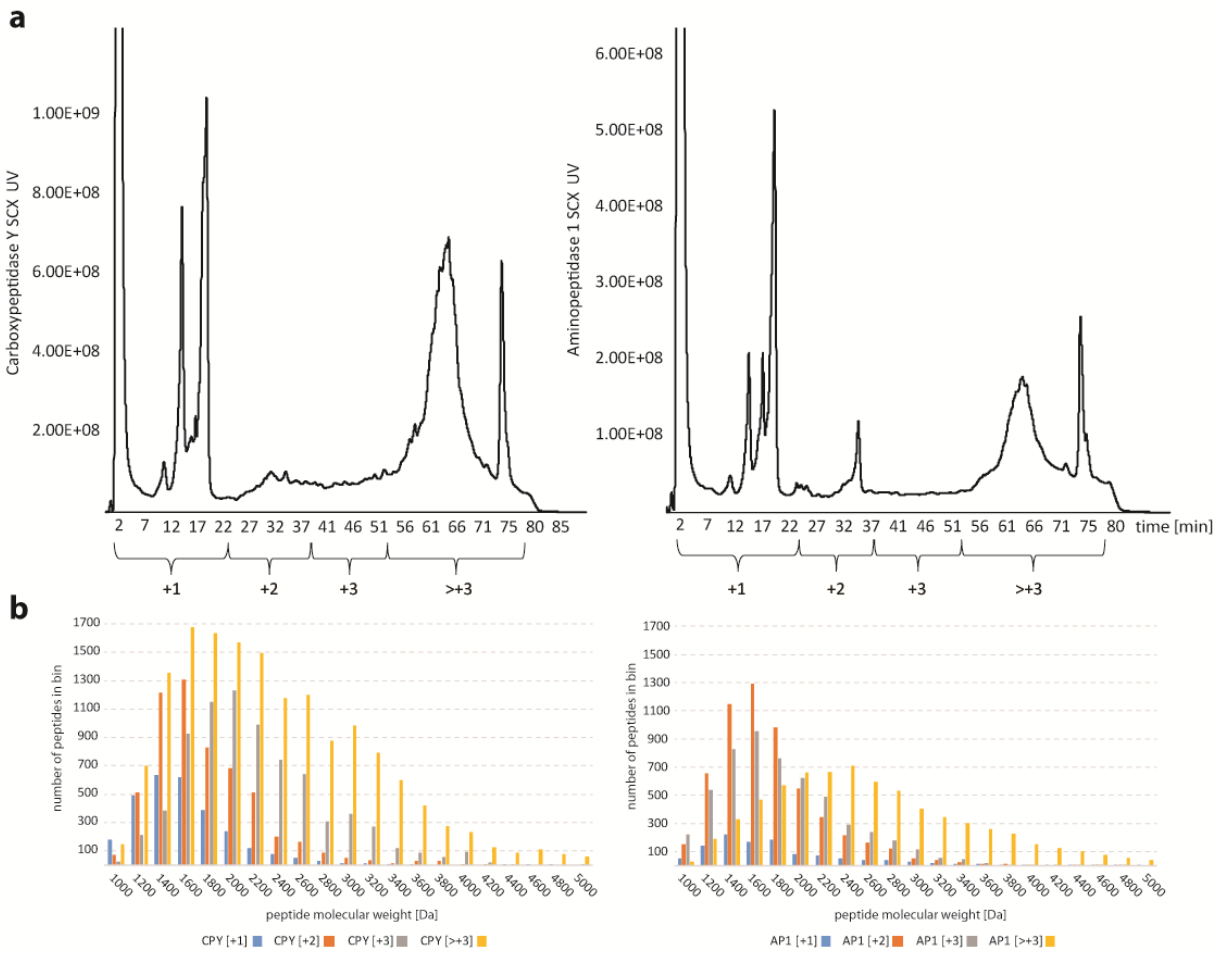
Christian-Albrechts-Universität zu Kiel

24105 Kiel, Germany

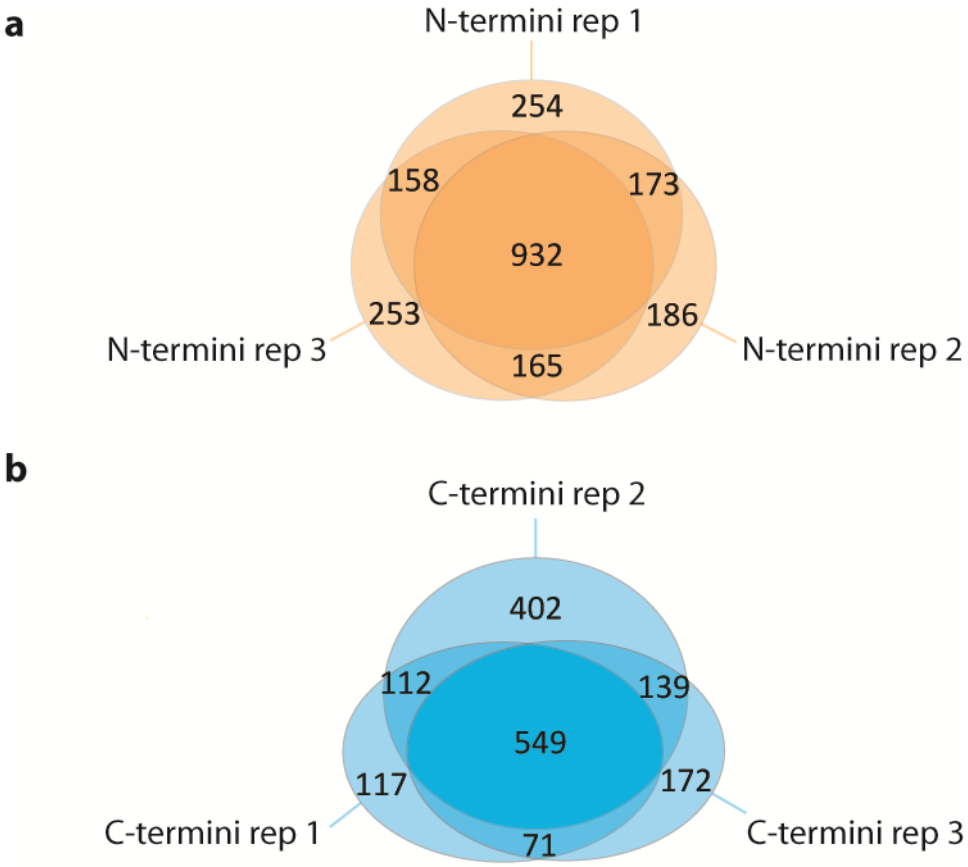
Phone: #49 (431) 500 30300; Fax: #49 (431) 500 30308

E-mail: a.tholey@iem.uni-kiel.de

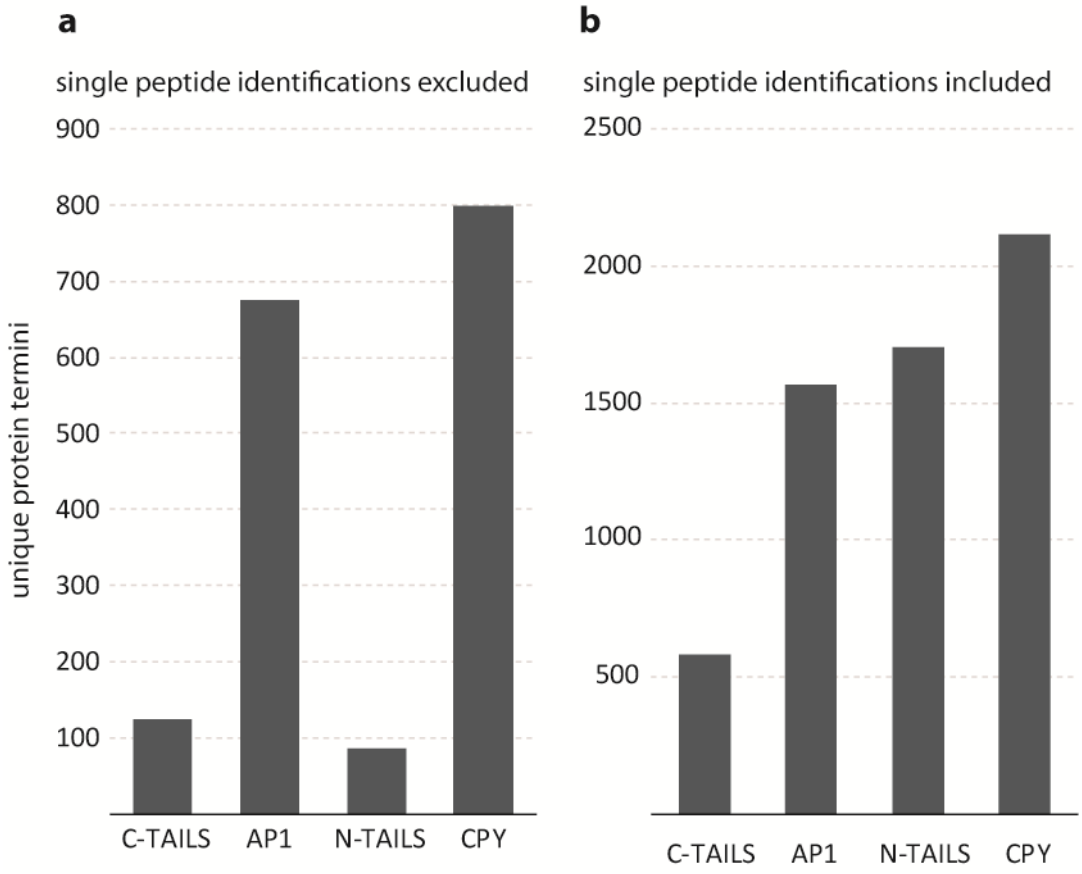
Content:



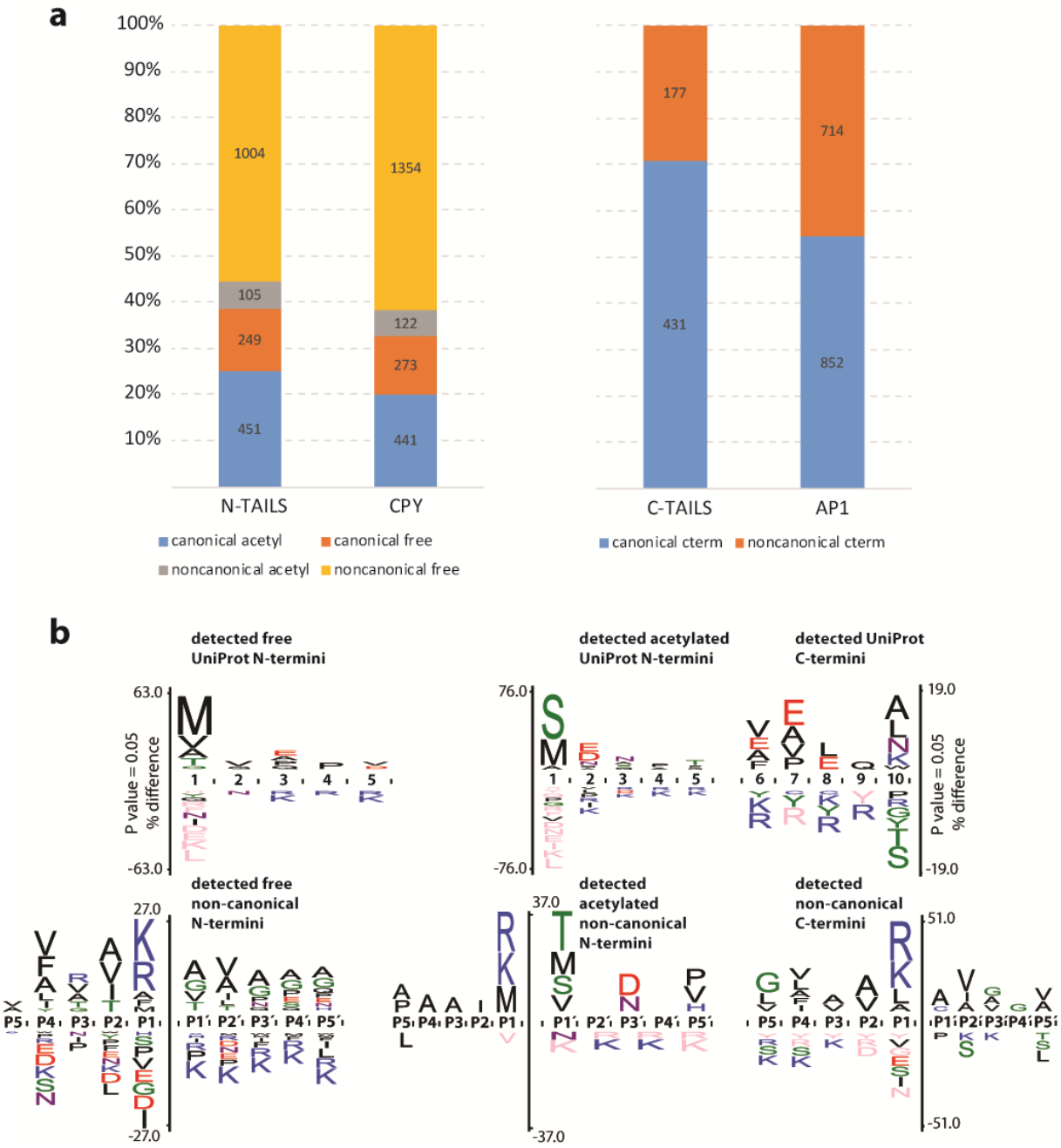
Supplementary Figure 1. (a) SCX UV traces recorded at 214 nm for CPY and AP1 digestions. Time course aliquots were combined prior to SCX chromatographic separation. Collection windows of charge-state fractions are indicated. (b) The bottom histograms show the size distribution of peptides identified in the different charge-state fractions. It can be observed that peptide size increases with higher charge state up to 5,000 Da for both CPY and AP1 digestions.



Supplementary Figure 2. Overlap of unique N- and C-termini identified in three different biological replicates of the (a) CYP and (b) AP1 digestions.

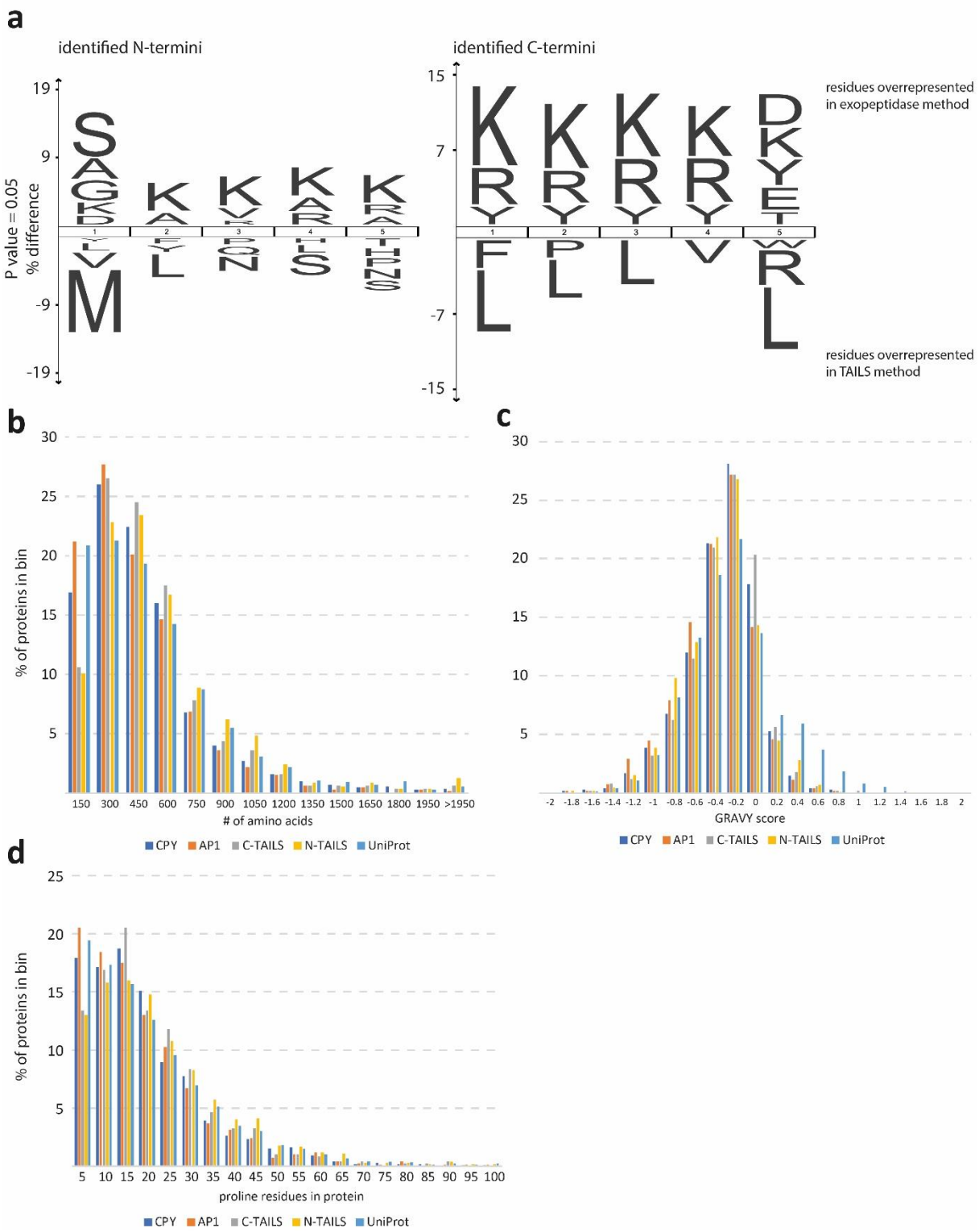


Supplementary Figure 3. Identified unique N- and C-termini from exopeptidase and N-/C-TAILS approaches. (a) When single peptide identifications are excluded, AP1 and CPY digestions yielded 6 – 9 times the number of terminal identification compared to previously established methodologies. (b) When single peptide identifications are included, exopeptidase digestions also led to increased numbers of identifications, especially in the case of C-terminomics (three-fold increase).



Supplementary Figure 4. (a) Comparison of *S. cerevisiae* protein termini detected by exopeptidase digestions and N-/C-TAILS. Distribution of canonical/non-canonical, as well as, free/acetylated N-termini was comparable between the different approaches. For C-termini, a similar ratio of canonical versus non-canonical termini was observed. (b) The sequence logos represent *S. cerevisiae* protein termini detected by N- and C-TAILS. The top row highlights detected canonical protein termini that correspond the genetically encoded proteoform as derived from UniProt. Due to ubiquitous cleavage

of the initial methionine, the penultimate residue of the annotated proteins is also considered a start point of canonical N-terminal sequences. For protein N-termini, a distinction between free and N-acetylated protein termini is made. In the bottom row, detected non-canonical termini are displayed. Non-prime sequences for N-termini and prime sequences for C-termini are indicated (free and acetylated N-termini separately shown).



Supplementary Figure 5. Comparsion of results achieved by exopeptidase-terminomics and TAILS. (a) Sequence logo comparison of *S. cerevisiae* protein N-termini (left panel) and C-termini (right panel) detected by exopeptidase digestions and N-/C-TAILS. Canonical and non-canonical termini were

combined for each method. Overrepresented amino acids are displayed for the different approaches. (b) Histogram displaying the distribution of protein sizes identified by the two terminomics approaches. The exopeptidase approach identifies more low molecular weight protein termini compared to TAILS. (c) Histogram displaying the distribution of protein hydrophobicity of proteins identified by each terminomics method. Highly hydrophobic proteins are underrepresented in all approaches. (d)) Histogram displaying the amount of proline residues in proteins identified in the used terminomics methods.

Supplementary Experimental Section

Trypsin digestion

500 µg of reduced and alkylated S. cerevisiae proteins were dissolved in 100 µl 6 M GndHCL, 50 mM ammonium bicarbonate pH 7.5. After dilution of the GndHCL to 0.6 M using 50 mM ammonium bicarbonate pH 7.5, 10 µg of trypsin were added and proteins were digested over night at 37 °C. Peptides were purified using the SepPak (Waters, UK) column system (50 mg of C_{18} material). After binding on the column, peptides were washed using 5 % formic acid and eluted with two times 100 µl 80 % and 100 % acetonitrile. Peptides were lyophilized and reconstituted in 50 µl 5 % formic acid.

CNBr digestion

500 µg of reduced and alkylated *S. cerevisiae* proteins were dissolved in 100 µl 6 M GndHCL. The sample was acidified using 100 µl 100 % acetic acid. After the addition of 50 mg cyanogen bromide, the mixture was incubated for 20 hours at room temperature in the dark. Peptides were purified using the SepPak (Waters, UK) column system (50 mg of C_{18} material). After binding on the column, peptides were washed using 5 % formic acid and eluted with two times 100 µl 80 % and 100 % acetonitrile. Subsequently, peptides were lyophilized and reconstituted in 50 µl 5% formic acid.

N-TAILS

Experiments were performed in triplicates. 1 mg of ethanol-precipitated reduced, alkylated and dimethylated *S. cerevisiae* protein lysate was dissolved in 100 µl of 6 M GndHCL 100 mM HEPES pH 7.5. After diluting the GndHCL to 0.6 M, proteins were digested using 20 µg of trypsin overnight at 37 °C. After digestion, peptides were purified using the SepPak (Waters, UK) column system (50 mg of C_{18} material) and redissolved in 2 M GndHCL 100 mM HEPES pH 7.5. 50 mg of high molecular weight

dendritic hyperbranched polyglycerol-aldehyde (Flintbox, University of British Columbia, Canada) were added and the coupling reaction started by the addition of 30 mM cyanoborohydride. The samples were incubated overnight at 37 °C. The reaction subsequently quenched using 100 mM ammonium bicarbonate and incubation for 1 h. Samples were applied to Amicon Ultra Centrifugal Filters (Merck Millipore, USA) with a 10 kDa molecular weight cut-off and N-terminal peptides were obtained by centrifugation at 21,000 g for 20 min. Filters were washed once using 100 mM ammonium bicarbonate. The filtrates were combined, acidified to pH 2 with 5 % formic acid and peptides were purified using the SepPak (Waters, UK) column system (50 mg of C_{18} material). After lyophilization, peptides were reconstituted in 30 µl 5 % formic acid.

C-TAILS

Experiments were performed in triplicates. 1 mg of ethanol-precipitated, reduced, alkylated and dimethylated *S. cerevisiae* protein lysate was dissolved in 200 µl of 6 M 500 mM MES pH 5 containing 1 M ethanolamine. 10 mM N-Hydroxysuccinimide (NHS) and 20 mM 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimid (EDC) were added and the reaction incubated at 25 °C for 1 h. The EDC addition was repeated twice during an incubation period of 20 hours. Proteins were subsequently precipitated using ethanol precipitation as described before and the proteins re-solubilized in 100 µl of 6 M GndHCL in 20 mM HEPES pH 7.5. After dilution of the GndHCl to 0.6 M, proteins were digested with 20 µg of trypsin overnight at 37 °C. Free N-termini formed during digestion were dimethylated by adding 20 mM formaldehyde / 20 mM cyanoborohydride and incubation overnight at 25 °C. Peptides were subsequently purified using the SepPak (Waters, UK) column system (50 mg of C_{18} material), lyophilized and dissolved in 400 mM MES pH 5, 2 M GndHCL, 10% acetonitrile and 1 mM poly-allylamine (~ 56 kDa molecular weight). Coupling was initiated by the addition of 50 mM EDC. The EDC addition was repeated twice during an incubation period of 20 h at 25 °C. Samples were applied to Amicon Ultra Centrifugal Filters (Merck Millipore, USA) with a 10 kDa molecular weight cut-off and C-

terminal peptides were obtained by centrifugation at 21,000 g for 40 min at 37 °C. Filters were washed once using water. The filtrates were combined, acidified to pH 2 with 5 % formic acid and peptides were purified using the SepPak (Waters, UK) column system (50 mg of C_{18} material). After lyophilization, peptides were reconstituted in 30 µl 5 % formic acid.