Supplementary Material

to

Chemoselective Fractionation for Shotgun Proteomics: Global Assessment of Cysteinyl Peptide Capture*

De Lin, Jing Li, Robbert J.C. Slebos and Daniel C. Liebler

Synthesis of IBB

Figures S1-S10.

SYNTHESIS OF IBB

^{1}H and ^{13}C NMR spectra were acquired with a Bruker DPX-300 or AV-400 instrument (Bruker Instruments, Billerica, MA). All reactions were carried out under argon. Anhydrous THF, DMF and Et_3N were either commercial available or dried under in house solvent purification system. Purification by column chromatography was carried out on silica gel and TLC plates were visualized by UV and stained with phosphomolybdic acid. Compounds **2, 3** and **4** were prepared according to reported procedures. {Rock, 1996 #34}

3-(tert-butyldimethylsilyloxy)benzaldehyde (2). To a solution of 3-hydroxylbenzaldehyde (6.1 g, 50 mmol) in THF (150 mL) was added tert-butyldimethylsilyl chloride (TBDMSCl, 9.0 g, 60 mmol). The solution was cooled to 0 ^{o}C and triethylamine (8.3 mL, 60 mmol) was added dropwise. A precipitate formed immediately. After stirring overnight, the reaction mixture was filtered and the THF removed under reduced pressure. The crude product was purified by silica gel chromatography (eluent hexane-EtOAc 9:1), affording **2** as a slightly yellow oil (10.1 g, 86%). ^{1}H-NMR (300 MHz, CD_3Cl) δ 0.24 (s, 6H), 1.02(s, 9H), 7.10-7.16 (m, 1H), 7.28-7.53 (m, 3H), 9.98 (s, 1H); ^{13}C NMR (75 MHz, CD_3Cl) δ -4.50, 18.14, 25.55, 119.81, 123.51, 126.50, 130.02, 137.86, 156.35, 192.08.

1-Hydroxy-1-[3-(tert-butyldimethylsilyloxy)phenyl](2-phenyl-1,3-dithian-2-yl)ethane (3). To a solution of 2-phenyl-1,3-dithiane (7.84g, 40 mmol) in THF (100 mL) at 0 ^{o}C under an argon atmosphere was added 16 mL of n-butyllithium (2.5 M in hexane, 40 mmol) dropwise. After 30 min, **2** (9.44g, 40 mmol) was added. The solution was stirred at 0 ^{o}C for 1 h, and then quenched by 6N HCl (10 mL). After evaporating the organic solvent, the residue was dissolved

in 200 mL CH_2Cl_2 and washed with brine (2 x 100 mL). The organic phase was dried with $NaSO_4$ and evaporated under reduced pressure. The resulting oil was crystallized from ethanol/water, affording **3** as white powder (15.7g, 91%). ^{1}H-NMR (300 MHz, CD_3Cl) δ 0.13 (s, 1H), 0.96 (s, 1H), 1.88-2.00 (2H), 2.61-2.81 (4H), 2.92 (d, 1H, *J* = 3.8 Hz), 4.96 (d, 1H, *J* = 3.6 Hz), 6.39-6.48 (2H), 6.66-6.73 (1H), 6.98 (t, 1H, *J* = 7.8 Hz), 7.23-7.37 (3H), 7.68-7.77 (2H); ^{13}C NMR (75 MHz, CD_3Cl) δ -4.42, 18.07, 24.73, 25.63, 26.96, 27.26, 66.43, 80.87, 119.71, 119.93, 121.17, 127.49, 127.81, 128.12, 130.46, 137.37, 138.72, 154.54.

Methyl 2-[3-(hydroxy(2-phenyl-1,3-dithian-2-yl)methyl)phenoxy]acetate (4). To a solution of **3** (10.8g, 25 mmol) and methyl bromoacetate (6.1g, 40 mmol) in dry THF (100 mL) under an argon atmosphere was added 27 mL TBAF (1 M in THF, 17 mmol) dropwise. The solution was stirred overnight, and then the organic solvent was evaporated. The residue was dissolved in 200 mL EtOAc and washed with brine (2 x 100 mL). The organic phase was dried with $NaSO_4$ and evaporated under reduced pressure. The resulting oil was crystallized from EtOAc/Hexane, affording **4** as white powder (9.1g, 93%). ^{1}H-NMR (300 MHz, CD_3Cl) δ 1.88-1.98 (2H), 2.60-2.82 (4H), 3.00 (d, 1H, *J* = 3.4 Hz), 3.80 (s, 3H), 4.39 (s, 2H), 4.98 (d, 1H, *J* = 3.2 Hz), 6.33 (s, 1H), 6.58 (d, 1H, *J* = 7.6 Hz), 6.82 (dd, 1H, *J* = 8.2 and 2.6 Hz), 7.07 (t, 1H, *J* = 7.8 Hz), 7.26-7.37 (3H), 7.66-7.74 (2H); ^{13}C NMR (75 MHz, CD_3Cl) δ 24.66, 26.91, 27.21, 52.10, 65.12, 80.55, 113.63, 115.25, 121.75, 127.47, 127.93, 128.06, 130.46, 137.36, 138.99, 156.54, 169.22.

N-[2-(2-(2-aminoethoxy)ethoxy)ethyl]-2-[3-(hydroxy(2-phenyl-1,3-dithian-2-yl)methyl)phenoxy]acetamide (5). To a solution of **4** (3.9g, 10 mmol) in 400 mL methanol was added 2,2'-(Ethylenedioxy)bisethylamine (5.9 g, 40 mmol. The solution was stirred overnight,

and then the organic solvent was evaporated. The crude product was purified by silica gel chromatography (eluent CH_2Cl_2-MeOH-NH_4OH 50:50:1), affording 6 as a slightly yellow oil (3.8 g, 75%). ^{1}H-NMR (300 MHz, MeOH-d_4) δ 1.76-1.96 (2H), 2.52-2.82 (6H), 3.47 (t, 2H, *J* = 5.4 Hz), 3.52 (t, 2H, *J* = 5.3 Hz), 3.59 (t, 2H, *J* = 5.3 Hz), 3.62 (s, 4H), 4.23 (s, 2H), 4.98 (s, 1H), 6.37 (s, 1H), 6.58 (d, 1H, *J* = 7.6 Hz), 6.83 (dd, 1H, *J* = 8.2 and 2.6 Hz), 7.04 (t, 1H, *J* = 8.0 Hz), 7.26-7.32 (3H), 7.71-7.77 (2H); ^{13}C NMR (75 MHz, MeOH-d_4) δ 24.7, 26.7, 26.8, 38.5, 40.7, 65.3, 66.7, 69.0, 69.8, 69.9, 72.1, 80.3, 114.0, 114.5, 122.0, 126.9, 127.4, 127.5, 130.9, 137.6, 140.8, 156.4, 169.6.

N-[2-(2-(2-aminoethoxy)ethoxy)ethyl]-2-[3-(hydroxy(2-phenyl-1,3-dithian-2-yl)methyl)phenoxy]acetamide (6). To a solution of biotin (540 mg, 2.2 mmol) in 10 mL DMF was added 1,1'-carbonyldiimidazole (360 mg, 2.2 mmol). After 30 min, the amine **5** (1.01 g, 2 mmol) was added to the reaction mixture and stirred overnight. The solution was concentrated under high vacuum. The crude product was purified by silica gel chromatography (eluent CH_2Cl_2-MeOH 9:1), affording 6 as a slightly yellow oil (820 mg, 56%). ^{1}H-NMR (300 MHz, MeOH-d_4) δ 1.37-1.50 (2H), 1.52-1.80 (4H), 1.82-1.95 (2H), 2.22 (t, 2H, *J* = 7.2 Hz), 2.54-2.75 (4H), 2.71 (d, 1H, *J* = 12.6 Hz), 2.93 (dd, 1H, *J* = 12.6 and 4.9Hz), 3.15-3.25 (m, 1H), 3.47 (t, 2H, *J* = 5.4 Hz), 3.55 (t, 2H, *J* = 5.4 Hz), 3.62 (t, 2H, *J* = 5.4 Hz), 3.62 (s, 4H), 4.24 (s, 2H), 4.30 (dd, 1H, *J* = 7.8 and 4.4 Hz), 4.49 (dd, 1H, *J* = 7.8 and 4.8 Hz), 4.98 (s, 1H), 6.38 (s, 1H), 6.58 (d, 1H, *J* = 7.6 Hz), 6.83 (dd, 1H, *J* = 8.2 and 2.6 Hz), 7.04 (t, 1H, *J* = 8.0 Hz), 7.26-7.34 (3H), 7.71-7.77 (2H); ^{13}C NMR (75 MHz, MeOH-d_4) δ 24.7, 25.4, 26.7, 26.8, 28.1, 28.3, 35.3, 38.5, 38.9, 39.7, 55.6, 60.2, 61.9, 65.3, 66.7, 69.0, 69.2, 69.9, 80.3, 114.0, 114.5, 122.0,126.9, 127.4, 127.5, 130.9, 137.6, 140.7, 156.3, 164.6, 169.6, 174.7.

N-(2-(2-(2-(2-(3-(hydroxy(2-phenyl-1,3-dithian-2-yl)methyl)phenoxy)acetamido)ethoxy)ethoxy)ethyl)-5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide (7). To a solution of **6** (510 mg, 0.7 mmol) in 40 mL 9:1 (v/v) acetonitrile/water was added mercury(I) perchlorate tetrahydrate (790 mg, 2.1 mmol). After stirred for 5 min at room temperature, the solution was quenched by NaS_2O_4 (5%, 10mL), diluted by methanol and filtered, and then washed by NaHCO3 and brine. The crude product was purified by silica gel chromatography (eluent CH_2Cl_2-MeOH 8:1), affording **7** as a white solid (310 mg, 69%). ^{1}H-NMR (300 MHz, MeOH-d_4) δ 1.35-1.49 (2H), 1.52-1.80 (4H), 1.82-1.95 (2H), 2.20 (t, 2H, *J* = 7.2 Hz), 2.70 (d, 1H, *J* = 12.8 Hz), 2.91 (dd, 1H, *J* = 12.8 and 4.9Hz), 3.14-3.23 (m, 1H), 3.46 (t, 2H, *J* = 5.2 Hz), 3.53 (t, 2H, *J* = 5.4 Hz), 3.57 (t, 2H, *J* = 5.4 Hz), 3.58 (s, 4H), 4.28 (dd, 1H, *J* = 7.6 and 4.4 Hz), 4.47 (dd, 1H, *J* = 7.6 and 4.8 Hz), 4.50 (s, 2H), 6.12 (s, 1H), 6.90 (d, 1H, *J* = 8.0 Hz), 7.06-7.13 (m, 2H), 7.28 (t, 1H, *J* = 8.0 Hz), 7.44 (t, 1H, *J* = 8.0 Hz), 7.56 (t, 1H, *J* = 7.2 Hz), 7.99 (d, 1H, , *J* = 6.6 Hz); ^{13}C NMR (75 MHz, MeOH-d_4) δ 25.37, 28.02, 28.29, 29.77, 35.27, 38.44, 38.84, 39.59, 55.53, 60.14, 61.88, 66.81, 68.98, 69.14, 69.81, 75.75, 113.88, 114.20, 120.90, 128.27, 128.62, 129.74, 133.19, 134.47, 140.85, 158.00, 164.63, 169.53, 174.68, 198.00. MS (ESI) calcd for $C_{32}H_{43}N_4O_8S$ (MH+) 643.28, found 643.33.

N-(2-(2-(2-(2-(3-(1-hydroxy-2-oxo-2-phenylethyl)phenoxy)acetamido)ethoxy)ethoxy)-ethyl)-5-(2-oxohexahydro-1H-thieno[3,4-d]imidazol-4-yl)pentanamide (8). To a solution of **7** (320 mg, 0.5 mmol) and iodoactic acid (280 mg, 1.5 mmol) in 20 mL CH_2Cl_2 was added DCC (270 mg, 1.3 mmol) and DMAP (6 mg, 0.05 mmol). After 2.5 h, the resulting suspension was filtered, and the filtrate was diluted with CH_2Cl_2. The organic phase was washed with $NaHCO_3$ and brine, dried with $NaSO_4$. The crude product was purified by silica gel chromatography (eluent CH_2Cl_2-MeOH 9:1), affording **8** as a white solid (330 mg, 81%). ^{1}H-NMR (300 MHz,

$CDCl_3$) δ 1.32-1.47 (2H), 1.55-1.78 (4H), 2.20 (t, 2H, *J* = 7.2 Hz), 2.70 (d, 1H, *J* = 12.7 Hz), 2.86 (dd, 1H, *J* = 12.8 and 4.8 Hz), 3.04-3.15 (m, 1H), 3.37-3.62 (12H), 3.85 (s, 2H), 4.26 (dd, 1H, *J* = 7.1 and 4.8 Hz), 4.41-4.49 (m, 1H), 4.48 (s, 2H), 5.79 (s, 1H), 6.66 (s, 1H), 6.73 (t, 1H, *J* = 5.3 Hz), 6.84 (s, 1H), 6.91 (dd, 1H, *J* = 8.1 and 2.3 Hz), 7.07 (s, 1H), 7.09-7.18 (m, 2H), 7.32 (t, 1H, *J* = 7.8 Hz), 7.42 (t, 1H, *J* = 7.8 Hz), 7.55 (t, 1H, *J* = 7.0 Hz), 7.93 (d, 1H, , *J* = 8.0 Hz); ^{13}C NMR (75 MHz, MeOH-d_4) δ 25.55, 28.02, 28.20, 35.92, 38.76, 39.04, 40.46, 55.61, 60.12, 61.70, 67.60, 69.55, 69.86, 69.98, 70.05, 78.21, 115.21, 115.34, 122.22, 128.78, 130.54, 133.84, 134.15, 134.65, 157.62, 164.07, 167.93, 168.16, 173.35, 192.84. MS (ESI) calcd for $C_{34}H_{43}IN_4NaO_9S$ (M + Na) 833.17, found 833.17

Figure S1. MS/MS spectrum of model peptide (**I**), Ac-AVAGCAGAR (Ac-TpepC).

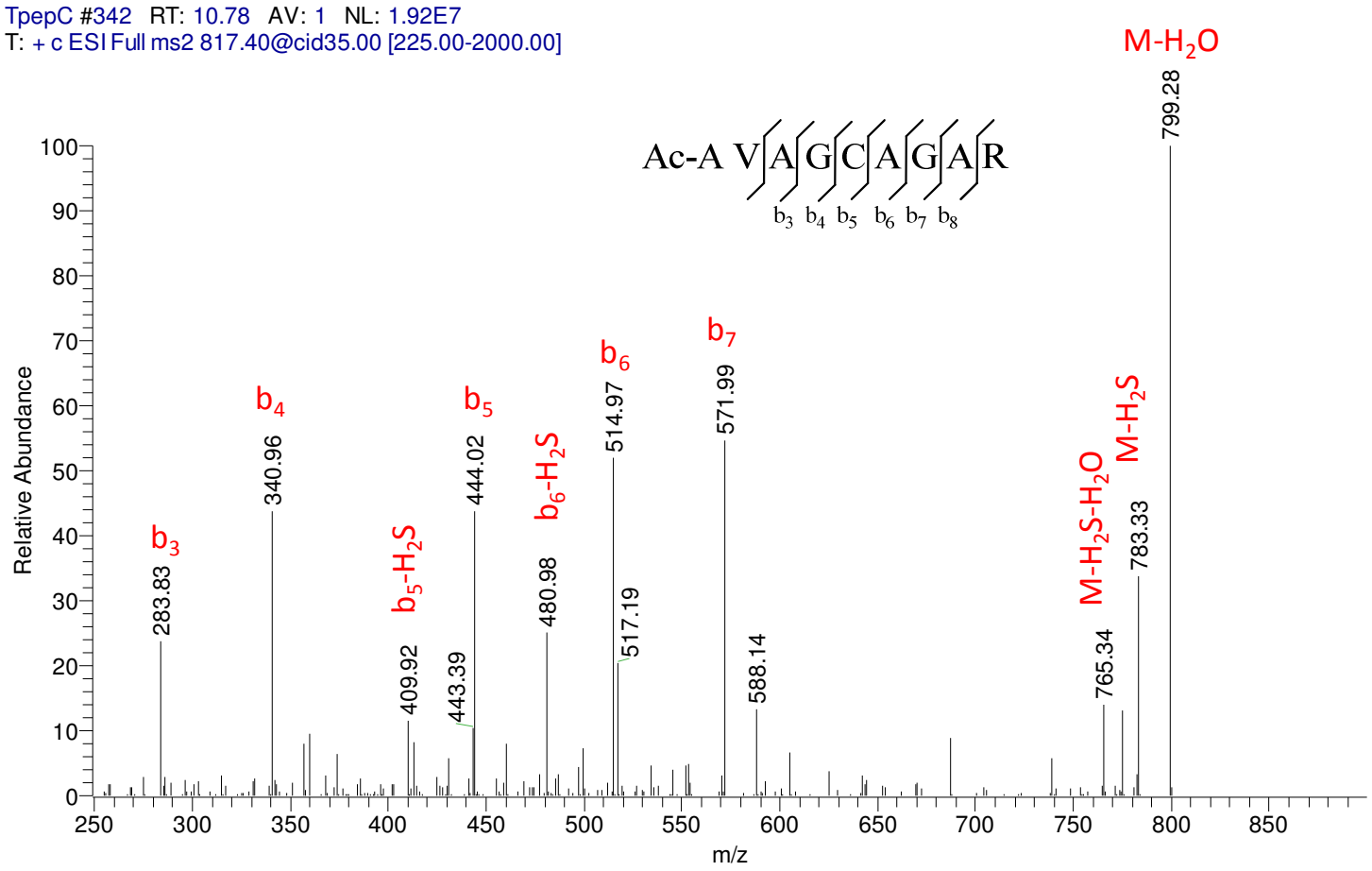

Figure S2. MS/MS spectrum of model peptide IBB conjugate (**II**, Ac-TpepC-IBB).

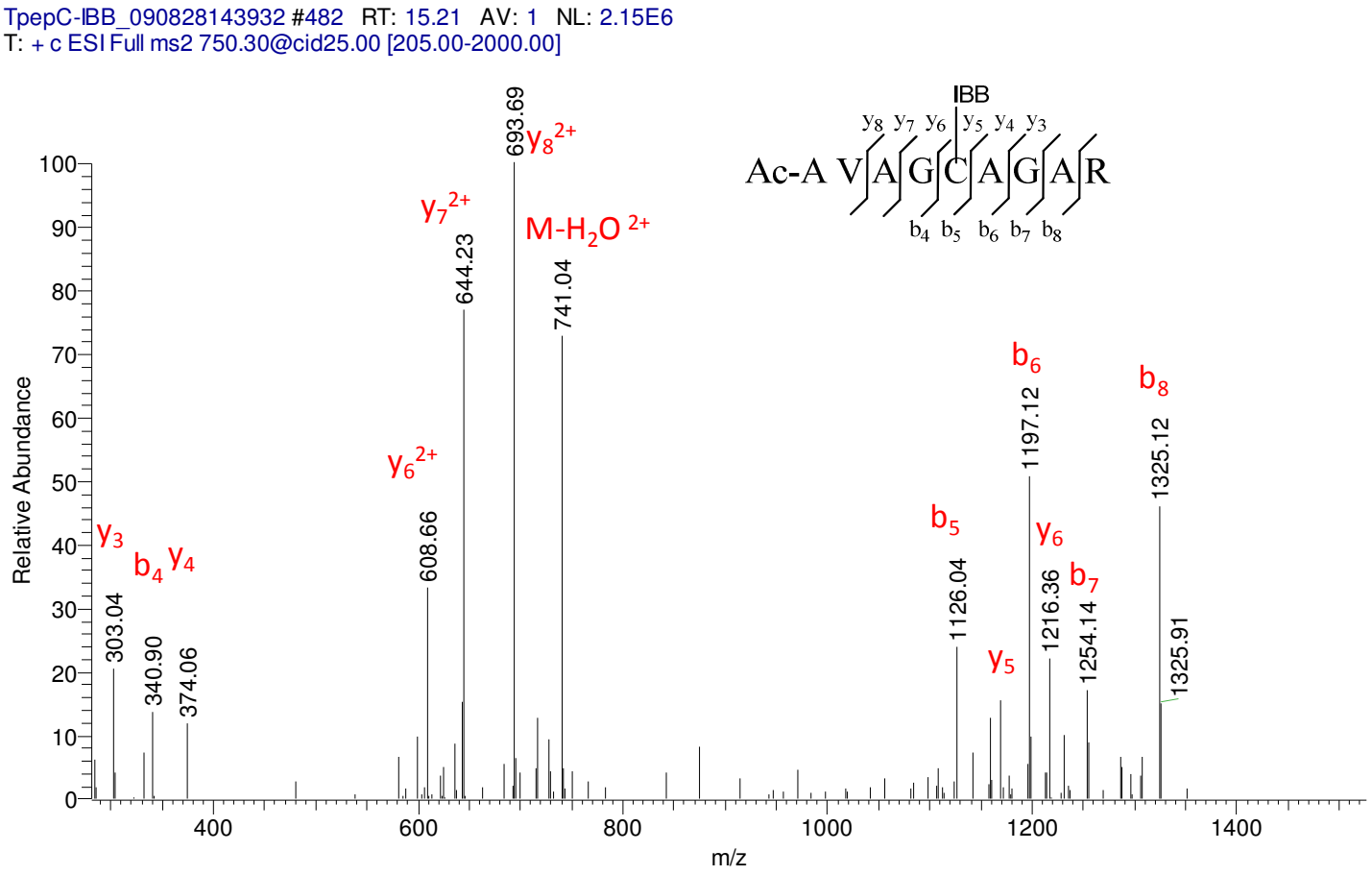

Figure S3. MS/MS spectrum of model peptide IAA conjugate (**V**, Ac-TpepC-IAA).

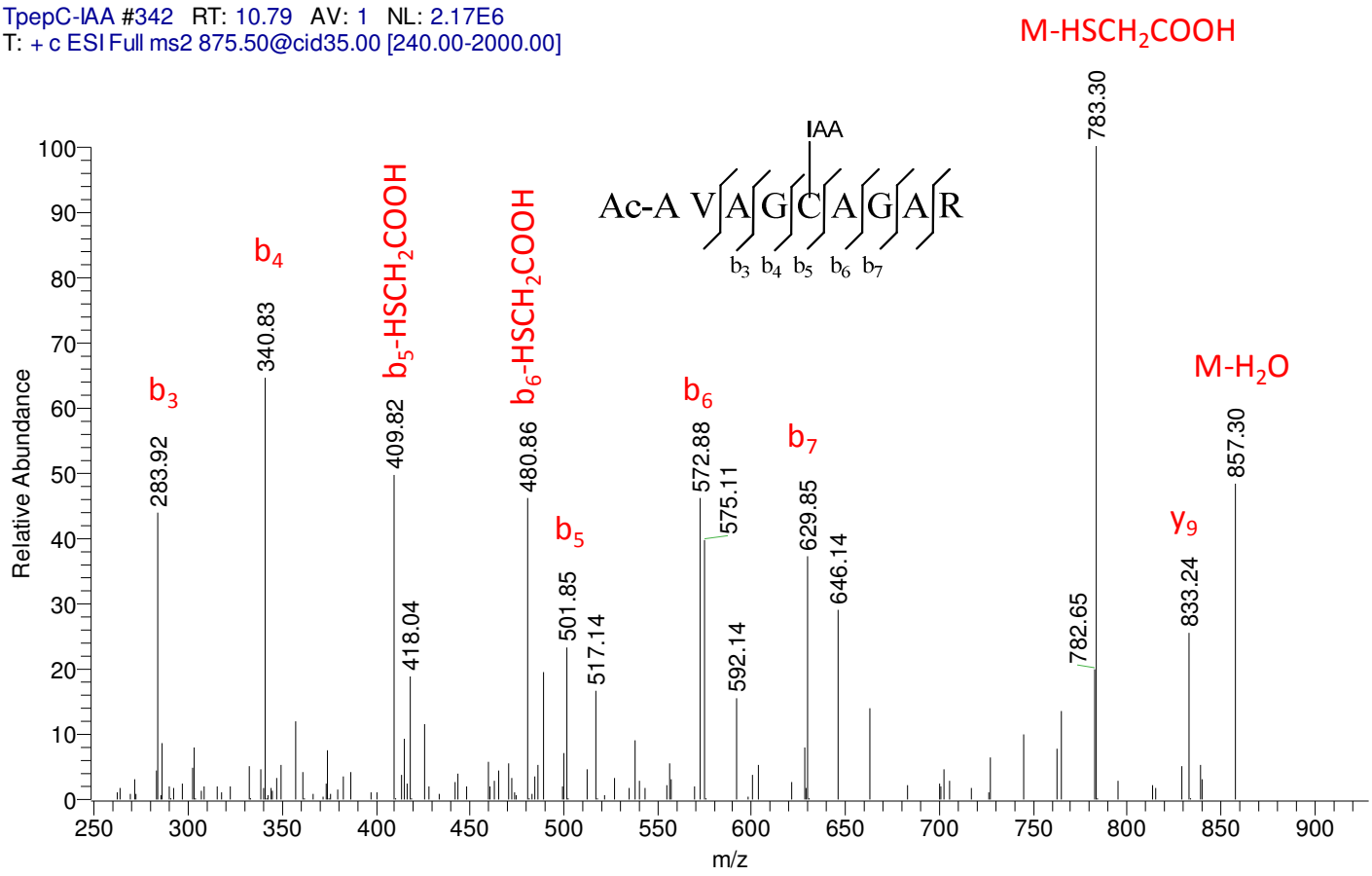

Figure S4. Distribution of theoretical amino acid containing peptides level from *in silico* tryptic digestion of the IPI Human database version 3.37 (■) and the *Saccharomyces* Genome database (●).

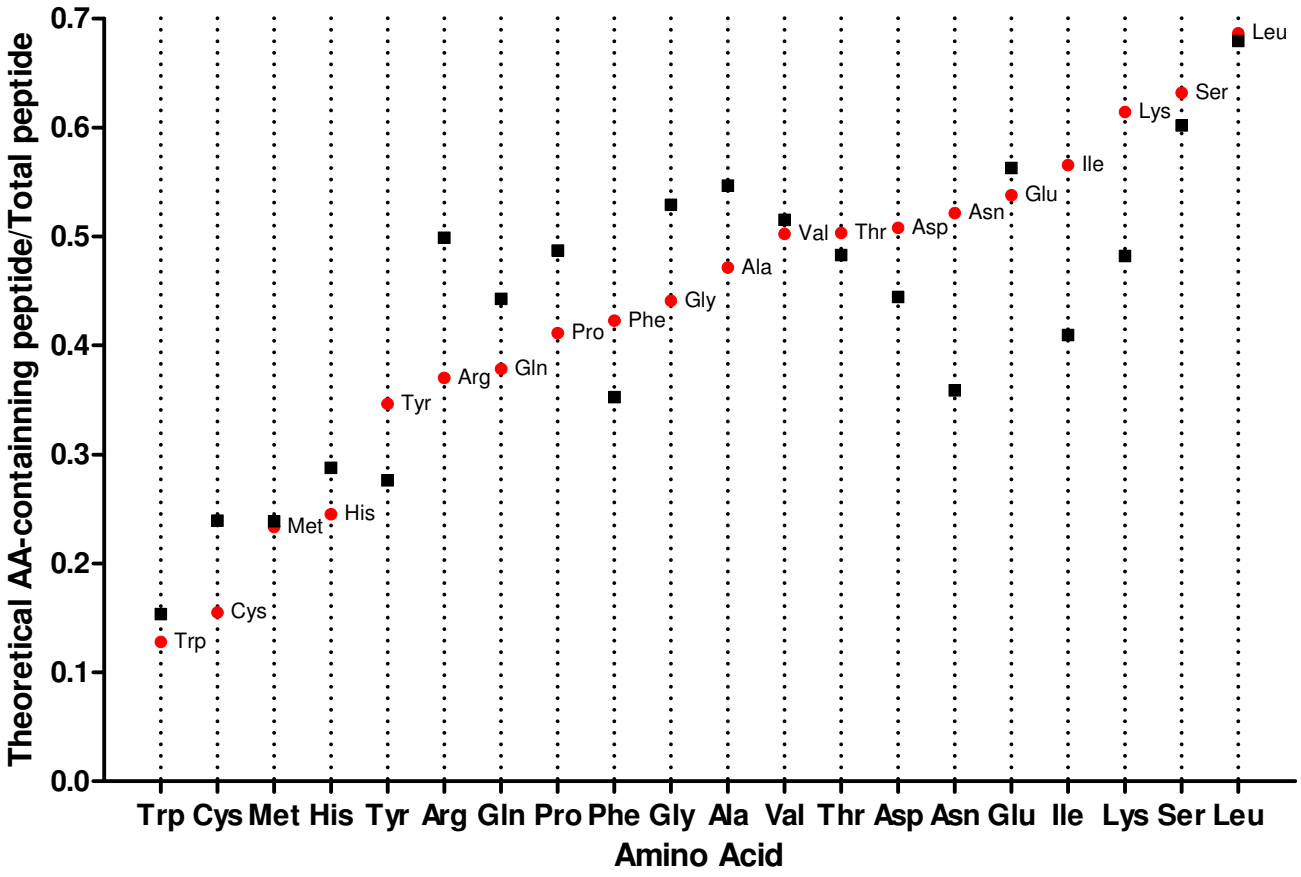

Figure S5. Relationship between IBB enrichment effect and Met-peptide abundance in yeast (A) and RKO cell (B) proteins. the protein index (x-axis) was sorted by increasing detected Met-peptide abundance. The listed r values are Spearman correlation coefficients for proteins with at least one predicted or detected Met-peptide.

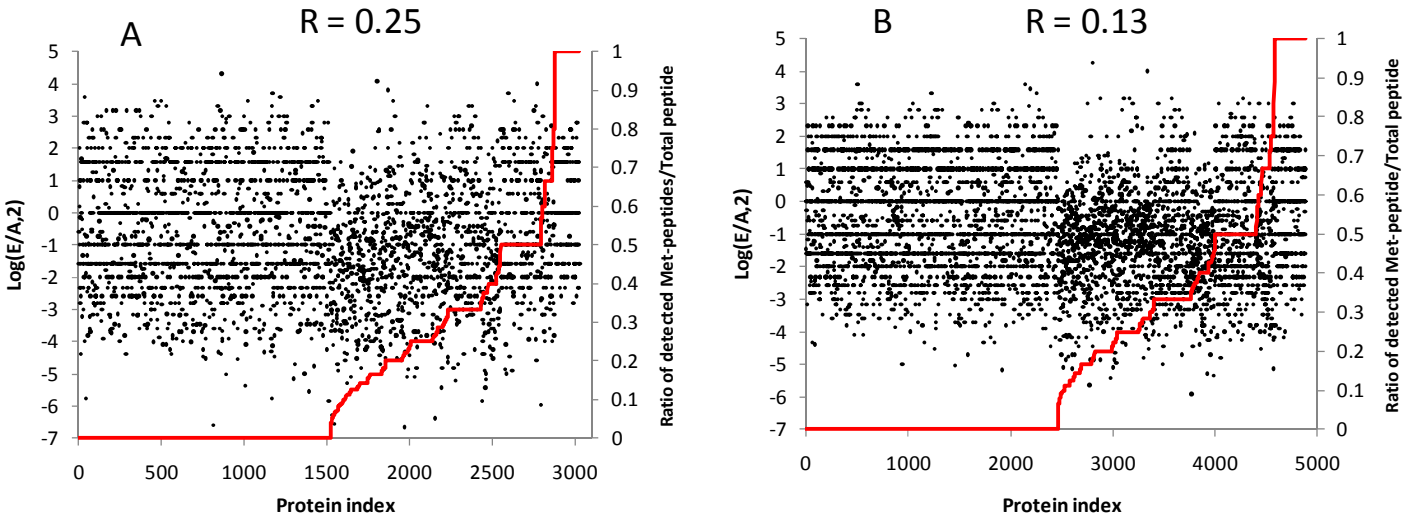

Figure S6. Correlation between protein abundance (cell copy number) and spectral counts in yeast global (G) fraction. Protein protein expression data are from (Ghaemmaghami et al. (2003) *Nature* 425: 737-741). Spearman r = 0.55, p (two-tailed) < 0.0001.

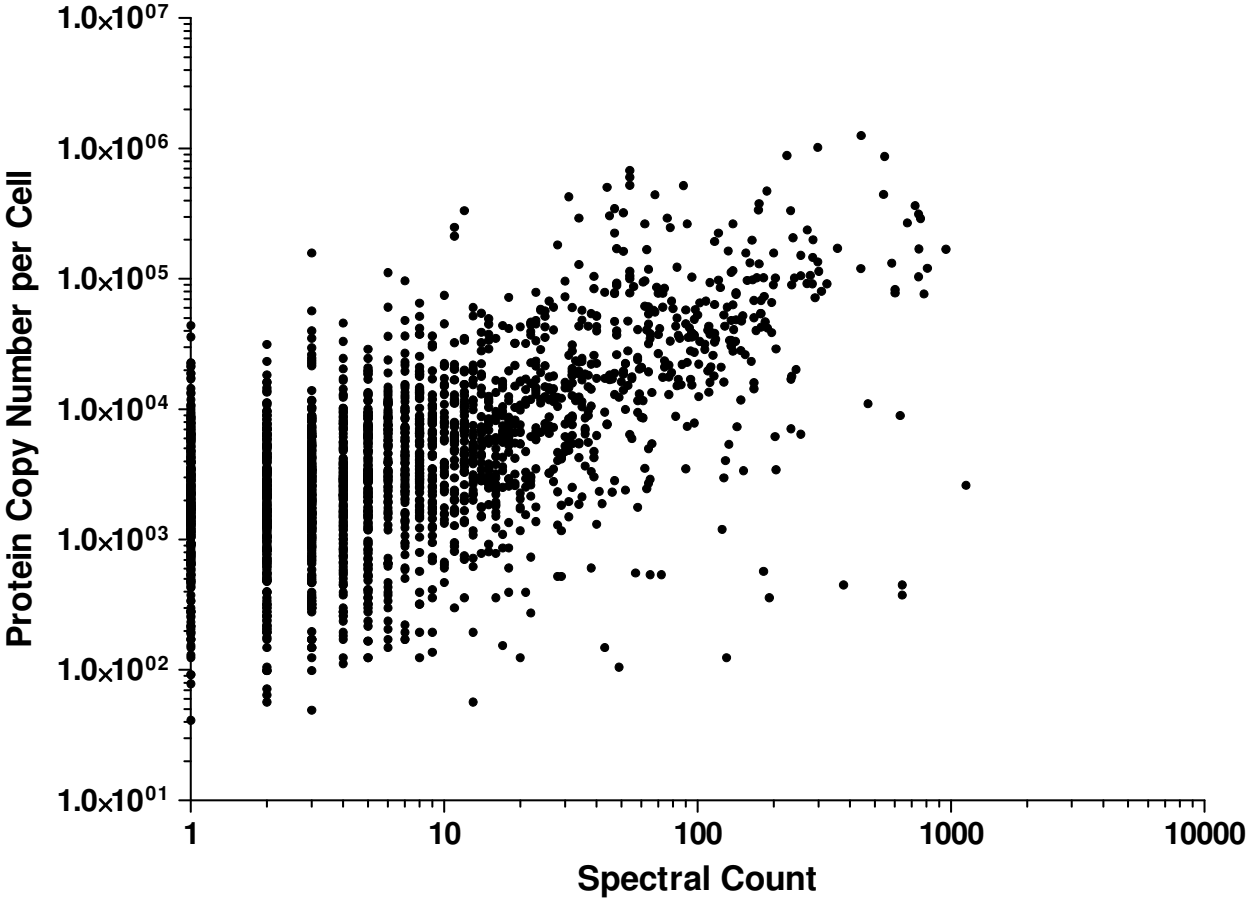

Figure S7. Correlation between theoretical Cys-peptide fraction and protein abundance (cell copy number) and in yeast global (G) fraction. Protein protein expression data are from (Ghaemmaghami et al. (2003) *Nature* 425: 737-741). Spearman r = -0.097.

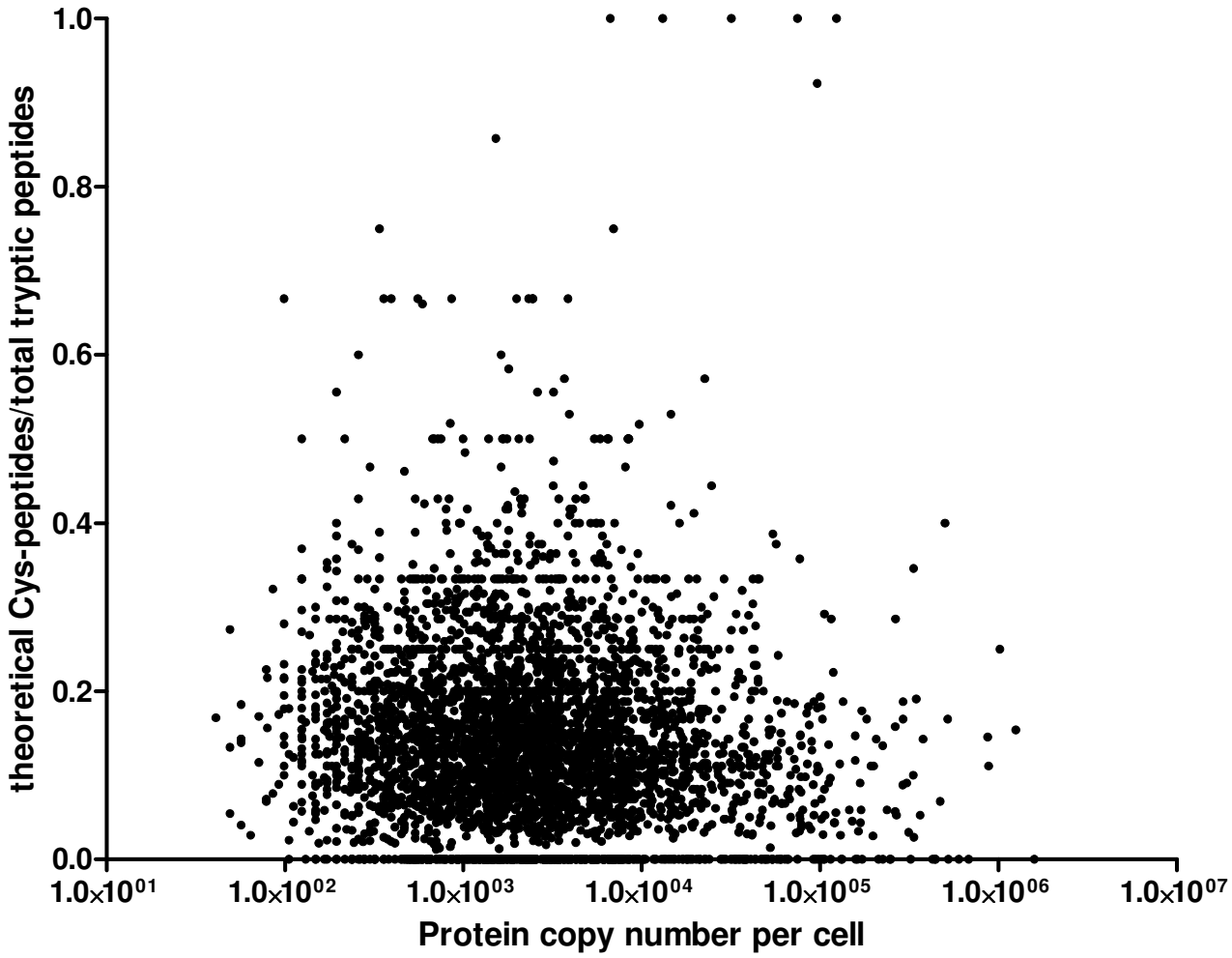

Figure S8. Comparison of distributions across cellular components for identified RKO cell proteins from global (G) and elute (E) fractions. Assignment to GO classifications was done with the WebGestalt gene analysis tool (Zhang et al. (2005) *Nucleic Acids Res* **33**, W741-748). A total of 4887 indentified proteins were annotated to 4668 unique genes, among which 4173 genes were identified in global sample and 3407 genes were identified in elute sample.

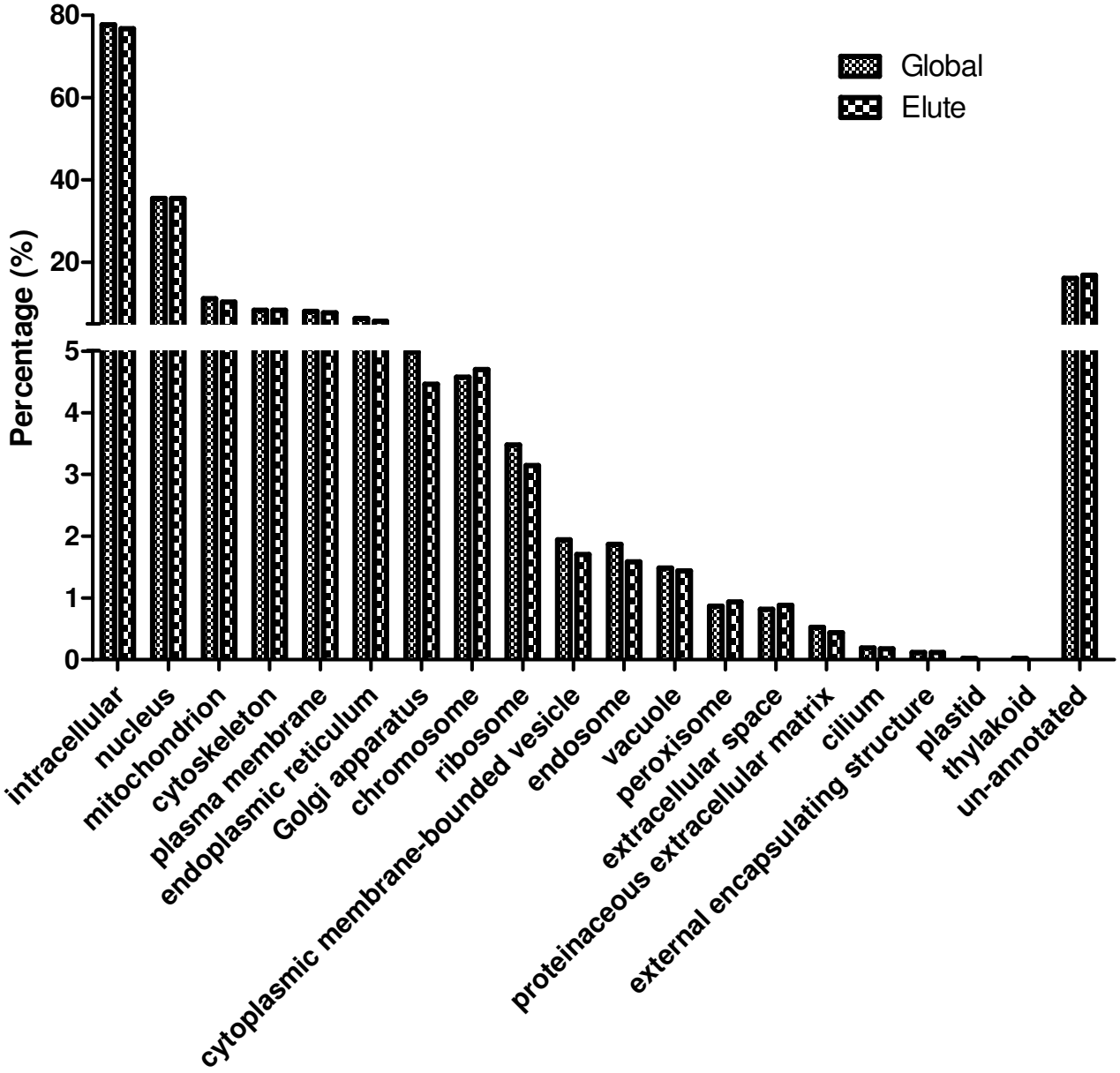

Figure S9. Comparison of distributions across biological processes for identified RKO cell proteins from global (G) and elute (E) fractions. Assignment to GO classifications was done with the WebGestalt gene analysis tool (Zhang et al. (2005) *Nucleic Acids Res* **33**, W741-748).

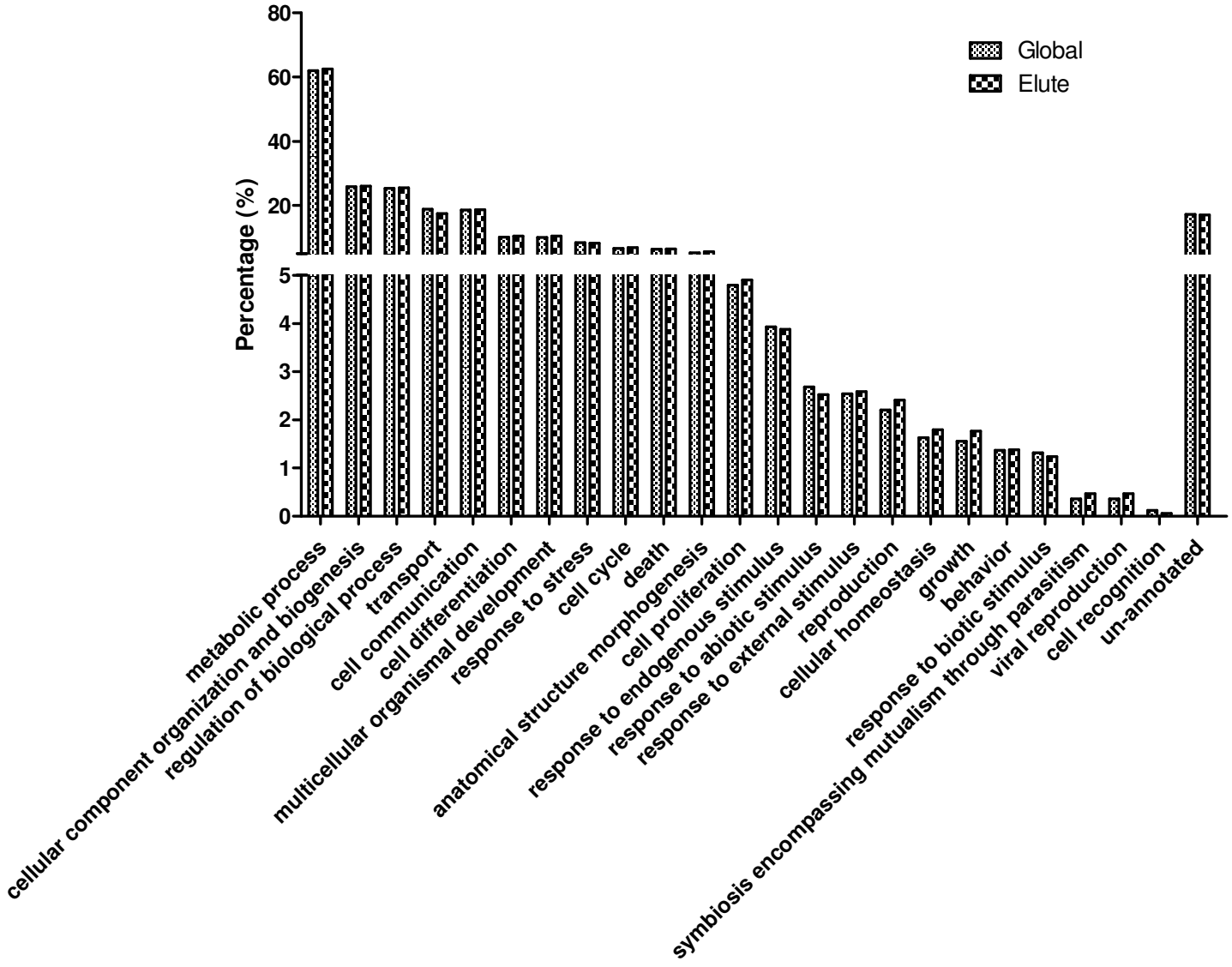

Figure S10. Comparison of distributions across molecular functions for identified RKO cell proteins from global (G) and elute (E) fractions. Assignment to GO classifications was done with the WebGestalt gene analysis tool (Zhang et al. (2005) *Nucleic Acids Res* **33**, W741-748).