SUPPORTING INFORMATION for

Chemo-Selection Strategy for Limited Proteolysis Experiments on the Proteomic Scale

Renze Ma,[1] He Meng,[1] Nancy Wiebelhaus,[1] and Michael C. Fitzgerald[1,*]

[1]Department of Chemistry, Duke University, Durham, North Carolina 27708

*Corresponding Author (email: michael.c.fitzgerald@duke.edu)

CONTENTS:

The **SUPPORTING INFORMATION** includes **Supplemental Text**, **Figures S-1** and **S-2**, as well as **Table S-1**, which is provided in this document, and **Tables S-2** - **S-7**, which are provided as excel spreadsheets. The **Supplemental Text** includes detailed information about the yeast culture, details on how the chemo-selection efficiency of the STEPP protocol was evaluated, more detailed information on how protein-ligand dissociation constants (i.e., K_d values) were determined from the STEPP-PP data. **Figure S-1** shows the global distribution of semi-tryptic peptide fold changes of the STEPP-PP experiments. **Figure S-2** includes a comparison of MALDI mass spectra acquired on a tryptic digest of cytochrome c before and after the chemo-selection step in the STEPP protocol. **Table S-1** summarizes the peptide sequences assigned to the ion signals observed in **Figure S-2**. **Tables S-2** summarizes the assayed semi-tryptic peptides, fold changes at each denaturant concentration, and the hit selection results from the STEPP-PP CsA binding experiments. **Table S-3** summarizes the assayed semi-tryptic peptides, fold changes at each denaturant concentration, and the hit selection results from the STEPP-PP geldanamycin binding experiments. **Table S-4** summarizes proteins identified in the STEPP-LiP MCF-7 versus MCF-10A cell line comparison with single digestion. **Table S-**

5 summarizes semi-tryptic peptides identified in the STEPP-LiP MCF-7 versus MCF-10A cell line comparison with double digestion. **Table S-6** summarizes the proteins identified in the assayed semi-tryptic peptides, normalized fold changes, p-values, and the hit selection results from the STEPP-LiP experiment on the MCF-7 and MCF-10A cell lines. **Table S-7** summarizes the overlapping assayed proteins and protein hits in the STEPP-LiP experiments described here and the normal LiP experiments described in reference 6.

SUPPLIMENTARY TEXT

Yeast Cell Culture Protocol

The yeast GAL1 overexpression strain YDR155C from Y258 host library was purchased from Open Biosystems. A yeast colony was incubated in 50 mL YEPD medium at 30 °C overnight to reach an OD600 of ~1.6. A 20 mL portion of the overnight culture was inoculated with the YEPD medium (1 L) to give an OD600 of ~0.3. The inoculated medium was incubated at 30 °C until the OD600 of the solution was between 1.2-2.0. Yeast pellets were generated by centrifuging 250 mL fractions of the final YEPD medium. The pellets were stored at -20 °C.

Evaluation of Chemo-Selection Efficiency in STEPP Protocol

The efficiency of the chemo-selection strategy employed in the STEPP protocol developed here was evaluated using a mixture of tryptic peptides generated from cytochrome c. The mixture contained an equimolar mixture of tryptic peptides with free amines (non-labeled tryptic peptides) and tryptic peptides that had been reacted with methacrylic acid NHS ester (i.e., labeled tryptic peptides). The methacrylic acid NHS ester was a TMT reagent surrogate that labeled and blocked all the free amines.

The cytochrome c tryptic peptide mixture was prepared by digesting 50 µg of cytochrome c with 1 µg trypsin in 100 µL TEAB buffer (0.5 M, pH 8.5) overnight. Half of the digest was reacted with 0.22 mg of methacrylic acid NHS ester dissolved in 20 µL ACN for 1.5 hours at room temperature. This amount of protein and methacrylic acid NHS ester was the same as the amount of total protein and TMT reagent (respectively)

that was used in the STEPP protocol. The labeling reaction was quenched with addition of 4 µL 5% v/v hydroxylamine, the same conditions used to quench the TMT labelling reaction in the STEPP protocol. Equimolar portions of the labeled and non-labeled mixtures of cytochrome c tryptic peptide mixtures were combined and desalted using a C18 column according to manufacturer's protocol. The desalted sample was subjected to the same chemo-selection reaction conditions used in the STEPP protocol described in the Experimental Section. Briefly, the desalted sample was re-dissolved in 100µL of 0.1M TEAB solution (pH 8.5). NHS-activated agarose resin and 50µL of 0.5M NaCl were added to the sample solution. The NHS-activated agarose resin to total peptide ratio was approximately 150:1 (w:w). The reactions were allowed to proceed for 1.5 hours at room temperature before acidified by 2% TFA solution and desalted using C18 columns.

MALDI spectra of the samples before and after the chemo-selection were acquired on an autoflex speed LRF MALDI-TOF mass spectrometer (Bruker Daltonics) and compared. Subjecting the peptide mixture to the STEPP protocol should remove all the non-labeled tryptic peptides with free amines. Indeed, a MALDI analysis of the tryptic peptide mixture before and after the STEPP protocol showed ion signals from the tryptic peptides with free amines completely disappeared, while the ion signals of labeled tryptic peptides all remained **(Figure S-2 and Table S-1)**.

K_d Value Determination

The median fold change of the four semi-tryptic peptides from HSP82 binding-domain in the geldanamycin binding STEPP-PP experiment were fit to **equation S-1** using a Mathematica 11.0 script:

$$log_2\ fold\ change = log_2\ \frac{\frac{1}{1+Exp\left(-\frac{\Delta G_f + m\times[Urea]}{RT}\right)} + A}{\frac{1}{1+Exp\left(-\frac{\Delta G_f + \Delta\Delta G_f + m\times[Urea]}{RT}\right)} + A} \quad \text{(S-1)}$$

In **equation S-1**, the $Fold\ change$ is the (-)/(+) semi-tryptic peptide TMT reporter ion intensity ratio at each denaturant concentration, ΔG_f is the folding free energy, $\Delta\Delta G_f$ is the observed ligand-induced protein folding free energy change (i.e., the binding free energy), m is $\frac{d\Delta G_f}{d[Urea]}$, R is the ideal gas constant; T is temperature in Kelvin, *A* is a customized constant representing the experimental noise (see below for explanation and derivation of **equation S-1**). Ultimately, the $\Delta\Delta G_f$ value obtained from fitting the $\log_2$(fold-change) versus [Urea] data to **equation S-1** was used in **equation S-2** to calculate the dissociation constant, K_d as described by Schellman (*Biopolymers* **1975**, *14* (5), 999-1018):

$$K_d = \frac{[L]}{e^{-\Delta\Delta G_f/RT} - 1} \quad \text{(S-2)}$$

In equation SI-2, [L] is the free ligand concentration, $\Delta\Delta G_f$ is as described above, R is the ideal gas constant, and T is the temperature in K.

Derivation of Equation S-1

As described by Myers and co-workers (*Protein Sci* **1995,** *4* (10), 2138-48) the folding free energy of a protein, $\Delta G_{f,(-)}$, in chemical denaturant $[Urea]$ can be expressed as

$$\Delta G_{f,(-)} = \Delta G_f + m \times [Urea] = -RTlnK_{f,(-)} \quad \text{(S-3)}$$

, where $\Delta G_{f,(-)}$ is the folding free energy of the protein in the absence of ligand, and $K_{f,(-)}$ is the equilibrium constant between the folded and unfolded state of the protein in the

absence of ligand, which is defined as

$$K_{f,(-)} = \frac{[Folded]_{(-)}}{[Unfolded]_{(-)}} \text{ (S-4)}$$

where $[Folded]$ is the concentration of the folded protein, and $[Unfolded]$ is the concentration of the unfolded protein.

Equations S-3 and **S-4** can be combined we get

$$\frac{[Folded]_{(-)}}{[Unfolded]_{(-)}} = Exp\left(-\frac{\Delta G_f + m \times [Urea]}{RT}\right) \qquad \text{(S-5)}$$

The total concentration of the protein can be expressed as, $[Protein] = [Folded]_{(-)} + [Unfolded]_{(-)}$. Therefore, **equation S-5** can be expressed as,

$$\frac{[Unfolded]_{(-)}}{[Protein]} = \frac{1}{1+Exp\left(-\frac{\Delta G_f + m \times [Urea]}{RT}\right)} \text{ (S-6)}$$

Similarly, the folding free energy of a protein in the presence of ligand, $\Delta G_{f,(+)}$, in chemical denaturant can be expressed as,

$$\Delta G_{f,(+)} = \Delta G_f + \Delta\Delta G_f + m \times [Urea] = -RTlnK_{f,(+)} \text{ (S-7)}$$

Similarly, we can have

$$\frac{[Unfolded]_{(+)}}{[Protein]} = \frac{1}{1+Exp\left(-\frac{\Delta G_f + \Delta\Delta G_f + m \times [Urea]}{RT}\right)} \qquad \text{(S-8)}$$

Dividing **equation S-5** by **S-8** and taking the logarithm of both sides yields,

$$log_2 \frac{[Unfolded]_{(-)}}{[Unfolded]_{(+)}} = log_2 \frac{\frac{1}{1+Exp\left(-\frac{\Delta G_f + m \times [Urea]}{RT}\right)}}{\frac{1}{1+Exp\left(-\frac{\Delta G_f + \Delta\Delta G_f + m \times [Urea]}{RT}\right)}} \text{ (S-9)}$$

As the all unfolded protein are digested by thermolysin and result in semi-tryptic peptides, the ratio $\frac{[Unfolded]_{(-)}}{[Unfolded]_{(+)}}$ is the same as the (-)/(+) semi-tryptic peptide TMT reporter ion intensity ratios. As the STEPP-PP is a TMT-based experiment, there is a constant noise

in target peptide quantification from co-isolation and co-fragmentation of other interfering TMT-labeled peptides. To account for this noise in TMT quantification, we add the constant, *A,* to both numerator and denominator. This constant term is set to be the median isolation interference of the data to be fitted (e.g. in fitting of the HSP82 hit peptides in geldanamycin-binding experiment, this constant *A* is set to be the median isolation interference of the four HSP82 hit peptides, which is 0.17).

By substituting $fold\ change$ for $\frac{[Unfolded]_{(-)}}{[Unfolded]_{(+)}}$ and adding the constant *A*, we get **equation S-1**:

$$log_2\ (fold\ change) = log_2\ \frac{\frac{1}{1+Exp\left(-\frac{\Delta G_f + m\times[Urea]}{RT}\right)} + A}{\frac{1}{1+Exp\left(-\frac{\Delta G_f + \Delta\Delta G_f + m\times[Urea]}{RT}\right)} + A}$$

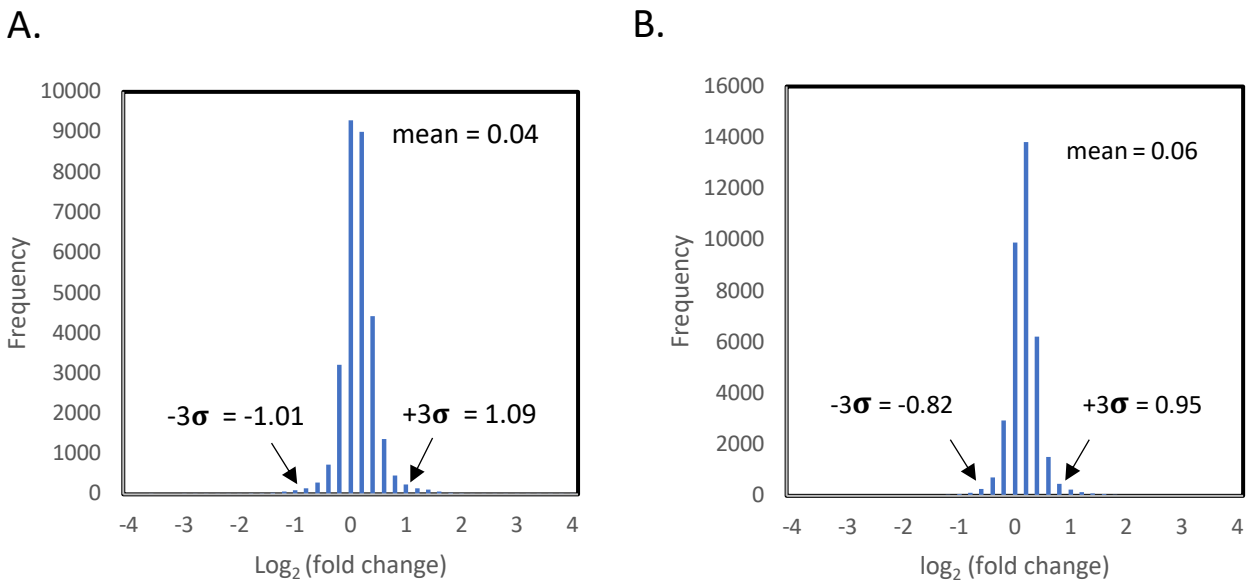

Figure S-1. Distribution of semi-tryptic peptide fold-change values in STEPP-PP CsA binding (**A**) and geldanamycin binding experiment (**B**).

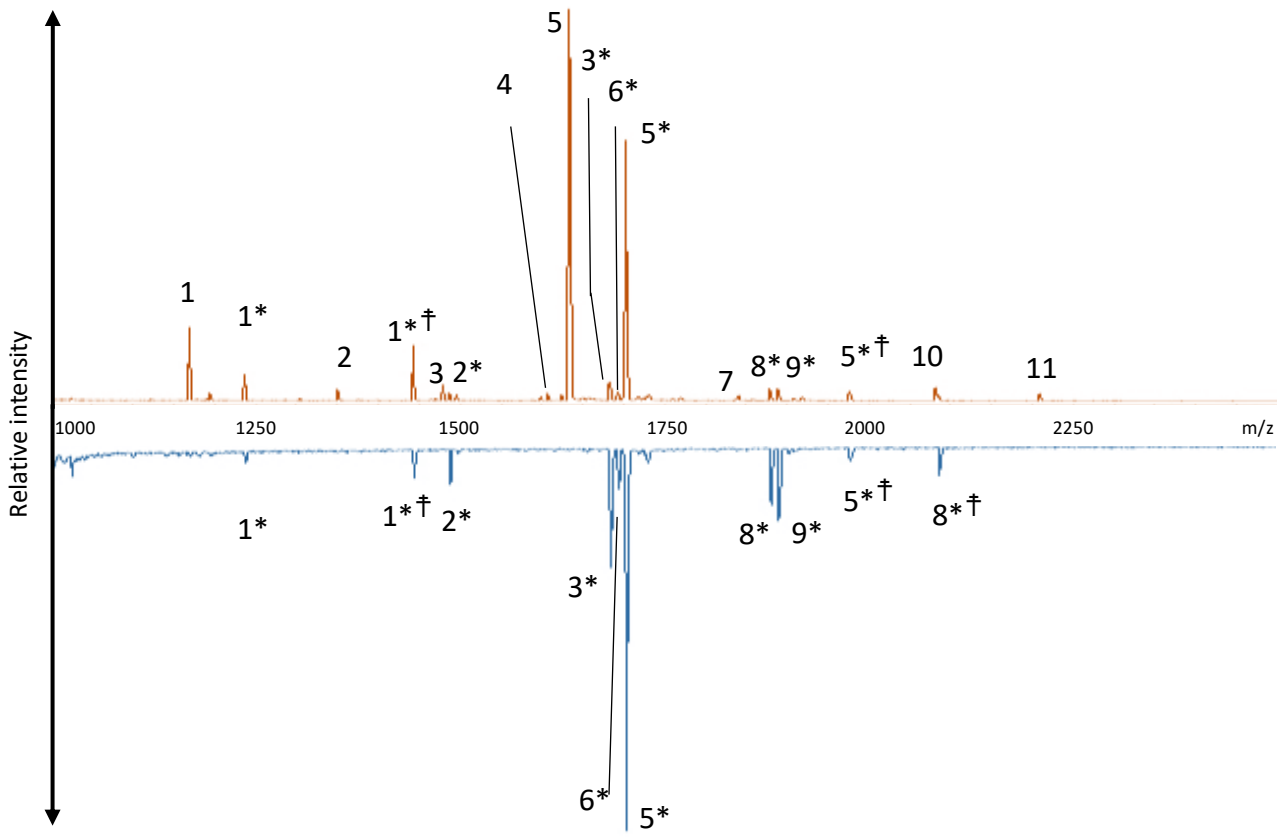

Figure S-2. MALDI spectra of the labeled and non-labeled cytochrome c tryptic digest mixture before (red plot on the top) and after (blue plot on the bottom) the chemo-selection procedures (see **Supplemental Text**). Ion signals of cytochrome c tryptic peptides are assigned with a number according to increasing m/z value. The signals of their corresponding labeled peptides are those with an asterisk (*). The † sign denotes the ion signal is from a peptide-matrix adduct (for signal m/z values and amino acid assignments, see **Table S-1**). After the chemo-selection procedure (blue plot on the bottom), all the non-labeled tryptic peptide signals disappeared, while all the labeled signals (i.e., the signal indicated with a "*") remained.

Table S-1. List of all detected cytochrome c tryptic peptide signals and their amino acid sequence assignment in **Figure S-2**.

Signal number	m/z measured	m/z expected	Amino acid sequence assignment [a,b]
1	1168.6	1168.6	TGPNLHGLFGR
1*	1236.6	1236.7	(MA)TGPNLHGLFGR
2	1350.6	1350.7	TEREDLIAYLK
1*†	1442.6	1442.7	(MA)TGPNLHGLFGR + sinapic acid - H_2O
3	1478.7	1478.8	KTEREDLIAYLK and/or TEREDLIAYLKK
2*	1487.0	1486.9	(MA)TEREDLIAYLK(MA)
4	1598.7	1598.8	KTGQAPGFTYTDANK
5	1633.5	1633.6	CAQCHTVEK | | Heme
3*	1682.8	1683.0	(MA)K(MA)TEREDLIAYLK(MA) and/or (MA)TEREDLIAYLK(MA)K(MA)
6*	1694.2	1694.2	(MA)MIFAGIK(MA)K(MA)K(MA)TER
5*	1701.6	1701.7	(MA)CAQCHTVEK and/or CAQCHTVEK(MA) [c] | | | | Heme Heme
7	1840.7	1840.9	KTGQAPGFTYTDANKNK
8*	1879.9	1880.2	(MA)GGK(MA)HK(MA)TGPNLHGLFGR
9*	1889.2	1889.1	(MA)K(MA)TEREDLIAYLK(MA) + sinapic acid - H_2O and/or (MA)TEREDLIAYLK(MA)K(MA) + sinapic acid - H_2O
5*†	1975.8	1975.8	(MA)CAQCHTVEK + sinapic acid - H_2O and/or CAQCHTVEK(MA) + sinapic acid - H_2O [c] | | | | Heme Heme
10	2080.9	2081.0	GITWKEETLMEYLENPK and/or YIPGTKMIFAGIKKKTER
8*†	2085.3	2085.2	(MA)GGK(MA)HK(MA)TGPNLHGLFGR + sinapic acid - H_2O
11	2209.0	2209.1	GITWKEETLMEYLENPKK

[a] The "(MA)" in the peptide sequence assignment indicates a methyl acrylic modification on the peptide N-terminus or lysin side chain(s).

[b] The "Heme" is the heme group covalently bonded to the two cysteines in cytochrome.

[c] Although these labeled peptides of signal 5* and 5*† had a non-labeled amine group, this amine group was considered as blocked (non-free), since it did not react with the NHS ester during the NHS ester click reaction. The blockage from the NHS ester reaction might be because of its coordination to the heme group. Because these peptides did not have free amine groups, they were not removed in the chemo-selection step and remained in the sample solution.

Tables S-2. Summary of the assayed semi-tryptic peptides and proteins and the hit selection results from the STEPP-PP CsA binding study.

Table S-3. Summary of the assayed semi-tryptic peptides and proteins and the hit selection results from the STEPP-PP geldanamycin binding study.

Table S-4. Summary of the proteins identified in the STEPP-LiP MCF-7 versus MCF-10A cell line comparison with single digestion.

Table S-5. Summary of the semi-tryptic peptides identified in the STEPP-LiP MCF-7 versus MCF-10A cell line comparison with double digestion.

Table S-6. Summary of the the assayed semi-tryptic peptides and and the hit selection results from the STEPP-LiP experiment on the MCF-7 and MCF-10A cell lines.

Table S-7. Summary of the overlapping assayed proteins and protein hits in the STEPP-LiP experiments described here and the normal LiP experiments described in reference 6.