Supporting Information of

Discovery of Protein-Ligand Interactions by a Combined Energetic-Based and Structural-Bioinformatics-Based Approach
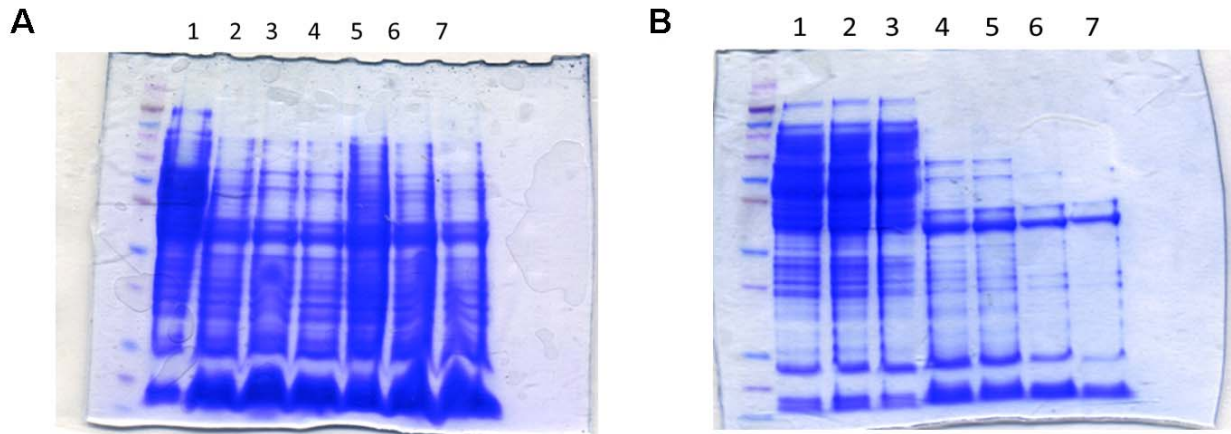
Lingfei Zeng, Woong-Hee Shin, Xiaolei Zhu, Sung Hoon Park, Chiwook Park, W. Andy Tao, and Daisuke Kihara

Contact: watao@purdue.edu, dkihara@purdue.edu

Supplemental Figure S-1. Flowchart of the computational prediction of NAD binding proteins using Patch-Surfer2.0.
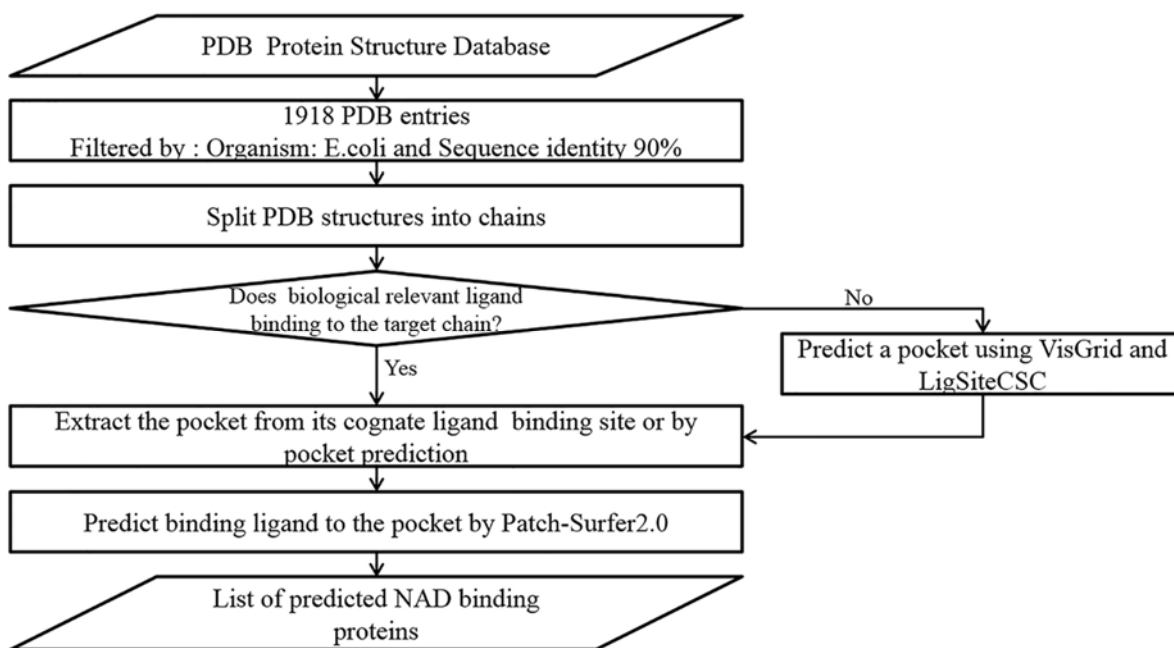
Supplemental Figure S-2. Testing Pulse-proteolysis condition: Thermolysin and urea

concentration.

A) 1. without thermolysin; 2. thermolysin 0.08mg/ml, 1min; 3. thermolysin 0.08mg/ml, 5min; 4.

thermolysin 0.08mg/ml, 10min; 5. thermolysin 0.04mg/ml, 1min; 6. thermolysin 0.04mg/ml,

5min; 7. thermolysin 0.04mg/ml, 10min.

B) 1. without thermolysin, without urea; 2. without thermolysin, 3M urea; 3. without thermolysin,

6M urea; 4. thermolysin 0.2mg/ml, 10min; 3M urea; 5. thermolysin 0.2mg/ml, 10min; 4M urea.

6. thermolysin 0.2mg/ml, 10min; 5M urea; 7. thermolysin 0.2mg/ml, 10min; 6M urea.

Supplemental Figure S-3. Pulse-proteolysis condition testing with different molecular weight cut-off filters.
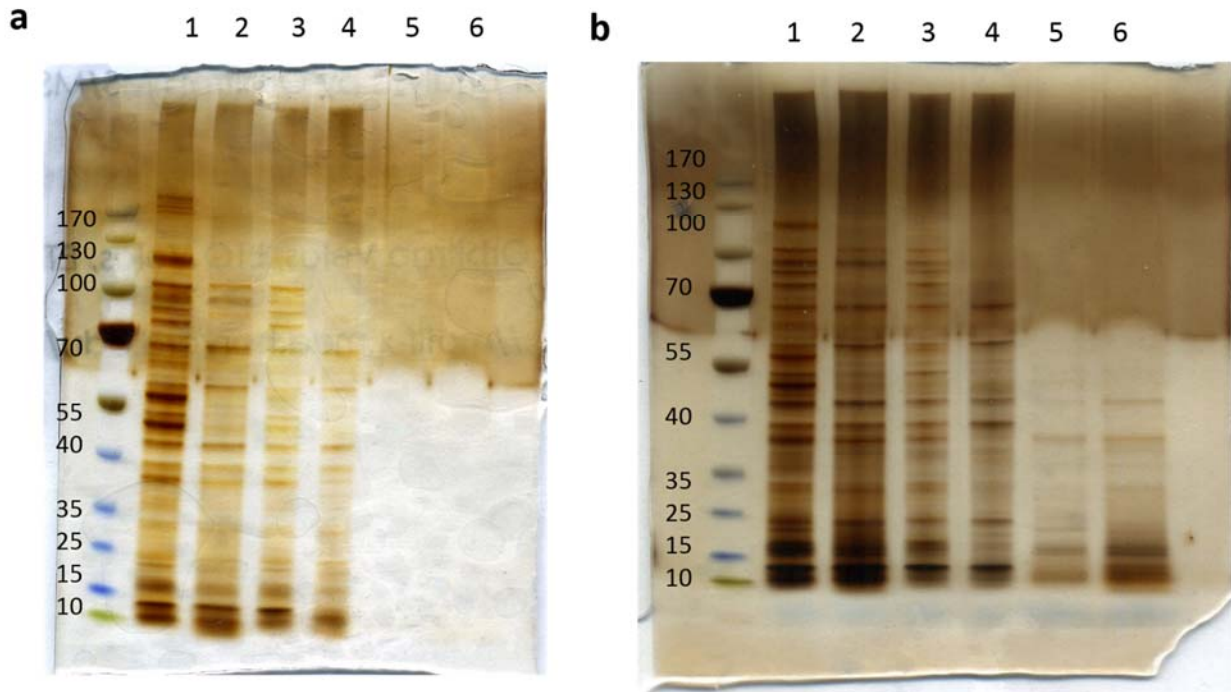
A) 30K filter; **B**) 100K filter. Sample are labeled as the treatment and the portion (top: in the filter; bottom: out of the filter, in the collection tube). Specifically, they are:

1. 50ug protein + 3M urea

2. 50ug protein + 3M urea + thermolysin

3. 50ug protein + 3M urea + filter, top

4. 50ug protein + 3M urea + filter, bottom

5. 50ug protein + 3M urea + thermolysin + filter, top

6. 50ug protein + 3M urea + thermolysin + filter, bottom

Supplemental Figure S-4. Flowchart of pulse-proteolysis-quantitative mass spectrometry using TMT-duplex labeling.
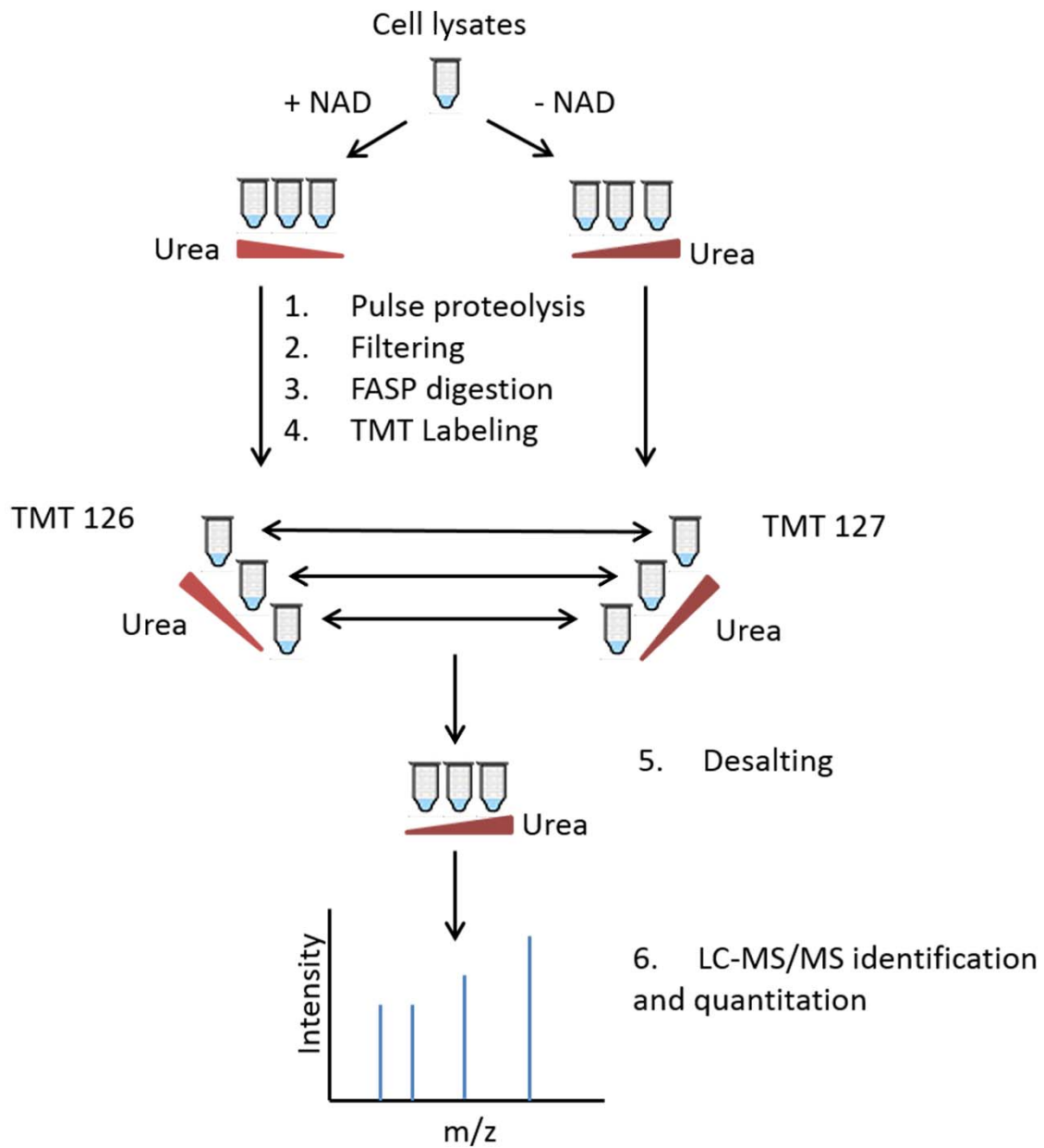
Supplemental Figure S-5. Protein identification and quantification with LC-MS. The number of protein IDs are shown for a range of protein ratios (+NAD/-NAD) in a log scale. Three urea concentrations, 3.5M, 4.0M, and 4.5M were used, for each of which three biological replicates were produced.
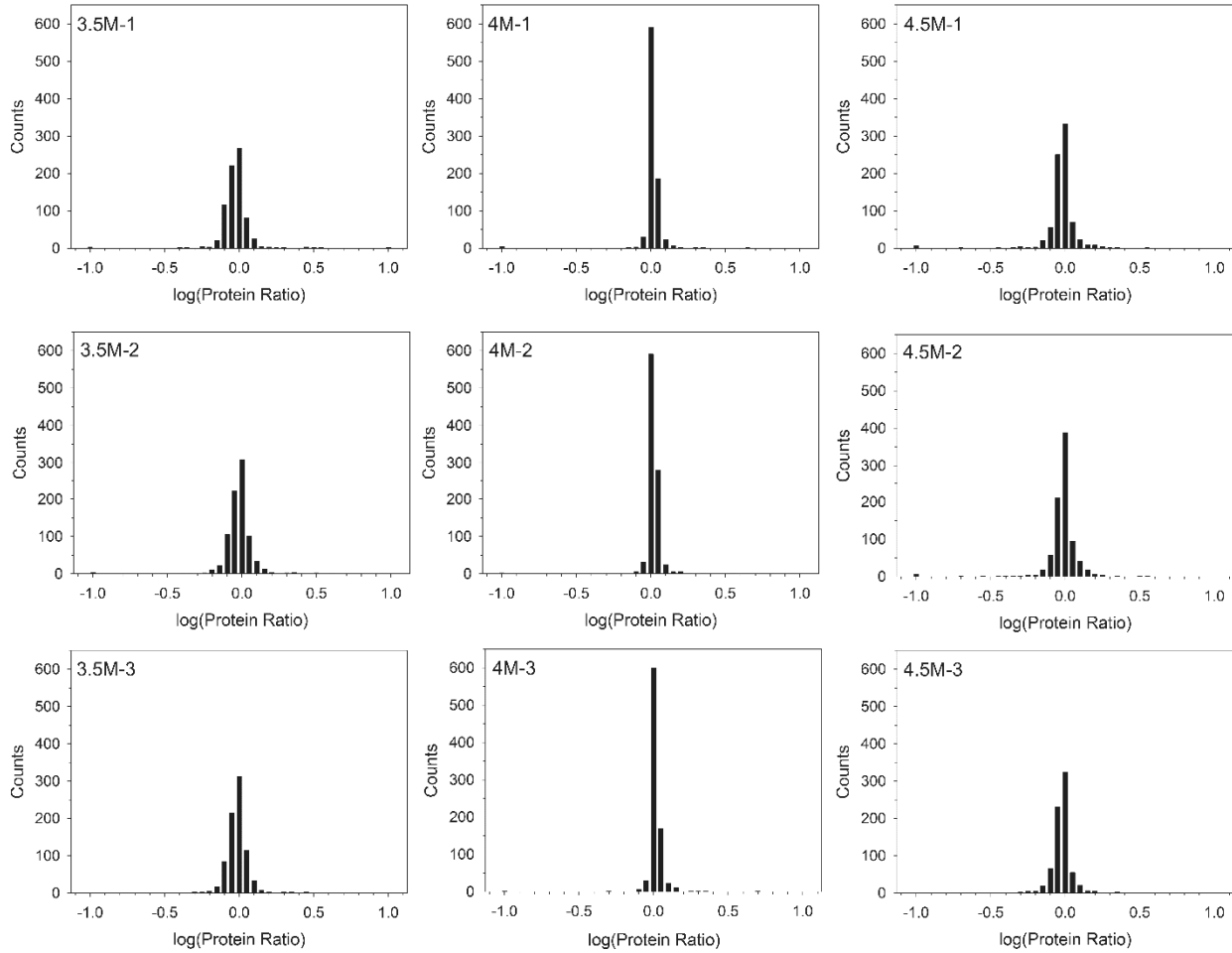
Supplemental Figure S-6. The number of proteins detected in three replicates using the three urea concentrations, 3.5M, 4.0M, and 4.5M.

Figure 2 in the main text shows the number of detected NAD binding proteins and Table 1 summarizes the total number of proteins as well as the detected NAD binding proteins.
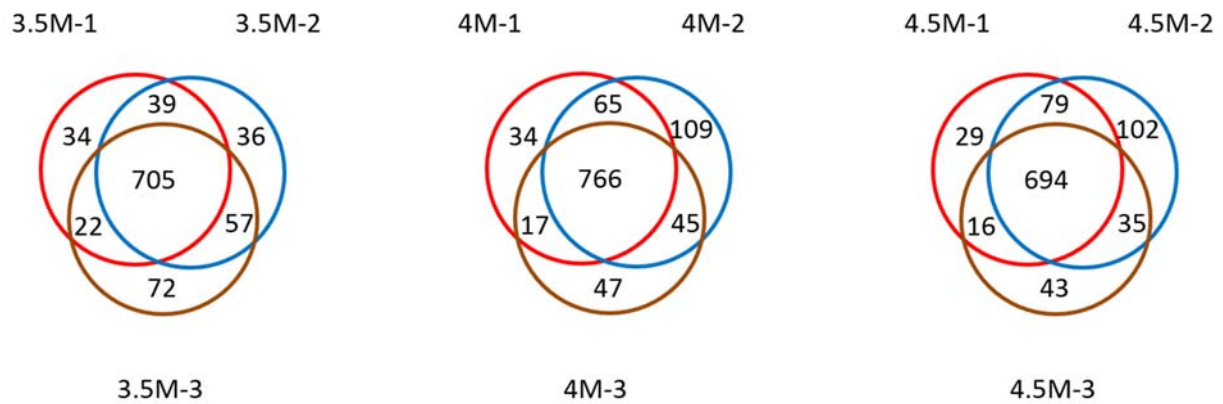
Table S-1. Pairwise sequence identity (%) between predicted NAD binding proteins.

	mtn	deoD	yghU	ppiA	panB	fabD	ushA
galT	16.06	9.42	17.11	18.45	13.82	17.92	16.18
mtn		21.40	11.01	10.67	18.64	20.65	16.58
deoD			16.50	9.59	15.45	21.98	18.04
yghU				22.93	13.04	15.00	19.37
ppiA					12.94	12.24	22.03
panB						13.69	24.39
fabD							21.18

Clustal Omega (Sievers, Higgins, Curr. Protoc. Bioinformatics, 2014) was used to compute the sequence identity.

Supplemental Table S-2. Mass spectrometry identification and quantification data (Excel file)

For sheets 3.5M-1, 3.5M-2, 3.5M-3, 4M-1, 4M-2, 4M-3, 4.5M-1, 4.5M-2 and 4.5M-3, they are data of original MS identification and quantitation. They include protein and peptides (sub rows under each proteins). For the protein rows:

1. Group description is the protein name.
2. Ratio: F1: (126, Sample) / (127, Control) is the protein ratio. In the experiment, we always use TMT[126] for samples with NAD, and TMT[127] for samples without NAD (control).

For sheets 3.5M overlap, 4M overlap and 4.5M overlap, they are overlapped proteins with ratio>1.25 between replicates. Different columns show overlaps between different replicates, and the combination of them are shown in the last column. Combined proteins are those consistently have ratio larger than 1.25 (at least in 2 out of 3 replicates) and considered as "significant proteins".