Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Energetics-Based Discovery of Protein–Ligand Interactions on a Proteomic Scale

Pei-Fen Liu[1,2], Daisuke Kihara[2,3,4,5,6] and Chiwook Park[1,2,6]*

[1]*Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 47907, USA*
[2]*Interdisciplinary Life Science Program, Purdue University, West Lafayette, IN 47907, USA*
[3]*Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA*
[4]*Department of Computational Science, Purdue University, West Lafayette, IN 47907, USA*
[5]*Markey Center for Structural Biology, Purdue University, West Lafayette, IN 47907, USA*
[6]*Bindley Bioscience Center, Purdue University, West Lafayette, IN 47907, USA*

Biochemical functions of proteins in cells frequently involve interactions with various ligands. Proteomic methods for the identification of proteins that interact with specific ligands such as metabolites, signaling molecules, and drugs are valuable in investigating the regulatory mechanisms of cellular metabolism, annotating proteins with unknown functions, and elucidating pharmacological mechanisms. Here we report an energetics-based target identification method in which target proteins in a cell lysate are identified by exploiting the effect of ligand binding on their stabilities. Urea-induced unfolding of proteins in cell lysates is probed by a short pulse of proteolysis, and the effect of a ligand on the amount of folded protein remaining is monitored on a proteomic scale. As proof of principle, we identified proteins that interact with ATP in the *Escherichia coli* proteome. Literature and database mining confirmed that a majority of the identified proteins are indeed ATP-binding proteins. Four identified proteins that were previously not known to interact with ATP were cloned and expressed to validate the result. Except for one protein, the effects of ATP on urea-induced unfolding were confirmed. Analyses of the protein sequences and structure models were also employed to predict potential ATP binding sites in the identified proteins. Our results demonstrate that this energetics-based target identification approach is a facile method to identify proteins that interact with specific ligands on a proteomic scale.

© 2011 Elsevier Ltd. All rights reserved.

Corresponding author. Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, IN 47907, USA. E-mail address: chiwook@purdue.edu.

Introduction

Biochemical functions of proteins commonly involve interactions with small molecules, which act as substrates, signaling molecules, and allosteric regulators. Discovery of novel interactions between proteins and metabolites provides valuable information on the function and regulation of proteins, as well as on the biochemical roles of metabolites.[1] Bioactive small molecules are frequently discovered from phenotype-based assays without knowing their molecular targets, and it is critical to identify their

targets to decipher the mode of action.[2–4] Additionally, the identification of off-targets is essential to understanding the mechanisms of the side effects of drugs.[5] This "target identification problem" or "target deconvolution problem" is actually a significant bottleneck in drug discovery.[6–9]

Protein–ligand interactions have been conventionally investigated with individual proteins by performing hypothesis-driven biochemical and biophysical assays, which are typically labor-intensive and low-throughput. Recent advances in genomics and proteomics, however, prompt the development of high-throughput approaches to discover novel protein–ligand interactions at a systems level. One of the most common methods employed to identify proteins interacting with a ligand is to capture binding proteins by affinity-based separation such as affinity chromatography.[5,10–12] Although well-established and popular, affinity-based separation still has many technical limitations.[6] Small molecules need to be attached to a solid matrix or a detection tag for the separation of the binding proteins. This chemical modification may result in the altered affinity or specificity of the molecule. Also, the requirement for chemical modification limits the application of affinity-based separations only to molecules with reactive functional groups for attachment. Another issue with affinity-based separation is that the control of stringency is not feasible. Proteins bound nonspecifically are removed by extensive washing. However, this process frequently separates proteins by their dissociation kinetics, not by their affinity.[6]

Genetic approaches using mutant libraries or arrays have recently gained popularity as a drug target identification method.[13–15] When a mutant has a distinct drug response from wild type, it is likely that the product of the mutated gene is somehow involved in the action of the drug in the cell. Because of the ease of high-throughput applications, genetic approaches appear to be a promising alternative to biochemical approaches. However, genetic approaches also have shortcomings. Construction of mutant libraries is costly and time-consuming. Mutant libraries are mostly available only for lower organisms, and drug targets in humans must be inferred from experimental results obtained with these model organisms. Genetic approaches are not useful for drugs with subtle pharmacological effects because inhibition of growth is the typical readout in genetic screens. Also, genetic approaches are limited to xenobiotic-drug-like molecules and cannot be used for metabolites. Finally, genes identified from genetic screens do not necessarily interact with the drug directly. There are many possible indirect mechanisms through which mutants have distinct drug responses even when the direct drug targets are still functional.[13]

Due to the pros and cons of current approaches, it is important to develop various tools to identify target proteins for small molecules. Target identification methods based on different principles are complementary to each other, and the combined use of several different approaches may provide a more complete picture of systemwide interactions between proteins and small molecules. Energetics-based target identification is a promising new strategy for this purpose. When a protein forms a complex with a ligand, the bound conformation is stabilized according to the dissociation free energy of the complex at a given ligand concentration. This stabilization of the complex results in changes in the energetic properties of the target protein, such as an increase in thermodynamic stability, a decrease in unfolding rates, and a change in the dynamics of the native proteins. Energetics-based target identification exploits these changes in the energetic properties of the target proteins to identify proteins that interact with test molecules in a mixture of proteins, such as a cell lysate.[16,17]

Here we report an energetics-based target identification method using 'pulse proteolysis.'[18] Pulse proteolysis determines the fraction of folded proteins under a given condition by a brief incubation with a protease. We have shown that this method is a reliable quantitative approach to determining the thermodynamic stability and unfolding kinetics of proteins under various circumstances.[18–21] The use of proteolysis as a structural probe enables us to monitor the urea-induced unfolding of a multitude of proteins in a cell lysate simultaneously without isolating individual proteins for biophysical characterization. By comparing the amounts of protein remaining after pulse proteolysis in the presence and in the absence of a ligand, we can identify targets stabilized by ligand binding from the mixture of proteins.

As proof of principle, we identified ATP-binding proteins in the *Escherichia coli* proteome by combining pulse proteolysis and two-dimensional (2D) gel electrophoresis. The schematic diagram of the experimental procedure is shown in Fig. 1. A cell lysate is incubated with a ligand in urea, while a control sample is incubated under identical conditions without the ligand. After the incubation, pulse proteolysis is performed in an identical manner for both samples. The remaining proteins are analyzed by 2D gel electrophoresis. The comparison of the two 2D gels reveals spots whose intensities are influenced by the presence of the ligand. With this approach, we identified known ATP-binding proteins and proteins whose affinity for ATP has not been reported. We characterized several of the identified proteins and confirmed that the biophysical principle of the methodology is valid. This result demonstrates that our energetics-based target identification by pulse proteolysis is a facile and
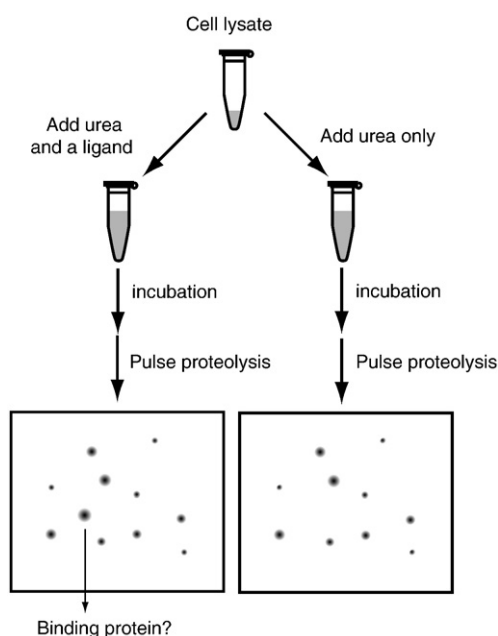
Fig. 1. Energetics-based target identification by pulse proteolysis. A cell lysate is incubated with a ligand and urea. A control reaction is prepared in an identical manner, but without the ligand. After incubation, unfolded proteins are digested by pulse proteolysis, and the remaining proteins are analyzed by 2D gel electrophoresis. Proteins in the spots showing differential intensities on the 2D gels of the two reactions are identified and characterized further.

powerful method for identifying proteins that interact with specific ligands.

Results

Identification of ATP-binding proteins

To identify ATP-binding proteins, we incubated an *E. coli* lysate for 2 h in a buffer containing 3.0 M urea and 1.0 mM adenosine 5′-[γ-thio]triphosphate (ATPγS). We also incubated a control without ATPγS under identical conditions. ATPγS was used, instead of ATP, to minimize the loss of the ligand by enzymatic hydrolysis during the incubation. After the incubation, the reactions were treated with 0.20 mg/mL thermolysin for 1 min (pulse proteolysis), and the resulting proteins were analyzed by 2D gel electrophoresis. Figure 2 shows a representative pair of the resulting 2D gels. Overall, the two gels look quite similar, indicating that ATPγS did not affect most proteins shown on the gels. Still, a careful comparison revealed spots that show different intensities between the two gels. To rule out possible false positives resulting from the variability in sample preparations and 2D electrophoresis, we performed three replicate experiments. For further analysis, we selected only the spots showing obvious and consistent changes in intensities from the three pairs of 2D gels. This screen was performed not to search exhaustively for ATP-binding proteins but to prove the principle of the
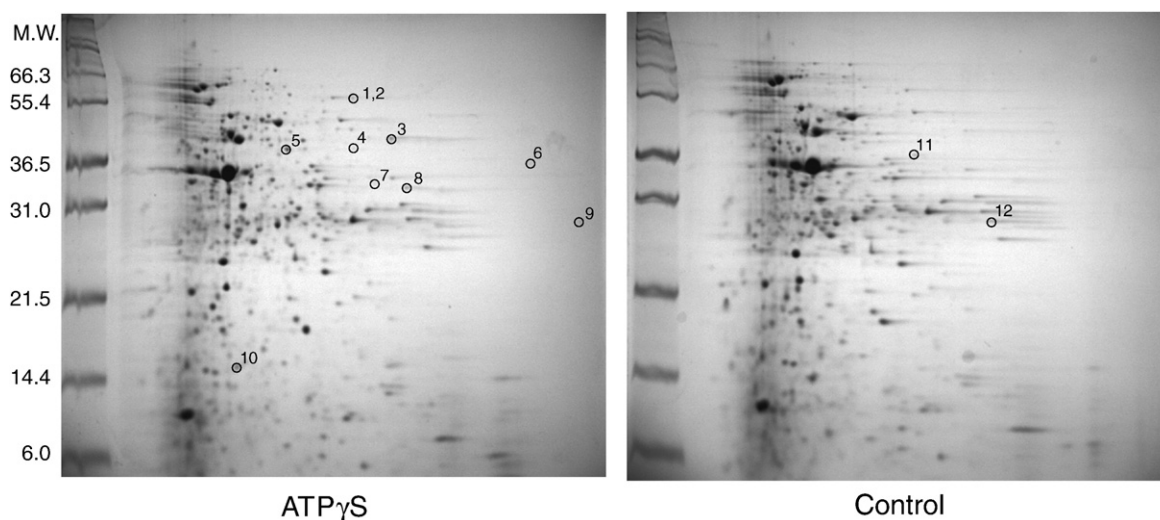

Fig. 2. Two-dimensional gel electrophoresis of an *E. coli* lysate after pulse proteolysis in 3.0 M urea with and without ATPγS. An *E. coli* cell lysate was incubated with 1.0 mM ATPγS in 3.0 M urea for 2 h before pulse proteolysis. A control sample was also prepared under identical conditions, but without 1.0 mM ATPγS. Spots with different intensities between the two gels were identified. Spots selected for in-gel digestion are labeled by numbers. Spots 1 and 2 are distinguishable on the actual gel but look as a single spot on this image. Two spots (spots 11 and 12) only appeared in the control gel. The major spot near 36.5 kDa corresponds to thermolysin.

approach. Through this selection process, we chose 12 spots for protein identification. Ten spots had higher intensities on the gels of ATPγS-treated samples than the control gels without ATPγS (Fig. 2, spots 1–10), and the other two spots had higher intensities on the control gels than the gels of ATPγS-treated samples (Fig. 2, spots 11 and 12). A spot present only in the control gels without ATPγS may occur when conformational changes associated with ATP binding make a protein susceptible to pulse proteolysis or when the spot contains a partially cleaved protein produced in the absence of ATPγS.

The identities of the proteins in the selected spots were determined by using in-gel digestion, followed by matrix-assisted laser desorption/ionization tandem time-of-flight mass spectrometry (Table 1; Table S1). Ten proteins were successfully identified from the 12 spots (Table 1). The comparison of molecular weights estimated from the location of the spots with the known molecular weights of the identified proteins suggested that nine spots contain intact proteins and three spots contain fragments (Table 1). Interestingly, intact glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was found only on the control gels. This observation rules out the possibility that GAPDH is digested partially only in the absence of ATPγS.

Functions of the identified proteins were surveyed by using the information available on EcoCyc,[28] a genomics database of *E. coli* K-12. Seven out of 10 proteins are annotated with a Gene Ontology (GO) term, 'ATP binding' (GO 0005524). Through literature search, we found experimental evidence of ATP binding to six out of these seven proteins (Table 1). The K_d values for five proteins collected from the literature range from 0.025 μM to 600 μM (Table 1), which are all smaller than the concentration of ATPγS used for the screen (1.0 mM). Apparently, yncE is annotated with ATP binding without experimental evidence based on the existence of P-loop, ATP/GTP binding site motif (pattern PS00017 in the PROSITE database[29]). The three identified proteins that are not annotated with ATP binding are dihydrolipoamide dehydrogenase (E3), GAPDH, and a periplasmic-binding protein (mlaC).

Validation of results from the proteomic screen

To validate the proteomic screen results, we chose three proteins with no documented ATP binding ability (GAPDH, E3, and mlaC) and the hypothetical protein annotated as an ATP-binding protein (yncE). The proteins were overexpressed in *E. coli* after cloning, and the effects of ATP on their unfolding in urea were monitored by pulse proteolysis. The cell lysates containing the overexpressed proteins were used directly for validation without purification. It is noteworthy that, even when the identified protein is a true positive, validation with the overexpressed proteins may not exactly reproduce the observation made with the 2D gel electrophoresis. In case of a multimeric protein, the apparent stability would be dependent on protein concentration and can be significantly increased when the protein is overexpressed. Also, if a protein is a component of a protein complex formed with other proteins, overexpression of the single component is not likely to allow the formation of the proper quaternary structure. Therefore, the purpose of the validation with the overexpressed protein is to examine the effect of ATP on the thermodynamics and kinetics of unfolding, not to reproduce the proteomic screen results exactly.

To compare with the proteomic screen, we incubated the proteins in crude cell lysates under identical conditions used for the screen. The cell lysates were incubated in 3 M urea for 2 h at 25 °C with 1.0 mM ATPγS. Controls were also incubated

Table 1. Identified proteins

Spot	Gene name[a]	Description[a]	GO[b]	Experimental evidence[c]	K_d (μM)
			ATP binding		
1	atpA	ATP synthase, F1 complex, α subunit	✔	✔	0.1[22]
2	Lpd	Lipoamide dehydrogenase (E3 monomer)			
3[d]	glnS	Glutaminyl-tRNA synthetase	✔	✔	190[23]
4, 7[d]	pheS	Phenylalanyl-tRNA synthetase, α-chain	✔	✔	600[24]
5	pfkA	6-Phosphofructokinase-1	✔	✔	0.025[25]
6	yncE	Hypothetical protein	✔		
8	sucD	Succinyl-CoA synthetase, α subunit	✔	✔	90[26]
9	mlaC	Periplasmic-binding protein of the phospholipid ABC transporter			
10	groS	GroES	✔	✔[27]	
11, 12[d]	gapA	GAPDH-A			

[a] The gene name and the description for each protein were collected from the EcoCyc (*E. coli* functional genomics) database.
[b] Genes annotated with the GO term 'ATP binding' (GO 0005524) in the EcoCyc database.
[c] Proteins whose ATP binding has been confirmed experimentally.
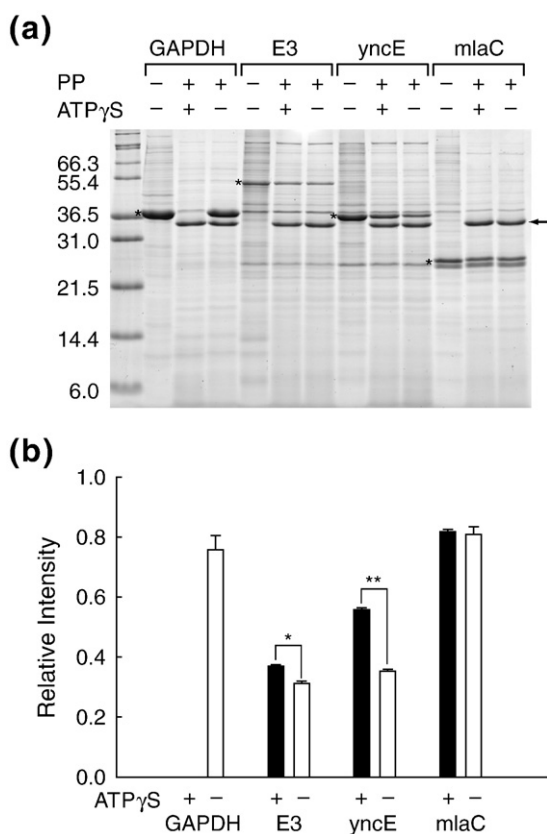[d] The spot was found to contain a fragment of the identified protein.

Fig. 3. Validation of the proteomic screen with recombinant proteins. (a) Pulse proteolysis of recombinant GAPDH, E3, yncE, and yrbC in *E. coli* lysates with and without 1.0 mM ATPγS. PP, pulse proteolysis. The bands corresponding to overexpressed proteins are marked with asterisks. The bands corresponding to thermolysin are marked with an arrow. (b) Relative intensities of the proteins remaining after pulse proteolysis with (black bar) and without (white bar) 1.0 mM ATPγS. The band intensities of overexpressed proteins on the gel shown in (a) were quantified and expressed as ratios to the band intensity of the undigested protein on the same gel. Error bars indicate the standard deviation of triplicate experiments. *$p = 0.0013$; **$p < 0.0001$.

under identical conditions, but without 1.0 mM ATPγS. After pulse proteolysis, the remaining proteins were analyzed by SDS-PAGE (Fig. 3a). For comparison, the amount of protein remaining after pulse proteolysis was normalized as the relative ratio to the amount of the intact proteins without pulse proteolysis (Fig. 3b). The averages of the relative ratios were determined from triplicate experiments. Examination with the overexpressed protein suggests that ATPγS indeed affects the unfolding of GAPDH, E3, and yncE. As observed on the 2D gels, no detectable amount of GAPDH survives pulse proteolysis when ATPγS is present. E3 shows a small but still statistically meaningful

increase in band intensity in the presence of ATPγS ($p = 0.0013$). The observed difference is somehow much less evident than that observed on the 2D gels. The fact that E3 is a component of multicomponent enzyme systems may explain this marginal effect of ATPγS on the overexpressed protein. YncE shows a statistically significant increase in band intensity in the presence of ATPγS ($p < 0.0001$), which confirms the proteomic screen result. However, the test with overexpressed mlaC is not consistent with the proteomic screen result. Whether or not ATPγS is present, mlaC is resistant to pulse proteolysis, and the relative intensities are not statistically different. The discrepancy with the proteomic screen result suggests that mlaC may be a false positive or that the experiment with overexpressed mlaC may not truly mimic the experiment performed with the endogenous mlaC due to the reasons described above.

Glyceraldehyde-3-phosphate dehydrogenase

To elucidate the physical origin of the effect of ATP on susceptibility changes in the identified proteins, we further investigated the effect of ATP on the thermodynamic stability and unfolding kinetics of the proteins. By using pulse proteolysis, we monitored the urea-induced unfolding of the overexpressed proteins in cell lysates without purification.

GAPDH is a glycolytic enzyme that catalyzes the reduction of glyceraldehyde-3-phosphate using NAD$^+$ as cofactor. ATP binding to this enzyme has not been previously known. Unfolding of GAPDH in varying concentrations of urea for 24 h clearly showed a significant decrease in the apparent C_m value (the midpoint of the unfolding transition) from 2.47 ± 0.02 M in the absence of ATPγS to 1.70 ± 0.01 M in the presence of 1.0 mM ATPγS (Fig. 4a). The thermodynamic stability of GAPDH as a tetrameric protein is likely to be dependent on the protein concentration. These C_m values are, therefore, apparent C_m values under the given conditions.[30] Interestingly, both C_m values are lower than 3.0 M urea, which is the urea concentration used for the proteomic screen. The comparison of the unfolding of GAPDH in 3.0 M urea for 2 h (~20%; Fig. 3) and the unfolding of GAPDH in 3.0 M urea for 24 h (~90%; Fig. 4a) indicates that the protein does not reach its conformational equilibrium under the conditions employed for the proteomic screen. Another interesting observation is that, under native conditions, GAPDH is resistant to pulse proteolysis even in the absence of ATPγS. This result rules out the possibility that binding of ATPγS induces a conformational change in GAPDH, which makes the protein susceptible to pulse proteolysis.

Reliable determination of the fraction of folded proteins (f_{Fold}) by pulse proteolysis requires that the
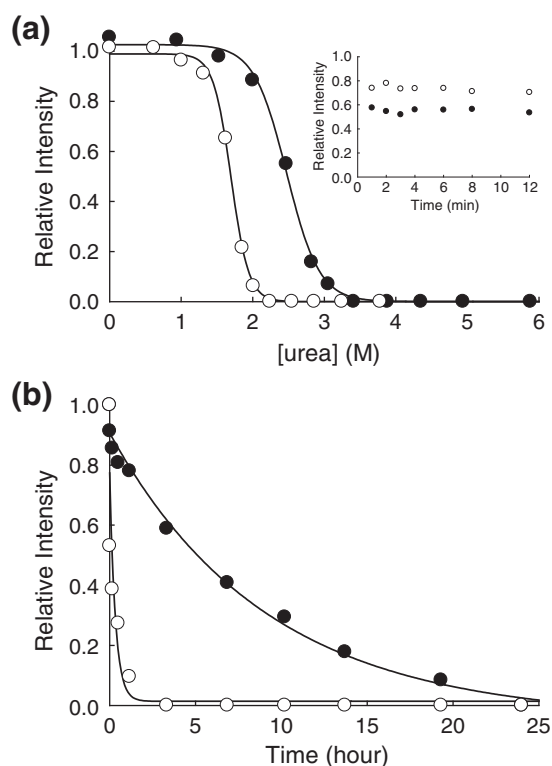
Fig. 4. Effect of ATP on the stability of GAPDH. (a) Equilibrium unfolding of GAPDH in urea probed by pulse proteolysis with (○) and without (●) 1.0 mM ATPγS. Relative intensities are the ratio of the band intensities of the proteins remaining after pulse proteolysis to the band intensity of the undigested protein. The apparent C_m values were determined by fitting the relative intensities to Eq. (2). Inset: Prolonged incubation of GAPDH with 0.2 mg/mL thermolysin in 2.7 M urea without ATPγS (●) and in 1.8 M urea with 1.0 mM ATPγS (○). (b) Unfolding of GAPDH in 3 M urea with (○) and without (●) 1.0 mM ATPγS, monitored by pulse proteolysis. Unfolding kinetic constants were determined by fitting the relative intensities to a first-order rate equation.

folded protein is not digested during the 1-min pulse. To confirm that the folded protein is not susceptible to pulse proteolysis, we typically monitor proteolysis at a urea concentration near the apparent C_m beyond 1 min. The inset to Fig. 4a shows that folded GAPDH is resistant to proteolysis near its apparent C_m whether or not ATPγS is present. Therefore, pulse proteolysis is a valid probe for determining the f_{Fold} of this protein. This result also suggests that, even in the transition zone, ATPγS does not make the folded GAPDH susceptible to pulse proteolysis. Instead, ATPγS seems to decrease the population of native GAPDH in urea by an unknown mechanism. One possible explanation for this unusual destabilizing effect of ATP is that ATP binds and accumulates a nonnative form of GAPDH,

which is susceptible to proteolysis. This hypothesis is currently being tested in our laboratory.

The unfolding kinetics of GAPDH also demonstrates that ATPγS increases the unfolding rate of this protein significantly (Fig. 4b). The unfolding rate constant of GAPDH in 3.0 M urea is increased by ∼30-fold from $(3.0 \pm 0.5) \times 10^{-5} \, s^{-1}$ to $(8 \pm 3) \times 10^{-4} \, s^{-1}$ by 1.0 mM ATPγS. Typically, ligand binding stabilizes the native conformations of proteins and slows protein unfolding. This increase in the unfolding rate by a ligand is quite unusual and may suggest the stabilization of the unfolding transition state, not the native form of GAPDH, by ATPγS. Also, the unfolding kinetics of GAPDH corroborates that, in the absence of ATPγS, the conformational equilibrium of this protein in 3.0 M urea cannot be achieved in 2 h (Fig. 4b).

Dihydrolipoamide dehydrogenase

E3 is a component of three multicomponent enzyme complexes: pyruvate dehydrogenase multienzyme complex, 2-oxoglutarate dehydrogenase complex, and glycine cleavage system.[31–33] These multicomponent enzymes are large protein complexes that are composed of 12–24 monomeric units of several enzymes. For example, the pyruvate dehydrogenase multienzyme complex contains 12 E1 dimers, a 24-subunit E2 core, and 6 E3 dimers. When isolated, E3 exists as a dimeric form bound to flavin adenine dinucleotide (FAD).[34] E3 uses FAD as a cofactor for electron transfer to the final acceptor NAD⁺. From our literature search, we could not find any report showing ATP binding to this enzyme. Because E3 functions as a component of these large enzyme complexes, it is plausible that our proteomic screen has monitored the effect of ATPγS on the stability of the complexes, not that of the dimeric form of E3. Therefore, as seen in Fig. 3, the energetic properties observed with the overexpressed E3 dimer can be different from the property that we observed in the proteomic screen.

Unfolding of E3 probed by pulse proteolysis after 24 h of incubation shows an evident increase in C_m from 2.63 ± 0.05 M to 3.02 ± 0.05 M by 1.0 mM ATPγS (Fig. 5a). This increase in C_m by ATPγS is consistent with the increase in band intensity after pulse proteolysis upon incubation with ATPγS in 3.0 M urea (Fig. 3b). Prolonged incubation with the protease near C_m beyond 1 min confirmed that folded E3 is resistant to pulse proteolysis and that f_{Fold} is determined faithfully by pulse proteolysis (inset to Fig. 5a).

To demonstrate that this ∼0.4 M change in C_m is meaningful, we performed control experiments with proteins that do not interact with ATP, *E. coli* maltose-binding protein, and *E. coli* dihydrofolate reductase (Fig. S1). In both cases, the differences in C_m values in the presence and in the
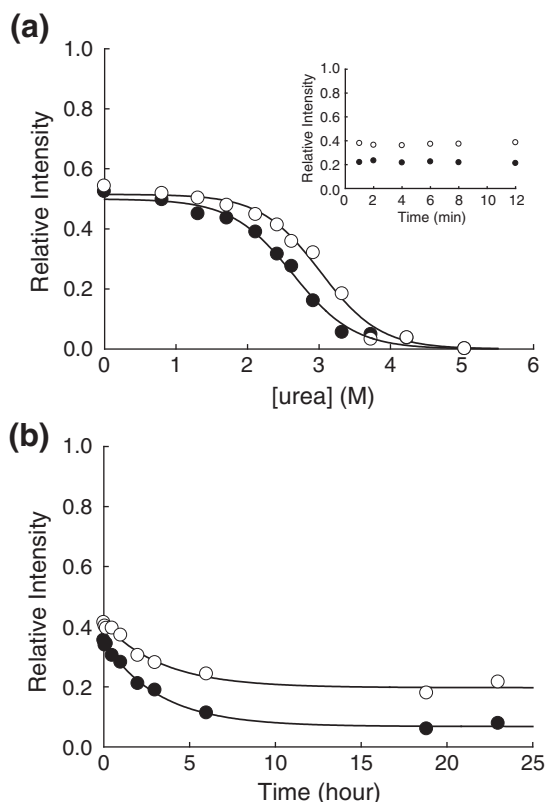
Fig. 5. Effect of ATP on the stability of E3. (a) Equilibrium unfolding of E3 in urea probed by pulse proteolysis with (○) and without (●) 1.0 mM ATPγS. Relative intensities are the ratio of the band intensities of the proteins remaining after pulse proteolysis to the band intensity of the undigested protein. The apparent C_m values were determined by fitting the relative intensities to Eq. (2). Inset: Prolonged incubation of E3 with 0.2 mg/mL thermolysin in 2.8 M urea with (○) and without (●) 1.0 mM ATPγS. (b) Unfolding of E3 in 3 M urea with (○) and without (●) 1.0 mM ATPγS, monitored by pulse proteolysis. Unfolding kinetic constants were determined by fitting the relative intensities to a first-order rate equation.

absence of ATPγS were less than 0.1 M. This small variation in C_m values is consistent with typical standard errors in triplicate measurements of C_m by pulse proteolysis. This result clearly shows that the change in the C_m value of E3 by ATPγS is statistically significant.

The relaxation kinetics of E3 under proteomic screen conditions is not affected by the presence of ATPγS. The relaxation rate constants of E3 in 3.0 M urea are $(8.7 \pm 0.7) \times 10^{-5}$ s^{-1} and $(8.3 \pm 1.3) \times 10^{-5}$ s^{-1} in the absence and in the presence of 1.0 mM ATPγS. This result indicates that the observed ATPγS effect does not result from the relaxation kinetics, but from the increase in f_{Fold}. As for GAPDH, the 2-h incubation for the proteomic screen was not long enough for E3 to reach its conformational

equilibrium in 3.0 M urea, but still produced discernable differences in f_{Fold} (Fig. 5b).

Interestingly, although folded E3 is resistant to pulse proteolysis (inset to Fig. 5a), about 50% of E3 in the cell lysates is readily digested by pulse proteolysis even under native conditions (Fig. 5a). The nature of this fraction of E3 that is susceptible to pulse proteolysis is not clear. One possible explanation is that this susceptible population is misfolded E3 without FAD. FAD is known to be necessary for the maturation of E3 into a stable dimer.[34] However, incubation of the overexpressed E3 in the cell lysate with added FAD for 2 h did not rescue the susceptible fraction (data not shown).

Encouraged by the experimental result suggesting an interaction between E3 and ATP, we investigated if E3 has any structural motif for ATP binding. Although the structure of E3 has not been solved, a structural model is available in MODBASE,[35] which is based on a homologous protein structure from *Pseudomonas putida* [Protein Data Bank (PDB) ID: 1lvl]. For this structural model, the ProFunc server,[36] which matches the structural templates of the functional sites of known proteins in a query structure, identified a match with an ATP binding site of *Pyrococcus horikoshii* L-proline dehydrogenase (PDB ID: 1y56) at around residues 10–20 and 145 (Fig. 6a). ProFunc identified 22 identical residues and 15 similar residues between the two proteins at the local sites. Indeed, the structure of the whole N-terminal domain (residues 1–156) of the model of E3 overlaps well with the structure of a domain (residues 100–213) in the α subunit of L-proline dehydrogenase with a root-mean-square deviation (RMSD) of 3.5 Å (Fig. 6b). This structural analysis and the observed stabilization by ATP suggest that E3 may have a potential ATP binding site. It is notable that this potential ATP binding site in E3 is part of the tentative FAD binding site; according to the structure of the homologous protein from *P. putida*, the identified ATP binding site overlaps with the binding site of the ADP moiety of FAD. This model suggests that ATP may compete with FAD for the same binding site and interfere with the regular function of the protein.

yncE

yncE is a hypothetical protein with no known function. Still, this protein is annotated as an ATP-binding protein in EcoCyc due to the presence of a P-loop, an ATP-binding sequence motif at positions 320–327. A preliminary crystal structure of the protein has been reported in the literature,[37] but the coordinate has not been deposited in the PDB.[38] Still using homology models of yncE available at the EcoliProteins database,[39] we found that the ATP-binding motif region is located at an exposed loop. The location of the binding motif on the surface of the
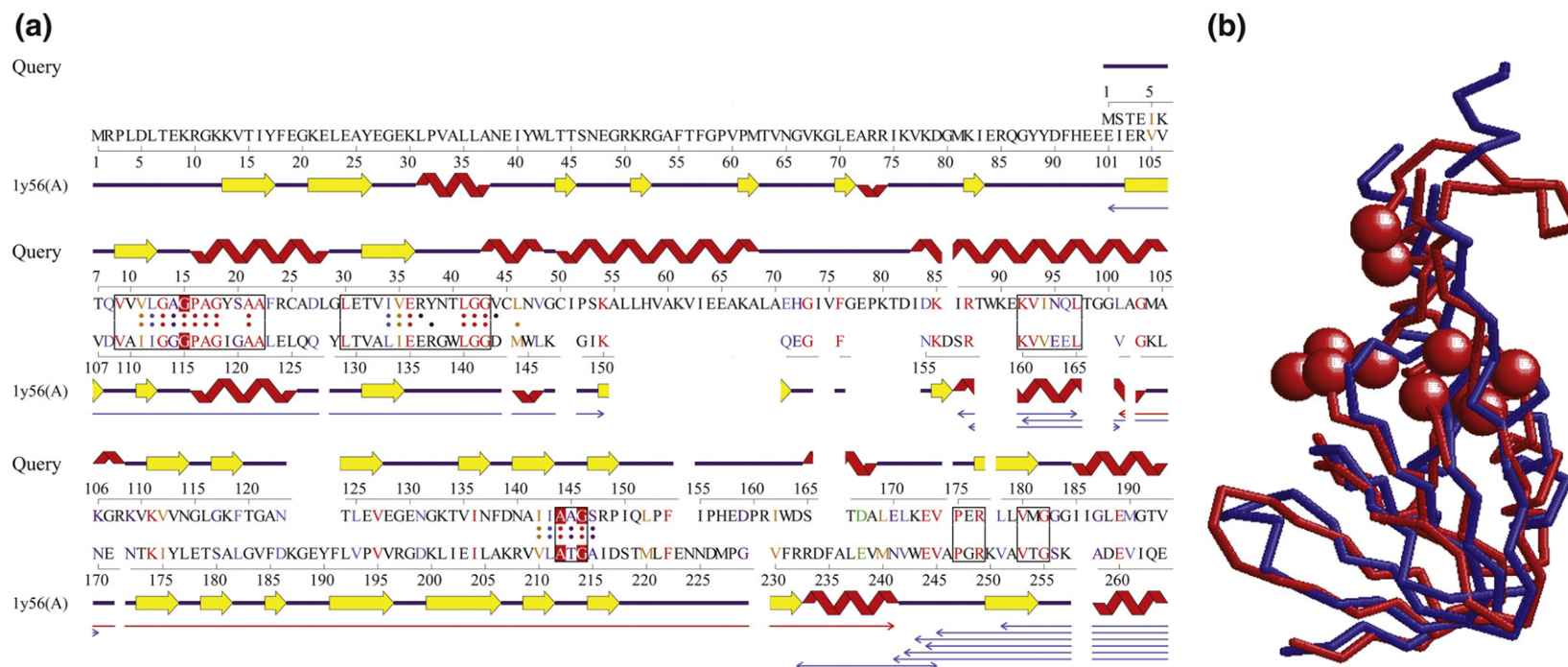
Fig. 6. Structure similarity between the N-terminal domain of E3 and the L-proline dehydrogenase α subunit. (a) Structural alignment between the N-terminal domain of the model of E3 (Query) and the L-proline dehydrogenase α subunit (PDB ID: 1y56A) computed by the ProFunc server.[36] The homology model of E3 was built with MODBASE[35] based on a template structure (PDB ID: 1lvl). The boxes on the sequences of the two proteins represent segments where the sequence identity exceeds 35%. Dots assigned to each sequence indicate residues within 10 Å of the center of the template protein (PDB ID: 1y56A) and hence considered in computing the alignment. The three residues in red, which have been initially matched in the ProFunc search, compose the ATP binding site of the template. The thin arrows below the alignments are structurally similar regions between the two proteins, which can be superimposed within an RMSD of 3.0 Å. The red arrow shows the longest such segment. For more details, see the ProFunc web site (http://www.ebi.ac.uk/thornton-srv/databases/profunc/). (b) Structural superimposition of the N-terminal domain of the model of E3 (blue; residues 1–156) and the homologous region of L-proline dehydrogenase (red; residues 100–213). Spheres indicate the ATP-binding residues of L-proline dehydrogenase. The RMSD of the two structures is 3.48 Å.

protein confirms that the P-loop is likely involved in ATP binding.

Our experimental validation also supports that yncE is indeed an ATP-binding protein. Equilibrium unfolding of yncE in urea shows that ATPγS increases the C_m of this protein from 1.72 ± 0.06 M to 2.11 ± 0.04 M (Fig. 7a). Without ATPγS, a small fraction of yncE ($\sim$10%) is digested by pulse proteolysis even under native conditions. However, prolonged incubation with the protease near C_m beyond 1 min confirmed that folded yncE is resistant to pulse proteolysis whether or not ATPγS is present (inset to Fig. 7a). The digestion of the small fraction of the protein under native conditions may suggest that some fraction of the protein is, in nonnative conformations, susceptible to proteolysis, and ATPγS may have converted this fraction of protein into the native conformation that is resistant to proteolysis. ATPγS slows the unfolding of yncE. The unfolding rate constant of yncE in 3.0 M urea is decreased from $(9.7 \pm 0.5) \times 10^{-5}$ s^{-1} to $(5.0 \pm 0.2) \times 10^{-5}$ s^{-1} by 1.0 mM ATPγS (Fig. 7b). The relaxation of this protein is relatively slow; even in the absence of ATPγS, the relaxation of yncE in 3 M urea is incomplete after 2 h of incubation. This incomplete relaxation explains the significant amount of protein remaining after pulse proteolysis in 3 M urea (Fig. 3), which is actually greater than its C_m values whether or not ATPγS is present.

Discussion

Energetics-based discovery of ATP-binding proteins

Using ATPγS as test molecule, we demonstrated the validity of our energetics-based target identification approach. Out of 10 identified positives, six proteins are already known to interact with ATP based on experimental evidence in the literature. To validate the ATP binding of the four remaining proteins, we investigated the effect of ATP on their thermodynamic stability and unfolding kinetics individually. Except for mlaC, we confirmed that ATP indeed affects their energetic properties (Figs. 3–7). This result clearly demonstrates that our energetics-based target identification approach is a reliable way to excavate putative ATP-binding proteins. Although recombinant mlaC was not confirmed to interact with ATP directly, it is still possible that this protein is stabilized by forming a complex with ATP-binding proteins when the protein exists in a stoichiometric amount.

According to GO terms in EcoCyc, the products of 361 out of 4144 protein-coding genes are annotated as ATP-binding proteins (GO 0005524). Because ATP binding to some proteins still may not be known, ATP-binding proteins are roughly estimated to be $\sim$10% of the proteome. Considering that typically $\sim$500 proteins are observable by 2D gel electrophoresis with the staining method employed in our study, we expect the $\sim$50 ATP-binding proteins to be expressed enough to be identified by 2D gel electrophoresis. Because quite a few proteins still remain folded in 3.0 M urea (the control gel in Fig. 2) and ATP binding to these folded proteins would not make any difference upon pulse proteolysis, the upper limit of the number of identifiable ATP-binding proteins under our experimental conditions would be 20–30 proteins. Although the positive spots were not searched exhaustively, identification of 10 ATP-binding proteins from selected spots, which are well isolated and reproducible from triplicate experiments, suggests that the coverage
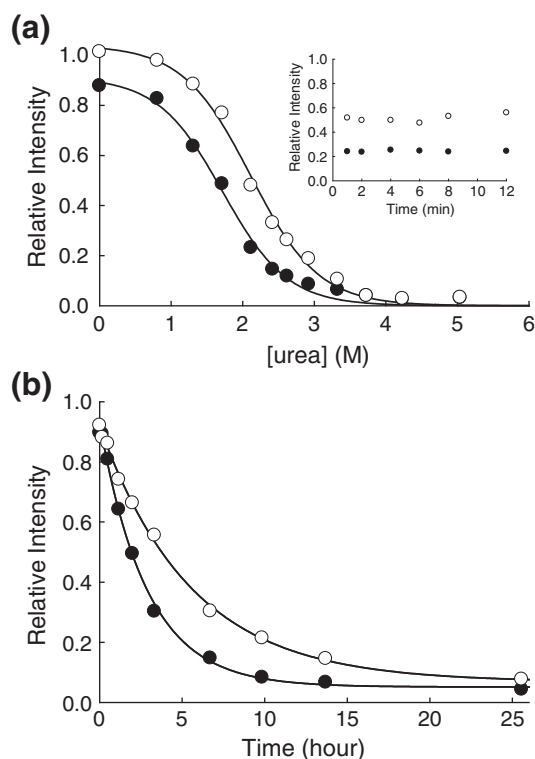

Fig. 7. Effect of ATP on the stability of yncE. (a) Equilibrium unfolding of yncE in urea probed by pulse proteolysis with (○) and without (●) 1.0 mM ATPγS. Relative intensities are the ratio of the band intensities of the proteins remaining after pulse proteolysis to the band intensity of the undigested protein. The apparent C_m values were determined by fitting the relative intensities to Eq. (2). Inset: Prolonged incubation of yncE with 0.2 mg/mL thermolysin in 1.9 M urea with (○) and without (●) 1.0 mM ATPγS. (b) Unfolding of yncE in 3 M urea with (○) and without (●) 1.0 mM ATPγS, monitored by pulse proteolysis. Unfolding kinetic constants were determined by fitting the relative intensities to a first-order rate equation.

of this approach is still within a reasonable range. The limitation in the coverage seems due to the resolving power of 2D gel electrophoresis. More quantitative proteomics platforms, such as differential gel electrophoresis,[40] or various quantitative mass spectrometry approaches, such as stable isotope labeling with amino acids in cell culture (SILAC),[41,42] may improve the proteome coverage significantly.

The result presented shows that our target identification approach is applicable to the identification of enzymes using a metabolite as substrate. Four confirmed ATP-binding proteins from our proteomic screen (glutaminyl-tRNA synthetase, phenylalanyl-tRNA synthetase, succinyl-CoA synthetase, and 6-phosphofructokinase) are known to use ATP for their catalytic functions. Recently, a powerful new proteomic method, dubbed 'activity-based protein profiling,'[43,44] has been introduced to identify proteins with specific enzymatic activities by using mechanism-based conjugation reagents. This approach has been shown to be quite effective in identifying novel enzymes and in monitoring changes in enzymatic activities on a proteomic scale. Our target identification provides an alternative way to identify enzyme–substrate interactions without creating mechanism-based conjugation reagents.

Our target screen identifies not only enzymes that utilize the test molecule as substrate but also proteins that interact with the test molecule through noncatalytic binding sites, which may have regulatory roles. Our screen identified ATP synthase F1-α subunit as an ATP-interacting protein (Table 1). The cytosolic F1 complex of ATP synthase contains two subunits, F1-α and F1-β. The catalytic subunit is F1-β, in which ATP is synthesized. F1-α is not involved in catalysis, but this subunit has been shown to bind ATP with high affinity.[22,45] Although the biochemical role is still unknown,[46,47] ATP binding by F1-α may have a regulatory function. Our screen also identified GroES (Table 1), which forms a heptameric lid to cap the chaperone protein GroEL.[48] GroES has been shown to bind ATP by azido-ATP labeling.[27] While GroEL utilizes ATP for its chaperone function, the role of ATP binding by GroES is not yet known. Still, this ATP binding may have a regulatory role. The identification of ATP synthase F1-α subunit and GroES demonstrates that our target identification approach may discover protein–ligand interactions at noncatalytic sites as well as active sites. It is still noteworthy that F1-α and GroES may be stabilized by ATP indirectly by forming complexes with F1-β and GroEL, respectively.

Our screen identified several novel ATP interactions that have not been reported previously. E3 was not known to interact with ATP. YncE is a hypothetical protein with an unknown function, and its interaction with ATP was suggested only by

a sequence analysis. Validation with recombinant proteins corroborates our proteomic identification of E3 and yncE as putative ATP-binding proteins (Fig. 3). In addition, a potential ATP binding site was identified in the structural model of E3. These proteins exemplify the utility of our target identification and subsequent validation processes, using unfolding energetics and bioinformatics, in discovering the novel biochemical functions and regulatory roles of metabolites. The putative interactions discovered through this approach would provide highly reliable leads, which are valuable in generating hypotheses for further rigorous biochemical and biophysical investigations.

Interestingly, we also found that GAPDH is apparently destabilized in the presence of ATPγS. Our preliminary characterization suggests that natively folded GAPDH does not bind ATPγS, but nonnative conformations of GAPDH may interact with ATPγS. Through this interaction with nonnative conformations, ATP seems to modulate the energetics of unfolding of GAPDH. This unusual effect of ATP on the unfolding of this protein is quite interesting because the ATPγS concentration used in our screen is close to the physiological concentration of ATP in *E. coli* cytosol.[49] It is plausible that ATP may exert a similar effect on this protein *in vivo*. Recent advances in proteomics have enabled investigations of the energetic properties of proteins on a proteomic scale, which uncover proteins with exceptional energetic properties such as resistance to proteolysis[50] or resistance to denaturation by SDS.[51] These studies demonstrate surprising diversity in the energetic properties of proteins in cells and provide valuable insights on how the energetic properties are linked with the functions of the proteins. The finding of the effect of ATP on GAPDH exemplifies that our screen not only reports protein–ligand interactions but also reveals their unusual energetic consequences.

Experimental parameters to be considered

Our screen discovers protein–ligand interactions based on the effect of ligand binding on the conformational energies of target proteins. Due to the diversity in the thermodynamic stabilities and binding affinities of target proteins in a proteome, a single experimental condition would not guarantee the identification of every target. However, by choosing experimental conditions tactfully, one may maximize the efficiency of the screen. Several experimental parameters are worth careful consideration.

First, the concentration of urea is an important parameter that determines the experimental window of proteome coverage. To be identified under a given experimental condition, a target protein needs to be unfolded in the absence of the ligand, but to be

folded in the presence of the ligand (Fig. 8a). This window of urea concentration for identification is unique to each target. If the urea concentration is outside of this window, the target protein would not show significant difference in f_{Fold} upon ligand binding. As a proof-of-principle experiment, we performed our screen at a single urea concentration (3.0 M). Performing screens at several different urea concentrations may further increase the coverage of the proteome. The exceptional stability of thermolysin allows pulse proteolysis to be performed in up to 8 M urea.[18]
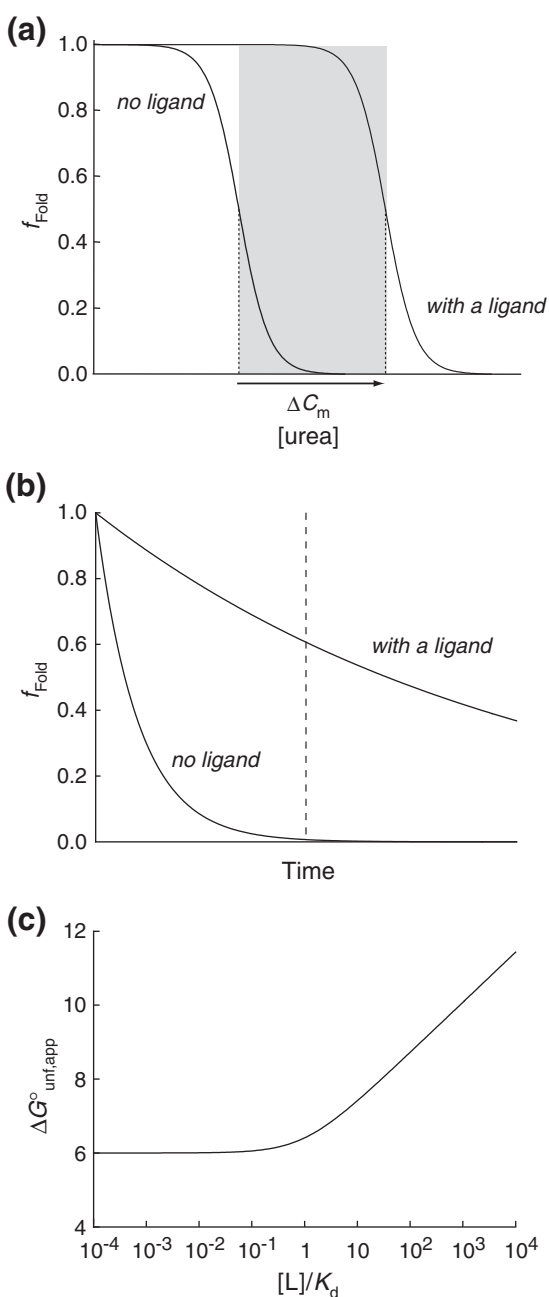
Second, the duration of the incubation in urea before pulse proteolysis is also a parameter that affects screening results. Our screen does not require a complete equilibration of the proteins under a given condition. Ligand binding frequently slows the unfolding of target proteins. Even before conformational equilibrium is achieved, target proteins may show a difference in f_{Fold} in the presence of the ligand due to the effect on unfolding kinetics (Fig. 8b). YncE is an example of this case. This protein does not reach its conformational equilibrium in 2 h under our experimental conditions. However, the decreased unfolding rate of the protein in the presence of ATP still results in a difference in the f_{Fold} of the protein (Fig. 7b). Actually, a prolonged incubation of this protein under the experimental conditions would lead to a complete unfolding of the protein regardless of the presence of the ligand, and no yncE would be detectable after pulse proteolysis even in the presence of ATP. Moreover, a prolonged incubation of a cell lysate at room temperature may cause adverse effects on the integrity of the proteins in the lysate. Considering these factors, we chose to incubate the lysate for 2 h. Again, this duration of incubation can be adjusted according to the conformational energetic characteristics of the proteins in a given proteome.

Third, the concentration of the ligand is another important parameter that determines the stringency of the screen. When a target protein binds a ligand through its native conformation, the thermodynamic stabilization of the protein becomes a function of the ligand concentration, as shown below:

$$\Delta G^{\circ}_{unf,app} = \Delta G^{\circ}_{unf} + RT\ln\left(1 + \frac{[L]}{K_d}\right) \quad (1)$$

where $\Delta G^{\circ}_{unf,app}$ is the apparent thermodynamic stability of the protein in the presence of the ligand,

Fig. 8. Experimental parameters in energetics-based target identification. (a) The window of urea concentrations allowing the detection of ligand binding by energetics-based target identification. The gray box indicates the range of urea concentrations between the C_m value of a protein in the presence of a certain concentration of a ligand and the C_m value of a protein in the absence of a certain concentration of a ligand. In this range, the target would show a significant difference in the fraction of folded protein (f_{Fold}), as probed by pulse proteolysis. (b) Incubation time as an experimental parameter. Even when the system does not reach equilibrium at the time of pulse proteolysis (broken line), the target may show a significant difference in f_{Fold} due to the slow unfolding in the presence of the ligand. For the demonstration of the principle in this example, the unfolding rate constant is decreased by 10-fold in the presence of ligand. (c) Ligand concentration as an experimental parameter. The apparent thermodynamic stability ($\Delta G^{\circ}_{unf,app}$) of a protein with the global stability (ΔG°_{unf}) of 6.0 kcal/mol is calculated with varying ratios of ligand concentration to dissociation equilibrium constant ($[L]/K_d$) using Eq. (1).

$\Delta G^\circ_{\text{unf}}$ is the thermodynamic stability of the protein in the absence of the ligand, [L] is the ligand concentration, and K_{d} is the dissociation equilibrium constant of the complex. This quantitative relationship between $\Delta G^\circ_{\text{unf,app}}$ and [L] is shown in Fig. 8c. When [L] $\gg K_{\text{d}}$, the apparent stability of the target protein is linearly proportional to ln[L]. As the target protein is stabilized further, the effective window of the experimental parameters for identification (shaded area in Fig. 8a) is broadened. The higher ligand concentration is desirable to maximize the coverage of the screen. However, as the ligand concentration is increased, the chance of observing nonspecific interactions may also increase. Figure 8c also demonstrates how the stringency of the screen can be controlled by the ligand concentration. A target with a K_{d} value much smaller than the ligand concentration ([L]/$K_{\text{d}} \gg 1$) would have a better chance for identification due to its greater stabilization with the given concentration of the ligand. Targets with K_{d} values greater than the ligand concentration ([L]/$K_{\text{d}} < 1$) do not experience any significant stabilization. Therefore, a suitable ligand concentration should be determined based on the desired stringency of the screen.

Comparison with alternative approaches

Our energetics-based target identification approach by pulse proteolysis overcomes several limitations of traditional target identification methods. First, this method is performed without modifying the test molecules, and possible adverse effects of the modification on binding are eliminated. The direct use of test molecules in target identification significantly reduces experimental time and cost. Second, the stringency of the screen can be modulated by changing the ligand concentration in our screen. This modulation of the stringency is quite valuable in minimizing nonspecific binding. Also, because the identification is based on the effect of ligand binding on the conformational energies of the target proteins, the method is universally applicable to any protein–ligand interaction, regardless of the functional consequences of the binding events. Our result suggests that the method can be used to identify not only interactions with substrates but also interactions with regulatory molecules, signaling molecules, inhibitors, and drugs.

The concept of energetics-based target identification has been demonstrated recently by other laboratories also by identifying drug targets from changes in the proteolytic susceptibility[16] and methionine oxidation[17] of target proteins upon the addition of a drug. Our approach using pulse proteolysis is based on physical and chemical principles distinct from those of the two approaches. Although proteolysis is employed as a probe, our approach differs in several aspects from the one

based on the change in proteolytic susceptibility.[16] Proteolytic susceptibility is determined by accessibility to a nonnative high-energy conformation (frequently partially unfolded forms) under native conditions.[52,53] Our approach is not strongly dependent on the changes in proteolytic susceptibility due to the brief incubation (1 min) with the protease. As shown with overexpressed proteins, folded GAPDH, E3, and yncE are not digested through pulse proteolysis even in the absence of ATP (insets to Figs. 4, 5, and 7). The observed changes in the amount of proteins remaining after proteolysis are not derived from the difference in susceptibility but from the difference in the thermodynamics and kinetics of unfolding. Moreover, ligand binding may not result in a change in proteolytic susceptibility. Our previous study on *E. coli* maltose-binding protein clearly showed that maltose does not affect the proteolytic susceptibility of this protein.[54] The proteolysis of this protein occurs through partial unfolding that does not involve the maltose binding site. The protein is digested without releasing the bound maltose, and the rates of proteolysis in the presence and in the absence of maltose are identical. Still, the thermodynamic stability and unfolding kinetics of maltose-binding protein are strongly dependent on the ligand concentration.[18,19] Therefore, our method, based on the changes in unfolding behaviors in urea, may identify the targets that the susceptibility-based methods cannot identify.

The target identification method based on methionine oxidation[17] is similar to our approach in that the effect of ligand binding is monitored by the change in the f_{Fold} of the target proteins in chemical denaturants. By employing quantitative mass spectrometry, they monitored the change in methionine oxidation at multiple denaturant concentrations, while we used only a single urea concentration in our approach with 2D gel electrophoresis. Also, their mass spectrometry approach using isobaric mass tags allowed a quantitative comparison between samples with the ligand and samples without the ligand. However, this approach covers only the proteins containing buried methionine residues whose susceptibility to oxidation increases significantly upon unfolding. Our approach uses nonspecific proteases, which can digest any unfolded protein regardless of its amino acid composition or sequence.

As in any other proteomic methods, our approach cannot cover the complete proteome. Approximately 20% of *E. coli* soluble proteins are susceptible to pulse proteolysis even under native conditions (Y. R. Na and C.P., unpublished result). Unless the association with the test molecule confers the proteins resistance to pulse proteolysis, our method would not identify these proteins. Also, quite a few *E. coli* proteins, possibly including membrane proteins, remain folded even in 8 M urea (Y. R. Na and

C.P., unpublished result). Our method would not be able to identify targets in this category of proteins either. Therefore, these energetics-based target identification approaches (proteolytic susceptibility, methionine oxidation, and pulse proteolysis) are actually complementary to each other. The application of two or more of these approaches may increase the proteome coverage significantly.

Here we demonstrate the principle of our energetics-based target identification approach by employing pulse proteolysis and 2D gel electrophoresis. The results clearly show that this approach is valid and powerful in discovering unknown interactions between proteins and ligands (metabolites, signaling molecules, drugs, and so forth) on a proteomic scale. When combined with more powerful proteomics tools, the proteome coverage of the approach would be improved significantly. Although we demonstrated its feasibility by using *E. coli* lysates in this study, we expect this approach to be also applicable to other proteomes, including human proteomes, in which many medically significant questions can be pursued with unprecedented efficiency.

Materials and Methods

E. coli lysate preparation

E. coli K-12 (MG1655) cells were grown at 37 °C to reach log phase ($OD_{600} = 0.6$) and harvested by centrifugation. The cell pellet was washed with ice-cold water and resuspended in 20 mM Tris–HCl buffer (pH 8.0) containing 10 mM ethylenediaminetetraacetic acid (EDTA) and 1 mM DTT. The resuspended cells were lysed on ice by sonication. Cell debris was removed by centrifugation. To remove metabolites, which may interfere with the following procedures, we first dialyzed the lysates against 20 mM Tris–HCl (pH 8.0) containing 1.0 M NaCl and 0.1% β-mercaptoethanol, and then dialyzed them against 20 mM Tris–HCl buffer (pH 8.0) containing 50 mM NaCl and 0.1% β-mercaptoethanol. Finally, the lysate was dialyzed against 20 mM Tris–HCl buffer (pH 8.0) containing 50 mM NaCl and 1 mM DTT. Dialysis was performed at 4 °C. The total protein concentration was estimated spectrophotometrically by using absorbance at 260 nm and 280 nm to subtract the contribution of nucleic acid.[55,56] The resulting cell lysate was aliquoted and stored at −80 °C for further experiments.

Pulse proteolysis

To identify proteins stabilized by ATP binding, we incubated the cell lysate at 25 °C for 2 h in 20 mM Tris–HCl buffer (pH 8.0) containing 1.0 mM ATPγS (Sigma-Aldrich, St. Louis, MO), 50 mM NaCl, 5 mM $MgCl_2$, 1 mM tris(2-carboxyethyl)phosphine (TCEP), and 3.0 M urea. To prevent the ligand from being hydrolyzed by endogenous enzymes, we used ATPγS, a nonhydrolyzable ATP analogue. A control reaction was prepared in an identical manner, but without ATPγS. After the incubation, pulse proteolysis was performed as described previously.[18,57]

Briefly, thermolysin was added to the final concentration of 0.20 mg/mL, and the reaction was quenched after 1 min by 16 μM phosphoramidon and then again by 10 mM EDTA. Phosphoramidon, a competitive inhibitor for thermolysin, was used to suppress the autolysis of thermolysin,[58] which complicates 2D gel analysis by producing multiple fragments of the protease. EDTA inactivates thermolysin by chelating Ca^{2+}, which is essential for the structural integrity of the enzyme.[59]

Two-dimensional gel electrophoresis

To remove salts prior to 2D gel electrophoresis, we exchanged the buffer of the quenched reaction with 8.0 M urea using a Protein Desalting Spin Column (Pierce, Rockford, IL). The resulting solution in 8.0 M urea was then mixed with an equivalent volume of a 2-fold concentrated stock solution for isoelectric focusing to make the final solution of 8.0 M urea, 2% 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate, 20 mM DTT, 0.002% bromophenol blue, and 0.5% IPG buffer (3–10 NL) (GE Healthcare, Pittsburgh, PA). The solution was then centrifuged to remove any precipitant. Of the resulting solution, 250 μL (~ 500 μg of total protein) was used to rehydrate a 13-cm 3–10 NL DryStrip (GE Healthcare). The hydrated strip was focused with Ettan IPGphor II (GE Healthcare) according to the manufacturer's instruction. Separation on the second dimension was performed in 15% (wt/vol) continuous SDS-PAGE gel with SE 600 Ruby Complete (GE Healthcare). The gels were stained with colloidal Coomassie staining solution.[60] Spots showing different intensities on the gels of a sample and its control were identified by visual inspection. To rule out false positives from artifacts in 2D gel electrophoresis, we performed the same experiment in triplicate. Spots that were identified consistently in the repeated experiments were selected for further characterization.

In-gel digestion

In-gel digestion was performed with selected spots on 2D gels, as described previously.[50] Peptides extracted from gel pieces were desalted using Ziptip$_{μ-C18}$ (Millipore, Bedford, MA). One microliter of the desalted eluant was mixed with an equal volume of 5 mg/mL α-cyano-4-hydroxy cinnamic acid in 60% acetonitrile containing 0.1% trifluoroacetic acid and allowed to dry on a matrix-assisted laser desorption/ionization target plate. Mass spectra were acquired on a 4800 Plus MALDI TOF/TOF™ Analyzer (Applied Biosystems, Foster City, CA). Tandem mass spectrometry analysis of each sample was performed on the top 10 peaks from each mass spectrum. Proteins were identified from the Swiss-Prot database by using the MASCOT database search engine[61] in GPS Explorer (Applied Biosystems). All identified proteins have significant MASCOT scores (>100) and GPS Explorer protein confidence indices (>95%).

Confirmation with recombinant proteins

To confirm the results from the proteomic screen, we cloned GAPDH, yncE, E3, and mlaC into a pAED4

expression vector by amplifying the corresponding genes from *E. coli* K-12 genomic DNA by polymerase chain reaction. The cloned proteins were overexpressed in BL21 (DE3) pLysS by induction with IPTG. After cell lysis, the supernatants were collected by centrifugation. The effect of ATPγS on the unfolding of the recombinant proteins in urea was determined by pulse proteolysis under the same conditions as the proteomic screen with the *E. coli* cell lysate. The lysates of cells overexpressing the cloned proteins were used for pulse proteolysis without further purification, as described previously.[18,19] Briefly, the lysates were incubated in 20 mM Tris–HCl buffer (pH 8.0) containing 50 mM NaCl, 5.0 mM MgCl$_2$, 1.0 mM TCEP, and 3.0 M urea with and without 1.0 mM ATPγS at 25 °C for 2 h. Pulse proteolysis was performed to digest the unfolded protein. The remaining proteins were determined by quantifying the band intensities on SDS-PAGE gels by ImageJ, an image analysis software†.

Determination of C_m

To determine the influence of ATPγS on global stability, we determined the C_m values by pulse proteolysis, as described previously.[18] Briefly, the recombinant proteins were incubated for 24 h at 25 °C in 20 mM Tris–HCl buffer (pH 8.0) containing 50 mM NaCl, 5.0 mM MgCl$_2$, 1.0 mM TCEP, and varying concentrations of urea. The lysates of the cells overexpressing the recombinant proteins were used for this experiment without purification. Pulse proteolysis was performed with 0.20 mg/mL thermolysin for 1 min to digest the unfolded protein. The amount of remaining protein in each reaction was determined by quantifying the band intensities of intact proteins on SDS-PAGE gels with ImageJ.

The C_m values were determined by fitting the band intensities to the following equation:

$$I = I_0 \left(\frac{1}{1 + \exp(m(C_m - [\text{urea}]) / RT)} \right) \quad (2)$$

where I is the observed band intensity, I_0 is the band intensity of the protein digested by pulse proteolysis under native conditions, and m is the dependence of global stability on urea.

Relaxation kinetics

The relaxation kinetics of the recombinant proteins in 3.0 M urea were determined by pulse proteolysis, as described previously.[19] The lysates of the cells over-expressing the recombinant proteins were used for this experiment without purification. Unfolding was initiated by adding urea to the cell lysates with the recombinant proteins. The resulting unfolding conditions were 20 mM Tris–HCl (pH 8.0), 50 mM NaCl, 5.0 mM MgCl$_2$, 1.0 mM TCEP, and 3.0 M urea. The progress of unfolding was monitored by performing pulse proteolysis with aliquots of the reaction at designated time points. The amounts of proteins remaining after pulse proteolysis were determined by quantifying band intensities on SDS-PAGE gels

with ImageJ. The relaxation kinetic constants were determined by fitting the band intensities to the first-order rate equation.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2011.02.026

References

1. Mercier, K. A., Baran, M., Ramanathan, V., Revesz, P., Xiao, R., Montelione, G. T. & Powers, R. (2006). FAST-NMR: functional annotation screening technology using NMR spectroscopy. *J. Am. Chem. Soc.* **128**, 15292–15299.
2. Luesch, H., Chanda, S. K., Raya, R. M., DeJesus, P. D., Orth, A. P., Walker, J. R. *et al.* (2006). A functional genomics approach to the mode of action of apratoxin A. *Nat. Chem. Biol.* **2**, 158–167.
3. Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S. *et al.* (2007). Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* **25**, 1035–1044.
4. Hantschel, O., Rix, U., Schmidt, U., Burckstummer, T., Kneidinger, M., Schutze, G. *et al.* (2007). The Btk tyrosine kinase is a major target of the Bcr-Abl inhibitor dasatinib. *Proc. Natl Acad. Sci. USA*, **104**, 13283–13288.
5. Missner, E., Bahr, I., Badock, V., Lucking, U., Siemeister, G. & Donner, P. (2009). Off-target decoding of a multitarget kinase inhibitor by chemical proteomics. *ChemBioChem*, **10**, 1163–1174.
6. Burdine, L. & Kodadek, T. (2004). Target identification in chemical genetics: the (often) missing link. *Chem. Biol.* **11**, 593–597.
7. Terstappen, G. C., Schlupen, C., Raggiaschi, R. & Gaviraghi, G. (2007). Target deconvolution strategies in drug discovery. *Nat. Rev. Drug Discov.* **6**, 891–903.
8. Ong, S. E., Schenone, M., Margolin, A. A., Li, X., Do, K., Doud, M. K. *et al.* (2009). Identifying the proteins to which small-molecule probes and drugs bind in cells. *Proc. Natl Acad. Sci. USA*, **106**, 4617–4622.

† http://rsbweb.nih.gov/ij/

9. Chan, J. N., Nislow, C. & Emili, A. (2009). Recent advances and method development for drug target identification. *Trends Pharmacol. Sci.* **31**, 82–88.

10. Godl, K., Wissing, J., Kurtenbach, A., Habenberger, P., Blencke, S., Gutbrod, H. *et al.* (2003). An efficient proteomics method to identify the cellular targets of protein kinase inhibitors. *Proc. Natl Acad. Sci. USA*, **100**, 15434–15439.

11. Rix, U. & Superti-Furga, G. (2009). Target profiling of small molecules by chemical proteomics. *Nat. Chem. Biol.* **5**, 616–624.

12. Sleno, L. & Emili, A. (2008). Proteomic methods for drug target discovery. *Curr. Opin. Chem. Biol.* **12**, 46–54.

13. Giaever, G., Flaherty, P., Kumm, J., Proctor, M., Nislow, C., Jaramillo, D. F. *et al.* (2004). Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl Acad. Sci. USA*, **101**, 793–798.

14. Parsons, A. B., Lopez, A., Givoni, I. E., Williams, D. E., Gray, C. A., Porter, J. *et al.* (2006). Exploring the mode-of-action of bioactive compounds by chemical–genetic profiling in yeast. *Cell*, **126**, 611–625.

15. Hoon, S., Smith, A. M., Wallace, I. M., Suresh, S., Miranda, M., Fung, E. *et al.* (2008). An integrated platform of genomic assays reveals small-molecule bioactivities. *Nat. Chem. Biol.* **4**, 498–506.

16. Lomenick, B., Hao, R., Jonai, N., Chin, R. M., Aghajan, M., Warburton, S. *et al.* (2009). Target identification using Drug Affinity Responsive Target Stability (DARTS). *Proc. Natl Acad. Sci. USA*, **106**, 21984–21989.

17. West, G. M., Tucker, C. L., Xu, T., Park, S. K., Han, X., Yates, J. R., III & Fitzgerald, M. C. (2010). Quantitative proteomics approach for identifying protein–drug interactions in complex mixtures using protein stability measurements. *Proc. Natl Acad. Sci. USA*, **107**, 9078–9082.

18. Park, C. & Marqusee, S. (2005). Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nat. Methods*, **2**, 207–212.

19. Na, Y. R. & Park, C. (2009). Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci.* **18**, 268–276.

20. Kim, M. S., Song, J. & Park, C. (2009). Determining protein stability in cell lysates by pulse proteolysis and Western blotting. *Protein Sci.* **18**, 1051–1059.

21. Schlebach, J. P., Kim, M.-S., Joh, N. H., Bowie, J. U. & Park, C. (2011). Probing membrane protein unfolding with pulse proteolysis. *J. Mol. Biol.* **406**, 545–551.

22. Dunn, S. D. & Futai, M. (1980). Reconstitution of a functional coupling factor from the isolated subunits of *Escherichia coli* F1 ATPase. *J. Biol. Chem.* **255**, 113–118.

23. Bhattacharyya, T., Bhattacharyya, A. & Roy, S. (1991). A fluorescence spectroscopic study of glutaminyl-tRNA synthetase from *Escherichia coli* and its implications for the enzyme mechanism. *Eur. J. Biochem.* **200**, 739–745.

24. Kosakowski, H. M. & Holler, E. (1973). Phenylalanyl-tRNA synthetase from *Escherichia coli* K10. Synergistic coupling between the sites for binding of L-phenylalanine and ATP. *Eur. J. Biochem.* **38**, 274–282.

25. Johnson, J. L. & Reinhart, G. D. (1992). MgATP and fructose 6-phosphate interactions with phosphofructokinase from *Escherichia coli. Biochemistry*, **31**, 11510–11518.

26. Joyce, M. A., Fraser, M. E., Brownie, E. R., James, M. N., Bridger, W. A. & Wolodko, W. T. (1999). Probing the nucleotide-binding site of *Escherichia coli* succinyl-CoA synthetase. *Biochemistry*, **38**, 7273–7283.

27. Martin, J., Geromanos, S., Tempest, P. & Hartl, F. U. (1993). Identification of nucleotide-binding regions in the chaperonin proteins GroEL and GroES. *Nature*, **366**, 279–282.

28. Keseler, I. M., Collado-Vides, J., Gama-Castro, S., Ingraham, J., Paley, S., Paulsen, I. T. *et al.* (2005). EcoCyc: a comprehensive database resource for *Escherichia coli. Nucleic Acids Res.* **33**, D334–D337.

29. Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S. *et al.* (2006). The PROSITE database. *Nucleic Acids Res.* **34**, D227–D230.

30. Park, C. & Marqusee, S. (2004). Analysis of the stability of multimeric proteins by effective ΔG and effective *m*-values. *Protein Sci.* **13**, 2553–2558.

31. Pettit, F. H. & Reed, L. J. (1967). Alpha-keto acid dehydrogenase complexes: 8. Comparison of dihydrolipoyl dehydrogenases from pyruvate and alpha-ketoglutarate dehydrogenase complexes of *Escherichia coli. Proc. Natl Acad. Sci. USA*, **58**, 1126–1130.

32. Guest, J. R. & Creaghan, I. T. (1972). Lipoamide dehydrogenase mutants of *Escherichia coli* K 12. *Biochem. J.* **130**, 8P.

33. Steiert, P. S., Stauffer, L. T. & Stauffer, G. V. (1990). The lpd gene product functions as the L protein in the *Escherichia coli* glycine cleavage enzyme system. *J. Bacteriol.* **172**, 6142–6144.

34. Lindsay, H., Beaumont, E., Richards, S. D., Kelly, S. M., Sanderson, S. J., Price, N. C. & Lindsay, J. G. (2000). FAD insertion is essential for attaining the assembly competence of the dihydrolipoamide dehydrogenase (E3) monomer from *Escherichia coli. J. Biol. Chem.* **275**, 36665–36670.

35. Pieper, U., Eswar, N., Davis, F. P., Braberg, H., Madhusudhan, M. S., Rossi, A. *et al.* (2006). MODBASE: a database of annotated comparative protein structure models and associated resources. *Nucleic Acids Res.* **34**, D291–D295.

36. Laskowski, R. A., Watson, J. D. & Thornton, J. M. (2005). ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res.* **33**, W89–W93.

37. Baba-Dikwa, A., Thompson, D., Spencer, N. J., Andrews, S. C. & Watson, K. A. (2008). Overproduction, purification and preliminary X-ray diffraction analysis of YncE, an iron-regulated Sec-dependent periplasmic protein from *Escherichia coli. Acta Crystallogr. Sect. F*, **64**, 966–969.

38. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. *et al.* (2000). The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242.

39. Yang, Y. D., Spratt, P., Chen, H., Park, C. & Kihara, D. Sub-AQUA: real-value quality assessment of protein structure models. *Protein Eng. Des. Sel.* 23, 617–632.

40. Ünlü, M., Morgan, M. E. & Minden, J. S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*, **18**, 2071–2077.

41. Aebersold, R. & Mann, M. (2003). Mass spectrometry-based proteomics. *Nature*, **422**, 198–207.

42. Mann, M. (2006). Functional and quantitative proteomics using SILAC. *Nat. Rev. Mol. Cell Biol.* **7**, 952–958.

43. Leung, D., Hardouin, C., Boger, D. L. & Cravatt, B. F. (2003). Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nat. Biotechnol.* **21**, 687–691.

44. Evans, M. J., Saghatelian, A., Sorensen, E. J. & Cravatt, B. F. (2005). Target discovery in small-molecule cell-based screens by *in situ* proteome reactivity profiling. *Nat. Biotechnol.* **23**, 1303–1307.

45. Senda, M., Kanazawa, H., Tsuchiya, T. & Futai, M. (1983). Conformational change of the α subunit of *Escherichia coli* F1 ATPase: ATP changes the trypsin sensitivity of the subunit. *Arch. Biochem. Biophys.* **220**, 398–404.

46. Perlin, D. S., Latchney, L. R., Wise, J. G. & Senior, A. E. (1984). Specificity of the proton adenosinetriphosphatase of *Escherichia coli* for adenine, guanine, and inosine nucleotides in catalysis and binding. *Biochemistry*, **23**, 4998–5003.

47. Myers, J. A. & Boyer, P. D. (1983). Catalytic properties of the ATPase on submitochondrial particles after exchange of tightly bound nucleotides under different steady state conditions. *FEBS Lett.* **162**, 277–281.

48. Hunt, J. F., Weaver, A. J., Landry, S. J., Gierasch, L. & Deisenhofer, J. (1996). The crystal structure of the GroES co-chaperonin at 2.8 Å resolution. *Nature*, **379**, 37–45.

49. Schneider, D. A. & Gourse, R. L. (2004). Relationship between growth rate and ATP concentration in *Escherichia coli*: a bioassay for available cellular ATP. *J. Biol. Chem.* **279**, 8262–8268.

50. Park, C., Zhou, S., Gilmore, J. & Marqusee, S. (2007). Energetics-based protein profiling on a proteomic scale: identification of proteins resistant to proteolysis. *J. Mol. Biol.* **368**, 1426–1437.

51. Xia, K., Manning, M., Hesham, H., Lin, Q., Bystroff, C. & Colon, W. (2007). Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE. *Proc. Natl Acad. Sci. USA*, **104**, 17329–17334.

52. Hubbard, S. J. (1998). The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta*, **1382**, 191–206.

53. Park, C. & Marqusee, S. (2004). Probing the high energy states in proteins by proteolysis. *J. Mol. Biol.* **343**, 1467–1476.

54. Chang, Y. & Park, C. (2009). Mapping transient partial unfolding by protein engineering and native-state proteolysis. *J. Mol. Biol.* **393**, 543–556.

55. Grimsley, G. R. & Pace, C. N. (2004). Spectrophotometric determination of protein concentration. *Curr. Protoc. Protein Sci.*; Unit 3.1.

56. Mach, H., Middaugh, C. R. & Denslow, N. (2001). Determining the identity and purity of recombinant proteins by UV absorption spectroscopy. *Curr. Protoc. Protein Sci.*; Unit 7.2.

57. Park, C. & Marqusee, S. (2006). Quantitative determination of protein stability and ligand binding by pulse proteolysis. *Curr. Protoc. Protein Sci.* **46**, 20.11.14–20.11.21.

58. Wildes, D., Anderson, L. M., Sabogal, A. & Marqusee, S. (2006). Native state energetics of the Src SH2 domain: evidence for a partially structured state in the denatured ensemble. *Protein Sci.* **15**, 1769–1779.

59. Dahlquist, F. W., Long, J. W. & Bigbee, W. L. (1976). Role of calcium in the thermal stability of thermolysin. *Biochemistry*, **15**, 1103–1111.

60. Neuhoff, V., Arold, N., Taube, D. & Ehrhardt, W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie brilliant blue G-250 and R-250. *Electrophoresis*, **9**, 255–262.

61. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, **20**, 3551–3567.