Identifying the subproteome of kinetically stable proteins via diagonal 2D SDS/PAGE

Ke Xia*†, Marta Manning*†, Helai Hesham‡, Qishan Lin§, Christopher Bystroff†‡, and Wilfredo Colón*†¶

Departments of *Chemistry and Chemical Biology and ‡Biology, and †Center for Biotechnology and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180; and §UAlbany Proteomics Facility, Center for Functional Genomics, University at Albany, Rensselaer, NY 12144

Edited by S. Walter Englander, University of Pennsylvania School of Medicine, Philadelphia, PA, and approved September 17, 2007 (received for review June 8, 2007)

Most proteins are in equilibrium with partially and globally unfolded conformations. In contrast, kinetically stable proteins (KSPs) are trapped by an energy barrier in a specific state, unable to transiently sample other conformations. Among many potential roles, it appears that kinetic stability (KS) is a feature used by nature to allow proteins to maintain activity under harsh conditions and to preserve the structure of proteins that are prone to misfolding. The biological and pathological significance of KS remains poorly understood because of the lack of simple experimental methods to identify this property and its infrequent occurrence in proteins. Based on our previous correlation between KS and a protein's resistance to the denaturing detergent SDS, we show here the application of a diagonal 2D (D2D) SDS/PAGE assay to identify KSPs in complex mixtures. We applied this method to the lysate of *Escherichia coli* and upon proteomics analysis have identified 50 nonredundant proteins that were SDS-resistant (i.e., kinetically stable). Structural and functional analyses of a subset (44) of these proteins with known 3D structure revealed some potential structural and functional biases toward and against KS. This simple D2D SDS/PAGE assay will allow the widespread investigation of KS, including the proteomics-level identification of KSPs in different systems, potentially leading to a better understanding of the biological and pathological significance of this intriguing property of proteins.

electrophoresis | kinetic trap | proteomics | sodium dodecyl sulfate | stability

Kinetic stability (KS) refers to the presence of an unusually high energy barrier that traps proteins in their native conformations, thereby dramatically decreasing their unfolding rate under a variety of conditions (1). This barrier allows kinetically stable proteins (KSPs) to maintain their native fold and activity for longer periods of time than would be expected for typical proteins, even in harsh environments. This unique feature is likely to play important biological roles, including the regulation of protein turnover, the protection of some proteins from proteolysis (1), and the entrapment of aggregation-prone proteins in their native state (2). Perhaps the best examples to illustrate the importance of KS are the proteins α-lytic protease and transthyretin (TTR). α-Lytic protease is synthesized with a pro-region that is essential for folding (1). Upon folding, α-lytic protease becomes active and cleaves off the pro region, leaving itself kinetically trapped in an active metastable conformation that is less stable than the unfolded state (3). In the case of TTR, the high thermodynamic and kinetic stability (KS) of the native tetramer is often compromised by mutation, leading to the population of a monomeric conformation that is the precursor of the aggregated species linked to the disease familial amyloid polyneuropathy (4). The large difference in the thermodynamic stability of α-lytic protease and TTR shows that KS need not correlate to thermodynamic stability because the former depends only on the unfolding rate, whereas the latter is defined by the ratio of the folding and unfolding rates.

Despite recent progress (5–7), the structural and chemical-physical basis of protein KS remains poorly understood, perhaps because of the small number of KSPs known and the lack of a simple method to identify proteins with this property. The most common method for probing the KS of proteins involves measuring their unfolding rates at different concentrations of denaturant and then extrapolating to estimate the unfolding rate in the absence of denaturant. This method is not accessible to many laboratories, and therefore very few KSPs have been identified to date (5). The enlargement of the database of KSPs with known 3D structures would facilitate studies to elucidate the chemical and physical basis of KS. This knowledge could then be applied to protein engineering endeavors aimed at increasing the KS of proteins for various applications.

In previous work (5), we demonstrated a correlation between KS and the resistance of proteins to denaturation by SDS, resulting in a simple PAGE-based assay that is very effective for probing the KS of proteins. Here, we show the extension of this method via a diagonal 2D (D2D) SDS/PAGE assay combined with MS for identifying potential KSPs from complex mixtures. Because this method does not require protein purification, it can be used in a high-throughput manner, and therefore we have applied it to the cell lysate of *Escherichia coli*. The analysis of the resulting database of putative KSPs in *E. coli* revealed preliminary insight about some of the structural and functional biases in favor of and against KS.

Results and Discussion

D2D SDS/PAGE as a High-Throughput Method for Identifying KS Proteins. Our previous study, which showed that KSPs are resistant to SDS, became the basis of a simple assay consisting of comparing the migration distance on a gel of two identical protein samples containing SDS, one boiled just before loading and the other left unheated (Fig. 1) (5). We expanded this SDS-resistance assay to a D2D SDS/PAGE assay for the high-throughput identification of KSPs from complex mixtures of proteins, such as cell lysates. This D2D SDS/PAGE method is very similar to previous ones used for the detection of protease susceptibility (8) and more recently to identify stable oligomeric protein complexes in the inner membrane of *E. coli* (9). In the first step of our KS assay, the unheated sample containing a mixture of proteins was analyzed in the first dimension by SDS/PAGE (Fig. 2*A*). The gel lane containing the protein was cut out, and the gel strip was then incubated in SDS/PAGE sample buffer and boiled for 10 min (Fig. 2*B*) before being placed above a larger gel for the second-dimension run (Fig. 2*C*). Most proteins were denatured by SDS even without heating and thus migrated the

Author contributions: K.X. and M.M. contributed equally to this work; K.X., M.M., and W.C. designed research; K.X., M.M., and Q.L. performed research; K.X., M.M., H.H., Q.L., C.B., and W.C. analyzed data; and K.X., M.M., Q.L., C.B., and W.C. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: D2D SDS/PAGE, diagonal 2D SDS/PAGE; GuHCl, guanidine hydrochloride; KS, kinetic stability; KSP, kinetically stable protein; 4°, quaternary; 2°, secondary; MS/MS, tandem MS.

¶To whom correspondence should be addressed. E-mail: colonw@rpi.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0705417104/DC1.

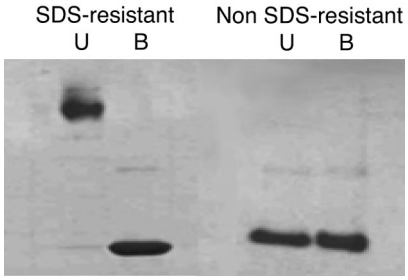

Fig. 1. A simple assay for KS involves comparing the migration of identical unheated (U) and boiled (B) protein samples. Boiled and unheated samples of streptavidin, a KSP, will migrate differently on SDS/PAGE, whereas for β2-microglobulin, which is not kinetically stable, they will have the same migration.

same distance in both dimensions of our analysis, resulting in a diagonal line of spots with a negative slope across the gel (Fig. 2*D*). However, SDS-resistant proteins traveled a shorter distance in the first-dimension gel and therefore after the second SDS/PAGE ended up migrating to a region below the gel diagonal, separated from the bulk proteins.

Identifying Putative KSPs in *E. coli* by D2D SDS/PAGE. To test whether the D2D SDS/PAGE method could detect KSPs from complex mixtures, we applied it to analyze the cell lysate of *E. coli*. The D2D SDS/PAGE assay exhibited the expected diagonal pattern resulting from the same migration in both dimensions of nonSDS-resistant (i.e., nonkinetically stable) proteins (Fig. 3). However, dozens of spots were seen below the gel diagonal, and these represent the putative KSPs present in the cell lysate of *E. coli*. To identify these proteins, each spot was cut out and subjected to trypsin digestion and proteomics analysis by using liquid chromatography tandem MS (LC-MS/MS). The resulting MS/MS data were searched against the *E. coli* protein database by using the algorithm Mascot 2.1 (10). As reasonable criteria for the accurate identification of proteins, we included only proteins that had at least two peptide hits with a P value of < 0.05, thereby resulting in the identification with high confidence of 50 nonredundant proteins (Table 1). *E. coli* expresses

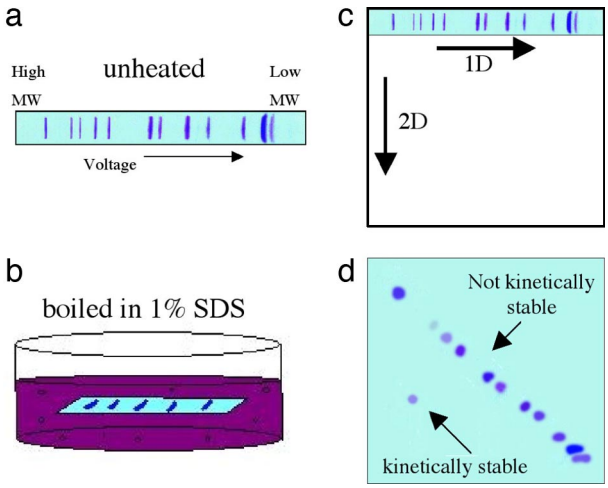

Fig. 2. Diagram of D2D SDS/PAGE method. (*a*) First-dimension SDS/PAGE separation is performed followed by excision of a gel strip containing the relevant lane. (*b*) The gel strip is incubated for ≈10 min in a boiling water bath of buffer containing 1% SDS. (*c*) The gel strip is placed above a larger gel, and a second-dimension SDS/PAGE separation is performed. (*d*) The gel is stained, revealing a diagonal band resulting from the equal migration of nonKSPs in both dimensions of the gel. Because KSPs migrate less in the first (unheated) dimension, they end up migrating to the left of the gel diagonal.

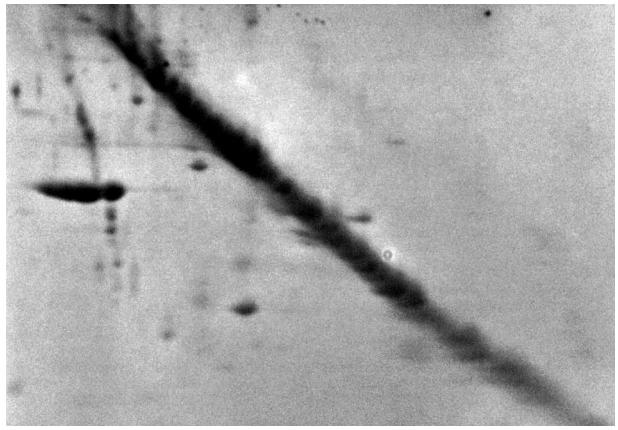

Fig. 3. Analysis of the cellular lysate of *E. coli* by D2D SDS/PAGE. The separation in both dimensions was performed as described in Fig. 2 and *Materials and Methods*. The visible spots to the left of the gel diagonal represent the soluble putative KSPs in *E. coli*.

≈884 water-soluble proteins that are visible on a traditional 2D gel (11), and therefore our results show that most *E. coli* proteins are not kinetically stable.

Fig. 3 shows a few unexpected bands and smearing above the gel diagonal. Because we do not use reducing agent (e.g., DTT) in our experiment, we hypothesized the proteins that migrate above the diagonal may result from disulfide bond formation during the heating step. To test this, we repeated our experiment and added DTT just before the heating step, and most of the bands and smearing disappeared [supporting information (SI) Fig. 7], except for one very consistent band near the middle of the diagonal. Analysis of this band by liquid chromatography MS/MS revealed that it corresponds to the outer-membrane protein OmpA, which has two cysteine residues and is know to be SDS-resistant (12). It appears that OmpA would have migrated well below the diagonal but may have formed a buried intermolecular disulfide bond that was not reduced by DTT. Thus, proteins that migrate above the diagonal may be caused by oligomerization via disulfide or other covalent bond formation during the heating step.

Are All SDS-Resistant *E. coli* Proteins Identified by D2D SDS/PAGE Kinetically Stable? Unlike chemical denaturation, SDS appears to denature proteins by irreversibly trapping them during the transient times in which proteins are unfolded (5), and because KSPs rarely escape their native state, they are virtually immune to SDS-induced denaturation. Since our initial study (5), we have analyzed dozens of other proteins and have not seen an exception to this observation. However, there may be other reasons independent of KS that may result in SDS resistance. For example, proteins that are highly negatively charged may repel SDS. The 50 SDS-resistant *E. coli* proteins we identified in this study have isoelectric points that range from 4 to 10, and therefore none is expected to electrostatically repel SDS. Also, proteins that are not KS in themselves but may be part of KS complexes could cause false-positives in our assay. A literature search of the proteins listed in Table 1 revealed several that form complexes with GroEL, including *S*-adenosylmethionine synthase, elongation factor Tu, RNA polymerase α-chain, and 50S ribosomal protein L7/L12 (13). Interestingly, the GroEL complexes have been shown to be SDS-resistant, whereas GroEL itself and some of its binding partners are known to lack SDS resistance (13). Thus, D2D SDS/PAGE may also be a useful method for identifying kinetically stable complexes arising from the interaction of proteins that may not in themselves be kinetically stable.

Although it would be desirable to confirm the KS of all of the *E. coli* proteins identified by the D2D SDS/PAGE assay, this is not

Table 1. Nonredundant subset of SDS-resistant proteins

GenBank identifier	Name	No. of residues	PDB*	2°	4°
15804817	Inorganic pyrophosphatase	176	1INO	α/β	6
15802070	Superoxide dismutase, iron(ii)	193	1ISA	α/β	2
9507572	Chloramphenicol acetyltransferase	219	1NOC	α/β	3
443293	Triosephosphate isomerase Tim	255	1TRE	α/β	2
16131680	Uridine phosphorylase	253	1K3F	α/β	6
14488510	Ompf porin	340	1HXX	β	3
51247607	Glycerophosphoryl diester phosphodiesterase	336	1T8Q	α/β	2
75196280	Periplasmic glycerophosphoryl diester phosphodiesterase	371	1YDY	α/β	2
11514297	Elongation factor, Tu	393	1D8T	$\beta+\alpha/\beta$	1
15804731	Aspartate ammonia-lyase (aspartase)	493	1JSW	α	4
6435772	Outer membrane protein Ompx	148	1QJ8	β	3
15803823	30S ribosomal protein S4	206	1VS5	α/β	21
16131215	Bacterioferritin	158	1BCF	α	24
112489962	Modulator of drug activity B (MdaB)	204	2AMJ	α/β	2
13786833	Pyridoxine 5′-phosphate synthase	242	1HO1	α/β	8
2914323	Enoyl reductase	261	1DFG	α/β	4
6730179	Reduced thioredoxin reductase	320	1CL0	α/β	2
26248038	Glyceraldehyde-3-phosphate dehydrogenase	334	1S7C	α/β	4
1421289	*S*-adenosylmethionine synthetase	383	1XRB	α/β	2
4557950	β-Ketoacyl-acp synthase Ii	412	2GFW	α/β	2
14278152	2-Amino-3-ketobutyrate CoA ligase	401	1FC4	α/β	2
15804373	Transcription termination factor Rho	419	1PVO	$\alpha+\beta$	12
1310928	Maltoporin Lamb	421	1MAL	β	3
9256952	Outer membrane protein TolC	428	1EK9	$\alpha+\beta$	3
15804852	Leucyl aminopeptidase	503	1GYT	α/β	6
91213467	Glycerol kinase	537	1GLA	β	8
42146	Unnamed protein†	733	1ZO1	α/β	2
146264	Xanthine guanine phosphoribosyltransferase	152	1NUL	α/β	4
30065622	Purine nucleoside phosphorylase	239	1ECP	α/β	6
223571	Protein L12	272	2AW4	α/β	31
75175990	Transaldolase	317	1UCW	α/β	1
42810	RNA polymerase α-subunit N-terminal domain	329	1BDF	$\alpha+\beta$	2
16132149	Isoaspartyl dipeptidase	390	1YBQ	α/β	8
38491472	GroEL	548	2EU1	α/β	7
75233972	L-fucose isomerase	591	1FUI	α/β	3
110643069	Fructose-bisphosphate aldolase class II	421	1ZEN	α/β	2
15800320	Alkyl hydroperoxide reductase, C22 subunit	187	1YEP	α/β	10
15804455	Glutamine synthetase	469	2GLS	α/β	12
15799927	Phosphoheptose isomerase	192	1X94	α/β	4
15804539	Catalase; hydroperoxidase HPI(I)	726	1MWV	α	2
15799800	Dihydrolipoamide dehydrogenase	474	1OJT	α/β	2
42377	Unnamed protein	549	1IAT	$\alpha+\beta$	2
15803853	Elongation factor EF-2	704	1DAR	$\beta+\alpha/\beta$	1
15799862	(3R)-hydroxymyristoyl ACP dehydratase	151	1U1Z	$\alpha+\beta$	6
75237743	Predicted GTPase	490			
148247	Proline dipeptidase	443			
15802566	Galactitol-1-phosphate dehydrogenase	346			
38704050	Fructose-bisphosphate aldolase	350			
15801691	Putative receptor	353			
91211384	Hypothetical protein UTI89_C2371	374			

*Nearest-homology Protein Data Bank (PDB) structures with high percent identity: From 1INO to 1GLA, 100%; from 1ZOE to 1ZEN, 99%; 1YEP, 98%; 2GLS, 97%; 1X94, 78%; 1MWV to 1IAT, 64%; 1DAR, 59%; and 1U1Z, 53%.
†The name of its 99% identity PDB structure is IF2, IF1, and tRNA fitted to cryo-EM data of *E. coli* 70S initiation complex.

feasible. However, we wanted to verify the SDS resistance–KS correlation by selecting a protein at random from our pool and analyzing its KS using fluorescence spectroscopy. We selected inorganic pyrophosphatase based on its commercial availability and its unknown KS. As shown in Fig. 4*A*, even at 6.0 M guanidine hydrochloride (GuHCl), inorganic pyrophosphatase took minutes to unfold. Plotting the unfolding rate constants of the protein against GuHCl concentration and extrapolating to 0 M yielded a native unfolding rate of 6.44×10^{-11}·sec^{-1}. Thus, we have demonstrated the high KS of inorganic pyrophosphatase, thereby attesting to the effectiveness of our D2D SDS/PAGE method for identifying KSPs in complex mixtures.

Functional Characteristics of KSPs in *E. coli*. To probe whether specific functions in *E. coli* may have a bias toward or against KS, we compared the functions of a nonredundant subset of the *E. coli*

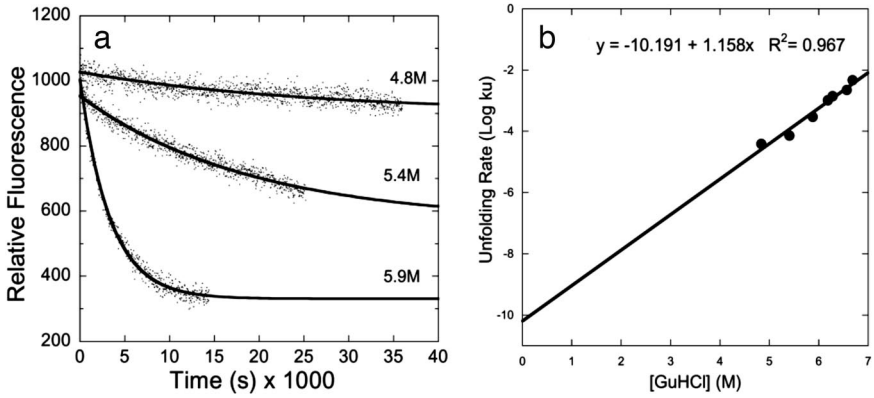

Fig. 4. Probing the KS of inorganic pyrophosphatase. (*a*) The unfolding kinetics of the protein was examined at 4.8–6.7 M GuHCl, and the rate constants were determined by fitting to a single exponential function. (*b*) The log of the rate constants was plotted against GuHCl and fitted to a linear function. Extrapolation of the plot to 0 M GuHCl yielded a native unfolding rate of 6.44×10^{-11}·sec^{-1} (i.e., half-life of 346 years).

proteome with the putative KSPs we identified (Table 1). The most notable difference is that, whereas ≈32% of all of the proteins in *E. coli* are enzymes, >70% of the KSPs are enzymes (Fig. 5*A*). To further extend this comparison, we determined the enzyme family distributions of the *E. coli* proteome (Fig. 5*B*). Within the six families of enzymes listed, there were no major differences, and in some cases, it is not possible to reach any reliable conclusion because the number of kinetically stable enzymes in our database is very low for some families (e.g., ligases). We speculate that oxidoreductases may have a real bias toward KS because of the common presence of metals and cofactors, and because they often involve the formation of potentially damaging free radicals. In the case of ligases, it is plausible that their function is incompatible with KS, because they may need to be tightly regulated. Ligases also interact with other proteins or nucleic acids, thereby requiring flexibility and the ability to undergo conformational changes (14). The virtual absence (Fig. 5*A*) of kinetically stable transporters and regulators is noteworthy but may be necessary for the efficient regulation of these proteins. In particular, the function of regulatory proteins, such as transcription factors, seems biologically incompatible with KS, because they often must rapidly be turned on and off.

Other nonenzyme proteins that were identified as kinetically stable were the periplasmic protein glycerophosphoryl diester phosphodiesterase and the outer membrane proteins OmpF, maltoporin Lamb, TolC, and OmpX. Interestingly, each periplasmic and outer-membrane protein in *E. coli* is synthesized in the cytoplasm as a precursor protein that contains an N-terminal signal sequence. During translocation, the signal sequence is proteolytically cleaved, generating the mature form of the protein, which because of its localization outside the cytoplasm may require KS.

These observations with a small database of proteins represent a step toward understanding how KS may be related to protein function. As the database of KSPs increases in the future, it will be interesting to determine how the function of the enzyme, its location within the cell, and other factors determine the biological role of KS.

Probing the Structural Features of KSPs in *E. coli*. Of the 50 putative KSPs identified (Table 1), 44 have known 3D structures or are linked to close homologs of known structures (15). This is a remarkably high percentage (88%) because, of 4,400 genes in *E. coli* (16), only 382 (≤10%) protein structures are found in the RSCB Protein Data Bank, after removing sequences that are ≥90% identical. This suggests that KSPs may be easier to purify and/or crystallize, perhaps because of their resistance to degradation and aggregation.

To identify any salient structural feature among the KSPs in *E. coli*, their secondary (2°) structures were compared with the *E. coli* proteome by using the classification obtained by the CATH database (17). As shown in Fig. 6*A*, there was a small difference in the percentage of β structures when compared with the bulk proteins in *E. coli* but a significant difference in the percentages of α and α/β proteins. The percentage of KSPs with all-α-helical structure was lower, suggesting a structural bias against KS. We hypothesize that monomeric predominantly helical proteins cannot easily attain the topological complexity that may be required for KS. In the case of α/β proteins, there seems to be a bias in favor of KS, perhaps because mixtures of 2° structure lead to more complex topologies.

Analysis of the quaternary (4°) structures of the KSPs in *E. coli* shows that the percentage of monomeric KSPs in *E. coli* is significantly lower compared with the whole proteome, whereas the percentage of oligomeric proteins with five or more subunits is significantly higher (Fig. 6*B*). Protection of surface residues from the solvent, increase in rigidity, and the stabilization of specific regions of a protein are potential effects of oligomerization that may increase KS. Thus, it is intriguing to consider the possibility that there may have been evolutionary pressure to favor certain oligomeric proteins on the basis of KS. Although oligomerization may be an important factor in the KS of some proteins, there are many oligomeric proteins that are not kinetically stable, and therefore nature has evolved other unknown factors to endow proteins with KS.

Implications of D2D SDS/PAGE for Investigating Protein KS. We have shown that the D2D SDS/PAGE method described here is simple, fast, inexpensive, and can be applied in high-throughput fashion for the proteomics-level detection of KSPs. Recently, Park *et al.* (18) described the application of a 4-day protease susceptibility screen toward the selection of proteins in *E. coli* that are resistant to proteolysis, and they suggested that KS is likely responsible for the proteins that remain intact. However, only six proteins overlapped in our study and theirs, and they also identified 34 other protease-resistant proteins. There are major differences between the two methods that may explain the different results. The SDS-induced denaturation of proteins involves its irreversible trapping via the interaction of the SDS micelle with hydrophobic residues exposed during the transient unfolding of the protein. This promiscuous interaction between SDS and the protein, aided by the high concentration of SDS (35 mM in our study), depends almost exclusively on the unfolding rate of the protein. In contrast, proteases must interact with the protein backbone, limited by its residue specificity, protease concentration, and intrinsic proteolysis rate. In the Park *et al.* study (18), the concentration and the intrinsic

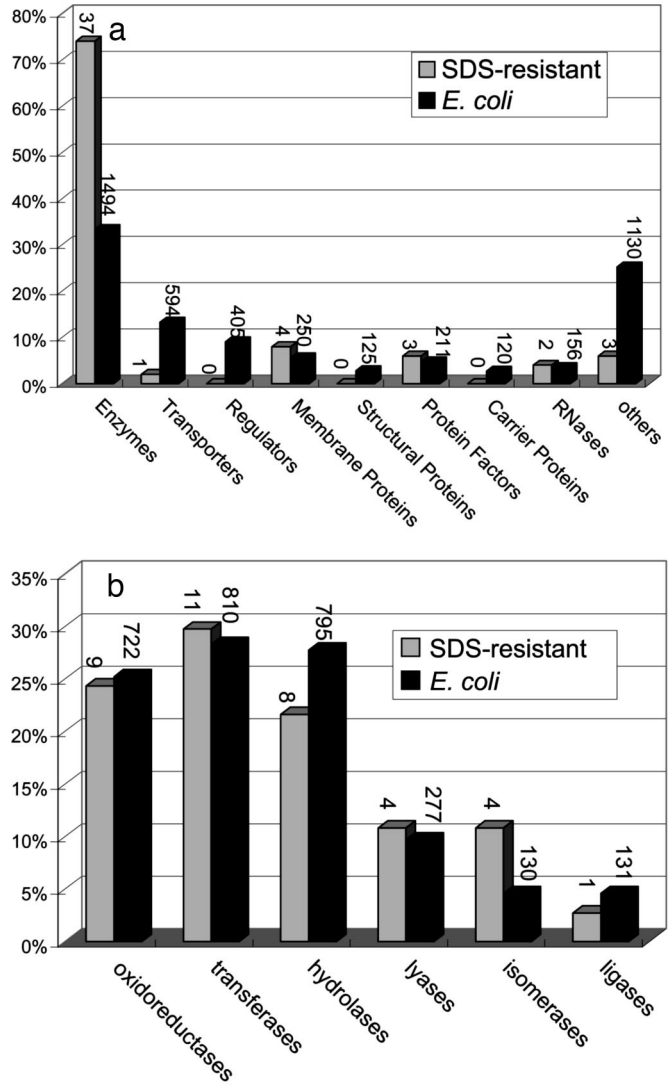

Fig. 5. Distribution of protein (*a*) and enzymatic (*b*) functions for a nonredundant subset of the *E. coli* proteome vs. our set of KSPs. (*a*) The kinetically stable subproteome has significantly more enzymes ($P < 0.0001$) but fewer transporters ($P = 0.0076$) and regulators ($P = 0.0082$). Other changes were not statistically significant at the 95% confidence level. Functional assignments were made by using the *E. coli* genome and proteome database (16). ''Other'' refers to the following functions: leader peptides, external origin, cell processes, lipoproteins, pseudogenes, phenotypes, unknown functions, unclassified proteins, and sites. (*b*) Distribution of the six most common enzyme functions does not show statistically significant differences at the 95% confidence level. Enzyme functions were obtained by using the BRENDA web site (21).

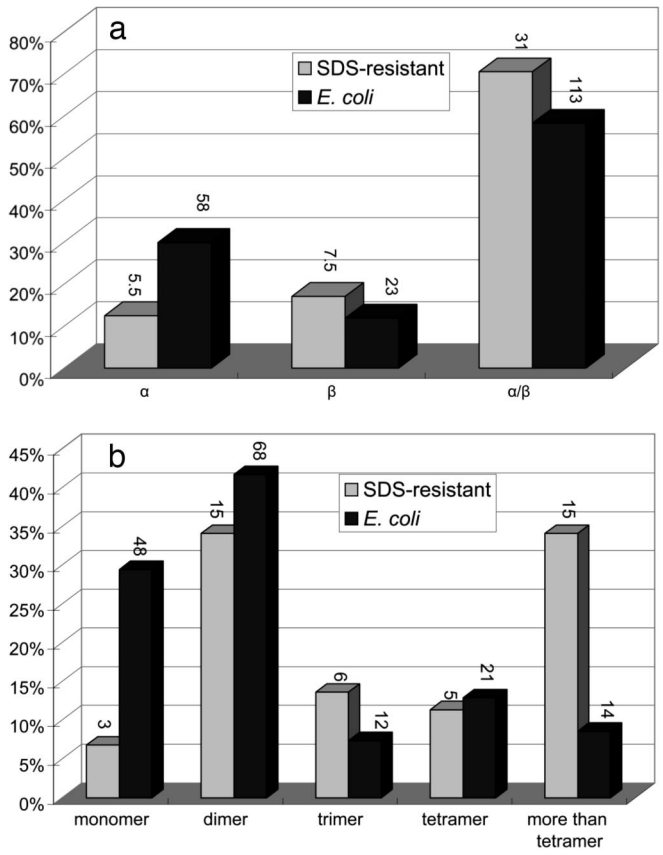

Fig. 6. 2° (*a*) and 4° (*b*) structure distribution of the nonredundant *E. coli* proteome and its subproteome of KSPs. (*a*) The KSPs have fewer ($P = 0.0034$) all α-helical proteins compared with the rest of the *E. coli* proteome. Structure classifications were made by using the CATH database (17). (*b*) The KSPs include significantly fewer monomers ($P = 0.0002$) and significantly more large oligomeric structures with at least five subunits ($P < 0.001$). Dimers and tetramers occur at approximately the same frequencies in the two sets. 4° structure information was obtained from the PQS Protein Quaternary Structure database (20).

proteolysis rate for thermolysin were 12 μM and 8.8 s^{-1}, respectively, and therefore, some thermodynamically stable proteins with a very fast folding rate may have escaped proteolysis even in the absence of high KS. Also, because proteases may cleave exposed loops and highly flexible regions within their native state or during their partial and global unfolding, some KSPs may have been cleaved over the 4-day incubation period without undergoing partial or global unfolding. Finally, different experimental conditions, such as incubation time (10 min vs. 4 days), pH (6.8 vs. 8.0), and electrophoresis method (D2D SDS/PAGE vs. conventional 2D electrophoresis) most likely have also contributed to the different proteins identified by each method. In summary, the D2D SDS/PAGE method seems more accurate for identifying proteins with high KS and is more convenient for high-throughput analysis. In contrast, the protease susceptibility method identifies proteins that are protease-resistant because of high thermodynamic and/or KS and can conveniently allow the quantitative analysis of protein stability over a wide range of energy landscapes.

In conclusion, the D2D SDS/PAGE method described here is simple, accessible, and can be applied in high-throughput fashion for the proteomics-level identification of KSPs. In addition to the analysis of cellular lysates from a variety of microorganisms, the D2D SDS/PAGE method may be used for the analysis of other proteomes, including those from plasma, cells, and organelles. Such studies will increase the database of KSPs and will facilitate investigation of the structural and functional basis of KS. Furthermore, they will stimulate research to understand the biological and pathological roles of the abnormal gain or loss of KS in proteins.

Materials and Methods

Preparation of Bacterial Cellular Lysates. *E. coli* cells (BL21 DE3 strain; Invitrogen, Carlsbad, CA) were grown overnight on LB/agar plates. Individual colonies were grown overnight in LB at 37°C to an OD of ≈0.9. The cell culture was centrifuged at 2,000 × *g* for 10 min, and the pellet was resuspended in 25 mM potassium phosphate buffer, pH 6.8. Sodium chloride was added to a final concentration of 450 mM, and the cells were lysed by using 10 min of sonication. The cell debris was removed by centrifugation, and the total protein

concentration in the cell lysate was estimated by using the Lowry Assay.

D2D SDS/PAGE. The *E. coli* cell lysate was diluted 5-fold and incubated for 5 min in SDS sample buffer (pH 6.8) to a final concentration of 45 mM Tris·HCl/1% SDS/10% glycerol/0.01% bromophenol blue). A 250-μl aliquot of the lysate solution was loaded without prior heating onto a well of a 12% acrylamide gel (16 cm × 14 cm × 3 mm). Electrophoresis was performed in a Protean II xi cell (Bio-Rad, Hercules, CA) by using 480 V and 120 mA. The gel was kept at 15°C by using a circulating water bath. Running buffer contained 25 mM Tris base, 0.2 M glycine, and 0.1% SDS. After the first-dimension run (Fig. 2*A*), the gel strip was cut out and incubated for 10 min in equilibration buffer (50 mM Tris·HCl/1% SDS/15% glycerol/0.02% bromophenol blue, pH 6.8) at ≈100°C (Fig. 2*B*). The gel strip was drained briefly and placed on top of a 12 cm × 14 cm × 3 mm 12% acrylamide gel (Fig. 2*C*). A small amount of 12% acrylamide solution was used to fuse the strip to the resolving gel. The second-dimension separation was performed under the same conditions as the first-dimension run (Fig. 2*D*). Gels were stained by using Coomassie (Bio-Rad Biosafe), and the protein spots below the diagonal protein band were picked by using a One Touch 2D gel spotpicker (1.5 mm).

MS and Protein Identification. The protein spots were washed, reduced, alkylated, and followed by in-gel tryptic digestion overnight. The peptide mixture was extracted, dried in a Speed-Vac, and dissolved in 10 μl of 5% formic acid. The liquid chromatography MS/MS experiments were performed on a Q-TOF 2 mass spectrometer (Waters, Milford, MA) equipped with the CapLC system. The stream select module was configured with a 180 μm ID × 50-mm trap column packed with 10 μm R2 resin (Applied Biosystems, Foster City, CA) connected in series with a 100 μm ID × 160-mm capillary column packed with 5 μm C18 particles. Peptide mixture (10 μl) were injected onto the trap column at 12 μl/min and desalted for 6 min before being flushed to the capillary column. The peptides were then eluted by the application of a series of mobile phase B gradients (5–10% B for 4 min, 10–30% B for 61 min, 30–85% B for 5 min, 85% B for 5 min). The final flow rate was 250 nl/min. Mobile phase A consisted of 0.1% formic acid/3% acetonitrile/0.01% TFA, whereas mobile phase B consisted of 0.075% formic acid/0.0075% TFA in 98/2 acetonitrile/water solution. The mass spectrometer was operated in a data-dependent acquisition mode while following the interrogation of MS data; ions were selected for MS/MS analysis based on their intensity and charge state + 2, +3, and + 4. The MS survey scan was from m/z 400–1,600 with an acquisition time of 1 sec, whereas the MS/MS fragmentation scan was from m/z 100–2,000 with an acquisition time of 2.4 sec. Mascot 2.1 (Matrix Science, London, U.K.) was used to search all of the MS/MS against the *E. coli* protein database obtained from National Center for Biotechnology Information nonredundant database with MS and MS/MS mass tolerance of 1.2 and 0.6 Da, respectively. PKL files were created by using Masslynx 3.5 from Waters. The parameters used for the searches were as follows: trypsin-specificity restriction with one missing cleavage site and variable modifications including oxidation (M), deamidation (NQ), and alkylation (C).

Unfolding Kinetics. The unfolding kinetics of inorganic pyrophosphatase (Sigma–Aldrich, St. Louis, MO) induced by GuHCl were monitored by using an F-4500 fluorescence spectrophotometer (Hitachi, Danbury, CT). A ≈1 mg/ml solution of the protein (25 mM potassium phosphate buffer, pH 7.0/0.1 M NaCl) was manually mixed with a GuHCl solution to obtain protein samples with a final denaturant concentration ranging from 4.8 to 6.7 M. Excitation/emission wavelengths for the time scans were 280/330 nm. Kinetic traces were analyzed by fitting to a first-order exponential function.

Data Analysis. Using Mascot 2.1, 85, proteins were identified that contained more than one unique peptide sequence. These 85 corresponded to a nonredundant set of 50 sequences with no more than 90% sequence identity, identified by using a BLAST search of the UniRef90 database (www.ebi.ac.uk/blastall) (19). Of the 50 KSP sequences identified, 26 were found to have known structures, 12 were close homologs (% identity > 90), and an additional six were distant homologs of high confidence (> 50% identity, e-value < 1E-40) to proteins of known structure (www.ncbi.nlm.nih.gov/blast) (15). For the statistical analysis presented here, we define the *E. coli* proteome as the nonredundant set of all proteins in the *E. coli* genome with at most 25% identity with any other member of the set. This method avoids overcounting paralogous protein families within the genome. No attempt was made to correct for different expression levels. Proteins that could not be assigned a structure, including natively unstructured proteins, were ignored in the structural analyses.

Protein 2° structure classes were assigned based on CATH classification, version 3.0 (www.cathdb.info) (17). 4° structures were assigned by using the EMBL-EBI PQS database (http://pqs.ebi.ac.uk) (20). Protein function assignments and enzyme function assignments were based on the *E. coli* genome and proteome database (http://genprotec.mbl.edu) (16) and the BRENDA enzyme database release 2007.1 (www.brenda-enzymes.info), respectively (21).

The significance of any differences in the distributions of function, enzyme function, and 2° and 4° structures within the two sets of proteins was calculated by using the resampling method (22) with 10,000 repetitions. A significance value of $P < 0.001$ means that the observed deviation between the probability observed in the *E. coli* proteome and the probability observed in the KSP set was found in random samples drawn from the *E. coli* proteome probability distribution fewer than one time in 1,000.

This work was supported by the National Science Foundation [Grants NSF MCB-9984913 (to W.C.) and MCB-0519507 (to W.C. and C.B.)]. M.M. was supported by the National Institutes of Health (Training Grant 5T32GM067545-02). Proteomics analysis work was supported by National Center for Research Resources Shared Instrumentation Grant 1S10RR01782-01A1 (to Q.L.)

1. Cunningham EL, Jaswal SS, Sohl JL, Agard DA (1999) *Proc Natl Acad Sci USA* 96:11008–11014.
2. Plaza del Pino IM, Ibarra-Molero B, Sanchez-Ruiz JM (2000) *Proteins* 40:58–70.
3. Sohl JL, Jaswal SS, Agard DA (1998) *Nature* 395:817–819.
4. Kelly JW, Colón W, Lai Z, Lashuel HA, McCulloch J, McCutchen SL, Miroy GJ, Peterson SA (1997) *Adv Protein Chem* 50:161–181.
5. Manning M, Colón W (2004) *Biochemistry* 43:11248–11254.
6. Rodriguez-Larrea D, Minning S, Borchert TV, Sanchez-Ruiz JM (2006) *J Mol Biol* 360:715–724.
7. Kelch BA, Eagen KP, Erciyas FP, Humphris EL, Thomason AR, Mitsuiki S, Agard DA (2007) *J Mol Biol* 368:870–883.
8. Nestler HP, Doseff A (1997) *Anal Biochem* 251:122–125.
9. Spelbrink RE, Kolkman A, Slijper M, Killian JA, de Kruijff B (2005) *J Biol Chem* 280:28742–28748.
10. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) *Electrophoresis* 20:3551–3567.
11. Sigdel TK, Cilliers R, Gursahaney PR, Crowder MW (2004) *J Biomol Tech* 15:199–207.
12. Pautsch A, Vogt J, Model K, Siebold C, Schulz GE (1999) *Proteins* 34:167–172.
13. Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU (1999) *Nature* 402:147–154.
14. Verdecia MA, Joazeiro CA, Wells NJ, Ferrer JL, Bowman ME, Hunter T, Noel JP (2003) *Mol Cell* 11:249–259.
15. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) *Nucleic Acids Res* 25:3389–3402.
16. Serres MH, Riley MH (2000) *Microb Comp Genom* 5:205–222.
17. Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM (1997) *Structure (Cambridge, UK)* 5:1093–1108.
18. Park C, Zhou S, Gilmore J, Marqusee S (2007) *J Mol Biol* 368:1426–1437.
19. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, *et al.* (2006) *Nucleic Acids Res* 34:D187–D191.
20. Henrick K, Thornton JM (1998) *Trends Biochem Sci* 23:358–361.
21. Schomburg I, Chang A, Ebeling C, Gremse M, Heldt C, Huhn G, Schomburg D (2004) *Nucleic Acids Res* 32:D431–D433.
22. Good P (2006) *Resampling Methods* (Birkhauser, Basel, Switzerland), 3rd Ed.