Metabolomics Approach for Enzyme Discovery

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · SEPTEMBER 2006

Impact Factor: 4.25 · DOI: 10.1021/pr0600576 · Source: PubMed

8 AUTHORS, INCLUDING:

Natsumi Saito
Keio University
16 PUBLICATIONS **396** CITATIONS
SEE PROFILE

Martin Robert
Tohoku University
28 PUBLICATIONS **1,736** CITATIONS
SEE PROFILE

Takaaki Nishioka
Nara Institute of Science and Technology
173 PUBLICATIONS **7,675** CITATIONS
SEE PROFILE

Masaru Tomita
Keio University
533 PUBLICATIONS **17,323** CITATIONS
SEE PROFILE

research articles Journal of proteome research

Metabolomics Approach for Enzyme Discovery

Natsumi Saito,[†] Martin Robert,[†] Sayaka Kitamura,[†] Richard Baran,[†] Tomoyoshi Soga,[*,†,‡] Hirotada Mori,[†,§] Takaaki Nishioka,[†] and Masaru Tomita[†,‡]

Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, 997-0017 Japan, Human Metabolome Technologies Inc. Tsuruoka, Yamagata, 997-0052 Japan, and Nara Institute of Science and Technology, Ikoma, Nara, 630-0192 Japan

The search for novel enzymes is an important but difficult task in functional genomics. Here, we present a systematic method based on in vitro assays in combination with metabolite profiling to discover novel enzymatic activities. A complex mixture of metabolites is incubated with purified candidate proteins and the reaction mixture is subsequently profiled by capillary electrophoresis electrospray ionization mass spectrometry (CE−MS). Specific changes in the metabolite composition can directly suggest the presence of an enzymatic activity while subsequent identification of the compounds whose level changed specifically can pinpoint the actual substrate(s) and product(s) of the reaction. We first evaluated the method using several *Escherichia coli* metabolic enzymes and then applied it to the functional screening of uncharacterized proteins. In this manner, YbhA and YbiV proteins were found to display both phosphotransferase and phosphatase activity toward different sugars/sugar phosphates. Our approach should be broadly applicable and useful for enzyme discovery in any system.

Keywords: metabolomics • enzyme • functional genomics • screening • capillary electrophoresis • mass spectrometry

Introduction

In the post-genomic era, systematic elucidation of gene and protein function is an indispensable task. Even in the model microorganism *Escherichia coli*, nearly half of the protein-coding open reading frames (ORFs) have no experimentally confirmed function.[1,2] Enzymes constitute an important class of proteins, playing fundamental roles in all cellular systems and metabolic processes and they are expected to represent a significant fraction of these uncharacterized ORFs. Determining the novel activities and substrate specificity of enzymes are interesting not only to better understand the cell's metabolic systems but also to reveal attractive drug targets or as resources to develop new biocatalysts for industrial applications.[3] Thus, the development of high-throughput and generic methods to screen for novel enzyme activities is highly desirable.

Traditionally, enzymes have been identified by assaying cells or tissue extracts for a particular biological activity and then purifying the active protein to homogeneity. This method is a canonical approach in enzymology, but it is a tedious and time-consuming process that cannot be easily accomplished on a proteome scale. However, with the availability of genome sequence information and experimental resources such as expression clone libraries for all ORFs, it is tempting us to use the reverse approach, using purified recombinant proteins and trying to associate them to specific activities. Moreover, to discover completely novel activities in an unbiased manner it is desirable to develop a system to monitor any type of reaction since there may be no prior information available concerning either the substrate(s) or the product(s).

Recently, different functional proteomics methods focusing on enzymes, using activity-based screening tools such as protein chips[4] and chemical probes[5] have been reported. In addition, class-specific enzymatic screens using in vitro assays, as reported by Proudfoot et al., allowed the identification of three nucleotidases among unknown proteins of *E. coli*.[6] Although these techniques are powerful for high-throughput classification of enzymes in complex proteome, they require specific reagents and tend to target specific classes of enzymes thus retaining an inherent bias.

The potential of mass spectrometry for the characterization of enzymatic activities has previously been demonstrated.[7−11] The method provides a generic read-out that can potentially be used to monitor any reaction in which a change in mass occurs. These studies revealed that mass spectrometry is suitable and effective for determining substrate specificity and enzyme kinetic parameters. In addition, metabolite profiling of cells in wild-type and enzyme-inactivated mutants by LC−MS has been successfully applied to assign endogenous substrates to specific enzymes in vivo.[12]

Capillary electrophoresis electrospray ionization mass spectrometry (CE−MS) has emerged as a powerful tool for the global detection of charged chemical species,[13] peptides,[14,15] and for quantitative metabolome analysis.[16−19] Recently, we intro-

* To whom correspondence should be addressed. Phone: +81-235-29-0528. Fax: +81-235-29-0530. E-mail: soga@sfc.keio.ac.jp.

† Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata, 997-0017 Japan.

‡ Human Metabolome Technologies Inc. Tsuruoka, Yamagata, 997-0052 Japan.

§ Nara Institute of Science and Technology, Ikoma, Nara, 630-0192 Japan.

duced the application of CE−MS to determine the dynamic biochemical properties of in vitro reconstituted glycolysis.[20]

Here, we describe the development of a direct, generic, and nontargeted approach to screen for enzymatic activities based on in vitro assays and metabolite profiling by CE−MS. We first show that the assays allow us to detect different types of enzymatic activities and identify the substrate(s) and product(s) of the enzyme reaction in a complex mixture of metabolites. In addition, we demonstrate the method's potential for the functional screening of uncharacterized proteins and report the identification of phosphatase/phosphotransferase activities in two uncharacterized *E. coli* proteins, YbhA and YbiV.

Experimental Section

Chemicals. *S*-(5′-adenosyl)-L-methionine (SAM), thiamine pyrophosphate (TPP), flavin mononucleotide (FMN), pyridoxal 5′-phosphate, glycerol 3-phosphate and 3-aminopyrrolidine were purchased from Sigma-Aldrich (St. Louis, MO); 2-morpholinoethanesulfonic acid (MES) and 3-morpholinopropanesulfonic acid (MOPS) from Dojindo (Kumamoto, Japan); methionine sulfone from Avocado (Lancashire, UK). Bacto tryptone, Bacto yeast extract, Bacto malt extract and Nutrient Broth from Becton, Dickinson and Company (Franklin Lakes, NJ), and isopropyl-β-D-thiogalactopyranoside (IPTG) from Promega (Madison, WI). All other reagents obtained from Wako Pure Chemicals Ind., Ltd. (Osaka, Japan).

Expression and Purification of Recombinant Proteins. We selected a few target proteins from the complete set of cloned genes from *E. coli* K-12 W3110[21] (ASKA library, GenoBase, http://ecoli.aist-nara.ac.jp/) for protein expression. Each of the full-length ORFs are cloned in an archive expression vector pCA24N (GenBank; AB052891) containing a 6 × His-tag at the amino-terminal of the ORF.[21] Expression of recombinant proteins were induced in *E. coli* AG1 cell (Stratagene, La Jolla, CA). Cells carrying the expression plasmid were inoculated in LB medium (1% Bacto tryptone, 0.5% Bacto yeast extract, 0.5% NaCl) and incubated at 37 °C overnight. An aliquot of the culture (180 μL) was inoculated into 6 mL of fresh LB medium and the incubation continued until the optical density at 600 nm (OD_{600}) reached 0.6. IPTG (0.5 mM) was added and the culture was incubated at 25 °C for 8−12 h with shaking. Cells were collected by centrifugation and stored at −30 °C until further use. The recombinant proteins were purified using cobalt-based immobilized TALON metal affinity chromatography resins and gravity-flow column (BD Biosciences Clontech., Palo Alto, CA) according to the manufacturer's instructions. Finally, the proteins were eluted from the column using phosphate buffer (pH 7.0) containing 200 mM imidazole. The protein solution was ultrafiltrated with a 10 000 nominal molecular weight limit (NMWL) filter (Millipore, Billerica, MA) to exchange the buffer to the 20 mM MOPS/KOH buffer (pH 7.2) containing 10 mM $MgCl_2$ and KCl, and concentrate the proteins. Glycerol was added to a final concentration of 50% and samples stored at −30 °C until use. For further purification and molecular weight estimations, purified his-tagged proteins were separated by gel filtration on a Superdex 200 HR 10/30 column (GE Healthcare Bio-Sciences Corp., 800 Centennial Ave, Piscataway, NJ) equilibrated with 50 mM MOPS buffer (pH 7.2) containing 10 mM $MgCl_2$ and KCl using an ÄKTA FPLC system (GE Healthcare Bio-Sciences Corp.). Fractions were collected at flow rate of 0.4 mL/min. Calibration was performed using Bio-Rad's gel filtration molecular weight standards for size exclusion chromatography columns. The recombinant protein concentration was determined using a protein assay kit (Bio-Rad, Hercules, CA) using bovine serum albumin as a standard and purity was estimated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE).

Preparation of Metabolite Pool. As potential sources of metabolites, the Bacto yeast extract, Bacto malt extract, and Nutrient Broth were tested. Solutions of 0.4% (w/v) of the above were prepared in Milli-Q water containing 200 μM methionine sulfone, 200 μM 3-aminopyrrolidine, 200 μM MES and 200 μM trimesic acid as internal standards. The solution was centrifugally filtrated through a Millipore ultrafiltration membrane with a 5000 NMWL to remove macromolecules, large peptides and cell debris. Aliquots of filtrate were lyophilized and redissolved in 100 μL of reaction buffer before use.

In Vitro Enzyme Assays. Reactions were performed in 20 mM MOPS/KOH buffer (pH 7.2), 10 mM $MgCl_2$, 10 mM KCl, containing the metabolite pool from yeast extract and a 0.1 mM supplement mixture of general enzyme cofactors as follows; β-nicotinamide adenine dinucleotide oxidized/reduced form (NAD^+/NADH), nicotinamide adenine dinucleotide phosphate oxidized/reduced form ($NADP^+$/NADPH), TPP, pyridoxal 5′-phosphate, biotin, SAM, coenzyme A (CoA), FMN, flavin adenine dinucleotide (FAD), acetyl-CoA, adenosine 5′-triphosphate (ATP), guanosine 5′-triphosphate (GTP). Following the addition of 1 μg of purified protein to 100 μL of the above assay solution, the mixture was incubated for 30 min at 37 °C, and then filtrated through a Millipore ultrafiltration membrane with 5000 NMWL to remove proteins and stop the reaction. The ultrafiltrate was directly analyzed by CE−MS in both positive and negative modes for the cationic and anionic metabolites, respectively. Phosphotransferase and phosphatase activity toward several compounds was determined by measuring their change in abundance using CE−MS. Glycerol in the protein stock solutions was removed by ultrafiltration prior to assay phosphotransferase activity. The mixture for the phosphotransferase assay contained 100 mM glycerol, 1 mM phosphate as substrates and 100 μM of MES as an internal standard in 20 mM MOPS/KOH buffer (pH 7.2), 10 mM $MgCl_2$, and 10 mM KCl. The reaction mixture for the phosphatase assay contained 1 mM of glycerol-3-phosphate, glucose-6-phosphate and ribose-5-phosphate as substrates, and 100 μM of MES in the same reaction buffer. Reactions were initiated by addition of protein (1 μg) to 100−200 μL of solution and carried out at 37 °C. Samples were taken at different time points over a 120 min time-course and reactions were stopped with 30 mM sodium hydroxide. The concentration of reactants in the reaction mixture was measured by CE−MS. Specific activities were determined as the rate of change in substrates or product concentration (μmol/min/mg).

Instrumentation and Analytical Procedures. All metabolome analyses were performed using CE−MS. Most CE−MS experiments were performed using an Agilent CE Capillary Electrophoresis System equipped with an Agilent 1100 series MSD (single quadrupole) mass spectrometer and an Agilent 1100 series isocratic HPLC pump, a G1603A Agilent CE-MS adapter kit and G1607A Agilent CE−MS sprayer kit (Agilent Technologies, Waldbronn, Germany). System control, data acquisition and MSD data evaluation were performed using the G2201AA Agilent ChemStation software for CE−MSD. All CE−MS conditions for ionic compounds were as previously described.[13,16−18] Data analysis was performed by selected ion monitoring (SIM) mode for every 1 m/z interval (70, 71, 72, ..., 430) and peaks areas were then integrated using Agilent

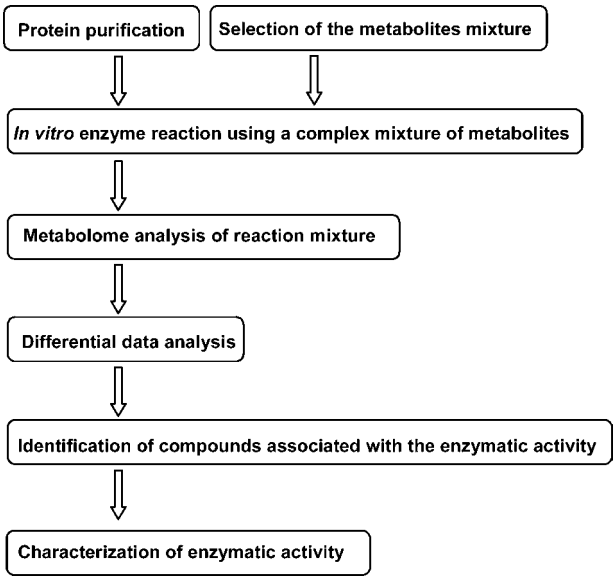

Figure 1. Basic strategy for discovering enzymatic activities using generic assays and metabolite profiling.

ChemStation software. Signal intensity normalization between samples was performed using the relative peak area, defined as the peak area divided by that of the internal standard. Metabolites in samples were quantified using the calibration curves made by analyzing standards. Under the analytical conditions used, the detection limits for compounds are in the range of 0.3 to 11 μM at a signal-to noise ratio of 3. Previously reported calibration curves using the same analytical setup showed good signal linearity with correlation coefficients between 0.991 and 0.999 at 10, 20, 50, and 100 μM.[13,16,18] In addition, the calibration curves for amino acids are linear between 10 and 500 μM with correlation coefficients between 0.996 and 0.999.[13] The compounds detected in SIM were identified by matching both their molecular weight and relative migration times (with respect to internal standards) with that of metabolite standards. Data analysis was also performed using a differential visualization tool that can highlight differences in the metabolite composition of two or more complex samples.[22] For chemical compound identification, capillary electrophoresis-time-of-flight mass spectrometry (CE-TOFMS) was also performed using Agilent 1100 series MSD TOF mass spectrometer combined with Agilent CE Capillary Electrophoresis System. The MSD-TOF detector was controlled with Agilent TOF software, whereas data acquisition and data analysis were performed with the Analyst QS (Applied Biosystems, Foster City, CA). Data acquisition was performed in the range of m/z 50−1000 with a scan rate of 0.15 cycles/s. The CE conditions in CE-TOFMS were the same as for CE−MS quadrupole system.

Results and Discussion

Strategy to Screen for Enzymatic Activity. A schematic representation of the activity screening and analytical workflow for the discovery of enzymatic function is summarized in Figure 1. Recombinant proteins are first purified as described in Experimental Section and used for in vitro reaction using a complex metabolite mixture. After the reaction, the mixture is ultrafiltrated to remove proteins and then comprehensively analyzed by CE−MS. By comparing the metabolite profiles of the reaction mixture incubated in the presence or absence of candidate protein, the presence of enzymatic activity can be monitored. A compound whose level specifically decreased following incubation with a protein can be considered a likely substrate for the reaction. Accordingly, a compound whose level increased during the incubation is likely product of the reaction. Subsequent identification of these compounds can directly pinpoint the enzymatic activity. The procedure is straightforward and should be effective for discovering and identifying enzymes—whose substrate(s) and/or product(s) bear an electric charge under the analytical conditions—in uncharacterized proteomes without any prior information about the type of catalytic activity.

Choice of a Suitable Metabolite Source. The choice of a suitable metabolite source is important to optimize the chances of discovering new enzymatic functions. Obviously, a complex metabolite soup is desirable. Intracellular metabolites are expected to include a large variety of compounds generated by biological processes or incorporated from the environments. The advantages of using natural and complex metabolite mixtures as potential substrate sources are as follows: (i) ease of preparation and low cost, yet a large variety of compounds, (ii) useful for the screening of unknown type of reactions, (iii) provide an environment closer to physiological conditions than standard in vitro assay conditions. While extracts from living cells might represent a suitable source, large scale availability and reproducibility of the metabolite pool are often difficult to realize. For these reasons, we tested commercially available nutrient mixtures, commonly used for the cultivation of microorganism, as potential rich sources of metabolites.

The metabolite content of yeast extract (YE), malt extract (ME), and nutrient broth (NB) were analyzed for both cationic and anionic compounds ranging from 70 to 430 in m/z values using CE−MS. To maximize sensitivity, MS analysis using a single quadrupole detector was performed in SIM mode by successively monitoring adjacent windows of 30 individual m/z values to cover the range of 70 to 430 m/z. Since excessive ionic strength and viscosity of samples can result in poor separation and unstable migration times of compounds in CE, the concentration of the solutions was limited to 0.4% (w/v). We first attempted to partially characterize the mixtures by comparing the amounts of cationic and anionic compounds that were identified between YE, NB, and ME. As shown in Figure 2A, YE contained larger amounts (2 to 5-fold more) of amino acids than NB while most amino acids were undetectable in ME. Cysteine is known to be easily oxidized,[13] and this may explain failure to detect it in the metabolite mixture. The contents of anionic compounds in the various extracts were similarly distributed between the three extracts, although the actual concentrations varied and were much lower in NB and ME (Figure 2B). All compounds that could be readily identified in YE, NB, and ME are listed in Tables 1 and 2. Although the composition and concentrations of metabolites were different between the different extracts, about 50 compounds could be identified in both YE and NB. A total of up to 530 (330 cations and 200 anions) compounds were detected in YE, compared to 340 (240 cations and 100 anions) compounds in NB. The analysis of ME resulted in the detection of less than 100 ionic compounds. These tables also include all yet unidentified peaks, some of them possibly representing metal adducts or other derivatives which may be formed in ESI−MS. Because of its high content in amino acids, bases, nucleosides and other stable metabolites, YE was selected as the most appropriate

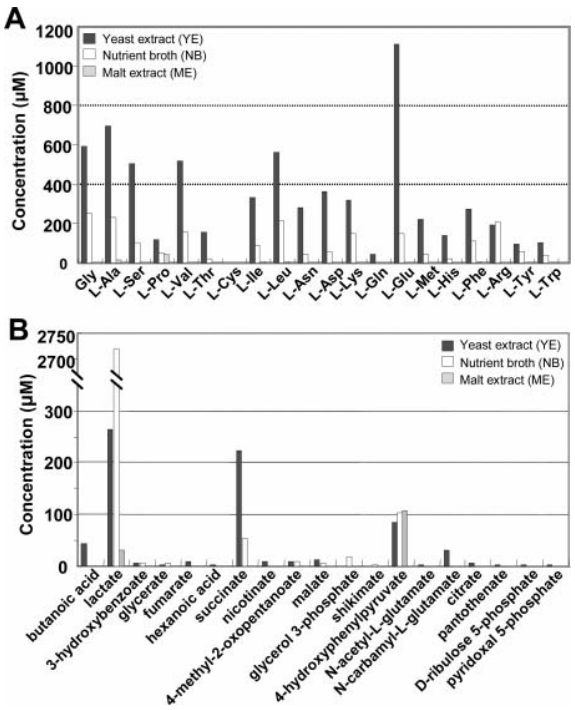

Figure 2. Levels of various metabolite profiling in yeast extract, nutrient broth and malt extract. All compounds were determined by CE−MS (quadrupole). Amino acids (A) and anionic metabolites (B). The concentrations reported correspond to a 0.4% (w/v) solution of each extract.

source of metabolites for the screening assays. However, we realized that the apparent absence of common nucleotide cofactors and other unstable but common metabolites might negatively affect the chances of detecting several important types of enzymatic reactions and thus elected to supplement the YE metabolite pool with a mixture of some of the most common cofactors used by multiple metabolic enzymes (see Experimental Section).

Method Validation: Detecting Enzymatic Activities. To evaluate the utility of our method to detect various types of enzymatic activities, we tested whether the activities of well characterized enzymes could be easily monitored using our generic assay system. We selected five metabolic enzymes from *E. coli* K-12 W3110 that represent four different enzyme classes as listed in Table 3 (decarboxylase, deaminase, hydrolase, and transferase). Reactions were initiated by adding the purified enzyme to the metabolite mixture. A control reaction was performed without addition of protein. After the reaction, both mixtures were analyzed by CE−MS to identify changes in their composition that might be associated with the enzymatic activity. CE−MS analyses were performed by monitoring the m/z range from 70 to 430 in both positive and negative modes, followed by differential analysis of metabolites in the control and enzyme reaction mixtures. Figure 3 shows the total ion chromatograms (TICs) for cations and the results of the visualization of differences in the metabolite profiles for the 251−280 m/z value range of the reaction mixture incubated with or without adenosine deaminase (EC 3.5.4.4). For this particular enzyme, the impact of the reaction on the mixture composition was readily apparent even in the total ion chromatogram (TIC), whereas the resulting product was difficult to find in TIC (Figure 3A). To achieve rapid, systematic, and sensitive analysis of the complex CE−MS data sets, we used a differential visualization tool developed in-house.[22] Differences between the two samples (control and enzyme reaction) can

Table 1. Quantification of Cationic Metabolites in Different Metabolite Sources

		μM[a]		
compound	m/z	YE	NB	ME
methylguanine	74		7	
glycine	76	595	249	
L-alanine	90	696	228	15
cadaverine	103		8	
4-aminobutanoate	104	52		8
2,3-diaminopropionate	105		61	
L-serine	106	500	107	
cytosine	112	0.6		1
creatinine	114		194	
5,6-dihydrouracil	115	47	72	
L-proline	116	120	49	41
L-valine	118	515	154	
L-homoserine	120	8	44	
L-threonine	120	153	21	
purine	121	0.7	16	
nicotinamide	123	16		
5-methylcytosine	126	2		2
N-acetylputrescine	131	2		
creatine	132		107	2
L-isoleucine	132	332	91	5
L-leucine	132	562	212	8
4-hydroxy-L-proline	132		4	
L-ornithine	133	49	20	
L-asparagine	133	281	43	
L-aspartate	134	361	57	
adenine	136	18	2	5
hypoxanthine	137	5	28	
tyramine	138	3	16	
anthranilate	138	10		
spermidine	146	8	6	
4-guanidinobutanoate	146	2	32	1
L-lysine	147	322	146	
L-glutamine	147	46		
L-glutamate	148	1111	145	
L-methionine	150	220	44	
L-histidine	156	142	24	
tryptamine	161		1	
L-carnitine	162	0.7	73	0.2
methylguanine	166	2		3
L-phenylalanine	166	274	113	5
pyridoxamine	169	1		
pyridoxine	170	0.7		
L-arginine	175	192	204	3
N-acetylornithine	175	8	5	
L-citrulline	176	42	17	
L-tyrosine	182	96	61	
spermine	203		0.8	
O-acetyl-L-carnitine	204	1	1	
L-tryptophane	205	105	38	
β-alanyl-L-lysine	218	22		
6-benzylaminopurine	226	212	95	
carnosine	227	6	198	
thymidine	243	48	46	
cytidine	244	6	49	
D-octopine	247	11	14	
deoxyadenosine	252			1
thiamin	265	41		
adenosine	268	38	12	
inosine	269		65	
Glu-Glu	277	25		
guanosine	284		26	
5′-methylthioadenosine	298	30		

[a] All metabolites were determined in 0.4% (w/v) yeast extract (YE), nutrient broth (NB) and malt extract (ME) prepared as described in Experimental section. Missing values mean that the compounds were undetectable under the conditions used.

Table 2. Quantification of Anionic Metabolites in Different Metabolite Sources

compound	m/z	μM[a]		
		YE	NB	ME
butanoic acid	87	45		
lactate	89	265	2712	30
3-hydroxybenzoate	103	6	7	
glycerate	105	3	8	
fumarate	115	9		
hexanoic acid	115	3		
succinate	117	223	54	
nicotinate	122	9		
4-methyl-2-oxopentanoate	129	9	9	0.8
malate	133	14	7	
glycerol 3-phosphate	171		18	
shikimate	173		3	
4-hydroxyphenylpyruvate	179	86	103	108
N-acetyl-L-glutamate	188	3		
N-carbamyl-L-glutamate	189	33		
citrate	191	7		
pantothenate	218	4	0.6	
D-ribulose 5-phosphate	229	5		
pyridoxal 5-phosphate	246	3		

[a] All metabolites were determined in 0.4% (w/v) yeast extract (YE), nutrient broth (NB) and malt extract (ME) prepared as described in Experimental section. Missing values mean that the compounds were undetectable under the conditions used.

be clearly visualized using two-dimensional representation of the CE−MS data (migration time and m/z) and directly highlight the compound peaks while amplifying the signal from the more subtle signal changes in the TICs (Figure 3B and Supporting Information Figure S1). Figure 3B shows a clear decrease in the area of the adenosine peak (m/z 268) in the YE and a concomitant increase in area of a peak corresponding to inosine (m/z 269), the product of the reaction. The whole data set can be found in Supporting Information Figure S1. The above changes were the only significant ones observed in this reaction mixture. These results demonstrate that a metabolomics-based screening in conjunction with careful data analysis can allow the detection of specific reactions in complex mixtures.

To verify whether the method could be broadly applicable to different types of reactions, we tested four other enzymes belonging to different categories of reactions (Table 3). Figure 4 shows the selected ion electropherograms for specific m/z values corresponding to the predominant changes in those mixtures as revealed by the differential visualization tool. Each electropherogram clearly shows that the levels of the known substrates of the reactions decreased while those of the corresponding products increased after reaction with the mixture of metabolites. No significant differences were detected in other peaks over the whole monitoring range (70−430 m/z) except for peaks representing sodium or other metal adducts, ions derived from ammonium elimination during ionization, or peaks corresponding to isotopes of the substrates and products (Supporting Information Figure S2). Figure S2 shows the complete data set resulting from the differential analysis of metabolite profiles obtained for the lysine decarboxylase reaction (EC. 4.1.1.18) and the corresponding mass spectra of the L-lysine and cadaverine electrophoretic peaks, acquired in the positive ion mode by scanning from m/z 70 to 300. Although sodium adduct formation and ammonium elimination are apparent in these spectra of L-lysine (substrate) and cadaverine (product), and may complicate data analysis/compound identification, careful examination of the data can allow the identification of the reactant ions. As far as we could observe they do not compromise the detection of activity.

Since both adenosine deaminase (EC 3.5.4.4) and ribonucleoside hydrolase (EC 3.2.2.1) utilize adenosine as a substrate (m/z 268), their activities cannot be distinguished simply on the basis of decrease in substrate peak intensity alone. However, we readily differentiated the activities by simultaneously observing the increase of signals in peaks corresponding to products of the reactions, inosine for adenosine deaminase (m/z 269) and adenine for ribonucleoside hydrolase (m/z 136). Although a significant decrease in the level of L-methionine (m/z 150), the substrate of methionine adenosyltransferase (EC 2.5.1.6) was not apparent, the corresponding increase in the products, SAM and inorganic phosphate were clearly observed. Failure to detect a decrease in L-methionine is likely related to the difference of detector sensitivity for L-methionine and products in MS. We observed that the relative peak areas of L-methionine and SAM (relative to the internal standard, methionine sulfone) were 0.08 and 0.35 at 10 μM suggesting a marked difference in sensitivity that made SAM level changes much easier to detect. Moreover accurate quantitation of L-methionine can be complicated due to its tendency to be easily oxidized.[13] This particular example demonstrates that even if changes in some compounds participating in the reaction cannot be readily observed, the presence of enzymatic activity can nonetheless be detected. Using the same generic assays and substrate mixture, we successfully detected the activity of five enzymes in Table 3, by simply monitoring specific metabolite changes in the reaction mixture. These results demonstrate the potential of the approach to screen for a wide variety of enzymatic activities using a common metabolite pool and standardized conditions to simultaneously and directly identify the charged substrate(s) and product(s) of a reaction without any prior knowledge of the biochemistry of a reaction.

Screening of Functionally Uncharacterized Proteins in the *E. coli* W3110 Proteome. While in the above validation step we used enzymes with known activity, we reasoned that the strategy could be readily applied to screen functionally uncharacterized proteins for enzymatic activity. As a proof-of-principle, we tested proteins of unknown function present in the *E. coli* W3110 strain proteome. A large target set of functionally uncharacterized proteins (encoded by so-called y-genes[23]) was first selected from the GenoBase 4.0 database as proteins of unknown or hypothetical function. We surveyed these proteins using the NCBI Conserved Domain Search (rpsblast, http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) to limit the selection to proteins carrying likely catalytic domains. Among these we chose 36 candidates, and eventually examined a subset of 25 hypothetical proteins which could be successfully purified.

Reactions were first performed using pools of five purified recombinant proteins together with the YE metabolite mixture supplemented with cofactors. Following the reaction, mixtures were profiled by CE−MS in the same way as described above. When significant changes were observed in pooled samples, a second round of screening was performed using single proteins and specific monitoring by CE−MS in SIM mode using the specific m/z values of peaks that showed significant differences during the first round of screening. In this manner, we successfully detected two enzymatic activities associated specifically with YbhA and YbiV. Both activities were detected from the appearance of a specific peak at m/z 171 that was not

Table 3. List of Enzymes Used for Method Validation

gene	enzyme	enzyme code	primary reaction[a]
cadA	lysine decarboxylase	4.1.1.18	L-lysine = cadaverine + CO_2
ade	adenine deaminase	3.5.4.2	adenine + H_2O = hypoxanthine + NH_3
add	adenosine deaminase	3.5.4.4	adenosine + H_2O = inosine + NH_3
rihC	ribonucleoside hydrolase	3.2.2.1	*N*-D-ribosylpurine + H_2O = purine + D-ribose[b]
metK	methionine adenosyltransferase	2.5.1.6	L-methionine + ATP + H_2O = phosphate + diphosphate +*S*-adenosyl L-methionine

[a] As defined in enzyme database BRENDA; http://www.brenda.uni-koeln.de/. [b] *N*-D-ribosylpurine; adenosine and guanosine, purine; adenine and guanine.

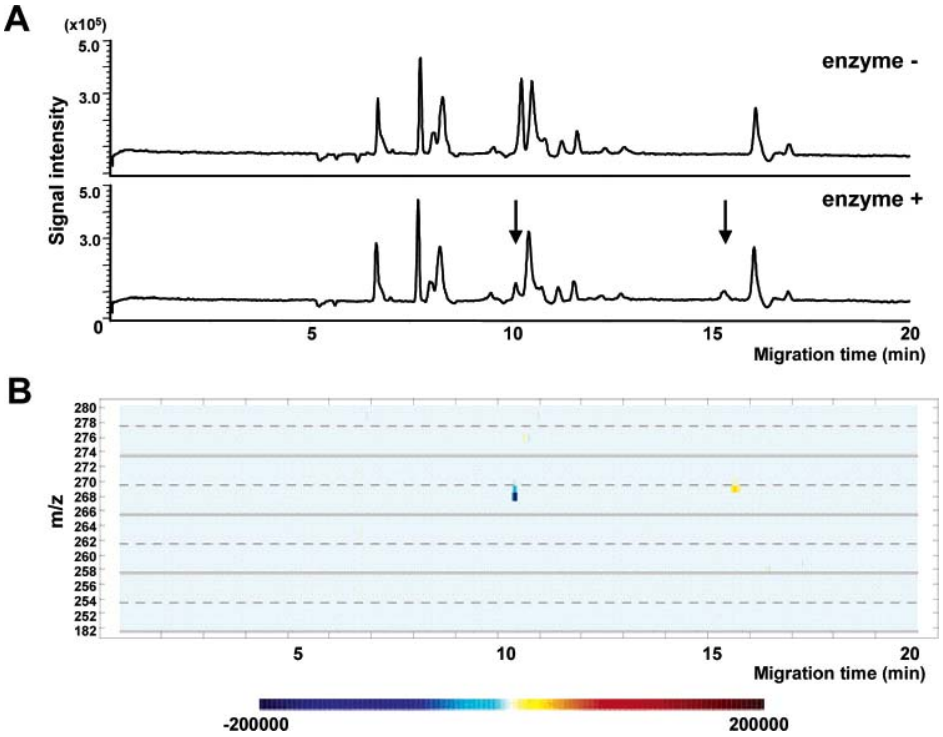

Figure 3. Detection of compounds associated with the activity of adenosine deaminase. (A). Total ion chromatograms (TICs) of the metabolite mixture after reaction with (enzyme +) or without (enzyme −) enzyme over a 251−280 *m*/*z* value range in positive mode CE-MS. The arrows mark distinct peaks that decreased or increased after the enzyme reaction. (B). Differential visualization of the mixtures' contents in the 251−280 *m*/*z* value range in cationic mode analysis. Peaks of equivalent intensity between the control and reaction mixtures are invisible in such plots while peaks whose intensity varies are colored according to the direction of the change. The numbers under the color range correspond to the absolute ion count differences between the two samples multiplied by the relative change in peak intensity between two samples; ex. A value of 8100 will be derived from the intensities of two peaks of 10 000 and 1000 counts (9000 × 9000/10 000). Light to dark blue plots represents decreased ion intensities while yellow to red represents increased ion intensities after enzyme reaction.

prominent in the control reaction performed without protein (Figure 5). The products for the two reactions were presumed to be identical since both *m*/*z* value and migration time were the same (Figure 5). Moreover, both YbhA and YbiV proteins exhibit significant amino acid similarity (42%), and both carry a hydrolase domain. We measured the accurate mass of the reaction product by CE-TOFMS (172.0136), and compared it with the theoretical mass of compounds in the KEGG LIGAND database (http://www.genome.ad.jp/ligand/).[24] Among the candidate compounds the closest one, whose mass was within 10 ppm of the measured one, was glycerol phosphate (172.0144, $C_3H_9O_6P$). We then also compared glycerol phosphate's relative migration time by CE with that of the reaction product and found them to be very close at 0.82 vs 0.81, respectively. We thus concluded that the unknown peak of *m*/*z* 171 was glycerol phosphate. We could not detect any significant decrease in peak intensity in the CE−MS profiles suggesting that the substrate of reaction might be undetectable by CE−MS under conditions used (e.g., neutral compound).

To predict the enzymatic activity of YbhA and YbiV from the above information, we surveyed the reactions associated with glycerol phosphate using KEGG LIGAND database. Two candidate reactions, glycerol phosphate phosphatase and glycerol kinase were selected as the most likely reactions since their possible substrates were present in the reaction mixture used. We thus next investigated these possibilities using biochemical analysis to clarify the activity of YbhA and YbiV.

Characterization of YbhA and YbiV Proteins. While this work was in progress, the function of *E. coli* YbiV was reported using structural and sequence information as well as biochemical experiments[25] and the protein classified as a member of

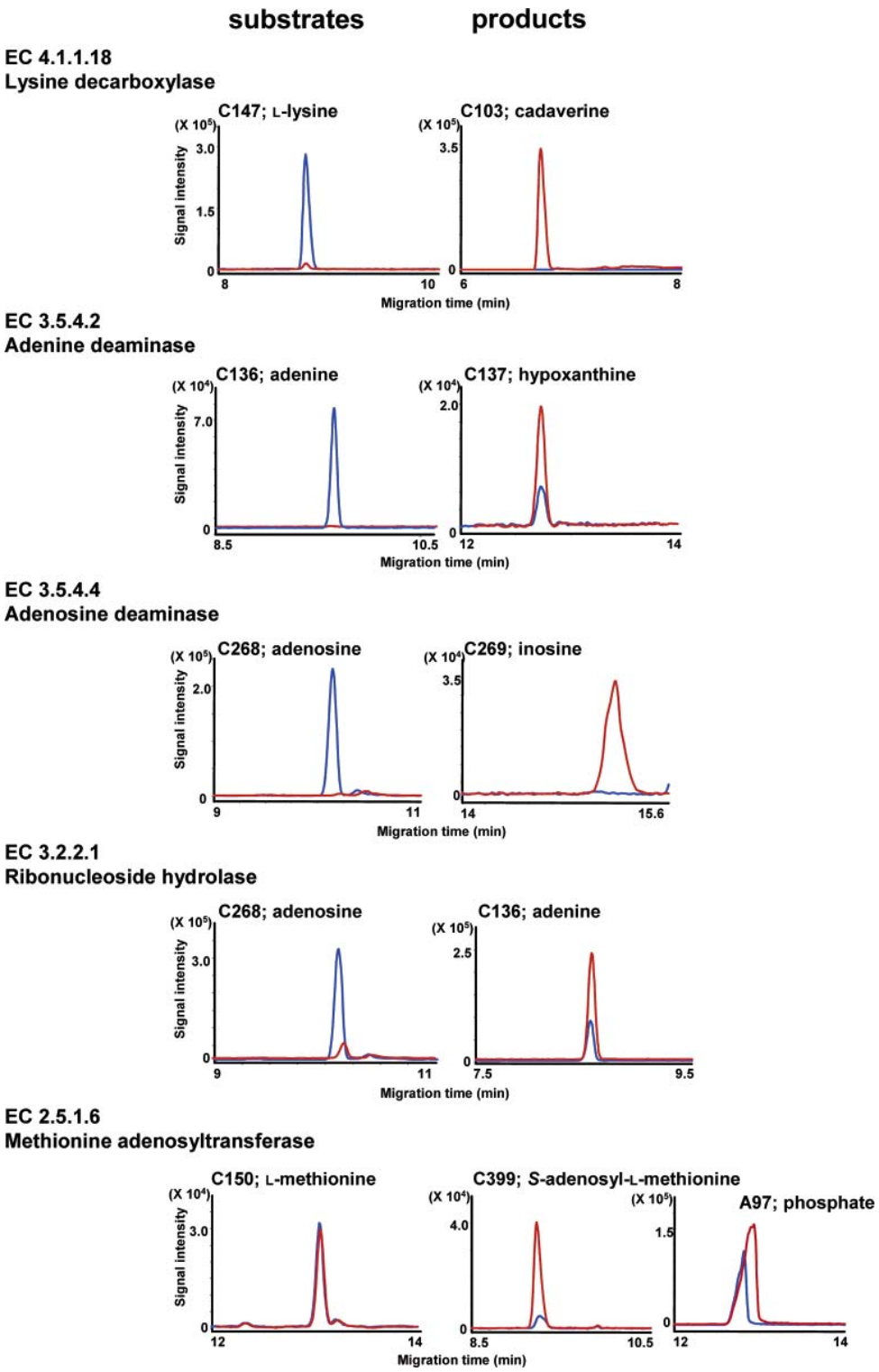

Figure 4. Specific changes in metabolite profiles for different enzymes of known activity. Selected ion electropherograms corresponding to the *m*/*z* of the predominant peak intensity changes after enzyme reactions. Signals from each electropherogram are overlaid: reaction with (red) and without (blue) enzyme. The symbols C and A represent cations and anions, respectively, while the adjacent number corresponds to the *m*/*z* value.

the haloacid dehydrogenase (HAD) hydrolase family and more specifically a sugar phosphatase. In addition, Kuznetsova et al. recently reported, in a review article containing previously unpublished work, the discovery of multiple novel phosphatases in the *E. coli* proteome, including a phosphatase activity for YbhA and YbiV.[26] However, there was no report of a possible phosphotransferase activity corresponding to our findings.

To confirm the validity of our findings in light of these reports, we performed gel filtration of the two recombinant

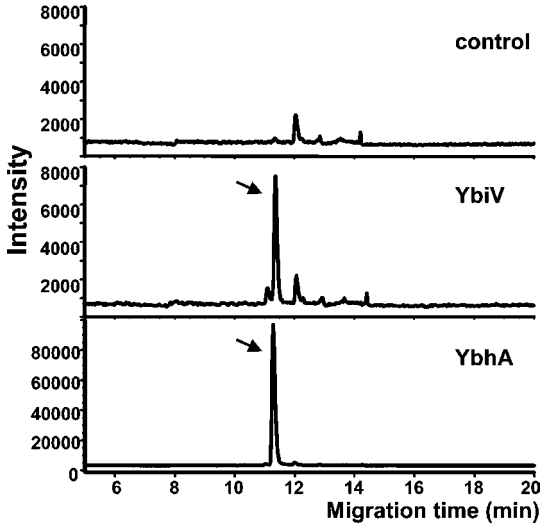

Figure 5. Changes in the metabolite incubated with uncharacterized *E. coli* proteins. Selected ion electropherograms (*m*/*z* 171) showing the same reactant peak in control (without protein; upper panel), YbiV (middle), and YbhA (lower) samples. Peaks indicated by arrows correspond to the compounds specifically produced by in the presence of YbiV and YbhA.

Table 4. Specific Activity of YbhA and YbiV Proteins toward Different Substrates

phosphatase[a] and phosphotransferase[b] activity (μmol/min/mg)		
substrates (products)	YbhA	YbiV
ribose-5-phosphate[a]	2.6	3.0
glucose-6-phosphate[a]	2.0	1.4
glycerol-3-phosphate[a] 2.3	0.7	
glycerol (glycerol phosphate)[b]	2.0×10^{-3}	5.0×10^{-3}

[a] Phosphatase activity was measured as the rate of substrate-decrease (μmol/min) per mg of protein. [b] Phosphotransferase activity was measured as the rate of product-increase (μmol/min) per mg of protein.

proteins to remove possible contaminants in the preparation and also to estimate the molecular weights of the proteins. Figure S3 shows the results of this experiment as a semilog plot of the molecular weights of standard proteins vs the K_{av} ($V_e - V_0/V_t - V_0$) value. From the K_{av} value of 0.4 and 0.5 for two main peaks of YbiV, the native molecular weight was calculated for the two forms as 51 000 and 26 000. The molecular weight of YbhA protein was similarly estimated as 65 000 and 32 000 from K_{av} of 0.39 and 0.48. The predicted molecular weight of YbiV and YbhA calculated from their amino acid sequences are 30 407 and 30 188, respectively. Therefore, the results of the gel filtration and SDS-PAGE analysis suggest that both proteins could exist in monomeric and homo-dimeric forms.

We next examined both the phosphotransferase and phosphatase activities of YbhA and YbiV using in vitro assays with gel-filtrated proteins (dimer and/or monomer). Table 4 shows the specific activities of YbhA and YbiV toward different potential substrates. The substrate specificity for the phosphatase activity of YbiV appears similar to that recently reported by Roberts et al.[25] with an apparent preference for ribose-5-phosphate. On the other hand, the level of phosphatase activity of YbhA for three different substrates tested was very similar, suggesting that the enzyme displays loose substrate specificity. The phosphotransferase activity was also examined in vitro through the formation of glycerol phosphate from glycerol and inorganic phosphate. Although the level of phosphotransferase activity in both YbhA and YbiV appeared low compared to the phosphatase activity, under the assay conditions, it could nonetheless be detected due to the high sensitivity of the CE−MS detection (Table 4). We also tested other potential phosphate donors for the phosphotransferase reaction such as ATP, GTP, ADP, AMP, ribose-5-phosphate, and glucose-6-phosphate. While the mono-phosphates could act as phosphate donor, neither nucleoside di-,tri-phosphates displayed such property (data not shown). These results suggest that YbhA and YbiV can act as both sugar phosphatases/phosphotransferases of with possibly broad substrate specificity and that the physiological importance of each activity may be dictated by the changing intracellular metabolic demands. The fact that we failed to detect the apparently stronger phosphatase activity during screening is likely due to the fact that the sugar phosphates were absent or present at very low levels in the YE metabolite mixture (Table 2). On the other hand, we could detect the rather weak phosphotransferase activity (Figure 5), because of the abundance of glycerol in the reaction mixture which originated from the protein solution. Together these events reflect well both the advantages but also the possible limits of the method. Very weak enzymatic activities can be detected due to the sensitivity of the MS detection while at the same time strong activities may be missed when the amount of potential substrate in the metabolite mixture is low or absent. As such the use of various types of metabolite mixtures, complemented with mixtures containing controllable amounts of commercially available chemicals may prove useful.

During screening, some of the compounds that may be readily detected may be related to reactions products that are common in biochemical pathways such as ATP/ADP or NAD/NADH etc., while the actual substrate may be more elusive due to compound's neutrality, for example. In such cases, an alternative analytical method such as LC−MS, GC−MS or direct infusion-MS may help identifying the other substrate or product of the reaction. Therefore, while no single analytical method can allow to monitor the whole metabolome at once, we believe our CE−MS approach is powerful and useful for polar and charged metabolites and intermediates that are very common in the main metabolic pathways.

Conclusion

In this work, we introduced a strategy to discover novel enzymatic activities based on in vitro biochemical assays and metabolite profiling using CE−MS. Our method focuses on detection of enzymatic reactions associated with charged metabolites. Its applicability was demonstrated using known *E. coli* metabolic enzymes and a complex mixture of metabolites. We successfully monitored several enzymatic reactions by observing changes in the levels of substrate(s) and/or product(s) of the enzyme reaction in the same complex mixture of metabolites and successfully linked two similar phosphatase/phosphotransferase activities to YbhA and YbiV. The power of the method resides in the fact that no a priori knowledge about either the protein activity or the composition of the metabolite mixture is necessary. Moreover the use of a biochemical assay with purified proteins allows to link changes in the metabolite levels to a catalytic activity, with simultaneous elucidation of the reaction's substrate(s) and product(s).

Our results clearly demonstrate the potential of this combination of biochemical assays and CE−MS detection system to successfully discover new enzymatic activities in pools of

uncharacterized proteins. In addition, the same approach could potentially be used to assign new activities to known enzymes, or to screen for enzyme inhibitors. We are currently screening a larger number of uncharacterized proteins and this should allow us to uncover several new enzymatic activities. Our method is simple, generic, and represents a broadly applicable tool for functional proteomics.

Acknowledgment. The authors thank Takamasa Ishikawa, Yuki Ueno (Human Metabolome Technologies Inc.) and Yuji Kakazu (IAB Keio University) for technical support. We are also grateful to Kenji Higashi and Masahiro Sugimoto (IAB Keio University) for the development of data analysis tools and Tomoya Baba and Akio Kanai (IAB Keio University) for helpful suggestions. This work was supported in part by a grant from New Energy and Industrial Technology Development Organization (NEDO) of the Ministry of Economy, Trade and Industry of Japan (Development of a Technological Infrastructure for Industrial Bioprocesses Project), from a Grant-in-Aid for Scientific Research on Priority Areas, The Ministry of Education, Culture, Sports, Science and Technology (MEXT) and The 21st Century COE Program, MEXT, and by research funds from the Yamagata prefectural government and Tsuruoka city.

Supporting Information Available: Survey of global metabolite changes associated with the activity of adenosine deaminase (Supporting Information Figure S1). Example of adduct formation during ionization (Supporting Information Figure S2). Molecular weight estimation of YbhA and YbiV proteins by gel filtration (Supporting Information Figure S3). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Blattner, F. R.; Plunkett, G., 3rd; Bloch, C. A.; Perna, N. T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J. D.; Rode, C. K.; Mayhew, G. F.; Gregor, J.; Davis, N. W.; Kirkpatrick, H. A.; Goeden, M. A.; Rose, D. J.; Mau, B.; Shao, Y. *Science* **1997**, *277*, 1453−1474.

(2) Riley, M.; Abe, T.; Arnaud, M. B.; Berlyn, M. K.; Blattner, F. R.; Chaudhuri, R. R.; Glasner, J. D.; Horiuchi, T.; Keseler, I. M.; Kosuge, T.; Mori, H.; Perna, N. T.; Plunkett, G., 3rd; Rudd, K. E.; Serres, M. H.; Thomas, G. H.; Thomson, N. R.; Wishart, D.; Wanner, B. L. *Nucleic. Acids. Res.* **2006**, *34*, 1−9.

(3) Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. *Nature* **2001**, *409*, 258−268.

(4) Zhu, H.; Klemic, J. F.; Chang, S.; Bertone, P.; Casamayor, A.; Klemic, K. G.; Smith, D.; Gerstein, M.; Reed, M. A.; Snyder, M. *Nat. Genet.* **2000**, *26*, 283−289.

(5) Adam, G. C.; Sorensen, E. J.; Cravatt, B. F. *Nat. Biotechnol.* **2002**, *20*, 805−809.

(6) Proudfoot, M.; Kuznetsova, E.; Brown, G.; Rao, N. N.; Kitagawa, M.; Mori, H.; Savchenko, A.; Yakunin, A. F. *J. Biol. Chem.* **2004**, *279*, 54687−54694.

(7) Hsieh, F. Y.; Tong, X.; Wachs, T.; Ganem, B.; Henion, J. *Anal. Biochem.* **1995**, *229*, 20−25.

(8) Bothner, B.; Chavez, R.; Wei, J.; Strupp, C.; Phung, Q.; Schneemann, A.; Siuzdak, G. *J. Biol. Chem.* **2000**, *275*, 13455−13459.

(9) Ge, X.; Sirich, T. L.; Beyer, M. K.; Desaire, H.; Leary, J. A. *Anal. Chem.* **2001**, *73*, 5078−5082.

(10) Pi, N.; Leary, J. A. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 233−243.

(11) Zea, C. J.; Pohl, N. L. *Anal. Biochem.* **2004**, *327*, 107−113.

(12) Saghatelian, A.; Trauger, S. A.; Want, E. J.; Hawkins, E. G.; Siuzdak, G.; Cravatt, B. F. *Biochemistry* **2004**, *43*, 14332−14339.

(13) Soga, T.; Heiger, D. N. *Anal. Chem.* **2000**, *72*, 1236−1241.

(14) Banks, J. F. *Electrophoresis* **1997**, *18*, 2255−2266.

(15) Simo, C.; Gonzalez, R.; Barbas, C.; Cifuentes, A. *Anal. Chem.* **2005**, *77*, 7709−7716.

(16) Soga, T.; Ueno, Y.; Naraoka, H.; Ohashi, Y.; Tomita, M.; Nishioka, T. *Anal. Chem.* **2002**, *74*, 2233−2239.

(17) Soga, T.; Ueno, Y.; Naraoka, H.; Matsuda, K.; Tomita, M.; Nishioka, T. *Anal. Chem.* **2002**, *74*, 6224−6229.

(18) Soga, T.; Ohashi, Y.; Ueno, Y.; Naraoka, H.; Tomita, M.; Nishioka, T. *J. Proteome Res.* **2003**, *2*, 488−494.

(19) Sato, S.; Soga, T.; Nishioka, T.; Tomita, M. *Plant, J.* **2004**, *40*, 151−163.

(20) Itoh, A.; Ohashi, Y.; Soga, T.; Mori, H.; Nishioka, T.; Tomita, M. *Electrophoresis* **2004**, *25*, 1996−2002.

(21) Kitagawa, M.; Ara, T.; Arifuzzaman, M.; Ioka-Nakamichi, T.; Inamoto, E.; Toyonaga, H.; Mori, H. *DNA Res.* **2005**, *12*, 291−299.

(22) Soga, T.; Baran, R.; Suematsu, M.; Ueno, Y.; Ikeda, S.; Sakurakawa, T.; Kakazu, Y.; Ishikawa, T.; Robert, M.; Nishioka, T.; Tomita, M. *J. Biol. Chem.* **2006**, *281*, 16768−16776.

(23) Rudd, K. E. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 985−1019.

(24) Goto, S.; Okuno, Y.; Hattori, M.; Nishioka, T.; Kanehisa, M. *Nucleic Acids Res.* **2002**, *30*, 402−404.

(25) Roberts, A.; Lee, S. Y.; McCullagh, E.; Silversmith, R. E.; Wemmer, D. E. *PROTEINS* **2005**, *58*, 790−801.

(26) Kuznetsova, E.; Proudfoot, M.; Sanders, S. A.; Reinking, J.; Savchenko, A.; Arrowsmith, C. H.; Edwards, A. M.; Yakunin, A. F. *FEMS Microbiol. Rev.* **2005**, *29*, 263−279.

PR0600576