research articles Journal of proteome research

Proteomics of *Pyrococcus furiosus*, a Hyperthermophilic Archaeon Refractory to Traditional Methods

Alice M. Lee,[†] Joel R. Sevinsky,[‡,#] Jonathan L. Bundy,[‡] Amy M. Grunden,*[,†] and James L. Stephenson, Jr.*[,‡]

Department of Microbiology, P.O. Box 7615, North Carolina State University, Raleigh, North Carolina 27695, and Biomarkers and Systems Biology Center, Research Triangle Institute, 3040 Cornwallis Road, Research Triangle Park, North Carolina 27709

Received December 29, 2008

Pyrococcus furiosus is one of the most extensively studied hyperthermophilic archaea. Proteins from this hyperthemophile organism are extremely thermostable and are highly resistant to chemical denaturants, organic solvents and proteolytic digestion. This thermostability makes it difficult to apply traditional methods of enzymatically digesting a complex mixture of proteins, commonly a first step in peptide generation in most shotgun proteomics methods. Here, we have developed a simple shotgun proteomics approach for the global identification of the *P. furiosus* proteome. This methodology uses a detergent-based microwave assisted acid hydrolysis (MAAH) step coupled with an overnight trypsin digest to obtain peptides. Subsequent peptide fractionation by isoelectric focusing in immobilized pH gradients (IPG-IEF), followed by chromatographic separation with reverse phase nano-HPLC and electrospray ionization tandem mass spectrometry (ESI-MS/MS) of peptides enabled the identification of over 900 proteins representing over 44% of the proteome. In most functional classes, over 50% of the predicted proteins were identified, including a number of membrane proteins. This new sample preparation technique will enable extensive proteomics data to be obtained for this organism, thereby enabling the reconstruction of metabolic pathways and promoting a systems biology based understanding of this important extremophile.

Keywords: *Pyrococcus furiosus* • shotgun proteomics • archaea • hyperthermophile • high throughput • IPG-IEF • LC-MS/MS • archaeal proteins • microwave assisted acid hydrolysis

Introduction

The Archaea represent one of the three primary domains of life, distinct from Bacteria and the Eucarya.[1] They inhabit some of the most extreme environments on Earth, from acidic pools, arctic ice and hypersaline lakes to volcanic vents. Hyperthermophilic archaea, organisms that have an optimal growth temperature of 80 °C or higher[2] are of particular interest. Enzymes from these organisms exhibit optimal activity under high temperatures, and thus have been a valuable source of thermostable enzymes for many industrial applications.[3] A cataloging of the proteome of these organisms is important in order to facilitate our use of these enzymes for novel metabolic reactions and to fully relate other 'omic scale data on these organisms. An increased understanding of these proteins will allow us to better employ them for applications in fields such as biotechnology, bioremediation, energy, and biofuel production.

One of the most studied hyperthermophilic archaea is *Pyrococcus furiosus*, a strictly anaerobic heterotroph that grows optimally at 100 °C. *P. furiosus* is an excellent candidate for proteomics studies due to its relative ease of cultivation, and the sequencing and *in silico* annotation of its genome.[4] The accumulation of genomic sequence data for *P. furiosus* and other hyperthermophilic archaea has produced a large set of new protein sequences, most of which have been derived from the translation of predicted open reading frames[5] (ORFs) with no experimental evidence relating to their expression or function. Nearly 50% of the genes in *P. furiosus* are not assigned a role category, or are identified as conserved hypothetical genes.[4] Since it is difficult to predict coding regions accurately from genomic data, whole proteome analysis can validate whether these predicted ORFs actually encode functional proteins. In addition, due to genome annotation discrepancies[6] in *P. furiosus*, use of a proteomics based approach to provide experimental evidence that specific ORFs are functional and or encode stable proteins will enable us to clarify the true number of genes in the genome.

P. furiosus was one of the first archaeal species whose proteome was examined using mass spectrometry.[7,8] With the use of matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) and capillary liquid chromatography electrospray

* Authors to whom correspondence should be directed. For J. L. S.: phone, (919) 316-3978; fax, (919) 541-6161; e-mail, stephensonjl@rti.org. For A. M. G.: phone, 919-513-4295; fax, 919-515-7867; e-mail, amy_grunden@ncsu.edu.
[†] North Carolina State University.
[‡] Research Triangle Institute.
[#] Current address: Luca Technologies, 500 Corporate Circle, Suite C, Golden, CO 80401.

Published on Web 05/08/2009
10.1021/pr801119h CCC: $40.75

ionization tandem mass spectrometry on proteins extracted from two-dimensional gels (2DE), the identification of approximately 60 proteins was reported.[7] A very recent report by Menon et al.[9] identified multiprotein complexes from *P. furiosus* using nondenaturing, sequential column chromatography and nano-LC-ESI-MS/MS.

The "shotgun" approach, reverse-phase liquid chromatography (LC) separations of proteolytic digests of whole cellular fractions coupled online with ESI-MS/MS, has been developed for high-throughput proteome analysis.[10] This method has facilitated the identification of thousands of expressed proteins in a single series of experiments.[11] Generally, the first step to a successful shotgun proteomics strategy is to create peptides from a complex protein mixture using proteases, a major obstacle in *P. furiosus*. Proteins from hyperthemophilic organisms show an intrinsic degree of thermostability,[3] which is associated with a higher resistance to chemical denaturants such as detergents, organic solvents and other chemical reagents.[12] Some thermophilic proteins also resist proteolytic digestion at moderate temperatures (20−60 °C) and only become sensitive to proteolytic attack above 70 °C.[3] Therefore, it is difficult to apply the traditional methods of enzymatically digesting a complex mixture of proteins using trypsin alone to obtain peptides from *P. furiosus*.

One way to circumvent this problem is the use of a microwave assisted acid hydrolysis (MAAH) step before overnight digestion with trypsin. The application of acid-catalyzed hydrolysis to degrade proteins to generate peptides can be traced to Sanger and Thompson.[13,14] More recently, Trauger et al.[15] applied an acid cleavable detergent (Rapigest) along with trypsin to examine the proteome of *P. furiosus*. In another study, an application developed by Fenselau's group has demonstrated that the use of microwave and acid catalyzed hydrolysis produces peptide products useful in the identification of *Bacillus* spores.[16] Zhong et al.[14] reported a method of protein identification based on MAAH using 25% trifluoroacetic acid (TFA) for protein degradation followed by LC-MALDI MS/MS of the resultant peptides. Microwave irradiation has been known to accelerate chemical or enzymatic reactions,[17] and the use of MAAH can significantly increase the speed of protein processing for proteome analysis.

Here, we have developed a simple shotgun proteomics approach as an alternative to gel-based proteomic platforms to identify the diversity of proteins produced by *P. furiosus*. This technique uses detergent with acid in a microwave-assisted acid hydrolysis step before overnight bulk digestion of proteins with trypsin. This complex mixture of mostly tryptic peptides is then fractionated by isoelectric focusing in immobilized pH gradients (IPG-IEF), followed by chromatographic separation with reverse phase nano-HPLC and ESI-MS/MS analysis of peptides using an ion trap mass spectrometer.[18] This combined approach identified over 900 unique proteins from *P. furiosus*, which is the largest number of reported proteins identified from this organism to date using shotgun-based proteomic techniques.

Experimental Section

Reagents. All standard chemicals were purchased from Sigma-Aldrich (St. Louis, MO) unless otherwise noted. HPLC-grade water and acetonitrile were purchased from Burdick and Jackson (Muskegon, MI).

Growth of Microorganisms. *P. furiosus* (DSM 3638) was grown at 90 °C in a 1-L capacity Corning pyrex media bottle with a 1% inoculum as previously described.[19] In addition to the 500 mL of basic salts medium, 5.0 mL of 100 mM $Na_2WO_4 \cdot 2H_2O$, 2.5 mL of yeast extract stock (10% w/v), 5.0 mL of maltose stock (50% w/v), and 5.0 mL of potassium phosphate stock (0.1 M, pH 6.8) were added.

For all experiments, *P. furiosus* was grown for 15 h and harvested by centrifugation at 6000*g* in a Sorvall RC-5B centrifuge for 30 min at 4 °C. Cell pellets were collected and stored at −80 °C. Two cell pellet samples, referred to as Pfu-W1 and Pfu-W2 were used for the proteomics analyses.

Preparation of Digested Lysate. A *P. furiosus* cell pellet of 0.1 g was resuspended in 400 μL of 25 mM Tris-Cl, pH 8, 1% SDS, and 5 mM DTT, and vortexed for 2 min in a 2 mL microcentrifuge tube. The sample was then sonicated on ice at 60% amplitude using 2 s pulses in a Digital Sonifier (Branson) for 20 pulses or until the sample appeared to be completely free of solid cellular material. Cellular debris was removed by centrifugation at 25 000*g* for 30 min at 4 °C. The whole cell lysate supernatant was incubated in a 37 °C water bath for 1 h to completely reduce the sample. The protein content was assayed using the Bradford method (Biorad, Hercules, CA). Next, 1 mg of whole cell lysate was transferred to a 2 mL microcentrifuge tube and formic acid (50% stock) was added to a final concentration of 6%. The sample was then placed in the center of a domestic microwave (1.2 kW Daewoo), with the cap open, and irradiated for 2 min and 45 s at the highest power setting to achieve 40 °C. The pH was adjusted to pH 7.0 with 500 mM Tris-Cl, pH 8.0. Then 20 μg of sequencing-grade trypsin (Promega, Madison, WI) was added to the sample and incubated overnight at 37 °C. In preparation for peptide IPG-IEF, tryptic peptides were desalted using a C_{18} Sep-Pak (Waters, Milford, MA) according to the manufacturer's protocol and eluted in 2 mL of 85% acetonitrile and 15 mM ammonium formate. Residual SDS was removed using HILIC (hydrophilic interaction chromatography) SPE cartridges (The Nest Group, Southborough, MA) according to the manufacturer's protocol. The peptides were eluted off the HILIC cartridge in 600 μL of HPLC-grade water and were lyophilized to dryness using a Speedvac (Thermo Savant, Waltham, MA).

Peptide IPG-IEF. Peptide IPG-IEF was performed as previously described.[20] Briefly, peptides were resuspended in 450 μL of 8 M urea and 0.5% ampholytes buffer, pH 3.5−5.0 (Amersham Biosciences, Piscataway, NJ). A 24 cm pH 3.5−4.5 pH linear Immobiline Drystrip (Amersham Biosciences) was rehydrated overnight with the peptide solution. Peptides were focused using an Ettan IPGphor IEF system (GE Healthcare, Piscataway, NJ) according to the manufacturer's protocol for a linear pH 3.5−4.5 IPG strip condition (step to 300 V for 900 Vh, gradient to 1000 V for 3900 Vh, gradient to 8000 V for 13 500 Vh, step to 8000 V for 93 700 Vh). After focusing, the IPG strip was cut into 60 fractions 4 mm in length. Each fraction was extracted three times for 30 min with 200 μL of 0.1% TFA. The three extracts for each fraction were pooled and desalted using a 96-well SPE plate packed with Waters HLB resin. Peptides were sequentially eluted in 500 μL of 50% acetonitrile and 0.1% TFA and 500 μL of 100% acetonitrile and 0.1% TFA. The eluates for each fraction were pooled and dried in a Speedvac and then resuspended in 25 μL of 0.1% TFA.

Mass Spectrometry. For inline LC-MS/MS, Zorbax C_{18} media (Agilent Technologies, Santa Clara, CA) was used to make a 1.5 cm 360 μm o.d. × 100 μm i.d. fused silica trap contained in a column holder (Upchurch Scientific, Oak Harbor WA) and a 10 cm 360 μm × 75 μm i.d. fused silica analytical column that

was integral with the nanospray tip. The mobile phases employed were A, 0.1% formic acid; B, 70% acetonitrile and 0.1% formic acid. Samples (10 μL) were loaded onto trap columns using an Eksigent (Dublin, CA) Nano-2 D LC system with autosampler and automatic 10-port switching valve (Valco Instruments, Houston TX) at a flow rate of 3 μL/min with 98% A. Samples were washed with 98% A for 35 min before switching inline with the analytical column. Chromatography was performed as follows: 0−44 min, 5−50% B; 44−45 min, 50−100% B; 45−47 min, 100−5% B; 47−55 min, 5−5% B. MS/MS was performed on a ThermoFisher Scientific LTQ XL (San Jose, CA). The instrument was fitted with a New Objective Picoview (Woburn, MA) source for online nanospray. Precursor ion isolation width was set to 2 Da, with default charge state setting of two. The MS/MS parameters used selective mass range scans followed by data-dependent MS/MS (1 μscan for Pfu-W1 and 3 μscans for Pfu-W2) of the two or three most intense peaks (400−600 m/z, top two peaks; 600−700 m/z, top three peaks; 700−800 m/z, top three peaks, 800−900 m/z, top three peaks; 900−1300 m/z, top 2 peaks). Dynamic exclusion was enabled using a repeat count of 1 over duration of 60 s with a mass width of 1.5.

Database Searching. Tandem mass spectra were extracted by BioWorks version 3.3. Because of the large number of spectra collected for the Pfu-W1 samples (1 μscan per MS/MS spectrum as opposed to 3 μscans per MS/MS spectrum for Pfu-W2), a Spectral Quality filter setting of 0.2 was used during data generation in Bioworks. All MS/MS samples were analyzed using SEQUEST (ThermoFisher Scientific, San Jose, CA; version 27, rev. 12). A protein database of *P. furiosus* proteins (NC_003413, 2125 proteins) was inverted (abcxyz inverted is azbycx) and appended to the NC_003413 database. SEQUEST was configured to search the NC_003413_inverted database twice, once assuming trypsin digestion and once assuming nonspecific cleavages. SEQUEST parameters included a fragment ion mass tolerance of 1.00 Da and a parent ion tolerance of 2.0 Da. Oxidation of methionine was specified as a variable modification.

Criteria for Protein Identification. Protein and peptide identification utilized peptide p*I*, target/decoy, and probability filtering. SEQUEST results were initially filtered by peptide p*I*[21] and target/decoy methods.[22] For each fraction, all top ranked peptide identifications in each .out file, both target and decoy, were collected, separated by charge state, and the p*I* for each peptide sequence was calculated using a novel p*I* prediction algorithm.[21,23] The peptide p*I* range for each charge state was calculated using quartile filtering of all target peptides with XCorrs greater than the highest decoy database identification (median ± 1.5*interquartile range). The p*I* range for the entire fraction was determined by using the highest and lowest p*I* values for the three charge states within the fraction. In instances where there were too few peptides in the +1 charge state to accurately determine the p*I* range, only the +2 and +3 charge states were considered. All identifications within this p*I* range, both target and decoy, were then used for probability filtering. Probability filtering used Scaffold (version Scaffold-01_07_00, Proteome Software, Inc., Portland, OR) to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm.[24] Protein identifications were accepted if they could be established at greater than 95.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm.[25] Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. Since decoy database identifications were included in the Scaffold analysis, we were able to apply an additional layer of filtering. Initial Scaffold analysis produced a list of 1057 proteins, 69 of which were from the decoy database, for a 13.1% false positive rate. Since we had two replicates and two searches per replicate (tryptic and nonspecific), for a total of four database searches, a protein identification was only accepted if it was identified at greater than 95% probability, contained at least two identified peptides, and it was identified in at least two separate searches. This reduced the number of identified proteins to 948 with 16 decoy proteins (932 true identifications), for a false positive rate of 3.4%. All processing was done on a desktop computer with a Pentium 4, 2.8 GHz processor and 3 GB RAM. A comprehensive list of the peptides and proteins identified in this work is available as Supporting Information.

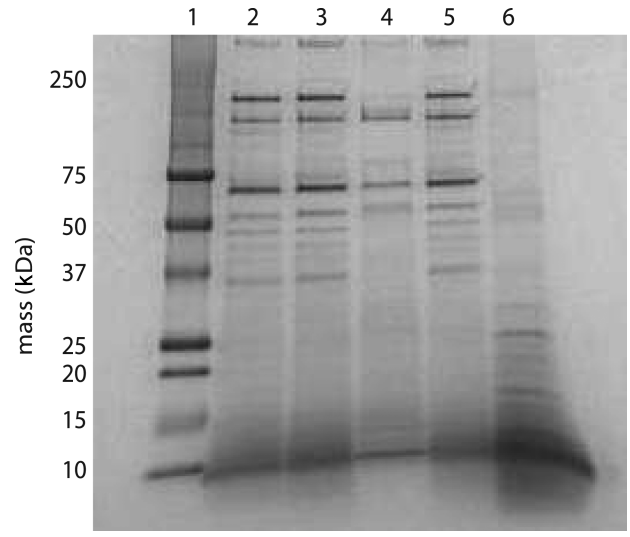

Figure 1. PAGE of the peptides resulting from the various digestion schemes applied to the *P. furiosus* cell pellets. (1) Precision plus blue standard; (2) 1 M Urea, 0.14% SDS, and 0.71 mM DTT; (3) 1.4% acetonitrile (ACN), 0.14% SDS, and 0.71 mM DTT; (4) 0.86% Formic Acid (microwave assisted acid hydrolysis) and 0.71 mM DTT; (5) 0.14% SDS and 0.71 mM DTT; (6) 0.86% Formic Acid/MAAH, 0.14% SDS, and 0.71 mM DTT.

Results and Discussion

Proteolytic Susceptibility of Hyperthermophilic Proteins. One of the major factors thought to affect proteolysis is the conformational mobility of protein structure.[26,27] The susceptibility of a peptide bond to proteolytic cleavage is partly determined by the flexibility of the protein chain region in which it is located, and the extent to which the bond is exposed.[28−30] This susceptibility can be influenced by high temperature, and the addition of denaturants or co-solvents, which increases protein flexibility.[3,31] In addition, recent studies have shown that there is also a direct correlation between the thermostability of a protein and its resistance to proteolytic degradation,[28] making traditional sample processing methods ineffective in obtaining peptides. This is supported by Figure 1 where we initially tried a variety of traditional tryptic digestion conditions to proteolyze *P. furiosus* proteins. The use of urea, acetonitrile, SDS, formic acid and DTT under various combina-

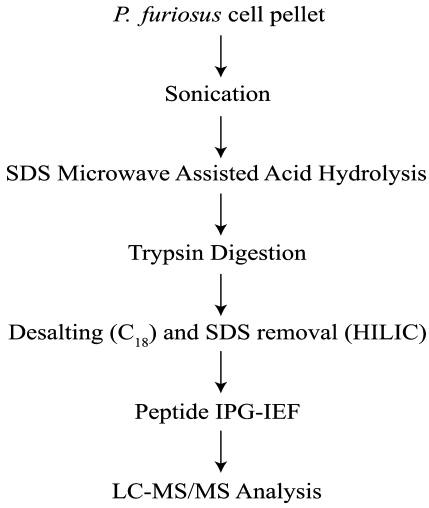

Figure 2. Workflow diagram for performing shotgun proteomics on *P. furiosus* via combined microwave assisted acid hydrolysis/proteolytic digestion.

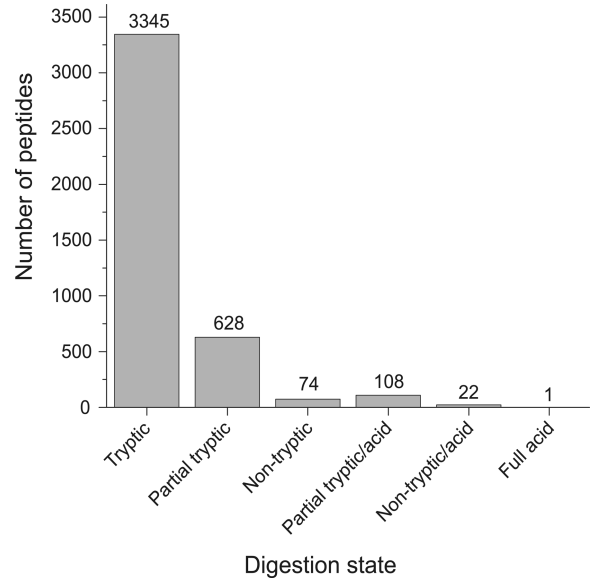

Figure 3. Numbers of peptides observed for the Pfu-W1 experiment categorized by tryptic digestion status.

tions failed to produce a complete digestion. As seen in Figure 1, lanes 2−5, most of the higher molecular weight proteins were not digested. Only when both formic acid and SDS were applied, Figure 1, lane 6, along with a microwave heat assisted step, did tryptic digestion of the higher molecular weight proteins occur. The SDS reduces excessive rigidity, and when coupled with an acid hydrolysis step under microwave conditions (MAAH), the elevated temperature further enhances protein flexibility and also induces thermal denaturation, thus, allowing for greater access to digestion sites for formic acid cleavage. These observations support previous studies by Park[32] and Zhong et al.[14] that enzymatic digestions of proteins that are resistant to proteolysis are significantly enhanced by thermal denaturation. Additional studies have shown *in vitro* that model proteases show preferences for denatured protein substrates.[33] The subsequent overnight trypsin digestion of these denatured and partially acid cleaved proteins results in the production of sufficient digest fragments for protein identification. A summary of this digestion scheme is found in Figure 2.

Our findings support the idea that proteolysis is highly correlated to the conformational mobility of proteins, especially hyperthermophilic proteins. The perturbation of molecular structure by microwave irradiation and SDS causes the thermally stable *P. furiosus* protein structure to be compromised, allowing for acid hydrolysis and trypsin digestion to generate sufficient peptides. Further studies of how these experimental conditions specifically influence protein structure may reveal mechanisms that underlie extreme thermostability.

Digestion Efficiency Using SDS-Microwave-Assisted Acid Hydrolysis and Trypsin. In this study, digestion efficiency was greatly enhanced through the use of a combination of SDS, low formic acid concentration (6%) and a relatively short microwave period (less than 3 min) to produce peptide sequences of various sizes. Several studies support the use of low formic acid concentrations with microwave-assisted hydrolysis to generate efficient protein digestion for MS analysis.[16,34] Hua et al.[34] demonstrated that formic acid cleavage at aspartic acid (Asp) residues under microwave irradiation can be used for peptide mapping. Swatkoski et al.[16] reported a single step protein extraction and microwave-assisted formic acid digestion of proteins from *Bacillus* spores. They used a relatively low concentration of formic acid (6% v/v) and a short microwave (90 s) reaction time to minimize nonspecific side reactions. Formic acid digestion will cleave proteins preferentially at the C-terminal or sometimes N-terminal end of the Asp residues of a protein.[35] Specificity is also an important consideration for proteolysis in protein identification; trypsin was used after acid hydrolysis in an overnight digestion since it generates peptides with desirable ionization and fragmentation characteristics.[31] Figure 3 shows that, for the initial Pfu-W1 study, a total of 4178 peptides for all digestion states were obtained, yielding predominantly tryptic or partial tryptic peptides (3973 fully tryptic peptides corresponding to 912 proteins). In contrast, there were significantly fewer peptides, 131 total that were either partial tryptic/acid, nontryptic/acid, and full acid cleaved. On the basis of our results, the formic acid treatment produced some acid cleaved peptides, but it is the specific trypsin cleavage that is responsible for generating the majority of the peptide identifications. This data confirms that pretreatment of *P. furiosus* proteins with SDS-MAAH provides proteases with greater access to potential cleavage sites, and when trypsin is applied to the overnight digestion, an abundance of tryptic peptides results. The fact that tryptic peptides were the predominant form of peptides seen greatly enhances the specificity of protein identification. In a repeat trial (Pfu-W2), 4040 total peptides were obtained, of which 3564 were either tryptic or partial tryptic peptides (907 identified proteins). The observed digestion specificity is very high (~88%) with trypsin as evidenced by appearance of mostly tryptic peptides.

Shotgun Proteomic Analysis. A plot of the peptide p*I* versus number of peptides for the whole *P. furiosus* proteome digested with trypsin *in silico* is shown in Figure 4a. This theoretical data reveals that a narrow-range p*I* strip covering the 3.5−4.5 p*I* range will successfully cover a large proportion of tryptic peptides. This p*I* range will maximize the resolution of peptides and increase the number of identifications from this mixture. The experimental data, as indicated in Figure 4b, reveals nearly all the peptides were in this range, but the last five fractions in Pfu-W1 reflect a bias toward peptides with a much higher p*I*

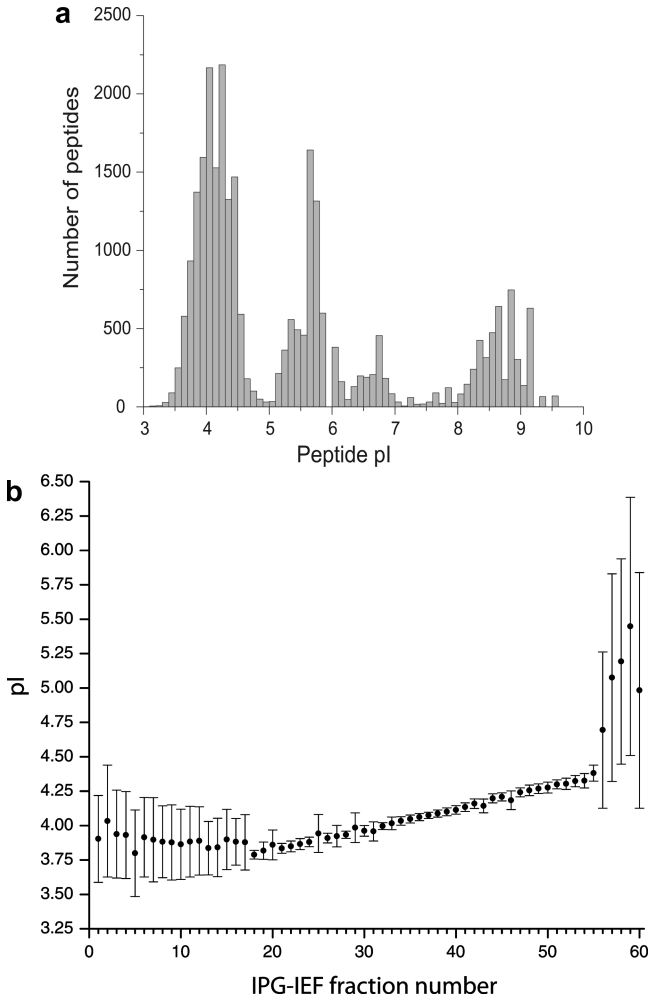

Figure 4. (a) Theoretical distribution of *P. furiosus* peptides as a function of isoelectric point. (b) Observed results of peptide IPG-IEF experiment (Pfu-W1), depicted as average p*I* ± standard deviation after quartile filtering.

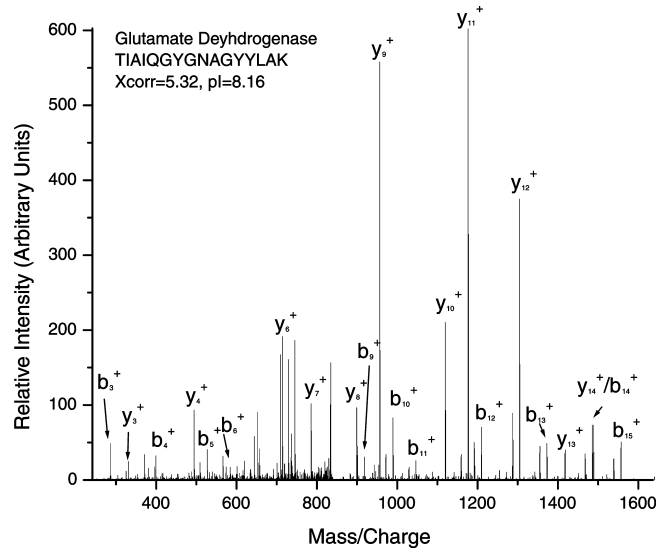

Figure 5. A representative ESI-MS/MS spectra of a peptide derived from glutamate dehydrogenase found at the extreme of basic p*I*.

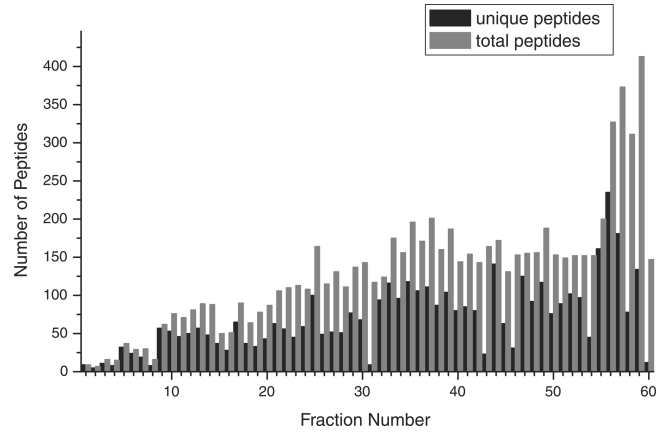

Figure 6. Observed number of unique and total peptides as a function of IEF fraction number for the Pfu-W1 data set.

range. The average p*I* range was 4.7−5.4 with a standard deviation range that covered p*I* values from 4.0 up to 6.5. The Pfu-W2 data, a biological repeat, illustrates that there were also several fractions toward the end of the strip that had high average p*I* values of 4.7−5.0 and a standard deviation range from 4.0 to 5.8. It is unusual to have such a large range of peptide p*I* values in a narrow range 3.5−4.5 IPG strip. These outlying peptides may originate from the focusing procedure itself, as has been observed in other studies[21] which have shown that the two ends of the gel strip tend to be poorly focused in comparison to the interior of the strip. It should be noted that in that study the ends of the strip had a lower percentage of unique peptide identifications. In contrast, we found higher numbers of unique peptides derived from the more basic end of the strip. One factor we have considered is that the presence of peptides with such aberrant p*I* values might result from residual SDS binding the peptides which would cause an acidic shift in the p*I* of the peptide. A more efficient removal of SDS, using chloroform/methanol/water[36] might correct for the distribution range of peptide p*I* values. It should be noted that the peptides with high p*I* values still have good XCorr's and can be identified to a specific protein. An example is depicted in the MS/MS spectrum shown in Figure 5. In our data, there were additional examples of peptides with p*I* values of over 8.0 that were readily identified, with strong database search engine scores.

A comparison between the total peptides and unique peptides found in each fraction can be seen in Figure 6 for Pfu-W1. It is apparent that a large number of different peptides are observed in all the fractions, but fractions 55−60 were particularly notable, having between 150 and 400 total peptides. A similar trend can also be seen in the Pfu-W2 data set (data not shown). It is thought that the propensity for a larger number of peptides to accumulate at the end of the IPG strip might be due to the residual SDS left after HILIC SPE cleanup. Any residual SDS will bind to the main peptide chain of very basic peptides, and effectively impart a negative charge of significantly greater magnitude than the original charge of the peptide which would influence movement of the peptide as it is focused. This may result in the total number of peptides obtained in each fraction being particularly skewed toward the last 5 fractions where a high concentration of "more basic" peptides was seen.

Global Identification of Proteins Expressed by *P. furiosus*. The data set consists of a list of expressed proteins derived by searching peptide tandem mass spectra against the

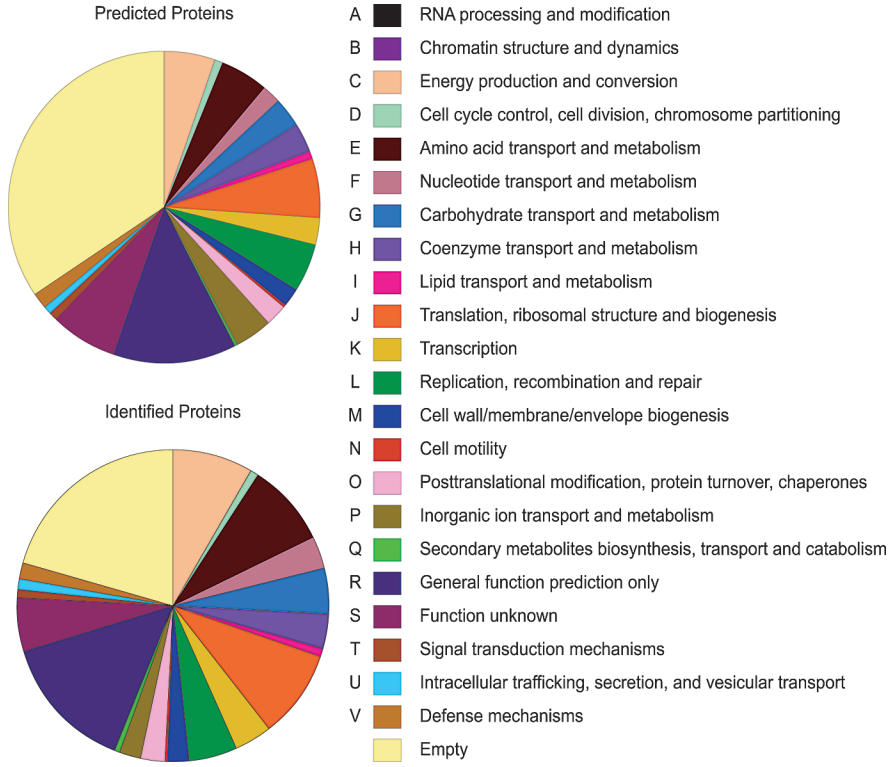

Figure 7. Pie charts showing observed and predicted proteins from *P. furiosus* based on functional category.

theoretical protein database of *P. furiosus*. A total of more than 8200 peptides (3717 unique) were identified from 2 samples using 2 mg of total protein as starting material (see Supporting Information). The reassembling of the unique peptides resulted in a total of 932 proteins, representing ~44% of the predicted proteins in *P. furiosus*. All of the proteins were identified based on at least two constituent peptides. The proteins identified in the study were sorted into functional categories using the "clusters of orthologous groups of proteins" (COG) classification system.[37,38] The proteins that did not belong to a COG were manually assigned as "empty". This is summarized in Figure 7, which shows the predicted distribution of the different functional classes of proteins from *P. furiosus* using COGs compared with the observed classes identified in this study. Proteins from all functional categories except RNA processing and modification were identified. A summary of the number of predicted ORFs and the identified proteins in each functional category is presented in Table 1. In most functional classes, over 50% of the predicted proteins were identified. It should be noted that these identified proteins include intracellular, extracellular and membrane bound proteins. Twenty-six percent of the proteins identified are annotated as empty or function unknown (20% empty, 6% function unknown); this is the largest functional class of proteins identified in this study. Clearly, such a large number of sequences identified among these classes of proteins suggests these molecules represent previously undocumented proteins synthesized in *P. furiosus*. Although the number of proteins identified in these classes was the highest, the relative percentage of the proteome observed is among one of the lowest, especially when compared to the classes of highly conserved protein families which were observed at much greater than 50%. This implies that current computational methods are capable of identifying highly conserved proteins, but new protein families that are specific to an organism or class of organisms that are not very well studied are difficult to predict. Fourteen percent of the proteins identified, the second largest percentage of proteins seen in this study, were classified as general function prediction only (i.e., conserved hypothetical proteins). Future studies to investigate the interaction of these genes and proteins in relation to genetic or environmental perturbations will be necessary to determine their function. A more thorough discussion of identified proteins in the context of *P. furiosus* biology is given in the Supporting Information.

Conclusions

Genome analysis can provide a list of all predicted genes, yet information regarding actual protein expression is difficult to measure. Thus, we have developed a shotgun proteomic approach using SDS-MAAH, IPG-IEF, nano-RPLC-MS/MS and computational analysis to characterize the peptides in complex mixtures of trypsin digested *P. furiosus* proteins. This approach requires a minimal amount of starting material, 1−2 mg, and significantly fewer separation and processing steps than previous reports. The results of this study include the identification of nearly 44% of the *P. furiosus* proteome over all functional classes. We believe that, in our method, the microwave irradiation induces higher temperatures, thus, increasing the flexibility of the protein structure, allowing for partial protein hydrolysis with formic acid. The addition of SDS further reduces protein rigidity, resulting in better proteolytic susceptibility than can be realized by either treatment separately. In our analysis following SDS-MAAH treatment, overnight trypsin digestion of these partially hydrolyzed proteins results in a nearly complete proteolytic digestion of *P. furiosus* proteins.

We anticipate this approach will work for other hyperthermophilic organisms recalcitrant to typical proteolytic methods. Researchers in the field of hyperthermophilic archaea can apply this technique to their proteomics studies with proteins that

Table 1. Relative Percentage of the Number of Proteins from Each Class That Was Identified[a]

functional classes (functional annotation)	total proteins predicted (predicted by genome)	total proteins observed (identified protein)	relative percentage of proteome observed
RNA processing and modification	1	0	0
Chromatin structure and dynamics	3	2	66.7
Energy production and conversion	111	76	68.5
Cell cycle control, cell division, chromosome partitioning	18	9	50
Amino acid transport and metabolism	107	80	74.8
Nucleotide transport and metabolism	43	32	74.4
Carbohydrate transport and metabolism	62	44	71
Coenzyme transport and metabolism	64	32	50
Lipid transport and metabolism	18	8	44.4
Translation, ribosomal structure and biogenesis	128	87	68
Transcription	63	34	54
Replication, recombination and repair	103	48	46.6
Cell wall/membrane/envelope biogenesis	45	20	44.4
Cell motility	6	2	33.3
Posttranslational modification, protein turnover, chaperones	43	25	58.1
Inorganic ion transport and metabolism	87	19	21.8
Secondary metabolites biosynthesis, transport and catabolism	6	5	83.3
General function prediction only	270	133	49.3
Function unknown	148	53	35.8
Signal transduction mechanisms	16	8	50
Intracellular trafficking, secretion, and vesicular transport	20	10	50
Defense mechanisms	34	15	44.1
Empty (Unclassified)	729	190	26.1
Total	**2125**	**932**	**44**

[a] Out of a predicted maximum of 2125 proteins, 932 unique proteins were identified.

normally resist proteolytic digestion with trypsin. With this technique, we have catalogued over 900 proteins; in particular, the proteins involved in metabolic pathways in *P. furiosus* cultured under maltose and higher tungsten conditions were investigated. The integrated computational and mass spectrometry data analysis in this study gives a better understanding of the expressed proteins. The information obtained in this study will be useful for subsequent work, such as expression of proteins of interest to investigate certain aspects of *Pyrococcus* biology.

As performed here, our method is not selective for secreted proteins (in the media supernatant), and insoluble proteins. Since it is unlikely that all the predicted putative proteins are expressed, we will certainly not detect the entire proteome. Even among the expressed proteins, post-translational or chemical modifications could lead to the formation of new or functionally different proteins that went undetected. This study was based on matching predicted proteins from the *P. furiosus* genome sequence and post-translational modifications were not considered. Furthermore, modified sample preparation protocols that target integral or membrane-associated proteins will enable even more expressed proteins to be identified in the future.

With this new proteomics technique for hyperthermophilic organisms, now extensive proteomics data can now be obtained and reconstruction of metabolic pathways can be attempted, thus, extending a systems biology understanding of this organism. The combination of novel genomic techniques and proteomic analysis will further enable us to follow metabolic and biochemical pathways, identify novel proteins with putative functions, and identify the pathways used by hyperthermophiles to cope with environmental stress.

Acknowledgment. The authors would like to acknowledge Catherine Fenselau, Steven Swatkoski and colleagues at the University of Maryland for helpful conversations and suggestions about acid catalyzed cleavage of proteins. The research reported here was funded by Internal Research and Development Funds from RTI (J.R.S., J.L.B. and J.L.S.), NASA's Institute for Advanced Concepts (A.M.L. and A.M.G.) and by the North Carolina Agriculture Research Station (A.M.G.).

Supporting Information Available: Comprehensive list of peptides identified in all experiments, with pertinent statistics in Excel Spreadsheet format. Discussion of proteins identified within the context of *P. furiosus* biology. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Woese, C. R. Bacterial evolution. *Microbiol. Rev.* **1987**, *51* (2), 221–271.
(2) Brock, T. D. *Thermophilic Microorganisms and Life at High Temperatures*; Springer: New York, 1978; Vol. xi.
(3) Vieille, C.; Zeikus, G. J. Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **2001**, *65* (1), 1–43.
(4) Robb, F. T.; Maeder, D. L.; Brown, J. R.; DiRuggiero, J.; Stump, M. D.; Yeh, R. K.; Weiss, R. B.; Dunn, D. M. Genomic sequence of hyperthermophile, Pyrococcus furiosus: implications for physiology and enzymology. *Methods Enzymol.* **2001**, *330*, 134–157.
(5) Yamazaki, S.; Yamazaki, J.; Nishijima, K.; Otsuka, R.; Mise, M.; Ishikawa, H.; Sasaki, K.; Tago, S.; Isono, K. Proteome analysis of an aerobic hyperthermophilic crenarchaeon, Aeropyrum pernix K1. *Mol. Cell. Proteomics* **2006**, *5* (5), 811–823.
(6) Poole, F. L., 2nd.; Gerwe, B. A.; Hopkins, R. C.; Schut, G. J.; Weinberg, M. V.; Jenney, F. E., Jr.; Adams, M. W. Defining genes in the genome of the hyperthermophilic archaeon Pyrococcus furiosus: implications for all microbial genomes. *J. Bacteriol.* **2005**, *187* (21), 7325–7332.
(7) Holden, J. F.; Poole II, F. L.; Tollaksen, S. L.; Giometti, C. S.; Lim, H.; Yates, J. R., III.; Adams, M. W. Identification of membrane proteins in the hyperthermophilic archaeon Pyrococcus furiosus using proteomics and prediciton programs. *Comp. Funct. Genomics* **2001**, *2*, 275–288.

(8) Lim, H.; Eng, J.; Yates, J. R.; Tollaksen, S. L.; Giometti, C. S.; Holden, J. F.; Adams, M. W.; Reich, C. I.; Olsen, G. J.; Hays, L. G. Identification of 2D-gel proteins: a comparison of MALDI/TOF peptide mass mapping to mu LC-ESI tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2003**, *14* (9), 957–970.
(9) Menon, A. L.; Poole, F. L.; Cvetkovic, A.; Trauger, S. A.; Kalisiak, E.; Scott, J. W.; Shanmukh, S.; Praissman, J.; Jenney, F. E.; Wikoff, W. R.; Apon, J. V.; Siuzdak, G.; Adams, M. W. Novel multiprotein complexes identified in the hyperthermophilic archaeon Pyrococcus furiosus by non-denaturing fractionation of the native proteome. *Mol. Cell. Proteomics* **2009**, *8*, 735–751.
(10) Hu, L.; Ye, M.; Jiang, X.; Feng, S.; Zou, H. Advances in hyphenated analytical techniques for shotgun proteome and peptidome analysis--a review. *Anal. Chim. Acta* **2007**, *598* (2), 193–204.
(11) Washburn, M. P.; Wolters, D.; Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–247.
(12) Sakuraba, H.; Ohshima, T. Novel energy metabolism in anaerobic hyperthermophilic archaea: a modified Embden-Meyerhof pathway. *J. Biosci. Bioeng.* **2002**, *93* (5), 441–448.
(13) Sanger, F.; Thompson, E. O. The amino-acid sequence in the glycyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *Biochem. J.* **1953**, *53* (3), 353–366.
(14) Zhong, H.; Marcus, S. L.; Li, L. Microwave-assisted acid hydrolysis of proteins combined with liquid chromatography MALDI MS/MS for protein identification. *J. Am. Soc. Mass Spectrom.* **2005**, *16* (4), 471–481.
(15) Trauger, S. A.; Kalisak, E.; Kalisiak, J.; Morita, H.; Weinberg, M. V.; Menon, A. L.; Poole Ii, F. L.; Adams, M. W.; Siuzdak, G. Correlating the transcriptome, proteome, and metabolome in the environmental adaptation of a hyperthermophile. *J. Proteome Res.* **2008**, *7* (3), 1027–1035.
(16) Swatkoski, S.; Russell, S. C.; Edwards, N.; Fenselau, C. Rapid chemical digestion of small acid-soluble spore proteins for analysis of Bacillus spores. *Anal. Chem.* **2006**, *78* (1), 181–188.
(17) Chiou, S. H.; W, K. T. Peptide and protein hydrolysis by microwave irradiation. *J. Chromatogr.* **1989**, *491* (2), 424–431.
(18) Cargile, B. J.; Sevinsky, J. R.; Essader, A. S.; Stephenson, J. L., Jr.; Bundy, J. L. Immobilized pH gradient isoelectric focusing as a first-dimension separation in shotgun proteomics. *J. Biomol. Tech.* **2005**, *16* (3), 181–189.
(19) Verhagen, M. F.; Menon, A. L.; Schut, G. J.; Adams, M. W. Pyrococcus furiosus: large-scale cultivation and enzyme purification. *Methods Enzymol.* **2001**, *330*, 25–30.
(20) Essader, A. S.; Cargile, B. J.; Bundy, J. L.; Stephenson, J. L., Jr. A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. *Proteomics* **2005**, *5* (1), 24–34.
(21) Cargile, B. J.; Talley, D. L.; Stephenson, J. L., Jr. Immobilized pH gradients as a first dimension in shotgun proteomics and analysis of the accuracy of pI predictability of peptides. *Electrophoresis* **2004**, *25* (6), 936–945.
(22) Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2003**, *2* (1), 43–50.
(23) Cargile, B. J.; Sevinsky, J. R.; Essader, A. S.; Eu, J. P.; Stephenson, J. L., Jr. Calculation of the isoelectric point of tryptic peptides in the pH 3.5−4.5 range based on adjacent amino acid effects. *Electrophoresis* **2008**, *29* (13), 2768–2778.
(24) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* **2002**, *74* (20), 5383–5392.
(25) Nesvizhskii, A. I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75* (17), 4646–4658.
(26) Hubbard, S. J.; Eisenmenger, F.; Thornton, J. M. Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci.* **1994**, *3* (5), 757–768.
(27) Fontana, A.; Fassina, G.; Vita, C.; Dalzoppo, D.; Zamai, M.; Zambonin, M. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. *Biochemistry* **1986**, *25* (8), 1847–1851.
(28) Mukherjee, S.; Guptasarma, P. Direct proteolysis-based purification of an overexpressed hyperthermophile protein from Escherichia coli lysate: a novel exploitation of the link between structural stability and proteolytic resistance. *Protein Expression Purif.* **2005**, *40* (1), 71–76.
(29) Hubbard, S. J. The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta* **1998**, *1382* (2), 191–206.
(30) Parsell, D. A.; Sauer, R. T. The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. *J. Biol. Chem.* **1989**, *264* (13), 7590–7595.
(31) Bark, S. J.; Muster, N.; Yates, J. R., III.; Siuzdak, G. High-temperature protein mass mapping using a thermophilic protease. *J. Am. Chem. Soc.* **2001**, *123* (8), 1774–1775.
(32) Park, Z. Y.; Russell, D. H. Thermal denaturation: a useful technique in peptide mass mapping. *Anal. Chem.* **2000**, *72* (11), 2667–2670.
(33) Pace, C. N.; Barrett, A. J. Kinetics of tryptic hydrolysis of the arginine-valine bond in folded and unfolded ribonuclease T1. *Biochem. J.* **1984**, *219* (2), 411–417.
(34) Hua, L.; Low, T. Y.; Sze, S. K. Microwave-assisted specific chemical digestion for rapid protein identification. *Proteomics* **2006**, *6* (2), 586–591.
(35) Inglis, A. S. Cleavage at aspartic acid. *Methods Enzymol.* **1983**, *91*, 324–332.
(36) Puchades, M.; Westman, A.; Blennow, K.; Davidsson, P. Removal of sodium dodecyl sulfate from protein samples prior to matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun. Mass Spectrom.* **1999**, *13* (5), 344–349.
(37) Tatusov, R. L.; Koonin, E. V.; Lipman, D. J. A genomic perspective on protein families. *Science* **1997**, *278* (5338), 631–637.
(38) Tatusov, R. L.; Natale, D. A.; Garkavtsev, I. V.; Tatusova, T. A.; Shankavaram, U. T.; Rao, B. S.; Kiryutin, B.; Galperin, M. Y.; Fedorova, N. D.; Koonin, E. V. The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* **2001**, *29* (1), 22–28.

PR801119H