

Part I: characterization of the extracellular proteome of the extreme thermophile *Caldicellulosiruptor saccharolyticus* by GeLC-MS²

David Muddiman · Genna Andrews · Derrick Lewis ·
Jaspreet Notey · Robert Kelly

Received: 3 May 2010 / Revised: 13 June 2010 / Accepted: 20 June 2010 / Published online: 10 July 2010
© Springer-Verlag 2010

Abstract The proteome of extremely thermophilic microorganisms affords a glimpse into the dynamics of microbial ecology of high temperature environments. The secretome, or extracellular proteome of these microorganisms, no doubt harbors technologically important enzymes and other thermostable biomolecules that, to date, have been characterized only to a limited extent. In the first of a two-part study on selected thermophiles, defining the secretome requires a sample preparation method that has no negative impact on all downstream experiments. Following efficient secretome purification, GeLC-MS² analysis and prediction servers suggested probable protein secretion to complement experimental data. In an effort to define the extracellular proteome of the extreme thermophilic bacterium *Caldicellulosiruptor saccharolyticus*, several techniques were considered regarding sample processing to achieve the most in-depth analysis of secreted proteins. Order of operation experiments, all including the C₁₈ bead technique, demonstrated that two levels of sample purification were necessary to effectively desalt the sample and provide sufficient protein identifications. Five sample preparation combinations yielded 71 proteins and the majority described, as enzymatic and putative uncharacterized proteins, anticipate consolidated

bioprocessing applications. Nineteen proteins were predicted by Phobius, SignalP, SecretomeP, or TatP for extracellular secretion, and 11 contained transmembrane domain stretches suggested by Phobius and transmembrane hidden Markov model. The sample preparation technique demonstrating the most effective outcome for *C. saccharolyticus* secreted proteins in this study, involved acetone precipitation followed by the C₁₈ bead method in which 2.4% (63 proteins) of the predicted proteome was identified, including proteins suggested to have secretion and transmembrane moieties.

Keywords Secretome · *Caldicellulosiruptor saccharolyticus* · GeLC-MS² · Thermophile · Sample preparation

Introduction

Extremophiles, bacterial microorganisms that thrive in extreme environmental conditions such as temperature, pressure, and pH extremes, offer unique resources for biocatalysis. Due to their habitats, extremophiles render robust enzymatic potential for processing materials under a variety of conditions [1, 2]. Hypotheses suggest and recent investigations evidence a variety of features and functions, including extracellular carbohydrate degradation [3–5] and antimicrobial potential [6, 7] of organisms with optimal growth temperature above 50 °C. One such example of enormous benefit, the thermostable Taq polymerase isolated from the extreme thermophile *Thermus aquaticus*, is widely used for DNA amplification in polymerase chain reaction experiments [8]. Furthermore, the significance and likely functions derived from the microbial secretome of thermophiles are grossly underestimated and command the positive identification of the

Electronic supplementary material The online version of this article (doi:10.1007/s00216-010-3955-6) contains supplementary material, which is available to authorized users.

D. Muddiman (✉) · G. Andrews
North Carolina State University,
Raleigh, NC 27695, USA
e-mail: david_muddiman@ncsu.edu
e-mail: DavidMuddiman@ncsu.edu

D. Lewis · J. Notey · R. Kelly
Chemical and Biomolecular Engineering,
North Carolina State University,
Raleigh, NC 27695, USA

secreted proteins. The secretome encompasses the proteins released by an organism into the surroundings as well as those associated with the outer cell wall [9]. This translocation can be instigated by different events such as communication between cells, interaction with various substrates, and environment stresses [10, 11]. Exploring the extracellular proteome is essential in understanding the microbial ecology of natural and bioprocessing environments.

Attributed to genome sequencing, considerable exploration of the genome, and growth on a variety of substrates, the extreme thermophile (optimal growth temperature >70 °C) *Caldicellulosiruptor saccharolyticus* offers a preview of extracellular thermophilic activity. *C. saccharolyticus*, originally isolated from a thermal spring in New Zealand, is a rod-shaped, gram-positive bacterium that grows optimally at 70 °C [12, 13]; more details pertaining to this organism are highlighted in Part II of this investigation. Defining this secretome and further examining thermophilic species afford a glimpse into microbial ecology. Transcriptional data offers inventory of probable translated gene products; however, mass spectrometry (MS) is capable of providing evidence of protein translation. Prior to liquid chromatography (LC)-MS interrogations, a robust sample preparation scheme must be established for sample cleanup due to interferences such as nutritional growth media and environmental contaminants present as the microbes grow and are handled. This method will provide proper identification of the secreted material and offer opportunity for further studies. The proteome, especially the extracellular proteome, of thermophilic microorganisms has not been largely studied, thus the absence of effective sample processing methods. Devoid of careful and well-defined methods, successful proteomic analysis becomes dubious as this initial sample handling propagates throughout the experimental workflow. Several factors must be considered such as compatibility with downstream LC-MS, sample handling, efficiency, and number of confident positive protein identifications, among others. Development of proteomic methods and strategies for analysis of the extracellular proteome will likely traverse other thermophiles and microorganisms grown in similar conditions in the quest to engineer the most efficient and effective thermostable biocatalysts applicable for industrial and pharmaceutical objectives.

Here, in Part I, the extracellular protein fraction of *C. saccharolyticus* was purified by several different methods prior to one-dimensional sodium dodecyl sulfate-polyacrylamide gel electrophoresis (1D SDS-PAGE). Acetone precipitation, trichloroacetic acid (TCA) precipitation, and phenol extraction were first used as a cleanup step. Additional experiments employed other techniques such as drop dialysis, extraction with stationary phase beads, and filtration by molecular weight cut-off (MWCO) filters. Order of operation or pairing of two methods for

increased sample purification was also investigated to provide sufficient protein purification. To evaluate the sample preparation techniques, the GeLC-MS² method [14], 1D SDS-PAGE followed by nano-flow LC coupled to electrospray ionization hybrid linear ion trap Fourier transform-ion cyclotron resonance (ESI-LTQ-FT-ICR) MS², afforded protein identification and label-free relative quantification of the *C. saccharolyticus* secretome. Part II probes the intracellular proteome of two thermophiles in different environmental conditions supplementing the secretome information from Part I.

Complementing experimental secretome investigations, prediction tools provide evaluation of the identified secreted proteins and transmembrane domains. Mechanisms by which proteins are secreted or exported to different locations in and outside the cell most frequently involve signal peptides [9]. These amino-terminal signal peptides are short chains of mostly hydrophobic amino acids that are cleavable after translocation. Bacterial cells can secrete proteins completely outside of the organism into the growth milieu through a provided pathway. Influenced by the type of bacteria, several pathways exist for protein secretion. SignalP 3.0 aims to recognize of secretory (Sec) pathway signal peptides and cleavage sites of these peptides after secretion of the mature protein into the extracellular space [15, 16]. The Sec pathway is noted as one of the most frequent means of protein secretion in gram-positive bacteria; a preprotein (protein with signal peptide) is directed to the translocation machinery at the cell membrane by different routes, the addition of a molecular chaperone, through the signal recognition particle pathway attributable to hydrophobic signal peptide chains, or due to the association with the translocase, a multicomponent domain [9, 17]. The translocation of the preprotein is promoted by the multicomponent membrane complex, and through the course of secretion, the signal peptide is cleaved. Another secretion pathway occurring very similarly to the Sec pathway, the twin-arginine translocation (Tat), also involves a signal peptide; however, it contains two consecutive arginine residues and can be predicted with TatP 1.0 [18]. Non-classical protein secretion, secretion of a species with no evident signal peptide, has been identified in eukaryotes [19, 20] and in bacteria [21, 22], and the SecretomeP server aims to predict these species [23]. Overall, the prediction models may support the secretion of a particular protein; however, secretion is not necessarily guaranteed.

Further investigating predicted protein secretion, transmembrane domains should be absent from the secreted entity as this would most likely result in the protein anchored to the plasma membrane, and thus not entirely released into the extracellular space. The transmembrane hidden Markov model (TMHMM) tool predicts transmembrane helices and anticipates protein topology [24]. Those

proteins identified with a secreted mature protein engaged in a transmembrane domain will be considered false positives. Several groups describe a weakness in this approach as most signal peptide and transmembrane domain predictors often mistake one feature for the other due to similar amino acid patterns [15, 25, 26]. Criteria of both predictors include pursuing the recognition of a positively charged and hydrophobic region among other measures [16, 24, 25]. A combined prediction server, Phobius, provides a prediction scheme in which the optimal feature, signal peptide or transmembrane domain, is selected [25, 26]. However, SignalP correctly predicts a greater percentage of signal peptides versus Phobius, as well as generated less false positives on a data set deficient of transmembrane domain-containing species [15].

Experimental

Thermophilic bacterial growth and secretome concentration

C. saccharolyticus DSM 8903 (ATCC 43494) was grown anaerobically on a modified DSMZ 640 media (DSMZ, Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany) containing yeast extract proteins and peptides ≤ 1 kDa and with glucose as the carbon source at 75 °C, similar to that previously described [27]. One thousand six hundred-milliliter cultures were harvested at mid-exponential phase (cell density approximately 5×10^7 cells/mL). Samples were filtered thru a 0.22- μ m Durapore polyvinylidene fluoride hydrophilic membrane filter with 47 mm diameter (Fisher Scientific, Waltham, MA, USA) to separate the culture media or secretome sample from the bacterial cells. The secretome fraction was concentrated by ultrafiltration and underwent dialysis in an effort to remove contaminants reducing the 1,600-mL secretome sample to approximately 10 mL employing a Millipore stirred cell and polyethersulfone ultrafiltration membranes with a 10-kDa MWCO (Fisher Scientific).

Protein purification techniques

Total protein concentration was approximated by a modified Bradford Assay method, the Coomassie Plus Assay (Thermo Scientific, Rockford, IL, USA), and 50 μ g of the *C. saccharolyticus* secretome sample were aliquoted for purification method evaluation. The technique employed for acetone precipitation was described in a technical resource note by Pierce [28] and illustrated in 1 in Fig. 1a. Briefly, four volumes of cold (-20 °C) acetone (Fisher Scientific) were added to a 50 μ g protein sample. The solution was vortexed and then incubated at -20 °C for 90–120 min. Centrifugation for 30 min at 11,000 rpm

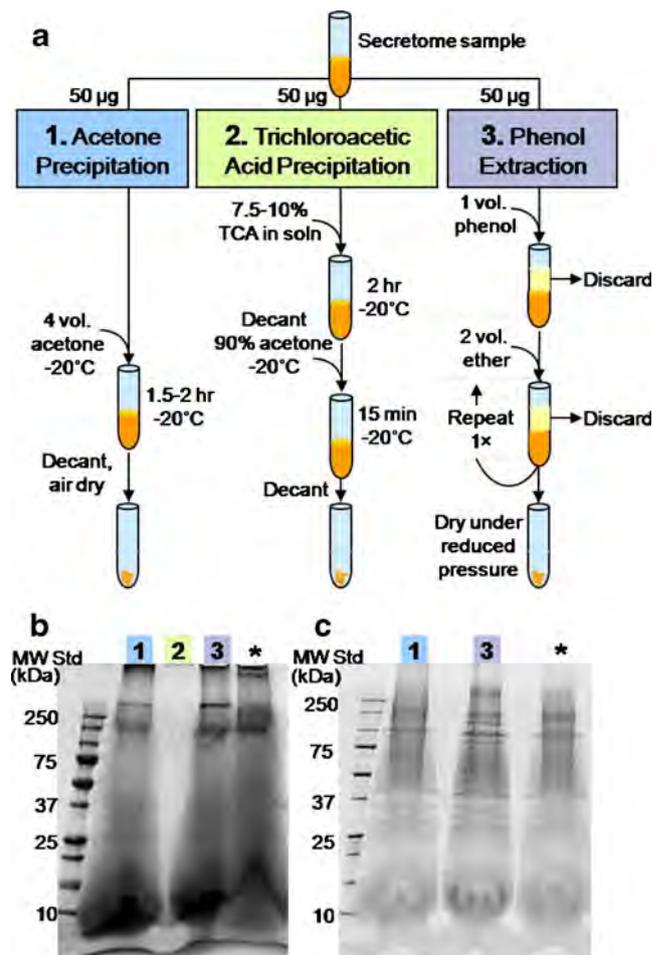


Fig. 1 Initial sample preparation employing three frequently used protein purification techniques and corresponding 1D gel images. **a** Schematic of acetone precipitation, trichloroacetic acid precipitation, and phenol extraction methods as performed on 50 μ g secretome samples. **b**, **c** 1D SDS-PAGE experiments of the purified protein samples visualizing with Bio-Safe Coomassie and Sypro Ruby, respectively. The numbered lanes correspond to the numbered purification methods. Asterisk represents the control sample in which the secretome aliquot was not processed prior to gel electrophoresis

allowed for decanting of the supernatant before air drying the protein pellet for 30 min. TCA precipitation was completed according to Jiang et al. [29] and is shown in 2 in Fig. 1a. TCA (Sigma-Aldrich, St. Louis, MO, USA) was added to an aliquot to a final concentration of 7.5–10%, and the solution was incubated for 2 h at -20 °C. This was centrifuged, and the supernatant was removed as in the acetone precipitation method. An aliquot (100 μ l) of 90% acetone (-20 °C) was used to wash the pellet during a 15-min incubation at -20 °C. Again, the solution was centrifuged, the supernatant was discarded, and the protein pellet was allowed to dry according to the procedure described above. Outlined by Sauv e et al. and illustrated in 3 in Fig. 1a, phenol extraction was completed by adding

1 vol. of liquefied phenol (Sigma-Aldrich) to a secretome aliquot; the solution was vortexed 20 s and centrifuged at 11,000 rpm for 5 min [30]. The upper aqueous phase was discarded, and 2 vol. of diethyl ether (Fisher Scientific) were added to the organic layer. After vortexing for 20 s, the solution was centrifuged as above, and the upper phase was discarded. This wash step was repeated, and the final lower phase dried under reduced pressure. The resulting dried protein pellets were stored at $-20\text{ }^{\circ}\text{C}$ before reconstituting sample with 20 mM Tris-HCl, pH 8.0 (Fisher) for analysis by GeLC-MS².

Drop dialysis 0.025- μm membrane filters (Millipore, Billerica, MA, USA) were moistened in a beaker of 100 mL high performance liquid chromatography (HPLC)-grade water (Burdick and Jackson, Muskegon, MI, USA) for 10 min prior to aspiration of a 50- μg secretome aliquot dried down to approximately 15 μl on the middle of the filter. After incubation (1, 3, and 6 h), the sample was collected and dried under reduced pressure (see 4 in Fig. 2a). For sample desalting using stationary phase beads, 5 in Fig. 2a outlines the method in which a methanolic slurry (100 μl) with 2.5–3 mg of 5 μm , 200 \AA silica Magic C₈ or C₁₈ AQ beads (Michrom BioResources, Auburn, CA, USA) was added to a 50- μg sample and vortexed for 5 min [31]. The mixture was centrifuged for 3 min at 14,000 rpm, 2 \times 20 μl of methanol were added, centrifugation was repeated, and the supernatant was discarded. The pellet was washed three times as follows: 1 mL of 0.1% trifluoroacetic acid (Sigma-Aldrich) was added, and the solution was vortexed for 1 min and centrifuged for 3 min as above. The supernatant was discarded. Lastly, for the MWCO filter technique (see 6 in Fig. 2a), a secretome aliquot was passed through a 10-kDa Microcon centrifugal filter (Sigma-Aldrich) for 20 min at 14,000 rpm. A 200- μl aliquot of 20 mM Tris-HCl pH 8.0 (Fisher Scientific) was used to rinse the retentate, concentrated sample on membrane, and was subjected to centrifugation for 10 min at 14,000 rpm; this rinsing was repeated twice. The ≥ 10 -kDa retentate was collected. For samples purified by a combination of two cleanup steps, reagent aliquots required for the second step were based upon the initial sample volume. Each desalted sample was stored at $-20\text{ }^{\circ}\text{C}$ prior to gel electrophoresis.

GeLC-MS²

Gel electrophoresis was performed with 10–20% Tris-HCl Criterion Gels (BioRad, Hercules, CA, USA) in 25 mM Tris-HCl, 192 mM glycine, and 0.1% SDS, pH 8.3 (BioRad). The resultant processed secretome samples were reconstituted in 20 mM Tris-HCl pH 8.0, loaded on the gel, and run at 200 V for 55 min. The gel was either stained with Bio-Safe Coomassie Blue G-250 (BioRad) for 60 min

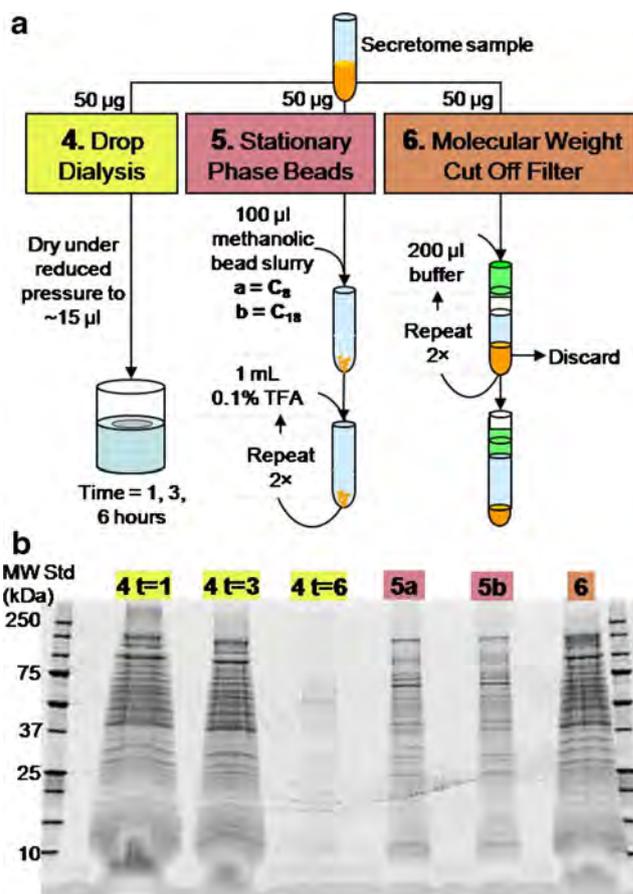


Fig. 2 Investigation of additional sample desalting and protein purification techniques evaluated by gel electrophoresis. **a** Schematic of drop dialysis, stationary phase beads, and molecular weight cut-off filter methods. **b** 1D SDS-PAGE experiment of secretome sample preparation methods. Sample loss occurred during drop dialysis over 6 h due to the drop exceeding the membrane area. The numbered lanes correspond to the numbered purification methods

or fixed, stained with Sypro Ruby (Invitrogen, Carlsbad, CA, USA) overnight, and rinsed according to the manufacturer's protocol. Each gel lane was cut into 32 bands, and eight bands were grouped for a total of four fractions per gel lane. The bands were excised using a disposable gridcutter (2 \times 7 mm lanes with 25 rows) from The Gelcompany (San Francisco, CA, USA) and further cut with a razor blade into 1 \times 2–3 mm pieces. An in-gel tryptic digestion was performed to the samples as adapted from Shevchenko et al. [32]. After dehydration with HPLC-grade acetonitrile (Burdick and Jackson), the samples were reduced with 10 mM dithiothreitol (BioRad) for 30 min at 56 $^{\circ}\text{C}$. This solution was removed, and acetonitrile was added to dehydrate the gel pieces. The acetonitrile was removed, and 55 mM iodoacetamide (Sigma-Aldrich) was added to alkylate for 20 min at room temperature in the dark. After the removal of the alkylation solution, the gel pieces were dehydrated with acetonitrile. The acetonitrile

was removed, and to digest the proteins, 1 ng/ μ l trypsin (Sigma-Aldrich) solution was added, and the samples were incubated overnight at 37 °C. The tryptic peptides were extracted by first incubating the gel pieces in 5% formic acid (Sigma-Aldrich):acetonitrile (1:2) at 37 °C and withdrawing the supernatant. The sample was dehydrated with acetonitrile, the supernatant was removed, and the gel pieces were hydrated with 100 mM ammonium bicarbonate:acetonitrile (2:1); these steps were repeated, and all extractions were pooled and dried under reduced pressure. Samples were stored at -20 °C prior to nanoLC-MS² interrogation.

NanoLC-ESI-LTQ-FT-ICR MS² interrogations were performed in triplicate with an Eksigent (Dublin, CA, USA) nanoLC-2D system in a continuous, vented column configuration as previously described [33], which was coupled to a hybrid LTQ-FT Ultra mass spectrometer (Thermo Scientific, Inc. San Jose, CA, USA) equipped with an actively shielded 7-Tesla superconducting magnet (Oxford Instruments, Concord, MA, USA). A 50- μ m thick stainless steel (SS) mesh screen and zero dead volume SS fitting from VICI Valco Instruments Co. Inc. (Houston, TX, USA) were included in the set-up between the transfer line and trap column. The trap, 100 μ m i.d. IntegraFrit capillary, and analytical, 75 μ m i.d. PicoTip capillary with 15 μ m i.d. tip, columns (New Objective, Woburn, MA, USA) were packed in-house with Magic C₁₈AQ stationary phase, 3.2 and 15 cm, respectively. Mobile phase solvents were purchased from Burdick and Jackson. Mobile phase A was water/acetonitrile (98/2), and mobile phase B was water/acetonitrile (2/98) each including 0.1% formic acid. Samples (approximately 1 μ g of protein) in 4 μ l were injected into the 10- μ l sample loop and then trapped and washed on the trap column by a 15- μ l metered injection consisting of 2% B from Channel 1 at 3 μ l/min.

At the conclusion of the metered injection, the switching valve positioned Channel 2, the nano-flow pumps, inline with the trap and analytical column where the analytical separations transpired with a flow-rate of 300 nl/min. Initially, the gradient elution profile was held at 2% B for 3 min, and then 10% B for 2 min preceded a linear gradient to 25% B over 40 min. The following 15 min ramped up to 50% B, then 95% B over 3 min and held for 2 min. Initial gradient conditions were reinstated linearly over 5 min and held 5 min for equilibration. The pulse sequence for the mass spectrometer consisted of seven events: a broadband survey scan 300–1,600 m/z in the ICR cell followed by data-dependent scans of the first through sixth most abundant ions having $\geq 2+$ charge state. The mass resolving power was set to 100,000_{fwhm} at m/z 400; the AGC limit was set at 1×10^6 for the ICR cell and 1×10^4 in the ion trap, dynamic exclusion restricted the selection of a m/z to twice within

30 s prior to its exclusion for 180 s. External calibration was performed as recommended by the manufacturer.

Data analysis

The bioinformatic platform included MASCOT Daemon, which integrates the analysis of RAW files by first processing through MASCOT Distiller, fitting experimental isotopic distributions with ideal isotopic distributions, followed by MASCOT (version 2.2.04, Matrix Science Ltd., London, UK) for database searching [34]. The protein database used for analysis was a compilation of the *C. saccharolyticus* UniProt protein sequence database and its reverse; the *C. saccharolyticus* protein sequences were reversed by a perl script and inserted into the FASTA file. In addition, the *Bos taurus* trypsin protein sequence (1 protein), the *Homo sapiens* keratin and keratin-related protein sequences (177 proteins; see Electronic Supplementary Material Table S1), and the *Saccharomyces cerevisiae* were incorporated into the database file for a total of 12,061 entries. The search parameters include cysteine carbamidomethylation, as a fixed modification attributable to protein alkylation with iodoacetamide, and variable modifications of methionine oxidation and asparagine and glutamine deamidation. A maximum of two missed trypsin cleavages were allowed, and tolerance measurements of ± 5 ppm for peptide mass and ± 0.6 Da for tandem MS measurements were set. One ProteoIQ (version 1.2.01, BIOINQUIRE, www.bioinquire.com) project was created merging data from each gel lane fraction and replicates, and offered comparative and statistical analysis. The proteins identified were arranged according to total protein score, and proteins were determined to be statistically significant maintaining a <1% false discovery rate (FDR) [35, 36]. Of those proteins having <1% FDR and identified with a single peptide, manual MS² validation was performed. Total spectral counts (SpC) and normalized SpC were extracted from the created ProteoIQ project.

The statistically significant proteins were exported from ProteoIQ into a FASTA file format for evaluation by SignalP 3.0, TMHMM 2.0, SecretomeP 2.0, TatP 1.0 (all available from CBS Prediction Server at <http://www.cbs.dtu.dk/services/>), and Phobius (available from Stockholm Bioinformatics Centre at <http://phobius.sbc.su.se/>). The FASTA formatted file was submitted to each prediction server for analysis, and gram-positive bacteria was chosen for those searches requiring selection. For evaluation by SignalP 3.0, the following parameters were selected: both neural networks and hidden Markov models employed for signal peptide prediction, standard output, and each sequence were truncated to the first 70 residues. The file was submitted to the TatP 1.0 server with the default settings for regular expression examining protein sequences

for the double arginine feature with a standard output and truncation of 100 residues.

Results and discussion

Evaluation of protein purification techniques and order of operation investigations

A fundamental importance exists in the production of a robust and reproducible sample preparation. Once developed, this method may expand to the analysis of extracellular proteomes of other thermophilic microorganisms leading to the investigation of thermostable biocatalysts for a variety of applications (e.g., biomass degradation and therapeutics requiring antimicrobial species). Because of the nature of the thermophilic secretome, containing many small molecules and other nutritional elements in the growth media (see Electronic Supplementary Material Table S2), sample purification was essential for proteomic evaluation. Furthermore, sample fractionation of the extracellular proteome prior to LC-MS analysis yields decreasing sample complexity anticipating increase of confident protein identifications.

Techniques for sample cleanup rely on different principles of purification. Protein precipitation results from lowering the analyte solubility, by modifying the solvent composition via addition of reagent. With this alteration, proteins are precipitated out of solution removing impurities that remain in the sample medium. Acetone precipitation, first mentioned by Zakowski in 1931 for the precipitation of a soybean protein [37], is still widely applied today on samples ranging from human urinary [38] to fungal proteins [39]. A second protein precipitation technique, TCA precipitation, was first described in 1930 by Sahyn and Alsber to prepare rabbit liver glycogen [40]. Again, this method is still currently used for protein precipitation on a variety of samples such as plasma proteins [29] and secreted proteins from myeloid cells [41]. Exploiting the physical properties of proteins afford an alternative sample cleanup of complex biological samples in contrast to manipulating the solvent composition. Extraction with liquefied phenol purifies samples by means of hydrophobicity and was initially used for antigenic bacterial carbohydrate complexes in 1940 [42]. Phenol extraction allows for proteins to remain in solution when purified. Addition of liquefied phenol to a protein sample and rapid vortexing denature proteins and “flips” the hydrophobicity. The proteins are compelled to exchange solution from aqueous to organic.

Preliminary experiments purifying the *C. saccharolyticus* secretome by acetone precipitation, TCA precipitation, and phenol extraction (see Fig. 1a) proved to be inadequate

as illustrated in Fig. 1b. This initial gel suggested that the Coomassie stain may not be sensitive enough for the secretome analysis, and precipitation by TCA was not effective. Subsequent experiments employed acetone precipitation and phenol extraction to purify the sample, and Sypro Ruby stain was used for protein visualization. Figure 1c illustrates increased band resolution with Sypro Ruby imparting greater sensitivity. Owing to the presence of vertical and horizontal diffusion, these experiments demonstrate that an interfering species or contaminant remains after an initial purification step.

It was necessary to incorporate a second level of secretome sample cleanup prior to GeLC-MS² experiments. Three techniques (see Fig. 2a) resulted in a diverse range of sample desalting as demonstrated in Fig. 2b. Employing a commercially available membrane or membrane device (i.e., centrifugal filter), dialysis and filtration can be performed fractionating the sample due to its physical properties. The separation of molecules in a sample as a result of differing diffusion rates through a semipermeable membrane, dialysis removes small molecules and salts. The drop dialysis technique is formatted for small volumes ($\leq 100 \mu\text{l}$) and is a simple and rapid method used for preparing samples for a variety of analysis (e.g., gel electrophoresis, isoelectric focusing, and electron microscopy) [43]. Drop dialysis reduced sample interferences over time; however, sample loss occurred between 3 and 6 h due to the limited membrane area as the droplet increased in size over time. Filtering with a 10-kDa MWCO centrifugal filter, often used for sample desalting, did not effectively eliminate the horizontal band diffusion as demonstrated (see Fig. 2b lane 6). MWCO centrifugal filters fractionate species according to the radius of gyration [44, 45]. With a 10-kDa MWCO filter, small interfering species and salts approximately 10 kDa and less will flow thru the membrane, while in principle, the retentate contains a desalted sample.

An alternative desalting technique not widely referenced, the addition of stationary phase beads directly into the sample of interest desalts as a function of protein/peptide interaction with the beads. As in reverse phase chromatography, the degree of hydrophobicity and the mobile phase induces partitioning of the analyte between the stationary and mobile phases. Prior to denaturing or linearizing a protein/peptide, the hydrophilic surface encapsulates hydrophobic moieties in aqueous solution. Integrating the methanolic slurry of C₁₈ stationary phase beads with a secretome aliquot and rapidly vortexing provides some denaturing of proteins and the analyte–bead interaction to occur. The beads have affinity towards the analyte, while the salts, small molecules, and impurities are washed from the sample. Although sample loss may occur due to extremely hydrophilic proteins or peptides as well as those not sufficiently denatured for interaction with the beads, the

gel image (Fig. 2b lanes 5a and 5b) reveals that visually, the bead technique is an improvement over more commonly used desalting techniques. Horizontal diffusion is almost completely eliminated employing the bead technique, particularly for lower molecular weight species which are potential small opening reading frames or small unknown opening reading frames and are of significant interest in future studies of *C. saccharolyticus* and other thermophiles [46]. For subsequent sample processing, Magic C₁₈AQ beads were utilized as it was anticipated that the longer chains would better retain smaller proteins.

In the case that one sample processing method, acetone precipitation or phenol extraction, improved purification in conjunction with using the Magic C₁₈AQ beads, order of operation experiments were performed. Figure 3 demonstrates that order does influence the 1D SDS-PAGE experiment visually. NanoLC-ESI-LTQ-FT-ICR MS² experiments of the in-gel digestion [32] affords evaluation of the secreted proteins and demonstrates the most effective secretome sample preparation of the techniques investigated in this study (see following section for clarification). Visual differences between the C₁₈ bead processed *C. saccharolyticus* secretome samples in Figs. 2b and 3 may be due to a variety of factors such as different *C. saccharolyticus* sample batches and inherently different volumes of secretome sample required to process approximately 50 µg of sample. Also, alterations to imaging parameters for sufficient visualization of sample lanes may attribute to some of the visual variations. Nevertheless, the order of operation experiments was performed on equal aliquots of the same *C. saccharolyticus* secretome sample providing relative comparison of the protein purification techniques.

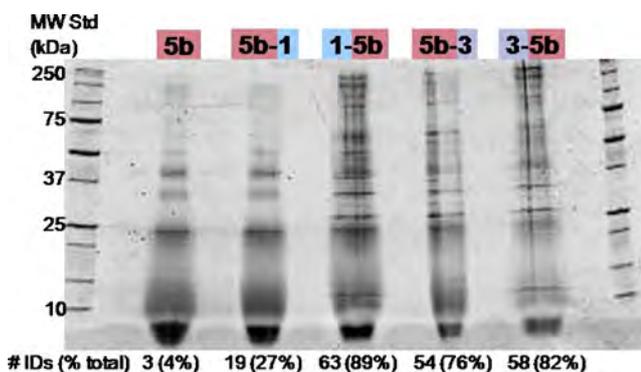


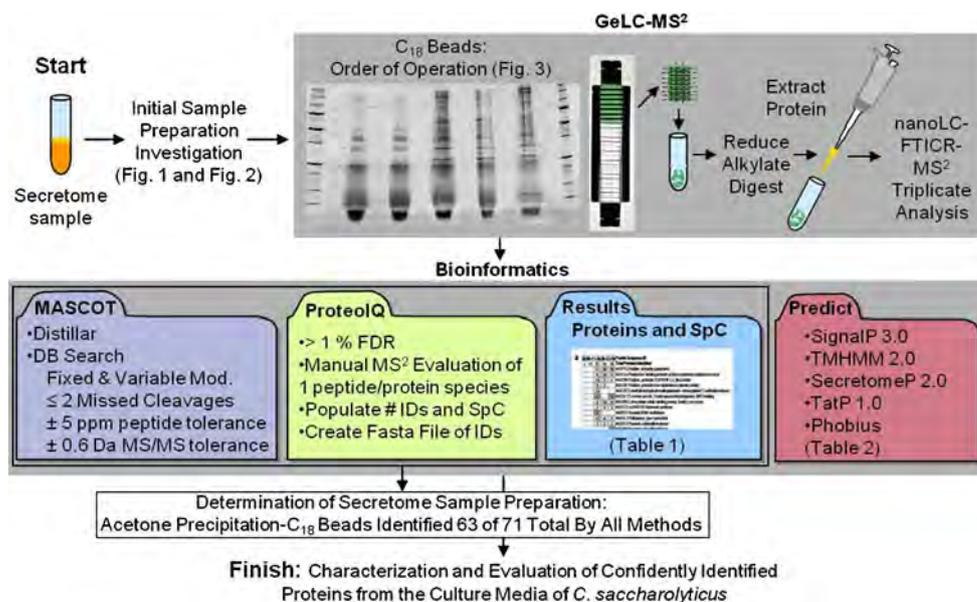
Fig. 3 Order of operation of sample processing techniques for optimal 1D SDS-PAGE experiment. The numbered lanes correspond to combinations of the described purification methods. The numbers below the gel indicate the number of proteins identified by each method as well as the corresponding percentage based on 71 uniquely identified proteins from the *Caldicellulosiruptor saccharolyticus* secretome sample

Protein identification attributed to sample preparation

Interrogations by nanoLC-ESI-LTQ-FT-ICR MS and evaluation by the MASCOT-ProteoIQ bioinformatic platform (see Fig. 4 for workflow) confidently identified 71 proteins in the *C. saccharolyticus* secretome sample from the order of operation experiments with <1% FDR (see Table 1). Fifteen proteins were identified with one peptide (Table 1, dark shaded boxes); however, after inspecting the MS² spectra and considering the protein probability value, one identified protein was eliminated from the protein inventory. Fourteen proteins were identified as putative uncharacterized entities, and the Protein Basic Local Alignment Search Tool 2.2.23 (Blastp) and Superfamily database of structural and functional protein annotation [47] were employed to determine possible functions listed in Table 1 following classification of putative. The number of protein identifications varies as a function of the purification method. Here, it is apparent that the C₁₈ bead method in combination with another protein cleanup technique provides the most advantageous outcome in terms of number of proteins identified in this study; those techniques incorporating two methods resulted in 27–89% of the total identified proteins, whereas the C₁₈ technique (5b) alone resulted in 4% or three proteins. Three techniques represented the majority (76–89%) of the proteins identified, acetone precipitation–C₁₈ beads (1-5b) and both combinations of phenol extraction and C₁₈ beads (5b-3, and 3-5b), resulting in nearly identical proteins despite differences in purification principles. Table 1 illustrates these similarities as well as includes a measurement of relative quantification suggesting explanation of protein recovery and most efficient sample preparation for *C. saccharolyticus* extracellular proteins.

Spectral counting relatively quantifies proteins by the number of MS² spectra from peptides corresponding to a particular protein [48–52]. This assessment of SpC is inherent in the collected data and is merited the most straightforward label-free quantification method. The overall total SpC for each order of operation experiment provides a reflection of total sample recovery and range from ten to 739 overall total SpC. Species identified with increased numbers of SpC, suggesting highly abundant proteins, were more likely to be identified by four to five order of operation experiments. Normalized SpC [53] (see Part II of this investigation for an in-depth discussion) were also computed in order for corroboration (data not shown for all identified proteins). For example, S-layer domain protein precursor (A4XM93) was identified by all five techniques, and the number of total SpC per function of purification method varies greatly from seven by the C₁₈ bead method (5b) to 63 SpC by the phenol extraction–C₁₈ bead method (3-5b) and is described in Table 2. This

Fig. 4 Experimental workflow from start to finish of secretome sample preparation and evaluations. Order of operation investigations analyzed by GeLC-MS² were assessed through a bioinformatic platform consisting of Mascot, ProteoIQ, and prediction servers. The sample preparation technique identifying the most proteins in conclusion of database searching (*black outlined box*) was suggested for use in future studies. Complementing the protein identification and spectral counts, prediction servers offered insight into the probable biological location of each protein



suggests that the S-layer domain protein precursor is in high abundance in the initial unprocessed sample and is sufficiently recovered by all order of operation methods except for C₁₈ beads alone (5b). A contrasting example of a lesser abundant species, as described by fewer overall SpC in Table 2 and comparison of normalized SpC with the S-layer domain protein, a putative, probable PKD domain-containing protein (A4XLH0) was identified by each combination method, and only one SpC was demonstrated by the C₁₈ bead–acetone precipitation method (5b-1), whereas the other methods resulted in six to 12 SpC. This pattern of total SpC corresponding to protein abundance and sample recovery is present in most of the data, although a few exceptions exist. It is evident that the three-combination order of operation experiments, acetone precipitation–C₁₈ beads (1-5b), phenol extraction–C₁₈ beads (3-5b), and C₁₈ beads–phenol extraction (5b-3), are nearly equivalent in performance identifying 63, 58, and 54 proteins, respectively, and providing 739, 737, and 618 overall total SpC, respectively.

Interestingly, the second step of the top three order of operation techniques concludes with a purification method reliant on hydrophobic interaction. This is presumably a factor in the number of homologous protein identifications. It can also be proposed that highly abundant proteins are better recovered following the sample purification methods. The flux in SpC of the homologous proteins can be an indicator of sample recovery as opposed to types of proteins identified by different techniques. For instance, the combination technique C₁₈ beads–acetone precipitation (5b-1) contributes 129 overall total SpC identifying few proteins. In general, these identifications had fewer SpC as compared to the top three-combination techniques. Further

probing this method (5b-1), a reason for so many fewer protein identifications may result from difficulty reconstituting the purified protein following acetone precipitation. The tertiary structure of proteins may be altered as the aqueous solution surrounding the analyte is replaced with acetone. In addition, drying down the sample to remove solvent, the structure is further vulnerable to modification and leads to reconstitution difficulties, and thus sample loss.

Biological importance and evidence for secretion

Noteworthy, *C. saccharolyticus* has the most active and stable cellulolytic enzymes, cellulases, in the genus *Clostridium*, which contributes to approximately 20% of open reading frames of known function encoding carbohydrate-degrading-type proteins/peptides [2, 13]. Accordingly, carbohydrate-degrading species were identified in the *C. saccharolyticus* secretome sample (see Table 1), for example, cellulase (A4XIF7 and A4XIF8), glycoside hydrolases (A4XIF5 and A4XIF6), and pullulanase (A4XHC3), as well as 14 putative uncharacterized proteins. The variety of enzymes, such as those mentioned, evidence the potential benefit of an individual thermophilic microorganism targeting complex carbohydrates for degradation and the production of bioenergy in a single step, or consolidated bioprocessing in bioenergy production, a process in which substrate utilization and product formation are combined [54, 55].

Almost all proteins resulting in prediction for secretion and/or transmembrane domains (28 of 30 total proteins) fell within the species identified by the acetone precipitation–C₁₈ beads technique (1-5b) and the phenol extraction–C₁₈

Table 1 Complete list of proteins identified in the study arranged alphabetically by sequence ID

5b	5b-1	1-5b	5b-3	3-5b	Protein Sequence ID
3	19	63	54	58	Total Proteins Identified
		17	50	28	A4XFY3 Putative, probable polyprotein
		6	29	17	A4XG54 Periplasmic binding protein/LacI transcriptional regulator precursor
		10	2	1	A4XG90 Putative, probable ISOPREN_C2_like protein
		9	2		A4XG91 Putative, probable transglutaminase domain protein
		14			A4XGB3 5-methyltetrahydropteroyltryptophan-homocysteine S-methyltransferase
		1		3	A4XGF7 D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding
		5	24	15	A4XGN5 Extracellular solute-binding protein, family 1 precursor
		1	3	3	A4XGZ2 or A4XH19 Glutamate synthase
		2			A4XHB1 Aspartyl-tRNA synthetase
		13	4	7	A4XHC3 Pullulanase, type I precursor
		3	4	1	A4XHF8 Pyruvate carboxyltransferase
				1	A4XH0 Pyruvate carboxyltransferase
		10	6	7	A4XH4 Aconitase
		1	1	2	A4XHL2 Transcriptional regulator / glucokinase
1	52	23	41	44	A4XHT4 Putative, probable 2-oxoacid ferredoxin/flavodoxin oxidoreductase protein
			1		A4XHV9 TIG Trigger factor
		2			A4XD6 Ser-tRNA(Thr) hydrolase / threonyl-tRNA synthetase
		39	2	16	A4XB6 EFG Elongation factor G
		21	22	51	A4XB7 EFTU Elongation factor Tu
		4	1	1	A4XM8 Aldehyde ferredoxin oxidoreductase
		5	3	2	A4XIF2 Putative, probably acriflavin resistance protein B
		8	8	6	A4XIF3 Putative, probable toxin complex protein
4	79	51	74		A4XIF5 Glycoside hydrolase, family 48 precursor
2	14	14	19		A4XIF6 Glycoside hydrolase, family 5 precursor
2	12	9	10		A4XIF7 Cellulose 1,4-beta-cellobiosidase precursor
2	35	17	20		A4XIF8 Cellulase, Cellulose 1,4-beta-cellobiosidase precursor
		2			A4XIG9 Cellulose phosphorylase
		4	3		A4XIL8 2-isopropylmalate synthase
		2	2		A4XIL9 2-isopropylmalate synthase
				3	A4XIQ5 Mannose-6-phosphate isomerase / glucose-6-phosphate isomerase
1	1	6	7		A4XIQ7 Fructose-bisphosphate aldolase
2	2	1	1		A4XJ08 CH10 10 kDa chaperonin
		44	25	28	A4XJ09 CH60 60 kDa chaperonin
		2		3	A4XJ20 L-aspartate aminotransferase / phosphoserine aminotransferase
2	23	4	16		A4XJ21 D-3-phosphoglycerate dehydrogenase
3	1	7	5		A4XJ27 Ig domain protein, group 2 domain protein precursor
		2	2		A4XJH3 Pyruvate ferredoxin oxidoreductase, gamma subunit
		2	5	6	A4XJH5 Pyruvate ferredoxin oxidoreductase, alpha subunit
		4	1	3	A4XJH6 Thiamine pyrophosphate enzyme domain protein TPP-binding
2		5	11	6	A4XJ8 Putative, probable zinc finger associated protein
		12	1	4	A4XJP0 Cell wall hydrolase/autolysin precursor
2		1	2	1	A4XJW3 Acyl carrier protein
		1	3	4	A4XJW4 3-oxoacyl-[acyl-carrier-protein] synthase II
		9			A4XJY3 Putative, probable phosphoenolpyruvate carboxykinase
		2	5		A4XK0 6-phosphofructokinase
		11	3	1	A4XKL3 NADH dehydrogenase
		10	11	6	A4XKL4 Hydrogenase, Fe-only
		2	1	1	A4XKR2 SY1 Isoleucyl-tRNA synthetase
2		11	4	14	A4XKR7 Putative, probable peptidase M56
		8		20	A4XKV0 ENO Enolase
		19		5	A4XKV2 Phosphoglycerate kinase / triosephosphate isomerase
		4	4	11	A4XKV3 Glyceraldehyde-3-phosphate dehydrogenase
3		87	67	62	A4XKV5 Pyruvate phosphate dikinase
			3	1	A4XKX0 ATP/ATP synthase subunit beta
			3	1	A4XL01 2-alkenal reductase
				1	A4XL36 TAL Probable transaldolase
				5	A4XL41 Phosphotransacetylase
		1	5	2	A4XL61 Serine hydroxymethyltransferase
1		6	12	12	A4XLH0 Putative, probable PKD domain-containing protein
		12	15	23	A4XLH1 Putative, probable polyamine oxidase
		4	5	13	A4XL2 Putative, probable polyamine oxidase
1					A4XLL5 Putative, conservative domain with unknown function
			2	3	A4XM01 EFTS Elongation factor Ts
		1		3	A4XM28 Putative, probable membrane protein
		7			A4XM70 Alpha amylase, catalytic region
5		3	2	1	A4XM72 Rubrerythrin
1		6	7	2	A4XM87 S-layer domain protein precursor
7	40	58	55	63	A4XM93 S-layer domain protein precursor
3		6	17	8	A4XMD5 Extracellular solute-binding protein, family 1 precursor
1		28	29	62	A4XME8 D-xylose ABC transporter, periplasmic substrate-binding protein precursor
		2	3		A4XMZ6 Ferritin, Dps family protein
10	129	739	618	737	Overall Total Spectral Counts

A shaded box indicates the identification of the particular protein by the described method, and the number within the box equals the number of total SpC for that particular identification. The darker shaded boxes indicate the protein was identified with one tryptic peptide, however still maintaining <1% FDR

Table 2 Comparison of SpC and normalized SpC of two proteins from the complete list

Protein ID	MW, kDa	5b	5b-1	1-5b	5b-3	3-5b
S-layer domain (A4XM93)	108	2 (1)	36 (0.918)	53 (0.733)	48 (0.515)	55 (0.516)
Putative uncharacterized (A4XLH0)	36	0 (0)	1 (0.082)	6 (0.267)	14 (0.485)	16 (0.484)

SpC is indicated first with normalized SpC in *parenthesis*

bead technique (3-5b) and are described in Table 3. Of all identified proteins, eight and 13 from 1-5b and 3-5b, respectively, were not identified by the said techniques, but did not result in output when evaluated by the various prediction tools. The combination technique involving the C₁₈ beads and phenol extraction (5b-3) identified two fewer proteins with sample preparation taking roughly half the time (~100 min) compared those methods involving acetone precipitation. Based upon the number of proteins identified, as well as prediction evaluation of the proteins, the acetone precipitation–C₁₈ beads method (1-5b) for secretome sample preparation offers the greatest benefit, with the combinations of C₁₈ beads and phenol extraction (5b-3 and 3-5b) following close behind.

Of the 19 proteins predicted for secretion, 17 were unanimously predicted to have signal peptides by Phobius; however, mixed results were generated with SignalP and TMHMM (see Table 4). Two enzymes were predicted for translocation to the extracellular milieu by the Tat pathway, thiamine pyrophosphate enzyme domain protein TPP-binding and pyruvate ferredoxin oxidoreductase gamma subunit, although the twin-arginine motif was not explicitly found in the later. Nine proteins were predicted to contain transmembrane domains with Phobius, while TMHMM suggested 24 proteins had this moiety. TMHMM reinforced most of the predictions proposed by Phobius but appeared to excessively suggest the presence of transmembrane domains when signal peptide recognition was the superior choice. In one instance, a putative uncharacterized protein precursor (A4XG91) was predicted by Phobius to contain both a signal peptide and several transmembrane domains; the same predictions were generated by SignalP and TMHMM, and thus no precise classification can be determined.

Conclusions

The investigations in Part I of this study provided molecular level proteomic insight into the surroundings of *C. saccharolyticus* and the material released into the extracellular space or culture medium. To adequately study the secretome, several preparation techniques were performed prior to nanoLC-MS² and the bioinformatic platform as illustrated in Fig. 4. Although requiring roughly twice as much time as other techniques (~200 min), the acetone precipitation–C₁₈ beads method (1-5b) gave the most protein identifications of the methods evaluated. This was followed by combinations of phenol extraction and C₁₈ beads techniques (3-5b and 5b-3) resulting in 7–12% fewer identifications and requiring less time (~100 min). To ensure the most thorough secretome analysis, it is essential to employ a sample preparation technique affording the greatest sample recovery and, consequently, many protein identifications. It was evidenced that the species identified by four to five purification methods had greater numbers of SpC, suggesting greater protein concentration in the unprocessed secretome sample as well as greater sample recovery from purification.

Approximately 19 of the 71 total proteins identified in the secretome of the extreme thermophile *C. saccharolyticus* were predicted to feature moieties highly suggesting and complementing their presence in the secretome. While the secretome should encompass only those species released out into the extracellular space, over 70% of proteins identified were deficient of a moiety for translocation as determined by the publicly available prediction servers (i.e., Phobius, SignalP, TMHMM, SecretomeP, and TatP). Several factors may cause this behavior. The particular prediction servers employed for protein evalu-

Table 3 Total proteins predicted for secretion and containing transmembrane domains as a function of preparation method

Technique	5b	5b-1	1-5b	5b-3	3-5b
Predicted secreted	2	9	19	17	19
Predicted transmembrane domain	1	3	9	9	9
Total	3	12	28	26	28

Summary of identified *Caldicellulosiruptor saccharolyticus* secretome proteins with output from prediction server evaluation

Table 4 List of proteins predicted to contain a signal peptide or transmembrane domain moiety

Protein sequence ID	Phobius signal peptide	Phobius trans. domain	SignalP	TMGMM	SecretomeP	TatP
Predicted secreted protein						
A4XIF5 glycoside hydrolase, family 48 precursor	Yes		Yes	1	Yes	
A4XIF7 cellulose 1,4-beta-cellobiosidase precursor	Yes		Yes	1	Yes	
A4XIF8 cellulase, cellulose 1,4-beta-cellobiosidase precursor	Yes		Yes		Yes	
A4XHC3 pullulanase, type 1 precursor	Yes		Yes	1	Yes	
A4XG90 putative, probable ISOPREN_C2_like protein	Yes		Yes	1	Yes	
A4XHT4 putative, probable 2-oxoacid:ferredoxin/flavodoxin oxidoreductase protein	Yes		Yes	1	Yes	
A4XLH1 putative, probable polyamine oxidase	Yes		Yes	1	Yes	
A4XLI2 putative, probable polyamine oxidase	Yes		Yes	1	Yes	
A4XM28 putative, probable membrane protein	Yes		Yes	1	Yes	
A4XKR7 putative, probable peptidase M56	Yes		Yes			
A4XM87 S-layer domain protein precursor	Yes		Yes	1	Yes	
A4XM93 S-layer domain protein precursor	Yes		Yes	1	Yes	
A4XGN5 extracellular solute-binding protein, family 1 precursor	Yes		Yes	1	Yes	
A4XMD5 extracellular solute-binding protein, family 1 precursor	Yes		Yes	1		
A4XJP0 cell wall hydrolase autolysin precursor	Yes		Yes	1	Yes	
A4XG54 periplasmic binding protein Lacl transcriptional regulator precursor	Yes		Yes	1	Yes	
A4XJ27 Ig domain protein, group 2 domain protein precursor	Yes		Yes	1	Yes	
A4XJH6 Thiamine pyrophosphate enzyme domain protein TPP-binding			Yes			Yes
A4XJH3 pyruvate ferredoxin oxidoreductase, gamma subunit ^a			Yes			Yes
Predicted anchored to transmembrane domain						
A4XIF6 glycoside hydrolase, family 5 precursor		1	Yes	1	Yes	
A4XJW4 3-oxoacyl-acyl-carrier-protein synthase II		1				
A4XME8 D-xylose ABC transporter, periplasmic substrate-binding protein precursor		1	Yes	1		
A4XL01 2-alkenal reductase		1		1	Yes	
A4XFY3 putative, probable polyprotein		1		1	Yes	
A4XIF2 putative, probably acriflavine resistance protein B		1	Yes	1	Yes	
A4XIF3 putative, probable toxin complex protein		1	Yes	1	Yes	
A4XJ18 putative, probable zinc figure-associated protein		1	Yes	1	Yes	
A4XJY3 putative, probable phosphoenolpyruvate carboxykinase		1				
A4XLH0 putative, probable PKD domain-containing protein		1		1	Yes	
A4XG91 putative, probable transglutaminase domain protein	Yes	3	Yes	4		

Proteins arranged by prediction features followed by protein type and sequence ID

^a Predicted as potential Tat signal peptide but motif absent

ation only detect three routes of translocation while several more exist. Also, the servers are trained on limited data quantities, due to limited experimental data, and may cause deficiencies and possible inaccuracies in prediction. Regarding the biological sample, cell lysis would expose intracellular material to the extracellular milieu and, consequently, their detection. Probability also exists that

detected peptides may be cleaved from proteins anchored to the plasma membrane. Overall, the sample preparation method provides satisfactory performance in defining the extracellular proteome of the thermophilic bacterial microorganism *C. saccharolyticus* with promise to traverse other species to continue examination of microbial applications and environments.

Acknowledgements We would like to acknowledge the financial support of the National Institutes of Health (Grant 5T32GM00-8776-08), which supports GLA in the North Carolina State University Molecular Biotechnology Training Program, and the W. M. Keck Foundation. RMK acknowledges support from the US National Science Foundation (CBT0617272) and the Bioenergy Science Center (BESC), a US DOE Bioenergy Research Center supported by the Office of Biological and Environmental Research. DLL acknowledges support from a US Department of Education GAANN Fellowship.

References

- Antranikian G, Egorova K (2007) Extremophiles, a unique resource of biocatalysts for industrial biotechnology. In: Gerday C, Glansdorff N (eds) Physiology and biochemistry of extremophiles. ASM Press, Washington, pp 361–406
- VanFossen AL, Lewis DL, Nichols JD, Kelly RM (2008) Polysaccharide degradation and synthesis by extremely thermophilic anaerobes. *Ann NY Acad Sci* 1125:322–337
- Vafiadi C, Topakas E, Biely P, Christakopoulos P (2009) Purification, characterization and mass spectrometric sequencing of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*. *FEMS Microbiol Lett* 296:178–184
- Barabote RD, Xie G, Leu DH, Normand P, Necsulea A, Daubin V, Medigue C, Adney WS, Xu XC, Lapidus A, Parales RE, Detter C, Pujic P, Bruce D, Lavire C, Challacombe JF, Brettin TS, Berry AM (2009) Complete genome of the cellulolytic thermophile *Acidothermus cellulolyticus* 11B provides insights into its ecophysiological and evolutionary adaptations. *Genome Res* 19:1033–1043
- Asada Y, Endo S, Inoue Y, Mamiya H, Hara A, Kunishima N, Matsunaga T (2009) Biochemical and structural characterization of a short-chain dehydrogenase/reductase of *Thermus thermophilus* HB8 A hyperthermostable aldose-1-dehydrogenase with broad substrate specificity. *Chem Biol Interact* 178:117–126
- Muhammad SA, Ahmad S, Hameed A (2009) Antibiotic production by *Thermophilic bacillus* specie Sat-4. *Pak J Pharm Sci* 22:339–345
- Esikova TZ, Temirov YV, Sokolov SL, Alakhov YB (2002) Secondary antimicrobial metabolites produced by thermophilic *Bacillus* spp. strains VK2 and VK21. *Appl Biochem Microbiol* 38:226–231
- Sato S, Hutchison CA, Harris JI (1977) Thermostable sequence-specific endonuclease from *Thermus aquaticus*. *PNAS* 74:542–546
- Tjalsma H, Bolhuis A, Jongbloed JDH, Bron S, van Dijk JM (2000) Signal peptide-dependent protein transport in *Bacillus subtilis*: a genome-based survey of the secretome. *Microbiol Mol Biol Rev* 64:515–547
- Zhou M, Boekhorst J, Francke C, Siezen RJ (2008) LocateP: genome-scale subcellular-location predictor for bacterial proteins. *BMC Bioinform* 9:173
- Hathout Y (2007) Approaches to the study of the cell secretome. *Expert Rev Proteomics* 4:239–248
- Rainey FA, Donnison AM, Janssen PH, Saul D, Rodrigo A, Bergquist PL, Daniel RM, Stackebrandt E, Morgan HW (1994) Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium. *FEMS Microbiol Lett* 120:263–266
- Blumer-Schuetz SE, Kataeva I, Westpheling J, Adams MWW, Kelly RM (2008) Extremely thermophilic microorganisms for biomass conversion: status and prospects. *Curr Opin Biotechnol* 19:210–217
- Rezaul K, Wu LF, Mayya V, Hwang SI, Han D (2005) A systematic characterization of mitochondrial proteome from human T leukemia cells. *Mol Cell Proteomics* 4:169–181
- Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 340:783–795
- Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng* 10:1–6
- van der Does C, Nouwen N, Driessen AJM (2003) The Sec translocase. In: Oudega B (ed) Protein secretion pathways in bacteria. Kluwer Academic Publishers, Boston, p 296
- Bendtsen JD, Nielsen H, Widdick D, Palmer T, Brunak S (2005) Prediction of twin-arginine signal peptides. *BMC Bioinform* 6:167
- Muesch A, Hartmann E, Rohde K, Rubartelli A, Sitia R, Rapoport TA (1990) A novel pathway for secretory proteins. *Trends Biochem Sci* 15:86–88
- Rubartelli A, Cozzolino F, Talio M, Sitia R (1990) A novel secretory pathway for interleukin-1-beta, a protein lacking a signal sequence. *EMBO J* 9:1503–1510
- Harth G, Clemens DL, Horwitz MA (1994) Glutamine synthetase of *Mycobacterium tuberculosis*: extracellular release and characterization of its enzymatic activity. *PNAS* 91:9342–9346
- Harth G, Horwitz MA (1997) Expression and efficient export of enzymatically active *Mycobacterium tuberculosis* glutamine synthetase in *Mycobacterium smegmatis* and evidence that the information for export is contained within the protein. *J Biol Chem* 272:22728–22735
- Bendtsen JD, Kiemer L, Fausboll A, Brunak S (2005) Non-classical protein secretion in bacteria. *BMC Microbiol* 5:58
- Krogh A, Larsson B, von Heijne G, Sonnhammer ELL (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 305:567–580
- Kall L, Krogh A, Sonnhammer ELL (2004) A combined transmembrane topology and signal peptide prediction method. *J Mol Biol* 338:1027–1036
- Kall L, Krogh A, Sonnhammer ELL (2007) Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* 35:W429–W432
- van de Werken HJG, Verhaart MRA, VanFossen AL, Willquist K, Lewis DL, Nichols JD, Goorissen HP, Mongodin EF, Nelson KE, van Niel EWJ, Stams AJM, Ward DE, de Vos WM, van der Oost J, Kelly RM, Kengen SWM (2008) Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. *Appl Environ Microbiol* 74:6720–6729
- Pierce Technical Resource (2004) TR00490. Pierce Technical Resource, Rockford, IL, USA
- Jiang L, He L, Fountoulakis M (2004) Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. *J Chromatogr A* 1023:317–320
- Sauve DM, Ho DT, Roberge M (1995) Concentration of dilute protein for gel-electrophoresis. *Anal Biochem* 226:382–383
- Winston RL, Fitzgerald MC (1998) Concentration and desalting of protein samples for mass spectrometry analysis. *Anal Biochem* 262:83–85
- Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* 1:2856–2860
- Andrews GL, Shuford CM, Burnett JC, Hawkridge AM, Muddiman DC (2009) Coupling of a vented column with splitless nanoRPLC-ESI-MS for the improved separation and detection of brain natriuretic peptide-32 and its proteolytic peptides. *J Chromatogr B* 877:948–954
- Perkins DN, Pappin DJC, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence

- databases using mass spectrometry data. *Electrophoresis* 20:3551–3567
35. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Stat Methodol* 57:289–300
 36. Weatherly DB, Atwood JA, Minning TA, Cavola C, Tarleton RL, Orlando R (2005) A heuristic method for assigning a false-discovery rate for protein identifications from mascot database search results. *Mol Cell Proteomics* 4:762–772
 37. Zakowski J (1931) On the purification of soya-urease through precipitation with acetone and carbonic acids. *Hoppe-Seyler's Z. Physiol Chem* 202:67–82
 38. Thongboonkerd V, McLeish KR, Arthur JM, Klein JB (2002) Proteomic analysis of normal human urinary proteins isolated by acetone precipitation or ultracentrifugation. *J Am Soc Nephrol* 13:119A–119A
 39. Collier TS, Hawkridge AM, Georgianna DR, Payne GA, Muddiman DC (2008) Top-down identification and quantification of stable isotope labeled proteins from *Aspergillus flavus* using online nano-flow reversed-phase liquid chromatography coupled to a LTQ-FTICR mass spectrometer. *Anal Chem* 80:4994–5001
 40. Sahyun M, Alsberg CL (1930) On rabbit liver glycogen and its preparation. *J Biol Chem* 89:33–39
 41. Chevallet M, Diemer H, Van Dorssealer A, Villiers C, Rabilloud T (2007) Toward a better analysis of secreted proteins: the example of the myeloid cells secretome. *Proteomics* 7:1757–1770
 42. Palmer JW, Gerlough TD (1940) A simple method for preparing antigenic substances from the typhoid bacillus. *Science* 92:155–156
 43. Marusyk R, Sergeant A (1980) A simple method for dialysis of small-volume samples. *Anal Biochem* 105:403–404
 44. Bergen HR, Vasmatzis G, Cliby WA, Johnson KL, Oberg AL, Muddiman DC (2003) Discovery of ovarian cancer biomarkers in serum using NanoLC electrospray ionization TOF and FT-ICR mass spectrometry. *Dis Markers* 19:239–249
 45. Johnson KL, Mason CJ, Muddiman DC, Eckel JE (2004) Analysis of the low molecular weight fraction of serum by LC-dual ESI-FT-ICR mass spectrometry: precision of retention time, mass, and ion abundance. *Anal Chem* 76:5097–5103
 46. Basrai MA, Hieter P, Boeke JD (1997) Small open reading frames: beautiful needles in the haystack. *Genome Res* 7:768–771
 47. Gough J, Karplus K, Hughey R, Chothia C (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* 313:903–919
 48. Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B (2007) Quantitative mass spectrometry in proteomics: a critical review. *Anal Bioanal Chem* 389:1017–1031
 49. Gao J, Friedrichs MS, Dongre AR, Opitck GJ (2005) Guidelines for the routine application of the peptide hits technique. *J Am Soc Mass Spectrom* 16:1231–1238
 50. Liu HB, Sadygov RG, Yates JR (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal Chem* 76:4193–4201
 51. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, Resing KA, Ahn NG (2005) Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol Cell Proteomics* 4:1487–1502
 52. Zybailov B, Coleman MK, Florens L, Washburn MP (2005) Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal Chem* 77:6218–6224
 53. Zybailov B, Mosley AL, Sardi ME, Coleman MK, Florens L, Washburn MP (2006) Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J Proteome Res* 5:2339–2347
 54. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS (2002) Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol Mol Biol Rev* 66:506–577
 55. Lynd LR, Elander RT, Wyman CE (1996) Likely features and costs of mature biomass ethanol technology. *Appl Biochem Biotechnol* 57–8:741–761