ANNUAL REPORT

OF THE EDGEWOOD CHEMICAL BIOLOGICAL CENTER IN-HOUSE LABORATORY INDEPENDENT RESEARCH PROGRAM

FY 09

RDECOM

ECBC Strategic Mission and Vision

The U.S. Army Edgewood Chemical Biological Center (ECBC) functions as the nation's principal research, development, and engineering resource for non-medical chemical and biological warfare (CBW) defense applications. ECBC's mission is *to provide integrated science, technology, and engineering solutions to overcome the CBW threat.* As such, ECBC is charged with understanding the comprehensive threat to the warfighter and providing the scientific knowledge, technology, and materiel required to protect and enable the warfighter to effectively operate and implement strategic decisions while in the presence of CBW contamination.

ECBC has a long and successful history of developing technology in the areas of detection, protection, and decontamination and prides itself on offering extensive support throughout the entire materiel lifecycle – from basic research through technology development, engineering design, equipment evaluation, product support, sustainment, field operations, and demilitarization.

ECBC Army ILIR Proposal Process

The In-House Laboratory Independent Research (ILIR) program is specifically designed to foster increased innovation within the ECBC Basic Research portfolio. It is ECBC's vision to become the Department of the Army's central research organization for chemistry and biology (biotechnology) solutions, while maintaining its role as the premier national resource for all aspects of chemical and biological defense. To carry out ECBC's critical mission while sustaining core science competencies for the future, the Center must maintain a well-trained and technically proficient workforce. The purpose of the ILIR is to fund innovative basic research projects that are high-risk and have high payoff for fulfilling Army needs.

The Department of Defense defines Basic Research as "the systematic study directed toward greater knowledge or understanding of the fundamental aspects of phenomena and of observable facts without specific applications toward processes or products in mind." ECBC views the ILIR as a critical part of its efforts at ensuring a high-level of basic science, to foster innovation in the areas of chemistry and biology, and to mentor junior investigators in the art and practice of the laboratory sciences. Specifically, ECBC is continuing the pursuit of new technological innovations and new phenomenology at the boundaries of chemistry, biology, mathematics, and physics that will gain additional insight and advances in support of CBRNE defense missions.

After review by a panel comprised of resident and external Department of the Army Senior Technologists (STs), Senior Scientists from other DoD organizations, and civilian and military faculty members at the United States Military Academy, eight ILIR proposals were awarded funding in FY09. This Annual Review contains technical reports from each of the eight ILIR-funded projects.

Table of Contents

Characterization of Volatile Organic Compound Profiles of Bacterial Threat Agents

Jennifer Horsmon
US Army Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD, USA, 21010

ABSTRACT

Volatile organic compound (VOC) profiles are potentially an underutilized class of threat agent signatures that may be exploited in the identification of threat agents. In the present study we first focused on determining if VOC profiles collected from liquid culture headspace could be utilized to differentiate between bacterium of different genus, in this case *Bacillus* and *Yersinia*. The second focus of this study was to determine is VOC profiles could effectively differentiate between species of the same genus, *Bacillus*.

Keywords: Volatile Organic Compound (VOC), headspace, threat agent, GC-MS, fingerprints, *Bacillus*, *Yersinia*

1. INTRODUCTION

Current protocols for field detection and identification of pathogenic bacteria include a two step process. A presumptive identification is first made based upon the results of polymerase chain reaction (PCR) assays or immunological methods (such as ECL, ELISA or hand-held assays). The results are then confirmed by laboratory culture, which is time-consuming. PCR and immunochromatographic (hand-held) assays can be performed in less than an hour, ELISA requires a few hours, and laboratory culture often requires 12-24 hours or longer. Reagentless methods are currently sought by DOD because of the reduced logistical support required relative to devices that employ "wet biochemistry" methods. The current technology is intended to be robust in its ability to rapidly identify threat agents when grown under optimal/suboptimal conditions as may be the case in a battlefield. Rapid, positive identification of a possible threat will allow commanders to take appropriate action in a significantly shorter time frame than that allowed by current technology.

A number of studies have shown that bacteria may be identified by an analysis of the volatile organic compounds contained in the headspace of bacterial cultures (Aathithan et al., 2001, Bunge et al., 2008, Casalinuovo et al., 2006, Lechner et al., 2005, Probert et al., 2004, Shnayderman et al., 2005). Such analyses, generally using gas chromatography / mass spectroscopy (GC-MS), generate a profile of the chemical compounds and relative concentrations of VOCs generated by a bacterial culture during growth. Such profiles have even been used to "diagnose" infections in humans and plants. Probert et al. (2004) identified the type of bacterial or viral species causing gastrointestinal infection in humans. Other researchers have used GC-MS to identify VOCs collected from potato tubers that were inoculated with fungi (de Lacy, Costello, et al. 2001) citing the results that some VOCs were specific to each pathogen. Similar methods could be developed to identify threat agent bacterial species in liquid culture using commercially available equipment and a customized software database.

2. MATERIALS AND METHODS

2.1. Bacterial Strains and Culture Methods

The strains used in this study were obtained from in-house culture collections at ECBC; most non-select agent strains are available from American Type Culture Collection (ATCC, Manassas, VA) or the Bacillus Genetic Stock Center (The Ohio State University, Columbus, OH). Strains of *Bacillus* spp. and *Yersinia pestis* (Table 1) were grown in liquid Luria Broth (Difco), prepared according to the manufacturer's instructions. Bacteria were grown overnight at 30°C (*Bacillus*) and 37°C (*Yersinia*) in a shaking incubator at 200 rpm or on solid media. Glycerol stocks were made for each strain and were stored in LB broth and 10% glycerol at -80°C.

Table 1. Bacterial strains used in the study.

Bacillus anthracis VNR1-Δ1
Bacillus thuringiensis kurstaki
Bacillus subtilis V.niger
Bacillus cereus 1122
Bacillus cereus 13824
Bacillus myoides 6462
Bacillus licheniformis 14581
Yersinia pestis EV76
Yersinia pestis Harbin 35

2.2. Sampling and Volatile Compound Analysis

For headspace analysis of the culture one end of a multi-bed sorbent tube (Fig. 1) is attached to a 500mL filter flask and the other end is attached to a vacuum. The sample collection was carried out for 1 minute and the flow of 100mL/min, using a mass flow controller, with the sample being agitated at 100rpm. The thermal desorption into the GC was carried out by heating the multi-bed sorbent tube at 350°C for 3 minutes, then the sample was focused on a 3 bed focusing trap at 375° C for 5 minutes before being injected onto GC column.

The sample is thermally desorbed using an Dynatherm ACEM 900 (CDS Analytical, Oxford, PA) that is coupled to a gas chromatograph 6890 (Agilent Technologies, Palo Alto, CA) which is coupled to mass spectrometer 5973 (Agilent Technologies) to analyze volatile compounds. The VOCOL (Sigma 23313-U, 60 m x 0.75 mm x 1.50 μm) column was used for GC analysis. Splitless injection mode was applied, and helium gas was used as the carrier gas with a constant flow rate of 1.0 mL/min. The GC oven temperature was set at 30°C for 2 min and increased at a rate of 5°C/min to a final temperature of 155°C. The mass spectrometer (MS) was operated in electron impact (EI) mode. Data acquisition was performed in total ion (TI) mode with a total scan time of 28 minutes.

Selected GC-MS peaks were identified by comparing the mass spectra and the retention index of the peaks with those from the National Institute of Standards and Technology (NIST) mass spectral library (Wiley registry). The signal-to-noise threshold level was set at 10 for selection of major volatile compounds and all selected peaks were integrated. The integrated peaks will be imported into statistical software for multivariate statistical analysis at the completion of this study.

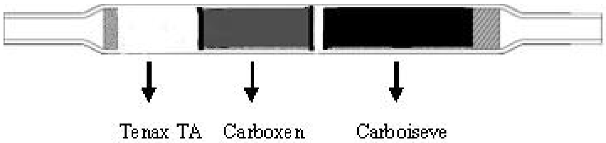


Figure 1. Multi-bed sorbent tube. Sorbent tubes are used for the collection and sampling of gases and vapors. The tubes we used were made of glass and contain 3 types of solid adsorbent materials, Tenax TA, Carboxen and Carboiseve. These materials make it possible to trap and retain compounds of interest.

3. RESULTS

3.1. Determination of Different Genus

The first task was to examine bacteria of different genus to determine if they produced unique spectra and differing concentrations of volatile compounds. For this experiment, *Bacillus anthracis* VNR-1Δ1 and *Yersinia pestis* EV-76 GC peak retention times and mass spectral patterns were examined. Diethyl ester decanedioic acid, decanal, naphthalene, pentanoic acid and toluene were particularly prevalent in the *Yersinia pestis* EV-76 samples while diisooctyl ester 1,2-benzenedicarboxylic acid, decene and tetradecane were the major components of *Bacillus* *anthracis* VNR1-Δ1 (Figure 2a & 2b). The data collected from the two samples provided sufficient evidence that the GC-MS is capable of differentiating between two different genus based on their VOC profiles (Table 2).

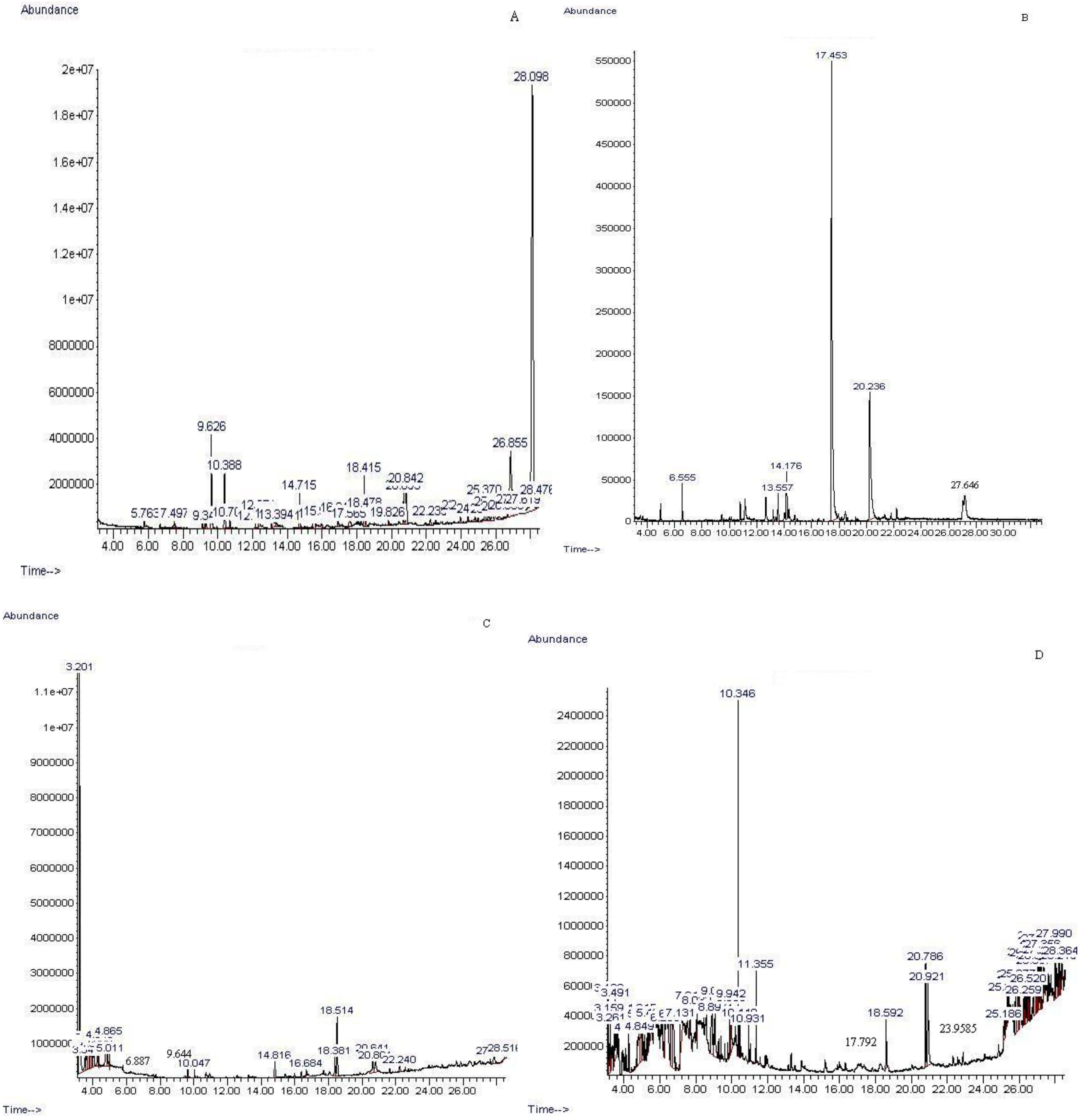
Table 2. Summary of unique VOC peaks identified for each bacterium tested.

RT (min)	Compounds	Bacteria								
		YP EV-76	YP Harbin 35	BA VNR1-Δ1	BT Kurstaki	BS V.niger	BC1122	BC 13824	BM 6462	BL 14581
9.1743	Disulfide, dimethyl				x	x		x	x	
9.6442	Disulfide, dimethyl		x							
9.6273	Toluene				x	x		x	x	
9.9161	Toluene	x								
10.9012	Toluene		x							
11.1347	Decene						x			
12.3677	o-Xylene							x		
13.0586	o-Xylene	x			x				x	
13.1719	Styrene					x		x	x	
13.3006	Pentadecane			x						
13.939	Tetradecane			x			x			
15.3462	Benzaldehyde	x			x	x				
15.3745	Dimethyl trisulfide							x	x	
16.841	Phenol	x								
17.4329	Decene			x			x			
17.7922	Undecane				x					x
18.1208	Acetophenone	x			x			x	x	
18.381	Nonanal		x							
20.2163	Decene			x			x			
20.782	Decanal	x	x		x					x
20.8612	Naphthalene	x			x	x		x	x	
21.2775	Benzenecarboxylic acid				x	x				x
23.9585	Tetradecane	x			x					x
24.7909	Biphenyl								x	
25.3458	2,2'-Dimethylbiphenyl	x				x		x	x	
26.4214	1,2-Benzenedicarboxylic acid, mono(2-ethylhexyl) ester							x		
26.9651	Decanedioic acid, diethyl ester	x								
27.0346	Phthalic acid					x	x		x	
27.646	1,2-Benzenedicarboxylic acid, diisooctyl ester			x						
28.4599	Pentanoic acid, 2,2,4-trimethyl-3-carboxyisopropyl	x								

Volatile compounds identified by GC-MS (extraction temp: RT, extraction time: 1 minute, 100 rpm).

3.2. Differences between Bacillus species

After determining that VOCs from the genus Yersinia and Bacillus could be differentiated visually by their chromatograms and by the major peaks detected by the GC-MS, the next step was to determine if unique VOCs from several different species of Bacillus could be determined. The following Bacillus species were utilized in this task, *Bacillus anthracis* VNR1-Δ1, *Bacillus thuringiensis* kurstaki, *Bacillus subtilis* V.niger, *Bacillus licheniformis* (ATCC 14581), *Bacillus mycoides* (ATCC 6462) *Bacillus cereus* (ATCC 13824) and *Bacillus cereus* (ATCC 1122). Results from this study indicate that GC-MS analysis is capable of differentiating bacterial species by their unique VOC fingerprints as shown below in Figure 2 and in Table 2.


A
Abundance
2e+07
1.8e+07
1.6e+07
1.4e+07
1.2e+07
1e+07
8000000
6000000
4000000
2000000
28.098
9.626
10.388
26.855
18.415
20.842
14.715
25.370
28.476
5.763
7.497
13.394
17.565
19.826
22.235
Time-->
4.00 6.00 8.00 10.00 12.00 14.00 16.00 18.00 20.00 22.00 24.00 26.00
B
Abundance
550000
500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
17.453
20.236
6.555
14.176
13.557
27.646
Time-->
4.00 6.00 8.00 10.00 12.00 14.00 16.00 18.00 20.00 22.00 24.00 26.00 28.00 30.00
C
Abundance
1.1e+07
1e+07
9000000
8000000
7000000
6000000
5000000
4000000
3000000
2000000
1000000
3.201
4.865
5.011
6.887
9.644
10.047
14.816
16.684
18.514
18.381
20.641
22.240
28.516
Time-->
4.00 6.00 8.00 10.00 12.00 14.00 16.00 18.00 20.00 22.00 24.00 26.00
D
Abundance
2400000
2200000
2000000
1800000
1600000
1400000
1200000
1000000
800000
600000
400000
200000
10.346
11.355
3.491
3.159
3.261
4.849
7.131
9.942
8.89
10.931
17.792
18.592
20.786
20.921
23.9585
25.186
26.520
26.259
27.990
28.364
Time-->
4.00 6.00 8.00 10.00 12.00 14.00 16.00 18.00 20.00 22.00 24.00 26.00

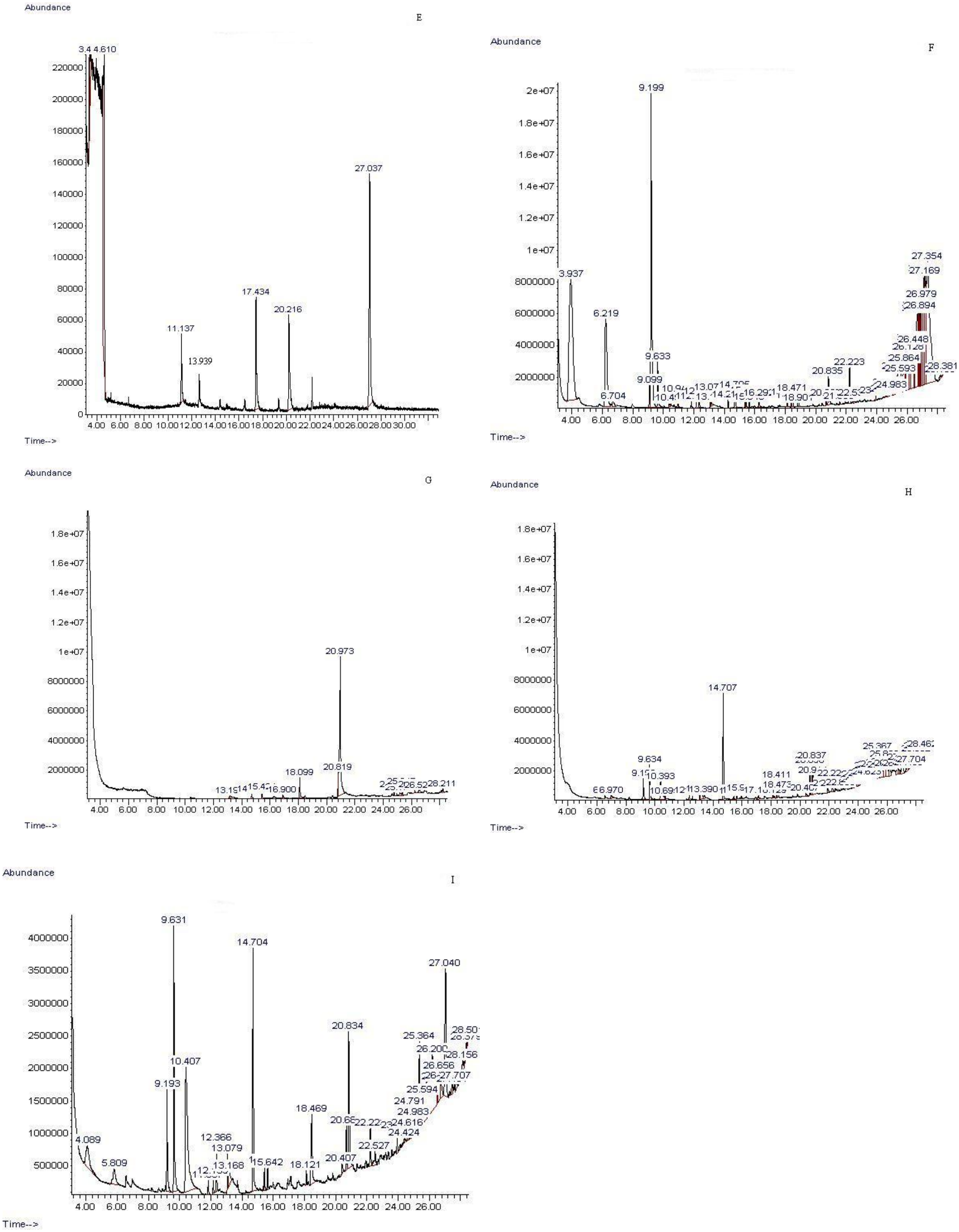


Figure 2. Typical examples of a gas chromatography-mass spectroscopy (GC-MS) chromatogram analysed on a VOCOL column for (A) *Bacillus anthracis* VNR1-Δ1, (B) *Yersinia pestis* EV-76, (C) *Yersinia pestis* Harbin 35, (D) *Bacillus licheniformis* 14581,

(E) *Bacillus cereus* 1122, (F*) Bacillus cereus* 13824, (G) *Bacillus myoides* 6462, (H) *Bacillus subtilis* V.niger, (I) *Bacillus thuringiensis* kurstaki.

3.3 Differences between Strains

After determining that VOCs from the genus Yersinia and Bacillus and several Bacillus species could be differentiated by the major peaks detected by the GC-MS and their unique chromatograms, the next step was to determine if unique VOCs from different strains of Bacillus and Yersinia could be determined. *Bacillus cereus* 13824 and 1122 along with *Yersinia pestis* EV76 and Harbin 35 were utilized in this task. Diethyl ester decanedioic acid, decanal, naphthalene, pentanoic acid and toluene were particularly prevalent in the *Yersinia pestis* EV-76 samples while dimethyl disulfide, toluene, nonanal and decanal were predominately found in *Yersinia pestis* Harbin 35. Decene and phthalic acid were most prevalent in Bacillus cereus 1122 while dimethyl disulfide, toluene, naphthalene and 1,2-benzenedicarboxylic acid, mono (2-ethylhexyl) ester were commonly found in *Bacillus cereus* 13824. Chromatograms for *Yersinia pestis* EV76 and Harbin 35 can be seen in figure 2a and 2c while *Bacillus cereus* 1122 and 13824 chromatograms can be found in figure 2e and 2f. Their corresponding peaks can all be found in Table 2.

4. CONCLUSIONS

The results from this study indicate that GC-MS analysis is capable of differentiating bacterial species and strains by their unique VOC fingerprints. Further research is currently underway to determine the unique VOC profiles for Yersinia pestis CO92 and Bacillus anthracis NH1. We are also testing the ability to determine if VOC profiles change in regards to passage number and if so, could this technology be used to determine genetic drift. The selected GC-MS peaks for each species/strain were identified by comparing the mass spectra and the retention index of the peaks with those from the National Institute of Standards and Technology (NIST) mass spectral library (Wiley registry). The integrated peaks will be imported into statistical software for multivariate statistical analysis at the completion of this study.

Future studies are planned to determine if this technology can further identify at the species and strain levels utilizing different bacterial collections. Culture media and cultivation conditions will also be investigated to determine which VOC fingerprints are uniquely characteristic of the microbial metabolism of each species and strain, regardless of the culturing process. If this is successful, a searchable database could be built that would allow for VOC analysis in the field to yield rapid and reliable identification of bacterial species grown in liquid culture.

ACKNOWLEDGEMENTS

The author would like to thank Kathy Crouse for her expertise and help with developing the protocols for sampling VOCs, Dr. Kevin O’Connell for his helpful discussions throughout the project and Ms. Heather Welsh for assisting in routine laboratory procedures.

REFERENCES

[1] Aathithan S, Plant JC, Chaudry AN, and French GL. 2001. Diagnosis of Bacteriuria by Detection of Volatile Organic Compounds in Urine Using an Automated Headspace Analyzer with Multiple Conducting Polymer Sensors. J. Clin. Micro. 39: 2590-2593.

[2] Bunge M, Araghipour N, Mikoviny T, Dunkl J, Schnitzhofer R, Hansel A, Schinner F, Wisthaler A, Margesin R, and Mark T. 2008. On-Line Monitoring of Microbial Volatile Metabolites by Proton Transfer Reaction-Mass Spectrometry. Appl. Environ. Microbiol. 74: 2179-2186.

[3] Casalinuovo IA, Pierro DD, Francesco PD, and Coletta M. 2006. Experimental use of a new surface acoustic wave sensor for the rapid identification of bacteria and yeasts. Letters in Appl. Micro. 42: 24-29.

[4] Ferreira, L, Perestrelo, R, Caldeira, M, and Camara, J.S. 2009. Characterization of volatile substances in apples from *Rosaceae* family by headspace solid-phase microextraction followed by GC-qMS. J. Sep. Sci. 32: 1875-1888.

[5] Lechner M, Fille M, Hausdorfer J, Dierich MP, Rieder J. 2005. Diagnosis of bacteria in vitro by mass spectrometric fingerprinting: a pilot study. Curr. Microbiol. 51: 267-269.

[6] Lee, J.E., Hong, Y.S., Lee, C.H. 2009. Characterization of Fermentative Behaviors of Lactic Acid Bacteria in Grape Wines through ^{1}H NMR- and GC-Based Metabolic Profiling. J. Agric. Food Chem. 57: 4810-4817.

[7] Probert CSJ, Jones PRH, and Ratcliffe NM. 2004. A novel method for rapidly diagnosing the causes of diarrhea. Gut. 53: 58-61.

[8] Shnayderman M, Mansfield B, Yip P, Clark H, Krebs MD, Cohen SJ, Zeskind JE, Ryan ET, Dorkin HL, Callahan MV, Stair TO, Gelfand JA, Gill CJ, Hitt B, Davis CE. 2005. Species-Specific Bacteria Identification Using Differential Mobility Spectrometry and Bioinformatics Pattern Recognition. Anal. Chem. 77:5930-5937.

Design and Testing of a Thermostable Platform for Multimerization of Single-Domain Antibodies

James P. Carney[a], Ellen Goldman[b], Alena M. Calm
[a]US Army ECBC 5183 Blackhawk Rd. RDCB-DRB-M, Aberdeen Proving Ground, MD, USA 21010
[b]Naval Research Laboratory, 4555 Overlook Ave. SW, Washington, DC, USA 20375

ABSTRACT

Immunoaffinity reagents have been demonstrated to be a powerful tool for the detection of biothreat agents. In spite of this success one limitation of these tools is the inherent thermal sensitivity of the molecules. Thus, the ability to maintain specific and sensitive recognition characteristics in combination with increased thermostability would provide an enhanced capability for fielding immunoaffinity based sensors. One potential technology capable of delivering this improved thermostability is single domain antibodies generated from sharks and camelids (camels and llamas). Largely due to their small size (12-14 kDa) these molecules have substantial thermostability as well as the ability to refold after exposure to extremes of temperature. However, the current lack of hybridoma technology for these species forces these antibodies to be selected by in vitro panning methods. This often leads to a general reduction in affinity towards the target antigen when compared to standard mouse monoclonal antibodies. One solution to this challenge is to multimerize antibodies leading to enhanced recognition. In this proposal we will use recombinant technology to multimerize thermostable single domain antibodies in a flexible, scalable and thermostable platform. Through collaboration with Dr. Ellen Goldman at the Naval Research Laboratory, we will incorporate two distinct ricin A chain binding single domain antibodies into a platform based on the structural maintenance of chromosome (SMC) proteins. The SMC portion of the molecule will be derived from the thermophilic organism Pyrococcus furiosus. The engineered molecules will be produced and their recognition characteristics determined. In addition, we will examine the thermostability of the molecules. Thus, we will combine the thermostability of the single domain antibodies with a thermostable platform to increase affinity of these molecules. This platform will be generally applicable to multimerization of single domain antibodies as well as have the potential for addition of signaling and/or mitigation activities.

Keywords: Single domain antibodies, Structural maintenance of chromosome proteins, Pyrococcus furiosus, Rican

1. INTRODUCTION

The development of antibody based detection systems have proven to be of great utility in the detection and initial characterization of biothreat agents. The specificity and sensitivity of antibodies has made them a valuable resource in the continued evolution of sensor development. Despite the successes of antibodies generated from mammalian sources (typically mouse, goat and rabbit) there are several drawbacks. The primary one relates to the need for maintaining the antibodies in a refrigerated environment as well as the inherent instability associated with these molecules leading to reduced shelf-life. This limits the ability to employ antibodies in the field.

Recent developments in the field of immunology have demonstrated the existence of single domain antibodies made up of a single heavy chain molecule from several organisms including sharks, camels and llamas (1). It has become possible to isolate small variable domains (12-14 kDa) that are highly specific to a particular antigen (2,3). The small size of the isolated variable domains combined with the physiology of these animals results in single domain antibodies that are extremely thermostable. The limitation of the techniques used for isolation of these single domain antibodies is that there are no tumor cell lines available to create hybridomas and the resulting molecules must be isolated in vitro using phage display technology. This often leads to reduced binding affinity. One method that has been developed to overcome this limitation is the multimerization of a given single variable domain to allow for enhanced affinity (4). While this has proven valuable it has not been conducted in a manner that would maintain the thermostability of the molecule, i.e. the domains used for dimerization are not thermostable resulting in a molecule that still suffers from the inherent limitations of standard mammalian antibodies (4). We have attempted to overcome these limitations by designing a platform for multimerizing single domain antibodies through the use of structural maintenance of chromosome (SMC) proteins. This conserved class of proteins forms a flexible intramolecular coiled-coil as well as intermolecular dimerization with a second SMC protein (Figure 1) (5). This

Family of proteins is highly conserved and much work has been done on these proteins from the archael kingdom, including the thermophiles (6).

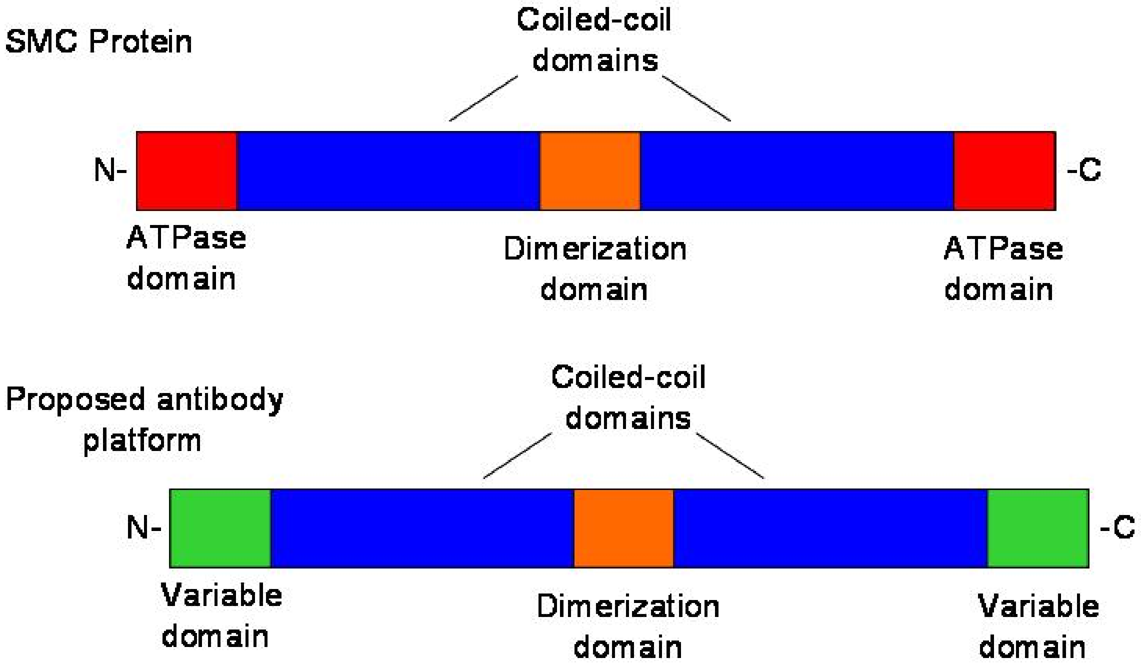


FIGURE. 1. Comparison of SMC molecules and the Anitbody Nanoassembly. Upper portion shows the schematic of a standard SMC protein. Lower portion indicates how the antibody nanoassembly was created by substituting the ATPase domain of the SMC protein with variable domains of a llama single domain antibody. This scaffolding allowed for two different variable domains to be incorporated into the molecule.

2. METHODS AND PROCESS DEVELOPMENT

Thermostable forms of the SMC molecule were used along with structural and functional data to design the construct (6-8). Dr. Ellen Goldman of the Naval Research Laboratory provided two individual Llama single domain antibodies that recognize distinct epitopes of the ricin A chain. These sequences were incorporated into the construct at the site of the SMC molecule where the ATPase domain was located (2,3,9).

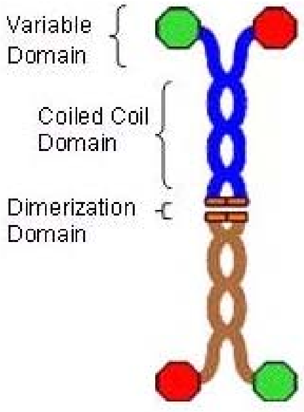


Figure 2. Illustration of C8smcF11. The red and green octagons represent the two different single domain ricin antibodies. The variable domains are tethered to the SMC coiled-coil backbone, which then dimerize. This allows for the creation of a ricin binding heteropolymeric quadravalent recognition moiety.

During the FY09 project year we were able to Design the Single domain ricin antibody SMC fusion construct (C8smcF11) (Figure 2). We tested different vectors and selected the one most suited for expression. Transformation and expression conditions were optimized. Purification of the expression product C8smcF11 was scaled up to one liter and optimized. We were also able to test the binding characteristics of the molecule and generate initial thermostability data.

In the design phase, the construct was created using sequences provided by Dr. Ellen Goldman along with published Sequences of Pyrococcus furiosus. The Program "Coils" was used from the Swiss EMBnet node server to help determine where the coils and the dimerization domain were located. The sequence was uploaded into the Coils program which then calculates the probability that the sequence will adopt a coiled-coil conformation. We chose areas with a low probability of forming a coiled-coil to insert the variable domains. DNA2.0 cloned the sequence into two different Vectors for us to test. The first construct that was delivered was cloned into pJexpress401 and was labeled p26739. This vector was supplied gratis by DNA 2.0 for testing, the company claimed that it routinely expressed better than traditional pET vectors and could be expressed in any strain of *E.coli.* The construct was later delivered in the pET28a(+) vector. Both plasmids were received lyophilized. Upon receipt they were rehydrated and transformed into DH5α and glycerol stocks were made. They were also transformed into BL21(DE3)SLyD- strains of *E.coli.* A pilot expression study was done to compare the expression levels between the two vectors in both DH5α and BL21. The results indicated that p26740 in BL21 was the better vector and strain to express C8smcF11 (Figure 3).

All subsequent work was done using p26740 in BL21 electrocompetent *E.coli.* Transformations were performed using electroporation. Briefly, 100ul of electrocompetent cells, and the p26740 plasmid were thawed on ice. 1ul of plasmid was added to the cells and triturated. The mixture was transferred to a chilled Bio-Rad electroporation cuvette which was placed immediately in the electroporator. 2.5kV of current was applied to the cells and they were immediately transferred into 1ml of SOC media. The culture was incubated at 37ºC for one hour with agitation. The culture was plated out on LB plates with appropriate antibiotic (30ug/ml kanamycin). Plates were incubated overnight at 37ºC.

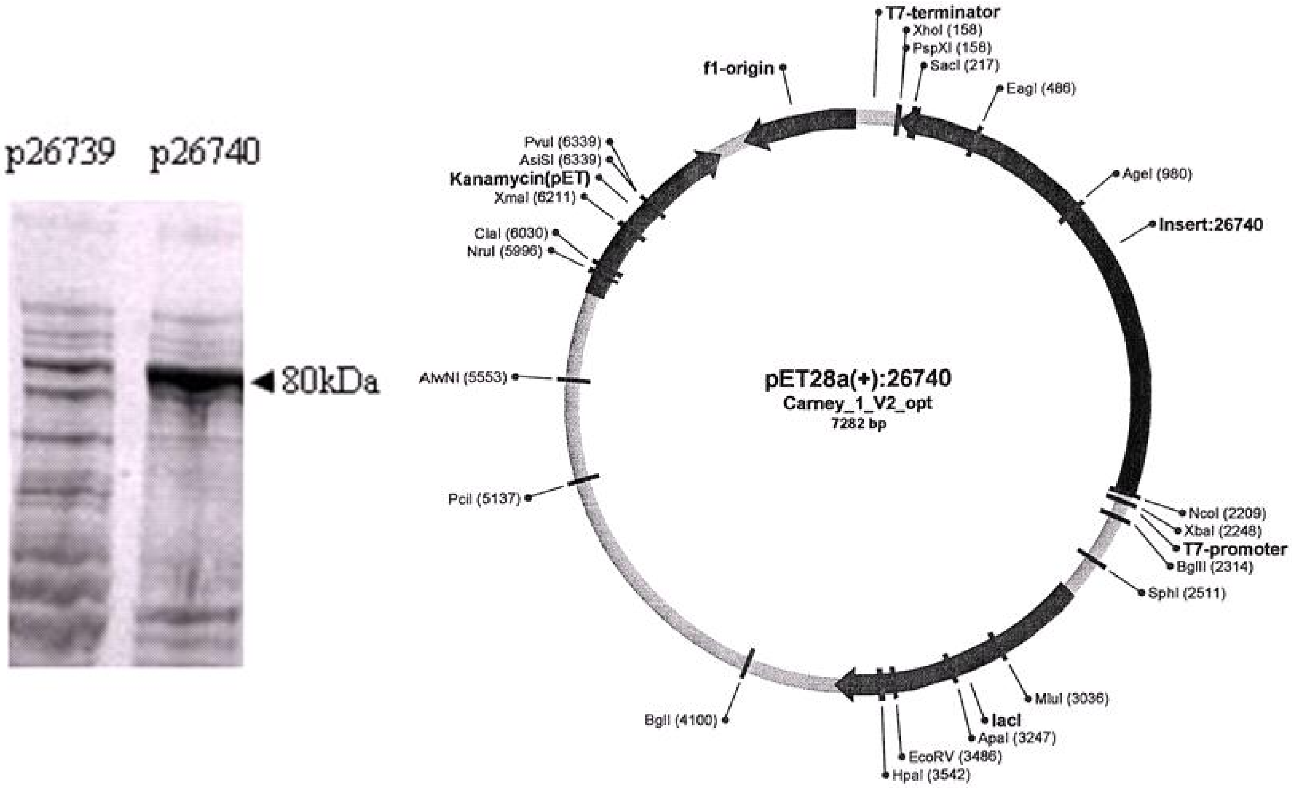


Figure 3. (Left) pilot expression comparing p26739 and p26740 in BL21(DE3)SLyD- *E.coli*. The 80kDa band highlighted corresponds to the predicted molecular weight of C8smcF11. (Right) Vector map supplied by DNA 2.0 of C8smcF11 in pET28a(+).

Inductions for expression of C8smcF11 were always started with single colonies that had been transformed within 24 hours. Our induction strategy utilized Auto Induction Instant TB media from Novagen. The media contains components that are metabolized differentially to promote high density cell growth and induces protein expression from *lac* promoters without the usual steps of monitoring cell density and adding IPTG. 10 ml of media plus kanamycin (30ug/ml) was inoculated with a single transformant and grown to log phase (0.6 O.D.$_{600}$) before being scaled up to a 500ml culture, these half liter cultures were then grown at 37ºC 300rpm for 18 hours. Typically 1ml of finished culture was saved to check expression levels via SDS PAGE (Figure 4). All cell paste was collected via centrifugation and stored at -80ºC. The construct was designed with a 6x his tag, so our first step in purification was IMAC. The initial attempt involved batch affinity purification. Briefly, the cell paste was thawed on ice and resuspended in lysis buffer. The cells were ruptured in a chilled French press and centrifuged to pellet the remaining cell debris. The clarified lysate was incubated with Ni-NTA agarose (Qiagen) with gentle agitation at 4ºC. The agarose was separated from the lysate via gentle centrifugation. The beads were washed and the protein was eluted off with 300mM imidizole.

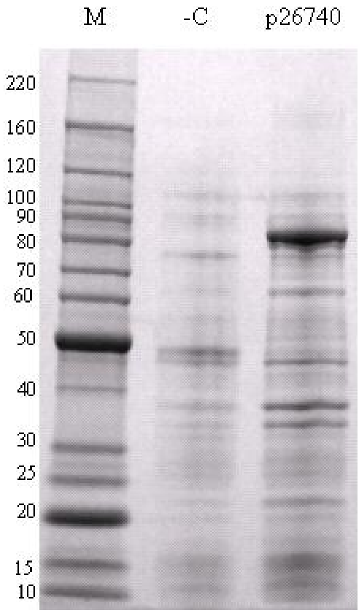


Figure 4. 4-15 % Tris-Glycine gel stained with SimplyBlue™. The 80kDa band in the p26740 lane represents typical C8smcF11 expression levels. The negative control lane for comparison is the induction of an empty pET28a(+) vector.

The initial product was noticeably degraded (Figure 5). The purification scheme was modified to make use of the AKTA purifier FPLC. The first columns we used were 1ml HisTrap ff pre-packed nickel columns (GE Healthcare). The cell paste was processed in the same way; with the addition of EDTA free protease inhibitors. The clarified lysate was pumped onto the column and washed before it was eluted off with 300mM imidizole, the product from these first FPLC runs were still degraded (Figure 5). Procedural modifications including, faster processing of cell paste and the use of extra protease inhibitors helped reduce proteolysis. We also switched to a 5ml HisTrap ff crude column; this helped alleviate a problem with high back pressure. The elution was changed so that the concentration of imidizole was stepped up to 300mM in 4 steps. After these modifications, the protein started coming of the nickel columns in tact. There was a considerable amount of non-specific protein that was coming off in the elution, to remediate this; ion exchange was added as a final step (figure 5). The best fractions from the HisTrap were pooled and desalted into 25mM Tris 7.6, 50mM NaCl using a 5ml HiTrap desalting column (GE Healthcare). The peak fractions from the desalting column were pooled and loaded onto a 1ml ff Q column (GE Healthcare). The protein was eluted using a salt concentration gradient from 50mM to 500mM NaCl over 10 column volumes. There is still minor protein contamination (figure 5) we believe this can be eliminated with the addition of a gel filtration column. We were able to produce and purify enough protein in FY09 to study the binding characteristics of the molecule and generate initial thermostability data.

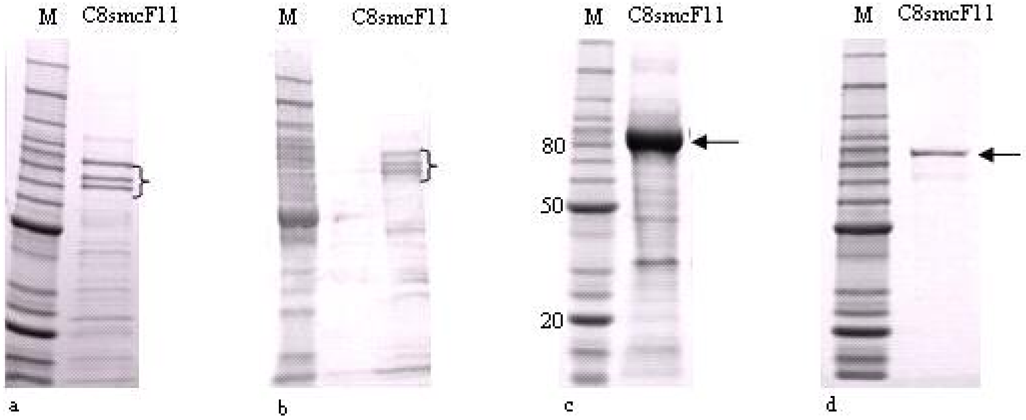


Figure 5. Gel images represent the progress made in purifying C8smcF11. a. results from batch purification with Nickle-NTA agarose, protein product is degraded. b. results from 1ml HisTrap ff column, protein product is degraded. c. results from 5ml crude HisTrap column, protein product is intact. d. results from quaternary amine (Q) ion exchange 1ml ff column, protein is intact and relatively pure.

The ProteOn (Bio-Rad) was used to assess weather or not our construct would recognize and bind to ricin. The ProteOn uses Surface Plasmon Resonance (SPR) to deliver data on affinity, specificity, and kinetics of protein interactions. The instrument uses microfluidics to load and bind ligands with amine groups to ProteOn sensor chip carboxyl surfaces via sulfo-NHS chemistry. Once a chip is made it can be used multiple times. We took a ProteOn GLC chip and loaded it with ricin, followed by dilutions of the purified C8smcF11. The results indicated that the single domain ricin antibodies on C8smcF11 do recognize ricin, and they bind tightly (Figure 6).

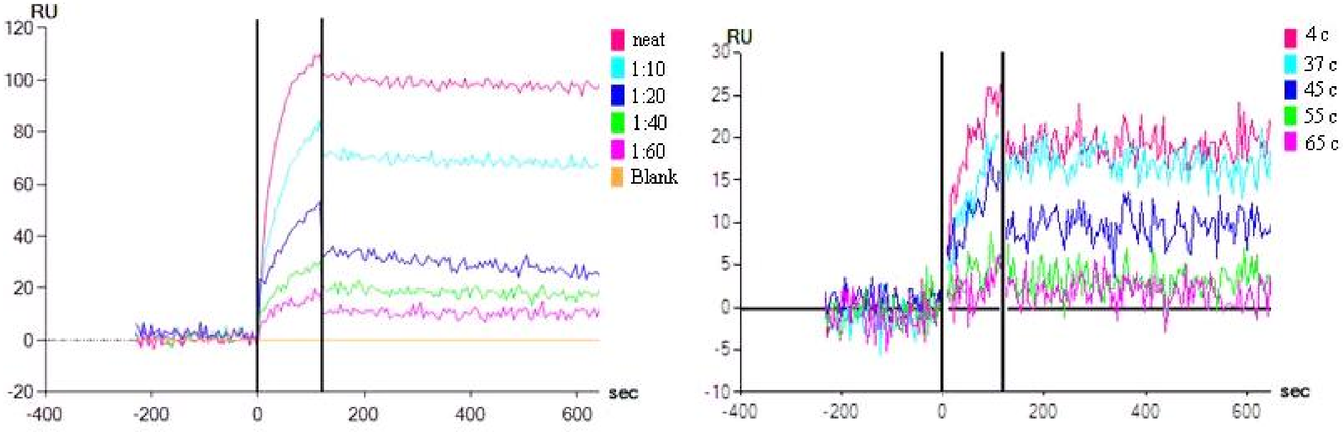


Figure 6. (Left) SPR data shows that C8smcF11 binds with high affinity to the ricin on the chip. -200-0 = baseline, 0-150 is the association phase when binding occurs; 150-600 is the disassociation phase. Notice the lack of downward slope in the disassociation phase; this indicates that C8smcF11 has a high association constant. The colors represent dilutions of protein applied to the chip. (Right) SPR data shows thermostability of C8smcF11. All samples were treated for 15 minutes, and then placed on ice before application to the chip.

More experiments need to be done to generate kinetic data, due to the very slow disassociation rate, it is impossible to analyze the current data. Initial thermostability studies indicate that C8smcF11 retains some activity up to 45ºC. We predict that by shortening the length of the coiled coil in the construct we can improve the thermostability of this molecule. We exposed C8smcF11 to five different thermal conditions ranging from 4ºC-65ºC for 15 minutes. This was followed by brief incubation on ice prior to being loaded on the chip (Figure 6). Differences in the response units (RU) between the two experiments can be attributed to protein concentration and the age of the chip.

3. CONCLUSIONS AND FUTURE DIRECTIONS

It can be concluded from the work completed in FY09 that a single domain antibody can be multimerized and tethered into a thermostable platform. We have constructed a recombinant protein in which the ATPase domain of an SMC protein has been replaced with single domain antibody variable domains of two distinct epitopes of the ricin A chain. This protein is able to be expressed in *E.coli* and purified. Upon testing, this molecule recognizes and binds to ricin and retains some activity up to 45ºC. In the future we would like to create a more thermostable form of the molecule by reducing the length of the coiled-coil. We would also like to test the activity of the protein via ELISA.

ACKNOWLEDGEMENTS

This work was made possible through a grant from the ECBC Research and Technology Directorate ILIR program.

REFERENCES

[1] Goldman, E.R., Anderson, G.P., Conway, J., Sherwood, L.J., Fech, M., Vo, B., Liu, J.L. and Hayhurst, A. (2008) Thermostable Llama Single Domain Antibodies for Detection of Botulinum A Neurotoxin Complex. Analytical chemistry.

[2] Goldman, E.R., Anderson, G.P., Liu, J.L., Delehanty, J.B., Sherwood, L.J., Osborn, L.E., Cummins, L.B. and Hayhurst, A. (2006) Facile generation of heat-stable antiviral and antitoxin single domain antibodies from a semisynthetic llama library. Analytical chemistry, 78, 8245-8255.

[3] Haering, C.H., Lowe, J., Hochwagen, A. and Nasmyth, K. (2002) Molecular architecture of SMC proteins and the yeast cohesin complex. Molecular cell, 9, 773-788.

[4] Harmsen, M.M. and De Haard, H.J. (2007) Properties, production, and applications of camelid single-domain antibody fragments. Applied microbiology and biotechnology, 77, 13-22.

[5] Lammens, A., Schele, A. and Hopfner, K.P. (2004) Structural biochemistry of ATP-driven dimerization and DNA-stimulated activation of SMC ATPases. Curr Biol, 14, 1778-1782.

[6] Liu, J.L., Anderson, G.P. and Goldman, E.R. (2007) Isolation of anti-toxin single domain antibodies from a semi-synthetic spiny dogfish shark display library. BMC biotechnology, 7, 78.

[7] Lowe, J., Cordell, S.C. and van den Ent, F. (2001) Crystal structure of the SMC head domain: an ABC ATPase with 900 residues antiparallel coiled-coil inserted. Journal of molecular biology, 306, 25-35.

[8] Nasmyth, K. and Haering, C.H. (2005) The structure and function of SMC and kleisin complexes. Annual review of biochemistry, 74, 595-648.

[9] Simmons, D.P., Abregu, F.A., Krishnan, U.V., Proll, D.F., Streltsov, V.A., Doughty, L., Hattarki, M.K. and Nuttall, S.D. (2006) Dimerisation strategies for shark IgNAR single domain antibody fragments. Journal of immunological methods, 315, 171-184.

Discrimination of Pathogenic vs. Non-pathogenic *Yersinia pestis* and *Escherichia coli* Using Proteomics Mass Spectrometry

Mary Margaret Wade[a], Rabih Jabbour[b]
[a]Research and Technology Directorate, Aberdeen Proving Ground, MD, USA, 21010
[b]Science Applications International Corporation, Gunpowder, MD, USA, 21010

PREFACE

The work described in this report was started in January 2009 and completed in September 2009. The use of either trade or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement. This report has been approved for public release.

1. INTRODUCTION

Recently mass spectrometry analysis has proven useful in the characterization and identification of biological agents using a proteomic approach (4). Therefore, the present study sought to determine whether proteomics mass spectrometry could be used to distinguish between pathogenic and non-pathogenic strains of the same organism. More specifically, discrimination between pathogenic and non-pathogenic organisms based on their outer membrane protein (OMP) composition, as determined by mass spectrometry, was investigated.

OMPs of gram-negative bacteria act as active mediators between the cell and its environment and are often associated with virulence in gram-negative pathogens. In pathogenic *Escherichia coli*, there are multiple OMPs present which are needed for intestinal colonization, as well as those that play a role in the type III secretion system responsible for delivering effector proteins to host cells (7, 8, 9, 10, 14). Virulent *Yersinia pestis* contains three plasmids encoding multiple OMPs that are required for virulence (1, 5, 6). For example the pCD1 plasmid encodes several *Yersinia* outer membrane proteins (YOPs) and a type III secretion system, which are needed for survival and entry into eukaryotic cells (3, 11). Additionally, the pPCP1 plasmid encodes an OMP plasminogen activator that interferes with clotting and complement (12). Avirulent strains often lack one or more of the plasmids or genes encoding proteins needed for virulence and it is these differences in OMP expression between virulent and avirulent strains of gram-negative bacteria that could potentially be exploited in order to distinguish among strains. Therefore, OMPs could prove to be excellent model biomarkers for strain differentiation among bacteria.

The objective of the present study was to establish the sequence-based identity of OMPs isolated from pathogenic and non-pathogenic strains of *Y. pestis* and *E. coli*. *Y. pestis* is classified as a Category A pathogen and is an important potential biowarfare agent. Pathogenic *E. coli*, such as *E. coli* O157:H7, is an important public health pathogen responsible for most common foodborne and waterborne illnesses in the United States. High-throughput proteomic analytical systems were applied, providing a rapid means of characterizing cellular proteins and producing amino acid sequence information for peptides derived from these proteins.

This one year basic research study aimed to: 1) isolate OMPs using ultracentrifugation and differential extractions, 2) determine sequence and post-translational modifications to amino acid residues composing membrane proteins using emerging high-throughput mass spectral proteomic systems, and 3) utilize bioinformatics modeling tools to establish strain differentiation methods based on the proteome differences among the *Y. pestis* and *E. coli* strains.

In addition to the aims described above, discrimination among strains of an additional agent of interest, *Bacillus anthracis*, was also investigated. Due to the fact that *B. anthracis* is a gram-positive organism and therefore lacks an outer membrane, total cellular proteins (whole cell lysates) were analyzed for discrimination via mass spectrometry rather than OMPs.

2. MATERIALS AND METHODS

2.1. Materials and Reagents

Ammonium bicarbonate, dithiotheritol, urea, acetonitrile-HPLC grade, and formic acid were purchased from Burdick and Jackson (St. Louis, MO). Sequencing grade modified trypsin was purchased from Promega (Madison, WI).

2.2. Bacterial Strains and Culture Conditions

Pathogenic strains employed in the present study were: *E. coli* O157:H7, *Y. pestis* Colorado 92 (CO 92) and *B. anthracis* Ames. Non-pathogenic strains employed were: *E. coli* K12, *Y. pestis* A1122 and *B. anthracis* Sterne. Working cultures were prepared by streaking cells from cryopreserved stocks onto tryptic soy agar (TSA) followed by incubation for approximately 18 hours at 37^{o}C for *E. coli* and *B. anthracis* strains and 30^{o}C for *Y. pestis* strains. After incubation, all working culture plates were stored at 4^{o}C. Cells from working cultures were used to inoculate broth cultures for each strain, which consisted of 100 ml of tryptic soy broth (TSB) for *E. coli* and *B. anthracis* strains and 100 ml of brain heart infusion (BHI) for *Y. pestis* strains. All cultures were incubated for approximately 18 hours at 37^{o}C for *E. coli* and *B. anthracis* strains and 30^{o}C for *Y. pestis* strains with rotary aeration at 180 rpm. After incubation, broth cultures were pelleted by centrifugation (2,300 RCF at 4^{o}C for 10 minutes), washed and resuspended in 10 ml HEPES buffer followed by heating at 95^{o}C for 1 hour to lyse cells. After heating, a portion of each sample was plated onto TSA and incubated for five days at appropriate temperature to ensure no growth prior to removing samples from the BSL-2 or BSL-3 laboratory for further processing. Total cellular protein samples (whole cell lysates) were complete after heating for the 1 hour and were transferred to Point Detection Branch for analysis after no growth on plates was confirmed. For OMP samples, samples were processed for OMP isolation as described below prior to being transferred to Point Detection Branch for analysis.

2.3. OMP Isolation

After lysis by heating at 95^{o}C for one hour, cell debris was pelleted by centrifugation at 2,300 RCF at 4^{o}C for 10 minutes. The supernatant was then centrifuged at 100,000 x g for one hour to pellet proteins. The pellet was resuspended in 1 ml of HEPES buffer and 1 ml of a 2% sarkosyl solution (N-Lauroylsarcosine sodium salt solution) was added and sample was incubated at room temperature for 30 mintues. Next samples were centrifuged at 100,000 x g for one hour and the pellet containing OMPs was resuspended in 1 ml of HEPES buffer and then transferred to Point Detection Branch for further processing and analysis as described below.

2.4. Processing of Whole Cell Lysates and OMP Samples

All protein samples were ultra-sonicated (20 seconds pulse on, 5 seconds pulse off, and 25% amplitude for 5 minutes duration) and a small portion of lysates was reserved for 1-D gel analysis. The lysates were centrifuged at 14,100 x g for 30 minutes to remove any debris. The supernatant was then added to a Microcon YM-3 filter unit (Millipore; Cat #: 42404) and centrifuged at 14,100 x g for 30 minutes. The effluent was discarded. The filter membrane was washed with 100 mM ammonium bicarbonate (ABC) and centrifuged for 15-20 minutes at 14,100 x g. Proteins were denatured by adding 8 M urea and 3μg/μl Dithiotheritol (DTT) to the filter and incubating overnight at 37°C on an orbital shaker set to 60 rpm. Twenty microliters of 100% of acetonitrile (ACN) was added to the tubes and allowed incubate at room temperature for 5 minutes. The tubes were then centrifuged at 14,100 x g for 30-40 minutes and washed three times using 150μL of 100mM ABC solution. On the last wash, ABC was allowed to sit on the membrane for 20 minutes while shaking, followed by centrifugation at 14,100 x g for 30-40 minutes. The micron filter unit was then transferred to a new receptor tube and proteins were digested with 5 μl trypsin in 240 μl of ABC solution + 5 μl ACN. Proteins were digested overnight at 37°C on an orbital shaker set to 55 rpm. Sixty microliters of 5% ACN/0.5% formic acid (FA) was added to each filter to quench the trypsin digestion followed by two minutes of vortexing for sample mixing. The tubes were centrifuged for 20-30 minutes at 14,100 x g. An additional 60 μl 5% ACN/0.5% FA mixture was added to filter and centrifuged. The effluent was then analyzed using the LC-MS/MS technique.

2.5. Protein Database and Database Search Engines

A protein database was constructed in a FASTA format using the annotated bacterial proteome sequences derived from fully sequenced chromosomes of 881 bacteria, including their sequenced plasmids (as of April 2009). A PERL program (http://www.activestate.com/Products/ActivePerl) was written to automatically download these sequences from the National Institutes of Health National Center for Biotechnology (NCBI) site (http://www.ncbi.nlm.nih.gov). Each database protein sequence was supplemented with information about a source organism and a genomic position of the respective ORF embedded into a header line. The database of bacterial proteomes was constructed by translating putative protein-coding genes and consists of tens of millions of amino acid sequences of potential tryptic peptides obtained by the *in silico* digestion of all proteins (assuming up to two missed cleavages).

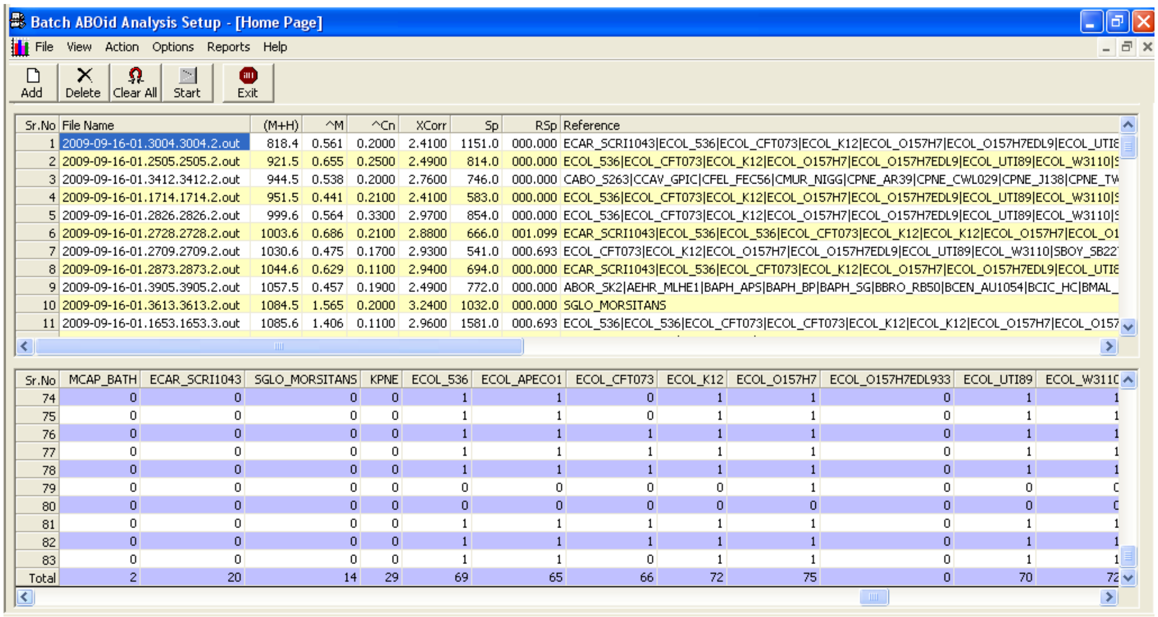
The experimental MS/MS spectral data of bacterial peptides were searched using SEQUEST algorithm against a constructed proteome database of microorganisms. The SEQUEST thresholds for searching the product ion mass spectra of peptides were Xcorr, deltaCn, Sp, RSp and deltaMpep. These parameters provided a uniform matching score of all candidate peptides. The generated outfiles of these candidate peptides were then validated using a peptide prophet algorithm. Peptide sequences with a probability score of 95% and higher were retained in the dataset and used to generate a binary matrix of sequence-to-bacterium assignments. The binary matrix assignment was populated by matching the peptides with corresponding proteins in the database and assigning a score or 1. A score of zero was assigned for a non match. The column in the binary matrix represents the proteome of a given bacterium and each row represents a tryptic peptide sequence from the LC-MS/MS analysis. Microorganisms were matched with the bacterium/bacteria based on the number of unique peptides that remained after further filtering of degenerate peptides from the binary matrix. Verification of the classification and identification of candidate microorganisms was performed through hierarchical clustering analysis and taxonomic classification.

The in-house developed software called “BACid” transformed results of searching MS/MS spectra of peptide ions against a custom protein database which was downloaded from NCBI with commercial software SEQUEST into a taxonomically meaningful and easy to interpret output. It calculated probabilities that peptide sequence assignment to a MS/MS spectrum was correct and used accepted spectrum–to-sequence matches to generate a sequence-to-bacterium (STB) binary matrix of assignments. Validated peptide sequences, differentially present or absent in various strains (STB matrices) were visualized as assignment bitmaps and analyzed by a BACid module that used phylogenetic relationships among bacterial species as a part of decision tree process. The bacterial classification and identification algorithm used assignments of organisms to taxonomic groups (phylogenetic classification) based on an organized scheme that begins at the phylum level and follows through classes, orders, families and genus down to strain level. BACid was developed in-house using PERL, MATLAB and Microsoft Visual Basic.

3. RESULTS AND DISCUSSION

The current project characterized and identified pathogenic and non-pathogenic strains of the same organism based on proteins present in whole cell lysates (global) versus OMP preparations (specific). All results are shown and discussed below. *B. anthracis* Ames and Sterne strains were also included to expand the project; however, analysis of whole cell lysates only were included and results discussed below.

Figure 1 below serves as an example to illustrate the typical output generated for the LC-ESI MS/MS analyses of bacterial proteins digest using bioinformatics tools to process the peptide sequence information for the bacterial differentiation and classification. The top window lists the identified unique proteins and their corresponding bacterium match. The lower window represents the binary matrix of the sequence-to-bacterium search matching. The total row, lower window, represents the total number of unique proteins identified for a given bacterium. Figure 2 also serves as an example and shows the histogram generated by plotting the number of unique proteins versus the bacterium matching in the database. The Y-axis represents the percentage of unique peptides matched with 95% confidence level for all the bacteria on the x-axis. In this example case, the identified bacterium at strain level is *Y. pestis* CO92. The horizontal redline is the threshold cutoff under which common degenerate peptides among various bacteria within the constructed proteome database are shown. These degenerate peptides are removed from the total number of unique peptides of the identified species.



Sr.No	File Name	(M+H)	^M	^Cn	XCorr	Sp	RSp	Reference
1	2009-09-16-01.3004.3004.2.out	818.4	0.561	0.2000	2.4100	1151.0	000.000	ECAR_SCRI1043|ECOL_536|ECOL_CFT073|ECOL_K12|ECOL_O157H7|ECOL_O157H7EDL9|ECOL_UTI
2	2009-09-16-01.2505.2505.2.out	921.5	0.655	0.2500	2.4900	814.0	000.000	ECOL_536|ECOL_CFT073|ECOL_K12|ECOL_O157H7|ECOL_O157H7EDL9|ECOL_UTI89|ECOL_W3110|
3	2009-09-16-01.3412.3412.2.out	944.5	0.538	0.2000	2.7600	746.0	000.000	CABO_S263|CCAV_GPIC|CFEL_FEC56|CMUR_NIGG|CPNE_AR39|CPNE_CWL029|CPNE_J138|CPNE_T
4	2009-09-16-01.1714.1714.2.out	951.5	0.441	0.2100	2.4100	583.0	000.000	ECOL_536|ECOL_CFT073|ECOL_K12|ECOL_O157H7|ECOL_O157H7EDL9|ECOL_UTI89|ECOL_W3110|
5	2009-09-16-01.2826.2826.2.out	999.6	0.564	0.3300	2.9700	854.0	000.000	ECOL_536|ECOL_CFT073|ECOL_K12|ECOL_O157H7|ECOL_O157H7EDL9|ECOL_UTI89|ECOL_W3110|
6	2009-09-16-01.2728.2728.2.out	1003.6	0.686	0.2100	2.8800	666.0	001.099	ECAR_SCRI1043|ECOL_536|ECOL_536|ECOL_CFT073|ECOL_K12|ECOL_K12|ECOL_O157H7|ECOL_O1
7	2009-09-16-01.2709.2709.2.out	1030.6	0.475	0.1700	2.9300	541.0	000.693	ECOL_CFT073|ECOL_K12|ECOL_O157H7|ECOL_O157H7EDL9|ECOL_UTI89|ECOL_W3110|SBOY_SB22
8	2009-09-16-01.2873.2873.2.out	1044.6	0.629	0.1100	2.9400	694.0	000.000	ECAR_SCRI1043|ECOL_536|ECOL_CFT073|ECOL_K12|ECOL_O157H7|ECOL_O157H7EDL9|ECOL_UTI
9	2009-09-16-01.3905.3905.2.out	1057.5	0.457	0.1900	2.4900	772.0	000.000	ABOR_SK2|AEHR_MLHE1|BAPH_APS|BAPH_BP|BAPH_SG|BBRO_RB50|BCEN_AU1054|BCIC_HC|BMAL_
10	2009-09-16-01.3613.3613.2.out	1084.5	1.565	0.2000	3.2400	1032.0	000.000	SGLO_MORSITANS
11	2009-09-16-01.1653.1653.3.out	1085.6	1.406	0.1100	2.9600	1581.0	000.693	ECOL_536|ECOL_536|ECOL_CFT073|ECOL_CFT073|ECOL_K12|ECOL_K12|ECOL_O157H7|ECOL_O157

Sr.No	MCAP_BATH	ECAR_SCRI1043	SGLO_MORSITANS	KPNE	ECOL_536	ECOL_APECO1	ECOL_CFT073	ECOL_K12	ECOL_O157H7	ECOL_O157H7EDL933	ECOL_UTI89	ECOL_W311
74	0	0	0	0	1	1	0	1	1	0	1	1
75	0	0	0	0	1	1	0	1	1	0	1	1
76	0	0	0	0	1	1	1	1	1	0	1	1
77	0	0	0	0	1	1	1	1	1	0	1	1
78	0	0	0	0	1	1	1	1	1	0	1	1
79	0	0	0	0	0	0	0	0	1	0	0	0
80	0	0	0	0	0	0	0	0	0	0	0	0
81	0	0	0	0	1	1	1	1	1	0	1	1
82	0	0	0	0	1	1	1	1	1	0	1	1
83	0	0	0	0	1	1	0	1	1	0	1	1
Total	2	20	14	29	69	65	66	72	75	0	70	72

Figure 1. The MS-based proteomic approach output. The upper section represents the matching algorithm results of the identified tryptic peptides resulted from the LC-MS/MS analysis. Lower section represents the binary matrix of sequence-to-bacterium scoring. Presence of a unique peptide corresponding with a protein in the given proteome of a bacterium is scored 1, non match score 0.

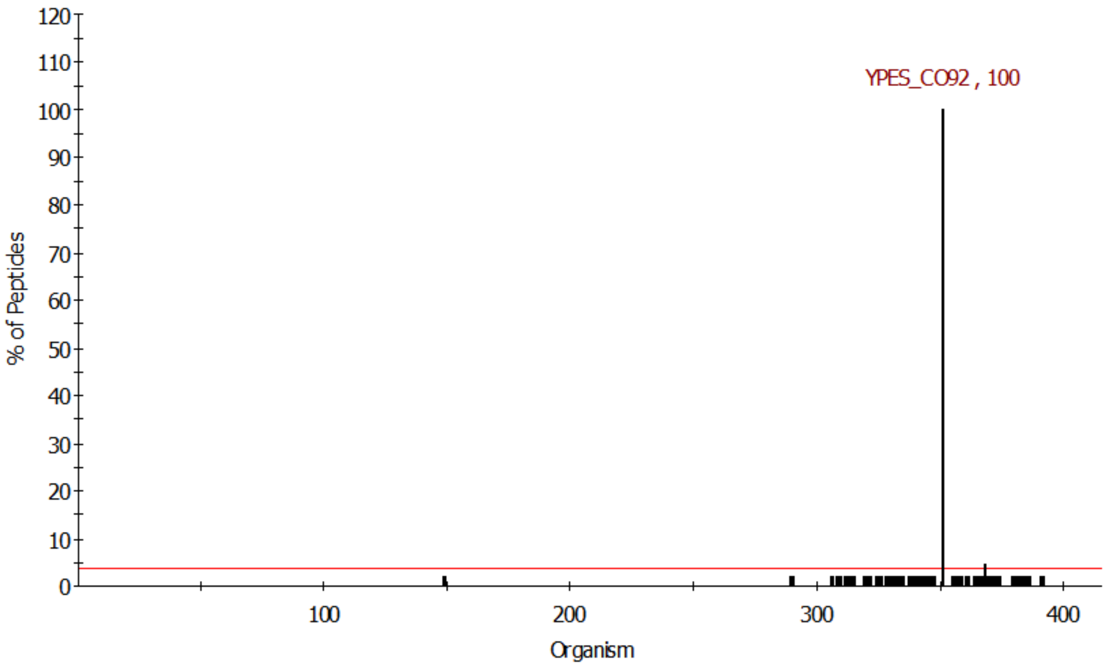


Figure 2. Histogram representing the output of the binary matrix of the unique peptides identified for a given bacterium at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non significant.

3.1. Differentiation of Pathogenic vs. Non-pathogenic E. coli Strains Using Whole Cell Lysates

Whole cell lysates of pathogenic and non-pathogenic *E. coli* strains, *E. coli* O157:H7 and *E. coli* K12, respectively, were prepared and analyzed by proteomic mass spectrometry as described above. Results showed correct identification at the strain level for both samples analyzed. The near neighbor analysis, using Euclidean distance linkage approach, for these lysed bacterial samples showed that the identified unique set of proteins had the closest match with the employed *E. coli* strains. Therefore, correct identification to the strain level was achieved for both bacterial whole cell lysates (Figures 3 a-b).

Figure 3a shows correct identification of one sample as *E. coli* O157:H7, with the next near neighbor being *E. coli* UTI89, the causative agent of human urinary tract infections. Although *E. coli* UTI89 is closely related to *E. coli* O157:H7, it is missing certain proteins such as the BAA35715 outer membrane and flagella related proteins that are distinctly expressed in *E. coli* O157:H7, but not in *E. coli* UTI189. Morevoer, the analyzed sample of the non-pathogenic *E. coli* K12, shown in figure 3b, was correctly identified as *E. coli* K12, yet had equal similarity with *E. coli* W3110, which is a nonpathogenic strain of *E. coli* genetically very closely related to *E. coli* K12.

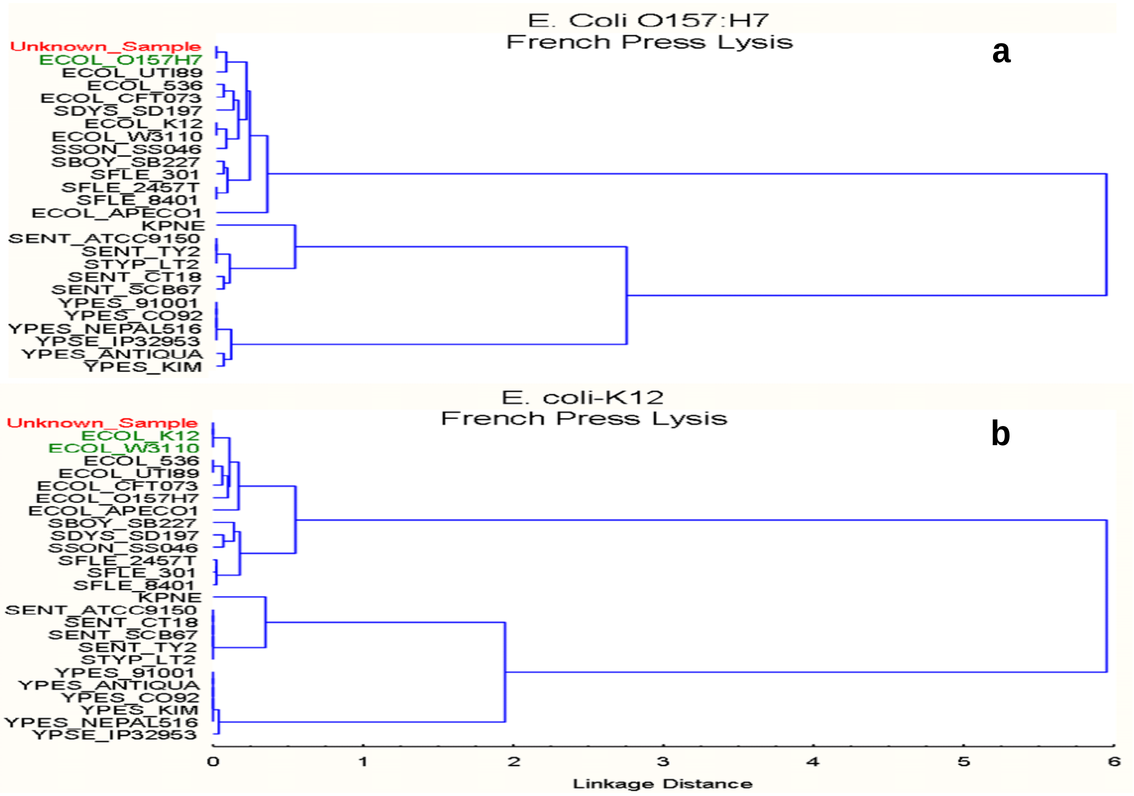


Figure 3 a-b. Near-neighbor classification of pathogenic *E. coli* O157:H7 (Figure 3a) vs. non-pathogenic *E. coli* K12 (Figure 3b) using whole cell lysates.

The number of unique proteins identified differed between the pathogenic and non-pathogenic *E. coli* strains, with the pathogenic strain having a relatively lower number of unique proteins (114) than that of non-pathogeic *E. coli* (139). Among the unique proteins of *E. coli* K12, most were shared with *E. coli* O157:H7; however, the few numbers of proteins that were not shared could potentially be used for strain level discrmination. The difference in number of unique proteins could be explained by the fact that *E. coli* K12 has been extensively studied more than any other *E. coli* strain and therefore has more genetic and biochemical information available to serve as a foundation for interpreting proteome sequences from other strains (2). This difference in the number of unique proteins between the two mentioned strains probably contributed to the difference in the similarity scoring for each respective strain as shown in Table 1.

Table 1. Unique Proteins in *E. coli* O157:H7 vs. *E. coli* K12 (Potential Biomarkers).

Protein Accesion #	Protein Info	***E. coli-K12 sub. HB101***	***E. coli-O157:H7***
NP_418358.1	stress-induced protein	✓	✗
NP_417795.1	bacterioferritin, iron storage and detoxification protein	✓	✗
YP_671573.1	putative cytoplasmic protein	✓	✗
NP_415386.1	lipoprotein	✓	✗
NP_755058.1	GnsAGnsB family protein	✓	✗
NP_668903.1	Chorismate synthase	✓	✗
YP_670276.1	Hypothetical protein	✓	✗
YP_669714.1	Aspartyl-tRNA synthetase	✓	✗
NP_312864	two-component sensor protein related to pathogenicity islands	✗	✓
NP_310689.1	Structural flagella protein	✗	✓
NP_290256	Secreted protein EspA, RELATED TO PATHOGENICITY ISLANDS	✗	✓
BAA35715	Outer memebrane protein	✗	✓
NP_286049	putative beta-barrel outer membrane protein	✗	✓

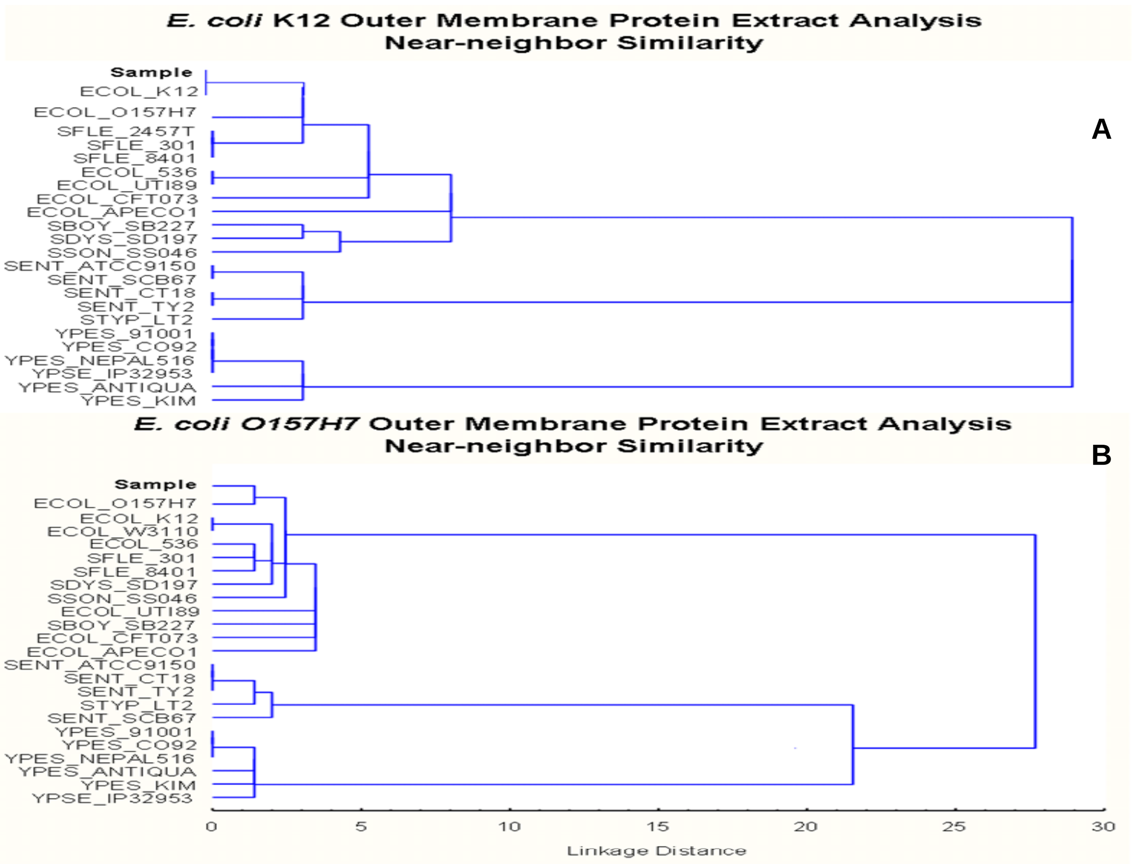
3.2. Differentiation of Pathogenic vs. Non-pathogenic E. coli Strains Using OMPs

The results of using OMPs as biomarkers for bacterial differentiation of pathogenic versus non-pathogenic *E. coli* strains are shown in figures 4a-b. Each *E. coli* strain was correctly identified with no near-neighbor strains sharing the strain level identification. Figure 4a shows the resultant near-neighbor similarity linkage analysis for OMP extracts from *E. coli* K12. The OMP extract resulted in a unique set of protein biomarkers that are capable of enhancing the differentiation at the strain level and resulted in complete similarity with *E. coli* K12 strain. No ambiguity was observed in the identification unlike that experienced when using the whole cell lysates in which an equal classification was shared between *E. coli* K12 and *E. coli* W3110 strains (Figure 3a). Although *E. coli* K12 and *E. coli* W3110 strains are genetically indistinguishable and their protein content appears very similar when analyzing whole lysates, a distinct difference was observed (significant dissimilarity) between the two closely related strains when using OMP extracts.

Figure 4b shows the near-neighbor similarity linkage results for the OMP extract of *E. coli* O157:H7. Better discrimination of *E. coli* O157:H7 was achieved using OMP extracts than that observed using whole cell lysates. The number of unique OMPs that could be identified was greater in the OMP extracts analyzed than that observed with whole cell lysates. However, this does not imply the absence of these OMPs from the whole cell lysate. Rather it is likely a higher abundance of non-OMPs in the whole cell lysate that is suppressing the detection of the OMPs in the whole lysate extracts. MS analysis has been reported to suffer ionization suppression due to the presence of large numbers of ionizable species. Generally, the whole cell lysate has a larger number of ionizable peptides and greater abundance of non-outer membrane tryptic peptides than that of OMP extracts and therefore is highly likely to experience ionization suppression during MS analysis.

Comparison of the whole cell and OMP extracts from *E. coli* O157:H7 showed distinct differences in the nature of the identified unique protein biomarkers. The whole cell lysate for *E. coli* O157:H7 had unique proteins it shared with its genetically closest strain, *E. coli* UTI89. However this was not the case when comparing the unique proteins for these two strains using the OMP extracts. For the OMP extracts, the difference in the number of strain unique protein biomarkers increased as compared with that of the whole lysate analysis.

Figure 4 a-b. Near-neighbor classification of non-pathogenic *E. coli* K12 (Figure 4a) vs. pathogenic *E. coli* O157:H7 (Figure 4b) using OMP extracts.



3.3. Differentiation of Pathogenic vs. Non-Pathogenic Y. pestis Strains Using Whole Cell Lysates and OMPs

Comparison of proteins present in whole cell lysates and OMP extracts of pathogenic versus non-pathogenic *Y. pestis* CO92 and *Y. pestis* A1122, respectively, was performed. Figure 5 a-b shows the near-neighbor analysis, using Euclidean distance linkage approach, for the bacterial identification based on OMP extracts. The identified unique sets of proteins had the closest match with the employed *Y. pestis* strains. However, it should be mentioned that the *Y. pestis* A1122 strain is not included in the current database due to the fact that its genome is not fully sequenced and publicly available. However, the constructed proteome database does include all the pathogenic *Y. pestis* strains that are listed as pathogenic strains in the DoD classification. Figure 5a. shows the dendogram for the identification of the avirulent *Y. pestis* A1122 sample. The result showed that this strain was identified at the strain level as *Y.*

pestis 91001, which is also an avirulent strain. This finding is encouraging since *Y. pestis* 91001 is the only avirulent strain in the proteome database among seven pathogenic *Y. pestis* strains currently included. Therefore, the absence of *Y. pestis* strain A1122 from the database provided an indirect test of the robustness of the proteomic approach in the classification of non-database bacteria. This also provides additional confidence in our findings in which identification at the species level was correct (Figure 5a-b). Based on these results, the unique sets of proteins for *Y. pestis* A1122 mostly resemble those found in the identified avirulent *Y. pestis* 91001 strain.

Figure 5b shows the identification result for the OMP extracts from *Y. pestis* CO92. This figure indicates a correct strain level identification of the studied sample. A closer look at the set of the unique protein biomarkers for *Y. pestis* CO92, shows the presence of biomarkers associated with virulence factors for *Y. pestis* CO92 strain. For example, proteins encoded by virulence plasmids in *Y. pestis* such as pPCP1 that encodes for plasminogen activator, pCD1 that encodes for low-calcium response and pMT1 that encodes for murine toxin, the structural gene for fraction 1 protein capsule, were present. The latter protein was present in higher abundance than that of the other mentioned protein biomarkers.

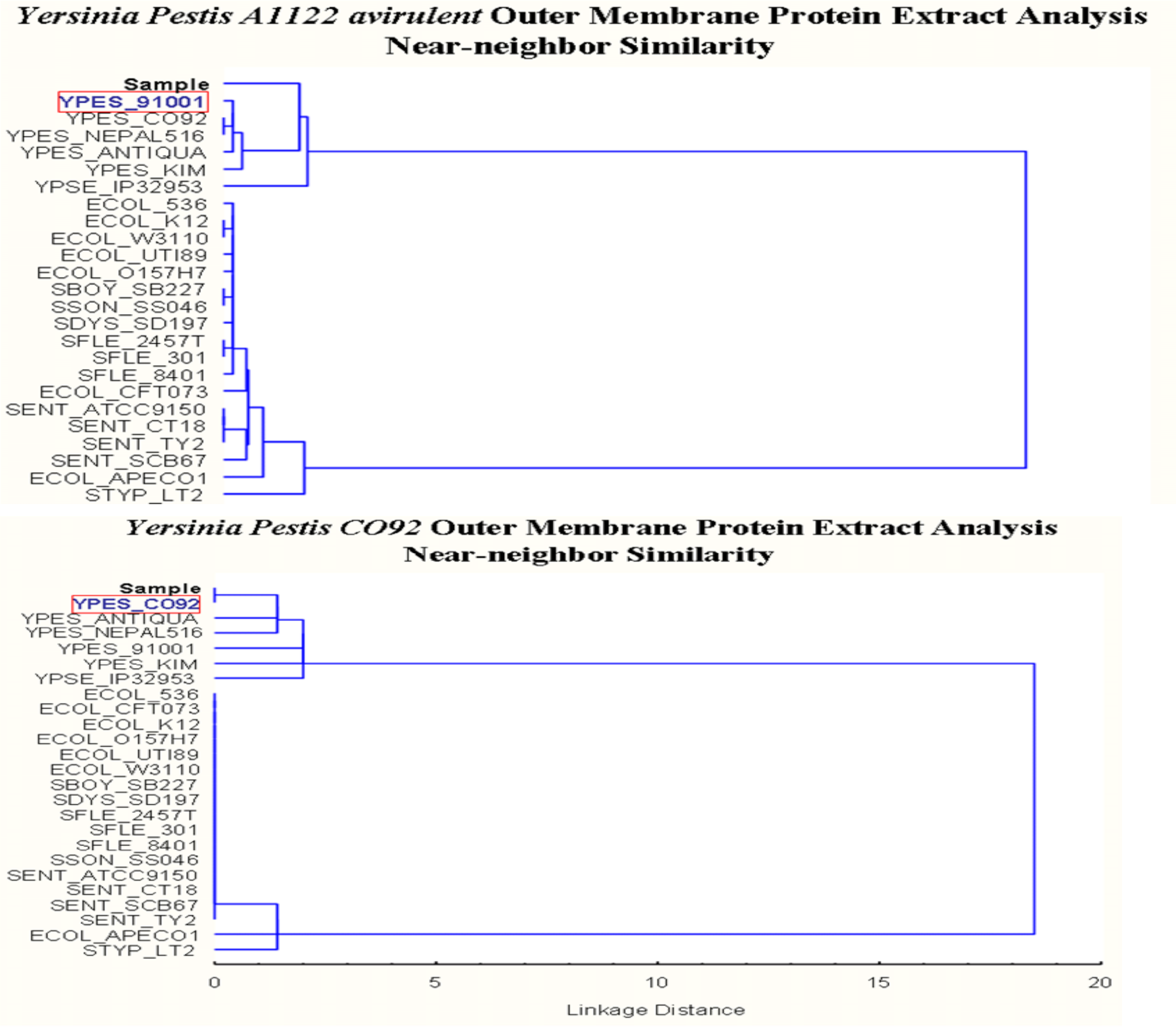


Figure 5 a-b. Near-neighbor classification of non-pathogenic *Y. pestis* A1122 (Figure 5a) vs. pathogenic *Y. pestis* CO92 (Figure 5b) using OMP extracts.

Comparing the number of unique proteins for the employed *Y. pestis* strains showed a difference between *Y. pestis* CO92 and *Y. pestis*. A1122. The former strain had 191 unique proteins versus 89 for the latter. Upon removing the highly conserved, house-keeping and energy transfer proteins from both sets, the number of strain unique proteins for *Y. pestis* CO92 was higher than that for *Y. pestis* A1122. The protein biomarkers that were observed for *Y. pestis* CO92 versus *Y. pestis* A1122 were present upon replicate analyses of the OMP extracts under different sample preparation conditions and instrumental analyses parameters.

Table 2 shows the comparison of the strain unique proteins for *Y. pestis* identified from the different cellular extracts. Also comparing whole cell lysate versus OMP extract showed a variation in the number of strain unique protein biomarkers in OMPs versus whole cell lysates. The number of strain unique proteins was slightly higher for pathogenic *Y. pestis* CO92 from OMPs extracts versus whole cell lysate. Few unique biomarkers were shared and virulence factors seemed to be present in higher abundance in the OMP extracts than in the whole cell lysates, which is in support of reported literature. (13).

Table 2. Unique Proteins for *Y. pestis* CO92 from Whole Cell Lysate vs. OMP Extracts

Y. pestis CO92 Unique proteins-Whole Cell Extract	*Y. pestis CO92* Unique proteins-OMPs Extracts
30S ribosomal protein S6	50S ribosomal protein L5
50S ribosomal protein L32	murine toxin
Acid-induced glycyl radical enzyme	attachment invasion locus protein
cationic 19 kDa outer membrane protein p	elongation factor Ts
cationic 19 kDa outer membrane protein pr	elongation factor G
chorismate mutase	putative outer membrane porin A protein
DNA-binding protein HU-alpha	hypothetical protein plu4065
ferritin	major outer membrane lipoprotein
hypothetical protein YP_0808	putative outer membrane porin A protein
hypothetical protein YP_1194	putative outer membrane porin A protein
hypothetical protein YP_1779	Acid-induced glycyl radical enzyme
hypothetical protein YP_1779	attachment invasion locus protein
hypothetical protein YP_pCD78	hypothetical protein y0163
inorganic pyrophosphatase	chaperonin GroEL
major outer membrane lipoprotein	phosphopentomutase
outer membrane protein X	plasminogen activator
periplasmic maltose-binding protein	enolase
Protein	Protein
PTS system, mannose-specific IIAB compone	outer membrane protein X
putative cystine-binding periplasmic pro	putative lipoprotein
putative lipoprotein	elongation factor Ts
putative lipoprotein	30S ribosomal protein S2
putative oxidoreductase	cationic 19 kDa outer membrane protein pr
secreted thiol:disulfide interchange prot	outer membrane protein 1b (Ib;c)
thiol peroxidase	attachment invasion locus protein
trigger factor	50S ribosomal protein L9
urease (urea amidohydrolase) beta subunit	attachment invasion locus protein
urease (urea amidohydrolase) beta subunit	attachment invasion locus protein
	hypothetical protein y2159
	hypothetical protein YP_3210
	malate dehydrogenase
	manganese superoxide dismutase

3.4. Differentiation of Pathogenic vs. Non-pathogenic B. anthracis Strains Using Whole Cell Lysates

Pathogenic and non-pathogenic *B. anthracis* Ames and Sterne, respectively, were analyzed by proteomic mass spectrometry for identification. Figure 6a-b shows the histogram for the sequence-to-bacterium binary matrix, with the number of unique peptides on the y-axis and bacterium proteome on the x-axis. As seen in this figure, correct identification of each strain was made, but with a higher confidence level for *B. anthracis* Ames than *B. anthracis* Sterne strain. This observation could likely be attributed to the fact that *B. anthracis* Ames has an additional plasmid lacking in *B. anthracis* Sterne strain (pX02) and therefore is not an expressed protein biomarker detectable in the Ms-Proteomic analysis for that of *B. anthracis* Sterne. In figure 6b, the presence of *B. anthracis* Ames slightly above the threshold cutoff of 95% confidence level supports such an observation. This was not the case with *B. anthracis* Ames samples where a distinct identification unadulterated by the presence of other *B. anthracis* strains was observed. The strain unique peptides observed with both *B. anthracis* strains are an indication of the application spectrum of such an approach. Although *B. anthracis* strains do not possess OMPs, using their whole cell lysate was sufficient to reveal the discrimination power of the MS-based proteomic approach. It is evidence that using higher concentrations and optimization of the lysis protocols could enhance the MS-based proteomic analysis and provide a utility for *Bacillus* differentiation at the strain level.

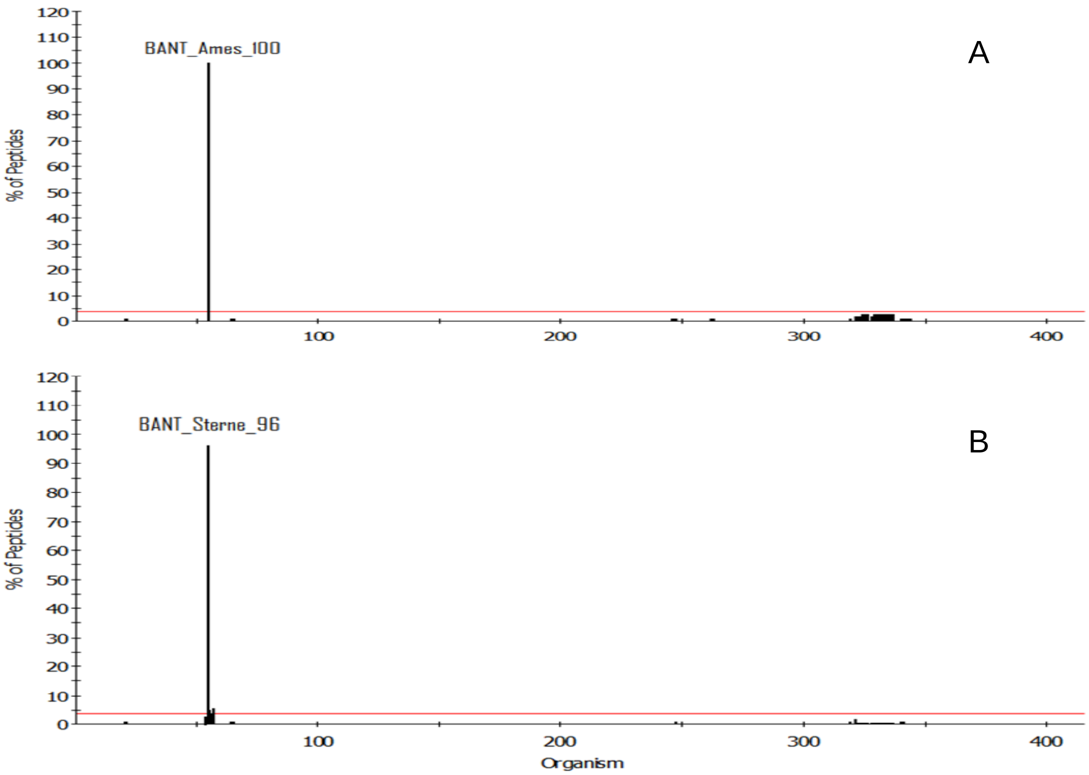


Figure 6 a-b. Bacterial differentiation of *B. anthracis* strains using whole cell lysates. Figure 6a represents the identification of *B. anthracis* Ames strain, while figure 6b represents the identification of *B. anthracis* Sterne strain. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

Further work should be conducted to investigate the strain unique peptides for *B. anthracis* strains in order to understand their characterization in a biologically meaningful medium. Doing so will increase our knowledge of the set of functional proteins responsible for strain virulence. Determining their corresponding set of genes that can be biological manipulated under different environmental conditions will be of great interest to measure the validity of strain differentiation.

4. CONCLUSIONS

This project revealed the advantage of using OMPs as unique biomarkers for bacterial differentiation of pathogenic versus nonpathogenic strains. The differentiation capability enhanced the confidence level of the discrimination process through the utilization of OMPs as biomarkers. OMPs provide a unique source of cellular variability and thus introduce biodiversity among cellular proteins for very similar bacterial strains and thereby provide distinct and unique protein biomarkers. The whole cell lysates did provide discrimination; however, the possible ionization suppression could shield the detection of important peptides that could be classified as unique biomarkers. On the other hand, whole cell lysates are an appropriate option for the differentiation of gram-positive bacterial strains and the reported results herein support their potential application in differentiation. Overall, an extension of this project to include a wider investigation of relevant pathogenic bacteria such as *Francisella tularensis, Burkholderia spp.* and other relevant strains could provide us with a more global outlook of the importance of OMPs with regard to pathogenicity and how we can confidently identify organisms at the strain level using protein biomarkers for bacterial classification and diagnostic purposes.

ACKNOWLEDGEMENTS

This work was funded by the 2009 U. S. Army Edgewood Chemical Biological Center's In-house Laboratory Independent Research program.

REFERENCES

[1] Ben-Gurion, R., et al. 1981. Plasmid 5: 183-187.
[2] Blattner F. R., et al. 1997. Science. 277:1453–1474.
[3] Cornelis, G. R. 2002. J. Cell Biol. 158:401-408.
[4] Dworzanski, J. P., et al. 2006. J. Proteome Res. 5: 76.
[5] Ferber, D. M., et al. 1981. Infect. Immun. 31:839-841.
[6] Filippov, A A. 1990. FEMS Microbiol Lett 67: 45-48.
[7] Garmendia, J., et al. 2005. Infect. Immun. 73:2573-2585.
[8] Ide, T., et al. 2001. Cell. Microbiol. 3:669-679.
[9] McDaniel, T. K., et al. 1997. Mol. Microbiol. 23:399-407.
[10] McDaniel, T. K., et al. 1995. PNAS. 92:1664-1668.
[11] Ramamurthi, K. S., et al. 2002. Annu. Rev. Cell Dev. Biol. 18:107-133.
[12] Titball, R. W., et al. 2003. Biochem. Soc. Trans. 31: 104-107.
[13] Tsukano, et al. 1996. Microbiol. Immunol. 40: 773-775.
[14] Wachter, C., et al. 1999. Mol. Microbiol. 31:1695-1702.

Electrospun Fiber Mats for Surface-Enhanced Infrared Spectroscopy

Alan C. Samuels[a], Bikas Vaidya[b], Ulf Drechsler[b],
Barry R. Williams[c], Ronald W. Miles, Jr.[a], and Melissa Hulet[c]
[a]Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD, USA 21010
[b]Lynntech, Inc., 7607 Eastmark Drive, Suite 102, College Station, TX, USA 77840
[c]Science Applications International Corp., Edgewood, MD, USA 21010

ABSTRACT

Infrared spectroscopy is routinely employed for the identification of organic molecules and, more recently, for the classification of biological materials. It is one of the only analytical techniques capable of analyzing matter regardless of its physical state (solid, liquid, or gas). We describe a novel sampling approach and proof of concept experimental result that demonstrates the collection and concentration of airborne pollutants into the interrogation volume of a Fourier Transform InfraRed (FTIR) spectrometer operating in diffuse reflectance (DR) mode. The approach is expected to be well suited to the detection and identification of vapor phase chemical agents, toxic industrial chemicals, and toxic industrial materials whether they present as gaseous vapors, liquid droplets, or solid aerosol particles.

1. INTRODUCTION

We describe a novel approach toward the sensing of trace organic vapors in air using a solid phase extraction (SPE) method coupled with diffuse reflectance infrared spectroscopy. Although the original intent of the proposed effort was to prepare fiber mats of this material using electrospinning techniques, the Principal Investigator (ACS) was unable to recruit the necessary graduate and/or summer student support staff to execute this material development. Instead, the PI secured nanostructured zinc sulfide powdered materials for investigation through a pre-existing partnership with a private entity (Lynntech, Inc.). This material proved adequate for an initial proof of concept for the proposed adsorptive gas analysis approach. We note that one group has already reported on the use of sorbent polymers in the extraction of airborne vapors for infrared analysis via attenuated total reflection (Bryant et al., 2007). The advantage of processing the substrates by diffuse reflection lies in the ease with which the substrates can be replaced or renewed vice the coating on an ATR crystal. Additionally, the metal sulfide is more benign than a polymer sorbent in the infrared in terms of potentially interfering absorption bands. A known concentration of dimethylmethyl phosphonate (DMMP) in nitrogen was passed over a diffuse reflectance cell containing powdered zinc sulfide specially formulated to contain high nanoscale porosity for efficient adsorption of organic compounds. The adsorption of DMMP on the surface of the powder was readily followed in real time using the diffuse reflectance measurement procedure. The spectral features of the adsorbed DMMP more closely matched those of the condensed phase infrared spectrum of the compound than of the gas phase spectrum, yet some absorption band perturbations suggest possible interactions between the adsorbate and the zinc sulfide matrix. The results demonstrate the efficacy of the zinc sulfide formulation as an adsorption substrate for organic vapor sensing, identification, and quantification using infrared diffuse reflectance spectroscopy.

2. MATERIALS AND METHODS

2.1. Samples

Powdered samples of nanostructured ZnS solid substrates were received from Lynntech, Inc. under a Cooperative Research and Development Agreement (CRADA). Details of the manufacture of these materials and their complete composition is considered intellectual property of the performer, Lynntech, Inc., and will not be described.

2.2. Vapor Generation and Measurement.

In order to develop a quantitative infrared spectral database, we have established a metrology facility in a chemical surety laboratory that allows us to carefully control the measurement environment, reproducibly and accurately generate known vapor concentrations, accurately and precisely measure the quantitative infrared spectral data and associated measurement uncertainty, and validate the accuracy of the measurement process. In addition, the basis for all our concentration data computations is the measurement of mass change, which is traceable to primary NIST standards. We also perform ancillary measurements to provide immediate feedback on any process errors that may occur during the data acquisition.

2.2.1. Vapor Generation

The first requirement in the development of a vapor phase infrared database is the reliable generation of a vapor of known concentration. Buchanan et al. developed just such a method for the quantification of agent vapor pressure properties[5]. The vapor generation device, hereafter referred to as the saturation cell, is described in detail in the vapor pressure reports, but the salient details of the system will be described here for completeness. A volume of the analyte is added to the saturation cell pictured and diagrammed in Figure 1. The cell is designed such that a high surface area ceramic wick is wetted with the analyte, and the carrier gas is forced to make three passes across the surfaces of the wick as it passes through the cell. At low flow rates, this vapor generation technique has been shown to establish a dynamic vapor-liquid equilibrium at the temperature of the saturation cell. The cell temperature is controlled using a constant temperature bath, and pure nitrogen from the boil-off of a liquid nitrogen dewar is exploited as a carrier gas and a downstream diluent. A Julabo F25 temperature control bath is employed to control the saturation cell temperature and Brooks model 5850S mass flow controllers are used to control the carrier and diluent gas flow rates.

The metered analyte vapor is passed through a multipass gas cell mounted on a Bruker IFS 66v/s (Bruker Optics, Billerica, MA) spectrometer as diagrammed in Figure 2, and the output stream from the gas cell delivered to the environmental diffuse reflection accessory of a Thermo-Nicolet SmartCollector™ Compound Parabolic Concentrator (CPC) installed in a Nexus 670 spectrometer (Thermo-Nicolet, Madison, WI). The detector used is an extended range MCT, and a Bruker proprietary software correction for nonlinearity is applied.

2.3. Diffuse Reflection Measurement.

We acquired the diffuse reflection spectra of the samples using a commercial Nicolet Nexus 670 FTIR spectrometer (Nicolet Instruments, Madison, WI) fitted with the SmartCollector™ diffuse reflectance accessory. A dual chamber environmental cell was used with the SmartCollector™ to house two samples of the powdered materials from Lynntech. The environmental cell was configured to allow us to control whether pure nitrogen or the effluent from the Bruker gas cell were flowed through each cell independently. A background single-beam spectrum was acquired using the pure powdered sample with nitrogen flowing through the cell, and then the flow through the cell was switched to the effluent flow from the Bruker gas cell. A single-beam spectrum of the material was acquired at three-minute intervals for 18 minutes, and each measurement ratioed to the blank powder to yield the diffuse reflection spectrum of the adsorbed analyte. Spectra were acquired at 4 cm^{-1} resolution taking an average of 128 scans and applying Happ-Genzel apodization. The experimental apparatus is pictured and shown schematically in Figure 3.

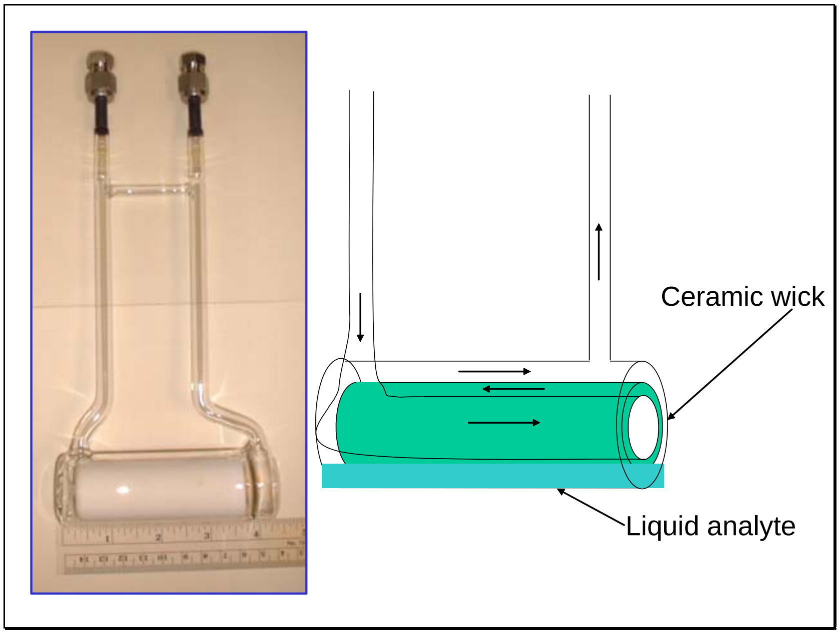


Figure 1. Saturation cell used to generate vapor-liquid equilibrium for gas phase measurements. The carrier gas is forced to make three passes over the saturated ceramic wick for a total resident time of over one minute at < 1 liter per minute flow rates. The mass change and total volume of carrier gas are used along with temperature and pressure data to compute the concentration of analyte vapor leaving the saturation cell.

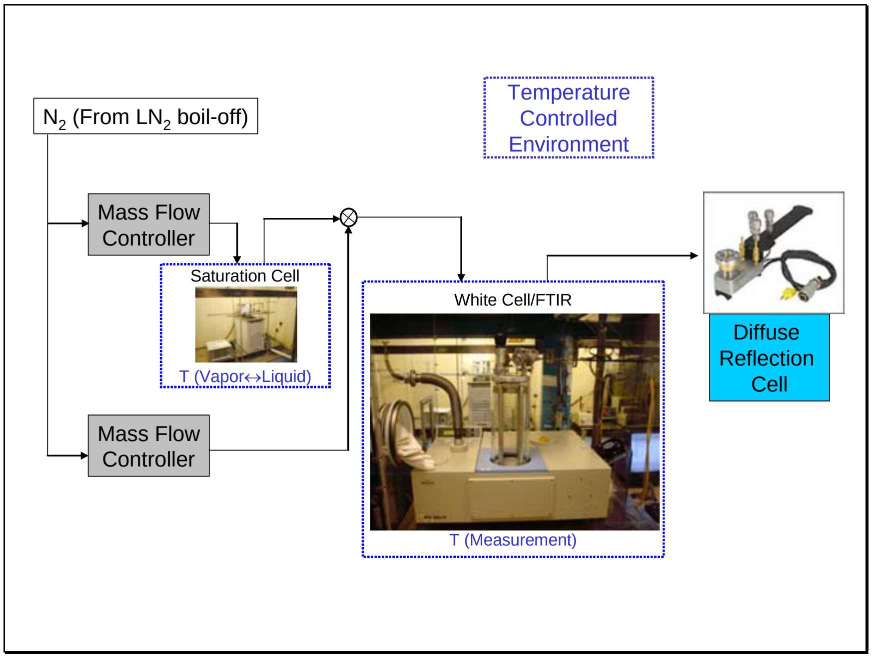


Figure 2. Experimental setup used to generate vapor phase quantitative FTIR spectra using the saturation cell. Dotted lines enclose the parts of the apparatus that are temperature-controlled.

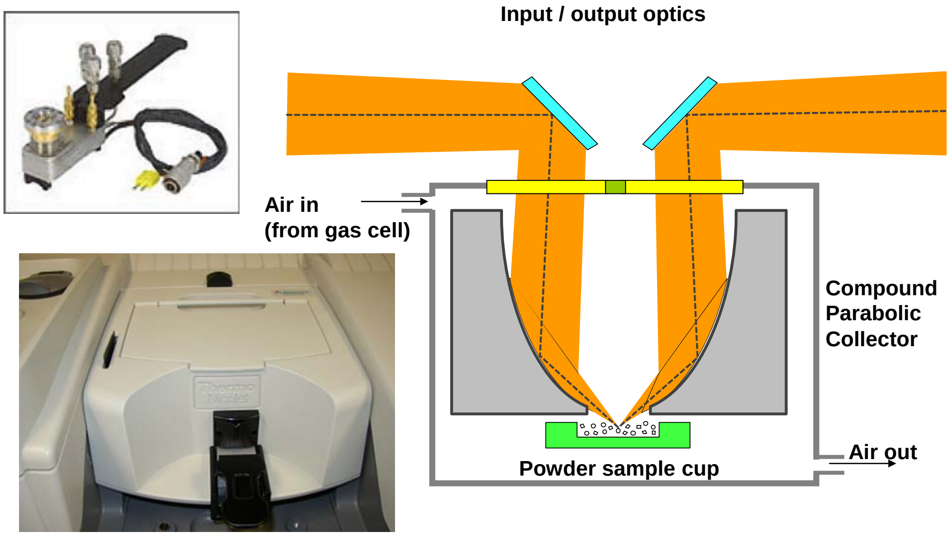


Figure 3. Experimental setup used to measure diffuse reflection spectra of adsorbed species using the SmartCollector accessory (lower left). The optics shown at the right are contained in an environmental cell (upper left) that is inserted in the SmartCollector port. The vapor laden nitrogen effluent from the Bruker white cell is passed through the environmental cell to allow the analyte to adsorbe onto the powder in the sample cup.

3. RESULTS

The concentration of DMMP vapor was calculated by comparing the observed absorption intensity of the 1050 cm^{-1} absorption feature with that of the quantitative cross-section in our spectral library, and by computing the mass loss in the saturation cell versus the total volume of nitrogen that was delivered to the apparatus by the mass flow controllers. The vapor phase spectrum is presented in Figure 4. The resulting vapor concentration is presented in Table 1 along with the computed mass deposited in the DR substrates. The time-dependent DR spectra of the adsorbed DMMP as measured by the SmartCollector apparatus is presented in Figure 5. A comparison of the liquid transmission spectrum of DMMP with that observed in the DR experiment is presented in Figure 6.

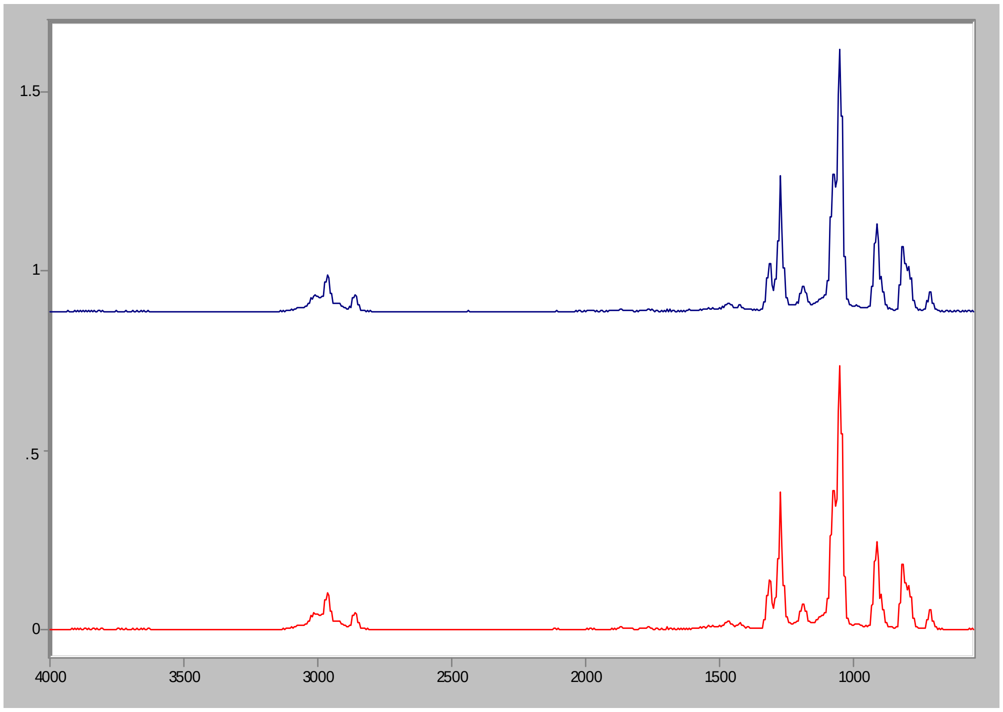


Figure 4. Red is the absorptivity coefficient of vapor phase DMMP, scaled to one of the spectra acquired during the DR experiments with the ZnS.

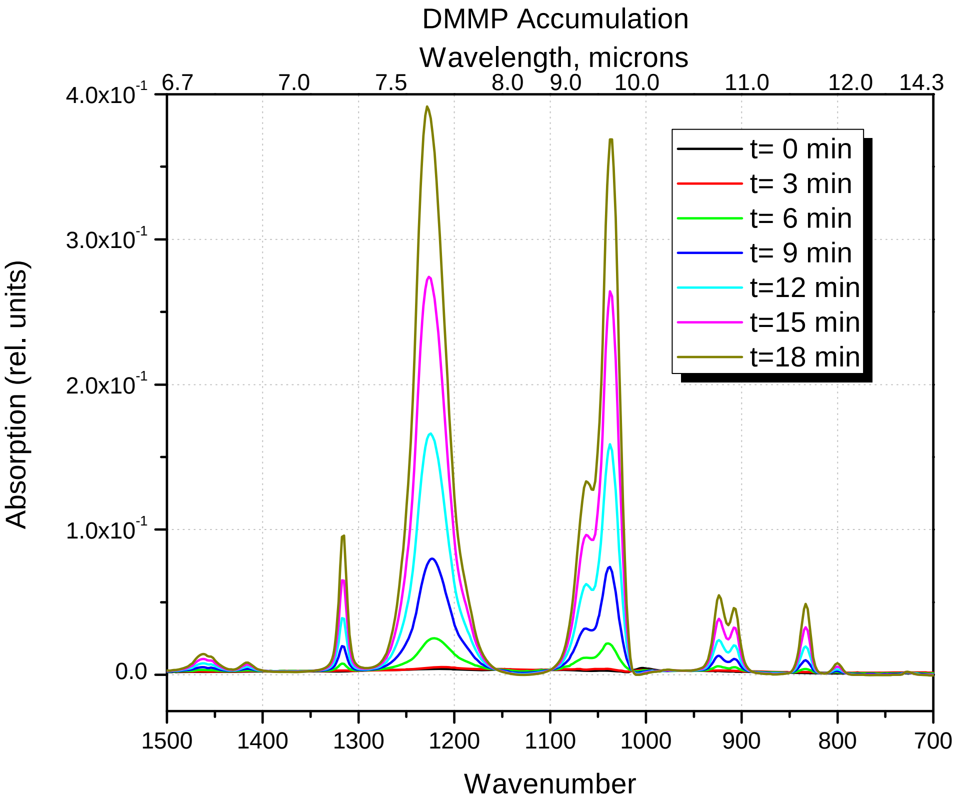


Figure 5. Diffuse reflection spectra (fingerprint region) of adsorbed DMMP on the porous ZnS substrate as a function of time of exposure to the DMMP vapor.

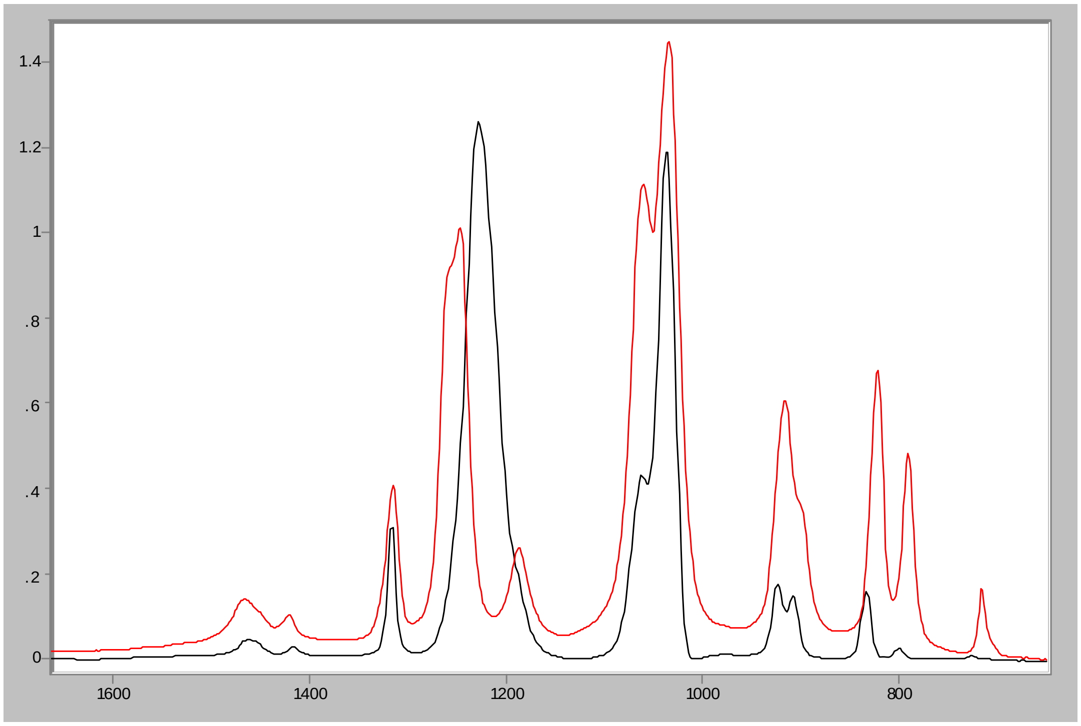


Figure 6. DMMP spectra (fingerprint region): liquid thin-film in red and diffuse reflectance from environmental chamber in black. The latter spectrum was re-scaled.

Table 1. Comparison of vapor measurements from Bruker to diffuse reflection.

alpha 1050*	A	CL	Path (m)	Conc (mg/m^3)
0.000512	0.259	505.8594	1.4045	360.2
*m^2/mg				
alpha 914	A	CL	Area (m^2)	mg
0.000177	0.0544	307.3	0.0000203	0.006239

The mass rate data from the saturator cell was very close to the calculated concentrations from the above table: 357.9 mg/min.

The total calculated mass of DMMP through the environmental chamber in the case of both the CTAB-ZnS and Pluronic-ZnS was the same, since the times and flow rates were matched: 0.258 mg. The mass load of DMMP at the time the flow to the environmental chamber stopped, (calculated from the diffuse reflectance spectra and the absorptivity coefficient of liquid DMMP) was 6.2 µg in the case of the first series (CTAB-ZnS) and 2.0 µg in the case of the second (Pluronic-ZnS). These calculated mass loads are estimates to within at most an order of magnitude. Note in Figure 6 that the intensities of the absorption bands in the two spectra are not proportional. This arises at least in part due to differences in phenomenology between absorption due to diffuse reflectance vs. transmission. In addition to the associated changes in band intensities that are known in DR versus transmission spectra, there are difficulties in determining the pathlengths of the measurements.

Also note the shifts in frequencies of several of the absorption bands in the DR vs. transmission spectrum. In the case of the P=O band, the band was lowered from 1246 to 1228 cm^{-1} in the DR spectrum (Figure 6). The shift could either be attributed to molecular bonding at the surface of the ZnS substrate, or a chemical reaction catalyzed by the surface. The C-H stretch region in the DR spectrum is much reduced in intensity relative to the other absorption features in the DMMP as compared to the thin-film spectrum. In combination with the shift to lower frequency in the phosphoryl band and the reduction in intensity in the P-O-C region, this could indicate that the DMMP has reacted chemically.

4. DISCUSSION

The concept of a sensor that would be capable of vapor, liquid aerosol, and solid aerosol sampling and analysis can potentially be realized by implementing the porous substrates described in this report as an impaction surface for an aerosol concentrator. Solid aerosol particles can be interrogated and analyzed directly by the DR apparatus as we have previously shown (Samuels et al., 2004).

5. CONCLUSION

Optical spectroscopic methods such as infrared and Raman spectroscopy provide unique spectral fingerprints of organic compounds, facilitating the unambiguous identification of CWAs and other toxic industrial chemicals; however, detection of trace vapors is a challenge for these methods. We present the product of a materials research effort aimed at developing SPE substrates that can collect and concentrate organic vapors onto a surface that is "spectroscopically benign" in that it either transmits or reflects mid infrared (radiation in the wavelength range from 2-25 microns) radiation. Such materials can be sintered or manufactured into mesoscale structures, and as demonstrated by the data presented here, are directly useful for the spectroscopic detection and analysis of volatile organic compounds in air. Additional efforts are underway to investigate the composition and manufacture of materials that exhibit both surface-enhanced infrared (SEIR) and surface-enhance Raman spectroscopy (SERS) by incorporating nanoparticulate silver and/or gold into the materials (Kellner et al., 1997).

ACKNOWLEDGMENTS

The principal author and PI would like to acknowledge the U.S. Army In-house Laboratory Independent Research (ILIR) program for supporting this initiative. The author also gratefully acknowledges the Lynntech Corporation for providing the zinc sulfide powder preparations that enabled the proof of concept study.

REFERENCES

[1] Alan C. Samuels, Diane M. Wong, Gerald J. Meyer, Geoffrey J. Roelant, Barry R. Williams, Ronald W. Miles, Jr., Christopher J. Manning, (2004)"Aerosol collection and analysis using diffuse reflectance infrared spectroscopy" Proc. SPIE Vol. 5416, Chemical and Biological Sensing V; Patrick J. Gardner; Ed., 224-232.

[2] Chet K Bryant, Peter T LaPuma, Gary L Hook, Eric J Houser. Chemical agent identification by field-based attenuated total reflectance infrared detection and solid-phase microextraction. *Analytical chemistry.* 2007 Mar 15;79(6): 2334-40

[3] Kellner, R., Mizaikoff, B., Jakusch, M., Wanzenböck, H.D, and Weissenbacher, N. (1997) Surface-enhanced vibrational spectroscopy: a new tool in chemical IR sensing? Appl. Spectrosc. 51, 495-503.

[4] Priebe, A., Fahsold, G., and Pucci, A. (2001) Surface enhanced infrared absorption of CO on smooth iron ultrathin films. Surface Science 90. 482-485.

Investigation of Molecule-Surface Interactions with Overtone Absorption Spectroscopy and Computational Methods

Jerry Cabalo
US Army Edgewood Chemical Biological Center, Aberdeen Proving Ground, MD, USA, 21010

1. INTRODUCTIONS

1.1. Objective

The energetics and dynamics of molecule-surface interactions play an integral part to chemical processes and chemical adsorption. Chemical reactions on the surface of a catalyst are a good example of these interactions.[1] A specific set of applications of molecule-surface interactions that is of interest to the warfighter is in chemical biological defense. For example, chemical personal protective equipment relies heavily on a material's ability to surface bind harmful compounds. Another type of study of interest to the Army's mission is the interaction of energetic material residues with various environmental surfaces. These interactions can have an impact on detectability of the energetic materials. A number of surface techniques have been applied to this problem, such as diffuse reflectance IR spectroscopy[2] and step scan FTIR spectroscopy.[1,3] These techniques have been effective at identifying decomposition intermediates and products even on the timescale of hundreds of nanoseconds. While these techniques are proven tools for the study of the chemical dynamics of decomposition, details of the molecule/surface interaction are still unknown. Vibrational overtones are uniquely sensitive to a molecule's local environment, and are potentially an effective tool in examining molecule/surface interactions.[4] It is the objective of this work to determine the most effective and general theoretical model of molecular-surface interactions. This will be done using vibrational overtone spectroscopy in tandem with molecular mechanics and ab initio calculations to investigate molecule-surface binding. Due to the importance of 1,3,5-trinitrotoluene (TNT) as a compound and the catalytic and optical properties of alumina, the system of TNT on an alumina surface serves as the test case for this study.

1.2. Background

The interaction of a molecule with its local environment is reflected in its vibrational spectrum, so that the vibrational spectrum can be used as an indicator of the local environment. Vibrational overtone spectroscopy is ideal for studying molecule-surface interactions because it involves higher vibrational energy levels. When the vibrational potential energy surface of a molecule is perturbed by the presence of a surface or solvent, the higher energy levels experience a more drastic change than the lower levels. This makes sense in view of the fact that at higher vibrational energy levels, the atomic displacement along the vibrational coordinate is greater, so that the molecule-environment interactions will be greater than at lower vibrational energies. These levels are not typically accessed with fundamental absorptions because almost all molecules are in the ground vibrational state at room temperature, so that optical absorption only promotes molecules to the first vibrationally excited state. This energy is quickly lost to equilibrium with the environment so that multiphoton processes are unlikely. As a result, the higher vibrational states are only accessible by overtone transitions, that is, transitions involving more than one quantum of vibrational energy. While molecules are typically modeled as harmonic oscillators, real molecules are anharmonic oscillators where transitions of more than one vibrational quantum are possible.[5] As a result, the anharmonic description of molecular vibrations is necessary for this study.

Most theoretical models use the harmonic description of molecular vibrational because it is computationally easier, and for studies of fundamental transitions is normally adequate. However, there are additional differences between the anharmonic and harmonic description of molecular vibrations that show that the harmonic description is not sufficient for this study. In the harmonic description, the potential energy surface goes to infinity at large atomic separation and does not account for the breakage of bonds. The result is that the calculated vibrational frequencies for a harmonic oscillator are integer multiples of the fundamental frequency that go up infinitely. In real anharmonic systems, the separation between energy levels tends to zero as the vibrational energy approaches the dissociation limit. This means that the overtone transition frequency will not be an integer multiple of fundamental transition frequency, and that a perturbation is likely to increase the deviation from the frequency predicted by the harmonic treatment. In other words, the harmonic oscillator description will consistently overestimate molecular fundamental and overtone vibrational frequencies. The anharmonic description is essential for adequate handling overtone vibrations.

There are several available approaches for modeling the interaction of a molecule with a surface, however, only some of these methods permit prediction of overtone frequencies. The simplest is molecular mechanics, which operates on a quasi-classical level.[6-9] The details of the electronic states are neglected and the energy of the system is treated as a sum of interactions through chemical bonds, such as bond stretches and bends, and non-bonding interactions such as electrostatic interactions and van der Waals energy. The overall description of the molecule as a sum of classical interactions is also known as a "force field". The strength of this approach is that it is not computationally expensive to perform, so that large systems can be modeled. A large assumption of this approach is that the interaction between two atoms in one molecule will be like that in another molecule. This assumption leads to two weaknesses of this approach: 1) empirical data is often necessary to build a complete force field, and 2) accuracy of the results depends on the similarity of the system under study to the systems that provided the empirical data. Predicted properties, such as vibrational spectra, are only accurate if the source of the empirical data is very much alike to the model system. While typical force fields permit use of the Morse Potential for describing bond stretches and anharmonic vibrations, the accuracy is most likely insufficient to make reliable predictions of overtone frequency shifts the can be used to infer the molecule/environment interaction.

The system can also be handled with ab initio quantum mechanical or density functional methods that do not require any empirical data. For ab initio methods, the electronic wavefunction of the system is calculated, and the vibrational potential energy surface is directly calculated from the electronic wavefunction.[7,10] For density functional methods, the square of the wavefunction is calculated and used to get the vibrational potential energy surface. For some quantum mechanical methods such as Hartree Fock, DFT, second and third order Moeller/Plesset (MP2/MP3), it is possible to obtain analytic second derivatives and numerical third derivatives of the vibrational potential energy surface. This allows more accurate anharmonic treatment of the potential energy surface, and calculation of the overtone frequencies. However, these approaches are computationally expensive, so that there is a strong limitation on the size of the theoretical model that can be handled. Fortunately, a number of theoretical modeling packages include the capability to perform hybrid calculations where part of a system is treated with quantum mechanics and another part is treated with molecular mechanics. Proper application of the hybrid treatment permits accuracy approaching that of a purely quantum mechanical method at a fraction of the computational cost.

The overtone frequency serves as the bridge between the molecule-surface interaction model and experiment.[11] A modeling approach will be considered valid if the predicted overtone frequencies accurately reproduce those measured from experiment. A number of important details in the surface/molecule interaction play a part in the simulation that will impact predicted overtone frequencies. For example, the orientation of the molecule relative to the surface will affect how its force field is perturbed, so that each possible orientation will have its own spectral shift. If no particular orientation is favored, then the observed shift in the overtone spectrum will be a convolution of transition energies of all possible orientations. If a particular orientation is favored, as in a catalytic surface, then the shift in the overtone spectrum will reflect this. Another important detail is the magnitude of the molecule-surface binding energy. The binding energy will govern transport dynamics. The binding energy is easily determined in the model by the difference between the energy of the molecule and the surface separated by infinite distance and the energy of the molecule-surface system at the optimized geometry. It is expected that a stronger binding energy will correspondingly perturb the molecular vibrations and result in more drastic shifts in the overtone absorption frequency relative to the isolated molecule.

2. EXPERIMENT

2.1. Computational Approach

Computational modeling of neat TNT and TNT molecules on an alumina surface are modeled utilizing Gaussian 2003 on a Sunstation V. This system has 32 Mbytes of RAM memory and 8 processors, although this implementation of Gaussian 2003 does not include the parallel processing code. Gaussian 2003 is a quantum mechanical modeling package capable of performing a wide range of methods such as HF, MP2/MP3/MP4, DFT, and hybrid quantum mechanical/molecular mechanical methods with the AMBER/DREIDING/UFF force fields.

The first task is to determine an appropriate level of theory for performing anharmonic frequency calculations on the TNT molecule. The criteria for selection will be computational expense and accuracy. Accuracy was determined by comparison of the calculated molecular dimensions and fundamental vibrational frequencies to published X-ray data as well as spectroscopic data. Anharmonic frequency calculations were run using HF, MP2, and B3LYP (DFT). These have available analytic second derivatives of the potential energy surface. Basis sets up to 6-311+G* were evaluated. More complex basis sets with more diffuse functions were not attempted due to computational constraints. Special attention was paid to the TNT C-H stretches because they occur at higher frequencies and the resulting overtone absorptions are accessible from wavelengths produced by the ECBC OPO laser system.

The second task in the computational component will be the determination of the optimal way to treat the surface on which the TNT molecule sits. Both neat TNT and alumina will have to be examined. For very thin films, i.e. submonolayer coverage, the predominant interaction will be between isolated TNT molecules and the substrate surface. For very thick coverage, the predominating interaction will be bulk, neat TNT. Both systems will need to be modeled. The systems were treated with pure quantum mechanics where neat TNT was treated as small clusters of TNT molecules and TNT on the alumina surface as a TNT molecule in proximity to an aluminum hydroxide molecule or small arrays of aluminum oxide. Hybrid QM:MM calculations were also attempted.

2.2. Laboratory Measurements

The goal of the experiment is to provide validation of the computational model of the molecule-surface interaction. The validation is to be provided by agreement of predicted overtone absorption frequencies to experimental measurements. The direct measurement of overtone absorption is accomplished by means of photoacoustic spectroscopy. When molecules absorb optical energy, they have an increase in internal energy and are out of equilibrium with the environment. The most efficient pathway to equilibrium with the environment is by emitting the excess energy as heat. When the heat is released into fluidic surroundings, such as air or solvent, a pressure wave is created that can be detected with an acoustic transducer.[12] The measurement is directly proportional to the absorbed energy and is completely independent of light scattering from the sample. The most important strength of this technique is that the measurement is unaffected by light scattering, so that sample morphology has little effect on the measurement of the absorption.[13] When the light source is modulated, the resulting acoustic response will have the same modulation frequency with some delay.

One problem with a photoacoustic approach that must be overcome is that overtone transitions tend to have cross sections one to two orders of magnitude less intense than the fundamental transitions. Comparison of between the methods of fundamental and overtone absorption show that this is surmountable. Although measurement of fundamental transitions occur in the mid to far IR portions of the spectrum where laboratory light sources are generally weak, useful measurements are routinely obtained in the laboratory. Overtone transitions tend to occur in the near IR and visible wavelengths, where there are more powerful light sources. The problem of weak signal is overcome with more intense light sources. In systems that are easily saturated with optical energy, the measurement is performed with additional sample to provide measurable photoacoustic signal.[14]

Absorption spectra are recorded utilizing the MTEC300, a commercial photoacoustic cell. Tunable excitation radiation is produced by a Nd:YAG pumped OPO laser system at ECBC, and introduced through a window in the cell. Because the excitation source is a pulsed laser, the acoustic signal will also be pulsed. Signal is acquired with a gated integrator to maximize signal to noise. Since the signal is not modulated, neither is a resonance condition for signal amplification required or digital signal processing. As a first shot at signal, graphite sample was measured with 100 μJ of impinging UV radiation.

Characterization of the sample is accomplished via FTIR measurements with the same cell. For reference, neat TNT samples were measured with a Perkin-Elmer near IR-Vis spectrometer, and with attenuated transmission/reflectance FTIR.

3. RESULTS

3.1. Computational Results

3.1.1. Selection of High Level Method

The overall approach is to calculate the change in the vibrational frequencies of TNT induced by the presence of a surface. The first task for the computational component will be the determination of the appropriate level of theory for treating the TNT molecule and calculating the overtone frequencies. In order to obtain the overtone frequencies of the TNT molecule, selection of the theoretical approach is reduced to methods that have analytic second derivatives of the energy with respect to nuclear position, which are Hartree/Fock (HF), Density Functional Theory (DFT), and Moeller/Plesset Theory (MP2/MP3). The anharmonic terms for the potential energy surface are calculated by taking the numerical third derivative of the energy with respect to nuclear position along the normal mode coordinate. The STO-3G, 3-21G, 6-31+G(d,p), and 6-311+G* basis sets were evaluated for geometry optimization and vibrational frequency calculation utilizing HF, MP2, and DFT, where the B3LYP hybrid functional was used for DFT. The results, such as the fundamental vibrational frequencies and optimized geometric structure, were compared to TNT FTIR spectra measured in the laboratory and published X-ray crystallographic data, respectively. The best results were obtained utilizing B3LYP and MP2 methods with the 6-31+G(d,p) basis set, where calculated bond lengths and angles compared well with the published crystal data,[15] and calculated C-H stretching vibrational frequencies compared well with the laboratory FTIR measurements (within 20 cm^{-1}). The HF level of theory produced differences of ~150 cm^{-1} between the predicted aromatic C-H stretching frequencies (mode 1 and 2 in Figure 2), and the highest frequency in the FTIR spectrum (which we assigned to the aromatic C-H stretching frequency). For the sake of comparison, conventional harmonic frequencies were computed, and were ~300 cm^{-1} higher in frequency than the experimental FTIR measurements, although these calculations had the same trend where higher levels of theory approached the value. For the first overtone frequencies the same trend was also followed where the higher levels of theory resulting in lower predicted frequencies (Table 1). Coincidentally, a small peak is visible at 6036 cm^{-1} in Figure 4, which is close to the overtone frequency for the aromatic asymmetric stretch predicted at 6161 cm^{-1}. While the B3LYP and MP2 methods proved to have equivalent accuracy in comparison to published data geometric data, the B3LYP method is the preferred method. The B3LYP calculation completed much faster than the MP2 calculation, and can be readily used with periodic boundary conditions for calculations on crystalline material (which will become important for some of the proposed calculations, especially neat TNT).

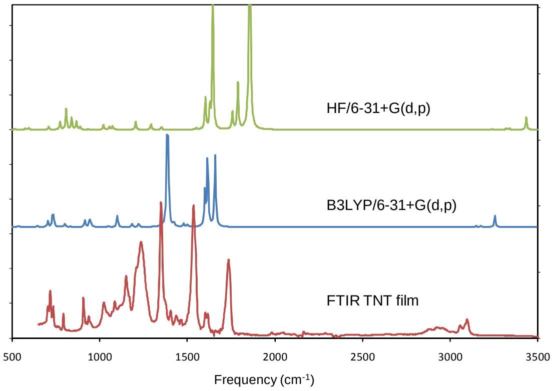
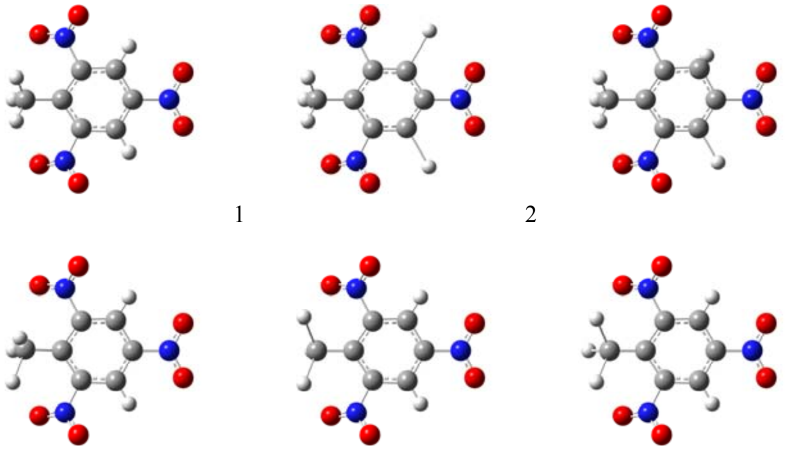


Figure 1. Comparison of calculated harmonic IR absorption spectra and a TNT film absorption spectrum measured with FTIR. The upper trace utilized the Hartree Fock Method and the middle trace utilized Density Functional Theory with the B3LYP functional. Both methods utilized the 6-31+G(d,p) basis set which has some diffuse functions. Note that the frequencies of the C-H motions between 3000-3500 cm^{-1} tend to approach the experimental value with the higher level of theory.



0 1 2

3 4 5

Figure 2. Description of vibrational motion, in order of decreasing frequency. 0) zero displacement, 1) symmetric aromatic C-H stretches, 2) anti-symmetric aromatic C-H stretches, 3) anti-symmetric aliphatic C-H stretch 1, 4) anti-symmetric aliphatic C-H stretch 2, 5) symmetric aliphatic C-H stretch.

	HF harmonic (cm^{-1})	HF anharmonic (cm^{-1})	First Overtones (cm^{-1})	DFT B3LYP harmonic (cm^{-1})	DFT B3LYP anharmonic (cm^{-1})	First Overtones (cm^{-1})
1	3434.5	3321.21	6580.28	3255.32	3110.4	6161.5
2	3434.5	3314.09	6570.02	3255.23	3114.6	6172
3	3337.9	3205.23	6349.26	3174.88	3003.4	5977.1
4	3319.6	3202.74	6349.26	3147.87	2963	5960.6
5	3240.2	3111.26	6362.47	3078.55	2963	5878

3.1.2. Molecule/Surface Modeling

There are several different approaches that have been evaluated, such as modeling with periodic boundary conditions (PBC), and hybrid molecular mechanics/quantum mechanics methods. A calculation using the PBC utility in Gaussian03 was attempted to determine the overtone absorptions in solid neat TNT. X-ray crystallographic measurements of the neat TNT crystal show an eight member unit cell and the spacing between the individual molecules.[15] The published geometry was used as a starting point for the calculation with HF and the STO-3G basis set, which is the lowest level method that can be invoked by the PBC convention. Although half of the Sunstation's resources were utilized for this calculation (4 processors and 16 GBytes of RAM memory), the calculation was terminated after 20 days without completing the geometry optimization. This approach was not feasible with the computational resources at hand.

The next attempt at modeling a cluster of TNT molecules utilized the AM1 Hamiltonian, which is a semi-empirical method. The semi-empirical methods are much less computationally expensive than ab initio methods, and were completed on a Windows desktop machine for clusters with three and eight members. The results of the geometric optimization did not compare well with published X-Ray data, and the difference between the resulting structure and the published TNT crystal structural data suggested that a very large cluster was necessary to replicate the published experimental data. The strategy had been to speed up the PBC calculation by providing a unit cell optimized with a semi-empirical method. This approach was not workable since the results showed significant deviation from the X-ray data.

To establish expected trends in the vibrational frequency, harmonic calculations were performed for a number of cluster systems utilizing purely quantum mechanical models. For the sake of computational simplicity anharmonic calculations were not attempted at this time, since this essentially doubles the time for calculating vibrational frequencies, and HF was the method used. The vibrational spectra of a TNT molecule next to another TNT molecule, of TNT next to aluminum hydroxide (an analog of the alumina surface), and of TNT next to a tetrahedral aluminum oxide/hydroxide surface, were calculated. Table 2 summarizes the vibrational frequencies for the C-H stretches (again only the C-H stretches are under discussion since the experimental setup does not permit access to the overtones of the other vibrational modes). For reference, the first column includes the C-H stretch frequencies of the lone TNT molecules. Not surprisingly for these small clusters, for hydrogens perturbed by the presence of a molecule, there were greater shifts in fundamental frequency. For example, comparing modes 1 and 2 (aromatic C-H stretches), in the unperturbed system these are practically degenerate in frequency, but when one of the aromatic hydrogens is perturbed by the presence of an external cluster, shifts of 10-30 cm-1 occurs in the fundamental absorption. As a very crude estimate of the shift in overtone frequency, it is expected that the frequency shift will be approximately a factor of two greater.

As a matter of course anharmonic frequency calculations were attempted. Although half of the computational resources were utilized with a relatively low level of theory (HF), after 30 days of operation, the calculation didn't complete after 30 days. This is most likely due to the fact that there are many more modes to be calculated since there are 3N-6 modes where N is the number of atoms. Gaussian 2003 attempts to calculate analytic third derivatives of the potential energy surface for each mode, and thus with the current computational resources and Gaussian 2003 calculations are ongoing.

Calculation utilizing the hybrid quantum mechanics/molecular mechanics (QM:MM) utility of Gaussian 2003 was also attempted utilizing combinations of HF:UFF and B3LYP:UFF. However difficulty arose when Gaussian 2003 attempted to compute anharmonic vibrational frequencies for the unified system, where accurate third derivatives of the potential energy surface couldn't be calculated for the portions of the system treated with molecular mechanics. Fortunately, Gaussian 2009 has a facility that should handle this problem. For the harmonic calculation, the system did not rapidly converge so that calculations are ongoing.

C-H stretch	HF harmonic TNT	HF harmonic TNT dimer	HF harmonic TNT/Aluminum hydroxide	HF harmonic TNT/tetrahedral aluminum oxide/ hydroxide surface
1	3434.5	3446.9	3450.9	3402.0
2	3434.5	3436.3	3433.6	3435.8
3	3337.9	3337.9	3344.9	3349.5
4	3319.6	3319.5	3320.7	3322.6
5	3240.2	3240.2	3240.8	3240.1

Table 2. C-H stretch fundamental harmonic frequencies of lone TNT molecule, TNT dimer, TNT and aluminum hydroxide cluster, and cluster of TNT and small aluminum oxide array. The most dramatic shifts occur for C-H stretches directed toward the aluminum oxide grouping. Frequencies are in units of cm^{-1}.

3.2. Experimental Results

The first phase of experimental data involved collection of TNT spectra with commercial analytical equipment. The mid IR absorption spectrum of TNT was recorded with an Attenuated Transmittance Reflectance (ATR) spectrometer with the data shown in Figure 3. This allowed the spectrum of a thin film of TNT to be recorded. Although several peaks are present in the recorded spectrum that are not visible in the published IR spectrum, there is excellent agreement for the C-H stretch frequencies around 3100 cm^{-1}. It was also attempted to view the overtone absorptions using a Perkin Elmer near IR spectrometer. The resulting spectrum is shown in Figure 4. Although a peak shows up around ~3100 cm^{-1} in this spectrum, corresponding to the aromatic C-H stretches, the signal to noise starts to become rather poor at ~5000 cm^{-1} where the light source becomes weak. There appears to be a small peak or divot in the spectrum occurring about where quantum mechanical calculations predict the overtone absorption. This may be an overtone absorption of neat TNT.

To demonstrate the feasibility of using photoacoustic detection of sample absorption, figure 5 shows the raw signal output of the MTEC300 microphone with a graphite target and 100 μJ/pulse of 280 nm laser radiation. A large signal is obtained that is 150 mV in amplitude and several ms in duration. The amount of signal collected with a low pulse energy suggests that it is feasible to collect signal from minute residues of TNT.

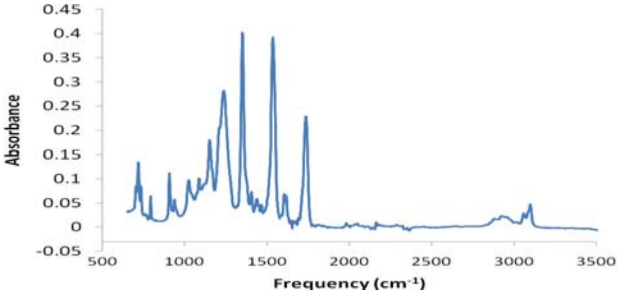


Figure 3. FTIR absorption spectrum of the fundamental vibrations of TNT.

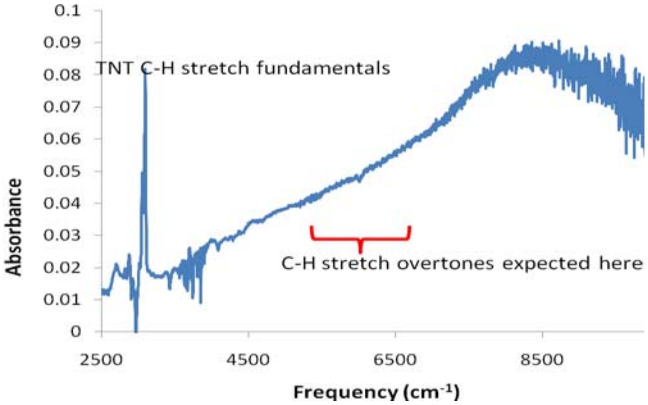


Figure 4. Near IR ATR absorption spectrum of TNT.

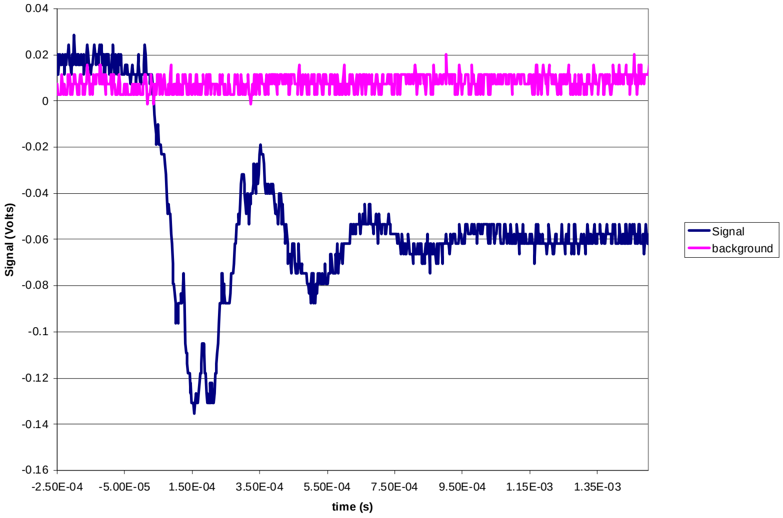


Figure 5. Raw photoacoustic signal from MTEC300 with graphite target and 100 µJ of UV pulse energy.

4. CONCLUSIONS

The preliminary results obtained from modeling the interaction of a molecule with a surface show that this is not an easy task. Of the methods attempted so far in this study, such as imposing periodic boundary conditions (PBC), modeling the system as an interaction between two (small) clusters entirely with quantum mechanics and semi-empirical methods, and with hybrid QM:MM methods, modeling the system as clusters has produced some modest results. These show that the fundamental vibrational frequencies will measurably shift in response to the presence of a physical perturbation of the molecule's environment. By extrapolation, it is expected that the overtone frequencies will experience even greater shifts in frequency in response to a disturbance.

In the course of this study, a number of hardware and software issues have imposed limitations. The primary limitation is that when Gaussian 2003 attempts to calculate the anharmonic frequencies of a system, it attempts to do this for the entire system, even for modes that are not of interest. In the study of molecular interaction with a surface, the vibrations in the surface itself are not of interest, especially when the substrate material is chosen to be transparent to the optical excitation. In fact, since the tunable laser light source used cannot access the mid IR wavelengths, only the C-H stretch overtones are of interest, and for the purposes of this study, all other modes can be neglected. Thus, the ability to select which modes to perform an anharmonic analysis can drastically reduce the computational cost. A secondary impediment is imposed by hardware limitations. Although half of the processors and memory of the Sunstation were used, anharmonic calculations of even isolated molecules can take several days. This study requires a great number of calculations (especially when considering multiple molecular orientations relative to a surface) that need to be completed in a timely manner. For the second phase of this project, access to a 32 processor cluster with a parallelized implementation of Gaussian 2009 has been obtained within our division at ECBC. With much more computational resources and an implementation of Gaussian that allows selection of only vibrational modes of interest, much more complex systems can be modeled than has been modeled to date.
Progress in the experimental area of this project has also had outside limitations. Laboratory measurements will depend on the availability of a tunable pulsed laser light source. ECBC's tunable OPO laser light source did not become available until the very end of the fiscal year.

REFERENCES

[1] Atkins, P. P. Physical Chemistry; Fourth ed.; W. H. Freeman and Company: New York, 1990.
[2] Cramer, C. J. *Essentials of Computational Chemistry*; Second ed.; John Wiley and Sons: New York, 2004.
[3] Menoza-Alvarez, K. G.; Cruz-Orea, A:; Zelaya-Angel, O.; Torres-Delgado, G.; Castanedo-Perez, R.; Hernandez, S. A.; Marquez-Marin, J. Characterization of TiO2 films for photocatalysis applications using photoacoustic spectroscopy. *Journal de Physique IV* 2004, *125*, 407-409.

[4] Morales, C. M.; Thompson, W. H. Mixed Quantum-Classical Molecular Dynamics Analysis of the Molecular-Level Mechanisms of Vibrational Frequency Shifts, *J. Phys. Chem. A* 2007, *111* (25), 5422-5433.

[5] Pipino, A. C. R.; Hoefnagels, J. P. M.; Watanabe, N. Absolute surface coverage measurement using a vibrational overtone. *Journal of Chemical Physics* 2004, *120* (6), 2879-2888.

[6] Skaare Rygh, L. E.; Gausemel, I.; Ellestad, O. H.; Klaeboe, P.; Nielsen, C.J.; Ryttere, E. Diffuse reflectance IR studies of bimetallic Fischer-Tropsch catalysts. *J. Mol. Struct.* 1995, *349*, 325-328.

[7] Zhang, P.; Urban, M. W. Photoacoustic FT-IR Depth Imaging of Polymeric Surfaces; Overcoming IR Diffraction Limits. *Langmuir* 2004, *20* (24), 10691-10699.

Pathway Signatures Analysis of Production Methods of Tetramethylenedisulfotetramine

George Hondrogiannis and David B. Cullinan
Edgewood Chemical and Biological Center Forensic Analytical Team
Research and Technology Directorate, Aberdeen Proving Ground, MD, USA 21010

ABSTRACT

The synthesis of tetramethylenedisulfotetramine (TETS) occurs via a condensation reaction between formalin and sulfamide in strong acid. TETS is a caged heteroadamantane structure with a high degree of symmetry, such that the $C_4H_8N_4S_2O_4$ structure shows only one signal each for ^1H and ^{13}C NMR. It is only slightly soluble in water, with a literature value of 0.25 mg/mL. The crude reaction products were found to be soluble in water < acetone < DMSO. When produced with a stoichiometric ratio of 2 to 1 formaldehyde to sulfamide, TETS was the principle reaction product. Analysis by GC/MS showed two compounds: one with a molecular ion of 240 amu (consistent with TETS) and the second with a molecular ion of 360 amu. Analysis by NMR showed two products in addition to TETS. One, consistent with the 360 amu product, showed a single ^{13}C NMR resonance and an AB pattern in the ^1H NMR spectrum. The second showed three resonances in the ^{13}C NMR spectrum with a 2:2:1 intensity ratio, each of which were associated with AB patterns in the ^1H spectrum. In comparison with literature findings, the 360 amu product was assigned to hexamethylenetrisulfohexamine (HEXS). No definite identification has been made for the third reaction product seen by NMR.

Keywords: Tetramethylenedisulfotetramine, TETS, 4-2-4, attribution, synthesis, sulfamide, formaline, NMR

1. INTRODUCTION

The purpose of this report was the synthesis of tetramethylenedisulfotetramine (TETS) under a variety of experimental conditions to investigate the impurities and side reactions that could be unique to each condition examined. Literature review of the synthesis of TETS reports high yields, without mentioning the nature of any impurities formed. The ability to trace back the conditions applied in the synthesis of TETS, based on the presence of signature compounds, would facilitate the attribution aspect of efforts by law enforcement agencies investigating a TETS-related crime. TETS is of particular concern to law enforcement and anti-terrorism agencies, due to its water solubility[1] and its extraordinarily high toxicity, which is five times greater than that of strychnine[2] as a poison targeting the central nervous system.

The toxicity of TETS, then an unknown compound, was first recognized during World War I in Germany, when several cases of severe poisonings were reported concerning workers handling fibers impregnated with sulfamide and formaldehyde[3]. In 1949, G. Hecht and H. Henecka[4] were able to isolate the toxic compound present in the fibers, characterize it, and synthesize it by the condensation of sulfamide and formaldehyde. Additional industry related poisonings were reported in the 1950s in a furniture pillow factory, attributed to sulfamide-treated fabrics.[3]

In China, where TETS was used as a rodenticide, TETS has been used in the commission of murders, such as the case in 2002 where a food vendor contaminated a rival's wares with TETS, resulting in about 400 poisonings and almost 40 deaths. In the United States, TETS is relatively unknown as a poison, yet in 2003 a 15-month old child was hospitalized with seizures after being found playing with a white powdery rodenticide her parents had brought back from China. The powder was later identified as containing TETS.[5-11]

2. EXPERIMENTAL

2.1. Materials

The sulfamide used as one of the starting materials in the synthesis of TETS was purchased from Spectrum (CAS # 7803-58-9) and used as received. Formalin, the second component of the condensation reaction, was purchased from Sigma as a 37% aqueous solution (CAS # 50-00-0) stabilized by 10-15% methanol. Methanol is added as a stabilizer to reduce precipitation of paraformaldehyde and to assist in the breakdown of high molecular weight polyoxymethylene glycols oligomers to form more soluble compounds.

2.2. Representative preparation[3]

At room temperature, in a 25 mL round-bottom flask equipped with a magnetic stirring bar and a condenser, sulfamide (5 mmol) was added, followed by 2.5 mL of concentrated HCl. To this solution, 37% formalin solution (10 mmol) was added dropwise via a syringe. Precipitation of a white, finely divided substance occurred after about a minute. The amount of precipitate increased and thickened after 1 hour. After stirring for 2 hours, the white precipitate was filtered and washed with 20 mL of ice water. The product was then allowed to air dry.

2.3. Synthetic differentiation

The condensation of formalin with sulfamide in the appropriate stoichiometric ratio of 2:1 was performed under stirring. One batch was allowed to stir for 2 hours and the second for 1 week. In addition, a stoichiometric ratio of 1:1 was used with 1 week of stirring,.. The analysis of reaction products was conducted in three solvents: H_2O (NMR & LC/MS), acetone (NMR & GC/MS), and DMSO (NMR & GC/MS).

2.4. NMR analysis

NMR analysis of solution-state reaction products was carried out at 25 °C using a Bruker DRX-300 and a Bruker DRX-500 NMR spectrometer. The DRX-300 was equipped with a QNP probe, while the DRX-500 was equipped with a cryogenic TCI probe with enhanced detection of ^1H and ^{13}C. Experiments utilized included 1-dimensional ^1H and ^{13}C NMR spectroscopy, as well as 2-dimensional inverse-detected ^1H-^{13}C HSQC and HMQC spectroscopy.

2.5. GC/MS analysis

GC/MS analysis was performed using an Agilent 6890N/5973 GC/MS equipped with a DB-5MSD column (30 m x 0.25 mm (ID) x 0.25 mm film thickness, 95% dimethyl- / 5% diphyl-polysiloxane). The injection port was maintained at 250 °C, detector at 280 °C, and an oven ramp was held at 40 °C, then increased at 10 °C/minute to 280 °C, where it held for 10 minutes. The column used helium with a flow rate of 0.9 mL/min.

2.6. LC/MS analysis

A Waters Corporation (Beverly, MA) Synapt HDMS equipped with an Acquity Ultra Performance Liquid Chromatograph (UPLC) was used for the detection of TETS. The Synapt HDMS is a high resolution quadrupole – time-of-flight mass spectrometer equipped with an ion mobility cell. Nitrogen gas was used as the nebulizer and drying gas within the source. The spray voltage was held constant at 3.2 kV. An Acquity C18 column (50 × 2.1 mm i.d., 1.7µm) was used for UPLC separations. Liquid chromatography conditions were as follows: mobile phase A – 0.1% formic acid in water; mobile phase B – Acetonitrile. The LC time program consisted of the following steps: Time 0-3.0 min, 1 → 75% B; Time 3.0-3.1 min, 75 → 95% B; Time 3.1-3.3 min, 95 → 1% B; Time 3.3-5.0 min, 1%B. The LC flow rate was 0.3 mL/min, and all injections were 5 µL. Positive electrospray ionization mode (+ESI) was used for analysis. The m/z range was from 10-600 m/z. The instrument was controlled and data analyzed using Waters MassLynx v4.1 software.

3. RESULTS

3.1. Reaction product in D2O

The crude reaction product of the stoichiometric ratio 2:1 formaldehyde:sulfamide after 1 week of stirring was examined by NMR using D_2O, DMSO-d6, and acetone-d6 as solvents. The ^1H NMR spectrum in D_2O showed a singlet resonance at 5.54 ppm due to TETS, and several lower-intensity resonances which could not be identified due to significant overlap with the H_2O/OH resonance. The resonance attributed to TETS by ^1H NMR was confirmed by an HMQC NMR experiment, which correlated the proton signal at 5.54 ppm in the ^1H NMR to a carbon signal at 69.9 ppm in the ^{13}C NMR. This carbon resonance was the only signal detected by ^{13}C NMR in D_2O at maximum solubility after a 20 hour accumulation on a cryogenically enhanced probe at 500 MHz. As a result of this analysis, water was determined unsuitable as a solvent for identifying trace impurities of TETS by NMR.

Analysis of this aqueous solution of TETS by LC/MS indicated the presence of two compounds, shown in Figure 1. However, the chromatography was characterized by low sensitivity, which resulted in rather noisy mass spectral data. It was determined that this data could contribute little the investigation into TETS attribution products.

3.2. Reaction product in DMSO

Analysis of the same crude product in DMSO by ^1H NMR showed the presence of two compounds, one a singlet at 5.56 ppm due to TETS (57 mole%), and the other which exhibited an AB quartet at 5.19, 5.24, 5.57, 5.62 ppm (43 mole%). In comparison with literature findings, this compound was attributed to hexamethylenetrisulfohexamine (HEXS). A number of other minor resonances, several of which also appear to have an AB pattern, were also detected. (The mole percents were normalized between the two known compounds, because the others were not identified.) Comparison of this spectrum with that of formalin in the same solvent indicated the absence of signals at 3.2-3.3 ppm due to $–O(CH_2)_nH$ attributed to hemiacetals in the starting material. A few minor resonances due to formalin itself could potentially be present.

GC/MS analysis of this reaction mixture confirmed the NMR data for the two major species. The GC/MS spectrum showed the presence of two compounds assigned the molecular masses at 240 amu (consistent with TETS) and 360 amu. However, no trace of any other compounds were detected, contrary to the NMR data.

3.3. Reaction product in acetone

Analysis by ^1H NMR in acetone-d6 of the crude reaction product of the stoichiometric ratio 2:1 formaldehyde:sulfamide after 1 week of stirring was similar to the analysis in DMSO, but the molar ratio of TETS to HEXS was 77% to 33%. This is attributed to differential solubility between the two compounds in these solvents.

In acetone-d6, TETS is recognized by NMR with a singlet in the ^1H NMR spectrum at 5.58 ppm (Figures 2 and 3, peak A). In the ^{13}C NMR spectrum a ^1H-decoupled experiment yields a singlet at 71.4 ppm (Figure 4, peak A), and a ^{13}C spectrum without decoupling revealing the signal to be a triplet (indicative of a CH_2) (Figure 5, peak A). GC/MS analysis of the reaction products in acetone gave similar results to those seen for the reaction products in DMSO. The primary response on the GC occurred at 17.95 minutes (Figure 6), and the mass spectral analysis showed that this eluent corresponded to a mass of 240 amu (Figure 7, top), consistent with its identification as TETS.

HEXS was the main impurity in the synthesis of TETS, and was identified by multinuclear NMR and GC/MS. HEXS gave a singlet in the ^1H-decoupled ^{13}C NMR spectrum at 61.5 ppm (Figure 4, peaks B), and an AB quartet at 5.12, 5.17, 5.73, and 5.78 ppm in the ^1H NMR spectrum (Figures 2 and 3, peaks B). In a ^{13}C NMR spectrum without ^1H decoupling, the resonance for this compound appears as a triplet of triplets, with the center triplet split again (Figure 5, peaks B and inset). By GC/MS, HEXS was seen in the chromatogram at 29.02 minutes (Figure 6), and mass spectral analysis of this peak corresponded to a mass of 360 amu (Figure 7, bottom). These data are in agreement with published results characterizing HEXS.[12]

The crude reaction product of the stoichiometric ratio 2:1 formaldehyde:sulfamide after only 2 hours of stirring indicated the formation of the same type of products observed during the 1 week reaction. The only difference was in the mole percent ratio of TETS to HEXS, which was found to be 93% to 7% and 77% to 33% for the 2 hr and 1 week reactions respectively, as determined by GC/MS. When the synthesis was repeated using a molar ratio of 1:1 formaldehyde:sulfamide for 1 week, GC/MS analysis showed a substantial increase in the formation of HEXS, with a mole percent ratio of 62% HEXS, and only 38% TETS.

In each of these syntheses (2:1 for 1 week, 2:1 for 2 hours, and 1:1 for 1 week), a third product was seen by NMR in both DMSO (as described above) and acetone-d6. In the ^{13}C NMR spectra in acetone, there were three signals seen at 62.3, 59.5, and 57.4 ppm. In each spectrum in which these were seen, they were present in a ratio of 2:2:1 (respectively). The consistency of this ratio led us to believe that these signals derive from a single compound. In the ^{1}H spectrum, a series of doublets appear, with no obvious pattern among them. However, a two-dimensional ^{1}H-^{13}C HSQC spectrum, which correlates the ^{1}H with their attached ^{13}C, clearly reveals that each of the three ^{13}C signals are associated with an AB pattern in the ^{1}H spectrum (Figure 8, peaks C1, C2, & C3). It is assumed that these signals in the NMR correspond to a single compound, due to the consistency in different solvents and syntheses of the ratio of 2:2:1 among their ^{13}C resonance intensities. No compound was seen by GC/MS that could correspond to these signals. At this time, this product has not been identified.

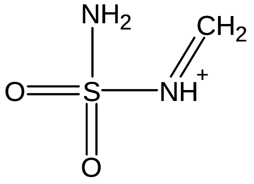
3.4. Decontamination

Based on information found in the literature[13] for TETS, it was determined prior to synthesis that our decontamination procedure for all TETS-containing solutions would consist of at least 24 hour exposure to concentrated nitric acid. To test the effectiveness of this decontamination technique, a small sample of TETS (approx 3 mg from the 2:1 reaction mixed for 1 week) was dissolved in concentrated nitric acid, stirred, centrifuged to remove particulates, and this sample was transferred to an NMR tube. ^{1}H NMR analysis of this sample showed that after 1 hour and 24 hours, there was only a minor change in the concentration of TETS (Figure 9). When the crude product was analyzed 1 hour after mixing with nitric acid, the ratio of TETS to HEXS was 53% to 47%. Analysis after 24 hours indicated a change in the molar ratio to 43% to 57%.

Further experimentation will be required to determine a more effective means of decontamination, perhaps including heating in nitric acid, or the use of sulfuric acid, or a strong base as an alternative decontamination agent. If the final product of the decontamination reaction of TETS in nitric acid is the formation of HEXS, it would be necessary for further study to be conducted into the potential hazards of HEXS.

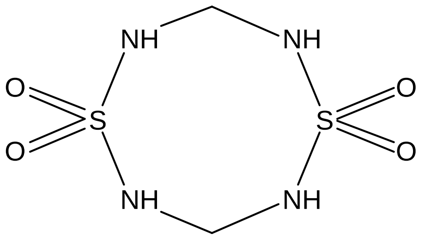
4. REACTION MECHANISM

The condensation of sulfamide with formalin in strong acid to form TETS involves the reaction of sulfamide with formaldehyde in the form of a hemiacetal or an oligomer to form a methylene-substituted sulfamide in the form of an iminium ion (**1**).[2]



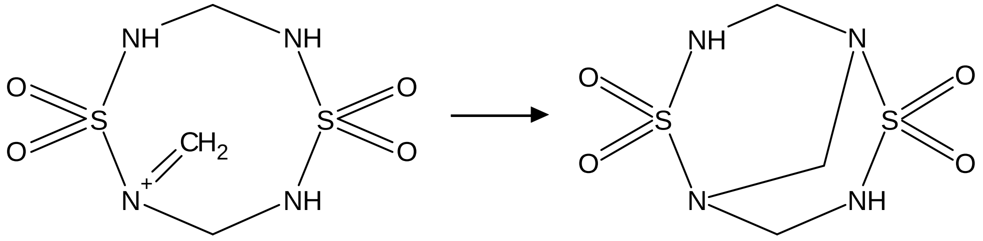
1

This ion undergoes an inter-molecular α-sulfamidoalkylation reaction followed by a dimerization to yield the eight-membered ring perhydro-dithatetrazocine-tetraoxide (**2**).



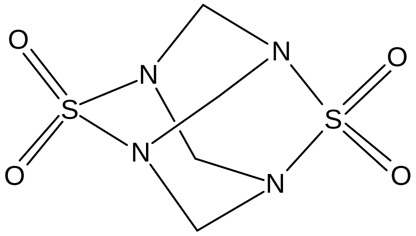
2

This adduct can then react with additional formalin to form the cyclic iminium ion (**3**). The methylene-substituted iminium ion in 3 can then undergo an intramolecular α-sulfamidoalkylation by the N atom across the 8-membered ring to form an additional 6-membered ring (**4**).



3 **4**

If the remaining two nitrogen atoms react with formalin in the same fashion – a sysmmetric adamantane-like molecule is formed which contains four 6-membered rings in a single cage compound (**5**), TETS.



5

The mechanism of HEXS (the secondary reaction product) formation is detailed in Figure 10. It involves the formation of the 8-membered ring **3** shown above in the formation of TETS. However, this species, instead of undergoing an intramolecular α-sulfamidoalkylation, reacts intermolecularly instead with another molecule of sulfamide. This, in turn, via a sequence of intramolecular α-sulfamidoalkylations induced by the formation of two seperate iminium ion intermediates, yields HEXS.

In agreement with this mechanism, when the TETS synthesis was repeated in the presence of excess sulfamide (the 1:1 ratio of formaldehyde to sulfamide with 1 week of mixing time), HEXS was the major reaction product, while TETS formation was greatly reduced.

5. CONCLUSIONS

The synthesis of TETS by the condensation of formalin with sulfamide at room temperature is a facile, high-yield, rapid reaction. It requires no special laboratory equipment and it can easily be carried out outside a laboratory environment. TETS is highly toxic to humans and animals, and, in this regard, it is of special interest to anti-terrorist government agencies.

From the attribution point of view, the synthesis of TETS is always accompanied by HEXS, the molar ratio of which varies according to the stoichiometry ratio of the starting materials. The amount of HEXS observed when acetone was used as a solvent varied from 7% when a formaldehyde to sulfamide ratio of 2:1 was employed, to 62% when a 1:1 ratio was used. Thus, the amount of HEXS in TETS can be used as a marker of the origin of TETS, and the sophistication of its synthesis.

A suitable method for the safe destruction of TETS in environmental waste still remains to be developed. TETS, with its heteroadamantane N-C-N-C amide nitrogen structure, should possess less than moderate stability to strong acids at room temperature. However, a solution of TETS in concentrated nitric acid observed by ^{1}H NMR retained a high degree of TETS, having only partially converted to HEXS after 24 hours. Further investigation into decontamination methods of TETS and HEXS must be performed.

ACKNOWLEDGEMENTS

The authors would like to thank Tom Rusek, Dr. Joy Ginter, and John Tokarz for their assistance with GC/MS and LC/MS analysis.

REFERENCES

[1] Brennan, R.J. Waeckerle, J.F., Sharp, T.W., et. al. Chemical Warfare Agents: medical and emergency public health issues. Ann Emerg Med (1999) 34: 191-204
[2] Budavari, S. ed. (1996) The Merck Index, An encyclopedia of chemicals, drugs, and biologicals, 12th edition., Whitehouse, NJ, Merck & Co., Inc.
[3] Cao, C., (1991) Chi J Int Med 4(4):46
[4] Esser, T., Karu, A.E. Toia, R.F., and Casida, J.E. (1991) Chem Res Toxicol 4:162-167
[5] Fan, C. (2001) China J Misdiagnosis 1(2): 266-267
[6] Hao, F., (1992) Chi J Ind Hyg Occp Dis 10(1): 34-35
[7] Hecht, G., Henecka, H. (1949) Angew. Chem 61: 365-366
[8] Hoffman, F.W. Technical command interim report TCIR 634, Chemical Division technical command Army chemical center, MD, 7 February 1951
[9] Lee, C., Kohn, H. J Org Chem (1990) 6098-6104
[10] Ning, P., He, Q., Yu, F., Feng, Z., Deng, P., Jia, T. (1997) Chi J Ind Hyg Occp Dis 15(2):108-109
[11] Wu, Q., Zhang, M-S., Lan Z-R. (2002) Chromatography 20(4): 381-382
[12] Xu, L., Mo, W., Liu, J. (1998) Emerg Med 7:232
[13] Zhao, L. (1999) J Law and Med 6(1):34-35

FIGURES and DATA

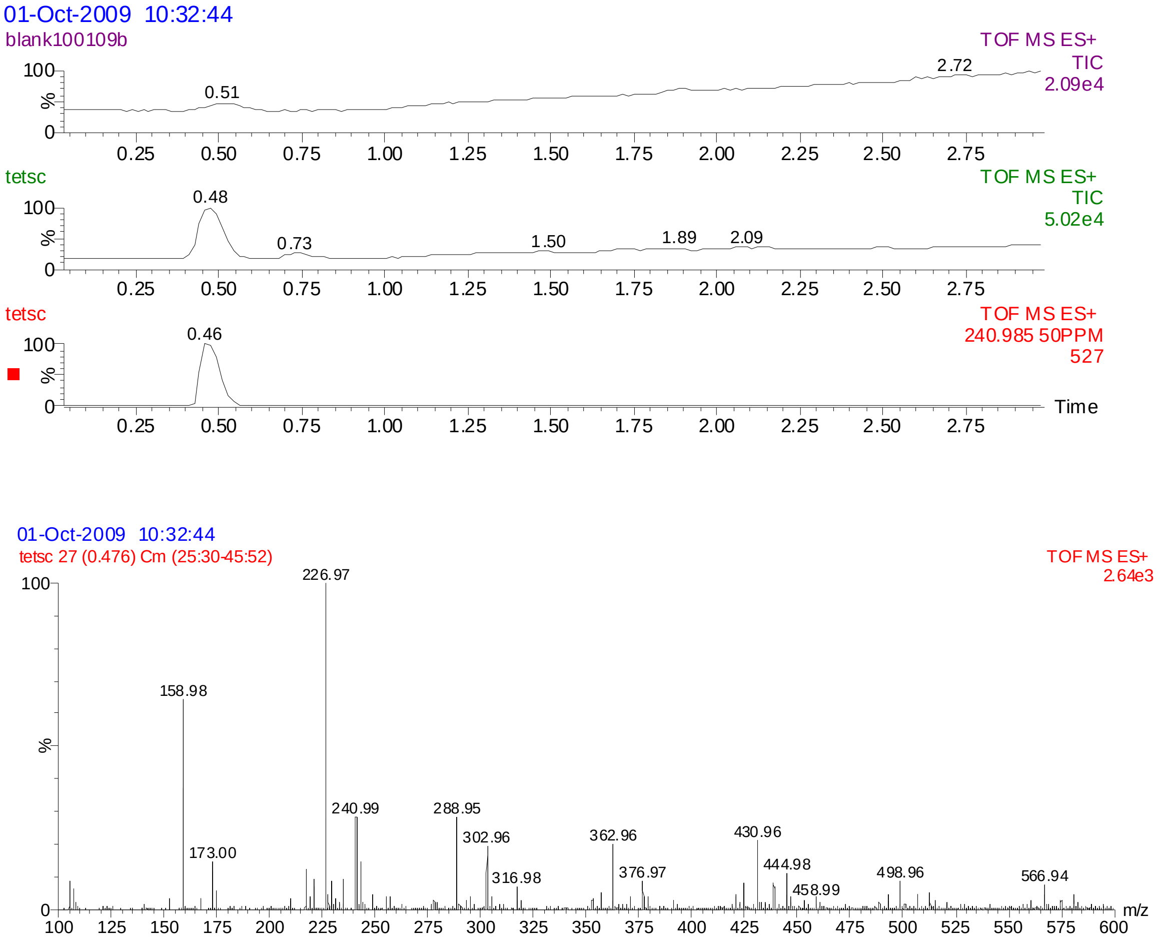


Figure 1. UPLC analysis (top) of the reaction product (molar ratio of 1:1 formaldehyde to sulfamide with 1 week of mixing/reaction time) in water, and the mass spectral analysis of the peak at 0.46 minutes (bottom).

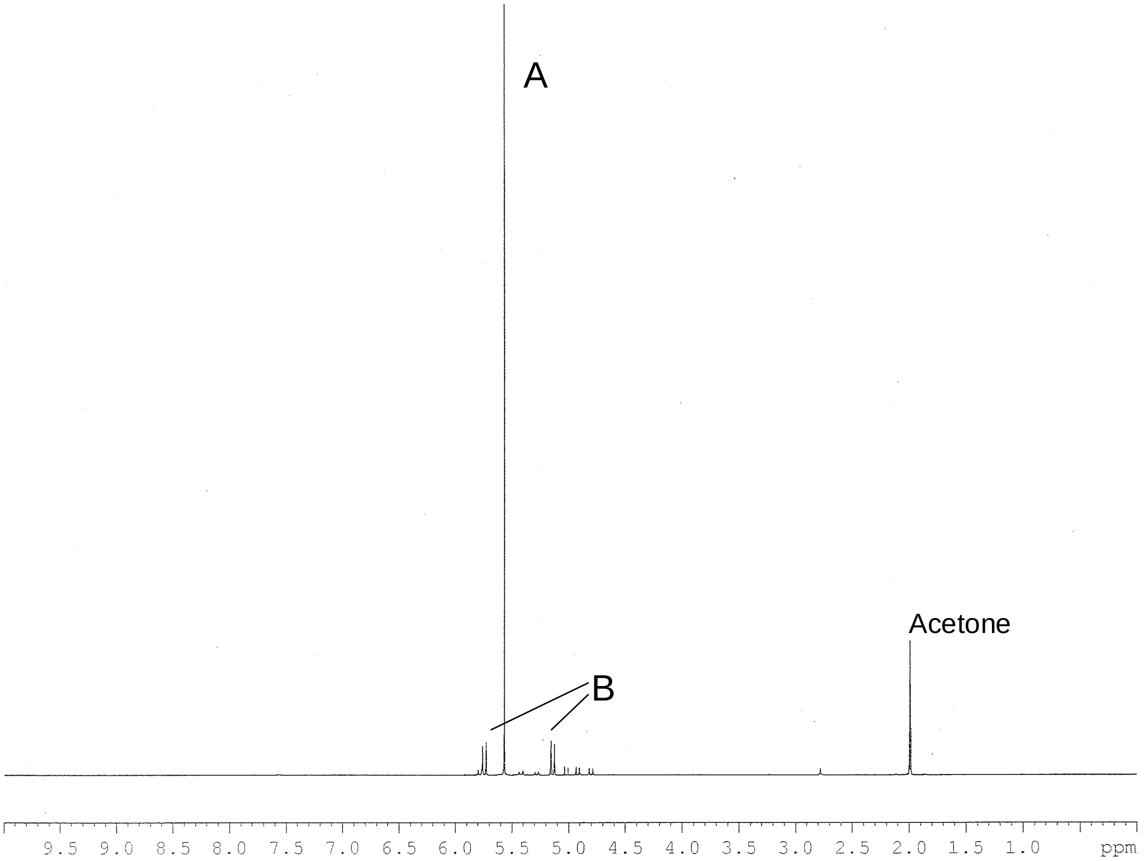


Figure 2. ^{1}H NMR spectrum of the reaction product (molar ratio of 2:1 formaldehyde to sulfamide with 1 week of mixing/reaction time) in acetone-d6. Marked peaks are: A) TETS singlet, B) AB pattern of HEXS. The unmarked peak at 2.8 ppm is from residual water. The region from 4.5 to 6.0 ppm is expanded in Figure 3.

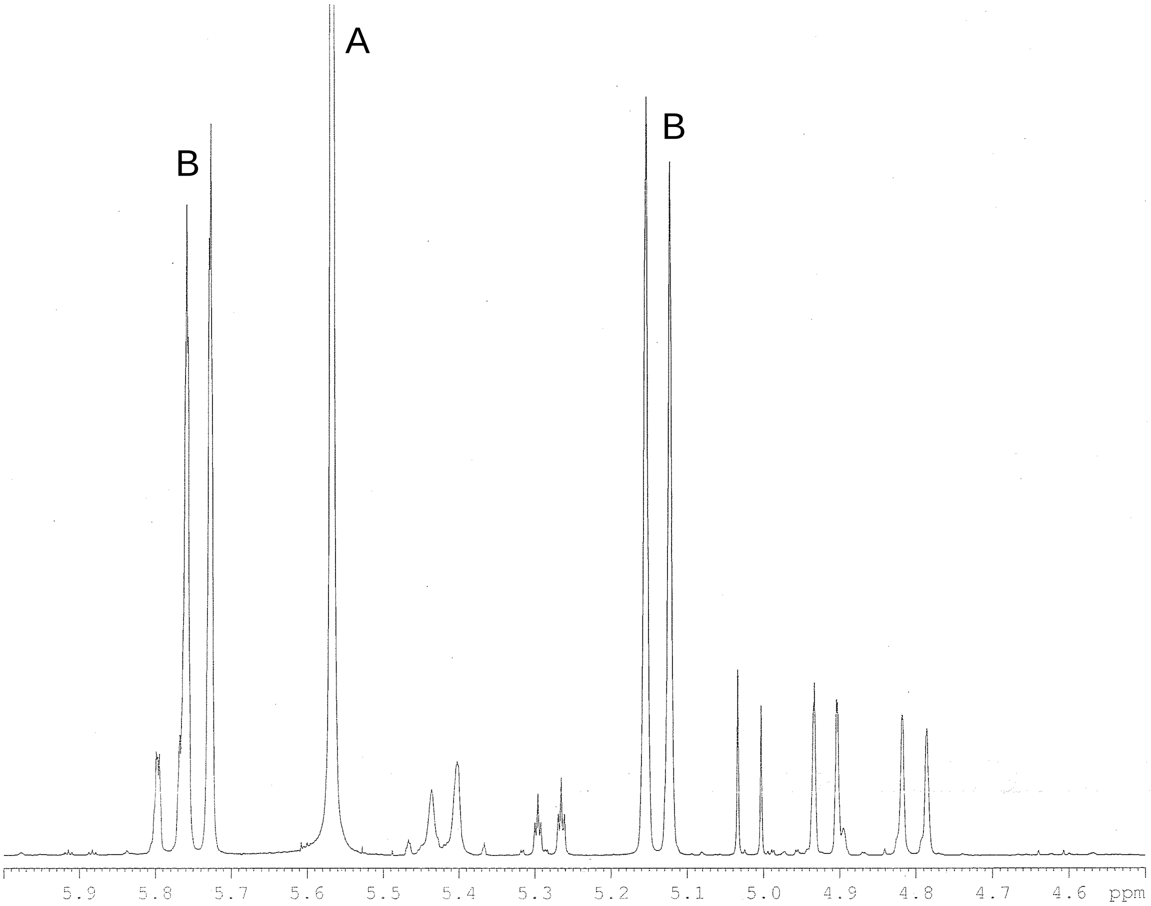


Figure 3. Expansion of the ^{1}H NMR spectrum from Figure 2. Marked peaks are: A) TETS singlet, B) AB pattern of HEXS.

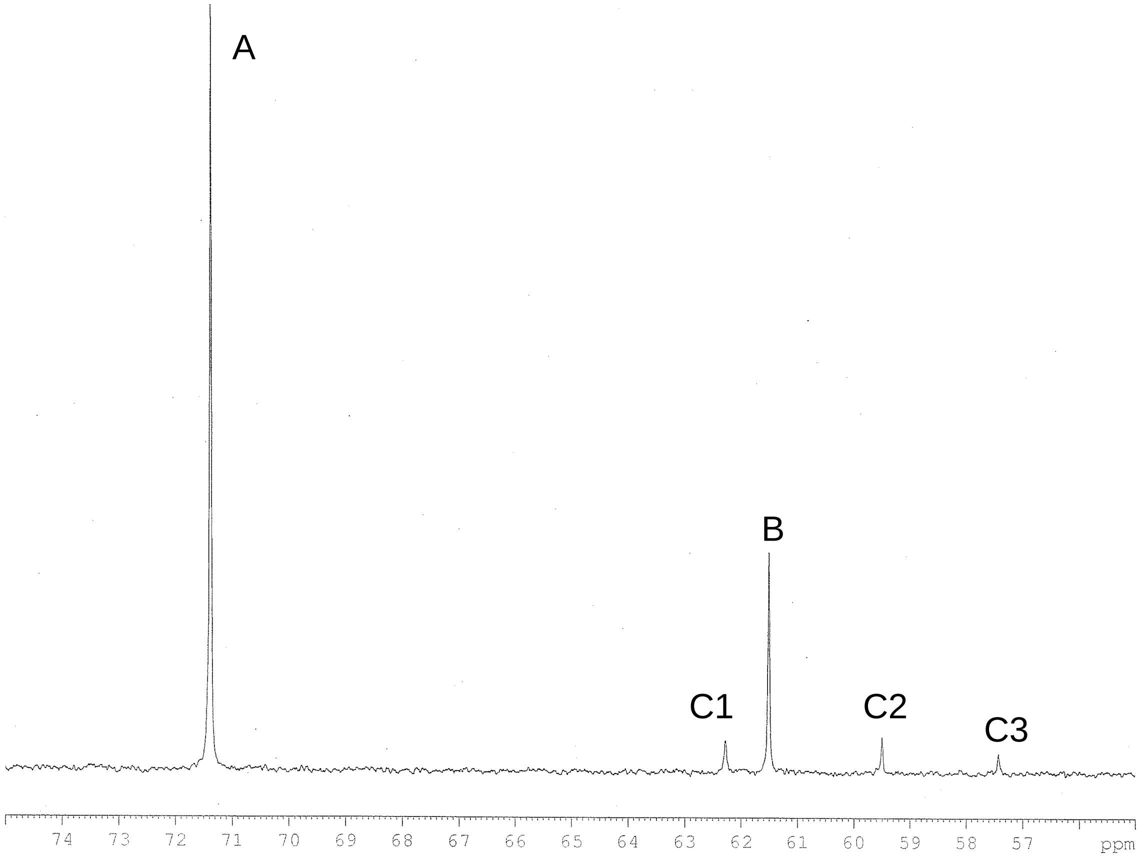


Figure 4. ^{13}C NMR spectrum (^{1}H-decoupled) of the reaction product (molar ratio of 2:1 formaldehyde to sulfamide with 1 week of mixing/reaction time) in acetone-d6. Marked peaks are: A) TETS singlet, B) HEXS singlet , C1, C2, C3) singlets consistently seen with a 2:2:1 ratio among their intensities, from a compound(s) not seen by GC/MS or LC/MS.

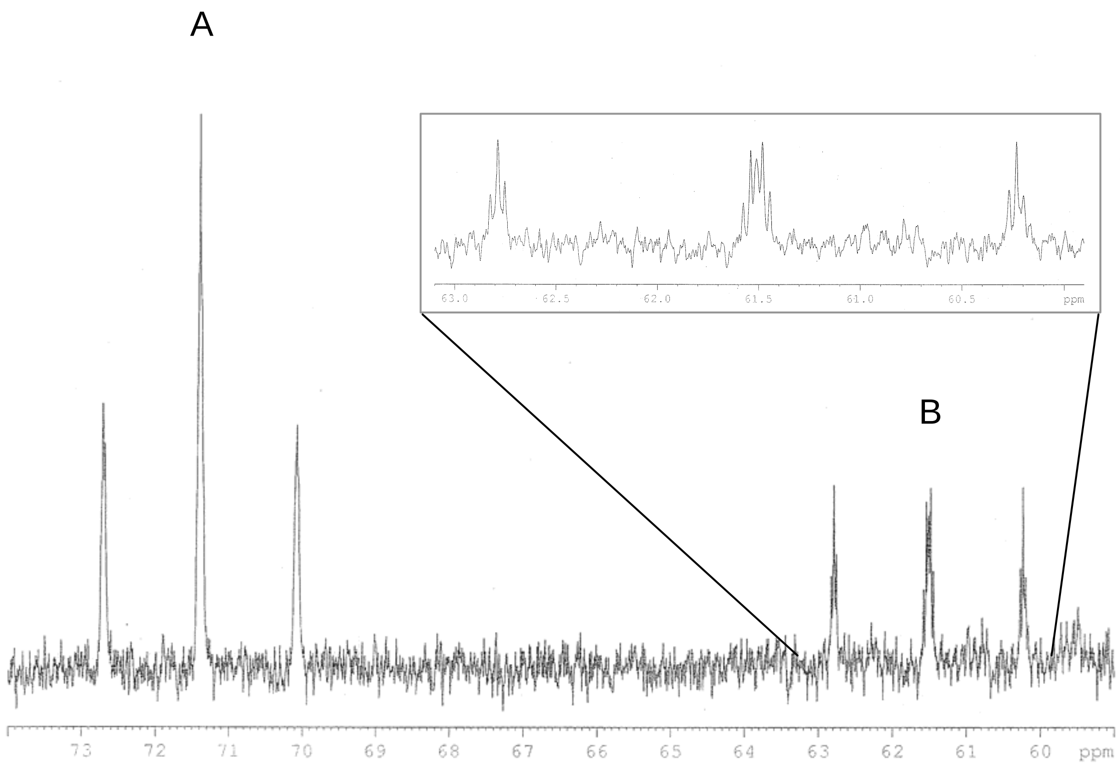


Figure 5. ^{13}C NMR spectrum of the reaction (molar ratio of 2:1 formaldehyde to sulfamide with 1 week of mixing/reaction time) in acetone-d6. Coupling with the bonded protons (CH_2) splits the singlets seen in Figure 4 into triplets. Marked peaks are: A) TETS, B) HEXS. In this spectrum, the weaker signals from C1, C2, and C3 are not seen.

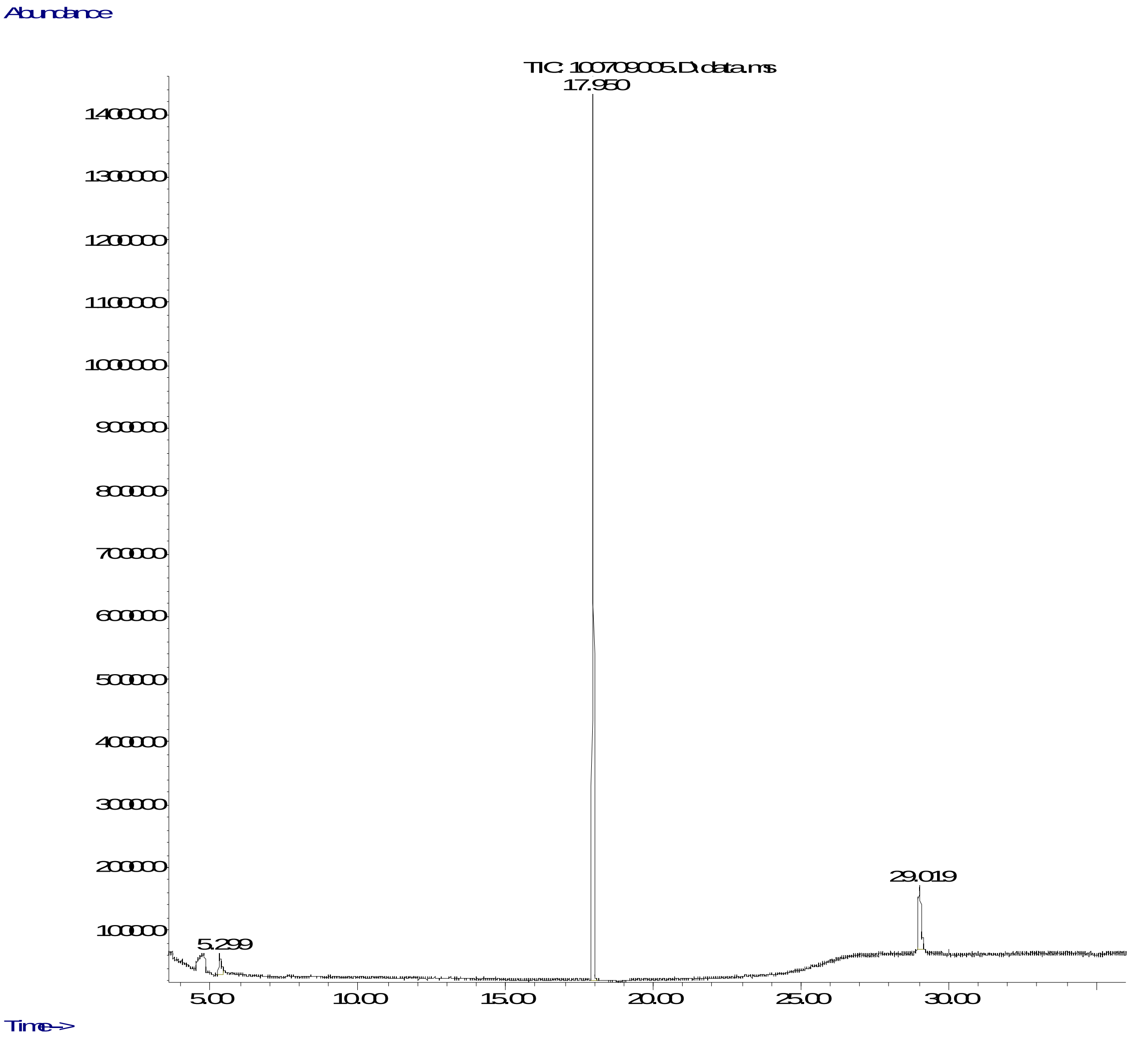
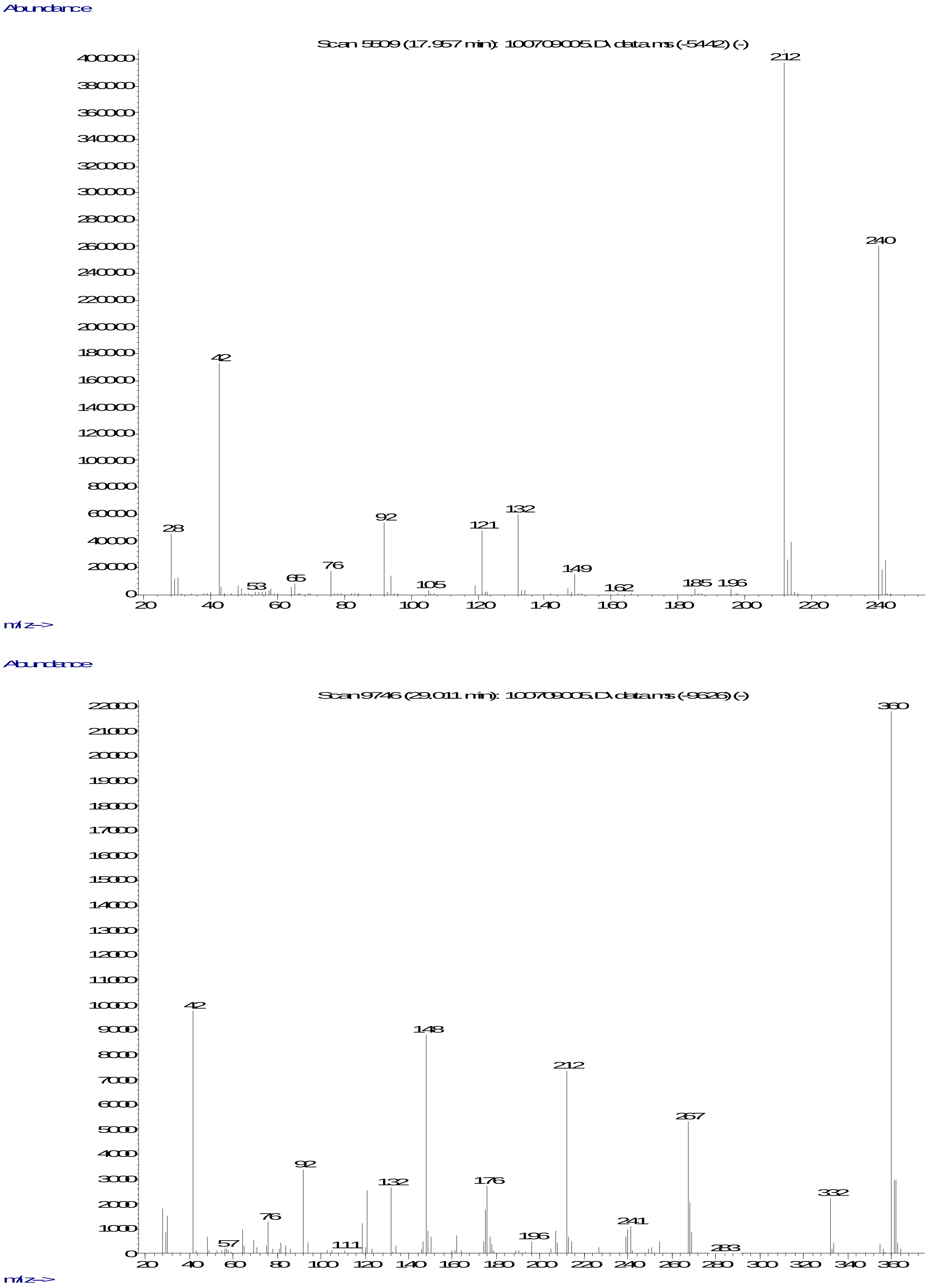


Figure 6. GC chromatogram of the reaction product (molar ratio of 2:1 formaldehyde to sulfamide with 1 week of mixing/reaction time) in acetone.



Scan 5509 (17.957 min): 100709005.D\ data.ms (-5442) (-)

Scan 9746 (29.011 min): 100709005.D\ data.ms (-9626) (-)

Figure 7. Mass spectral analysis of the 17.95 minute (top) and 29.01 minute (bottom) chromatogram traces, showing molecular ions at 240 and 360 amu, respectively.

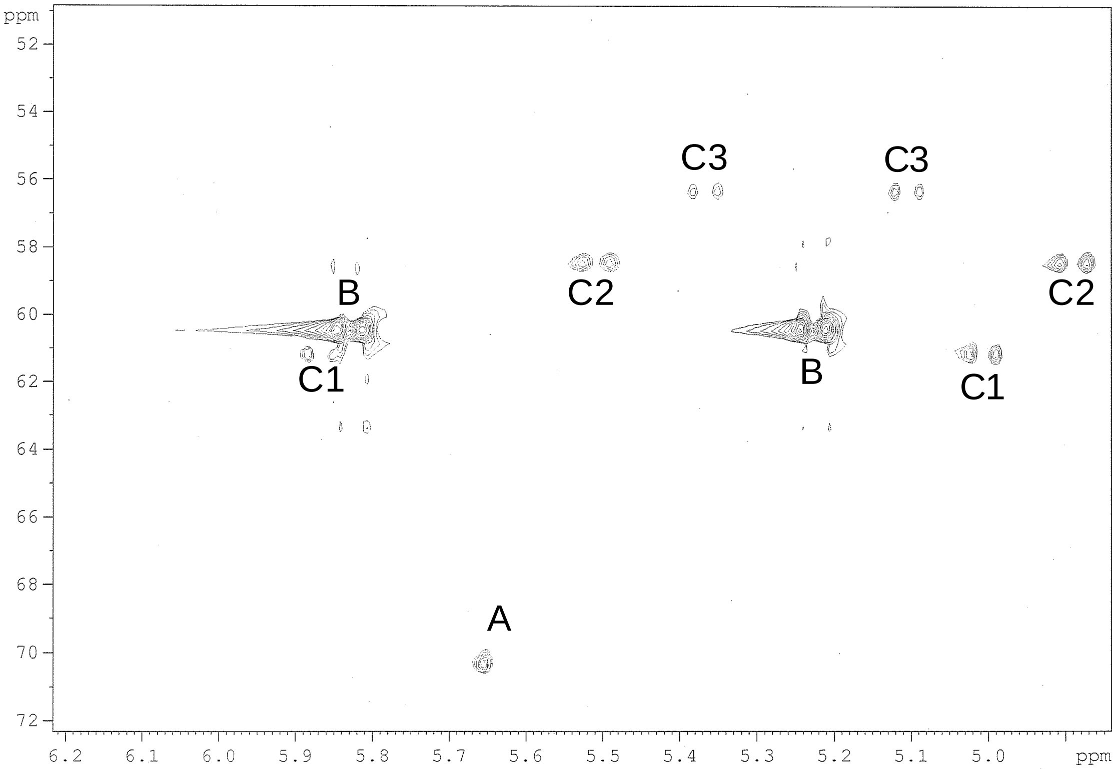


Figure 8. ^{1}H-^{13}C HSQC NMR spectrum of the reaction product (molar ratio of 1:1 formaldehyde to sulfamide with 1 week of mixing/reaction time) in acetone-d6. In this sample, the chemical shifts are slightly different relative to those before, likely due to a difference in sample pH. Note that, with the 1:1 reactant ratio, the TETS production (A) is greatly reduced, while production of the HEXS (B) is dominant. Marked peaks are: A) TETS, B) AB pattern from HEXS, C1, C2, C3) AB pattern from a compound(s) not seen by GC/MS or LC/MS. Note that these pairs of doublets can be directly seen in the ^{1}H NMR spectrum expanded in Figure 3.

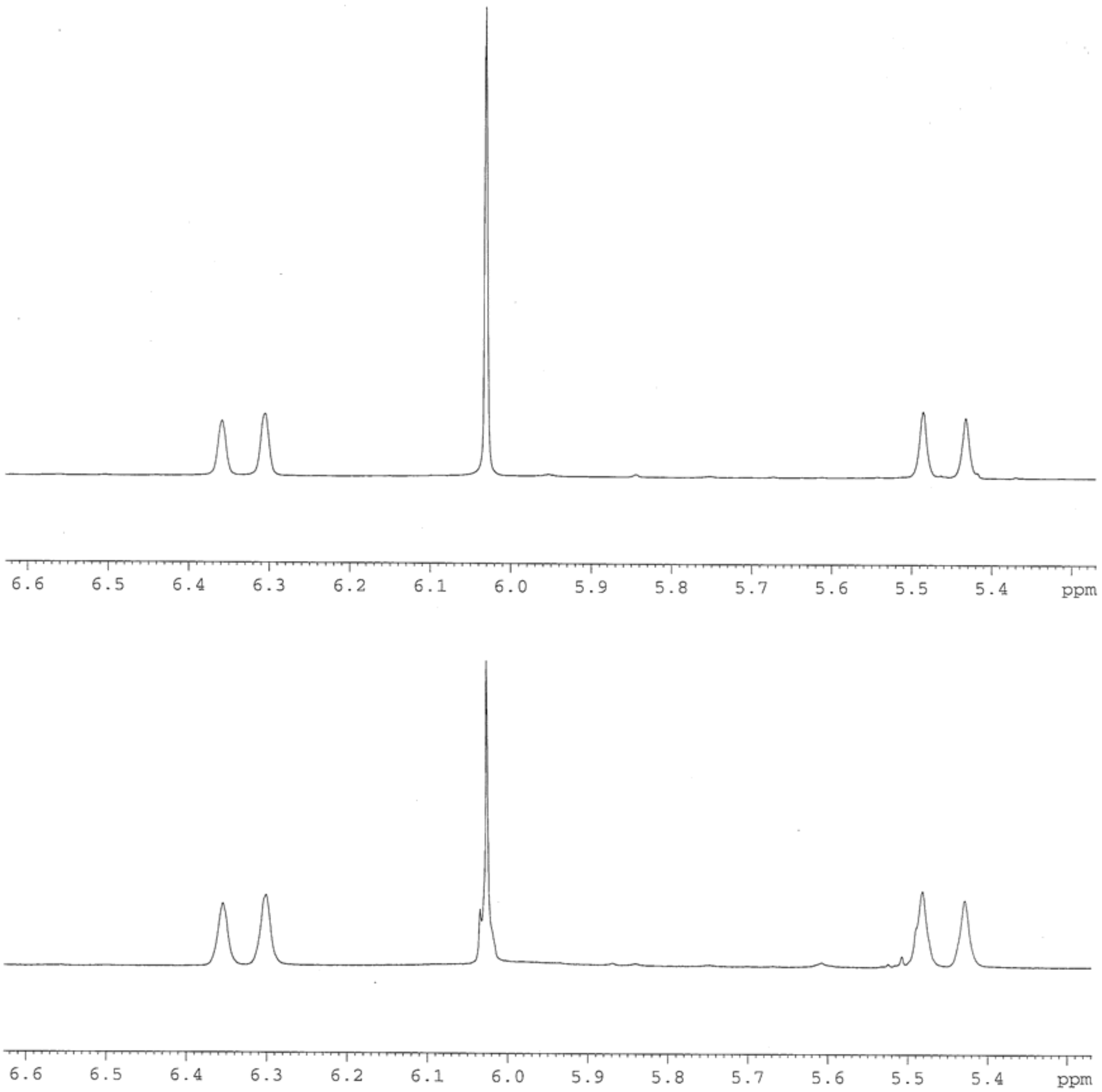


Figure 9. ^{1}H NMR spectrum of the reaction product (molar ratio of 2:1 formaldehyde to sulfamide with 1 week of mixing/reaction time) dissolved in nitric acid. The similarity between the spectra after 1 hour (TOP) and after 24 hours (BOTTOM) indicates that concentrated nitric acid at room temperature will not be sufficient to decontaminate TETS or HEXS. Analysis indicates a relative change of approximately 10% in favor of HEXS.

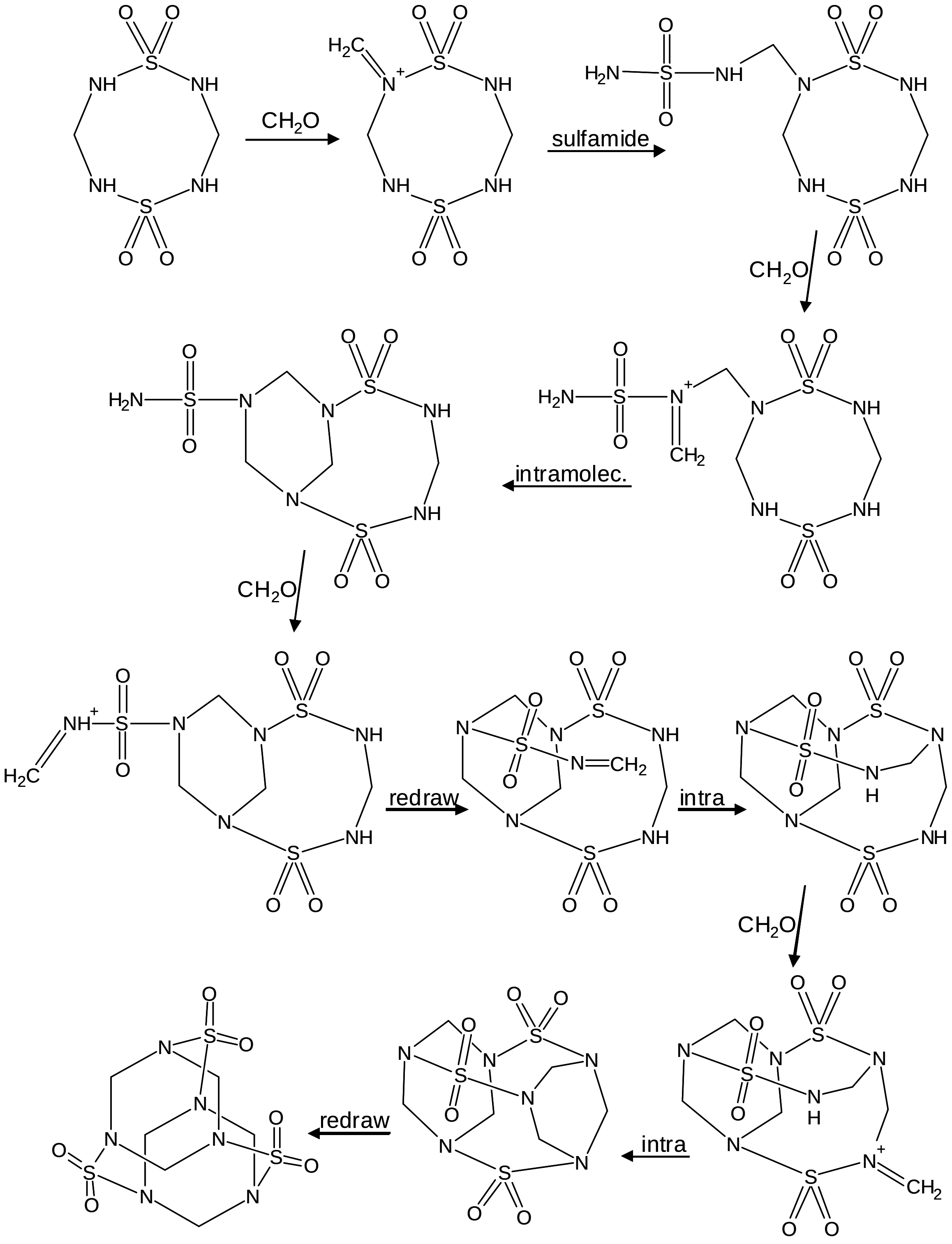


Figure 10. Schematic diagram showing one of the pathways by which HEXS can be formed from the same intermediates as TETS in this reaction.

Ricin Toxicity in BALB/c 3T3 Cells: Correlation of Total Proteins with Dose Level by Mass Spectrometry and Proteomics

Vicky L. H. Bevilacqua and Janna S. Madren-Whalley[a], Rabih Jabbour[b], Lisa M. Reilly[c],
Jeffrey S. Rice[d]
[a]Research and Technology Directorate, Aberdeen Proving Ground, MD, USA, 21010
[b]Science Applications International Corporation, Abingdon, MD, USA, 21009
[c]Bethany College, Bethany, WV, USA, 26032
[d]Elona Biotechnologies, INC. Greenwood, IN, USA, 46143

PREFACE

The work described in this report was authorized under Project No. 61110191a00 (In-House Laboratory Independent Research). This work was started in January 2009 and completed in September 2009. The use of either trade or manufacturers' names in this report does not constitute an official endorsement of any commercial products. This report may not be cited for purposes of advertisement. This report has been approved for public release. Registered users should request additional copies from the Defense Technical Information Center; unregistered users should direct such requests to the National Technical Information Service.

1. INTRODUCTIONS

Traditional cell toxicity assays lead to an IC_{50} or EC_{50} value based on the determination of cell death. However, additional information is desirable for understanding a toxin's total effects. NMR and/or mass spectrometry (MS) in combination with chemometrics analysis can yield physiological details in the form of metabolome or proteome biomarker information, but model-building with IC_{50}/ED_{50} values obtained from traditional cytotoxicity assays has been required to correlate the biomarker concentration with toxicity [1, 2]. We hypothesize that NMR or MS can be used for the direct determination of EC_{50} or IC_{50} in addition to biomarker identification. Here we report on a study to explore this hypothesis using liquid-chromatography and tandem mass spectrometry (LC/MS/MS) with ricin dosed BALB/c 3T3 murine fibroblasts (CCL-163 American Type Culture Collection [ATCC], Manassas, VA, USA) as a model system. An MS-based chemometrics procedure is also expected to have advantages including: 1) no dyes or labeled compounds, and 2) high sample throughput.

2. METHODS

Note: Ricin is extremely toxic. Its use is controlled under the Biological Select Agents and Toxins program in the United States by the Centers for Disease Control and Prevention (Atlanta, GA). Handling of ricin should follow strict safety procedures determined in collaboration with the safety office of the research laboratory's organization.

2.1. Ricin Preparation

Ricin communis agglutinin II (ricin, Vector Laboratories, Burlingame, CA) was dialyzed into 10 mM sodium phosphate buffer (pH 7.0, PB) over a period of ~ 24 hrs with gentle stirring using three volumes of ~ 600 mL PB. Dialysis was carried out on ice using regenerated cellulose or cellulose ester Dispodialyzers® (Spectrum Laboratories, Rancho Dominguez, CA) with 5000 or 8000 molecular weight cutoff. Dialyzed ricin was stored at 0–4 ºC. The dialyzed ricin concentration was determined at 25 ºC by ultraviolet (UV) absorbance employing either a JASCO Model J-810 Spectropolarimeter (JASCO Analytical Instruments, Easton, MD) equipped with a PTC-423S Peltier thermoelectric temperature control system or a NanoDrop® ND-1000 spectrophotmeter (NanoDrop® Technologies, Inc., Thermo Fisher Scientific, Waltham, MA). For the JASCO method, fifteen absorbance measurements on one aliquot of the solvent blank solution (no protein) were recorded at intervals of 1 second and averaged. Fifteen measurements were then recorded on one aliquot of the ricin sample and averaged. The blank average was subtracted from the sample average. The concentration was calculated using Beer's Law with $E^{0.1\%}{}_{280nm} = 1.4$ [3]. For the Nanodrop® method, blank-subtracted measurements were recorded on an aliquot of the ricin sample. The concentration was calculated using Beer's Law with $E^{0.1\%}{}_{280nm} = 1.4$ [3]. The Nanodrop® measurement procedure and concentration calculation procedure were carried out three times, each using fresh aliquots of blank and ricin sample, and the concentration results averaged. The results of the two procedures were within error of each other. The dialyzed ricin was sterile filtered prior to use with the cell cultures and the concentration verified by the Nanodrop® method.

2.2. Ricin Exposure of BALB/c3T3 Murine Fibroblasts

The fibroblasts (CCL-163 American Type Culture Collection [ATCC], Manassas, VA, USA) were passaged a minimum of three times after thawing prior to testing. Seventy-five- cm^2 cell culture flasks were seeded at 8 x 10^3 cells/cm^2 and maintained in culture at 37 ºC in a humidified atmosphere of 5% CO_2 in air for 24 hrs prior to treatment. Exposures were performed 24 ± 2 hr after seeding the 75-cm^2 flasks. Three concentrations of ricin were prepared by diluting the dialyzed ricin in cell culture medium. The concentrations corresponded to IC_{20}, IC_{50}, and IC_{80} concentrations previously determined by Neutral Red Uptake assay [4, 5]. Six replicate sets of flasks (2 flasks per set) were prepared per ricin concentration. In addition, flasks were prepared containing untreated cells (Vehicle Controls, VC), and containing media with and without ricin. After ricin exposure, the flasks were incubated at 37 ºC in a humidified atmosphere of 5% CO_2 in air for another 48 hrs.

Cell harvesting was carried out 48 hrs ± 0.5 hr post-exposure. Medium from each set of flasks (two flasks per concentration of ricin and vehicle controls) was removed. The attached cells remaining in the flasks were rinsed twice with Hank's Balanced Salt Solution (HBSS) and then incubated at room temperature for 3 min with trypsin-versine (Product # 17-161E, Lonza Walkersville, Inc., Walkersville, MD). The trypsin-versine was neutralized with the addition of 37 ºC Routine Culture Medium (RCM) containing serum. The cells were removed from the flasks, centrifuged for 5 min at 1200 rpm. The supernatant was then discarded and the pellet was resuspended in 4 mL of 37 ºC PBS. A 700 µL sample was removed for analyses not reported here. The remaining cells were centrifuged for 5 min at 1200 rpm. The supernatant was discarded and the cell pellet was resuspended in 1 mL of 37 ºC PBS and then frozen at -80 ºC.

2.3. Liquid-Chromatography/Mass Spectrometry Sample Preparation

The cell samples were thawed and lysed by ultrasonication (25 s on, 5 s off, 4 min total) using a Branson Digital Sonifier® (Danbury, CT). The lysate was centrifuged at 14,000 rpm for 20 min at 10 ºC using a Beckman GS-15R centrifuge and F2402H Rotor. The supernatant was transferred to a Microcon® YM-3 filter unit (Millipore, Billerica, MA 01821) and centrifuged at 14,000 rpm and 10 ºC (500 µL for 30 min x 2, total volume of 1.00 mL). The filtrate was stored at -20 ºC for other analyses. The cellular proteins in the retentate were denatured overnight at 37 ºC with 300 µL of 7.2 M Urea and 3 ug/mL dithiothreitol in 100 mM ammonium bicarbonate (ABC). The urea was removed by centrifugation (14,000 rpm, 30 min, RT) and the retentate washed with 200 µL ABC followed by centrifugation using an Eppendorf centrifuge (5415C with rotor F-45-18-11 or 5415D with rotor F-45-24-11, Eppendorf North America, Westbury, NY 11590) at 14,000 rpm for 30 min (RT). The filter unit was then transferred to a new receptor tube and the proteins in the retentate digested at 37 ºC for 7 hr with 5 µL sequencing grade trypsin (Product # 511A, Promega, Madison, WI 53711) in 10 µL acetonitrile and 50 µL ABC. The tryptic peptides were isolated by centrifuging at 14,000 rpm, 15 min, RT (Eppendorf 5415C or D). The filtrate containing the tryptic peptides was stored at 4 ºC until analysis.

2.4. Liquid-Chromatography/Mass Spectrometry Experiments

For each sample, 10 µL of filtrate containing tryptic peptides (described above) was diluted with 90 µL of aqueous phase (95% H_2O, 5% acetonitrile). Ten microliters of this dilution was injected into a Thermo Electron Corp Finnigan Surveyor HPLC (Thermo Scientific, Waltham, MA 02454) and the peptides separated using a 0.1 x 150 mm C_{18} Hypersil GOLD KAPPA column (5 µ particle size, 175 Å pore size, Fisher Scientific International, Pittsburgh, PA 15275) with a linear gradient of 20 to 80% organic phase (100 acetonitrile, 0.1% formic acid). The column was connected to a Finnigan LTQ tandem ion trap mass spectrometer fitted with a nanospray ESI source operated at 1.82 kV with a collision energy of 25 V. A data-dependent mode and scan range of *m/z* 300-2000 were used. The full mass spectra were collected, followed by MS on the resulting five most intense ions.

2.5. Proteomics Analysis

A 3T3 protein database was constructed in a FASTA format using the annotated proteome sequences from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov). For this task, an in-house PERL (http://www.activatestate.com/Products/ActivePerl) program was used to automatically download the annotated 3T3 proteome sequences from the NCBI. The database was constructed by translating putative protein-encoding genes and contains amino acid sequences of potential tryptic peptides obtained by the *in silico* digestion of all 3T3 proteins, assuming up to two missed cleavages. The acquired mass spectra were searched against this database with the SEQUEST algorithm (Thermo Scientific). The SEQUEST thresholds for searching the product ion mass spectra were Xcorr, deltaCn, Sp, RSp, and deltampep. These parameters provide a uniform matching score of all candidate peptides. The generated output files of these candidate peptides were validated using the PeptideProphet™ algorithm (http://peptideprophet.sourceforge.net). Peptide sequences with probability scores of 95% and higher were retained.

3. RESULTS AND DISCUSSION

The LC/MS/MS experiments were carried out on 24 cellular protein extract samples as described above and summarized in Table 1. Typical LC/MS/MS data for a single sample is shown in Figure 1.

Table 1. Summary of Tryptic Peptide Samples analyzed by LC/MS/MS.

Sample Type	Ricin Concentration (μg/mL)	# of Replicate Samples
Vehicle Control Cells	---0---	6
Cells dosed with IC20 concentration of ricin	9.54×10^{-04}	6
Cells dosed with IC50 concentration of ricin	2.15×10^{-03}	6
Cells dosed with IC80 concentration of ricin	5.97×10^{-03}	6
Total	----	24

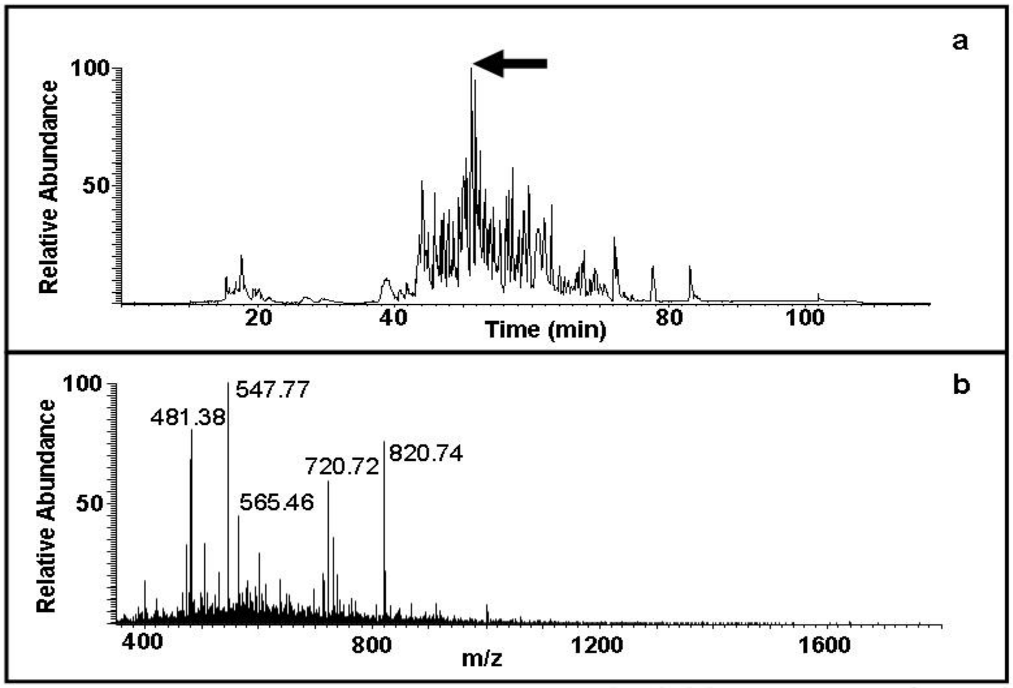


Figure 1. Liquid chromatography-mass spectrometry data from extracted cellular proteins. a) Total ion chromatogram. The arrow marks the peak having a 51-min retention time. b) Full MS spectrum for LC peak with 51-min retention time.

For this experiment, the attached cells recovered from each set of cell growth flasks (representing a single ricin dose) were suspended in 1 mL of PBS (see section 2.2 of this report). As the toxin dose and thus the percent inhibition are increased, the result will be fewer total cells recovered and correspondingly fewer total cells per mL suspended in PBS. It is expected that the total amount of recoverable protein per mL would be dependent on the total number of cells and that the signal-to-noise of the LC/MS/MS data from the tryptic peptides would therefore decrease corresponding with toxin dose, resulting in fewer proteins identified by proteomics analysis and forming the basis for toxicity analysis. Table II provides the average number of proteins identified by MS for each ricin concentration. The relative ratios for the proteins could be determined per sample type and protein expression relative to ricin dose could be examined using regression analysis. However, that type of analysis is more appropriately carried out after first adjusting either the cell concentration so that the total cells per sample from which proteins are extracted is constant, or the volume of cellular protein extract injected for LC/MS/MS analysis, to take into account the decrease in total cells in the original 1 mL of PBS. An analysis along this line is in progress, but is outside the scope of this report. For this report, we focus on the relationship of the total number of proteins identified relative to toxin dose.

Table 2. Average Number of Proteins Identified from Recovered Cellular Proteins.

Cell Sample Type	Average # of Proteins Identified	Normalized Average # of Proteins Identified
Vehicle Control	47.3	100.0
Dosed with IC20 concentration of ricin	39.2	82.8
Dosed with IC50 concentration of ricin*	30.6	56.7
Dosed with IC80 concentration of ricin	13.8	29.2

*Average of 5 (rather than 6) replicate samples. One sample was not included for the IC_{50} set due to a Microcon filter unit membrane leak that occurred during extraction.

As noted under section 2.2 of this report, the ricin concentrations employed for dosing corresponded to IC_{20}, IC_{50}, and IC_{80} concentrations that were calculated from previously carried out Neutral Red Uptake assay experiments [3]. Because the percent inhibition was not actually measured here, we will refer to percent inhibition values for these ricin doses as "calculated percent inhibition" values. A plot of the normalized average number of proteins identified versus ricin dose mirrors that of calculated percent inhibition versus ricin dose (Figure 2). Each parameter (normalized average number of proteins identified and calculated percent inhibition) is related logarithmically to ricin concentration. As would be expected based on the results in Figure 2, a plot of the average number of proteins identified versus the calculated percent inhibition (Figure 3) fits a linear curve ($R^2 = 0.998$) having a negative slope.

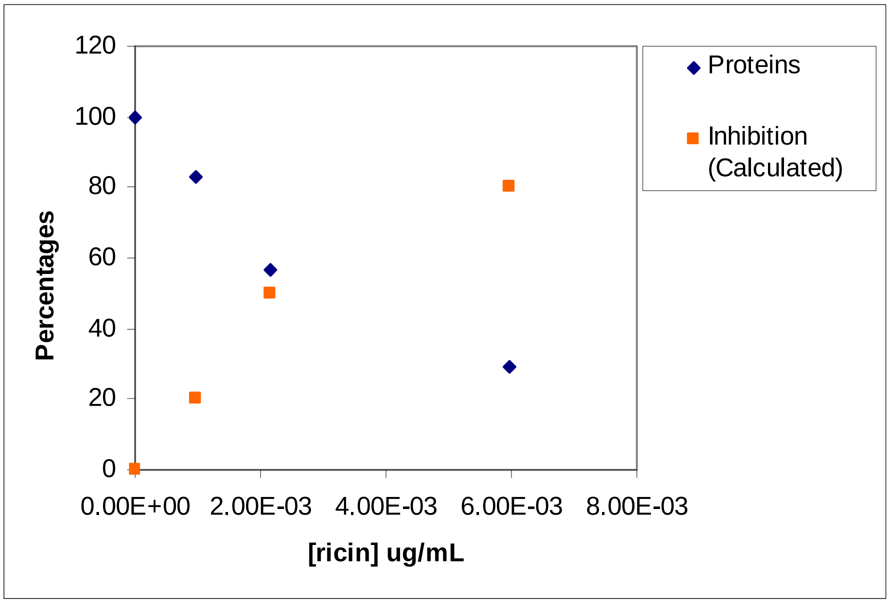


Figure 2. Calculated percent inhibition (orange squares) and normalized average number of proteins identified (blue diamonds) versus ricin dose.

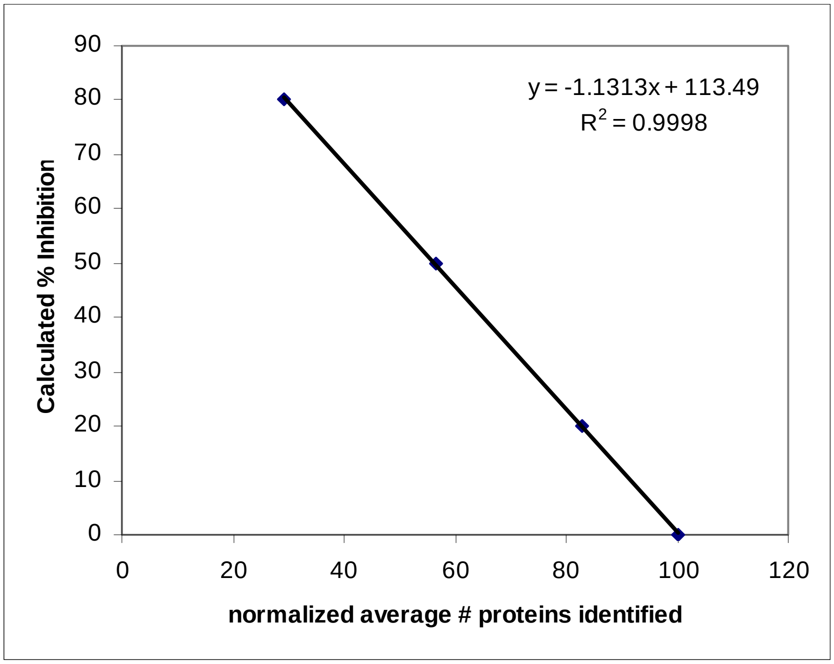


Figure 3. Calculated percent inhibition versus normalized average # of proteins identified by LC/MS/MS of cellular protein extracts from ricin dosed cells.

4. CONCLUSIONS

Proteomics studies of dosed cells typically involve adjusting the cell concentration so that it is constant across all samples immediately prior to protein extraction, or adjusting the volume of protein extract injected (for LC/MS/MS analysis) according to the cell concentration. These procedures are carried out to keep the overall MS signal-to-noise level constant from one sample to the next. The proteins are then identified and individual proteins quantitated in search of specific biomarkers related to dosing. The analysis requires model-building with IC_{50}/EC_{50} values obtained independently from cytotoxicity assays such as a neutral red uptake assay. Here we have analyzed cellular protein extracts from BALB/c 3T3 murine fibroblasts dosed with ricin. The cell concentration per dose was constant at the initial time of dosing. However, at the time for sample collection (48 hr post dosing) only attached cells were harvested and no adjustment was made for cell concentration. This procedure takes advantage of the decrease in MS signal-to-noise for cellular protein extracts that is expected as inhibition of cell growth increases with toxin dose. The assumption is that as the intensity for individual proteins drops below the limit of detection, fewer proteins are identified from samples that contained fewer cells at the time of harvest. Using this procedure, we have shown that the average number of cellular extract proteins identified and the calculated percent inhibition are each related logarithmically, but in an inverse manner, to ricin concentration. We have also shown that the average number of proteins identified is related linearly to the calculated percent inhibition. These results indicate that, if at least eight ricin concentrations are included as is done for the NRU assay [4, 5], toxicity could be determined and effective dose values obtained directly from LC/MS/MS data. We are planning to carry out such an experiment. If the results are in agreement with the 3-concentration study carried out here, they will be validated by measuring toxicity of the ricin preparation independently employing the neutral red uptake assay. If the validation is successful, it will prove that LC/MS/MS can be used for the direct determination of EC_{50} or IC_{50} in addition to biomarker identification, as originally hypothesized.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the In-House Laboratory Independent Research program for funding. The authors would also like to acknowledge Mr. Alan Zulich and Dr. A. Way Fountain for administrative support.

REFERENCES

[1] E.M. Lenz, et al., *J. Pharm. Biomed. Anal. 35* (2004) 599-608.
[2] Shimoda, T.; Funatsu, G. Binding of Lactose and Galactose to Native and Iodinated Ricin D. *Agric. Biol. Chem.*, 1985, *49,* pp 2125-2130.
[3] S.V. Vulimiri, et al., *Chem. Res. Toxicol.* (2009, Epub ahead of print).
[4] V.L.H. Bevilacqua, et al., ECBC-TR-702, U.S. Army ECBC, APG, MD 21010-5424; UNCLASSIFIED Report (2010).
http://iccvam.niehs.nih.gov/methods/acutetox/invitrocyto/invcyt_pr

Saccharification of a Potential Bioenergy Crop, *Phragmites australis* (Common Reed), by Lignocellulose Fractionation Followed by Enzymatic Hydrolysis at Decreased Cellulase Loadings

Noppadon Sathitsuksanoh[a,b], Zhiguang Zhu[a], Neil Templeton[a], Joseph A. Rollin[a],
Steven P. Harvey[c], and Y-H. Percival Zhang[a,b,d,e]
[a]Biological Systems Engineering Department, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA 24060
[b]Institute for Critical Technology and Applied Science (ICTAS), Virginia Polytechnic Institute and State University, Blacksburg, VA, USA 24060
[c]US Army ECBC, Research & Technology Directorate, Aberdeen Proving Ground, MD, USA 21020
[d]DOE BioEnergy Science Center (BESC), Oak Ridge, TN, USA 37831
[e]To whom correspondence should be addressed. E-mail: ypzhang@vt.edu. Tel.: (540) 231-7414. Fax: (540) 231-3199.

Published in *Industrial and Engineering Chemistry Research*, **2009**, *48(13)*, pp 6441-6447.

PREFACE

Cost-effective biological saccharification of nonfood lignocellulosic biomass is vital to the establishment of a carbohydrate economy. *Phragmites australis* (common reed) is regarded as an invasive perennial weed with a productivity of up to 18-28 tons of dry weight per acre per year. We applied the cellulose solventand organic solvent-based lignocellulose fractionation (COSLIF) to the stems and leaves of *Phragmites* and optimized the pretreatment conditions (e.g., temperature, reaction time, and biomass moisture content) through response surface methodology (RSM). The optimal pretreatment conditions were 85% (w/v) H_3PO_4, 50 °C, and 60 min, regardless of the biomass moisture contents from 5-15% (w/w). Glucan digestibility of the COSLIF-pretreated common reed was 90% at hour 24 at a low cellulase loading (5 filter paper units and 10 β-glucosidase units per gram of glucan). Under these conditions, the overall sugar yields were 88% for glucose and 71% for xylose, respectively. Cellulose accessibility to cellulase (CAC) was increased 93.6-fold from 0.14 ± 0.035 to 13.1 ± 1.1 m^2 per gram of biomass with the COSLIF pretreatment. Results showed that cellulase concentrations could be reduced by 3-fold with only a slight reduction in sugar yield. This study suggested that *Phragmites* could be used as a carbon-neutral bioenergy feedstock, while its harvesting could help control its invasive growth and decrease nutrient pollution in adjacent waterways.

1. INTRODUCTION

The production of biofuels and biobased products from renewable lignocellulosic biomass will promote rural economy, decrease greenhouse gas emissions, and enhance energy security.[1-3] Biomass saccharification usually involves two sequential steps: lignocellulose pretreatment/fractionation and enzymatic hydrolysis of cellulose. The largest technological and economical challenge for biomass biorefineries is the efficient release of fermentable soluble sugars from low-cost lignocellulosic biomass at competitive costs.[4-7] Currently, the production of second generation biofuels, that is, cellulosic ethanol, cannot compete with that made from corn grain and sugar cane, because of its high processing costs (ca. $1-3 per gallon of cellulosic ethanol), huge capital investment ($2-10 per annual gallon ethanol capacity), and relatively low revenues from ethanol ($2-3 per gallon of cellulosic ethanol).[4,7,8]

Recently, a new technology called cellulose solvent- and organic solvent-based lignocellulose fractionation (COSLIF) has been developed to separate lignocellulose components (cellulose, hemicellulose, lignin, and acetic acid) by using a cellulose solvent, an organic solvent, and water.[4] Different from other lignocellulose pretreatments, this technology can be conducted at modest reaction conditions (e.g., 50 °C and atmospheric pressure) for minimizing sugar degradation. The key ideas of COSLIF are (1) partial removal of lignin and hemicellulose from cellulose, allowing more cellulose exposure to cellulase, (2) decrystallization of cellulose fibers (allowing cellusase to work more efficiently), and (3) modest reaction conditions (i.e., a decrease in sugar degradation, less inhibitor formation, lower energy requirement, and less capital investment). Higher glucan digestibility of the COSLIF-pretreated biomass was attributed to greater cellulose accessibility and more lignin removal, as compared to the dilute acid pretreatment.[9] In addition, COSLIF can separate lignocellulose components on the basis of their different solubilities in solvents and exhibit coutilization of lignocellulose components such as lignin.[4,7,10]

The DOE cellulosic ethanol workshop has summarized three distinct goals associated with potential bioenergy feedstocks: (1) maximizing the total amount of biomass produced per acre per year, (2) producing sustainable biomass with minimal inputs (e.g., pesticides, fertilizers, seeds, and harvesting), and (3) maximizing the amount of biofuels that can be produced per unit of biomass.[11] A yield of 20 dry tons per acre per year may be considered as a reasonable target in an area with adequate rainfall and good soil.[11]

Pharagmites australis (common reed) is a widespread perennial grass that grows in wetlands or near inland waterways throughout the world. Although it is harvested for thatched roofs, ropes, baskets, pulping feedstock, etc., in some areas of the world, common reed is typically regarded as an invasive weed, due to its vigorous growth and difficulty of eradication. *Phragmitez*, a C4 photosynthesis plant, can grow as high as 18 feet, with enormously high productivities of 18-28 tons of dry biomass per acre per year.[12] This productivity is approximately three to five times higher than a dedicated bioenergy crop switchgrass. Since it produces seeds in addition to its growth from rhizomes, large-scale planting would be easier as compared to another potential bioenergy plant *Miscanthus*. Judging from annual inputs, the use of common reed as a bioenergy plant would have several advantages: growth from rhizomes after initial establishment from seeds or rhizomes, no or low fertilizer requirement, no irrigation (growing in wetlands), and low pesticides needed. Since common reed is regarded as an invasive weed by the U.S. Environmental Protection Agency (EPA), annual harvesting of common reed as a bioenergy feedstock can be regarded as weed control. In addition, existing strands of *Pragmites* are huge in the USA, and its further planting as a bioenergy crop seems promising. In fact, growth features of common reed are very good for biomass harvesting. Its winter standing allows a much longer harvesting time, as compared to corn stover. Also, harvesting of standing naturally dried strands with decreased moisture contents of ~5-15% would save drying costs and biomass transportation costs. Before winter, it can recycle its nutrients to rhizomes for the growth in next year. Common reed usually grows in neutral pH or alkaline tropical and temperate water lands or wetlands, which are not suitable for most crops. Because it can take up nutrients efficiently, harvesting of existing strands will effectively remove phosphorus and nitrogen from inland waterways, and prevent algal blooms and other microbial pollution.[13-16] The features and associated advantages of common reed are presented in Table 1.

Table 1. The Features of *Phragmites* and Its Potential Advantages as a Bioenergy Crop.

Features	**Advantages**
Perennial grass	Harvested yearly, long canopy duration
C4 photosynthesis	High photosynthesis efficiency, e.g., 18-28 tons of dry biomass per acre per year
Having seeds	Easy large-scale planting at the beginning
Growing from rhizomes	No yearly replanting
Removing pollutant nutrients	No fertilizers, waste water treatment
Growing in wetland	No irrigation
Few pests	Low pesticide needed
Invasive	Harvesting = weed control
Winter standing	Long harvesting time (several months), low moisture content, feedstock for winter harvesting
Having rhizomes	Recycling nutrients to rhizomes in winter
Temperate and tropical regions	Worldwide
Marginal lands	No competition for arable land, preferring neutral and even alkaline wetlands

In this study, we investigated the feasibility of applying the COSLIF technology to common reed. We also sought to further improve the COSLIF technology by replacing the organic solvent (acetone) with ethanol for reductions in processing costs and capital investment for recycling of organic solvent. We optimized key pretreatment conditions by using response surface methodology (RSM), studied the release of soluble sugars from this potential bioenergy plant at decreased enzyme loadings, and analyzed potential economic benefits associated with low use of costly enzyme.

2. MATERIALS AND METHODS

2.1. Chemical and Materials

All chemicals were reagent grade and purchased from Sigma-Aldrich (St. Louis, MO), unless otherwise noted. Phosphoric acid (85%) and 95% ethanol were purchased from Fisher Scientific (Houston, TX). The *Trichoderma* cellulase (Novozyme 50013) and *β*-glucosidase (Novozyme 50010) were gifts from Novozymes North American (Franklinton, NC). They had activities of 84 filter paper units (FPUs) per mL and 270 *β*-glucosidase units per mL, respectively.

Common reed was obtained from the U.S. Army Edgewood Chemical Biological Center (Aberdeen, MD) in the early winter of 2007. The naturally dried common reed was milled into small particles by using the Pallmann counter-rotating knife ring flaker (Clifton, NJ). The resulting particulates were screened to the sizes of less than 40 mesh (i.e., smaller than 0.420 mm) and greater than 60 mesh (i.e, larger than 0.250 mm). The milled materials were slowly dried to a moisture content of ~5% at room temperature, whose moisture contents were determined by complete drying in a convection oven, at 105± 3 °C for 4 h or longer, until a constant weight was achieved. The different moisture contents of biomass samples were prepared by mixing ~5% moisture content biomass with water, and then equilibrating in a closed container at room temperature overnight.

2.2. COSLIF Procedure

The COSLIF pretreatment for common reed was conducted as described elsewhere,[4,10] with some modifications. Acetone was replaced with 95% (v/v) ethanol. One gram of dry common reed with a moisture content, varying from 5%, 10% to 15%, was mixed with 8 mL of 85% phosphoric acid at different temperatures (40, 50, and 60 °C) for different lengths of time (30, 60, and 90 min) in 50-mL plastic centrifuge tubes. The biomass dissolution and weak hydrolysis reactions were stopped by adding 20 mL ethanol. After mixing well, solid/liquid separation was conducted in a swinging bucket centrifuge at 4500 rpm at room temperature for 15 min. After the supernatant was decanted, an additional 40 mL of ethanol was mixed with the slurry containing cellulose and hemicellulose. The solid/liquid separation was again conducted by centrifugation. After the supernatant was decanted, the pellets were resuspended and washed twice with 40 mL of water. The residual amorphous solid pellet was neutralized to pH 5-7 with a small amount of 2 M sodium carbonate.

2.3. Carbohydrate and Lignin Assays

The structural carbohydrate composition of the biomass was determined with a modified quantitative saccharification (QS) procedure.[17] In the modified QS, the secondary hydrolysis was conducted in the presence of 1% (w/w) sulfuric acid, rather than 4% sulfuric acid at 121 °C, for 1 h for more accurate determination of acidlabile carbohydrates (e.g., xylan and arabinan).[17] Monomeric sugars were measured by a Shimadzu HPLC, with a Bio-Rad Aminex HPX-87P column (Richmond, CA), at 65 °C with a distilled water as a mobile phase at a rate of 0.6 mL per min.[17] Lignin and ash were measured according to the standard NREL biomass protocol.[18] The concentrations of glucose and xylose in the enzymatic hydrolysate were measured by a Shimadzu HPLC with a Bio-Rad Aminex HPX-87H chromatography column by using 0.1% (v/v) sulfuric acid as a mobile phase at a flow rate of 0.6 mL per minute and a column temperature of 65 °C.[4]

2.4. Enzymatic Hydrolysis

The pretreated common reed samples were diluted to 10 g glucan per liter in a 50 mM sodium citrate buffer (pH 4.8) with supplementary addition of 0.1% (w/v) NaN3, which prevented the growth of microorganisms.[4] All hydrolysis experiments were carried out in a rotary shaker at 250 rpm and 50 °C. Four enzyme loadings were tested: (1) 5 FPUs cellulase and 30 units of _-glucosidase per gram of glucan; (2) 10 FPUs cellulase and 30 units of *β*-glucosidase per gram of glucan; (3) 15 FPUs cellulase and 30 units of *β*-glucosidase per gram of glucan; (4) 5 FPUs cellulase and 10 units of glucosidase per gram of glucan. Eight hundred microliters of well-mixed hydrolysate were removed, followed by immediate centrifugation at 13 000 rpm for 5 min. Then exactly 500 *μ*L of the supernatant was transferred to another microcentrifuge tube and incubated at room temperature for 30 min, enabling the conversion of (nearly) all cellobiose to glucose, by glucosidase in the supernatant. The supernatant was acidified by adding 50 *μ*L of 10% (w/w) sulfuric acid, and then was frozen overnight. The thawed liquid samples were mixed well and then centrifuged at 13 000 rpm for 5 min, to remove any solid sediment. The clear supernatants were used for determination of the released glucose by HPLC. After 72-h hydrolysis, the remaining hydrolysate was transferred to a 50 mL centrifuge tube, and centrifuged at 4500 rpm for 15 min. After decanting, the pellet was resuspended in 20 mL of water and centrifuged to remove soluble sugars. Following centrifugation, the remaining sugars and lignin in the lyophilized pellets were measured by QS. The soluble glucose and xylose (including galactose and mannose) in the enzymatic hydrolysate were measured by HPLC using a Bio-Rad HPX-87H column, as described above.

The enzymatic glucan digestibility (*X*)[19] can be calculated in percent as:

$$G = \frac{G_f}{(180/160)G_i}100 \qquad (1)$$

where *G*f, is the amount of soluble glucose plus cellodextrins in the liquid phase after hydrolysis (g glucose equivalent, GE) and Gi, is the initially added glucan in solid cellulosic samples before hydrolysis (g).

For biomass pretreatment and subsequent enzymatic hydrolysis, the biomass input (stream 1) generated two streams (pretreatment hydrolysates-stream 2 and pretreated biomass-stream 3) and then enzymatic hydrolysis (stream 3) produced the solid residue (stream 4) and the enzymatic hydrolysate (stream 5).[20] The overall glucose yield (Y_{Glu}), during the COSLIF pretreatment and enzymatic cellulose hydrolysis, is calculated in percent as:

$$Y_{Glu} = \frac{Glu_2 + Glu_5}{(180/162)Glu_1}100 \qquad (2)$$

where Glu_2 and Glu_5 are mass amounts of glucose equivalent in streams 2 (the pretreatment liquid hydrolysate) and 5 (the enzymatic hydrolysate), respectively. Glu_1 is the initial glucan content before pretreatment. It is worth noting that commercial cellulase and *β*-glucosidase solutions contain very high concentrations of sugars (~20-100 g glucose per liter of enzyme solution).[21] Therefore the glucose concentration in stream 5 needs to be reduced by the amount of sugars already present in the enzyme solutions.[21]

Since a significant amount of xylooligosaccharides that cannot be measured in the presence of cellodextrins, by the regular HPLC columns, could exist in stream 5, the overall xylose yield Y_{Xyl} can be calculated in an alternative way as:

$$Y_{Xyl} = \frac{Xyl_2 + (150/132)(Xly_3 - Xly_4)}{(150/132)Xly_1}100 \qquad (3)$$

where Xyl_3 and Xyl_4 are mass amounts of xylan in steams 3 and 4, respectively. Xylan contents in streams 1, 3, and 4 were measured by the HPLC HPX-87P column after quantitative saccharification.

2.5. Scanning Electron Microscopy (SEM)

The biomass materials were imaged with a Zeiss-DSM940 (Carl Zeiss, Okerkochen, Germany). All samples were sputter-coated with gold and imaged by SEM, as described elsewhere.[22]

2.6. Substrate Accessibility Assays

The total substrate accessibility to cellulase (TSAC) was determined on the basis of the maximum adsorption capacity of the TGC protein.[9] The TGC protein is a nonhydrolytic fusion protein, containing a green fluorescence protein and cellulose-binding module.[23] The recombinant TGC fusion protein was produced in *Escherichia coli* BL21 (pNT02),[23] and purified by affinity adsorption on regenerated amorphous cellulose,22 followed by modest desorption using ethylene glycol (EG).[24] EG was removed by membrane dialysis in a 50 mM sodium citrate buffer (pH 6.0). The TGC protein solution was reconcentrated using a 10000 Da molecular weight cutoff centrifugal ultrafilter column (Millipore, Billerica, MA). Mass concentration of the nonadsorbed TGC protein was measured on the basis of a fluorescent reading using a BioTek multidetection microplate reader, as described elsewhere.[23] Cellulose accessibility to cellulase (CAC, m2/g biomass) can be measured on the basis of the maximum TGC adsorption capacity after the blocking by a large amount of BSA (e.g., 5 g/L). Noncellulose accessibility to cellulose (NCAC, m^2/g biomass) was calculated as NCAC) TSAC - CAC.[9]

3. RESULTS

The common reed sample was harvested at the Aberdeen Proving Ground of Maryland. After complete drying, it contains 32.7 ± 2.5% glucan, 18.1 ± 2.2% xylan, 1.2 ± 0.2% galactan, 2.5 ± 0.3% arabinan, 22 ± 2.0% lignin, as well as 20 ± 3.6% mass weight for extractives, ashes, proteins, and so on.

3.1. Modified COSLIF Technology

The original version of COSLIF used a highly volatile organic solvent (acetone) between a cellulose solvent (concentrated phosphoric acid) and water. The functions of this organic solvent are (1) to partially remove lignin by dissolving it, (2) to decrease cellulose solvent recycling costs, and (3) to separate water-soluble depolymerized hemicellulose fragments and water-insoluble amorphous cellulose.[4,10] Low boiling- point acetone can be recycled easily by simple flashing, but it must be recycled with very high yields (e.g., > 99.99%). Any loss in acetone would negatively impact the economics of COSLIF implementation.

Here we replaced acetone by using ethanol for the modified COSLIF. This modification brought several benefits such that (1) a much lower recycling efficiency of ethanol is acceptable because the remaining ethanol in the hydrolysate and cellulose phase can be recycled after ethanol fermentation, (2) ethanol is more chemically stable than acetone, and (3) ethanol is less corrosive to the following membrane-based separations. Furthermore, we decreased organic solvent use nearly 2-fold, from 100 volumes to 60 volumes.

3.2. Optimization of COSLIF Pretreatment Conditions

The yield of fermentable sugars from the lignocellulosic biomass is a critical factor for evaluating the overall performance of the saccharification process, because sugar yields correlate closely with revenue.[10,19,20] Biomass saccharification usually involves two sequential steps: pretreatment and enzymatic hydrolysis. The COSLIF pretreatment conditions (temperature, time, and biomass moisture content) were optimized by using RSM.[25] The pretreatment temperature (T, 40, 50, and 60 °C), reaction time (t, 30, 60, and 90 min), and biomass moisture content (MC, 5, 10, and 15%) were chosen as independent variables (Table 2). The experimental design consisted of a 3-factor 2-level pattern with 20 experimentss14 combinations with 6 replications of the central point. The statistical software Design-Expert 6.0 (Stat-Ease Inc., Minneapolis, MN) was used to analyze the experimental results. The glucan retention after the COSLIF pretreatment, glucan digestibility, and glucan yield are presented in Table 2. The quadratic equation was obtained for the maximum glucose release, from pretreatment and enzymatic hydrolysis as:

$$\begin{aligned}\text{glucan yield }(\%) = {} & -155.68 + 8.66T + 1.94\text{MC} + 1.44t - \\ & 0.083T^2 - 0.025(\text{MC})^2 - 4.09 \times 10^{-3}t^2 - 0.026T(\text{MC}) - \\ & 0.018Tt - 6.69 \times 10^{-3}(\text{MC})t\end{aligned} \tag{4}$$

Table 2. Pretreatment Conditions and Experimental Results for Glucan Retention after the COSLIF Pretreatment, Glucan Digestibility after Enzymatic Hydrolysis, And Glucan Yield That Equals Glucan Retention × Glucan Digestibility[a].

run	*T*	MC	*t*	glucan retention (%)	glucan digestibility (%)	glucan yield (%)
1	33.2	10	60	98.0	82.5	80.9
2	40	5	30	93.9	87.8	82.4
3	40	5	90	95.1	92.6	88.1
4	40	15	30	95.1	87.3	83.0
5	40	15	90	91.1	91.9	83.8
6	50	1.6	60	88.6	93.1	82.5
7	50	10	9.6	92.4	73.1	67.5
8	50	10	60	92.7	93.8	86.9
9	50	10	60	91.9	93.9	86.3
10	50	10	60	94.3	92.3	87.0
11	50	10	60	94.5	93.6	88.4
12	50	10	60	91.6	93.5	85.6
13	50	10	60	91.4	93.8	85.8
14	50	10	110.5	85.9	93.2	80.1
15	50	18	60	92.8	88.9	82.5
16	60	5	30	84.4	93.2	78.6
17	60	5	90	68.7	89.9	61.8
18	60	15	30	79.3	92.4	73.3
19	60	15	90	59.0	89.9	53.3
20	66.8	10	60	47.5	85.3	40.5

[a] All hydrolysis experiments were carried out at the same enzyme loading of 15 FPUs of cellulase and 30 units of β-glucosidase per gram of glucan for 24 h.

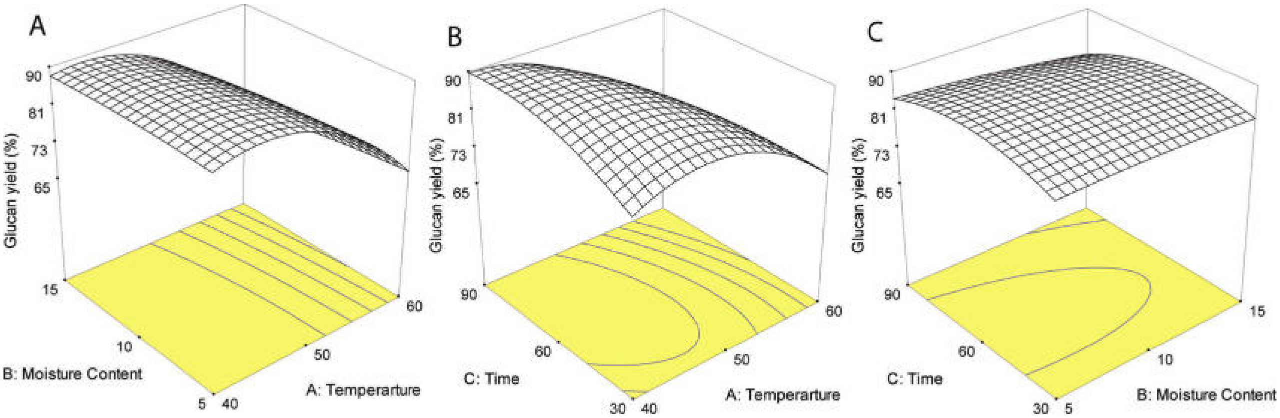


Figure 1. Response surface for the glucose yield from common reed pretreated by COSLIF at (a) temperature and moisture content, (b) temperature and reaction time, (c) moisture content and reaction time, followed by enzymatic cellulose hydrolysis (15 FPUs of cellulase per gram of glucan).

Figure 1A shows the effects of reaction temperature and biomass moisture content on glucan yield that equals glucan retention multiplied by glucan digestibility. Regardless of temperature, the moisture content alone, between 5 and 15% had little effect on glucan yield. Reaction temperatures, between 40 and 50 °C, did not differentially impact glucose yields, while the higher reaction temperature (60 °C) resulted in a much lower sugar yield. As shown in Figure 1B, at a long reaction time (90 min), increasing reaction temperature significantly decreased the sugar yield, mainly due to overhydrolysis of polysaccharides. There was a maximum glucan yield at approximately 50 °C for a short reaction time (30 min). Low reaction temperatures prevented overhydrolysis of the glucan, resulting in high glucan retention. However, when pretreatment conditions were not sufficient, the enzymatic digestibilities were much lower than those of well-pretreated samples (90%) (Table 2). Therefore, a trade-off between pretreatment and hydrolysis was identified (i.e., maximum glucan yield) at two points: 40 °C for 90 min and 50 °C for 60 min. Figure

1C suggests that the maximum sugar yields were obtained when the reaction time was approximately 60 min, regardless of biomass moisture contents. All data suggested that biomass with a moisture content ranging from 5-15% did not affect pretreatment efficiency. The optimal pretreatment conditions for common reed were found to be 50 °C and 60 min, regardless of moisture content between 5 and 15%. After COSLIF treatment, ~93% of the glucan was retained, while 65% of the xylan and 28% of the lignin were removed.

3.3. Effect of Enzyme Loading and Mass Balance

Since cellulase is still a relatively costly biocatalyst accounting for a significant fraction of the processing costs for cellulosic ethanol production (approximately 30-100 cents per gallon of ethanol),[24] we studied the effects of an enzyme decrease from 15 to 5 FPUs per gram on glucan digestibility. Figure 2 shows the glucan digestibility profiles of the common reed pretreated under the optimum condition (50 °C, 1 atm, and 60 min) at different enzyme loadings. Since enzymatic cellulose hydrolysis involves a rate-limiting primary cellulose hydrolysis (soluble cellodextrinrelease from solid cellulose), and a fast secondary cellulose hydrolysis (glucose generation from cellodextrins mainly mediated by β-glucosidase),[26] we first tested the effects of decreased cellulase loadings from 15 to 10 to 5 FPUs per gram of glucan, with a fixed high β-glucosidase loading (30 units per gram of glucan). High β-glucosidase loading can prevent any possible cellobiose inhibition. At a high enzyme loading (15 FPUs per gram of glucan), glucan digestibility reached 94% at hour 24 before leveling off (Figure 2A). When cellulase loading was decreased, glucan digestibility decreased slightly. At a low cellulase loading (5 FPUs per gram of glucan), glucan digestibilities were 87% at hour 12, 90% at hour 24, and nearly leveled off after 24 h.

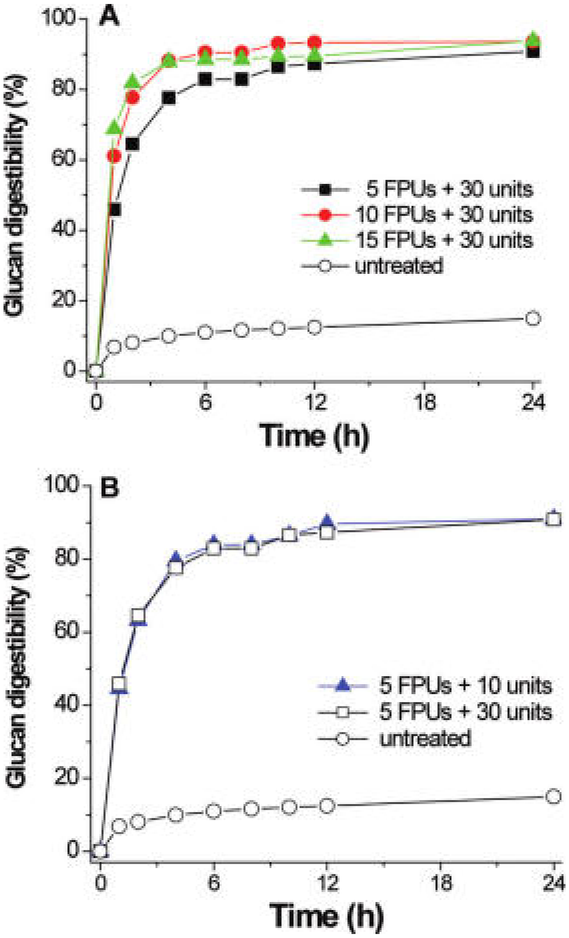


Figure 2. Enzymatic cellulose hydrolysis profiles for the COSLIF-pretreated common reed at different enzyme loadings (A, various cellulase and 30 units of beta-glucosidase; and B, 5 FPUs of cellulase and 10 or 30 units of beta-glucosidase).

Furthermore, we investigated the effect of 3-fold reduction in β-glucosidase from 30 to 10 units of per gram of glucan on glucan digestibility at 5 FPUs per gram of glucan. As shown in Figure 2C, a 3-fold reduction in total cellulase loading resulted in only 1-2% decrease in final glucan digestibility, and decreased hydrolysis rate only during the first 12 h. There was no significant difference in hydrolysis rates and final glucan digestibilities.

Figure 3 presents the mass balance of common reed pretreated by the COSLIF process and hydrolysis with 5 FPUs of cellulose as well as 10 units of β-glucosidase per gram of glucan. The enzymatic digestibilities at a low enzyme loading were 90% for glucan and 46% for xylan, respectively. The overall glucose and xylose yields, including enzymatic hydrolysis and pretreatment (water stream), were 88% and 71%, respectively.

3.4. Surface Morphology and Substrate Accessibility

Figure 4 shows the surface morphology changes in intact and COSLIF treated common reed samples. The intact plant cell wall structures of common reed presents its plant cell vascular bundles and its fibril structure (Figure 4A). Concentrated H_3PO_4 can overcome biomass recalcitrance by dissolving crystalline cellulose fibers, accompanied with increasing cellulose accessibility.[4,9,22] The sequential washing by the organic solvent can partially remove lignin.[9] A well-treated lignocellulose sample (85% H_3PO_4, 50 °C and 60 min) shows no fibrous structure (Figure 4C), suggesting that all fibrous structures of the lignocellulose were completely disrupted. However, this disruption required sufficient reaction time at the set temperature.[9] Figure 4B shows that 20 min reaction time at 50 °C looks to break large fibrils of common reed but is not as efficient as that in Figure 4C. We further measured the substrate accessibility before and after the COSLIF pretreatment. This measurement was based on adsorption of a nonhydrolytic fusion protein TGC containing green fluorescent protein and a cellulose-binding module.[23] Through the COSLIF pretreatment, the total substrate accessibility to cellulase (TSAC) increased from 0.35 ± 0.056 to 16.1 ± 1.3 m^2 per gram of biomass (Table 3). To eliminate interference from the remaining lignin and other noncellulose components, cellulose accessibility to cellulase (CAC) was measured on the basis of the adsorption of TGC after blocking with BSA. The CAC values of the intact common reed and pretreated common reed were 0.14 ± 0.035 and 13.1 ± 1.1 m^2 per gram of biomass, respectively. This result suggested that COSLIF can increase substrate accessibility 93.6-fold and yield a cellulosic product with high substrate digestibility mediated by cellulase and fast enzymatic hydrolysis rate even at a low enzyme loading. A 14.4- fold increase in noncellulose accessibility (NCAC), from 0.21 to 3.03 m^2 per gram of biomass, was much lower than a 93.6- fold increase in CAC, suggesting the importance of increasing cellulose accessibility through biomass pretreatment.

4. DISCUSSION

This study showed that very high overall yields (88% for glucose and 71% for xylose) were achieved for the COSLIFpretreated common reed at a low cellulase loading (5 FPUs of cellulase and 10 units of β-glucosidase per gram of glucan) within 24 h hydrolysis. The optimal pretreatment conditions through surface response methodology were 50 °C, 1 atm, and 60 min in the presence of 85% H_3PO_4, regardless of the moisture contents of the feedstock, from 5 to 15% (w/w). Glucan digestibility (94%) of the pretreated common reed at a high enzyme loading was slightly lower than the previous results (i.e., 96-97%) for corn stover, switchgrass, poplar, and hemp hurds. This small difference was attributed to less efficient lignin removal (28% of overall lignin) in the modified COSLIF as compared to those achieved (40-50% of overall lignin)[4,10] by decreasing the use of organic solvent.

Water in ~5-15% moisture content biomass did not dilute concentrated phosphoric acid significantly below the critical values (e.g., 80-83%) as a cellulose solvent.[10,22] The moisture contents of harvested biomass range widely from ~5 to 40% w/w, depending on the harvesting season and biomass type.[27] Winter harvesting of the standing bioenergy plants after natural drying to 5-15% moisture contents, such as common reed, would save feedstock transportation costs as compared to that of freshly cut wet biomass feedstock. This study suggested the technological feasibility of efficient sugar release from a perennial grass, the common reed.

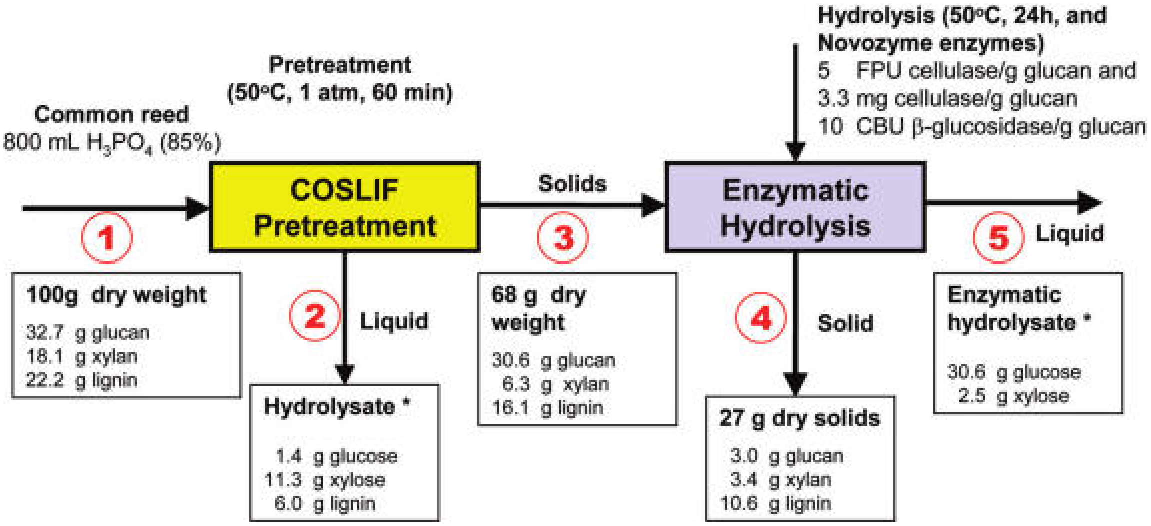


Figure 3. Mass balance for common reed pretreated by COSLIF followed by enzymatic hydrolysis at a low enzyme loading (5 FPUs of cellulase and 10 units of *β*-glucosidase per gram of glucan).

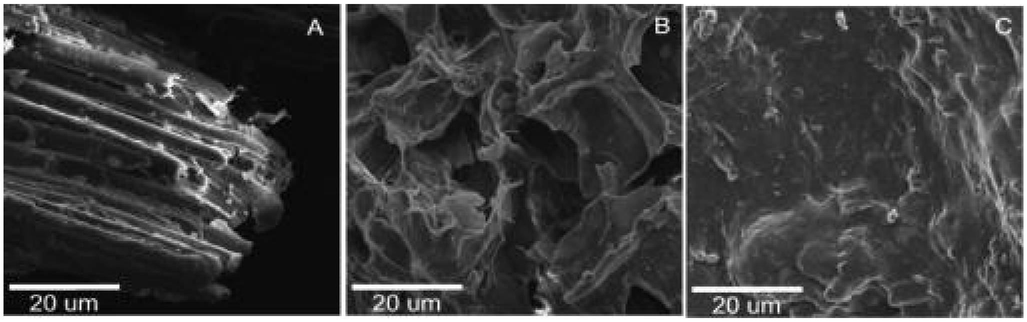


Figure 4. SEM images for the common reed samples before (A) and after the COSLIF pretreatment (B, 20-min dissolution; and C, 40 min dissolution).

Cost analysis associated with enzyme costs and sugar-toethanol revenues suggests that decreasing cellulase use would compensate for the slight revenue loss, resulting from a slightly low overall sugar yields at a decreased enzyme loading (Figure 5). It was estimated that approximately 79.4 and 75.4 gallons of cellulosic ethanol per ton of common reed could be produced at high (15 FPUs per gram of glucan) and low (5 FPUs pergram of glucan) cellulase loadings, respectively. If cellulose costs are \$0.30 per gallon based on an enzyme use of 15 FPUs per gram of glucan, a 3-fold reduction in cellulase use can save \$0.20 per gallon and decrease the ethanol revenues of \$0.126 per gallon, resulting in a net savings of 7.4 cents per gallon of ethanol, \$5.55 per ton of common reed, or a \$3.7 million of annual cost savings for a biorefinery processing 2000 tons of biomass per day. The cost saving would increase drastically to \$0.27 and \$0.53 per gallon of ethanol, if the cellulase costs were \$0.60 and \$1.20 per gallon, respectively. On the other hand, if cellulase costs were decreased to \$0.15 per gallon of ethanol, a saving in enzyme cost could not be enough to compensate for the sugar loss.

Table 3. The Substrate Accessibilities (TSAC, Total Substrate Accessibility to Cellulase; CAC, Cellulose Accessibility to Cellulase; and NCAC, Noncellulose Accessibility to Cellulase) for Intact and COSLIF-Pretreated Biomass.

sample	TSAC (m^2/g biomass)	CAC (m^2/g biomass)	NCAC (m^2/g biomass)
intact	0.35±0.056	0.14±0.035	0.21±0.066
pretreated	16.1±1.3	13.1±1.1	3.0±1.7

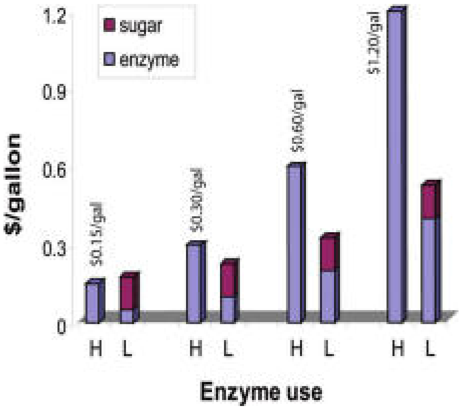


Figure 5. Enzyme cost analysis under the low (L, 5 FPUs of cellulase per gram of glucan) and high (H, 15 FPUs of cellulase per gram of glucan) cellulase loadings. Given an assumption of ethanol fermentation yield) 95% of theoretical yield for glucose and xylose, the overall ethanol yields were 79.4 and 75.4 gallons per ton of dry common reed, at the high and low enzyme loadings, respectively. The selling price of cellulosic ethanol was assumed to be $2.50 per gallon.

The COSLIF technology may be regarded as a nearly generic pretreatment. It has previously been shown to efficiently increase the glucan digestibility of a relatively broad range of feedstocks, including corn stover, switchgrass, hemp hurds, and poplar.[4,10] This study extended the range of feedstocks to the common reed and also made improvements in the COSLIF process. Different from widely studied dilute acid pretreatment,[28-31] which substantially removes hemicelluloses thereby disrupting the linkages among cellulose, hemicellulose, and lignin, the COSLIF pretreatment not only partially removes lignin and hemicelluloses, but also substantially disrupts the fibrillar structure of biomass. The resulting fast hydrolysis rates and high glucan enzymatic digestibilities of the COSLIF-pretreated common reed are attributed to (i) more efficient biomass structure destruction, qualitatively shown by SEM images (Figure 4), and (ii) higher substrate accessibility to cellulose (Table 3).

5. CONCLUSION

The pretreatment conditions were optimized for the common reed through surface response methodology for the maximal release of soluble sugars. At a low enzyme loading (5 FPUs of cellulase and 10 units of β-glucosidase), the overall glucose and xylose yields were 88% and 71%, respectively. Low use of costly cellulase would significantly improve the overall economics of cellulosic ethanol production. Since the COSLIF technology is still at a relatively early stage of development, more detailed economic analyses, based on rigorous Aspen-plus models are needed to understand its potential for practical applications. This study clearly suggests that currently growing *Phragmitez*, an invasive weed, can be used for bioenergy feedstock and would be planted as a bioenergy crop at marginal wetlands in the future.

ACKNOWLEDGMENTS

This work was supported mainly by a DOD grant (W911SR- 08-P-0021), and partially by the USDA-sponsored Bioprocessing and Biodesign Center, and the DOE BioEnergy Science Center (BESC). NS was partially supported by the ICTAS scholar program. NT was supported by the NSF REU program. The authors are grateful for the free cellulase samples provided from Novozymes North American.

5.1. Note Added After ASAP Publication

The version of this paper that was published on the ASAP website May 28, 2009 had errors in the presentation of data for run 20 in Table 2. The corrected version of this paper was reposted to the Web June 2, 2009.

NOMENCLATURE

BSA = bovine serum albumin
CAC = cellulose accessibility to cellulase
EG = ethylene glycol
GE = glucose equivalent
NCAC = noncellulose accessibility to cellulase

MC = moisture content (%)
QS = quantitative saccharification
RSM = response surface methodology
TGC = a nonhydrolytic fusion protein, containing a green fluorescence protein and a cellulose-binding module
TSAC = total substrate accessibility to cellulose

REFERENCES

[1] Bernardez, T. D.; Lyford, K.; Hogsett, D. A.; Lynd, L. R. Adsorption of *clostridium thermocellum* cellulases onto pretreated mixed hardwood, Avicel, and lignin. *Biotechnol. Bioeng.* 1993, *42* (7), 899–907. Ooshima, H.; Burns, D. S.; Converse, A. O. Adsorption of cellulose from *trichoderma reesei* on cellulose and lignacious residue in wood pretreated by dilute sulfuric-acid with explosive decompression. *Biotechnol. Bioeng.* 1990, *36* (5), 44

[2] Cullis, I. F.; Saddler, J. N.; Mansfield, S. D. Effect of initial moisture content and chip size on the bioconversion efficiency of softwood lignocellulosics. *Biotechnol. Bioeng.* 2004, *85*, 413–421.

[3] Davies, L. C.; Carjas, C. C.; Novais, J. M.; Martins-Dias, S. Phytoremediation of textile effluents containing azo dye by using *Phragmites australis* in a vertical flow intermittent feeding constructed wetland. *Ecol. Eng.* 2005, *25*, 594–605.

[4] Demain, A. L.; Newcomb, M.; Wu, J. H. D. Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Re*V. 2005, *69*, 124–154.

[5] DOE; Office of Energy Efficiency and Renewable Energy; Office of Science, Breaking the Biological Barriers to Cellulosic Ethanol: A joint Research Agenda. A Research Roadmap Resulting from the Biomass to Biofuels Workshop. http://www.doegenomestolife.org/biofuels/2006.

[6] Duke, J. Ecosystematic data on economic plants. *Quart. J. Crude. Drug. Res* 1979, *17* (3-4), 91–110.

[7] Fortman, J. L.; Chhabra, S.; Mukhopadhyay, A.; Chou, H.; Lee, T. S.; Steen, E.; Keasling, J. D. Biofuel alternatives to ethanol: pumping the microbial well. *Trends Biotechnol.* 2008, *26* (7), 375–381.

[8] Hong, J.; Ye, X.; Zhang, Y. H. P. Quantitative determination of cellulose accessibility to cellulase based on adsorption of a nonhydrolytic fusion protein containing CBM and GFP with its applications. *Langmuir* 2007, *23* (25), 12535–12540.

[9] Hong, J.; Ye, X.; Wang, Y.; P., Z. Y.-H. Bioseparation of recombinant cellulose binding module-protein by affinity adsorption on an ultra-high-capacity cellulosic adsorbent. *Anal. Chem. Acta* 2008, *621*, 193– 199.

[10] Li, J.; Wen, Y.; Zhou, Q.; Xingjie, Z.; Li, X.; Yang, S.; Lin, T. Influence of vegetation and substrate on the removal and transformation of dissolved organic matter in horizontal subsurface-flow constructed wetlands. *Biores. Technol.* 2008, *99*, 4990–4996.

[11] Lynd, L. R.; Laser, M. S.; Bransby, D.; Dale, B. E.; Davison, B.; Hamilton, R.; Himmel, M.; Keller, M.; McMillan, J. D.; Sheehan, J.; Wyman, C. E. How biotech can transform biofuels. *Nat. Biotechnol.* 2008, *26* (2), 169–172.

[12] Lynd, L. R.; Weimer, P. J.; van Zyl, W. H.; Pretorius, I. S. Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Re*V. 2002, *66*, 506–577.

[13] Moxley, G.; Zhang, Y.-H. P. More accurate determination of acidlabile carbohydrate composition in lignocellulose by modified quantitative saccharification. *Energy Fuels* 2007, *21*, 3684–3688.

[14] Moxley, G. M.; Zhu, Z.; Zhang, Y.-H. P. Efficient sugar release by the cellulose solvent based lignocellulose fractionation technology and enzymatic cellulose hydrolysis. *J. Agric. Food Chem.* 2008, *56* (17), 7885–7890.

[15] Ryan, T. P. *Modern Experimental Design; WileyInterscience*: New York, 2007.

[16] Schell, D. J.; Farmer, J.; Newman, M.; McMillan, J. D. Dilutesulfuric acid pretreatment of corn stover in pilot-scale reactor: investigation of yields, kinetics, and enzymatic digestibilities of solids. *Appl. Biochem. Biotechnol.* 2003, *105/108*, 69–85. (29) Lloyd, T. A.; Wyman, C. E. Combined sugar yields for dilute sulfuric acid pretreatment of corn stover followed by enzymatic hydrolysis of the remaining solids. *Biores. Technol.* 2005, *96*, 1967–1977.

[17] Shardendu, Salhani, N.; Boulyga, S. F.; Stengel, E. Phytoremediation of selenium by two helophyte species in subsurface flow constructed wetland. *Chemosphere* 2003, *50*, 967–973.

[18] Sim, C.; Yusoff, M.; Shutes, B.; Ho, S.; Mansor, M. Nutrient removal in a pilot and full scale constructed wetland, Putrajaya city, Malaysia. *J. En*V*iron. Manage* 2008, *88*, 307–317.

[19] Sluiter, A.; Hames, B.; Ruiz, R.; Scarlata, C.; Sluiter, J.; Templeton, D.; Crocker, D. Determination of structural carbohydrates and lignin in biomass. http://devafdc.nrel.gov/pdfs/9572.pdf. In Laboratory Analytic Procedure LAP-002, 2006.

[20] Wyman, C. E.; Dale, B. E.; Elander, R. T.; Holtzapple, M.; Ladisch, M. R.; Lee, Y. Y. Comparative sugar recovery data from laboratory scale application of leading pretreatment technologies to corn stover. *Biores. Technol.* 2005, *96*, 2026–2032.

[21] Wyman, C. E. What is (and is not) vital to advancing cellulosic ethanol. *Trends Biotechnol.* 2007, *25* (4), 153–157.

[22] Zhang, Y.-H. P.; Berson, E.; Sarkanen, S.; Dale, B. E. Sessions 3 and 8: Pretreatment and Biomass Recalcitrance: Fundamentals and Progress. *Appl. Biochem. Biotechnol.* 2009, *153*, 80–83.

[23] Zhang, Y.-H. P.; Cui, J.-B.; Lynd, L. R.; Kuang, L. R. A transition from cellulose swelling to cellulose dissolution by o-phosphoric acid: Evidences from enzymatic hydrolysis and supramolecular structure. *Biomacromolecules* 2006, *7* (2), 644–648.

[24] Zhang, Y.-H. P.; Ding, S.-Y.; Mielenz, J. R.; Elander, R.; Laser, M.; Himmel, M.; McMillan, J. D.; Lynd, L. R. Fractionating recalcitrant lignocellulose at modest reaction conditions. *Biotechnol. Bioeng.* 2007, *97* (2), 214–223.

[25] Zhang, Y.-H. P.; Himmel, M.; Mielenz, J. R. Outlook for cellulose improvement: Screening and selection strategies. *Biotechnol. Ad*V. 2006, *24* (5), 452–481.

[26] Zhang, Y.-H. P.; Lynd, L. R. Toward an aggregated understanding of enzymatic hydrolysis of cellulose: Noncomplexed cellulase systems. *Biotechnol. Bioeng.* 2004, *88*, 797–824.

[27] Zhang, Y.-H. P. Reviving the carbohydrate economy via multiproduct biorefineries. *J. Ind. Microbiol. Biotechnol.* 2008, *35* (5), 367– 375.

[28] Zhang, Y.-H. P.; Schell, D. J.; McMillan, J. D. Methodological analysis for determination of enzymatic digestibility of cellulosic materials. *Biotechnol. Bioeng.* 2007, *96* (1), 188–194.

[29] Zhu, Z.; Sathitsuksanoh, N.; Vinzant, T.; Schell, D. J.; McMillan, J. D.; Zhang, Y.-H. P. Comparative study of corn stover pretreated by dilute acid and cellulose solvent-based lignocellulose fractionation: Enzymatic hydrolysis, supramolecular structure, and substrate accessibility, *Biotechnol. Bioeng*. 2009, published online, DOI: 10.1002/bit.22307.