TECHNOLOGY DRIVEN. WARFIGHTER FOCUSED.

EDGEWOOD CHEMICAL BIOLOGICAL CENTER

# IN-HOUSE LABORATORY INDEPENDENT RESEARCH PROGRAM

# ANNUAL REPORT FY10

U.S. ARMY 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 FY10, with two receiving funding directly from the Army through the Competitive ILIR review process. This Annual Review contains technical reports from each of the ten ECBC produced, ILIR-funded projects.

# ECBC FY10 ILIR Technical Report Executive Review

Ten projects were supported by Edgewood Chemical Biological Center (ECBC) and Army Competitive In-house Laboratory Independent Research (ILIR) funds in FY10. Together, the ECBC and Army ILIR projects covered multiple areas of interest across the chemical, biological, and physical sciences. Of the awarded projects, three fit into the "Molecular Toxicology" area of interest (Wade, Dorsey, Bevilacqua), two projects in the area of "Rational Molecular and Nano-System Design" (Sekowski, Banks), two in "Surface Science" (Cabalo, Karwacki), and one project in "Novel Concepts in Nano-Scale Chemical and Biological Sensing" (Christesen). Two ILIR continuation projects fell outside the areas of interest for FY10, but fit under the general catagory of chemical and biological basic science (Hondrogiannis, Calm respectively). These projects were presented at the Assistant Secretary of the Army (Acquisition, Logistics, and Technology) (ASA(ALT)) Basic Research Review, in Arlington, VA, Summer 2009.

Total FY10 funding in support of ILIR projects performed at ECBC totaled approximately $1.351M. The projects are briefly summarized below:

"Analysis of production methods of tetramethylenedisulfotetramine based on sulfamide-formaldehyde molar ratios" (PI: Hondrogiannis). In this study, a highly potent neurotoxin, tetramethylenedisulfotetramine (TETS), was synthesized and characterized by spectral analysis. The ratio of sulfamide to formaldehyde and the choice of acid used in the reaction had the greatest effect on the reaction products and resultant yield. Hydrochloric acid afforded high yields of TETS, while sulfuric acid afforded high yields of HEXS with little or no TETS.

"Elucidating the mechanism of protein transport for the type III secretion system in *Burkholderia mallei*" (PI: Dorsey). Researchers used site directed mutagenesis of cloned *B. mallei* BsaS to examine the effect of conformational changes on ATP hydrolysis. If successful, the approach opens a potential screening approach for identifying treatments for glanders which targets the T3SS ATPase.

"Biotic-abiotic interfaces within a nanostructured polymer matrix platform: Towards a completely abiotic system" (PI: Sekowski). Chemically functionalized polystyrene-b-poly(2-vinyl-pyridine) (PS-b-P2VP) block copolymers were used to develop glucose responsive and ovalbumin-specific color-changing thin films. Such functional abiotic polymers could be designed for use as autonomous sensor platforms or as integrated functional materials on or in the warfighter's kit.

"Azomethine ylide click chemistry applications to nanotechnology and biochemistry" (PI: Banks). This work aims to discover efficient chemical reactions that occur under mild conditions that can be used to produce biologically relevant molecules and nano-materials. Here, substituted aziridines that readily form azomethine ylides were identified. Aziridines with carbon or nitrogen anionic centers were also found to have accelerated reaction rates.

"Investigation of molecule-surface interactions with overtone adsorption spectroscopy and computational methods" (PI: Cabalo). In this study, empirical overtone spectra were compared to spectra calculated from density functional theory (DFT) as a means of validating computational models of intermolecular and molecule-surface interactions. With the enhanced ability to model intermolecular and molecule-surface interactions, the warfighter's ability to detect explosives, chemical and biological agents, determine environmental fates of these materials, as well as predict surface mediated catalysis, can be improved.

"Design and testing of a thermostable platform for multimerization of single-domain antibodies" (PI: Calm). Single domain antibodies specific for ricin were used to create two proteins consisting of binding sites tethered to a coiled-coil backbone with a dimerization domain derived from the SMC protein of the thermophile *Pyrococcus furiosus.* Both proteins displayed an increase in avidity for ricin. This platform demonstrates a 'plug and play' approach to multimerization of single domain antibodies and can be used for the addition of signaling and/or mitigation activities.

"Asymmetric torsion equations for large molecules" (PI: Karwacki). Simulations and modeling were used to calculate physical properties of chemical nerve agents and stimulants. The model correctly reproduced the density and vapor pressure within 10% of empirical measurements. These simulations may reduce the risk of future experimentation on dangerous chemical agents.

"Understanding the role of physical and chemical adsorption on the Raman enhancement from metallic nanoparticles and nanostructured surfaces" (PI: Christesen). Tethered gold nanoparticle substrates were fabricated and used to analyze molecules that form charge transfer complexes, (e.g., benzenethiol) and those that do not have strong charge transfer characteristics (e.g., aminoethanol). A preliminary analysis of the binding kinetics of aminoethanol to the gold nanoparticle surface was evaluated and preliminary density functional theory calculations to simulate the interaction between the molecule and a SERS substrate were developed.

"Discrimination of pathogenic vs. non-pathogenic *Fancisella tularensis* and *Burkholderia pseudomallei* using proteomics mass spectrometry" (PI: Wade). In this study, outer membrane proteins (OMP) were examined using mass spectrometry and used as biomarkers to discriminate between organism types within the same species. The variety and variability of the OMP provided a unique and enhanced method for pathogenic discrimination that could be extended for use in discriminating subtypes of other species.

"Ricin toxicity in BALB/c 3T3 cells: peptide biomarkers of exposure" (PI: Bevilacqua). In this work, nuclear magnetic resonance (NMR) and/or mass spectrometry was used in combination with chemometric analysis to reveal biomarker information from ricin-treated cells. Identification of the relevant proteins and metabolic pathways may reveal the survival mechanism for cells and provide insight into the related 3T3 physiological changes required to overcome ricin dosing.

# Table of Contents

# Analysis of production methods of tetramethylenedisulfotetramine based on sulfamide-formaldehyde molar ratios

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

## ABSTRACT

The products of the reaction between sulfamide and formaldehyde, obtained at room temperature from reaction mixtures with different sulfamide to formaldehyde molar ratios, were systematically examined by multinuclear NMR and GC/MS. The reaction systems studied were sulfamide-formalin in concentrated hydrochloric acid and sulfamide-dimethoxymethane in trifluoroacetic acid. In the sulfamide-formalin reaction system, the choice of the acid catalyst played a vital role in the yield of the formation of TETS.

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

## 1. INTRODUCTION

It is well known that the synthesis of TETS (tetramethylenedisulfotetramine) can be accomplished in high yields by the room-temperature condensation of sulfamide and formaldehyde at a 1:2 molar ratio in concentrated hydrochloric acid for 2 hours. Under these conditions, the reaction yields up to 94% TETS, with the remaining 6% resulting from the formation of HEXS (hexamethylenetrisulfohexamine). However, very little is known about the reaction when different molar ratios of sulfamide to formaldehyde are used and the reaction times varied.

In the present paper, we describe the products formed when the reaction is run with an excess of sulfamide and formaldehyde, respectively. Two sources of formaldehyde and acid catalyst were examined: formalin (37% formaldehyde in water) with concentrated hydrochloric acid and dimethoxymethane with trifluoroacetic acid. In addition, we examined the effect of sulfuric acid as an acid catalyst in the synthesis of TETS and the effect of nitric acid on the stability of TETS at room temperature and at 60°C.

## 2. EXPERIMENTAL

### 2.1. Materials

Sulfamide (CAS # 7803-58-9) was purchased from Spectrum. Formaldehyde was purchased from Sigma as a 37% aqueous solution (CAS # 50-00-0) stabilized by 12% methanol. Dimethoxymethane (CAS # 109-87-5) was purchased from Aldrich. Trifluoroacetic acid (CAS # 76-05-1) was purchased from Fisher Chemical.

### 2.2. Synthetic procedures

The reaction between formalin (37% formaldehyde, 12% methanol and 51% water) and sulfamide in hydrochloric acid was carried out at five different sulfamide to formaldehyde molar ratios: 1:1, 1:1.2, 1:1.7, 1:2, and 1:4. Reaction times of 2 hours, 1 day, 1 week, and 2 weeks were employed. The reactions between dimethoxymethane and sulfamide in trifluoracetic acid were each carried out for 20 hours, using three different molar ratios: 1:1, 1:2, and 1:4. All reactions were started at room temperature (23 ± 2°C) without any temperature control thereafter. After reaction, the solid product was gravity filtered, washed with ice-cooled water, and dried.

### 2.3. 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 (at

different times) a cryogenic TCI probe with enhanced detection of $^{1}H$ (inner coil) and $^{13}C$ or a cryogenic QNP probe with enhanced detection of $^{13}C$, $^{15}N$, or $^{31}P$ (inner coil) and $^{1}H$. Experiments utilized included 1-dimensional $^{1}H$, $^{13}C$, and $^{15}N$ NMR spectroscopy, as well as 2-dimensional inverse-detected $^{1}H$-$^{13}C$ HSQC and HMQC spectroscopy.

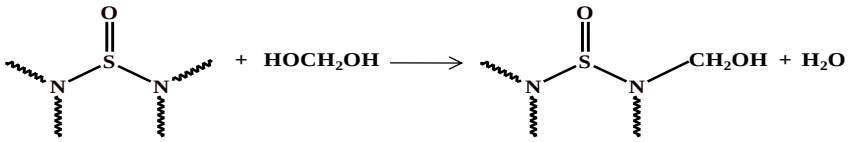
### 2.4. GC/MS analysis

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, and then increased at 10°C/minute to 280°C, where it was held for 10 minutes. The column used helium with a flow rate of 0.9 ml/min.

## 3. RESULTS

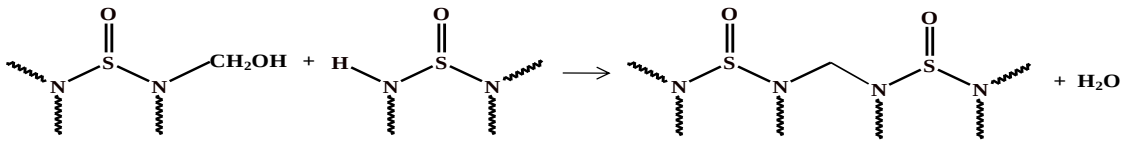
### 3.1. Reaction summary

The first step in the condensation reactions is believed to be the methylolation of the sulfamide under acid catalyzed conditions.



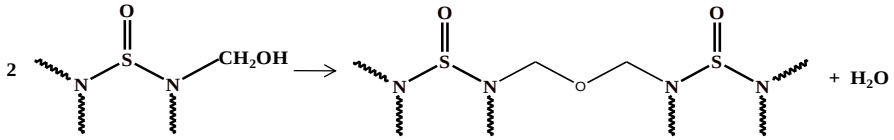

Under acidic conditions, methylol groups can undergo two major reactions. When excess sulfamide is present (sulfamide to formaldehyde ratios between 1:1 and 1:2), the formation of methylene linkages tends to predominate.

A: Formation of N-$CH_2$-N methylene linkages

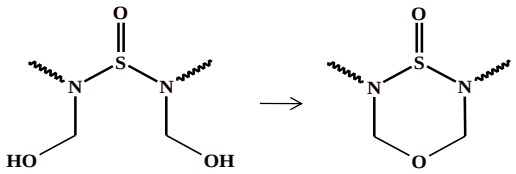



When formaldehyde is in excess (sulfamide to formaldehyde ratios between 1:2 and 1:4), the formation of dimethylene ether linkages becomes the dominant reaction pathway.

B: Formation of N-$CH_2OCH_2$-N dimethylene ether linkages
This can occur either intermolecularly:




or intramolecularly:




These reactions are all reversible, which can explain the tendency of TETS to rearrange in the presence of acids to form HEXS.

### 3.2. TETS synthesis at stoichiometric ratio

Figure 1 shows the $^{13}C$ NMR spectrum of the condensation products prepared from a mixture with a sulfamide to formaldehyde molar ratio of 1:2 at room temperature in a 24 hour reaction. The signal at 71.4 ppm indicates that TETS is the primary product of the reaction, while the HEXS signal is seen at 61.5 ppm.

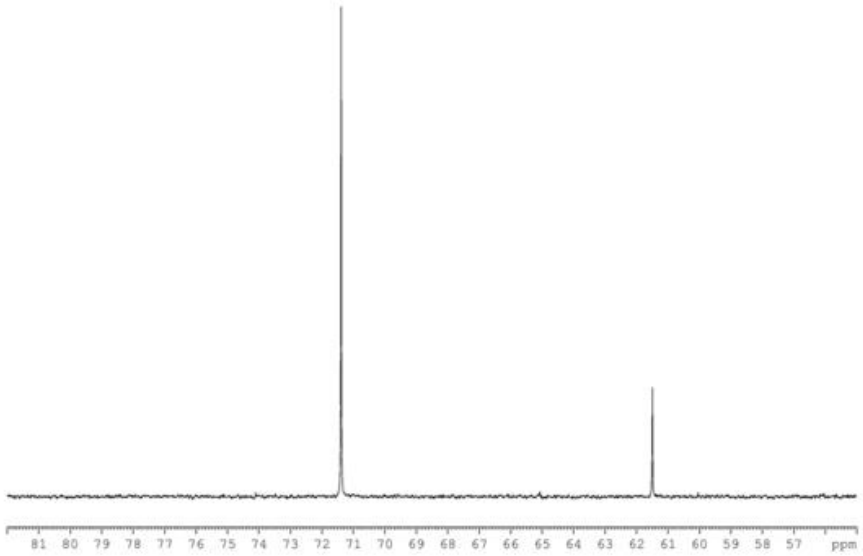



Figure 1: $^{13}C$ NMR spectrum of the product of the 1:2 molar ratio, sulfamide:formaldehyde

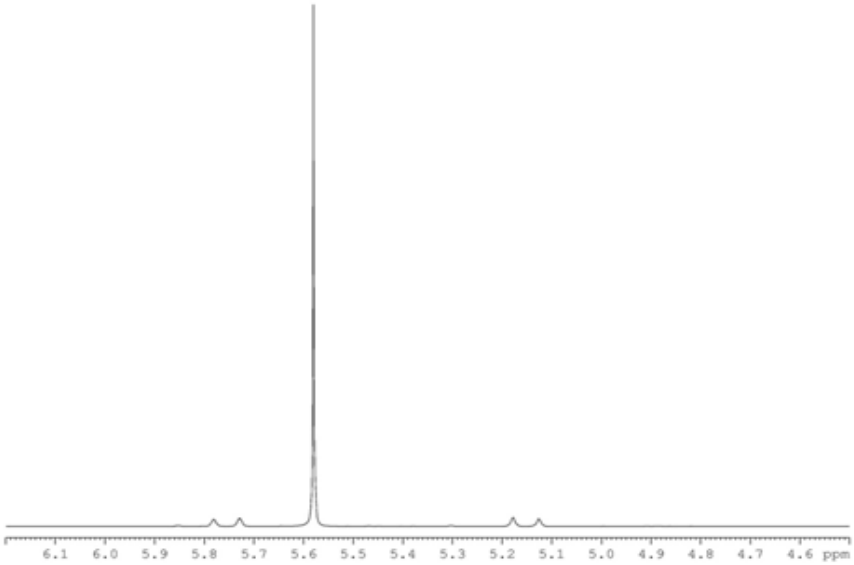



Figure 2: $^{1}H$ NMR spectrum of the product of the 1:2 molar ratio, sulfamide:formaldehyde

In the $^{1}H$ NMR spectrum, shown in Figure 2, TETS appears as the strong singlet at 5.58 ppm, while the HEXS is seen to have an AB pattern with peaks at 5.78, 5.73, 5.17, and 5.12 ppm. Analysis of the $^{13}C$ NMR integrals indicates that the product ratio is 86.3% TETS, 13.7% HEXS. In the $^{1}H$ NMR spectrum, other impurities are seen, but at much lower concentration (1-2% in total of the reaction products).

### 3.3. TETS synthesis with excess sulfamide

Figure 3 shows the $^{13}C$ NMR spectrum of the condensation products prepared from a mixture with a molar ratio of 1:1 at room temperature over a 24 hour reaction. The signal at 61.5 ppm indicates that HEXS is the primary product of the reaction, while the TETS signal is seen at 71.4 ppm. The other three major signals represent a compound that has been tentatively identified as pentamethylenedisulfopentamine (Figure 9, TETS + $NCH_2$). This compound is only present in the reaction mixture when sulfamide is present in concentrations above the stoichiometric ratio. Initially, it was not clear whether or not these signals represented one compound or several, but additional experiments made it clear that these are

all from one compound. In all TETS syntheses carried out with an abundance of sulfamide, these peaks were consistently present at a ratio of 2:2:1 for the signals at 62.3, 59.6, and 57.4 ppm, respectively.

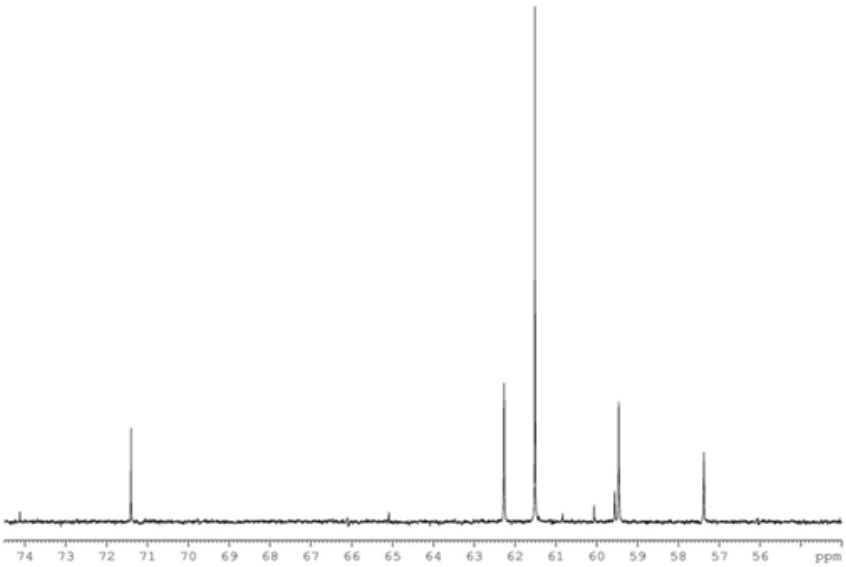



Figure 3: $^{13}$C NMR spectrum of the product of the 1:1 molar ratio, sulfamide:formaldehyde

As seen in the two-dimensional HMQC spectrum in Figure 4, the carbon signals from this compound correlate to $^{1}$H signals that have the AB pattern seen before with HEXS. The signal from TETS is seen at the correlation between 71.4 ppm in the $^{13}$C dimension and 5.58 ppm in the $^{1}$H dimension. The HEXS signals are again seen at 61.5 ppm in the $^{13}$C dimension, and the AB pattern $^{1}$H signals are seen at 5.78, 5.73, 5.17, and 5.12 ppm in the $^{1}$H dimension.

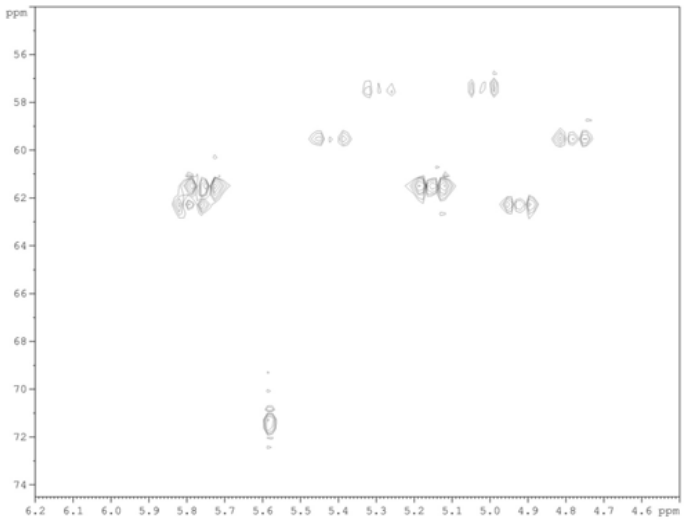



Figure 4: $^{1}$H-$^{13}$C HMQC spectrum of the product of the 1:1 molar ratio, sulfamide:formaldehyde

The signals attributed to the tentatively assigned compound are seen at (starting at the top of the spectrum):

57.4 ppm: 5.33, 5.28 (both narrowly split triplets), 5.05, and 5.00 ppm
59.6 ppm: 5.46, 5.41 (both broad doublets), 4.83, 4.78 ppm (both split singlets)
62.3 ppm: 5.82, 5.77 (both narrowly split multiplets), 4.95, and 4.90 ppm

The parenthetical descriptions of the peaks refer to their appearance in the one-dimensional $^{1}$H NMR spectrum, as shown in Figure 5.

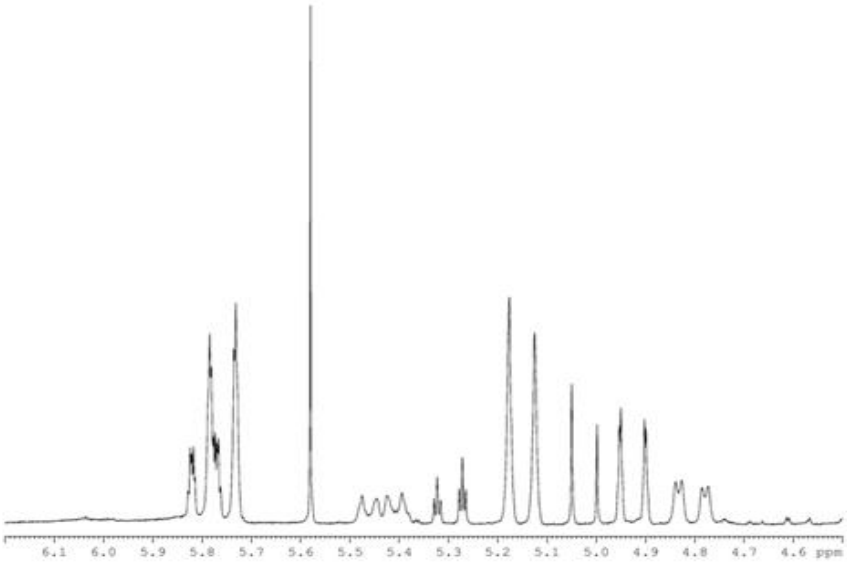



Figure 5: $^{1}$H NMR spectrum of the product of the 1:1 molar ratio, sulfamide:formaldehyde

Homonuclear decoupling experiments on these signals are what made it clear that they originate from a single compound, since when the decoupler was centered on the split multiplets of the 62.3 ppm signal (between 5.82 & 5.77 ppm), the fine coupling of the triplets in the 57.4 ppm signals (5.33 & 5.28 ppm) collapsed to form singlets. While the signals associated with 59.6 ppm cannot be directly linked to the others through observed splitting patterns, this establishes a physical link between the signals at 57.4 and 62.3ppm. One other piece of evidence as to the identity of this compound comes from the fact that a homodecoupling experiment in which the decoupler is focused on broad signals (indicative of an exchangeable proton, perhaps on a nitrogen) at 8.1 ppm (not shown in these spectra) causes a collapse of the two broad doublets associated with 59.6 ppm into singlets, while at the same time collapsing the splitting in the signals at 4.83 and 4.78 ppm.

Analysis of the reaction products of the 1:1 molar ratio synthesis by both $^{1}$H and $^{13}$C NMR yields a product distribution of 10.9% TETS, 52.3% HEXS, and 36.8% TETS + $NCH_2$ (tentative). GC/MS analysis of this sample gave major peaks for TETS at a retention time of 16.93 min and for HEXS at 27.56 min, and a minor signal at 22.78 min. The mass spectrum of the minor peak showed a molecular ion at 269 amu, and fragment ions at 241, 176, 121 (base peak), 92, and 42 amu. This molecular ion is in good agreement with the tentative assignment of TETS + $NCH_2$ (MW 269). When the sample was derivatized with BSTFA, in addition to TETS and HEXS, two other major peaks were detected with retention times of 13.24 min and 23.72 min. The former was attributed to the bis-trimethylsilyl derivative of sulfamide, based on comparison with authentic standard. The mass spectrum consisted of 225 ($M^{+\cdot}$-15), 146 (base peak), and 73 amu. The unidentified peak at 23.72 min can tentatively be assigned to the trimethylsilyl derivative of TETS + $NCH_2$, with a molecular mass of 341. The mass spectrum of this peak consisted of 326 ($M^{+\cdot}$-15), 206 (base peak), 190, 162, 120, 73, and 42 amu. The BSTFA derivatization of this compound at the NH moiety, if it occurs, would generate a molecular ion of 341. In the analogous derivatization of sulfamide under EI conditions, however, the highest mass detected was not that of the molecular ion (240 amu), but that of the $M^{+\cdot}$-15 ion (225 amu).

When the reaction was repeated with a molar ratio of 1:1.2 for one week, the amount of the unknown represented 16.5% of the reaction products as determined by NMR. Change of the molar ratio from 1:1.2 to 1:1.7 resulted in further reduction in the amount of unknown to 6.9 mole% of the reaction products. When the reaction was repeated using a molar ratio of 1:2, the unknown compound was not detected by either NMR or GC/MS, even after BSTFA derivatization. Thus, this unknown compound can be unambiguously used as a marker for all reactions run using a molar ratio between 1:1 and 1:2.

### 3.4. TETS synthesis with excess formaldehyde

Figure 6 shows a $^{1}$H NMR spectrum of the reaction products formed when excess formaldehyde was used at a molar ratio of 1:4 at room temperature for 24 hours. One obvious difference between this reaction product and that of the 1:1 or

1:2 molar ratios is that, while TETS is still strongly produced in this synthesis, there is no HEXS detectable by NMR. (In the $^{1}H$ NMR spectrum in Figure 6, TETS is clear as the strong singlet at 5.58 ppm, but the AB pattern of HEXS at 5.78, 5.73, 5.17, and 5.12 is missing.)

Two other impurity signals appeared in the spectrum, however, which can best be simplified by examining the $^{13}C$ NMR spectrum, shown in Figure 7. In that figure, the signal at 71.4 ppm is TETS, but the signals at 78.4 and 61.2 ppm are unknowns. The signal at 61.2 ppm is not HEXS - HEXS should be slightly downfield from that position, at 61.5 ppm. The $^{1}H$-$^{13}C$ HMQC experiment in Figure 8 correlates the $^{1}H$ and $^{13}C$ NMR signals, showing the TETS signal at 71.4 ppm in $^{13}C$ and 5.58 ppm in $^{1}H$ NMR. The signals at 61.2 and 78.4 ppm are seen to exhibit the AB pattern in $^{1}H$ NMR, where the resonance at 61.2 ppm in $^{13}C$ has $^{1}H$ correlations at 5.62, 5.56, 5.16, and 5.10 ppm, while the 78.4 ppm signal has correlations at 5.42, 5.38, 5.06, and 5.02 ppm. Note that the $^{1}H$ correlations of the 61.2 ppm signal confirm that it is not HEXS, since HEXS has $^{1}H$ correlations at 5.78, 5.73, 5.17, and 5.12 ppm. The $^{13}C$ NMR signal at 61.2 ppm is one-half the intensity of the signal at 78.4 ppm, which indicates that they could be part of the same compound (most likely with the signal at 78.4 ppm representing four carbons, while the signal at 61.2 ppm represents two).

This is confirmed by the fact that, in homodecoupling experiments, decoupling the area around the split doublet at 5.4 ppm (the signals from 78.4 ppm that correlate in the $^{1}H$ NMR spectrum to 5.42 and 5.38 ppm) causes the signals from 61.2 ppm that correlate at 5.78 and 5.73 ppm to collapse from triplets to singlets. Likewise, decoupling the area between the triplets causes the split doublets to collapse to singlets. The only explanation for this is that the two signals seen in the $^{13}C$ NMR spectrum, aside from TETS, both come from a single compound.

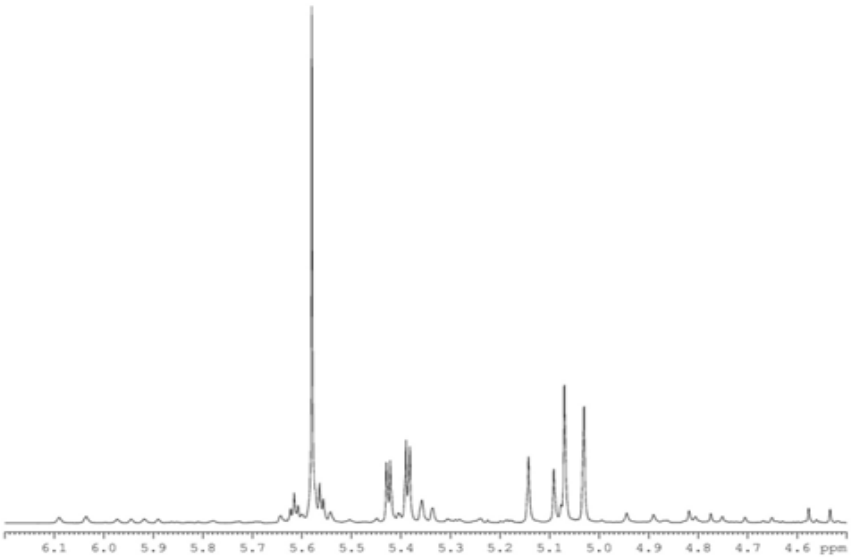



Figure 6: $^{1}H$ NMR spectrum of the product of the 1:4 molar ratio, sulfamide:formaldehyde

GC/MS analysis indicated two major peaks at 16.93 min and 22.66 min, corresponding to TETS and the unknown, respectively, and one minor peak at 20.41 min. This minor peak is in agreement with the NMR data, which shows low-abundance impurities, such as those seen in the $^{13}C$ NMR spectrum (Figure 7) as the peaks barely above the baseline at 84.4 ppm, 66.5 ppm, and possibly others. Any correlations from these signals in the HMQC, however, were too weak to be detected. GC/MS analysis indicated a product distribution of 36.3% TETS, 58.9% of the first unknown ($^{13}C$ at 51.2 and 78.4 ppm), and 4.8% of the second unknown.

By GC/MS, the first unknown (22.66 min) showed a molecular ion peak at 300 amu of very low abundance and fragments at 269, 241, 212, 178, 121 (base peak), 92, and 42 amu. The assignment of the molecular ion at 300 amu was confirmed by chemical ionization GC/MS. Based on the NMR and GC/MS data, this compound was tentatively assigned to hexamethylenedisulfodioxotetramine (TETS with the addition of two dimethylene ether linkages, in accord with the equations in section 3.1). The structure of this assignment is shown in Figure 9 (TETS + 2 $OCH_2$). In order to satisfy the symmetry elements observed by NMR, the two dimethylene ether linkages must be placed across from each other in the caged molecular structure.

The minor peak at 20.41 min, corresponding to 4.1% of the total area recorded, showed a high mass peak at 270 amu, and fragment peaks at 240 (base peak), 212, 121, 92, and 42 amu. This compound was tentatively assigned to pentamethylenedisulfomonooxotetramine (TETS + $OCH_2$ in Figure 9).

When the sample was derivatized by BSTFA, no additional peaks were detected in the GC chromatogram. When the reaction was repeated using the sulfamide to formaldehyde 1:4 molar ratio for 2 hours, NMR analysis gave the following product distribution: TETS 55.6%, TETS + 2 $OCH_2$ (300 amu) 35.6%, TETS + $OCH_2$ (270 amu) 6.9%, and HEXS 1.9%. It was very interesting to observe that, when the reaction time was extended to two weeks, the product distribution was shifted in favor of TETS. Most important was the substantial decrease in TETS + 2 $OCH_2$, which was not detected by NMR and was detectable only by GC as 0.40% of the total area observed. The percent product distribution for the two week reaction by GC/MS was TETS 92.6%, TETS + 2 $OCH_2$ 0.40%, TETS + $OCH_2$ 1.45%, and HEXS 5.5%. Again, the assignments of TETS + $OCH_2$ and TETS + 2 $OCH_2$ are tentative.

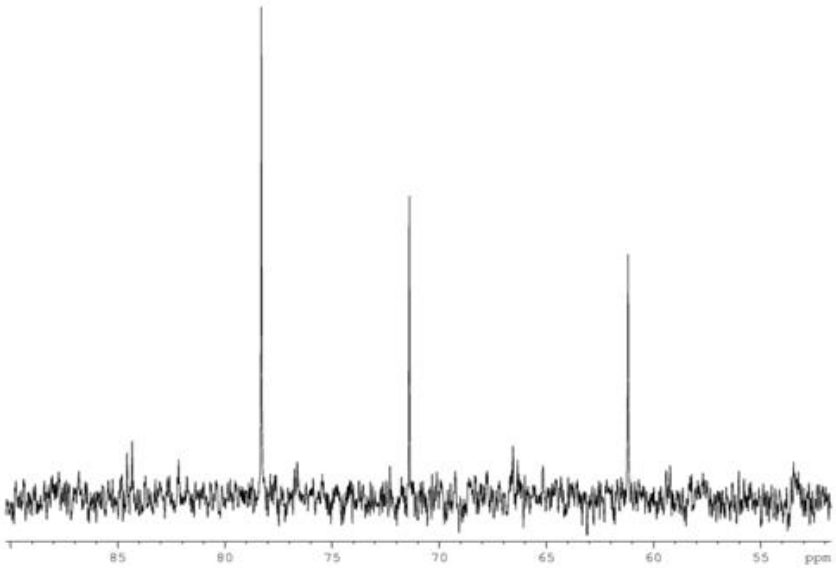



Figure 7: $^{13}C$ NMR spectrum of the product of the 1:4 molar ratio, sulfamide:formaldehyde

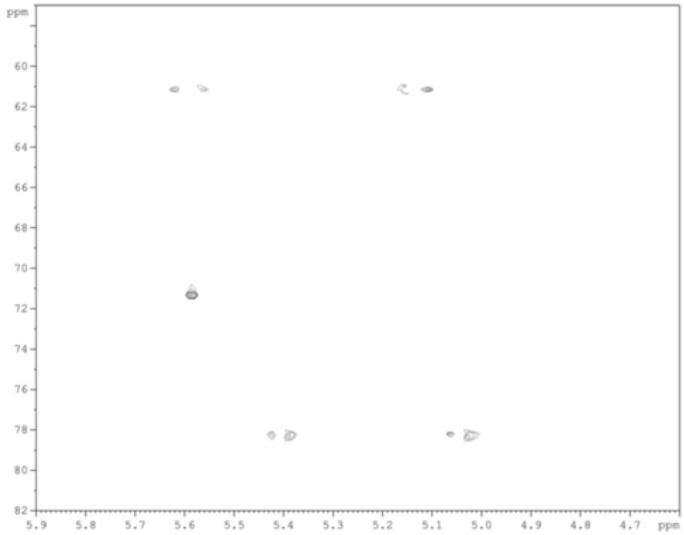



Figure 8: $^{1}H$-$^{3}C$ HMQC spectrum of the product of the 1:4 molar ratio, sulfamide:formaldehyde

Thus, the synthesis of TETS using a reactant ratio with an excess of formaldehyde leads to a product that is primarily TETS, but has one major impurity (others exist at much lower levels) that can be used as a marker for this synthetic ratio. This marker may also provide insight into how long the reaction was allowed to continue, given the drop in its concentration from a one-day synthesis to a two-week synthesis. The second unknown is also a useful marker, given that it remains at a fairly consistent level over a wide range of reaction times and is seen in all of the reactions performed with an abundance of formaldehyde, yet not seen in the absence of that excess.

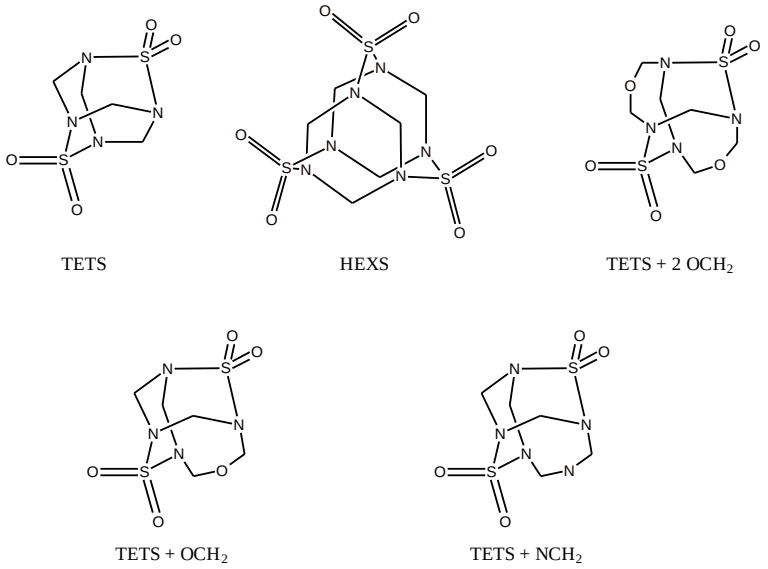



Figure 9: Chemical structures of TETS, HEXS, and the three tentative identification of unknowns with masses 300 amu (TETS + 2 $OCH_2$), 270 amu (TETS + $OCH_2$), and 269 amu (TETS + $NCH_2$).

The change in the product distribution with time under acidic conditions is attributed to the reversibility of these types of reactions, which is related to the hydration of the N-CH2-N methylene linkages to form the corresponding N,N'-dimethylolsulfamides, which can then re-condense to form different reaction products.

### 3.5. TETS synthesis under other reaction conditions

Since the effect of different molar ratios was so profound in the distribution of products of the condensation between 37% formaldehyde and sulfamide in concentrated hydrochloric acid, it was of interest to examine its effect in a different reaction system. Thus, the condensation reaction between dimethoxymethane, sulfamide, and trifluoroacetic acid was observed at room temperature for 20 hour reactions with molar ratios of 1:1, 1:2, and 1:4 of sulfamide to dimethoxymethane.

When the molar ratios of 1:4 and 1:2 were used, analysis by NMR and GC/MS of the reaction products indicated the presence of only TETS and HEXS, at 93-97% for the 1:4 and 92-98% for the 1:2 molar ratios. When the molar ratio of dimethoxymethane to sulfamide of 1:1 was used, the product distribution was 83.5% TETS, 15.4% HEXS, and 1.1% of the compound tentatively identified as TETS + $NCH_2$ previously with the excess of sulfamide (see section 3.3). As had been the case previously, by GC/MS this compound was only detected as the trimethylsilane derivative. From a forensics attribution perspective, it is worth noting that products formed from this reaction system are easily recognizable by running $^{19}F$ NMR. Residual fluorine from the trifluoroacetic acid is easily detected, as this is a highly sensitive nucleus by NMR.

These results indicate that the dependence of the reaction products on the molar ratios of the starting material is not universal, and depends on the reaction system. It is not clear whether these results are due to the choice of the dimethoxymethane as the source of the formaldehyde or due to the effect of trifluoroacetic acid in promoting the reversible reactions in this system.

It should be mentioned that the choice of the acid catalyst is very important. For example, when the reaction of sulfamide to 37% formaldehyde in a molar ratio of 1:2 was attempted in the presence of $H_2SO_4$ at room temperature for 24 hours, HEXS was formed instead of TETS to the extent of 98-99%. TETS was detected as a minor product.

### 3.6. Hydrolysis of TETS in concentrated nitric acid

Published literature on TETS indicates that it can be acid hydrolyzed by exposure to concentrated nitric acid at room temperature for 24 hours. A 50.2 mg sample of 83% TETS, 17% HEXS was dissolved in concentrated nitric acid, and the hydrolysis reaction was followed by $^1$H NMR. As can be seen in Figure 10, over the course of three days in the concentrated nitric acid, TETS gradually converted to HEXS. No other products were seen to be formed, and the reaction was far from complete in 24 hours at room temperature. When the sample was rechecked after seven days in acid, no TETS was detected, only HEXS.

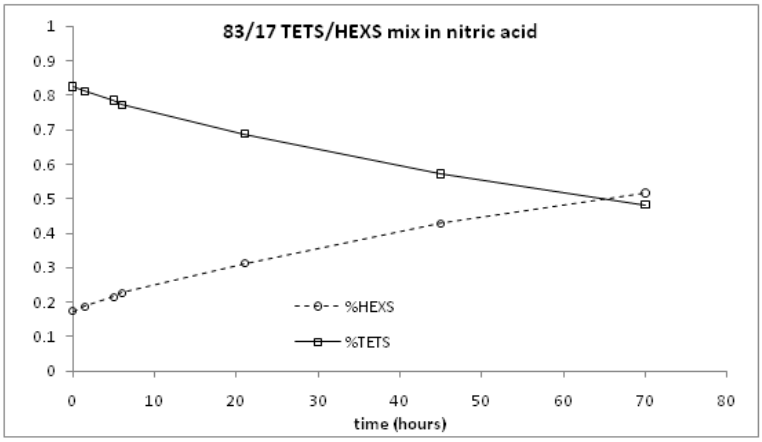


Figure 10: Time course of acid hydrolysis of TETS in concentrated nitric acid.

When TETS was dissolved in nitric acid and refluxed at 60°C, however, the conversion was much more rapid. After six hours under these conditions, conversion to HEXS was complete, with no TETS detected by NMR in the refluxed sample.

## 4. CONCLUSIONS

The products of the condensation reactions prepared at room temperature from sulfamide and formaldehyde in acid are highly dependent on the source of formaldehyde and the choice of acid employed in the reaction. The sulfamide-formalin-hydrochloric acid reaction yielded product distributions that were highly sensitive to the molar ratios of the reactants employed. When the reaction was performed with an excess of sulfamide, TETS, HEXS, and a compound tentatively identified as pentamethylenedisulfopentamine were the major reaction products. Reactions performed with an excess of formalin yielded, as major products, TETS and a compound tentatively identified as hexamethylenedisulfodioxotetramine. Only small amounts of HEXS were detected in this reaction product. On the other hand, the sulfamide-dimethoxymethane-trifluoroacetic acid reaction system was less sensitive to the molar ratios of the reactants employed, and all gave high yields of TETS contaminated with small amounts of HEXS. A general trend was observed in these reactions that, when sulfamide is in excess, an increase of HEXS is observed. In the sulfamide-formalin system, the choice of acid was crucial to the yield of TETS formed. Hydrochloric acid afforded high yields of TETS, while sulfuric acid afforded high yields of HEXS with little or no TETS.

## REFERENCES

[1] Chuang, I.; Maciel, G.E. *Macromolecules* Vol 25, No 12, (1992) p. 3204-3226

[2] Kang, J-B.; Thyagarajan, B.S.; Gilbert, E.E.; Siele, V. *Int J Sulfur Chem A* Vol 1, No 4 (1971) p. 261-268

[3] Lee, C.; Kohn, H. *J Org Chem* (1990) p. 6098-6104

# Elucidating the mechanism of protein transport for the type III secretion system in *Burkholderia mallei*

Russell Dorsey[a], Mark Guelta[a], Daniel Carmany[b], Mike Retford[a], and Wieslaw Swietnicki[c]
[a]US Army, Development and Engineering Command, Edgewood Chemical Biological Center, 5183 Blackhawk Rd, Aberdeen Proving Ground, MD, USA 21010
[b]Battelle Memorial Institute, Battelle Eastern Science and Technology Center, 1204 Technology Drive, Aberdeen, MD, USA 21001
[c]Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, MD, USA 20814

## ABSTRACT

*Burkholderia mallei* is a Gram-negative zoonotic pathogen of solipeds which can also secondarily infect human hosts causing glanders. The bacterium is highly infectious and has been used in the past as a potential biological weapon. In Western Europe and North America, glanders has been eradicated, but it persists in other areas. The two major bacterial virulence systems, the type III (T3SS) and type VI (T6SS) secretion systems, have been studied and found to be essential for the virulence in animal models of the disease. *B. mallei* has two T3SS, plant-like and animal-like, but only the latter is essential for virulence in animal models of glanders. In the current work, we have expressed, purified, and measured catalytic activity of the *B. mallei* BsaS ATPase in an *in vitro* assay.

**Keywords:** *Burkholderia mallei,* type III secretion system, BsaS ATPase

## BACKGROUND

*Burkholderia mallei*, a Gram-negative pathogen which infects horses and humans, is the causative agent of glanders [1]. The bacterium is classified as a select agent by the CDC due to its high infectivity rates and its past use as a biological weapon [2]. The pathogen uses two major virulence systems, the type III secretion system (T3SS) and the type VI secretion system (T6SS), to deliver virulence factors into host tissues [3-4]. The T3SS is found in many gram negative bacteria including *Yersinia*, *Salmonella*, *Shigella,* and enteropathogenic *E. coli* [5-9]. All the pathogens that contain T3SS are of interest to the US Army, and research into preventive and therapeutic measures is being actively pursued at the present time. The mechanism by which the T3SS recognizes and transports virulence factors extracellulary is not known, but depends on the presence of an active ATPase contained in the BsaS subunit [10-11]. The energy for the transport is thought to be derived from proton gradients but some controversy remains. Based on similarities between the T3SS proteins identified to date and the well known flagellar ATPase system, we hypothesize that the T3SS ATPase forms hexameric assemblies as catalytic units [12-15]. Interestingly, the flagellar ATPase can transition to a monomer, thus the hexameric form is not a necessary requirement for catalysis. A possible explanation is that the hexameric complex is required to uncouple the effectors from the chaperone-effector complex and the ATPase is contained in a single unit. Based on the *in silico* modeling results that we have conducted on homologous *Yersinia pestis* T3SS YscN ATPase, blocking full closure on the C-terminal of the protein or introduction of steric hindrances from adjacent subunits will render the oligomer either unable to bind or hydrolyze ATP. We will confirm this prediction by introducing small changes in the nucleotide binding site and the adjacent subunit through site directed mutagenesis of cloned *B. mallei* BsaS. If confirmed, it would help to explain how ATP hydrolysis may be coupled to molecular movements.

The T3SS in *B. mallei* is encoded chromosomally and the key components have been identified by analysis of genetic deletions in the T3SS operon [3]. Models predict that the T3SS is an $Mg^{+2}$-dependent ATPase. We propose to examine the validity of the models by expressing BsaS in *E. coli*, measuring its activity *in vitro* and correlating it with conformational changes. If successful, the approach opens a potential screening approach for identifying treatments for glanders that targets the T3SS ATPase. In the second year of the proposed project, we intend, in conjunction with our collaborators at the New York Structural Biology Center, to crystallize and determine the structure of the native and mutational inactive T3SS BsaS ATPase.

## METHODS

*Gene cloning*. The catalytic domain and the full length open reading frame (ORF) fragments of the *BsaS* gene from the *B. mallei* ATCC 233344 strain were optimized for codon usage and *E. coli* expression using *in silico* methods. The gene fragments were synthesized chemically, cloned into a pMAL (NEBiolabs) vector, and verified by sequencing (GenScript). The final construct included a 10xHis affinity tag at the C-terminus and a cleavable (enterokinase) maltose binding protein (MBP) fusion at the N-terminus.

*Protein expression*. The construct was transformed into *E. coli* T7 Express cells (NEBiolabs) and the protein expression was performed using the TB Auto (Novagen) auto induction medium supplemented with 50 mg/ml ampicillin at 22$^{o}$C. After an overnight growth, the cells were harvested by centrifugation and stored at -80$^{o}$C. To purify the full length protein, the cells were thawed, resuspended in a 2xTBS (10 mM Tris, pH=8.0, 150 mM NaCl), 40 mM imidazole, and 10% glycerol buffer, disrupted in a Microfluidizer (Microfluidics, Inc.), and the debris was removed by centrifugation at 28,000xg, 4$^{o}$C for two hours. The lysate was filtered through a 0.45 mm sterile filter and applied on a HisTrap crude FF (GE Healthcare) column connected to AKTAExpress (GE Healthcare) protein purification system. The protein was eluted from the HisTrap crude FF (GE Healthcare) column with a step gradient of 500 mM imidazole and loaded directly on a HiLoad Superdex 200 26/60 (GE Healthcare) column equilibrated in 2xTBS, 10% glycerol. The peaks containing the protein were pooled and applied to an MBPTrap HP (GE Healthcare) column and eluted with a linear gradient of 10 mM maltose. The fractions containing the protein were pooled; buffer exchanged into 2xTBS, 10% glycerol, and concentrated using Amicon Ultra 15 (10 kDa cutoff) concentrators. Purity was verified by SDS-PAGE. The concentrated protein was aliquoted and stored at -20$^{o}$C. The catalytic domain of BsaS was purified in analogous way, but due to the higher affinity for the amylose resin (MBPTrap HP column), the procedure was performed in an automated mode using the following four steps: HisTrap crude FF, desalting, MBPTrap HP, and HiLoad Superdex 200 26/60 columns.

*Enzymatic assays*. Enzymatic activity was measured by EnzCheck (Invitrogen) following phosphate release from ATP using a coupled assay relying on incorporation of the phosphate into a chromogenic substrate. Typically, the reaction was performed at 37$^{o}$C in 1xTBS buffer in the presence of 2 mM Mg+2. At specific time points, aliquots were withdrawn, quenched with 2 mM EDTA, and stored at -20$^{o}$C for further processing.

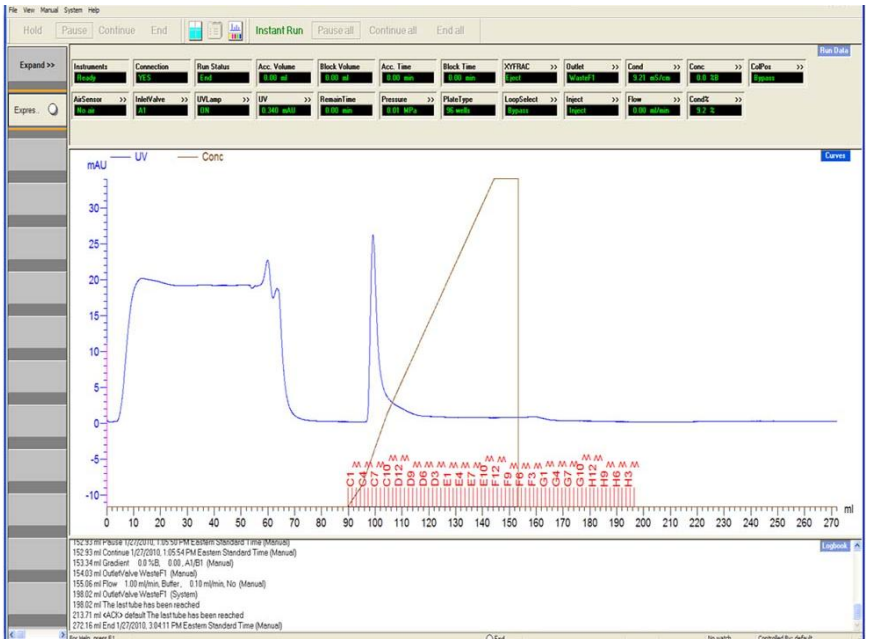



Figure 1: Last step of BsaS protein purification. The pooled fractions from size exclusion chromatography were applied on MBPTrap HP (GE Healthcare) affinity column and eluted with a linear gradient of 10 mM maltose.

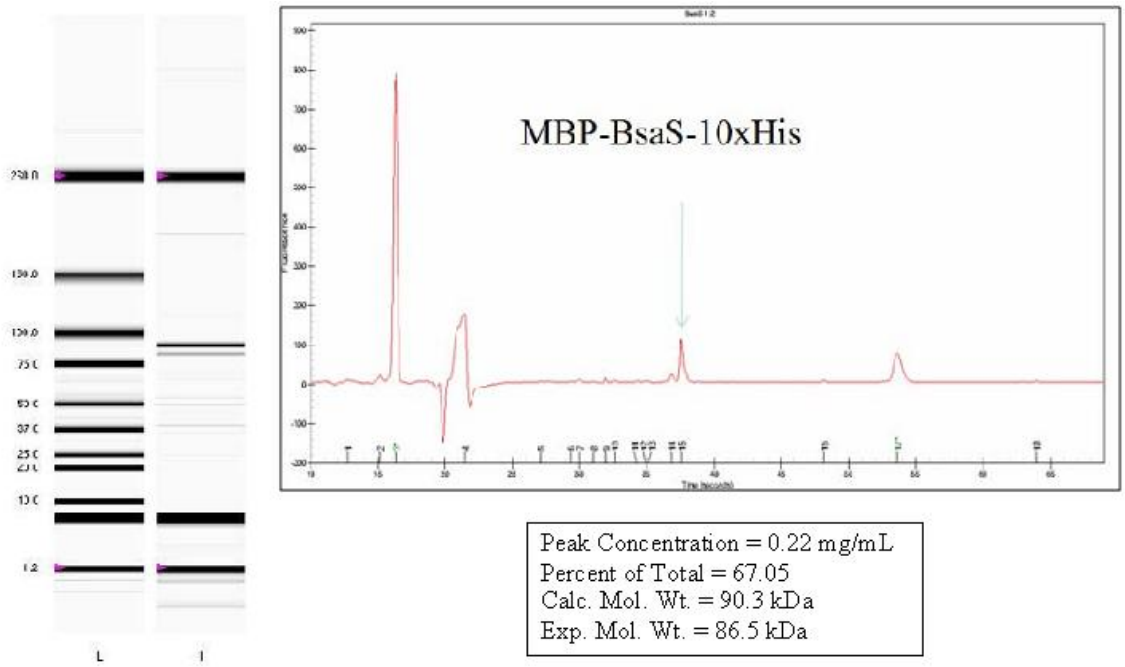



Figure 2: SDS-PAGE of purified BsaS protein. The gel was run on an Experion (BioRad) automated electrophoresis system.

## RESULTS

The catalytic domain of the BsaS protein was purified to homogeneity. The ATPase activity of the protein (Fig. 3) was found to be non-linear versus enzyme concentration (Fig. 4), and a simple kinetic model could not be applied to calculate the apparent *Km* and *Vmax*. The best approximation was obtained when the Lineweaver-Burk plot was fitted to each linear portion of the activity plot (Fig. 5).

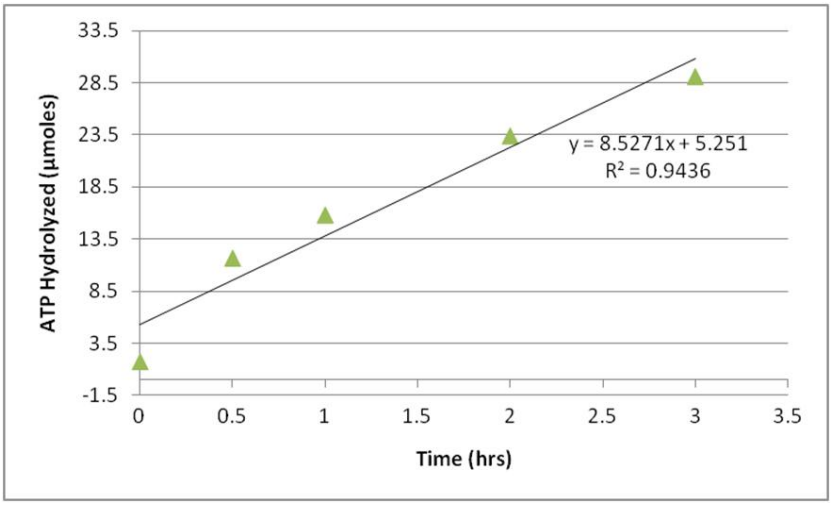



Figure 3: A representative kinetic trace of ATP hydrolysis by the BsaS protein. The hydrolysis was quantified by following phosphate release in a coupled reaction in 1xTBS, 2 mM Mg+2 at 37°C. Total amount of BsaS protein used was 6.3 µg.

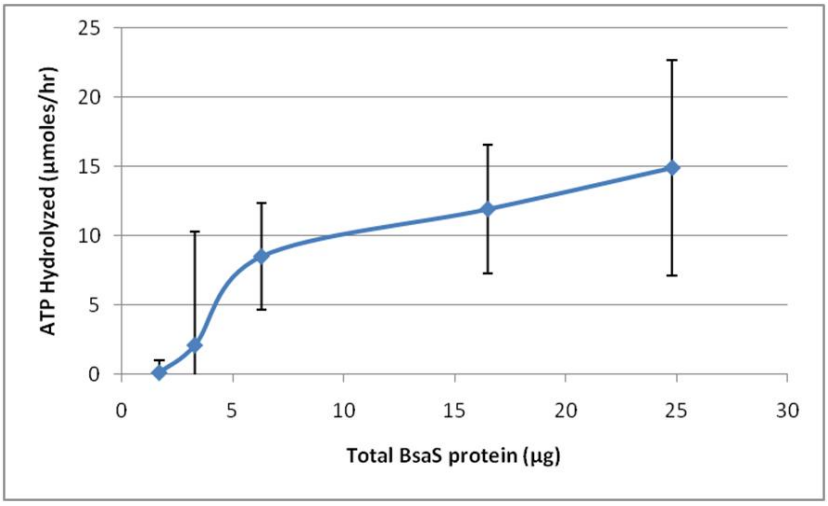



Figure 4: Non-linearity of initial ATP hydrolysis versus protein concentration. The hydrolysis was quantified by following phosphate release in a coupled reaction in 1xTBS, 2 mM Mg+2 at 37°C.

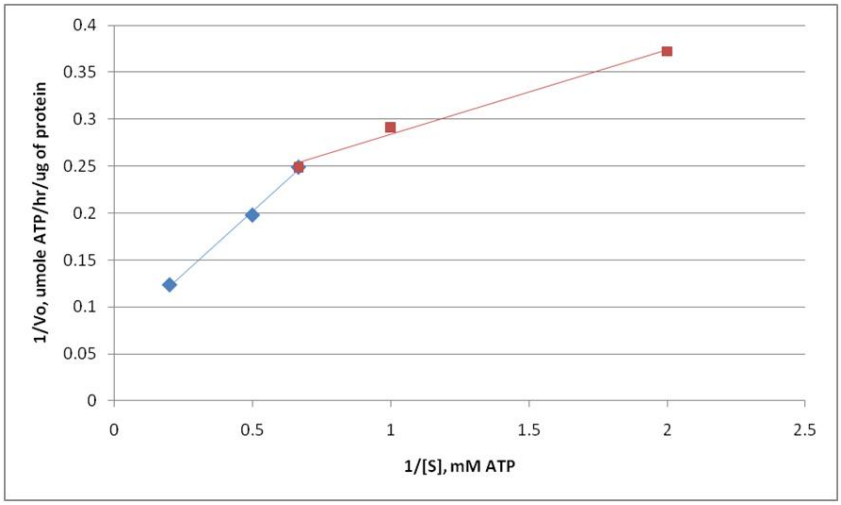



Figure 5: Allostery of ATP hydrolysis by BsaS. The apparent Km and Vmax values were determined from Lineaweaver-Burk plots for separate ranges of substrate concentrations. The hydrolysis was quantified by following phosphate release in a coupled reaction in 1xTBS, 2 mM Mg+2 at 37°C. Range of ATP concentration was 0.2-5 mM. Total concentration of BsaS protein was 6.3 µg.

## DISCUSSION

Initial attempts to express the protein fragments using native gene fragments were not successful due to extremely low yield of protein production by the *E. coli* (R. Dorsey, unpublished results). Therefore, the gene was optimized for *E. coli* expression and all cystine residues were replaced with serine to minimize protein aggregation. Model building of the BsaS protein suggested that none of the cystine residues formed intramolecular disulfide bridges (R. Dorsey, unpublished data). Also, reconstruction of the putative hexameric oligomer based on the rigid body protein docking and the *Salmonella* FliI putative hexamer did not suggest the cystine residues would be involved in intermolecular disulfide bridge formation (W. Swietnicki, unpublished data). Therefore, the replacements were judged to be justified.

The expressed BsaS protein was shown to have ATPase activity, in agreement with the computational predictions. In orthologous T3SSs, the ATPases are thought to be responsible for the release of virulence factors from their complexes

with specific chaperones before transport through the translocation pore from the bacterial cytosol directly into the mammalian cells. The N-terminal part on the ATPases is most likely involved in the release process but the details are not known. The catalytic and the N-terminal domains of the ATPases are apparently separated by a flexible linker based on the homology modeling and the available structural data. The release of the chaperones may require an oligomerization of the ATPase as the oligomeric states are observed in all homologous ATPases purified to date and the native conformation of plant-like T3SS HrcN ATPase is a dodecamer. The full length T3SS ATPases also show catalytic activity associated with the hexameric form of the protein. The non-linear dependence of activity on the enzyme concentration, also observed for the T3SS ATPases and in our experiments, may reflect the dependence of the catalytic activity on the tertiary arrangements of the complex. The tertiary arrangement may be influenced by the N-terminal domain of the protein and may involve conformational arrangements within the complex. The details of the mechanism are currently under investigation.

## REFERENCES


[1] Whitlock, G.C., D.M. Estes, and A.G. Torres, *Glanders: off to the races with Burkholderia mallei.* FEMS Microbiol Lett, 2007. **277**(2): p. 115-22.

[2] Rotz, L.D., et al., *Public health assessment of potential biological terrorism agents.* Emerg Infect Dis, 2002. **8**(2): p. 225-30.

[3] Ulrich, R.L. and D. DeShazer, *Type III secretion: a virulence factor delivery system essential for the pathogenicity of Burkholderia mallei.* Infect Immun, 2004. **72**(2): p. 1150-4.

[4] Schell, M.A., et al., *Type VI secretion is a major virulence determinant in Burkholderia mallei.* Mol Microbiol, 2007. **64**(6): p. 1466-85.

[5] Tree, J.J., et al., *Controlling injection: regulation of type III secretion in enterohaemorrhagic Escherichia coli.* Trends Microbiol, 2009. **17**(8): p. 361-70.

[6] Mota, L.J. and G.R. Cornelis, *The bacterial injection kit: type III secretion systems.* Ann Med, 2005. **37**(4): p. 234-49.

[7] McGhie, E.J., et al., *Salmonella takes control: effector-driven manipulation of the host.* Curr Opin Microbiol, 2009. **12**(1): p. 117-24.

[8] Parsot, C., *Shigella type III secretion effectors: how, where, when, for what purposes?* Curr Opin Microbiol, 2009. **12**(1): p. 110-6.

[9] Enninga, J. and I. Rosenshine, *Imaging the assembly, structure and activity of type III secretion systems.* Cell Microbiol, 2009.

[10] Andrade, A., et al., *Enzymatic characterization of the enteropathogenic Escherichia coli type III secretion ATPase EscN.* Arch Biochem Biophys, 2007. **468**(1): p. 121-7.

[11] Blocker, A., K. Komoriya, and S. Aizawa, *Type III secretion systems and bacterial flagella: insights into their function from structural similarities.* Proc Natl Acad Sci U S A, 2003. **100**(6): p. 3027-30.

[12] Blaylock, B., et al., *Characterization of the Yersinia enterocolitica type III secretion ATPase YscN and its regulator, YscL.* J Bacteriol, 2006. **188**(10): p. 3525-34.

[13] Akeda, Y. and J.E. Galan, *Chaperone release and unfolding of substrates in type III secretion.* Nature, 2005. **437**(7060): p. 911-5.

[14] Akeda, Y. and J.E. Galan, *Genetic analysis of the Salmonella enterica type III secretion-associated ATPase InvC defines discrete functional domains.* J Bacteriol, 2004. **186**(8): p. 2402-12.

[15] Claret, L., et al., *Oligomerization and activation of the FliI ATPase central to bacterial flagellum assembly.* Mol Microbiol, 2003. **48**(5): p. 1349-55.

# Biotic-abiotic interfaces within a nanostructured polymer matrix platform: towards a completely abiotic system

Jennifer Sekowski[a], Omar B. Ayyub[b], and Peter Kofinas[b]
[a]US Army RDECOM, Edgewood Chemical Biological Center, Gunpowder, MD, USA 21010
[b]University of Maryland, Fischell Dept of Bioengineering, College Park, MD, USA 20742

## ABSTRACT

In developing new responsive materials for the warfighter, inspiration for function often comes from the biological world. Yet, in order to preserve biological function in the battlefield, the structure must be made more robust than is possible using the original biological materials. Hybrid biotic/abiotic structures are an important step toward the design of totally artificial constructs with the functionalities of biotic systems. Such functional abiotic polymers could serve in many capacities, as autonomous sensor platforms or as integrated function materials in the warfighter's kit. Hybrid biotic/abiotic structures were fabricated using chemically functionalized polystyrene-b-poly(2-vinyl-pyridine) (PS-b-P2VP) block copolymers. These polymer films induce a visible color change when exposed to aqueous media. The P2VP block of the copolymer was functionalized with either 2-bromomethylphenylboronic acid or bromoethylamine. The 2-bromomethylphenylboronic acid functionalization allowed the films to respond to glucose with a change in color. When exposed to glucose the films changed from green to orange. Ovalbumin antibodies were attached to films functionalized with bromoethylamine. These films also responded to the Ovalbumin protein with a color change.

## 1. INTRODUCTION

In developing new responsive materials for the warfighter, inspiration for function often comes from the biological world. Yet, in order to preserve biological function in the battlefield, the structure must be made more robust than is possible using the original biological materials. Hybrid biotic/abiotic structures are an important step toward the design of totally artificial constructs with the functionalities of biotic systems. Such functional abiotic polymers could serve in many capacities, as autonomous sensor platforms, or as integrated function materials (e.g., on a sticker or in coatings and fabric) in the warfighter's kit. The overall goal of this research is to explore the basic rules governing the biotic-abiotic interfaces and other chemistry required to integrate different types of binding or capture reagents within a responsive, nanostructured block copolymer (BCP) matrix. The ease of processing the BCP at ambient temperature already allows the production of responsive shape-conforming thin films that can be processed, as coatings or sheets, and distributed over a large area substrate. The principles established from this work will ultimately lead to development of responsive flexible polymers that could form the basis for multiple or multiplexed "litmus test" polymers, configured as small "stickers", large coating sheets, or even be integrated into the fabric or coatings used in the warfighter's uniform and kit.

To create a polymeric material that can successfully recognize ricin, anthrax, or any other biological or chemical molecule, it is necessary to combine a selective recognition element with a measurable output signal. The material we used achieves these needs by using the tunable reflectance of swollen functionalized block copolymers (BCPs). Block copolymers are polymers consisting of two or more chemically distinct sequences of monomer repeat units linked together through a covalent bond. Upon evaporation from a solvent, BCPs will microphase separate into solid films displaying a number of different morphologies (e.g., hexagonal, cubic, gyroid, or lamellar) depending on the relative volume fraction of each block. BCPs in which both blocks are of equal molecular weight generally exhibit the lamellar morphology. In this work, we used an A-B diblock copolymer that microphase separates into a lamellar periodic stack. The A block is water-insoluble polystyrene (PS) and does not contain any functional groups on the polymer chain that can interact with the biological/chemical threat compound. The B block contains the 2-vinyl-pyridine (P2VP) functionality which can be further chemically modified to provide ionic interaction sites directly with a CB threat compound or with CB agent capture reagents.

Without modification, the refractive index contrast between subsequent A and B layers is too low to successfully reflect light in the visible wavelength. In addition, the spacing between these layers is not large enough to interact with visible light. Through selective chemical modification of the B block, the condition for visible light can be achieved, thus

creating lamellar structures with tunable reflective properties. Upon binding a CB agent, directly or indirectly (using a capture reagent), the B layers swell or contract, changing the spacing between the subsequent A layers. This changes the refractive index contrast, inducing a visible color change. The ranges of wavelengths reflected are highly dependent upon the spacing between the AB layers and the refractive indices of subsequent layers. The color of the self-assembled polymer films can be modulated by simply changing the spacing between subsequent layers.

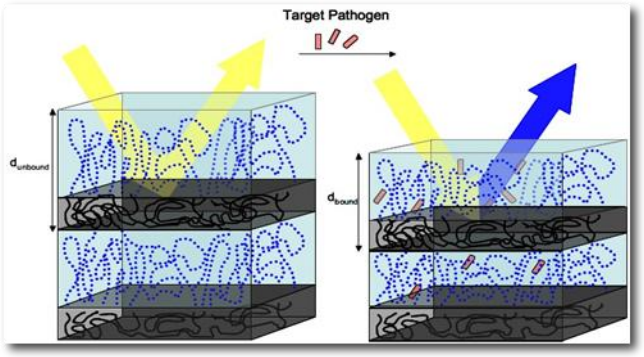



Figure 1: Schematic of BCP film in the unbound (left) and bound (right) state. The spacing (d), and refractive index contrast between subsequent layers changes as soon as the CB agent selectively binds to capture reagent in one of the layers.

Films of the block copolymer polystyrene-b-poly(2-vinyl-pyridine) (PS-b-P2VP) were fabricated to have a periodic lamellar structure, as seen in Figure 1. The films were then chemically functionalized. This functionalization only occurs on the poly(2-vinyl-pyridine) block of the copolymer placing a positive charge on the nitrogen atom of the pyridine ring, allowing the film to swell in water and change color. The block copolymer films were functionalized using either 2-bromomethylphenylboronic acid or bromoethylamine, as seen in Figure 2. When functionalized with the 2-bromomethylphenylboronic acid, the films respond to the glucose with a change in color. The bromoethylamine functionalization introduces a primary amine in the poly(2-vinyl-pyridine) block. The primary amine is then able to bind proteins, including capture reagents (e.g., antibodies) to the polymer film. When exposed to the binding target (e.g ricin) of the capture reagent (e.g anti-ricin antibody), the polymer film swells and changes color. To date, polyclonal antibody directed against ovalbumin has been successfully attached and bound (to ovalbumin protein) within the film. The ovalbumin-exposed films responded with a change in color.

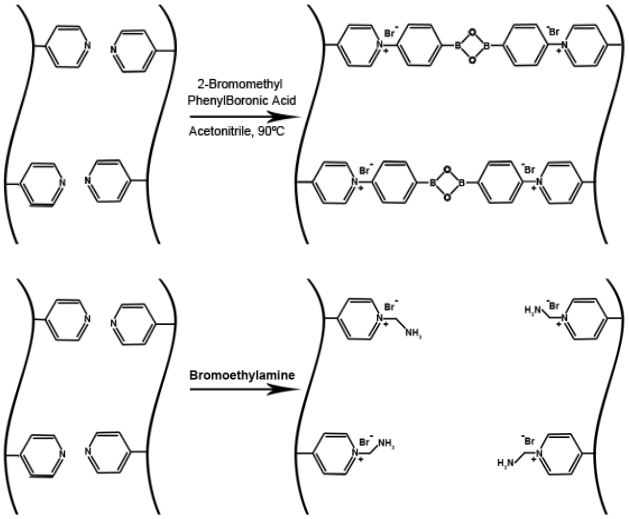



Figure 2: Functionalization of the 2-vinyl-pyridine block with either 2-bromomethylphenylboronic acid (top) or bromoethylamine (bottom).

## 2. METHODOLOGY

### 2.1 Fabrication of photonic BCP films

The procedure for fabrication of the PS-b-P2VP films was adapted from Kang et al, 2007. The PS-b-P2VP block copolymers were purchased from Polymer Source. The molecular weight of each block of the polymer was 133,000. A 5% weight/volume stock solution of PS-b-P2VP was prepared in propylene glycol monomethyl ether acetate (PGMEA). The solution was then passed through a 0.2µm teflon filter and degassed under vacuum for 10 minutes. The films were prepared by spin-casting 300µl of the PS-b-P2VP solution was onto 1''x1'' glass slides at 350rpm for two minutes. The glass slides were purchased from Ted Pella, and were functionalized with 3-Aminopropyl Triethoxysilane. The spin-cast block copolymer films were subsequently annealed in chloroform vapor at $50^{o}$C for 24 hours to allow them to self-assemble into a lamellar structure.

### 2.2 Functionalization

The films were functionalized with either 2-bromomethylphenylboronic acid or bromoethylamine depending on the chemical or biological target. The 2-bromomethylphenylboronic acid functionalization allows the films to target sugars such as glucose, while the bromoethylamine functionalization allows for further modification of the films. To functionalize the P2VP block of the copolymer, a solution of 40mg of either 2-bromomethylphenylboronic acid or bromoethylamine in 40ml of acetonitrile was prepared. The block copolymer films were then immersed in the solution for 24 hours at $90^{o}$C.

### 2.3 Glucose detection

Once the films were functionalized with 2-bromomethylphenylboronic acid, they were exposed to various concentrations of glucose to observe if the films would respond with a color change. Aqueous solutions of 50, 40, 30, 20, 10, 5 and 1mg/ml of glucose were prepared. To gather the baseline UV-Vis spectrum, the films were first soaked in water. The spectrum was again measured following exposure to each glucose solution.

### 2.4 Antibody attachment and protein detection

Films quaternized with bromoethylamine were modified by attaching polyclonal ovalbumin antibody. The attachment was created via a 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and N-Hydroxysuccinimide (NHS) reaction. 1 ml of ovalbumin antibody (1 mg/ml) was dialyzed into an (N-morpholino) ethanesulfonic acid buffer (pH 7). 1.1mg of NHS and 0.5µl EDC were allowed to incubate with the antibody solution 15 minutes at room temperature. A film functionalized with bromoethylamine was then immersed in the antibody solution for 3 hours. Once the antibody attachment reaction was complete, the film was introduced to a solution of 30mg/ml solution of ovalbumin protein.

## 3. RESULTS

Prior to testing, each type of film was analyzed by FTIR to verify that the particular type of quaternization reaction had been successful. Once verified, films with the bromoethylamine functionalization were further modified with ovalbumin antibodies, exposed to ovalbumin solution, and observed/quantified for color change by UV-Vis spectroscopy. Films functionalized with bromomethylphenylboronic acid were placed in various glucose concentrations and their responses were measured using the same method.

### 3.1 Verification of bromoethylamine attachment

To verify that the PS-b-2VP films were functionalized with bromoethylamine, FTIR was utilized to determine presence of amines. As seen in Figure 3 there is a peak at 3750$cm^{-1}$ which is associated with primary amines.

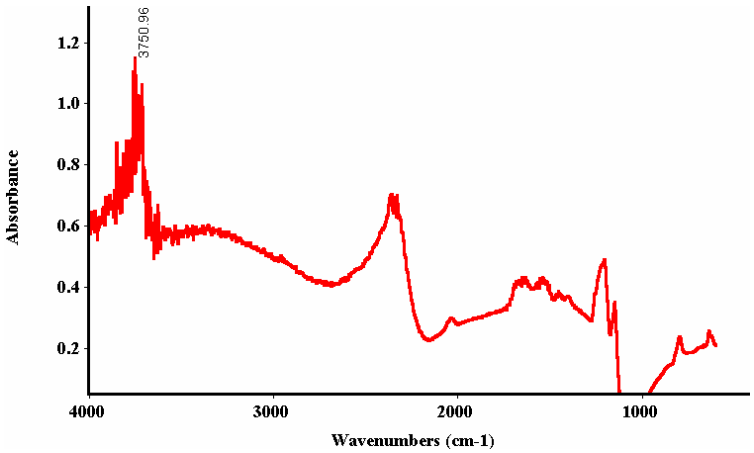



Figure 3: FTIR spectrum of film quaternized with 2-bromoethylamine.

### 3.2 Detection of ovalbumin

PS-b-2VP films quaternized with bromoethylamine were modified with polyclonal ovalbumin antibody using a carbodiimide reaction. The proposed mechanism for ovalbumin binding can be seen in Figure 4. Figure 5 shows that when the film swelled, changing its color from light orange to red, upon exposure to ovalbumin protein. The color change from orange to red was difficult to discern. Future experiments will involve optimizing a more obvious color change as well as measuring the shift in color using UV-Vis spectroscopy.

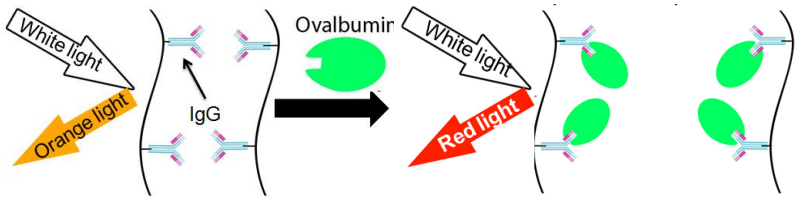



Figure 4: Upon exposure of the film to ovalbumin protein, the protein binds to the antibody causing a shift in mass and volume in the polymer film. This causes the film to swell and change color.

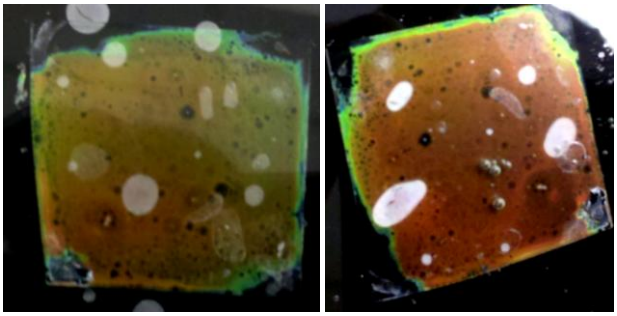


Figure 5: The PS-b-2VP film modified with antibody (left) is light orange before exposure to antibody target protein. After exposure to the target protein, the film is deep red orange (right).

### 3.3 Verification of boronic acid attachment

To determine if the 2-bromomethylphenylboronic acid was attached to the glucose sensing polymer, FTIR was again utilized. As seen in Figure 6, there is a peak at 3530cm$^{-1}$ which corresponds to the hydroxyl groups on the boron atom.

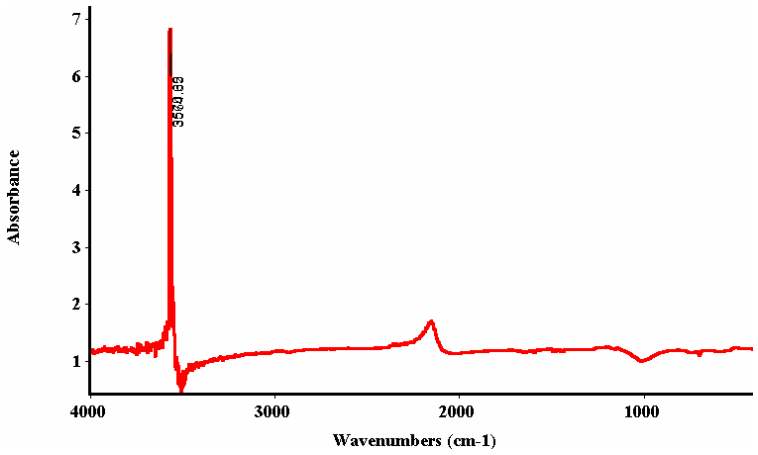



Figure 6: FTIR spectrum of film quaternized with 2-bromomethylphenylboronic acid

### 3.4 Detection of Glucose

The PS-b-2VP films functionalized with 2-bromomethylphenylboronic acid were exposed to a 50mg/ml aqueous solution of D-glucose. The films were initially green in pure deionized water due to the functionalization. Once exposed to the glucose solution, the films instantly swelled and became orange in color. This color change and proposed mechanism can be seen in Figure 7.

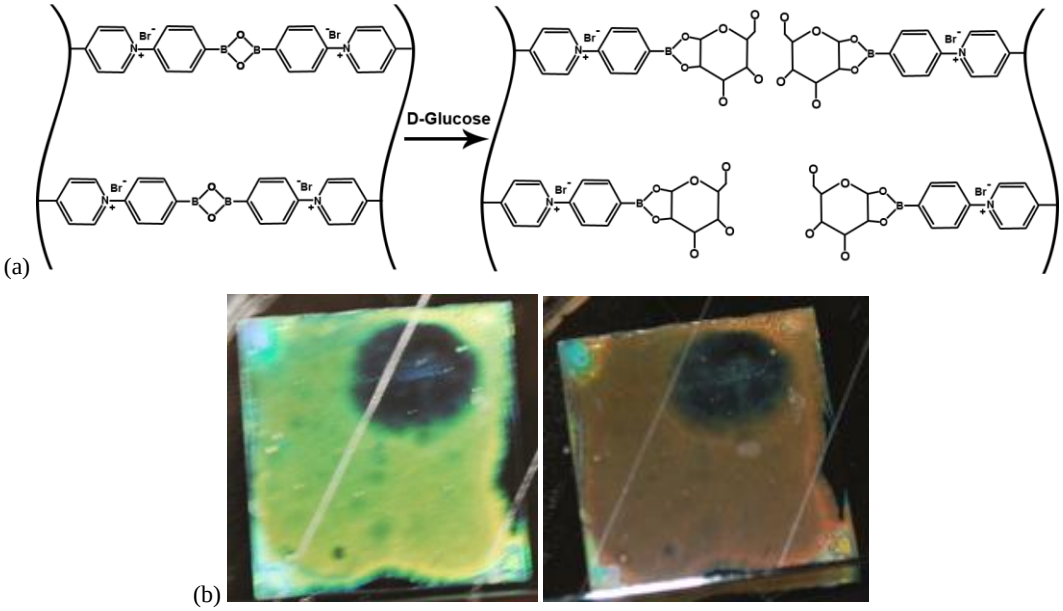



Figure 7: (a) Once the PS-b-2VP film is functionalized with 2-bromomethylphenylboronic acid the boronic acid groups will reversibly bind to each other cross-linking the film and inhibiting swelling. When exposed to glucose the boronic acid cross-links are broken as the boronic acid groups bind to glucose allowing the film to swell. (b)PS-b-2VP film functionalized with 2-bromomethylphenylboronic acid. Once exposed to a 50mg/ml D-glucose solution the film changed from green (left) to orange (right).

Films were exposed to various aqueous concentrations of glucose. To show a change in color, the visible spectrum of the films before and after exposure to glucose was measured using UV-Vis spectroscopy, as seen in Figure 8. The spectrum's peak should shift toward 700nm or the "red" end of the spectrum after exposure.

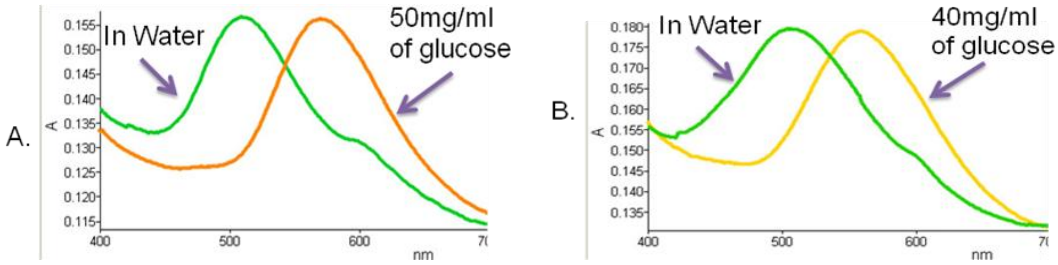



Figure 8: (a) UV-Visible spectrum of films in water and in a 50mg/ml glucose solution. (b) UV-Visible spectrum of film in water and a 40mg/ml glucose solution.

When exposed to concentrations of 50mg/ml and 40mg/ml the films shifted from green to orange or yellow respectively. This indicates that they became swollen. Exposure to lower concentrations had a different effect. As seen in Figure 9, for concentration of 30, 20, 10 and 5mg/ml, the films collapsed when exposed to the glucose solution. Their color shifted towards 400nm or the "violet" end of the spectrum. This effect of swelling at high concentrations and collapsing at lower concentrations has been noted in similar boronic acid based detection systems (Lee and Asher, 2000).

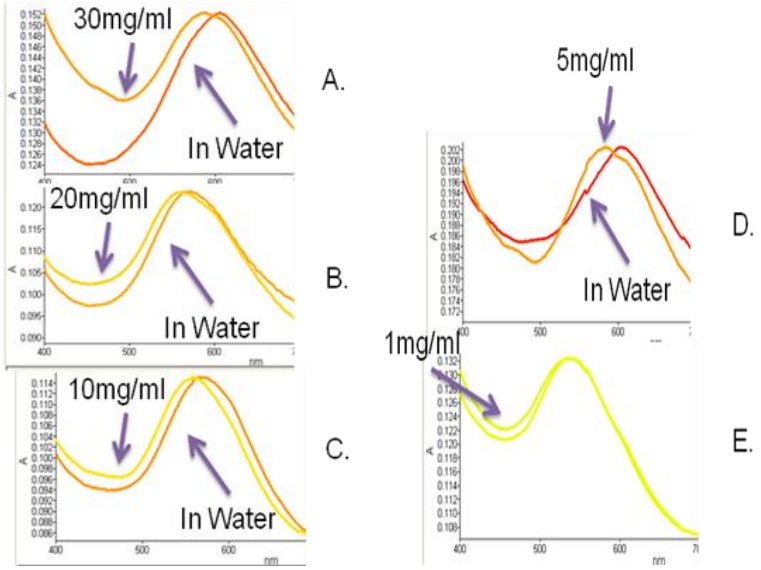



Figure 9: UV-Vis spectums of films in water and various glucose solutions. The film changed from (A) red to orange upon exposure to 30mg/ml, (B) red-orange to yellow upon exposure to 20mg/ml, (C) orange to yellow for 10mg/ml, (D) red to orange upon exposure to 5mg/ml, (E) and 1mg/ml had little effect upon the film.

The binding and color change to the sugar glucose demonstrates the concept of the PS-b-2VP films response to chemical or biological targets. This block copolymer functionalized with 2-bromomethylphenylboronic could eventually be modified to allow the PS-b-P2VP films to detect auto-inducer 2 (AI-2), which has been described as a "universal signal" for interspecies bacterial communication. AI-2 consists of molecules with chemical structures isomeric to simple sugars.

## 4. CONCLUSIONS

These two PS-b-P2VP films demonstrate early promise as thin, flexible, power-free, biologically-inspired materials with which to directly or indirectly detect chemical and biological threat compounds. The films functionalized with the 2-bromomethylphenylboronic acid demonstrated the ability to respond to glucose with a change in color without the use of any supplementary enzymes such as glucose oxidase. The polymer modified with polyclonal ovalbumin antibodies, demonstrated proof-of-concept that integration of antibodies and other types of capture reagents (e.g., single chain

antibodies, peptides, molecular imprinted polymers) into this structure may be a useful, reagent-free, power-free means of providing autonomous sensing of CB agents and other environmental cues. Ongoing experiments involve attaching other types of capture reagents (e.g., monoclonal antibody, single chain antibody, peptide, etc.) to the polymer as well as optimizing the resultant shift in color when exposed to the target protein.

## ACKNOWLEGEMENTS

This work was made possible through support from the ECBC ILIR Program.

## REFERENCES

[1] Kang, Y., Walish, J. J., Gorishnyy, T., & Thomas, E. L., 2007: Broad-wavelength-range chemically tunable block copolymer photonic gels, *Nature Materials*, 957-960.

[2] Lee, K., & Asher, S. A., 2000: Photonic Crystal Chemical Sensors: pH and Ionic Strength, *Journal of American Chemical Society*, 9534-9537.

# Azomethine ylide click chemistry applications to nanotechnology and biochemistry

Harold D. Banks
US Army Edgewood Chemical Biological Center, APG, MD, USA 21010

## ABSTRACT

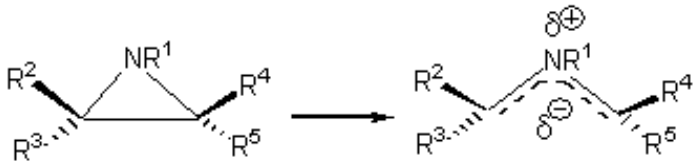
In order to construct novel materials for nanotechnology and biotechnology, molecules have been modified by means of conjugation to various functionalities. "Click chemistry," exemplified by the high yield reaction of an appended acetylene derivative with an organic azide, has been successfully applied. When the mild conditions critical for *in vivo* work are used, either potentially toxic Cu(I) salts, or commercially unavailable strained alkynes, cyclooctynes, are required. This research effort encompassed comprehensive, computational studies of a potentially superior click chemistry, 1,3-dipolar cycloaddition of azomethine ylides originating from aziridines with unstrained alkynes. While previously studied aziridines react too slowly under physiological conditions, it was predicted, based on previous results with the nucleophilic chemistry of aziridines that also proceeds through dipolar transition states, that appropriate substituents would increase reactivity to useful levels. A high level *ab initio* computational investigation [MP2(Full)/6-311++G(d,p)//MP2(Full)/6-31+G(d)] of these electrocyclic reactions was employed. This approach has the advantage of being able to arrive at useful results in a more expedient manner than an empirical laboratory investigation. Since cleavage of unsymmetrical aziridines has two possible conrotatory modes, torquoselectivities were determined. To obtain reasonable reactivity at 298 K, it is necessary for substituents to accelerate the rate by almost $10^{20}$.

**Keywords:** Click chemistry, 1,3-dipolar cycloaddition, aziridines, substituent effects, nanotechnology, biotechnology

## 1. INTRODUCTION

One of the current challenges of chemistry is the discovery of efficient reactions that occur under mild conditions to ensure fragile functionalities which might be found in nanotechnological or biological systems are not disrupted or destroyed. "Click chemistry" introduced by the Sharpless group[1] in 2001 is an attractive approach. The requirements for this new synthetic methodology are modular reactions that are wide in scope, produce very high yields, and produce inoffensive byproducts from readily available starting materials using no or benign solvents and affording simple product isolation. Of the several general reaction types that meet these requirements, 1,3-dipolar cycloadditions have received the most attention, in particular, the reaction of azides with alkynes or alkenes. Typical reactants require a catalyst, usually Cu(I), toxic in biological systems, to proceed under the mild conditions demanded by click chemistry. To avoid Cu(I), Bertozzi[2] cleverly used relief of ring strain in the transition state of the reaction of azides with highly strained cyclooctynes to obtain acceptable reaction rates. Azomethine ylides have been known for many years to constitute another family of 1,3-dipoles that react readily with dipolarophiles. One method of generating these reactive species is the thermally allowed conrotatory ring opening of aziridines, as illustrated by the reaction of *trans*-2-benzoyl-3-phenylaziridine and dimethoxycarbonylacetylene.[3] Prolonged heating at 140$^{o}$C was required for success; these conditions are incompatible with reactions for molecules containing sensitive functionalities or biological studies.

Having established that formation of the azomethine ylide from the precursor aziridine was the rate determining step, we were intrigued by the possibility that substituents might lower the free energy of activation to the point that azomethine ylide formation might occur at an acceptable rate at room temperature. Previous studies on nucleophilic attack on small ring heterocycles by ammonia established that halogen substituents can produce remarkable accelerations. The transition states for these reactions have considerable dipolar character; those resulting from ring cleavage of aziridines, also have charge separation with partial positive charge developing on nitrogen. The rate of ring cleavage of aziridines should also be sensitive to the nature of the substituent. Since computational studies permit a rapid screening of substituted aziridine possibilities in a safe, cost-effective manner as opposed to empirical laboratory studies, an investigation of the effect of substituents on the thermal $C_2 - C_3$ cleavage was undertaken. One possible contribution to reaction rate that had to be studied was the possibility of release of strain energy in the transition state. If this effect is obtained, the more strain energy released in the transition state, the lower the free energy of activation and the faster the reaction.




Since this cleavage reaction is an example of four-electron electrocyclic chemistry, it is well established that the course of the reaction will be conrotation (i.e., substituents at different carbons and found on opposite sides of the aziridine plane will rotate in the same direction). If the aziridine is unsymmetrical, clockwise or counterclockwise rotation entails other stereochemical consequences; the $R^2$ group can selectively move with an outward or inward modality. Since this *torquoselectivity*[6] is intimately related to the ultimate rate, its determination was the starting point of this research effort.

## 2. METHODOLOGY

*Ab initio* calculations were performed at the MP2(Full)/6-311++G(d,p)// MP2(Full)/6-31+G(d) level by means of the Gaussian 03 suite of programs[7] for determination of energies, the NBO charges[8, 9] (keyword, pop = SaveNBOs) and the Wiberg bond indices[10] (keyword, pop = NBOREAD; final line of input file = $nbo BNDINX $end). Gas phase calculations were deemed to be sufficient for calculation of relative energies since calculations using a SCI-PCM water solvent model for aziridine and 2-fluoroaziridine produced only modest changes in relative rates (< 25 kJ/mol). The minor effect of solvent on the rate azomethine ylide formation has also been observed experimentally.[11] A scaling factor of 0.9646 was used[12] for the thermal correction to the computed energies at a reaction temperature of 298.15 K. Criteria for finding the transition states and ground states were calculation of one and zero imaginary frequencies, respectively. Calculations using invertomers (*anti* or *syn*) at nitrogen resulted in insignificant changes in the free energies of activation. GaussView 3.09[13] was used for animation of the sole imaginary frequency and IRC calculations were used to confirm identification of the transition states. It was determined using 2-substituents such as $CH_3CO$ and $OCOCH_3$ with aziridine for which conformational isomerism was possible due to *syn* and *anti* invertomers had energy differences of <4kJ/mol. These differences were judged too small to justify the expense of an exhaustive study of the conformations of the ground and corresponding transition states for all pertinent compounds. Relative reaction rates were obtained from transition state theory.[14] Electrostatic energies, $E_{elst}$, in kJ/mol were calculated using a classical relationship[15]:

$$E_{elst} = \frac{(1.39 \times 10^3)(q_{C2})(q_{C3})}{(D_{eff})(d)}$$

where $\boldsymbol{q}_{C2}$ and $\boldsymbol{q}_{C3}$ are the calculated charges at $C_2$ and $C_3$ from the atomic NBO charges with hydrogens summed into the heavy atoms, $D_{eff}$ is the effective dielectric constant, $\boldsymbol{d}$ is the distance between these atoms in Angstroms, and $\mathbf{E_{elst}}$ is expressed in kJ/mol. For the calculations, $D_{eff} = 1.0$, that of a vacuum, was used.[16]

## 3. RESULTS AND DISCUSSION

The calculated[17] free energies of activation of a family of haloaziridines are presented in Table 1. It is clear that there is no correlation between strain energy and reaction rate. The most strained compound, 2,2-difluoroaziridine, for example, reacts at the slowest rate.

Table 1. Strain Energy Effects on the Relative Rates of Thermolysis of Haloaziridines[a].

| $R^2$ | $R^3$ | $k_{rel}$ | SE[b] (kJ/mol) | $[k_{rel}]$ from SE difference |
|---|---|---|---|---|
| H | H | 1.00 | 114 | 1.00 |
| F | H | $1.37 \times 10^{3}$ | 149 | $1.20 \times 10^{6}$ |
| F | F | $1.20 \times 10^{-5}$ | 179 | $2.66 \times 10^{11}$ |
| Cl | H | $4.47 \times 10^{4}$ | 138 | $1.77 \times 10^{4}$ |
| Cl | Cl | $7.13 \times 10^{0}$ | 140 | $3.47 \times 10^{4}$ |
| Br | H | $1.69 \times 10^{5}$ | 151 | $2.79 \times 10^{6}$ |
| Br | Br | $3.11 \times 10^{2}$ | 167 | $1.69 \times 10^{9}$ |

a) See Eq 1.
b) Calculated strain energy.

The free energies of activation, $\Delta G_{act}$, of almost fifty aziridines were determined. For unsymmetrical reactants the preferred reaction path is that corresponding to its torquoselectivity, and accordingly, the $\Delta G_{act}$ values were evaluated for outward and inward modes of conrotatory cleavage. These results are furnished in Table 2.

Table 2. Torquoselectivity in Thermal Cleavage of Substituted 3-Membered Heterocycles

| **Substituents** | $\Delta G^{\ddagger}_{out}$ | $\Delta G^{\ddagger}_{in}$ | $k_{out}/k_{in}$ |
|---|---|---|---|
| None | 194.6 | 194.6 | 1.00E+00 |
| 2-F | 176.7 | 213.7 | 3.00E+06 |
| 2-Cl | 168.1 | 207.8 | 9.13E+06 |
| 2-Br | 164.8 | 204.8 | 1.05E+07 |
| 2-$NH_2$ | 157.7 | 207.7 | 5.70E+08 |
| 2-$NHCOCH_3$ | 163.8 | 205.1 | 1.73E+07 |
| 2-OH | 167.7 | 216.3 | 3.27E+08 |
| 2-$OCH_3$ | 160.8 | 194.7 | 8.61E+05 |
| 2-$OCOCH_3$ | 137.5 | 185.7 | 2.81E+08 |
| 2-Vinyl | 163.4 | 195.8 | 4.61E+05 |
| 2-Ethynyl | 162.5 | 183.8 | 5.45E+03 |
| 2-Ph | 152.7 | 182.9 | 1.95E+05 |
| 2-(*o*-$FC_6H_4$) | 151.5 | 203.1 | 1.12E+09 |
| 2-CN | 167.2 | 179.3 | 1.33E+02 |
| 2-COOMe | 186.4 | 182.7 | 2.27E-01 |
| 2-CHO | 167.4 | 169.2 | 2.10E+00 |
| 2-(*E*)-CH=CHCHO | 156.0 | 186.6 | 2.30E+05 |
| 2-$CH_3CO$ | 161.7 | 159.5 | 4.09E-01 |
| 2-$CF_3$ | 187.0 | 193.9 | 1.59E+01 |
| 2-Me | 184.7 | 206.5 | 6.57E+03 |
| 2-t-Bu | 180.0 | 199.0 | 2.12E+03 |
| *trans*-2-F, 3-Ph | 141.0 | 203.0 | 6.98E+10 |
| *trans*-2-Cl, 3-Ph | 128.8 | 191.3 | 8.69E+10 |
| *cis*-2-Cl, 3-Ph | 146.9 | 161.0 | 2.95E+02 |
| *trans*-2,3-diF | 165.4 | 248.9 | 4.14E+14 |
| *cis*-2,3-diF | 194.3 | 194.3 | 1.00E+00 |
| *trans*-2,3-diCl | 147.1 | 222.2 | 1.42E+13 |
| *trans*-2,3-diMe | 181.2 | 221.8 | 1.30E+07 |
| *cis*-2,3-diMe | 191.0 | 191.0 | 1.00E+00 |
| *trans*-2,3-diPh | 113.4 | 165.1 | 1.16E+09 |
| *cis*-2,3-diPh | 142.3 | 142.3 | 1.00E+00 |
| *trans*-2,3-diCHO | 149.9 | 123.7 | 2.59E-05 |
| 2-COOH | 165.3 | 170.7 | 8.96E+00 |
| 2-$COO^-$ | 186.5 | 229.1 | 2.88E+07 |
| 2-CHΞ$C^-$ | 122.3 | 178.4 | 6.58E+09 |
| 2-$NH^-$ | 43.5 | 75.0 | 3.29E+05 |
| 2-$SO_3^-$ | 163.9 | 196.4 | 5.02E+05 |
| 2-$SO_3H$ | 181.0 | 182.9 | 2.21E+00 |
| 2-$O^-$ | 47.2 | 90.5 | 3.83E+07 |
| 2-$S^-$ | 115.6 | 181.2 | 2.98E+11 |
| 2-Li | 176.2 | 133.3 | 3.09E-08 |
| 2-Na | 196.6 | 183.8 | 5.82E-03 |
| 2-$BH_2$ | 163.6 | 97.3 | 2.44E-12 |
| 2-$Be^+$ | 111.7 | 17.8 | 3.62E-17 |
| 2-$NH_3^+$ | 186.9 | 198.3 | 9.67E+01 |
| 2-$SiH_3$ | 182.3 | 182.3 | 1.00E+00 |
| 2-$Si(CH_3)_3$ | 184.1 | 184.3 | 1.09E+00 |

Unlike the isoelectronic substituted cyclobutenes, outward rotation is the preferred mode of rotation in virtually all cases.[18] Inspection of Table 2 allows one to conclude that substituents do indeed play a significant role in the

torquoselectivity of azomethine ylides by means of aziridine $C_2 - C_3$ bond cleavage for aziridines that lack a σ plane passing through the heavy atoms of the ring. Electronegative or anionic functionalities were found to have a preference for outward rotation, while electropositive or cationic groups follow predominant inward rotation. After consideration of explanations in terms of frontier molecular orbitals[18g], and electrostatic effects, the most reasonable rationalization of the results is that outward rotation gives rise to a transition state that is stabilized by electron-withdrawal. For inward rotation, the dipole of the substituent encounters that of the ring; repulsion of negative groups or attraction by positive groups determine to likelihood of this rotational mode.[19]

A determination of torquoselectivity is necessary before the overall reaction rate can be determined. In most cases, there is a vast preference for one rotational mode as indicated by the $k_{out}/k_{in}$ values. In order for a reaction to satisfy the criterion of mildness with respect to temperature (298 K or lower), using the usual criterion that ten half-lives constitutes a complete (>99 %) reaction, and noting that the rate-determining ring cleavage is first-order, the ceiling for $\Delta G_{act}$ may be calculated using transition state theory, assuming an overnight (12 h) reaction time is acceptable. Those substituents that lower the $\Delta G_{act}$ of **1** from 194.6 kJ/mol to less than 94.8 kJ/mol will have acceptable reactivity. Of the compounds found in Table 1, only $R^2 = NH^-$, $O^-$, $BH_2$, and $Be^+$ fall into this range. There are, however, other requirements for azomethine ylide precursors. These compounds should be stable in aqueous media (the latter two are not) and should be close to neutral pH (the former two are quite basic.)

There are many additional conclusions that can be gleaned from studying the relative $\Delta G_{act}$ values of the monosubstituted aziridines in Table 2. For the halo substituents, Br < Cl < F. This trend suggests that polarizability has a greater transition state stabilizing effect than through bond electron-withdrawing ability. For atoms of the second period of the periodic table, $NH_2$ < OH < F. This suggests that electron donating ability, or basicity, is important in increasing the reaction rate. Acetylation of $NH_2$ results in a minor increase in $\Delta G_{act}$, but the effect with OH is a rather substantial decrease; this difference may be due to ground state stabilization of the ground state of the alcohol due to H-bonding. Conjugated systems stabilize the transition state (Ph < CΞCH < $HC=CH_2$). Furthermore, favorable changes occur when conjugation is extended [(*E*)-CH=CHCHO < $HC=CH_2$]. Substitution of methyl hydrogens produced only minor changes ($CH_3 \approx C(CH_3)_3 \approx CF_3$). Removal of a proton from a COOH group, as opposed to an $SO_3H$ group, produces dramatically different results; the former results in destabilization of the transition state, while the latter is calculated to proceed at a rate palpably faster than that of aziridine. The more reactive diastereomer in the 2,3-disubstituted examples studied here has the *trans* configuration. The most noteworthy results were found for negatively ($NH^-$, $O^-$, $S^-$), and one positively ($Be^+$), charged substituents attached to $C_2$. In these cases, all functionalities are capable of strong electron withdrawal, which seems to be an important factor in transition state stabilization.

In an attempt to further elucidate the effect of substituents on the rate of formation of azomethine ylides, an extensive study of haloaziridines was undertaken with particular emphasis on fluorine. Fluorine is an extremely useful probe of reaction mechanisms.[20] It is similar in size to hydrogen, but it has vastly different electronic effects because it is the most electronegative atom. The results of this investigation are presented in Table 3.

Table 3. Free Energies of Activation (kJ/mol) of Aziridines as a Function of Halogen Substitution

| Aziridines | $R^2$ | $R^3$ | $R^4$ | $R^5$ | $\Delta G_{act}$[a] (kJ/mol) $R^1 = H$ | $\Delta G_{act}$ (kJ/mol) $R^1 = CH_3$ | $\Delta\Delta G_{act}$ (kJ/mol) |
|---|---|---|---|---|---|---|---|
| **1** | H | H | H | H | 194.6 [194.6] | 182.9 | -11.7 |
| **2** | F | H | H | H | 176.7 [176.7] | 163.2 | -13.5 |
| **3** | F | F | H | H | 222.7 [195.8] | 204.0 | -18.7 |
| **4** | F | H | H | F | 165.4 [158.8] | 147.0 | -18.4 |
| **5** | F | H | F | H | 194.3 [195.8] | 175.4 | -18.9 |
| **6** | F | F | F | H | 224.7 [177.9] | 205.8 | -18.9 |
| **7** | F | F | F | F | 230.6 [197.0] | 225.8 | -4.8 |
| **8** | Cl | H | H | H | 168.1 [168.1] | 154.2 | -13.9 |
| **9** | Cl | Cl | H | H | 189.7 [181.3] | 171.1 | -18.6 |
| **10** | Cl | H | H | Cl | 147.1 [141.6] | 128.9 | -18.2 |
| **11** | Cl | H | Cl | H | 172.3 [181.3] | 157.4 | -14.9 |
| **12** | Br | H | H | H | 164.8 [164.8] | 144.8 | -20.0 |
| **13** | Br | Br | H | H | 180.4 [175.0] | 161.0 | -19.4 |

(a) Bracket values correspond to the calculation based on torquoselectivities (see text).

For example, the differences in free energy of activation for outward and inward rotation of **2** are -17.9 and +19.1 kJ/mol relative to **1**. The estimated $\Delta G_{act}$ of **3**, for example, is 194.6 – 17.9 + 19.1 = 195.8 kJ/mol. Inspection of Table 3 reveals that the estimated values are reasonably close to those computed in this manner with the conspicuous exceptions of **3**, **6** and **7**, fluoroaziridines that possess geminal fluoro groups. Since these aziridines, chosen to minimize the computation cost of this investigation, are unsubstituted at nitrogen, there looms the possibility that interaction between fluorine and the N – H bond in the form of strong electrostatic forces might play a role in these discrepancies. These forces are essentially absent for the remaining halogens studied; thus, the rate retardation observed for the geminally disubstituted difluoro compounds is significantly diminished in **9** and **13**. Accordingly, calculations were performed on the N – $CH_3$ compounds. It is significant that the $\Delta G_{act}$ results for the chloro compounds, **8** – **11**, and the bromo compounds, **12** and **13**, are comparable to those obtained for the fluoroaziridines. All N-methyl derivatives are calculated to be more reactive than their corresponding parent compounds. One explanation of these findings is that the transition state leading to the azomethine ylide has a formal positive charge on nitrogen. Electron release by methyl would lower the energy of this species. Exceptional behavior is demonstrated, however, by tetrafluoride **7**. It is clear that when the carbon positions of aziridine are completely fluorinated, stabilization by an N-methyl group is all but eliminated. Inspection of the geometry of the transition state of **7** reveals that unlike the other compounds studied, the planes described by the substituents at $R^2R^3C_2$ and $R^4R^5C_3$ are almost perpendicular, leading to minimal overlap of orbitals (see Eq 1).

Table 4. NBO Atomic Charges of Haloaziridines

| | $R^2$ | $R^3$ | $C_2$ | $R^4$ | $R^5$ | $C_3$ | $R^1$ | N |
|---|---|---|---|---|---|---|---|---|
| **1** | 0.195 | 0.218 | -0.228 | 0.198 | 0.207 | -0.290 | 0.427 | -0.725 |
| **2** | -0.429 | 0.214 | 0.431 | 0.200 | 0.213 | -0.370 | 0.425 | -0.691 |
| **3** | -0.415 | -0.402 | 0.966 | 0.248 | 0.235 | -0.309 | 0.411 | -0.734 |
| **4** | -0.422 | 0.215 | 0.343 | 0.215 | -0.422 | 0.343 | 0.440 | -0.711 |
| **5** | -0.418 | 0.208 | 0.355 | -0.427 | 0.201 | 0.355 | 0.453 | -0.725 |
| **6** | -0.388 | -0.384 | 0.956 | -0.417 | 0.208 | 0.355 | 0.453 | -0.782 |
| **7** | -0.404 | -0.388 | 0.931 | -0.388 | -0.404 | 0.930 | 0.419 | -0.696 |
| **8** | -0.043 | 0.252 | -0.069 | 0.217 | 0.203 | -0.295 | 0.431 | -0.696 |
| **9** | 0.010 | 0.024 | 0.007 | 0.226 | 0.208 | -0.214 | 0.425 | -0.686 |
| **10** | -0.027 | 0.255 | -0.112 | 0.255 | -0.112 | -0.027 | 0.433 | -0.666 |
| **11** | -0.004 | 0.259 | -0.046 | -0.045 | 0.232 | -0.144 | 0.430 | -0.681 |
| **12** | 0.004 | 0.257 | -0.143 | 0.219 | 0.205 | -0.280 | 0.433 | -0.694 |
| **13** | 0.076 | 0.092 | -0.157 | 0.228 | 0.211 | -0.188 | 0.429 | -0.691 |

Since fluorine exerts a powerful electron-withdrawing through-bond effect, it was not surprising that the NBO atomic charges[21] determined for $C_2$ and $C_3$ for geminally difluorinated **3**, **6**, and **7** (red values in Table 4) were quite positive (>0.9) both absolutely and relative to the remaining compounds of this series. It is reasonable that unfavorable charge development at the carbon atoms in the transition state is at least partially responsible for the significant rate retardation of the *gem*-difluoro compounds. A valid assumption is that the accumulation of positive charge at $C_2$ and $C_3$ of **7** is responsible for the considerable distortion of its transition state geometry and the elevation of its $\Delta G_{act}$ to the highest value of the series. Table 5 presents differences in NBO atomic charges relative to **1**. It is obvious that substitution of fluorine for hydrogen dramatically raises the positive charge of the attached carbon (compare, for example, the $C_2$ and $C_3$ NBO charges for **2**.) This negative effect does not express itself in an increase in $\Delta G_{act}$ until an additional fluorine replacement at this atomic center is made.

Table 5. Differences in NBO Atomic Charges of Haloazirdines with Respect to **1**

| | $R^2$ | $R^3$ | $C_2$ | $R^4$ | $R^5$ | $C_3$ | $R^1$ | N |
|---|---|---|---|---|---|---|---|---|
| **1** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **2** | -0.624 | -0.004 | 0.659 | 0.002 | 0.006 | -0.080 | -0.002 | 0.034 |
| **3** | -0.610 | -0.620 | 1.194 | 0.050 | 0.028 | -0.019 | -0.016 | -0.009 |
| **4** | -0.617 | -0.003 | 0.571 | 0.017 | -0.629 | 0.633 | 0.013 | 0.014 |
| **5** | -0.613 | -0.010 | 0.583 | -0.625 | -0.006 | 0.645 | 0.026 | 0.000 |
| **6** | -0.583 | -0.602 | 1.184 | -0.615 | 0.001 | 0.645 | 0.026 | -0.057 |
| **7** | -0.599 | -0.606 | 1.159 | -0.586 | -0.611 | 1.220 | -0.008 | 0.029 |
| **8** | -0.238 | 0.034 | 0.159 | 0.019 | -0.004 | -0.005 | 0.004 | 0.029 |
| **9** | -0.185 | -0.194 | 0.235 | 0.028 | 0.001 | 0.076 | -0.002 | 0.039 |
| **10** | -0.222 | 0.037 | 0.116 | 0.057 | -0.319 | 0.263 | 0.006 | 0.059 |
| **11** | -0.199 | 0.041 | 0.182 | -0.243 | 0.025 | 0.146 | 0.003 | 0.044 |
| **12** | -0.191 | 0.039 | 0.085 | 0.021 | -0.002 | 0.010 | 0.006 | 0.031 |
| **13** | -0.119 | -0.126 | 0.071 | 0.030 | 0.004 | 0.102 | 0.002 | 0.034 |

The presumed structure of the azomethine ylide suggests partial double bond character in the C – N bonds. To study this possibility and the gain structural data with respect to the transition states for these reactions, the bond distances were measured for this collection of haloaziridines (Table 6). Relative to their average bond distances, the results of these calculations confirmed that the C – N bond lengths were intermediate between single (amino) and double (imino) bonds. It can also be noted that the C – F distances of the geminally disubstituted fluoroaziridines were somewhat shorter than the average values. Only minor changes in bond distances are observed for the bonds to halogen.

Table 6. Transition State Distance in Conrotatory Cleavage of Aziridines.[a]

| Aziridine | Distances[b] (Å) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | $R^2$ — $C_2$ | $R^3$ — $C_2$ | $R^4$ — $C_3$ | $R^5$— $C_3$ | N — $C_2$ | N — $C_3$ | N — H | $C_1$— $C_2$ |
| **1** | 1.091 | 1.081[c] | 1.083[c] | 1.090[c] | 1.368 | 1.377 | 1.083 | 2.059 |
| **2** | 1.378 | 1.078[c] | 1.087[c] | 1.079[c] | 1.332 | 1.381 | 1.023 | 2.086 |
| **3** | 1.358 | 1.360[c] | 1.085[c] | 1.079[c] | 1.321 | 1.396 | 1.029 | 2.119 |
| **4** | 1.372 | 1.079[c] | 1.079[c] | 1.372[c] | 1.348 | 1.348 | 1.020 | 2.099 |
| **5** | 1.368 | 1.087[c] | 1.384[c] | 1.083[c] | 1.337 | 1.368 | 1.019 | 2.117 |
| **6** | 1.348 | 1.356[c] | 1.084[c] | 1.368[c] | 1.360 | 1.377 | 1.015 | 2.049 |
| **7** | 1.347 | 1.351[c] | 1.351[c] | 1.348[c] | 1.364 | 1.364 | 1.039 | 2.221 |
| **8** | 1.738 | 1.080 | 1.082 | 1.088 | 1.344 | 1.380 | 1.020 | 2.060 |
| **9** | 1.733 | 1.740[c] | 1.082 | 1.087 | 1.333 | 1.388 | 1.024 | 2.102 |
| **10** | 1.729 | 1.080 | 1.080 | 1.729[c] | 1.353 | 1.353 | 1.024 | 2.076 |
| **11** | 1.723 | 1.080 | 1.748[c] | 1.086 | 1.364 | 1.351 | 1.023 | 2.106 |
| **12** | 1.906 | 1.080 | 1.082 | 1.088 | 1.343 | 1.380 | 1.020 | 2.057 |
| **13** | 1.895 | 1.906[c] | 1.082 | 1.087 | 1.332 | 1.387 | 1.024 | 2.096 |

(a)Average bond distances[22] (Å): $C(sp^3)$ — F = 1.38; $C(sp^3)$ — Cl = 1.78; $C(sp^3)$ — Br = 1.94; Average distance $C(sp^3)$ – N = 1.47 Å; Average distance $C(sp^2)$ – N = 1.28 Å
(b) Outward rotation
(c) Inward rotation

An alternative means of scrutinizing the nature of bonding in these transition states is the use of Wiberg Bond Indices (WBI). These data permit the assignment of bond order in a straightforward manner. A WBI of 1.0 is equivalent to a single bond. The results presented in Table 7 agree nicely with the results from the determinations of bond distances. The WBI for the C – N bonds indicate that the C – halogen bonds are weakened in the transition state.

Table 7. Selected Wiberg Bond Indices for Haloaziridines

| **Aziridine** | **$C_2$ — N** | **$C_3$ — N** | **$C_2$ — $C_3$** | **$C_2$— X** | **$C_n$ — X[c]** |
|---|---|---|---|---|---|
| **1** | 1.165 | 1.139 | 0.756 | — | — |
| **2** | 1.266 | 1.131 | 0.661 | 0.804[a] | — |
| **3** | 1.253 | 1.168 | 0.627 | 0.818[a] | 0.822[b] |
| **4** | 1.177 | 1.177 | 0.656 | 0.814[a] | 0.814[a] |
| **5** | 1.128 | 1.204 | 0.662 | 0.822[a] | 0.803[b] |
| **6** | 0.823 | 1.106 | 0.711 | 0.823[a] | 0.857[a]; 0.849[b] |
| **7** | 1.209 | 0.838 | 0.496 | 0.838[a]; 0.844[b] | 0.884[a]; 0.838[b] |
| **8** | 1.202 | 1.129 | 0.685 | 1.030 | — |
| **9** | 1.029 | 1.146 | 0.624 | 1.029 | 1.020 |
| **10** | 1.178 | 1.178 | 0.613 | 1.043 | 1.043 |
| **11** | 1.159 | 1.068 | 0.624 | 1.019 [a] | 1.067 [a] |
| **12** | 1.206 | 1.130 | 0.692 | 1.017 | 0.933; 0.929 |
| **13** | 1.203 | 1.152 | 0.629 | 1.007 | 0.990 |

(a) Outward rotation
(b) Inward Rotation
(c) $C_n = C_2$ or $C_3$ depending on aziridine structure

A thorough search of the literature revealed only two reports of azomethine ylide formation from aziridines in which the kinetics of the reaction were studied.[23] Unfortunately, due to the need for highly derivatized aziridines that are relatively easy to handle and the experimental requirements of elevated temperatures and aromatic solvents, extrapolation of these results to our gas phase calculations is difficult. While temperature corrections to 298 K can be effected, perhaps the most troublesome choice for a calculation is that of a suitable solvent model.

Finally a recent in *ab inito* method, SCS-MP2[24], uses a different approach than the classical MP2 approach to evaluate the electron correlation energy, $E_c$:

$$E_{SCS\text{-}MP}{}^2 = E_{HF} + E_c$$

$$E_{SCS\text{-}MP}{}^2 = E_{HF} + 1.20(E_{\alpha\alpha} + E_{\beta\beta}) + 0.333(E_{\alpha\beta})$$

where $E_{\alpha\alpha}$ and $E_{\beta\beta}$ are the energy contributions of the parallel spins, and $E_{\alpha\beta}$ is the energy contribution of the antiparallel spins. This method is attractive since it is reported to produce more accurate results than MP2 calculation. Preliminary results for representative compounds in this study indicate that absolute differences are relatively small (<6 kJ/mol) calculated rates relative to aziridine are comparable.

## 4. CONCLUSIONS

The aim of this research project is to discover substituted aziridines that readily form azomethine ylides under extremely mild conditions (298 K and pH close to that found under physiological conditions). Calculations of several quantities for mechanistic elucidation were performed. Strain energy release, a potential contributor to lowering of the transition state energies for strained heterocycles was eliminated as a major contributor to the rate of thermal cleavage of substituted aziridines. The torquoselectivities of the aziridine derivatives were evaluated since the preferred rotational mode was an important component of calculation of the total rate. From the overall rates, conclusions could be drawn concerning the effects of structure on reactivity. The haloaziridines were employed to further refine the effects of substituents on the

cleavage reaction. Transition state distances, NBO charges and Wiberg bond indices were utilized to obtain mechanistic insights that are leading the way to the discovery of aziridines that fulfill the requirements of reaction under extremely mild conditions. Mechanistic elucidation by means of the fluoroaziridines has channeled calculations into directions that have succeeded in identifying substituted aziridines that satisfy these conditions. Aziridines with $R^1 = NH^-$, $O^-$, $BH_2$, and $Be^+$ each have acceptable $\Delta G_{act}$ values. Unfortunately, the first two derivatives are unacceptably basic, and the remaining compounds are too reactive in typical solvents. An exciting recent finding is that aziridines with carbon or nitrogen anionic centers have accelerated reaction rates. The thrust of current investigations is, therefore, to identify substituents that can lower the basicity of these compounds into the target range while at the same time retaining exceptional reactivity.

## REFERENCES


[1] Kolb, H.; Finn, M.; Sharpless, K. *Angew. Chem. Int. Ed*. **2001**, *40*, 2004–2031. (b) Sharpless, K. *Drug Discovery Today* **2003**, *8*, 1128. (c) Himo, F.; Lovell, T; Higraf, R.; Rostovtsev, V.; Noodleman, L.; Sharpless, K.; Folkin, V. *J. Am. Chem. Soc*. **2005**, *127*, 210–216. (d) Krasinski, A.; Radic, Z.; Manetsch, R.; Raushel, J.; Taylor, P.; Sharpless, K.; Kolb, H. *J. Am. Chem. Soc*. **2005**, *127*, 6686. (e) Agard, N.; Prescher, J.; Bertozzi, C. *J. Am. Chem. Soc*. **2004**, *126*, 15046. (f) Kumararswamy, G.; Ankamma, K.; Pitchaiah, A. *J. Org. Chem*. **2007**, *72*, 9822. (g) Campbell-Verduyn, L.; Elsinga, P.; Mirfeizi, L.; Dierckx, R.; Feringa, B. *Org. Biomol. Chem*. **2008**, *6*, 3461. (h) Shi, F.; Waldo, J.; Chen, Y.; Larock, R. *Org. Lett*. **2008**, *10*, 2409. (i) Moses, J.; Moorhouse, A. *Chem. Soc. Rev*. **2007**, *36*, 1249.

[2] Agard, N.; Prescher, J.; Bertozzi, C. *J. Am. Chem. Soc*. **2004**, *126*, 15046.

[3] Padwa, A.; Hamilton, L. *Tetrahedron Lett*. **1965**, 4363.

[4] Banks, H. *J. Org. Chem*. **2006**, *71*, 8089.

[5] Banks, H. *J. Org. Chem*. **2008**, *73*, 2510.

[6] Bachrach, S. *Computational Organic Chemistry*, Wiley: Hoboken, NJ; 177-190.

[7] Gaussian 03, Revision C.02, Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Montgomery Jr., J.; Vreven, T.; Kudin, K.; Burant, J.; Millam, J.; Iyengar, S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J.; Hratchian, H.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P.; Morokuma, K.; Voth, G.; Salvador, P.; Dannenberg, J.; Zakrzewski, V.; Dapprich, S.; Daniels, A.; Strain, M.; Farkas, O.; Malick, D.; Rabuck, A.; Raghavachari, K.; Foresman, J.; Ortiz, J.; Cui, Q.; Baboul, A.; Clifford, S.; Cioslowski, J.; Stefanov, B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R.; Fox, D.; Keith, T.; Al-Laham, M.; Peng, C.; Nanayakkara, A.; Challacombe, M.; Gill, P.; Johnson, B.; Chen, W.; Wong, M.; Gonzalez, C.; Pople, J.; Gaussian, Inc., Wallingford, CT; **2004**.

[8] (a) Reed, A.; Weinstock, R.; Weinhold, F. *J. Chem. Phys*. **1985**, *83*, 735 -746. (b) Reed, A.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736. (c) Reed, A.; Weinhold, F.; Curtiss, L.; Pochatko, D. *J. Chem. Phys.* **1986**, *84*, 5687. (d) Reed, A.; Curtiss, L.; Weinhold, F. *Chem. Rev*., **1988**, *88*, 899. (e) Carpenter, J.; Weinhold, F. *J. Am. Chem. Soc.* **1988**, 110, 368-72. (f) Goodman, L.; Pophristic, V.; Weinhold, F. *Acc. Chem. Res*. **1999**, *32*, 983. (g) Glendening, E.; Faust, R.; Streitwieser, A.; Vollhardt, K.; Weinhold, F. *J. Amer. Chem. Soc.* **1993**, *115*, 10952.

[9] NBO charges proved useful in our earlier studies of nucleophilic substitution reactions: (a) Banks, H. *J. Org. Chem*. **2008**, *73*, 2510. (b) Banks, H. *J. Org. Chem.* **2006**, 71, 8089. (c) Banks, H. *J. Org. Chem*. **2003**, *68*, 2639. It should be noted that the electrostatic energies obtained using this method are critically dependent on the population analysis used by Gaussian 03. Also, it should be understood that atomic charges are not physical observables, are not necessarily centered at the nuclei. For a discussion of atomic charge see: Wiberg, K.; Rablen, P. *J. Comput. Chem*. **1993**, *14*, 1504.

[10] (a) Mayer, I. *Comput. Chem.* **2006**, 28, 204 -221. (b) Cooper, D.; Ponec, R. *Phys. Chem.* **2008**, *10*, 1319 – 1329. (c) Wei, Y.; Wang, B.; Hu, S.; Chu, T.; Tang, L., Liu, X.; Wang, Y.; Wang, X. *J. Phys. Org. Chem*. **2005**, *18*, 625–631. (d) Wiberg, K.; Marquez, M. *J. Am. Chem. Soc*. **1998**, 120, 2932–2938. (e) Wiberg, K. *Tetrahedron* **1968**, *24*, 1083–1096.

[11] Lown, J., Padwa, A. *1,3-Dipolar Cycloaddition Chemisty*. **1984**, 1, 653–732.

[12] Foresman, J.; Frisch, A. *Exploring Chemistry with Electronic Structure Methods*, 2nd Ed., Gaussian Inc., Pittsburgh, PA.

[13] GaussView 3.0, Gaussian Inc., Carnegie Office Park, Blg. 6, Pittsburgh, PA 15106.

[14] See, for example: Lowry, T.; Richardson, K. *Mechanism and Theory in Organic Chemistry*, 2nd ed.; Harper & Row: New York, NY; p194.

[15] Cramer, C. *Essentials of Computational Chemistry*. Wiley: New York, NY **2004**; 32-33.

[16] An effective dielectric constant of 2.0 was used by Wiberg, K. *J. Org. Chem*. **2002**, *67*, 1613-1617 in his computational study of substituent effects on the acidity of bicyclic carboxylic acids. Due to the reciprocal relationship of $D_{eff}$ and $E_{elst}$, use of $D_{eff}$ = 1.0 generates the largest absolute $E_{elst}$ values and allows assessment of the largest electrostatic effect. The choice of $D_{eff}$ values for transition states however, is problematic.

[17] Dudev, T.; Lim, C. *J. Am. Chem. Soc*. **1998**, *120*, 4450–4458.

[18] (a) Kirmse, W.; Rondan, N.; Houk, K. *J. Am. Chem. Soc*. **1984**, *106*, 7989. (b) Rondan, N.; Houk, K. *J. Am. Chem. Soc.* **1985**, 107, 2099. (c) Nakamura, K,; Houk, K. *J. Org. Chem*. **1994**, *60*, 686. (c) Dolbier, W.; Kornoniak, H.; Houk, K.; Sheu, D. *Acc. Chem. Res.* **1996**, *29*, 471. (d) Um, J.; Xu, H.; Houk, K.; Tang, W. *J. Am. Chem. Soc*., **2009**, *131*, 6664. (e) Lee, P.; Zhang, X.; Houk, K. *J. Am. Chem. Soc*., **2003**, *125*, 5072. (f) Niwayama, S.; Kallel, E.; Sheu, C.; Houk, K. *J. Org. Chem*. **1996**, *61*, 2517. (g) Niwayama, S.; Kallel, E.; Spellmeyer, D.; Sheu, C.; Houk, K. *J. Org. Chem*. **1996**, *61*, 2813. (h) Houk, K. *Strain and Its Implications in Organic Chemistry*, de Meijere, A.; Blechert, S., Eds.; Kluwer Academic Publishers: Dordrecht, **1989**. (i) Rondan, N.; Houk, K. *J. Am. Chem. Soc.* **1985**, *107*, 2099. (j) Yasui, M.; Maruse, Y.; Inagaki, S. *J. Org. Chem*. **2004**, *69*, 7246.

[19] Banks, H. *J. Org. Chem*. **2010**, *75*, 2510-2517.

[20] (a) Chambers, R. *Fluorine in Organic Chemistry*: Blackwell, Baton Rouge, FL, **2004**. (b) Lemal, D. *J. Org. Chem*. **2004**, *69*, 1-11. (c) Snyder, J.; Chandrakumar, N.; Sato, H.; Lankin, D. *J. Am. Chem. Soc*. **2000**, *122*, 544-545. (d) Farnham, W.; Smart, B.; Middleton, W.; Calabrese, J.; Dixon, D. *J. Am. Chem. Soc*. **1985**, *107*, 4565-4567. (e) Goodman, L.; Gu, H.; Pophristic, V. *J. Phys. Chem. A.* **2005**, *109*, 1212-1219. (f) Dolbier Jr., W.; Battiste, M. *Chem. Rev*. **2003**, *103*, 1071-1098. (g) Hunter, L.; *Beil. J. Org. Chem*. **2010**, 6, No. 38.

[21] NBO charges proved to be useful in our earlier studies of nucleophilic substitution reactions: see refs 4, 5, and Banks, H. *J. Org. Chem*. **2003**, *68*, 2639. It should be noted that the electrostatic energies obtained using this method are critically dependent on the population analysis used by Gaussian 03. Also, it should be understood that atomic charges are not physical observables, are not necessarily centered at the nuclei. For a discussion of atomic charge see: Wiberg, K.; Rablen, P. *J. Comput. Chem*. **1993**, *14*, 1504.

[22] March, J. *Advanced Organic Chemistry*, 4th ed.; Wiley: New York, N.Y., **1992**; p21.

[23] (a) Derdour, A.; Texier, F. *Can. J. Chem*. **1985**, *63*, 3605-3612. (b) Oehlschlager, A.; Yim, A.; Aktar, M. *Can. J. Chem*. **1978**, *56*, 273-288.

[24] (a) Schwabe, T.; Grimme, S. *Acc. Chem. Res*. **2008**, *41*, 569-579. (b) Neese, F.; Hansen, A.; Wennmohs, F.; Grimme, S. *Acc. Chem. Res*. **2009**, *42*, 641-648. (c) Hyla-Kryspin, I.; Grimme, S. *Organometallics*, *2004*, **23**, 5581-5592. (d) Grimme, S.; Steinmetz, M.; Korth, M. *J. Org. Chem*. **2007**, *72*, 2118-2126. (e) Wheeler, S.; Moran, A.; Pieniazek, S.; Houk, K. *J. Phys. Chem. A*. **2009**, *113*, 10376-10384. (f) Grimme, S.; Mück-Lictenfeld, C.; Würthwein, E.; Ehlers, A.; Goumans, T.; Lammertsma, K. *J. Phys. Chem. Lett*., **2006**, *110*, 2583-2586. (g) Grimme, S. *J. Chem. Phys*. **2003**, *118*, 9095-9102.

# Investigation of molecule-surface interactions with overtone absorption spectroscopy and computational methods

J. Cabalo[a] and R. Sausa[b]
[a]Edgewood Chemical Biological Center, APG, MD, USA 21010
[b]Army Research Laboratory, APG, MD, USA 21015

## ABSTRACT

The objective of this work is to enhance computational modeling of intermolecular and molecule-surface interactions by means of comparing experimental overtone spectra and predicted overtone spectra. The relevance to the warfighter is that intermolecular interactions affect the detectability of explosives, chemical and biological agents, environmental fates of these materials, as well as surface mediated catalysis. The C-H stretch first overtone spectrum of 1,3,5-trinitrotoluene (TNT) is measured in the region of 5697 $cm^{-1}$ to 6663 $cm^{-1}$ with laser photoacoustic spectroscopy. The Harmonically Coupled Anharmonic Oscillator (HCAO) model is employed to predict overtone vibrational spectra from models of solid TNT that are then compared to the recorded spectrum. The models are single molecule models and a hybrid quantum mechanical/molecular mechanical cluster model to include the effect of intermolecular interactions on the overtone spectrum at the HF/6-31+G(d,p), B3LYP/6-31+G(d,p), X3LYP/6-31+G(d,p), MP2/6-31+G(d,p) levels of theory, and the Universal Force Field (UFF) to account for van der Waals and electrostatic effects from surrounding TNT molecules. The MP2 predicted spectra are closest to the recorded data for both single molecule and cluster models, although some peaks in the laser PAS spectrum are not accounted for by the model. The HF, B3LYP, and X3LYP methods predict small anharmonic constants and are in complete disagreement with results from the MP2 method. The HF and DFT methods do not account sufficiently for electron correlation effects due to the long range coulomb interactions.

## 1. INTRODUCTION

### 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 (Mendoza-Alvarez, J. G., 2004). 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 (Skaare-Rygh, L. E., 1995) and step scan FTIR spectroscopy (Mendoza-Alvarez, J. G., 2004; Zhang, P., 2004). 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 (Pipino, A. C. R., 2004). The objective is to obtain a method for predicting overtone spectra from molecules under the influence of intermolecular interactions. The capability to predict changes in the overtone spectrum due to molecule-bulk and molecule-surface interactions permits the refinement and validation of computational models of the intermolecular interactions. From these models, other important properties such as binding energies can be predicted. Progress to the objective is accomplished using theoretical calculations in tandem with spectroscopic measurements. Because 1,3,5-trinitrotoluene (TNT) is well studied, including its crystal structure (Carper, W. R., 1982), and is a highly relevant energetic material, it is used as a test case.

### 1.2 Background

Vibrational overtone spectroscopy is a potentially powerful tool for studying intermolecular and molecule-surface interactions because it involves the higher vibrational energy levels. When the vibrational potential energy surface of a molecule is perturbed by the presence of surrounding molecules, such as near a surface or within a solvent, the higher energy levels experience more change than the lower levels. The higher energy portion of the potential energy surface is located between the equilibrium position and surrounding molecules, and would experience a greater disturbance than the portions closer to the equilibrium position. 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 energy levels. The change in the potential energy surface also affects ground and excited state nuclear wavefunctions, which in turn affects the transition dipole moment and the absorption intensities. As a result, it is anticipated that magnitude of change in absorption frequencies and intensities for overtones will be greater than that for the fundamental absorptions.

## 2. EXPERIMENTAL

### 2.1 Theoretical considerations

Theoretical calculations are performed using Gaussian 2003 version B.04 on a Sunstation V, as well as a LAN Parallel cluster utilizing 48 Intel processors, and Gaussian 2009, version A.02. The Sunstation has 8 processors and 32 GBytes of RAM memory. Each node in the Gaussian 2009 cluster has 8 processors and 8 GBytes of RAM. Overtone vibrational frequencies of TNT molecules are calculated for single TNT molecules and for TNT molecules at various positions within a cluster of TNT molecules, as shown in Figure 1. The molecular cluster is used to simulate the effect of intermolecular interactions on the overtone spectrum. The geometry of the cluster is first optimized with the Universal Force Field (UFF) before any subsequent calculations, where the qEq algorithm is used to assign partial charges to all the atoms in the cluster. The cluster is then treated as a hybrid molecular mechanical/quantum mechanical model where the molecule undergoing vibrational analysis is treated with ab initio/density functional methods (i.e. Hartree Fock, second order Møller Plesset (MP2), and B3LYP) and the surrounding molecules are treated with the UFF. For Gaussian 2009 this is the Our own N-layered Integrated molecular Orbital molecular Mechanics (ONIOM) method. The ONIOM method allows partial charges and van der Waals forces to act on the part of the system in the quantum mechanical layer of the model. The numbered molecules in Figure 1 are the positions of the molecules treated quantum mechanically in separate calculations. Position 1 is chosen so that the methyl hydrogens are directed into the cluster, and positions 2 and 3 are chosen so that the aryl hydrogens interact with the interior of the cluster. The structure of the unit cell of a TNT crystal serves as the initial geometry of the molecular cluster. The hybrid cluster model geometries are further optimized with the molecule targeted for vibrational analysis treated at the higher level of theory.

There are two possible approaches that are utilized here for predicting overtone frequencies. The first is by applying anharmonic corrections to the frequencies calculated for the normal modes, which is well described in the literature and is available in the Gaussian code (Barone, V., 2005a; Barone, V., 2005b; Clabo, D. A., 1988; Miller, W. H., 1990). The second approach utilizes the local mode model (Child, M. S., 1984; Henry, B. R., 1977; Henry, B. R., 1983; Sowa, M. G., 1991), where individual bond stretches are treated as pseudo-diatomic oscillators. This approach is only valid for X-H stretches, where X is a heavier atom such as C, O, or N.

The Harmonically Coupled Anharmonic Oscillator (HCAO) model (Child, M. S., 1981) is a variant of the local mode model and is used to simulate overtone transitions from a target molecule, 1,3,5-trinitro-toluene (TNT), interacting with its surroundings. The HCAO model allows calculation of both absorption frequencies and intensities of the overtone transitions, and is shown in the literature to be successful at describing the absorptions of several molecules, including chemical agents. The model is well described in the literature (Kjaergaard, H. G., 1992; Kjaergaard, H. G., 1994; Petryk, M. W. P., 2005). For the work described here, the methyl and aryl C-H stretching modes are considered separately from each other, and a separate Hamiltonian is constructed for each set of modes. The wavefunctions are constructed using Morse Oscillator wavefunctions as a basis set.

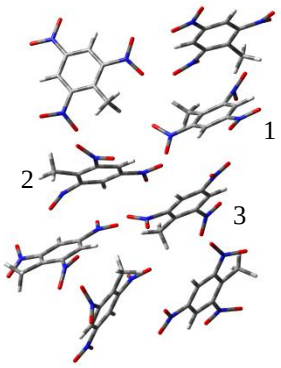



Figure 1: Molecular arrangement of the TNT molecular cluster model. The model is based on an eight molecule unit cell of the TNT crystal. The numbered positions indicate which molecules in the cluster are modeled with HCAO.

### 2.2 Spectral measurements

Overtone and fundamental vibrational spectra are collected using photoacoustic spectroscopy in a commercial MTEC 300 cell that is flushed with Argon gas to remove water vapor contribution to the signal. Photoacoustic spectroscopy is used because it is insensitive to analyte morphology, and is suitable for powdered samples. Fundamental absorption spectra are recorded in a Varian FTIR with the same cell. Tunable near IR laser light is obtained from the idler output of a 10 Hz Nd:YAG pumped Continuum Sunlite EX OPO laser system, while scanning the visible output from 465 nm to 445 nm, with resulting near IR output in the range of 1500 nm to 1755 nm (or 6663 $cm^{-1}$ to 5697 $cm^{-1}$). The pulse energy varied from 591 μJ/pulse at 6430 $cm^{-1}$ to 154 μJ/pulse at 5694 $cm^{-1}$. A Schott glass filter is used to remove residual visible light from the near IR beam. Laser pulse energy is measured using a Coherent Field-Max II meter with a J-10 energy sensor. The MTEC 300 has an onboard preamplifier with an amplification setting of 20. The output is directed to a Tektronix 5104B digital storage oscilloscope and a Stanford Research boxcar integrator, with a maximum gate setting of 3 μs and a delay of 10 ms. The output of the boxcar is digitized and recorded on a computer. Power spectra are recorded by scanning over the tuning range and monitoring the laser energy with a photodiode directed into the boxcar. To calibrate the power curve, the average pulse energy of the IR idler beam is measured with the pulse energy meter at 5 nm intervals of the visible output wavelength. All spectra are normalized with the power curve. Background spectra are recorded with an empty sample cup, normalized with the power curve, and subtracted from the sample data. Near IR and overtone spectra are collected from ~50 mg pure, recrystallized TNT obtained from the Weapons Materials Research Directorate at the Army Research Laboratory, and from ~100 mg munitions grade TNT obtained from the Pyrotechnics Branch at the Edgewood Chemical Biological Center. For preparation of TNT on silica, 12 mg of the munitions grade TNT is dissolved in 10 ml of acetone. 9 ml of this solution is added to 35 mg of Cabosil™ 5M fumed silica, and the acetone is evaporated, leaving a coating TNT on the fumed silica. According to the manufacturer, the fumed silica has a surface area of 200 $m^2/g$ of material. Given the 35 mg mass of silica used in the measurements, there should be 7.0 $m^2$ of amorphous silica surface. Assuming TNT has a diameter of 7 Å and an area of 3.8 x $10^{-19}$ $m^2$, 35 mg of TNT can cover approximately 35 $m^2$ in a monolayer. Assuming the TNT spreads evenly, the layer of TNT on the surface is approximately five molecules deep. Because benzoic acid is also an aromatic solid, an overtone spectrum is recorded for the sake of comparison. The TNT overtone spectra are compared to the predicted spectra from HCAO.

## 3. RESULTS AND DISCUSSION

As shown in Figure 2, both pure and munitions grade TNT spectra show the same features at 5891 $cm^{-1}$, 5925 $cm^{-1}$, 6036 $cm^{-1}$, and the shoulders at 6090 $cm^{-1}$ and 6175 $cm^{-1}$. There is a small shift in the strongest feature from 6131 $cm^{-1}$ to 6143 $cm^{-1}$, and reduction in intensity for the 5960 $cm^{-1}$ when going from pure TNT to the munitions grade. For the TNT adsorbed onto silica, there are strong features centered at 6119 $cm^{-1}$ and 5920 $cm^{-1}$. Most notably there is the absence of the peak at 6036 $cm^{-1}$ which is present in the munitions grade and pure TNT. The spectrum of benzoic acid is distinguishable from TNT, although there is a similar dominant feature. The dominant feature in benzoic acid is red shifted to 6096 $cm^{-1}$, with no additional distinct features in the spectrum.

Differences in the recorded vibrational overtone spectrum reflect not only differences between compounds as shown between the spectrum of TNT and benzoic acid, but also differences in the local molecular environment. There are noticeable differences between the three TNT spectra in Figure 2, both in peak absorption frequencies as well as relative intensities. Simulation of overtone spectra using both single molecules and an eight molecule clusters are performed to elucidate the specifics of the effect of intermolecular interactions. However, the different levels of theory predict radically different spectra.

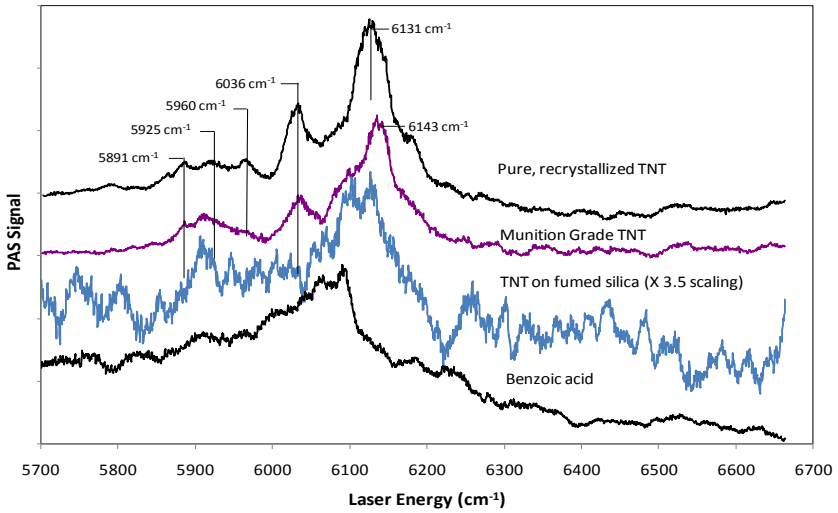



Figure 2: Near IR Laser Photoacoustic Spectra of pure, recrystallized TNT munition grade TNT, TNT solvent deposited on fumed silica powder, and benzoic acid, in the frequency range of 5700 $cm^{-1}$ to 6700 $cm^{-1}$ (1.493 μm to 1.754 μm).

For the single molecule model the HCAO predicted spectra from the levels of theory in Figure 3 have the same general shape as the laser PAS data, with a strong feature at higher frequencies, and two other slightly weaker features at lower frequencies. There appears to be a progression in agreement between the predicted single molecule model absorption frequencies and the level of theory, both for HCAO predicted frequencies and normal mode frequencies with anharmonic corrections, where the MP2/6-31+G(d,p) predicted spectrum is the closest to experiment, then B3LYP/6-31+G(d,p), and HF/6-31+G(d,p). Although results for both the single molecule and cluster molecule models at the MP2/6-31+G(d,p) level of theory are not successfully obtained, in a similar manner to the HCAO results, the HF/6-31+G(d,p) predicted frequencies, as shown in Figure 4, overestimate the aryl C-H stretching frequencies by hundreds of wavenumbers, and the B3LYP/6-31+G(d,p) predicted frequencies are closer to the measured spectra. This is explained by the fact that there is a progression of accuracy at reproducing the curvature of the potential energy surface (PES), from where the force constants are calculated. Also, the B3LYP/6-31+G(d,p) and MP2/6-31+G(d,p) levels of theory are more closely in agreement with the molecular geometries measured using X-ray crystallography.

The general shape of the DFT predicted spectrum is similar to the experimental spectrum, but the peaks due to the methyl C-H stretches are nearly as intense as the aryl C-H stretches, and there is about a 300 $cm^{-1}$ separation between the aryl and the methyl C-H stretches. The MP2/6-31+G(d,p) predicted aryl C-H stretch is within 4 $cm^{-1}$ of the large peak at 6131 $cm^{-1}$, and a methyl C-H peak 3 $cm^{-1}$ from the peak at 5925 $cm^{-1}$. However, not all of the prominent peaks in the experimental spectrum are predicted with the single molecule model. All of the spectra predicted from a single molecule model using HCAO predict a total of three features because the two of the aryl stretches are degenerate, as well as two of the methyl C-H frequencies. The predicted degeneracy can be explained by the symmetry of the molecule, but clearly the experimental spectra indicate that overall the symmetry is reduced because of intermolecular interactions. In addition, two variant geometries of TNT are reported from X-ray crystallography measurements, potentially increasing the number of distinct C-H stretching overtones to ten. Thus, a multi-molecular model is necessary to account for the additional absorption features. The same general pattern of degeneracy is observed for the anharmonic corrected normal mode analysis predicted frequencies, where two of the methyl C-H stretching modes are degenerate, and the two aryl C-H stretching modes are nearly degenerate (by 10-20 $cm^{-1}$).

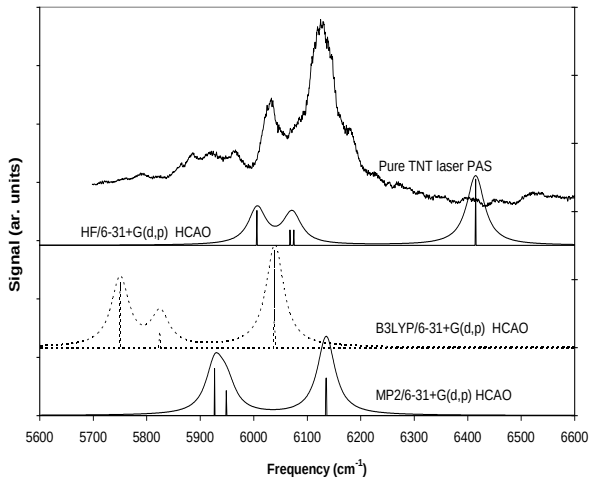



Figure 3: TNT overtone spectrum compared to predicted spectra from the local mode model for 1) HF/6-31+G(d,p), 2) DFT B3LYP/6-31+G(d,p), and 3) MP2/6-31+G(d,p). HCAO corrections to the predicted local mode frequencies are not shown in the predicted spectra.

The results of the predicted overtone spectra from the cluster model are shown in Figure 4. Both the MP2/6-31+G(d,p) and the B3LYP/6-31+G(d,p) levels of theory are employed for the molecule at position 1 in Figure 1, and the results are compared to the spectrum from pure TNT. The aryl C-H stretches point away from the core of the TNT cluster, and are expected to be the least perturbed by the presence of the cluster molecules. The B3LYP/6-31+G(d,p) level of theory not only predicts much different frequencies, but also much attenuated oscillator strengths. For the single molecule model, all three levels of theory predict similar oscillator strengths, certainly well within an order of magnitude. When using a hybrid cluster model, the B3LYP/6-31+G(d,p) predicts a large drop in overtone intensity, so that to compare the DFT predicted spectrum with the MP2 predicted spectrum, the DFT spectrum must be scaled by a factor of 33.

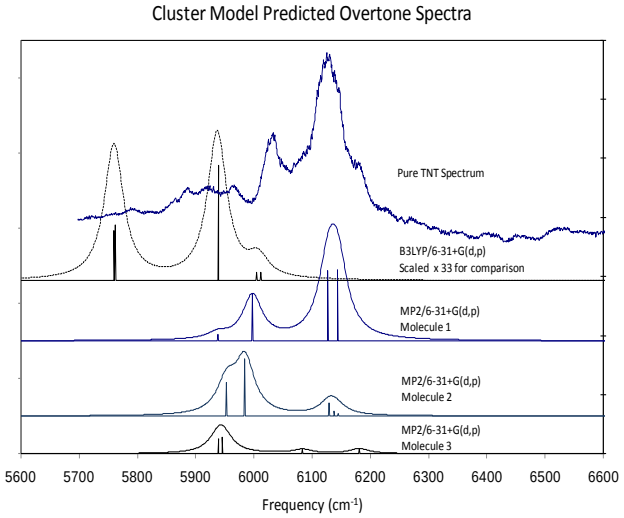



Figure 4: Overtone spectra produced from TNT molecule clusters.

For the single molecule model, the Morse parameters are roughly equivalent between the three levels of theory, where for all modes, anharmonicities of 83 to 87 $cm^{-1}$ are obtained. The results obtained from a model cluster for MP2 are about the same as the single molecule model result, even for the spectrum obtained from position 2. The case is different for the B3LYP/6-31+G(d,p) level of theory, where the anharmonicity constants become much smaller, 0.01-0.02 $cm^{-1}$ for the aryl C-H stretches, and 1-5 $cm^{-1}$ for the methyl C-H stretches. As might be expected, the overtone absorption intensities become correspondingly smaller and the overall shape of the absorption spectrum becomes radically different.

The reason for the reduction in anharmonicity becomes apparent when the potential energy surfaces for the different levels of theory are plotted together on the same scale. Figure 5 shows the HF/6-31+G(d,p), B3LYP/6-31+G(d,p), and MP2/6-31+G(d,p) predicted aryl C-H stretch PES for both single molecule and cluster models. A PES from an additional calculation using the X3LYP/6-31+G(d,p) is also presented on the same chart. For all levels of theory and a single molecule model, there is a clear non-quadratic component to the PES. This component contributes to the cubic force constant used to calculate the anharmonicity. For the cluster model, only the MP2/6-31+G(d,p) level of theory keeps the non-quadratic component. A consequence of both the HF and B3LYP predicted PES's losing the cubic component is that the anharmonicity is greatly reduced and overtone intensity is lost. Both B3LYP and HF inadequately take into account the electron correlation energy, and as a result, do not reproduce long range coulomb or van der Waals interactions well. This is an explanation for the fact that the cluster model/MP2 predicted spectra are in much better agreement with experiment than HF or B3LYP.

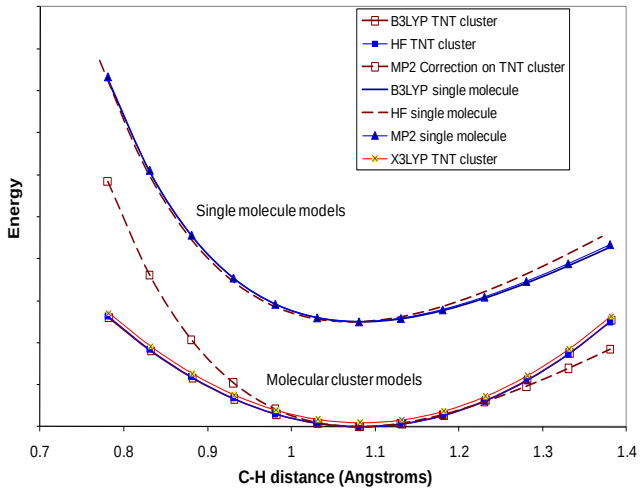



Figure 5: Plot of potential energy surfaces along the aryl C-H stretch coordinate for single molecules of TNT and TNT within an 8 molecule cluster. Neither the B3LYP nor X3LYP functional, nor HF methods reproduce the anharmonic potential energy surface, while the surface calculated with MP2 retains anharmonic character for all attempted calculations.

To further investigate the role of electron correlation effects, the cluster model is optimized and a potential energy surface scan is performed with the X3LYP functional. The X3LYP functional is an extended version of B3LYP that is designed to take into account longer range interactions, especially van der Waals forces. It agrees with experimental measurements of binding energies of noble gas dimers. The exchange functional within B3LYP is found to have some correlation effects, so that when it is combined with the LYP correlation functional, correlation contribution is overestimated. The X3LYP functional is intended to improve accuracy by removing correlation effects from the exchange functional, so that the exchange and correlation energies are properly partitioned. Instead, when the DFT calculation is permitted to interact with the partial charges in the molecular mechanics layer of the cluster, the predicted PES along the C-H stretch coordinate is exactly the same as that predicted by HF. Although the X3LYP functional accounts for van der Waals interactions, it appears the remaining discrepancy between the X3LYP PES and the MP2 predicted PES is due to interactions with the partial charges in the molecular mechanics layer of the ONIOM model of the cluster. The anharmonicity of the B3LYP PES is reduced further for a 64 member cluster, thus indicating that interaction with additional Coulomb charges introduces more electron correlation effects. Doubly hybrid DFT methods are not attempted because they involve the same computational cost as MP2 calculations.

The inclusion of harmonic coupling between anharmonic oscillators does not bring additional agreement with experiment. For the simulated spectra in Figures 3 and 4, only the pure local mode predicted frequencies are represented. Whenever these corrections are applied, differences with the experimental spectra are increased. This indicates that the effect of the surrounding molecules in the model far outweighs the effect of interaction between oscillators. Thus, a pure local mode treatment appears to be appropriate for molecules interacting with a matrix of surrounding molecules.

While use of a cluster model improves agreement between the simulated spectra and the experimental measurements, especially for particular features, the fact remains that a number of experimental spectral features are not reproduced. X-ray crystallography shows that there are two variants of TNT molecular structure present in the crystal. There are

potentially up to ten distinct C-H stretching overtone lines possible. Clearly, while the eight member cluster provides an improvement, it is still inadequate in reproducing the structure in the bulk TNT crystal. As shown in the simulated spectra between position 1 and 2, simulation of the absorption spectrum from different positions in the cluster produce some radical changes in frequency and intensity. The methyl C-H oscillators are more highly sensitive to the local environment than the aryl C-H oscillators, and experience an enhancement of intensity in going from position 1 to position 2 relative to the aryl C-H stretches. Although the aryl C-H stretching frequencies do not change much from position 1 to position 2, one of the methyl transitions changes frequency so much that it overlaps the aryl C-H stretching features. Because the spectrum from position 2 is in disagreement with the measured spectra, it is unlikely that the local environment at that location in the model is reflective of the actual local environment.

To obtain agreement between the predicted overtone spectra and experimental measurement, the molecular model must reproduce the local interactions. There is no position within the eight molecule cluster that completely cages the molecule undergoing HCAO analysis, nor does the arrangement of molecules reproduce the two variant forms of TNT. The overtone absorption spectrum should be successfully reproduced by adding the predicted overtone spectra of the two geometric variants. As the two forms occur in a one to one ratio in the TNT crystal, the contribution to the overall overtone absorption spectrum should also be one to one.

## 4. CONCLUSIONS

The local mode model is applied to the C-H oscillators in the TNT molecule, including HCAO corrections to the local mode frequencies. The model systems are single TNT molecules and eight member TNT clusters, where TNT molecules at various positions are treated with the HCAO analysis. HCAO predicted spectra are compared to experimental measurements of TNT overtone absorption spectra measured with near IR photo-acoustic spectroscopy. Comparison of the overtone spectra calculated from the single molecule model shows agreement with some features in the experimentally measured spectra. However, the single molecule model is inadequate because it predicts degeneracy for the two aryl C-H oscillators and two of the methyl C-H oscillators. There are at least six features in the experimental spectrum, and four of those features are not reproduced by the results of the single molecule model.

Progress towards reproducing the experimental spectra is obtained with a cluster of eight TNT molecules to take into account intermolecular interactions. Degeneracies are removed, so that the intensity and absorption frequency of each C-H oscillator reflect the presence of intermolecular interactions. Comparison of predicted spectra from three positions within the cluster model shows that the overtone spectrum is very sensitive to the surroundings. For the TNT molecule with the aryl C-H oscillators pointing out of the cluster, the aryl vibrational frequencies are close to the main feature in the experimental spectrum. For the positions that point the methyl C-H oscillators away from the cluster, it is the methyl C-H oscillator absorptions that become more intense, and they experience radical frequency shifts. Although two of the cluster predicted spectra do not agree with experiment, progress has been made because there is a way to include intermolecular interactions. Disagreement with the experimental spectrum exists because the current cluster model does not replicate the local environment in the TNT sufficiently. As a result, the overtone spectrum can be used to refine and validate the computational model.

The overtone spectra of the cluster model also show that the method used to calculate the HCAO spectrum must take into account electron correlation effects. PES's calculated from the HF/6-31+G(d,p) and B3LYP/6-31+G(d,p) levels of theory have a small third order contribution. In contrast, the results of the MP2/6-31+G(d,p) level of theory retains the third order contribution and reasonable anharmonicity constants. Furthermore, the MP2 results do not lose oscillator strength as the B3LYP and HF methods do. The difference between the HF, B3LYP, and MP2 methods is that the MP2 method handles electron correlation the most effectively. This result is consistent with published results (Kone, M., 2005; Wodrich, M. D., 2006; Xu, X., 2004) that show accounting for electron correlation is necessary to model nonbonding forces, such as van der Waals interactions and long range Coulombic interactions.

In summary, progress has been made towards utilizing the comparison of empirical overtone spectra to spectra calculated from theory as a means of validating computational models of intermolecular and molecule-surface interactions. It has been shown that laser overtone absorption spectroscopy of solid materials relevant to the Army in powdered form, as well as films deposited on surfaces, is straightforward. In addition, it is demonstrated that electron correlation must be dealt with properly in order to correctly model the targeted intermolecular interactions. We expect

the benefit warfighters will be the enhanced ability to model intermolecular and molecule-surface interactions that are relevant to the warfighter.

## ACKNOWLEDGEMENTS

This work is funded through the US Army's In-house Independent Research (ILIR) Program. We wish to thank Dr. Alan Samuels and Dr. Ronald Miles for kind use of their FTIR in the characterization of the TNT samples, and Dr. Rose Pesce-Rodriguez for providing the recrystallized TNT samples.

## REFERENCES

[1] Barone, V., 2005a: Anharmonic vibrational properties by a fully automated second-order perturbative approach. *Journal of Chemical Physics*, **122**, 014108-014117.

[2] Barone, V., 2005b: Vibrational zero-point energies and thermodynamic functions beyond the harmonic approximation. *Journal of Chemical Physics*, **120**, 3059-3065.

[3] Carper, W. R., L. P. Davis, and M. W. Extine, 1982: Molecular Structure of 2,4,6-Trinitrotoluene. *Journal of Physical Chemistry*, **86**, 459-462.

[4] Child, M. S. and L. Halonen, 1984: Overtone frequencies and intensities in the local mode picture. *Advances in Chemical Physics*, **57**, 1-58.

[5] Child, M. S. and R. T. Lawton, 1981: Local and Normal Vibrational States: a Harmonically Coupled Anharmonic-oscillator Model. *Faraday Discussion of the Chemical Society*, **71**, 273-285.

[6] Clabo, D. A., W. D. Allen, R. B. Remington, Y. Yamaguchi, and H. F. Schaefer III, 1988: A systematic study of molecular vibrational anharmonicity and vibration-rotation interaction by self-consistent-field higher-derivative methods - asymmetric-top molecules. *Chemical Physics*, **123**, 187-239.

[7] Henry, B. R., 1977: Use of local modes in the description of highly vibrationally excited molecules. *Accounts of Chemical Research*, **10**, 207-212.

[8] Henry, B. R., A. W. Tarr, O. S. Mortensen, W. F. Murphy, and D. A. C. Compton, 1983: Raman and Infrared Excitation of Local Mode States in Neopentane. *Journal of Chemical Physics*, **79**, 2583-2589.

[9] Kjaergaard, H. G. and B. R. Henry, 1992: The relative intensity contributions of axial and equatorial CH bonds in the local mode overtone spectra of cyclohexane. *Journal of Chemical Physics*, **96**, 4841-4851.

[10] Kjaergaard, H. G., B. R. Henry, H. Wei, S. Lefebvre, T. Jr. Carrington, O. S. Mortensen, and M. L. Sage, 1994: Calculations of vibrational fundamental and overtone band intensities of H2O. *Journal of Chemical Physics*, **100**, 6228-6239.

[11] Kone, M., B. Illien, J. Graton, and C. Laruence, 2005: B3LYP and MP2 Calculations of the Enthalpies of Hydrogen-Bonded Complexes of Methanol with Neutral Bases and Anions: Comparison with Experimental Data. *Journal of Physical Chemistry A*, **109**, 11907-11913.

[12] Mendoza-Alvarez, J. G., A. Cruz-Orea, O. Zelaya-Angel, G. Torres-Delgado, R. Castanedo-Perez, S. A. Hernandez, and J. Marquez-Marin, 2004: Characterization of TiO2 thin films for photocatalysis applications using photoacoustic spectroscopy. *Journal de Physique IV*, **125**, 407-409.

[13] Miller, W. H., R. Hernandez, N. C. Handy, D. Jayatilaka, and A. Willetts, 1990: Ab initio calculation of anharmonic constants for a transition state, with application to semiclassical transition state tunneling probabilities. *Chemical Physics Letters*, **172**, 62-68.

[14] Petryk, M. W. P. and B. R. Henry, 2005: CH Stretching Vibrational Overtone Spectra of tert-Butylbenzene, tert-Butyl Chloride, and tert-Butyl Iodide. *Journal of Physical Chemistry A*, **109**, 4081-4091.

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

[16] Skaare-Rygh, L. E., I. Gausemel, O. H. Ellestad, P. Klaeboe, C. J. Nielsen, and E. Rytter, 1995: Diffuse refelctance IR studies of bimetallic Fischer-Tropsch catalysts. *Journal of Molecular Structure*, **349**, 325-328.

[17] Sowa, M. G., B. R. Henry, and Y. Mizugai, 1991: Vibrational Overtone Study of Five-Membered Aromatic Heterocycles: Local Mode Interpretations. *Journal of Physical Chemistry*, **95**, 7659-7664.

[18] Wodrich, M. D., C. Cormeinboeuf, and P. von R. Schleyer, 2006: Systematic Errors in Computed Alkane Energies Using B3LYP and Other Popular DFT Functionals. *Organic Letters*, **8**, 3631-3634.

[19] Xu, X. and W. A. Goddard III, 2004: The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proceedings of the National Academy of Sciences*, **101**, 2673-2677.

[20] Zhang, P. and M. W. Urban, 2004: Photoacoustic FT-IR Depth Imaging of Polymeric Surfaces: Overcoming IR Diffraction Limits. *Langmuir*, **20**, 10691-10699.

# Design and testing of a thermostable platform for multimerization of single-domain antibodies

James P. Carney[a], Ellen Goldman[b], Candice Warner[c], and Alena M. Calm[a]
[a]US Army ECBC, RDCB-DRB-M, 5183 Blackhawk Rd, APG, MD, USA 21010
[b]Naval Research Laboratory, 4555 Overlook Ave, Washington, DC, USA 20375
[c]URS, 1309 Continental Drive, Abingdon, MD, USA 21009

## 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. Therefore, 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 (sdAb) 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 (thermoplasticity). 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, which increases the number of binding sites, leading to enhanced recognition. We have used recombinant technology to multimerize single domain antibodies in a flexible, scalable and thermostable platform. We utilized the sequences for two single domain antibodies which recognized unique epitopes on the ricin A chain, and tethered them to a platform based on the structural maintenance of chromosome (SMC) proteins. The SMC portion of the molecule was derived from the thermophilic organism *Pyrococcus furiosus*. The engineered molecules were produced in *E. coli* and their recognition characteristics were determined by Surface Plasmon Resonance (SPR). In addition, the thermostability and thermoplasticity of the molecules was tested up to 65°C. The engineered constructs exhibited very tight binding to ricin, and demonstrated an avidity increase over the sdAb alone. Upon exposure to high temperatures they maintained half of the biological activity. These results indicate that by multimerizing the single domain antibodies on a thermostable platform it is possible to increase avidity and maintain some thermoplasticity. This platform will allow for a 'plug and play' approach to multimerization of single domain antibodies, as well as have the potential for the addition of signaling and/or mitigation activities.

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

## 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 (sdAb) 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 results in single domain antibodies that are extremely thermostable. The use of single domain antibodies is limited by the lack of tumor cell lines to produce hybridomas, and the arduous isolation by phage display technology often leads to reduced binging 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 there are well characterized SMC proteins from the archael kingdom, including the thermophiles (6).

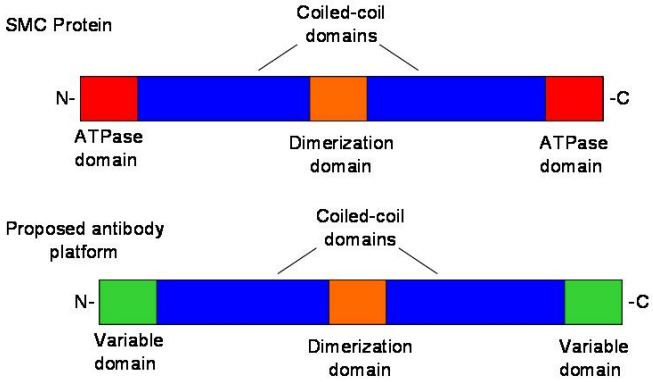



Figure 1. Comparison of SMC molecules and the Antibody Nanoassembly. Upper portion shows the schematic of a standard SMC protein. Lower portion indicates how the antibody nanoassembly was created by substituting the ATPase domain of the SMC protein with variable domains of a llama single domain antibody.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

## 2. METHODOLOGY

### 2.1 Design of C8smcF11

Published Sequences of *Pyrococcus furiosus* were used along with structural and functional data to design the construct (6-8). 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. Dr. Ellen Goldman of the Naval Research Laboratory (NRL) provided the sequences for α ricin C8, and α ricin F11; two single domain antibodies derived from llama that recognize distinct epitopes of the ricin A chain. These sequences were incorporated into the SMC construct at the site of the ATPase domain (2,3,9). In total, three constructs were designed and designated as C8smcF11:1-1, 1-2, and 1-3. Each construct contained the two sdAb sequences C8 and F11, with a varied length of coiled coil in each, where 1-1 is the full length version, 1-2 has a truncated length of coiled coil, and 1-3 consists of the dimerization domain only (Figure 2).

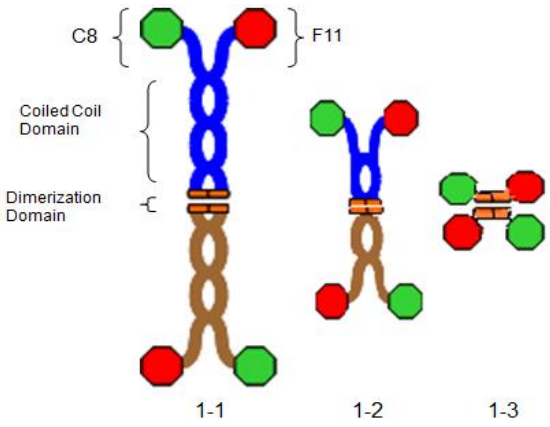



Figure 2. Illustration of C8smcF11. The red and green octagons represent the two different single domain ricin antibodies. The variable domains are tethered to the SMC coiled-coil backbone, which then dimerize. This allows for the creation of a ricin binding heteropolymeric quadravalent recognition moiety. Three constructs were designed, one with a full length coiled coil domain, one that was truncated, and one that contained the dimerization domain only.

Initially, DNA2.0 cloned the sequence for C8smcF11:1-1 into two different vectors, pJexpress401 and pET28a(+) for evaluation. Both plasmids were received lyophilized. Upon receipt they were rehydrated and transformed into BL21(DE3)SLyD- strains of *E. coli.* A pilot expression study was done to compare the expression levels between the two vectors, results indicated that pET28a(+) was the better vector (Figure 3). DNA2.0 then cloned the sequences for C8smcF11:1-2, and 1-3 in pET28a.

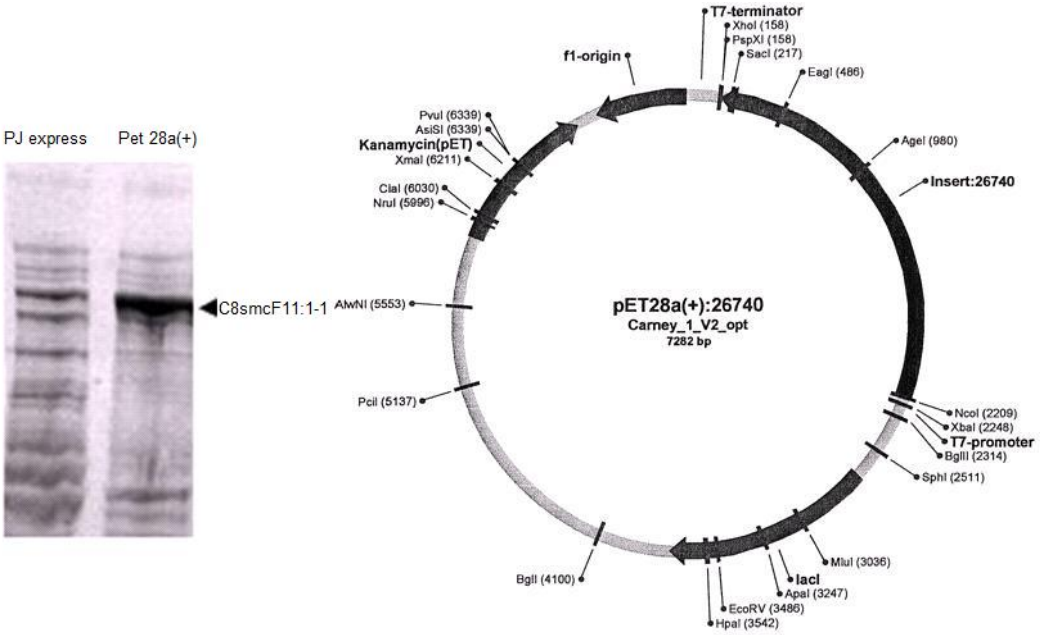



Figure 3. (Left) Pilot expression comparing pJexpress and pET 28a(+) in BL21(DE3)SLyD- *E. coli*. The 80kDa band highlighted corresponds to the predicted molecular weight of C8smcF11:1-1. (Right) Vector map supplied by DNA 2.0 of C8smcF11:1-1 in pET28a(+).

### 2.2 Protein expression and purification

Both α ricin C8, and α ricin F11 single domain antibodies were kindly provided by Dr. Ellen Goldman from NRL. The three C8smcF11 constructs were expressed in BL21(DE3)SLyD-electrocompetent *E. coli.* Transformations were performed using electroporation. Briefly, 100µl of electrocompetent cells, and the 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 on LB plates with appropriate antibiotic (30µg/ml kanamycin). Plates were incubated overnight at 37ºC. 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 (30µg/ml) was inoculated with a single transformant and grown to log phase before being scaled up to a 500ml culture, these half liter cultures were then grown at 37ºC, 300rpm for 20 hours. The cell paste was collected via centrifugation and stored at -80ºC. Crude analysis of the three constructs revealed that C8smcF11:1-2 had very poor expression levels. We focused our efforts on 1-1, and 1-3. The B-PER Bacterial protein extraction reagent (pierce) was used to determine if the proteins were soluble or insoluble. 1-3 was almost completely insoluble and had to be purified under denaturing conditions (8M Urea).

### 2.3 Soluble purification of C8smcF11:1-1

Each construct was designed with a 6x his tag to ease purification, therefore the 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 (20mM Sodium Phosphate, 500mM Sodium Chloride, 10mM Imidizole pH 8.0). The cells were ruptured by 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. The initial product was noticeably degraded. The purification scheme was modified to use the AKTA Purifier FPLC. The first columns employed were 1ml HisTrap FF pre-packed nickel columns (GE Healthcare). The cell paste was processed as described previously; with the addition of EDTA free protease inhibitors (Thermo Fisher). The clarified lysate was pumped onto the column and washed, prior to elution with 300mM imidizole, the product from these first FPLC runs were still degraded. Procedural modifications including, faster processing of cell paste and the use of extra protease inhibitors helped reduce proteolysis. We also switched to a 1ml HisTrap FF crude column; this helped alleviate problems with high back pressure. The elution was changed so that the concentration of imidizole was stepped up to 500mM in 4 steps. After these modifications, the protein eluted off the nickel columns intact. A considerable amount of non-specific protein came off in the elution. In order to remediate this, ion exchange was added as a final step. The peak fractions from the HisTrap were pooled (Figure 4) 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. Fractions were analyzed by SDS PAGE. The fractions were pooled and buffer exchanged into PBS-T 7.4 (Bio-Rad) for SPR analysis. Aliquots were made and snap frozen in liquid nitrogen.

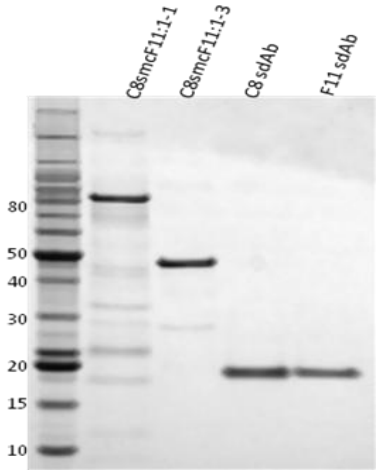



Figure 4. 4-12% SDS PAGE, lanes one and two are pooled fractions from IMAC, lanes three and four are samples that were supplied to us by Ellen Goldman.

**2.4 Insoluble purification of C8smcF11:1-3**

Frozen cell paste was thawed on ice and resuspended in lysis buffer (20mM Sodium Phosphate, 500mM Sodium Chloride, 10mM Imidizole pH 8.0). The cells were ruptured by French press and centrifuged to pellet the insoluble material. The lysate was discarded and the insoluble material was resuspended in lysis buffer containg 8M Urea. Once the cell material was solubilized, the lysate was clarified via centrifugation. The clarified lysate was pumped over a 1ml His trap crude column (GE Healthcare) at room temperature. The column was washed and the protein was refolded on the column by decreasing the concentration of Urea by 10 percent every ten column volumes. At 4M Urea, the column was attached to a refrigerated AKTA purifier to complete the refolding. Once the Urea was washed out of the column, the protein was eluted with 500mM imidizole. The yield was very poor, and upon further analysis it was discovered that about 50% of the protein remained in the flow through. To recover the protein, the flow through was incubated with 5ml of packed BD Talon resin for three hours at room temperature. The Talon resin was then washed extensively, and the protein was refolded on the beads by decreasing the concentration of urea in the buffer with each wash. Once the concentration reached 4M urea, the work proceeded on ice. After refolding, the protein was eluted off in eight 2ml elution steps using 500mM imidizole elution buffer with 1 minute incubation before centrifugation. The recovered protein was analyzed by SDS PAGE (Figure 4). The fractions were pooled and buffer exchanged into PBS-T 7.4 (Bio-Rad) for SPR analysis. Aliquots were made and snap frozen in liquid nitrogen.

### 2.5 Surface plasmon resonance

The ProteOn (Bio-Rad) was used to assess whether or not our constructs would recognize and bind to ricin and RCA 120, it was also used to measure the activity of the proteins after being exposed to thermal stress.

*Assay for protein activity*. 276RU of Ricin was covalently coupled to a ProteOn GLC sensor chip using amine chemistry. Three-fold serial dilutions of C8smcF11:1-1 and 1-3 were used as the analyte and were flowed across the surface of the chip at 100µl/minute for two minutes, with a ten minute dissociation.

*Thermal stress test*. 276RU of Ricin was tethered to a ProteOn GLC sensor chip using amine chemistry. A concentration scouting experiment was run using C8smcF11:1-1, C8smcF11:1-3, α ricin C8 sdAb, and α ricin F11 sdAb as analytes to obtain an $R_{Max}$ between 50-100RU. The proteins were diluted to the following concentrations: C8smcF11:1-1 (4 nM), C8smcF11:1-3 (12 nM), α ricin C8 (800 nM), and α ricin F11 (419 nM). Once diluted, the proteins were exposed to the following temperatures for fifteen minutes: 4°, 37°, 45°, 55°, and 65°C. After the exposure was complete, all samples were placed on ice and immediately run on the ProteOn at a flow rate of 100 µl/minute for 60 seconds with a ten minute dissociation. Binding was measured in Response Units.

*Assay for avidity*. 965RU of RCA 120 was tethered to a ProteOn GLC sensor chip surface using amine coupling chemistry. 50nM of C8smcF11:1-1, 150nM of C8smcF11:1-3, 3µM of the sdAb C8, and 1µM of the sdAb F11 were interacted with the surface of the RCA 120 chip for 60 seconds at a flow rate of 100µl/minute with a dissociation time of 10 minutes. Dissociation constants were determined using Scrubber software.

## 3. RESULTS AND DISCUSSION

*Assay for protein activity*. C8smcF11:1-1 and C8smcF11:1-3 were evaluated by SPR to determine if the engineered sdAb portion of the molecule would retain activity once tethered to the SMC backbone. Results indicate that both C8smcF11:1-1 and C8smcF11:1-3 recognize and bind to Ricin (Figure 5) and RCA 120.

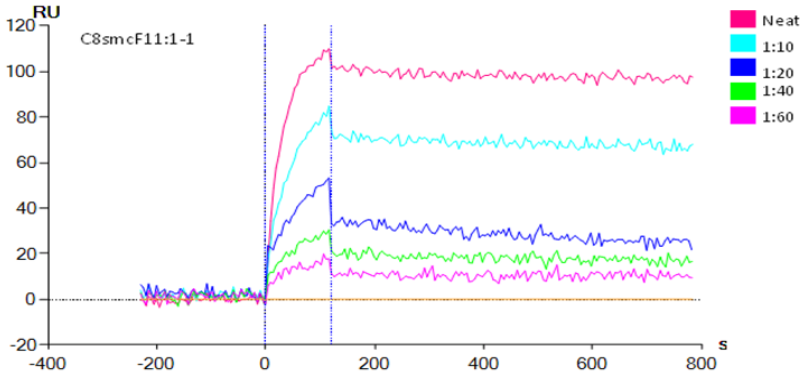



Figure 5. SPR Sensogram depicting a scouting experiment. C8smcF11:1-1 dilutions are shown binding to a chip that had ricin covalently linked to its surface. On the x-axis, -200-0 is the baseline, 0-150 is the association phase when binding occurs, and 150-600 is the dissociation phase.

*Assay for avidity*. One of our goals in constructing these proteins was to improve the binding of the sdAb. Due to the lack of hybridoma technology, the creation of some single domain antibodies relies on *in vitro* panning methods. This often leads to a general reduction in affinity towards the target antigen when compared to standard mouse monoclonal antibodies (1). By fusing C8 and F11 to a coiled-coil backbone, attached to a dimerization domain, we were able to create a tetrameric protein that has a greater avidity for RCA120 when compared to the sdAb alone. Analysis of SPR data (Figure 6) shows an order of magnitude decrease in dissociation constant between the constructs we made and the sdAb F11. This experiment illustrates that while both of our constructs and the sdAb F11 binds to RCA120, the C8smcF11:1-1 and C8smcF11:1-3 exhibits an increase in avidity which causes it to dissociate from RCA120 at a slower rate. In this experiment we also attempted to demonstrate this affect with sdAb C8, however we did not get a high enough $R_{Max}$ to determine the dissociation constant.

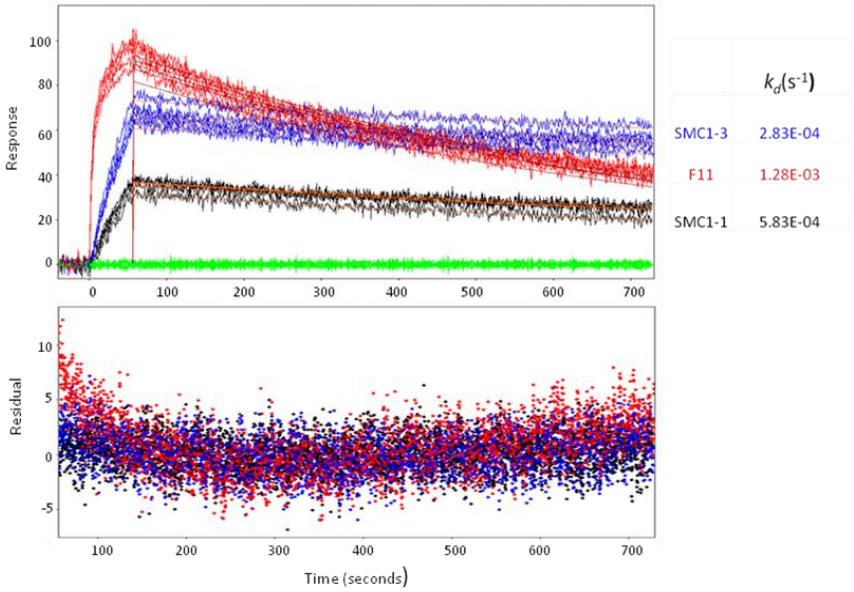



Figure 6. Sensogram with Langmuir 1:1 fit of C8smcF11:1-1, 1-3, and sdAb F11 binding to RCA 120 with residuals shown below. Dissociation constants are shown to the right.

*Assay for thermoplasticity*. A unique feature of single domain antibodies is their ability to unfold and refold efficiently upon heating and cooling. This thermoplasticity is likely the result of the molecules small size, and increased hydrophilicity (1). By constructing a protein that incorporates the structure of a single domain antibody on a platform that is derived from a thermophile we were trying to maintain thermoplasticity while increasing the avidity due to multimerization. The results from the thermal stress test were extrapolated from SPR data. Percent activity was determined by obtaining report points at $R_{Max}$, for each protein and temperature combination (Figure 7), and assigning one hundred percent activity to the 4°C sample. When percent activity is plotted against temperature (Figure 8), trend lines show that as temperature increases, the percent activity of C8smcF11:1-1 and C8smcF11:1-3 decreases. SdAb C8 and F11 retain 100% of its activity throughout the thermal stress test, and actually appears to exhibit an increase in specific activity. This is most likely due to instrument noise and is not a significant increase. Both C8smcF11:1-1 and C8smcF11:1-3 retain about 50% of their activity up to 55°C.

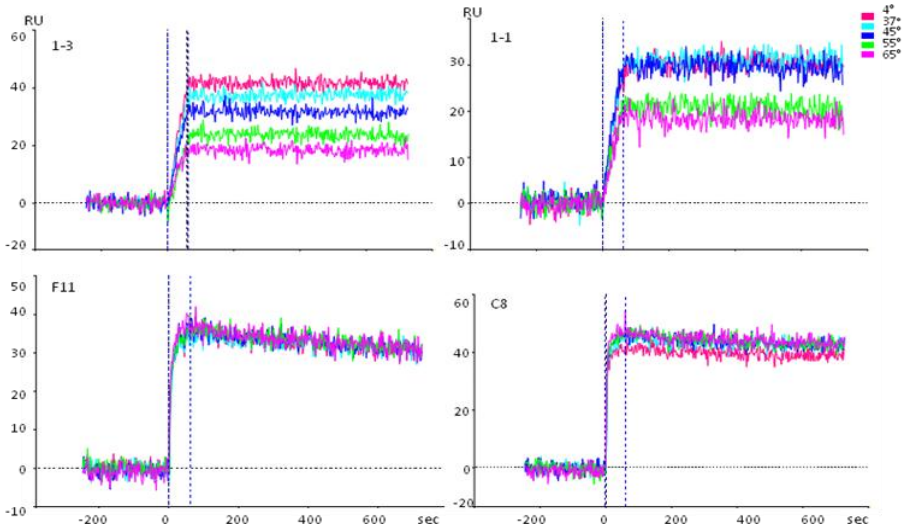



Figure 7. SPR sensograms of C8smcF11:1-1, 1-3, and sdAb's C8and F11 after the thermal stress test.

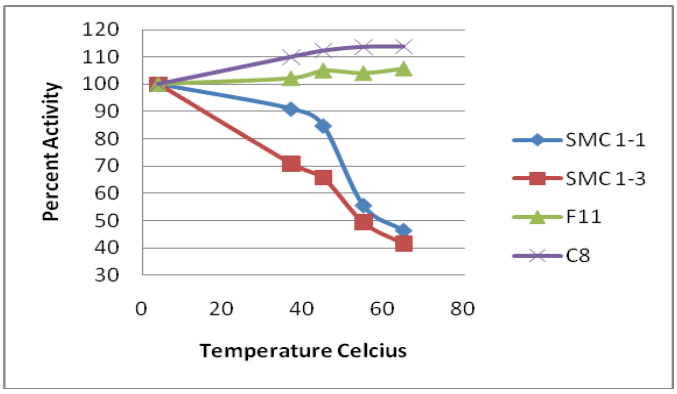


Figure 8. Percent activity of proteins after being exposed to various temperatures.

## 4. CONCLUSION

We have designed two proteins with two unique binding sites derived from single domain antibodies specific for ricin. The binding sites are tethered to a coiled-coil backbone with a dimerization domain derived from the SMC protein of the thermophile *Pyrococcus furiosus*. When compared to anti-ricin F11 sdAb, C8smcF11:1-1 and 1-3 both display an increase in avidity for ricin, demonstrated by the 10 fold decrease in dissociation rate. The avidity improvement is due to the heteropolymeric quadravalent nature of these unique ricin binding molecules. The thermoplasticity of the sdAb was compromised when tethered to the coiled-coil backbone. Both C8smcF11:1-1 and 1-3 experienced a loss of specific activity as they were exposed to increasing temperatures. The act of fusing the sdAb to the SMC backbone increased the size of the protein dramatically; this likely had an effect on its ability to refold after heating. There is also the possibility that the place where the sdAb and the SMC backbone are joined is critical for proper re-folding. The fact that both proteins constructed retain 50% activity after being heated to 55°C is promising. The results from this study indicate that by multimerizing the single domain antibodies on a thermostable platform it is possible to increase avidity and maintain some thermoplasticity. This platform will allow for a 'plug and play' approach to multimerization of single domain antibodies (1) as well as have the potential for the addition of signaling and/or mitigation activities. Future efforts will be focused on improving the construct design to maximize protein refolding after exposure to high temperatures, and testing other single domain antibodies using this platform.

## ACKNOWLEDGEMENTS

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

## REFERENCES

[1] 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.

[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] 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.

[4] 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.

[5] Nasmyth, K. and Haering, C.H. (2005). The structure and function of SMC and kleisin complexes. Annual review of biochemistry, 74, 595-648.
[6] 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.
[7] 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.
[8] 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.
[9] 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.

# Asymmetric torsion equations for large molecules

Bryan J. Schindler[a], Margaret M. Hurley[b], Alex Balboa[c]
[a]SAIC Gunpowder Branch, PO Box 68, APG, MD, USA 21010
[b]Edgewood Chemical Biological Center, 5183 Blackhawk Rd, APG, MD, USA 21010
[c]U.S. Army Research Laboratory, Weapons and Materials Research Directorate, APG, MD, USA 21005

## ABSTRACT

A new equation for torsion potentials for dihedral angles was fit for molecules where there is no symmetry. Quantum mechanical (QM) calculations were run to calculate the rotational barrier for two large molecules, dimethyl methyl phosphonate and Sarin, which were used to scan the potential energy surface (PES) for the dihedrals of interest. The torsion potentials were fit to two classical equations and a new equation based on a full Fourier series. The two classical equations fit a symmetric curve about the zero (0) dihedral angle of rotation while the proposed full Fourier series closely fits the asymmetric dihedral angles PES. In order to determine the appropriateness of the new equation, liquid densities were calculated with Monte Carlo (MC) simulations and compared with both previous MC simulations and experimental data.

**Keywords:** Potential, Quantum Chemistry, DMMP, Sarin, Monte Carlo, Force Field

## 1. INTRODUCTION

A burgeoning application of Monte Carlo (MC) or Molecular Dynamic methods is in probing the physical characteristics of chemical systems with the usage of classical force fields. Recently, force fields have been proposed by Sokkalingam et al.[1] and Visnyakov and Neimark[2] which are suitable for classical simulations of organophosphorus (OP) compounds such as dimethylmethylphosphonate (DMMP) and Sarin. These force fields have been tuned to reproduce thermal and physical characteristics of these compounds, and have demonstrated an excellent degree of transferability. DMMP and Sarin both have multiple conformers which have been studied experimentally[3]. As the conformational landscape is expected to play an important role in the interactions of these compounds with surfaces and within solutions, a thorough description of torsional asymmetry within these systems is in order. The current work attempts to improve upon these force fields by focusing on the asymmetry of torsions within these systems.

Multiple functional forms may be used to describe energetic barriers to dihedral rotation. Radom et al.[4] fit the energetic barriers to dihedral rotation for a series of molecules using a truncated Fourier cosine series shown in equation 1. Steele[5] later introduced equation 2, shown below. In a later paper by Jorgenson[6], equation 1 was used along with a full Fourier series version shown in equation 3, although this was used primarily for pseudorotations involving puckering and phase angle and has not been used to describe torsion behavior. For organophosphorus compounds, variations on equation 1 have been proposed by Sokkalingam et. al[1] and Visnyakov and Neimark[2]. Other work by Ji et al.[7] used Gaussian functions for molecules without rotational symmetry; however, the functions used are not periodic so they were not examined in this study.

$$V = V_0 + \sum_{i=1}^{N} V_i \left[1 - \cos(i\phi)\right] \tag{1}$$

$$V = V_0 + \sum_{i=1}^{N} V_i \cos\phi^i \tag{2}$$

$$V = V_0 + \sum_{i=1}^{N} A_i \cos(i\phi) + B_i \sin(i\phi) \tag{3}$$

## 2. METHODOLOGY

Both Møller–Plesset perturbation theory (MP2) and Hartree-Fock (HF) calculations[8] with a 6-31+G** basis set were used to perform a relaxed scan on the O=P-O-$CH_3$ dihedral angle in the DMMP and the P-O-CH-$CH_3$ and $CH_3$-P-O-CH dihedral for Sarin. These levels of theory were chosen because they have been demonstrated to be sufficient by previous results. Additionally, it was desirable to maintain consistency with previous results in order to demonstrate effects solely from the torsion potential and not from additional perturbations of the theory. Partial charges were calculated using the CHELPG analysis and all calculations were performed using the Gaussian 09 software suite. Figure 1 shows the results of the MP2/6-31+G** calculations with the best data fit using equations 1-3 for DMMP. These torsion data were fit using torsion equation(s) with thirteen terms using the least squares analysis of non-linear least squares. The number of parameters was chosen to be large to allow the torsion equations to assume an optimum fit to the torsion data.

## 3. DATA

The data shows good agreement with previous *ab initio* runs by Cuisset et al.[9], and is a mirror image of the work done by Sokkalingam et al.[1] The minimum energy structure in the figure corresponds very closely to the lowest energy conformer found previously.[3,10,11,12] Here, the MP2/6-31+G** global minimum corresponds to a conformer with O=P-O-CH3 dihedrals of -51.015 and -27.50, while the HF/6-31+G** structure is similar with dihedrals of -46.863 and -27.501.

## 4. RESULTS

This compares well to previously published results of -48.876 and -25.132[3] [HF/6-31G*], or -47.9 and -26.4[9] [B3LYP/6—311++G(3df,2pd)]. When the various torsion equations are fit to the QM dihedral scan data it can be seen that equations 1 and 2 overlay each other. Work done by Lim[13] showed that equations 1 and 2 are related to each other, thus resulting in their overlap. Equations 1 and 2 showed a maximum energy at angles of -123 and 123 degrees. Both have a maximum barrier energy value of 4.7 kcal/mol, where as the *ab initio* results have values of 5.718 and 3.746 kcal/mol at -124.67 and 125.32 degrees, respectively. Since equation 3 does not impose symmetry, it is able to more fully describe the potential energy surface (PES), and therefore allow the system to more properly probe the extrema. The *ab initio* results show two distinct minima. There is a global minimum at -24.67 degrees and a second minimum at 35.32 degrees with a value of 0.2049 kcal/mol. Equations 1 and 2 show some of the same behavior mentioned earlier, specifically that the minima are symmetric about 0 degrees at ±26.88 degrees with values of 0.090 kcal/mol. Equation 3 shows behavior that allows for different energies and is not symmetric about 0 degrees. The minimums for equation 3 are at -27.44 and 37.42 degrees and have values of 0.0094 and 0.1706 kcal/mol, respectively. The low barrier found here is in accord both with previous theoretical PES scans and with previous conformational studies, which have found conformer energies to be "essentially degenerate"[3].

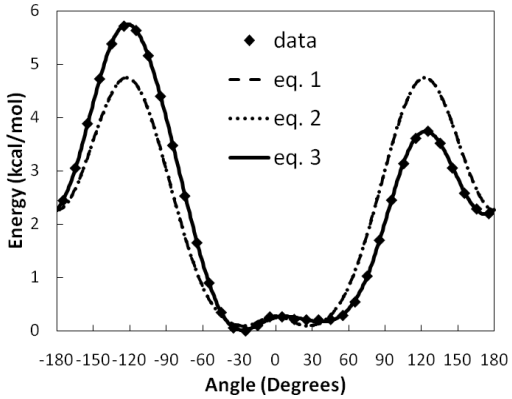



Figure 1: Torsional barrier for O=P-O-CH3 dihedral in DMMP.

The structure of Sarin was subsequently studied using the same methods and approaches on the P-O-CH-$CH_3$ dihedral. Four conformers were found with the first and second conformers being very close in energy. The structures of the first three conformers show good agreement with the data from Kaczmarek et. al[14]. While the rotational spectrum of conformer 4 has been obtained and related calculations performed at the MP2/6-311G** level, its structure has not been analyzed in this level of detail. The minimum energy structure (conformer 1) at MP2/6-31+G** shows P-O-CH-$CH_3$ dihedrals of -157.67 degrees and 79.987 degrees. Conformer 1 optimized at the HF/6-31+G** level has P-O-CH-$CH_3$ dihedrals of -147.67 and 89.58. The MP2 structure compares well with previous theoretical results[14] at the MP2/6-31G(d,p) showing a P-O-CH-$CH_3$ dihedral of -156.4. Equations 1-3 were fit to the PES dihedral angle scan energies and the results are shown in Figure 2. It is readily apparent that equations 1 and 2 lack the flexibility to represent the dihedral for Sarin adequately. Equations 1 and 2, which overlap each other again, are constrained by the need to be symmetric about the angle of rotation and energy, thus causing arbitrary broad shoulders in the ±50-130 degree range and missing the minimum and maximum energy conformers. Equations 1-2 have a maximum at 0.0 degrees of 3.184 kcal/mol, a global minimum at ±180.0 degrees of 0.429 kcal/mol, and a second minimum at ±65.3 degrees of 1.55 kcal/mol. Equation 3, lacking this constraint, shows good fit with the data, with two maximums at -1.28 degrees and 106.52 degrees with values of 3.24 and 3.299 kcal/mol, respectively. Equation 3 also shows minimums at -157.64 and -102.61 degrees with values of 0.081 and 0.0081 kcal/mol, respectively.

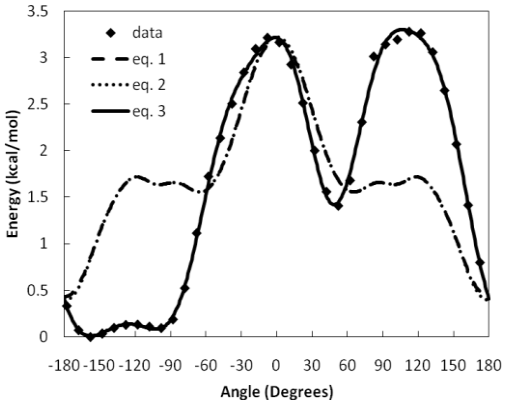


Figure 2: Torsion barrier for P-O-CH-$CH_3$ dihedral in Sarin.

Monte Carlo (MC) simulations were run to compare the effect of the different torsion equations. All MC simulations were performed with the MC simulation program MCCCS Towhee[15]. The parameters used are from Sokkalingam et al.[1] except for the torsion equation. NPT simulations were run to calculate the liquid density at temperatures of 303K and 373K for direct comparison with experimental data (Table 1). It is noted that the modified torsion, equation 3, shows somewhat improved agreement with experimental, relative to equation 1. Agreement of equation 3 results with experimental is comparable to results obtained with equation 2 at lower temperatures. Equation 3 density values are slightly off at higher temperature.

Table 1: Monte Carlo simulations of liquid DMMP at temperatures 303K and 373K

| | Density (g/ml) | | | |
|---|---|---|---|---|
| Temperature (K) | Eq. 1 | Eq. 2 | Eq. 3 | Experimental[16] |
| 303 | 1.2350 | 1.1539 | 1.1509 | 1.1507 |
| 373 | 1.1742 | 1.0769 | 1.1224 | 1.0717 |

Radial distribution functions (rdf) were then obtained from the MC simulation results using the analyze_movie utility of MCCCS Towhee. These are shown in Figure 3. Bin sizes were chosen to compare with previous theoretical results. It can be noted from the O(6)-C(7) and P(3)-P(3) plots (Figures 3a and 3b) that equation 1 gives rise to more short-range order than equations 2 and 3; see supporting information for schematic of DMMP. All four rdf plots shown below are in good agreement with previously published results. As has been previously noted, it is unfortunate that no experimental data is availiable for this system.

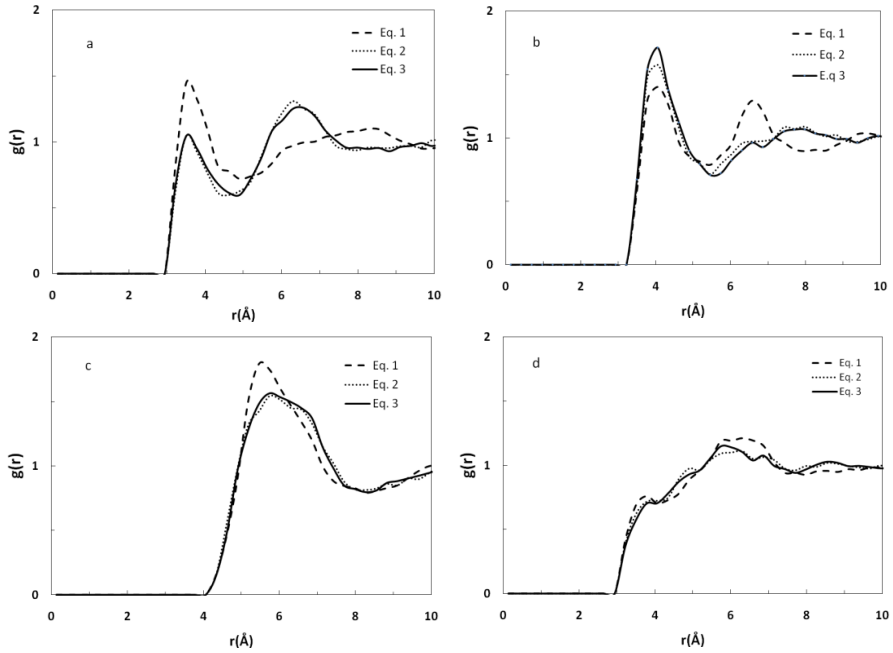



Figure 3. Radial distribution function for DMMP at 303 K.
a) O(6) – C(7); b) C(5) – C(5); c) P(3) – P(3); d) C(7) – O(2).

In addition, the dihedral angle distribution from the MCCS Towhee runs was obtained and shown in Figure 4. It can be seen in comparing the dihedral distribution in Figure 4 with the torsion profile of equation 3 shown in Figure 1 that these runs are indeed sampling the correct minima corresponding to the most likely conformers of this molecule.

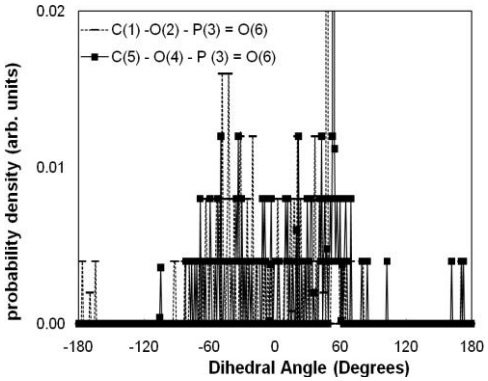



Figure 4. Normalized probablility density as a function of the C(1) – O(2) – P(3) = O(6) and C(5) – O(4) – P(3) = O(6) dihedral angles representing the orientation of the two methoxy arms in the DMMP molecule at 303K for eq. (3).

## 5. CONCLUSIONS

We have demonstrated the successful application of a truncated Fourier series torsion form to two representative asymmetric molecules, namely DMMP and Sarin, to fully describe their complex dihedral PES. In conjunction with previously developed classical parameters to describe the remaining energy terms, this form has been demonstrated to correctly reproduce physical properties such as the density, and the vapor pressure which preliminary data shows to be correct within 10% and is a topic of future work. Additionally, the preliminary structural data shown here in terms of the radial distribution functions and dihedral distributions for this current torsional form have been show to be consistent both with previous classical results and with current quantum mechanical data. In addition, work by Badenhoop and Weinhold[17] has shown that the use of natural bond orbital (NBO) steric analysis provides some indication of the importance of localized bond orbitals on the torsion barrier.

## ACKNOWLEDGEMENT

This work is supported by the Office of the Assistant Secretary of the Army for Acquisition, Logistics, and Technology. The authors kindly acknowledge Professor Jeffrey J. Potoff and Dr. Nandhini Sokklingam for providing the initial Towhee input files, and Mr. Leonard Buettner for his computer hardware support, and Dr. John Mahle for his computer software support.

## REFERENCES

[1] N. Sokkalingam, G. Kamath, M. Coscione *et al.*, "Extension of the Transferable Potentials for Phase Equilibria Force Field to Dimethylmethyl Phosphonate, Sarin, and Soman," Journal of Physical Chemistry B, 113(30), 10292-10297 (2009).

[2] A. Vishnyakov, and A. V. Neimark, "Molecular model of dimethylmethylphosphonate and its interactions with water," Journal of Physical Chemistry A, 108(8), 1435-1439 (2004).

[3] A. R. H. Walker, R. D. Suenram, A. Samuels *et al.*, "Rotational spectrum of Sarin," Journal of Molecular Spectroscopy, 207(1), 77-82 (2001).

[4] L. Radom, J. A. Pople, and W. J. Hehre, "MOLECULAR-ORBITAL THEORY OF ELECTRONIC-STRUCTURE OF ORGANIC COMPOUNDS .13. FOURIER COMPONENT ANALYSIS OF INTERNAL-ROTATION POTENTIAL FUNCTIONS IN SATURATED MOLECULES," Journal of the American Chemical Society, 94(7), 2371-& (1972).

[5] D. Steele, "AN ABINITIO INVESTIGATION OF THE TORSIONAL POTENTIAL FUNCTION OF N-BUTANE," Journal of the Chemical Society-Faraday Transactions Ii, 81(JUL), 1077-1083 (1985).

[6] W. L. Jorgensen, "THEORETICAL-STUDIES OF MEDIUM EFFECTS ON CONFORMATIONAL EQUILIBRIA," Journal of Physical Chemistry, 87(26), 5304-5314 (1983).

[7] X. B. Ji, L. M. Yan, and W. C. Lu, "New torsion potential expression for molecules without rotational symmetry," Journal of Chemical Physics, 128(22), 9 (2008).

[8] M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, [Gaussian 09, Revision A.02] Gaussian, Wallingford CT (2009).

[9] A. Cuisset, G. Mouret, O. Pirali *et al.*, "Gas-phase vibrational spectroscopy and ab initio study of organophosphorous compounds: Discrimination between species and conformers," Journal of Physical Chemistry B, 112(39), 12516-12525 (2008).

[10] A. M. McAnoy, M. R. L. Paine, and S. J. Blanksby, "Reactions of the hydroperoxide anion with dimethyl methylphosphonate in an ion trap mass spectrometer: evidence for a gas phase alpha-effect," Organic & Biomolecular Chemistry, 6(13), 2316-2326 (2008).

[11] P. F. Conforti, M. Braunstein, and J. A. Dodd, "Energetics and Dynamics of the Reactions of O(P-3) with Dimethyl Methylphosphonate and Sarin," Journal of Physical Chemistry A, 113(49), 13752-13761 (2009).

[12] L. Yang, R. M. Shroll, J. X. Zhang *et al.*, "Theoretical Investigation of Mechanisms for the Gas-Phase Unimolecular Decomposition of DMMP," Journal of Physical Chemistry A, 113(49), 13762-13771 (2009).

[13] T. C. Lim, "A note on mathematical relationships among bond-torsion force fields," Journal of Mathematical Chemistry, 31(4), 421-428 (2002).

[14] A. Kaczmarek, L. Gorb, A. J. Sadlej *et al.*, "Sarin and soman: Structure and properties," Structural Chemistry, 15(5), 517-525 (2004).

[15] M. G. Martin, [MCCCS Towhee]. http://towhee.sourceforge.net/index.html].

[16] G. M. Kosolapoff, "VAPOUR PRESSURES AND DENSITIES OF SOME LOWER ALKYLPHOSPHONATES," Journal of the Chemical Society, 2964-2965 (1955).

[17] J. K. Badenhoop, and F. Weinhold, "Natural steric analysis of internal rotation barriers," International Journal of Quantum Chemistry, 72(4), 269-280 (1999).

# Understanding the role of physical and chemical adsorption on the Raman enhancement from metallic nanoparticles and nanostructured surfaces

Jason Guicheteau[a], Steven Christesen[a], John Kiser[b], and Jerry Caballo[a]
[a]Edgewood Chemical Biological Center
[b]SMART Student, University of Maryland, Baltimore County

## ABSTRACT

We have demonstrated that it is possible to reproducibly (<20% RSD) fabricate tethered gold nanoparticle substrates. While work still needs to be done to improve intra- and inter-substrate reproducibility, we have also shown that these substrates can be used to analyze molecules that form charge transfer complexes, (e.g., benzenethiol), and molecules that do not have strong charge transfer characteristics (e.g., aminoethanol). Using these substrates, a preliminary analysis of the binding kinetics of aminoethanol to the gold nanoparticle surface was evaluated.

**Keywords:** Surface-enhanced Raman, nanostructures, nanoparticles

## 1. INTRODUCTION

Surface-enhanced Raman spectroscopy (SERS) is a vibrational spectroscopic method that is attractive for analytical applications because it can provide fingerprint spectra of analytes for specific identification. First observed from roughened Ag electrodes, it has since been determined that virtually any noble metal substrates roughened on the nanometer scale or a suspension of nanoparticles can provide enhancement of the Raman signal, though the magnitude of enhancement varies greatly.

The enhancement of the Raman scattering seen in SERS is generally agreed to result from two distinct mechanisms: chemical (CE) and electromagnetic enhancement (EM). EM is the result of an increase in the electromagnetic field of noble metal substrates due to surface plasmons produced by absorption of incident photons. More general than CE, EM accounts for the majority ($>10^4$) of the signal enhancement seen in SERS. Because the intensity of the Raman scattering from an adsorbed analyte molecule is proportional to the square of the local electromagnetic field (EMF), the increase in the magnitude of the EMF at the metal surface increases the number of photons that are Raman scattered. Due to the interaction of the electromagnetic fields that can be created, the largest calculated enhancements are found at the interstitial sites between metal particles.[1,2] The surface plasmon resonance of SERS substrates can be tailored to coincide with the excitation wavelength being used by manipulating the size, shape, and interparticle spacing of the substrate. Clusters of noble metal nanoparticles have been shown and calculated to provide the largest EMF enhancement because of the interaction of surface plasmons from the different nanoparticles. Even though Raman spectroscopy can be performed using virtually any excitation wavelength, the best SERS is accomplished through a coordination of the surface plasmon resonance of the substrate with the excitation wavelength being used.

Unlike EM, CE is related to analyte-specific interactions between the metal and the molecule of interest through formation of a molecule-metal charge-transfer complex that can resonantly enhance the Raman scattering signal from analyte molecules. Sites of atomic scale roughness, such as metal clusters or adatom defects, are thought to act as chemical active sites that facilitate the creation of charge transfer complexes, but the nature and identity of these sites has not been conclusively determined. CE is site and analyte specific and is thought to contribute an enhancement factor of at most $10^2$.

Even without the formation of a strict charge-transfer complex, the analyte molecule must adsorb onto the metal to take advantage of the EM which falls off rapidly with distance. Chemicals that don't adsorb to the metal nanoparticle will not be enhanced. This is why most SERS studies are performed with analytes that bind strongly to the metal particles or nanostructures. The nanoparticle surface charge plays a critical role in the determining the adsorptivity of the analyte as well as the stability of the colloidal suspension. This surface charge on the nanoparticle results in a relative increase in

the concentration of counter-ions close to the surface and the formation of an electrical double layer. The potential difference between the stationary layer of solvent and the mobile layer or slipping plane is referred to as the zeta potential and provides information about the charge on the nanoparticle which in turn affects the stability of the colloid and the binding of the analyte molecule.

The Langmuir isotherm can be used to describe both physical and chemical adsorption and will be used to characterize the binding of the analyte to the substrate:[3]

$$\theta = \frac{Kc}{Kc+1} \qquad (1)$$

In this equation, the analyte concentration is given by c, the fractional surface coverage by θ, and the adsorption equilibrium constant by K. The fractional coverage is calculated as the ratio of the SERS peak intensity divided by the intensity at full surface coverage (I/Imax). K can be determined from experimental data by fitting K and Imax using a nonlinear regression technique as shown in Figure 1a for trans-1,2-Bis(4-pyridyl)ethylene (BPE) on a gold nanostructured SERS substrate. The calculated adsorption equilibrium constant is 1.2 x 107 M-1. While the plot of SERS intensity vs. concentration yields the adsorption equilibrium constant, the rate constant for adsorption can be extracted from the time dependence of the SERS signal via the time-dependent Langmuir equation:[4]

$$\theta = 1 - \exp(-kt) \qquad (2)$$

where k is the rate constant and t the time delay between exposure of the SERS substrate to the analyte solution and data collection. Figure 1b shows the results for BPE adsorbed onto Klarite® (not the same substrate as in Figure 1a) and the fit of the data to Equation 2.

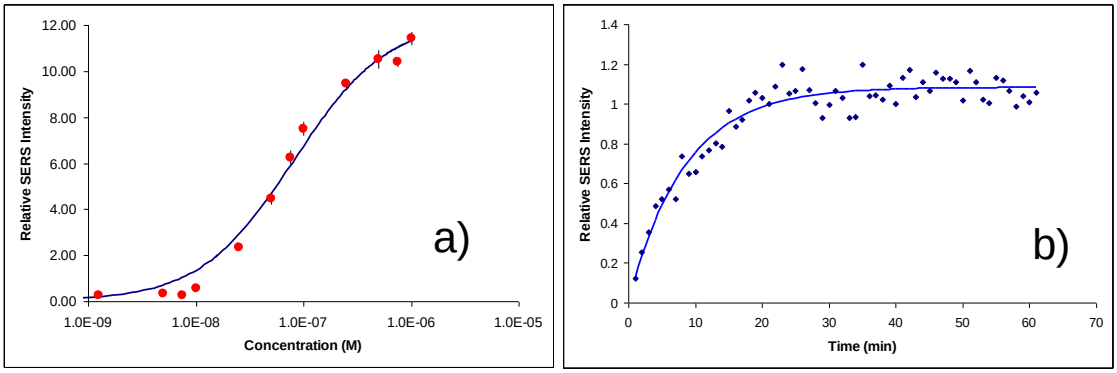



Figure 1: Concentration (a) and time (b) dependence of the SERS signal with fits to Langmuir isotherm and time-dependent Langmuir equation, respectively.

## 2. RESULTS

Our goals this year were to develop a 2-dimensional SERS substrate based on tethering metal colloids to a quartz slide, and using the substrate to evaluate the SERS enhancement obtainable for non-charge transferring analytes such as mercaptoacetic acid. The research this past year can be divided into four phases: a) synthesizing both gold and silver colloids, b) characterizing the UV-Visible absorption properties, the size of the particles, and the zeta potential of the colloids, c) fabricating colloids into the tethered planar substrates, and d) using the substrates to measure the binding kinetics of aminoethanol.

### 2.1 Silver colloids

The initial phase of this work was to fabricate planer substrates by functionalizing quartz slides with a mercaptan functional group and then binding silver nanoparticles to the slide though a "tethering" mechanism. To this end, four batches of silver colloids were synthesized using procedures already established at by our research group. The particle size and zeta potential were used to characterize the preparations. The UV-Visible absorption spectra in Figure 2 clearly show that in three cases the silver colloids have extremely similar absorption maxima and bandwidths. The fourth preparation had significant plating during the synthesis, which accounts for the broad absorption spectrum.

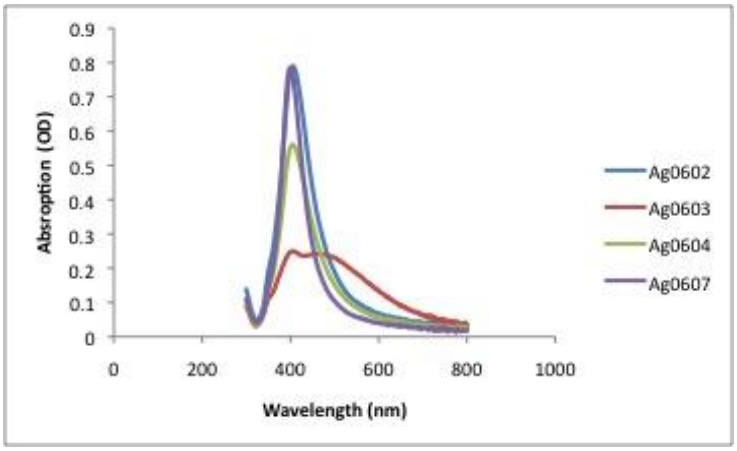



Figure 2: Visible absorption spectra for various preparations of Silver colloid solutions. The silver nanoparticle sizes are 37 nm, 83 nm, 45 nm, and 33 nm for the blue, red, green, and purple spectra, respectively.

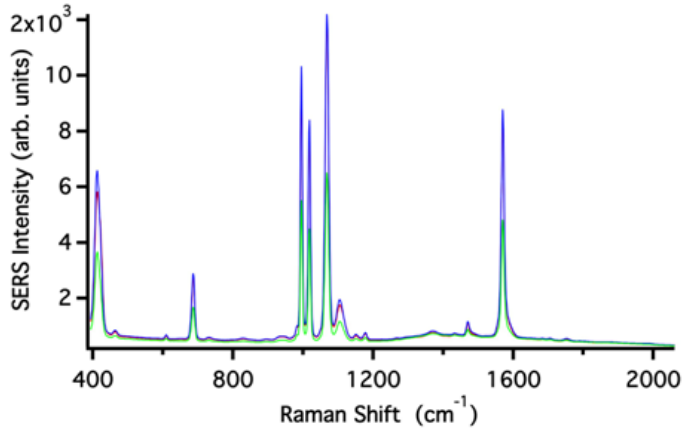



Figure 3: SERS spectra of benzenethiol when analyzed using sliver colloid preparations 0602 (red), 0604 (blue), and 0607 (green).

For the 0602, 0604, and 0607 preparations, the SERS response was evaluated by analyzing 19.6 µM solutions of benzenethiol without forced aggregation by sodium chloride, as shown in Figure 3. The data in Figure 3 demonstrate that all three preparations are SERS active for benzenethol, with preparations 0602 and 0604 being nearly identical and significantly better then preparation 0607. Preparation 0602 was chosen for planar substrate fabrication due to its very slight red shift in the plasmon absorption band.

### 2.2 Tethered silver nanoparticle substrate fabrication

Initial studies focused on determining the optimal ratio of functional group to solvent in the derivativization of the quartz substrates. Piranha (4:1 $H_2S0_4$: 30%$H_2O_2$) cleaned quartz slides were suspended in functionalization solutions prepared at 19:1, 9:1, and 4:1 methanol:mercaptopropyltrimethoxysilane and allowed to soak for 24 hours. The functionalized slides were then rinsed with copious amount of deionized water followed by methanol and allowed to dry. The SERS substrates were then fabricated by placing 1-ml of the silver colloid onto the derivatized slides and allowed to dry. The process was repeated two more times. During this drying process, the slides functionalized with more dilute solutions of mercaptopropyl trimethoxysilane exhibited less aggregation as determined by minimal blacking of the surface.

The analytical results obtained from this procedure were less than ideal. Of the 16 fabricated slides 12 delaminated when placed in ethanolic solutions for analysis. Of the substrates that did not delaminate, little to no response for trans-1,2-bis(4-pyridyl)ethylene or mercaptoacetic acid was observed. Some SERS response was obtained for benzenethiol but because of delaminating issues and poor signals due to silver oxidation, the focus was shifted to the analysis of substrates fabricated with gold nanoparticles.

## 2.3 Synthesis of gold nanoparticles

Gold nanoparticles were synthesized using a standard citrate reduction reaction of auric acid in boiling aqueous solution. In this work the target nanoparticle size was 100 nm in order to push the plasmon absorption band of the colloid solution as far into the red as possible. This should make the substrates more SERS active with the 785 nm excitation wavelength. To achieve this, ~50 ml of water with 25 µmol of auric acid heated until boiling and sodium citrate was then added so the mole ratio of gold:citrate was either 1.8:1 or 1.6:1. Both of these preparations resulted in variety size distributions. Due to the large particle size and large gold to citrate ratio, the solutions were not very reproducible and tended to plate. The UV-visible absorption spectra for typical gold colloid solutions are shown in Figure 4. This figure clearly shows a broad absorption band for all the trials, indicative of a large particle distribution. The preparation 0623p6 exhibited significant plating during the synthesis and was included to demonstrate the effect plating has on the nanoparticle absorption profile. It is interesting to note that changes in the average size of the nanoparticle between 35 nm and 150 nm do not significantly change the absorption maximum of the colloid solutions. This would suggest that substrates fabricated with different batches of Au colloids would have similar SERS responses.

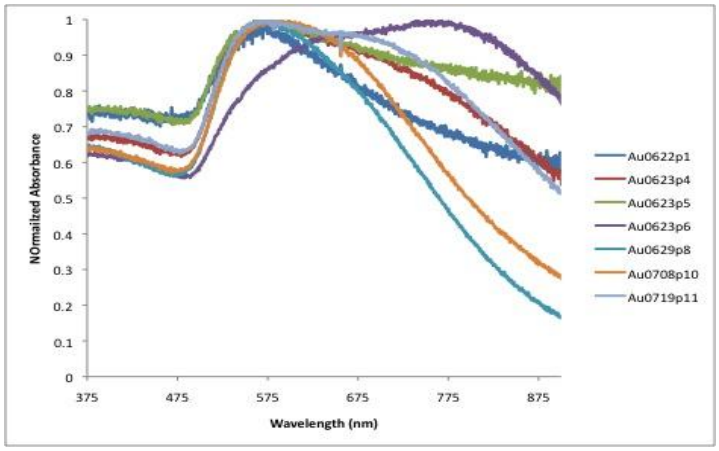



Figure 4: Visible absorption spectra for different solutions of Gold colloid. The size of the gold nanoparticles are 65 nm, 88 nm, 154 nm, and 88 nm, 36.12 nm, 54.6 nm, 89.0 nm; for Au0622p1, Au0623p4, Au0623p5, Au0623p6, Au0629p8, Au0708p10, and Au0719p11, respectively.

## 2.4 Tethered gold nano particle substrate reproducibility

In this study, three batches of gold substrates were fabricated using different colloid solutions. The slides were pre-cleaned as before except that they stayed in the bath overnight. Following cleaning, the slides were suspended in a 4:1 solution of methanol and aminopropyltrimethoxysilane. The quartz slides were then sonicated for 5 hours for batches 1 and 3, and 16 hours for batch 2. The sonication increases the motion of the solution and adds heat to the silanization procedure. After the derivatization, the slides were sonicated twice in 50-ml aliquots of methanol for at least 60 minutes each time to remove any unbound aminopropylsilane. Following the methanol cleaning, the slides were rinsed with copious amounts of water followed by methanol rinsing. The functionalized slides were then coated with three 1-ml volumes of gold colloid in the same manner as the silver slides. This technique produces substrates with reproducible plasmon resonance bands as shown in Figure 5. This figure shows the plasmon absorption bands for batch 1 before and after analysis with benzenethiol. What you will notice is the primary absorption band located at 550 nm is highly reproducible between each slide in the batch. During the washing step gold particle were visually observed in the wash solution. This accounts for the blue shift and narrowing of the plasmon bands measured after the SERS analysis, due to the loss of larger nanoparticle aggregates during the washing.

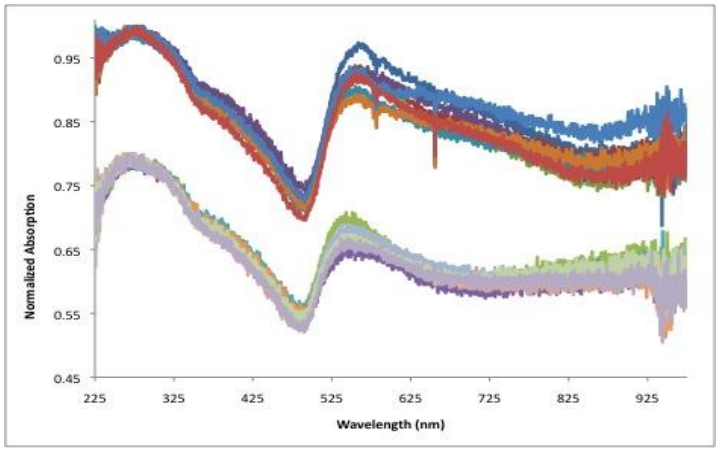



Figure 5: This figure shows the normalized absorption spectra for planer gold substrates pre (top set) and post (bottom set) analysis of benzenethiol. The post analysis spectra were off set (-0.25) for clarity.

### 2.5 SERS analysis of tethered gold nanoparticle substrates

In this study, slides fabricated from three colloid preparations were soaked in 0.1 mM ethanolic benzenethiol overnight, rinsed to remove any unbound gold or benzenethiol, and then allowed to dry. The SERS analysis was performed using a scanning Raman microscope with a 4X objective and 785 nm excitation. A typical transmission image of a gold substrate is shown in the inset of Figure 6. This inset image shows the relative uniformity of the substrate and the location where each of the benzenethiol SERS spectra shown in Figure 6 were measured. The dimensions of the substrate, shown in the inset image, are 14.6 mm by 11.2 mm, which accounts for 25% of the total surface area of the substrate. It is important to note that in the analysis of these substrates, the substrate was place under the 4X microscope objective, adjusted into the focal plane, and the map was determined automatically by the scanning system to insure a random sampling of data points across the surface. The SERS response for each substrate was calculated by averaging the total integrated intensity at 994 and 1016 $cm^{-1}$ for all 81 spectra. Batches of 3, 8, and 6 substrates were evaluated for a total of 17 trials. The relative standard deviation across all the substrates was 17.1 percent and 16.1 percent for the 994 $cm^{-1}$ and 1016 $cm^{-1}$ bands, respectively.

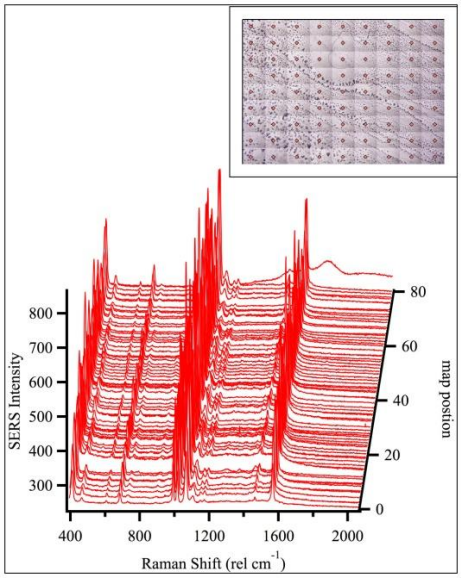



Figure 6: This figure shows the characteristic SERS spectra of Benzenethiol for each of the 81 data point obtained for a typical tethered gold nanoparticle substrate. Inset: A typical transmission image of the tethered gold nanoparticle substrate and the location where each spectrum was measured on the substrate.

## 2.6 Kinetic analysis of ethanolamine binding to tethered gold nanoparticle substrates

In this set of experiments, the binding of ethanolamine to the gold substrates were measured as a function of time by measuring the SERS response. These experiments were conducted using a scanning Raman microscope using a 4X objective and 785 nm excitation. Sample preparation consisted of placing the substrate into a Pyrex Petri dish and covering with 5 ml of aqueous ethanolamine solution.

Prior to the SERS kinetic analysis, normal Raman spectra were obtained using a plain quartz slide as a substrate. This was done to keep the optical path length consistent between the normal Raman measurements and the SERS measurements. The spectral results at concentrations between 1.0 and 10 M ethanolamine in water are shown in Figure 7. The vibrational bands at 865 $cm^{-1}$ and 2871 $cm^{-1}$ were integrated and are displayed as a calibration curves in the inset. The linear regression coefficients are 0.997 and 0.999 for the 865 $cm^{-1}$ and 2871 $cm^{-1}$ bands, respectively. A standard at 0.5M was analyzed, but the bands at both 865 $cm^{-1}$ and 2871 $cm^{-1}$ were not significantly different than the baseline. For this reason, the results were not included in the calibration curves and concentrations of less than 100 mM were used for all the SERS measurements in order to ensure that the signals measured for the kinetics experiments are a result of SERS and not normal Raman.

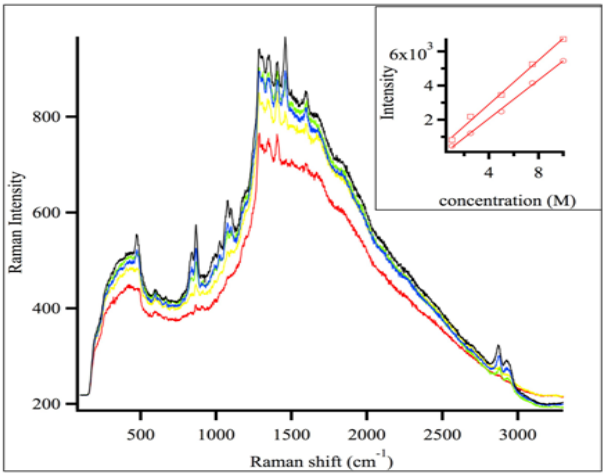



Figure 7: Normal Raman spectrum of ethanolamine at concentrations of 1.0 M (red), 2.5 M (yellow), 5.0 M (green), 7.5 M (blue), and 10.0 M (black). Inset: Calibration curves for the 865 $cm^{-1}$ (squares) and 2871 $cm^{-1}$ (circles) vibrational bands.

The binding kinetics of ethanolamine to the tethered-nanoparticle substrate was evaluated using a combination band that is unique to the SERS spectrum of aqueous ethanolamine. Figure 8 shows the raw SERS spectrum of a 25 mM solution of aqueous ethanolamine 1 minute (red) and 51 minutes (green) after the solution was added to the substrate. When compared with the normal Raman spectra, Figure 7, these spectra clearly show the appearance and growth of a new band at 2015 $cm^{-1}$. This band is believed to be from a Fermi resonance between the amine anti-symmetric deformation at 1592 $cm^{-1}$ and the torsional oscillation at 472 $cm^{-1}$.[9] Justification for this comes from the fact that SERS is a surface phenomenon and the amine interactions with gold will orient the group towards the metal nanoparticle surface. This will account for the enhancement of the amine vibrations. There is also a disappearance of both of these bands in the SERS spectrum. A spectrum of the substrate with a deionized water blank is also shown in Figure 8. This spectrum demonstrates that there are no background bands from the aminopropylsilane used in the fabrication of the SERS substrate occurring in the 1900-2100 $cm^{-1}$ region.

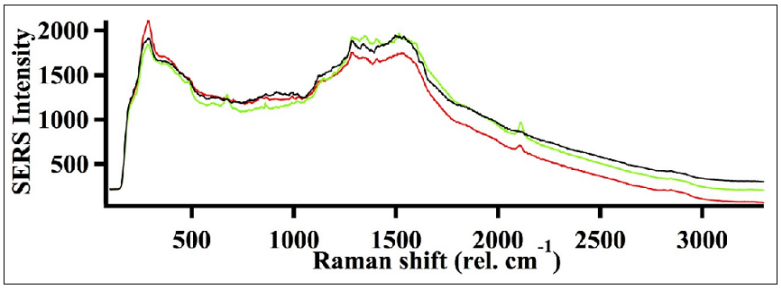



Figure 8: Typical SERS spectra of 25 mM aqueous ethanolamine obtained using a tethered gold-nanoparticle substrate. The red and green traces were taken after 1 min. and 50 min. of exposure to the sample. The black spectrum is of a water blank.

The integrated intensity at 2015 $cm^{-1}$ was used to construct time dependant intensity curves, as shown in Figure 9. The data shown in this figure was fit using a time dependent Langumir isotherm of the form I(t) = Imax(1-exp-kt). This equation was adapted from the work of Karpovich et al.[10] Using this equation, the observed rate constant was determined to be 0.0012 ± 0.0002 $s^{-1}$ at the 95% confidence interval. While this is a preliminary analysis, it will be possible to use this approach to determine the observed rate constant at different concentrations and use this information to determine the rate constants for adsorption and desorption of ethanolamine to the gold surface.

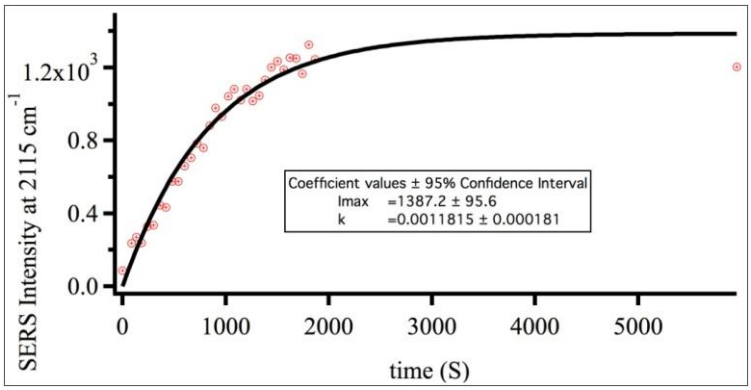



Figure 9: Kinetic analysis of a 50 mM solution ethanolamine fit to a Langmuir isotherm.

### 2.7 Density functional theory

Complementing the experimental research in this project, computational modeling is being used to understand how the interaction between an analyte molecule and the SERS substrate controls signal enhancement. Density functional theory (DFT) has been used to simulate the interaction between selected molecules, such as 2-aminoacetic acid (glycine) and a SERS substrate. The binding geometry of glycine interacting with a single unit cell of silver using the basis set B3LYP/LANL2D2, along with a frequency analysis of the moiety were calculated (Figure 10).

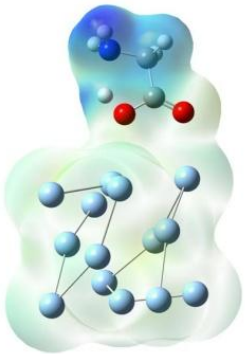


Figure 10: Gaussian calculation of binding geometry of glycine to a silver cell with acid group down.

The correlation between Raman shift enhancement and molecule/substrate orientation, ionic form, substrate nanostructure in the theoretical model, and adsorption energy will continue to be investigated. The relationship between these quantities and Raman shift enhancement is expected to give further insight into the mechanism for SERS enhancement. To simulate the influence of nanostructure, the substrate will be modeled as smooth crystal faces as well as small structures. Hybrid DFT and molecular mechanics methods will be employed to simulate larger substrate models that avoid edge effects. Correlation between situations that lead to SERS enhancement and the electron probability density is expected to also give fundamental insight into the mechanism for SERS enhancement.

## 3. CONCLUSIONS

We have demonstrated that it is possible to reproducibly fabricate tethered gold nanoparticle substrates. This work has also shown that these substrates can be used to analyze molecules that form charge transfer complexes (e.g., benzenethiol) and those that do not have strong charge transfer characteristics (e.g., aminoethanol). Using these substrates, a preliminary analysis of the binding kinetics of aminoethanol to the gold nanoparticle surface was evaluated. We also have performed a preliminary density functional theory calculation to simulate the interaction between the molecule and a SERS substrate.

## REFERENCES


[1] M. Moskovits, Rev. Mod. Phys. 57, 783 (1985).

[2] J.C. Hulteen, D.A. Treichel, M.T. Smith, M.L. Duval, T.R. Jensen, and R.P. Van Duyne, J. Phys. Chem. B 103, 3854 (1999).

[3] A.W. Adamson and A.P. Gast, Physical Chemistry of Surfaces (Wiley Interscience, New York, 1997), 6th ed., p. 599.

[4] Kevin B. Biggs, Jon P. Camden, Jeffrey N. Anker, and Richard P. Van Duyne, J. Phys. Chem. A, Article ASAP 16 March 2009.

[5] G. Levi, J. Pantigny, J.P. Marsault, and J. Aubard, J. Raman Spectrosc. 24, 745-752, (1993).

[6] Ramon A. Alvarez-Puebla and Ricardo F. Aroca, Anal. Chem. 81, 2280-2285 (2009).

[7] G.D. Pirngruber, G. Eder-Mirth, J.A. Lercher, J. Phys. Chem. B 101(4), 561-568 (1997).

[8] R.N. Roy, J.J. Gibbons, G. LaCross, C.W. Krueger, J. of Sol. Chem. 5(5), 333-343 (1976).

[9] Lin-Vien, D., Colthup, N., Fateley, W., and Grasselli, J., [The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules], Academic Press, San Diego, 172-175 (1991).

[10] Karpovich and Blanchard, Langmuir, 10(9), 3315-3322 (1994).

# Discrimination of pathogenic vs. non-pathogenic *Francisella tularensis* and *Burkholderia pseudomallei* using proteomics mass spectrometry

Mary Margaret Wade[a], Rabih Jabbour[b], Alan W. Zulich[a]
[a]ECBC, Research and Techonology Directorate, Aberdeen Proving Ground, MD, USA 21010
[b]Science Applications International Corporation, Gunpowder, MD, USA 21010

## ABSTRACT

Recently the utility of mass spectrometry analysis to characterize and identify biological agents using a proteomic approach has been demonstrated. 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 based on outer membrane proteins (OMPs). OMPs are often associated with virulence in gram-negative pathogens and could prove to be excellent model biomarkers for strain differentiation among bacteria. OMPs were isolated from pathogenic and non-pathogenic strains of *Francisella tularensis* and *Burkholderia pseudomallei* using ultracentrifugation and a sarkosyl extraction method followed by protein digestion and analysis using liquid chromatography-mass spectrometry/mass spectrometry (LC-MS/MS). The experimental MS/MS spectral data of bacterial peptides were searched using the SEQUEST algorithm against a constructed proteome database of microorganisms to determine the identity and number of unique peptides. Results showed that pathogenic strains could be distinguished from non-pathogenic strains for both organisms based on differences in OMP profiles. Virulent *F. tularensis* and *B. pseudomallei* expressed multiple OMPs that were not present in avirulent samples. In conclusion, proteomics mass spectrometry proved useful in discriminating between pathogenic and non-pathogenic strains, and OMPs unique to pathogenic strains could prove to be useful biomarkers for strain differentiation.

**Keywords:** *Francisella tularensis*, *Burkholderia pseudomallei*, outer membrane proteins, proteomics, mass spectrometry

## 1. INTRODUCTION

Recently mass spectrometry analysis has proven useful in the characterization and identification of biological agents using a proteomic approach (1). 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. 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 *Francisella tularensis* and *Burkholderia pseudomallei. F. tularensis*, classified as a Category A pathogen, is an important potential biowarfare agent and the causative agent of tularemia. *B. pseudomallei*, classified as a Category B pathogen, is also considered a potential biowarfare agent. *B. pseudomallei* is the causative agent of melioidosis and infects both humans and animals. Although not endemic to the United States like *F. tularensis*, *B. pseudomallei* is a risk to soldiers serving in areas such as Southeast Asia and Australia, where the disease is known to occur.

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 *F. tularensis* and *B. pseudomallei* strains.

## 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 *F. tularensis* Schu4 and *B. pseudomallei* 1026B. Non-pathogenic strains employed were *F. tularensis* LVS and *B. thailandensis* E264. Working cultures were prepared by streaking cells from cryopreserved stocks onto chocolate agar (CA) for *F. tularensis* strains and nutrient agar (NA) for *Burkholderia* strains, followed by incubation for approximately 18 hours at 37$^{o}$C. 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 brain heart infusion (BHI) supplemented with cysteine (1% final concentration) for *F. tularensis* and 100 ml of nutrient broth (NB) for *Burkholderia*. All cultures were incubated for approximately 18 hours at 37$^{o}$C 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 appropriate agar as described above and incubated for five days at appropriate temperature as listed above 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 the 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 1 hour to pellet proteins. The pellet was resuspended in 1 ml of HEPES buffer, 1 ml of a 2% sarkosyl solution (N-Lauroylsarcosine sodium salt solution) was added, and sample was incubated at room temperature for 30 minutes. Next samples were centrifuged at 100,000 x g for 1 hour and the pellet containing OMPs was resuspended in 1 ml of HEPES buffer and then transferred to the 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 3 μg/μl Dithiotheritol (DTT) to the filter, and then incubating overnight at 37°C on an orbital shaker set to 60 rpm. 20 μL of 100% of acetonitrile (ACN) was added to the tubes and allowed to 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 150 μL of 100 mM 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. 60 μL 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 engine

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; accessed April 2009) 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; accessed April 2009). 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 for 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 of 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. 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.

| 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_UTI8 |
| 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\|S |
| 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_TW |
| 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\|S |
| 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\|S |
| 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_UTI8 |
| 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_W3110 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 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*. 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.

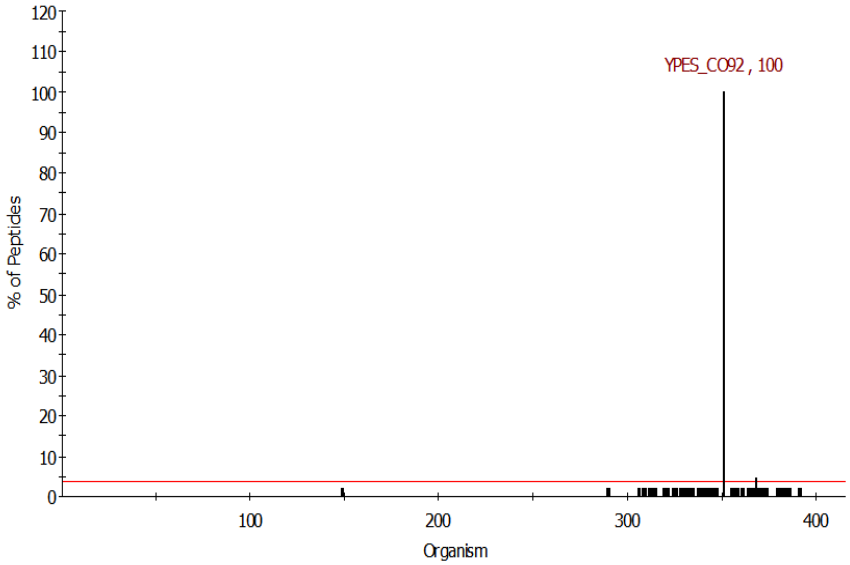



Figure 2. Histogram representing the output of the binary matrix of the unique peptides identified for a given bacterium at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non significant.

### 3.1 Differentiation of pathogenic vs. non-pathogenic *F. tularensis* strains using whole cell lysates and OMPs

Pathogenic and non-pathogenic *F. tularensis* Schu S4 and LVS, respectively, were analyzed by proteomic mass spectrometry for identification using OMPs and whole cell lysates. Figures 3 through 6 show 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 the figures, correct identification of each strain was achieved.

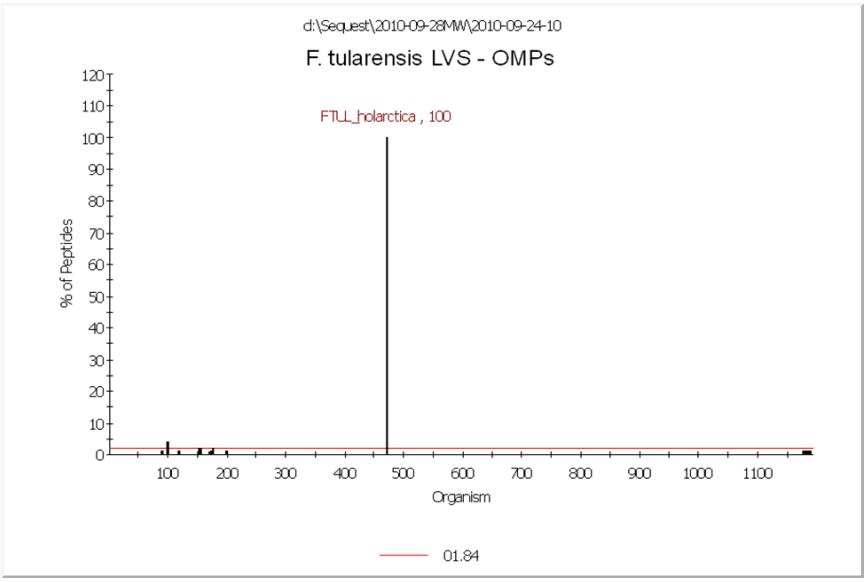



Figure 3. Bacterial differentiation of *F. tularensis* LVS using OMPs. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

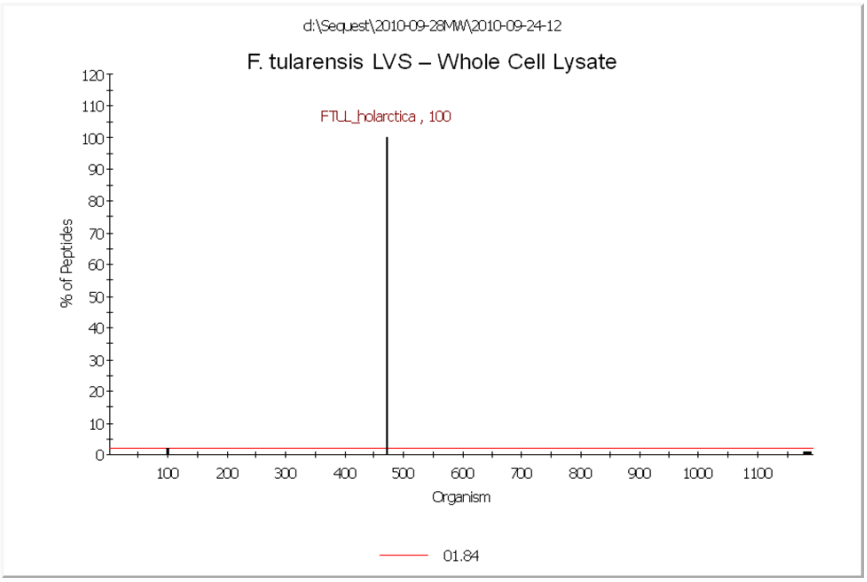



Figure 4. Bacterial differentiation of *F. tularensis* LVS using whole cell lysates. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

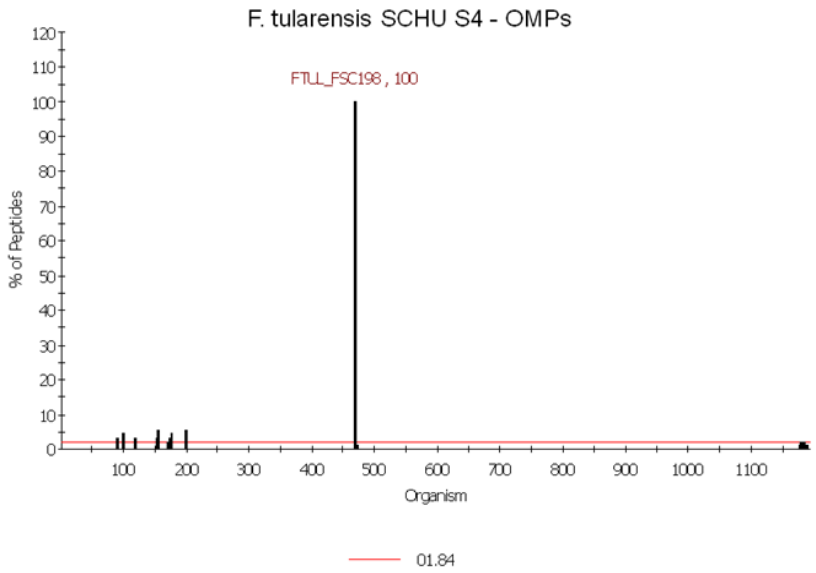



Figure 5. Bacterial differentiation of *F. tularensis* Schu S4 using OMPs. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

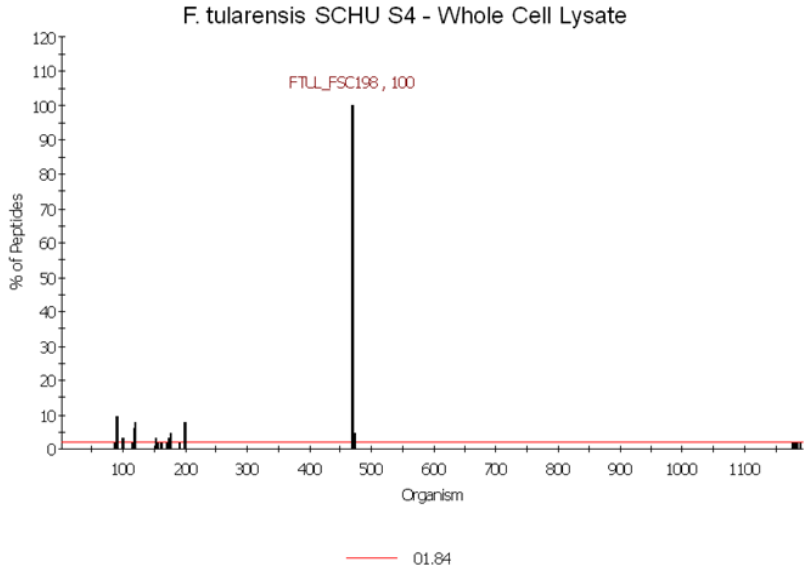



Figure 6. Bacterial differentiation of *F. tularensis* Schu S4 using whole cell lysates. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

OMPs of pathogenic and non-pathogenic *F. tularensis* Schu S4 and LVS, 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 *F. tularensis* strains. Therefore, correct identification to the strain level was achieved for both bacterial OMPs (Figures 7 and 8).

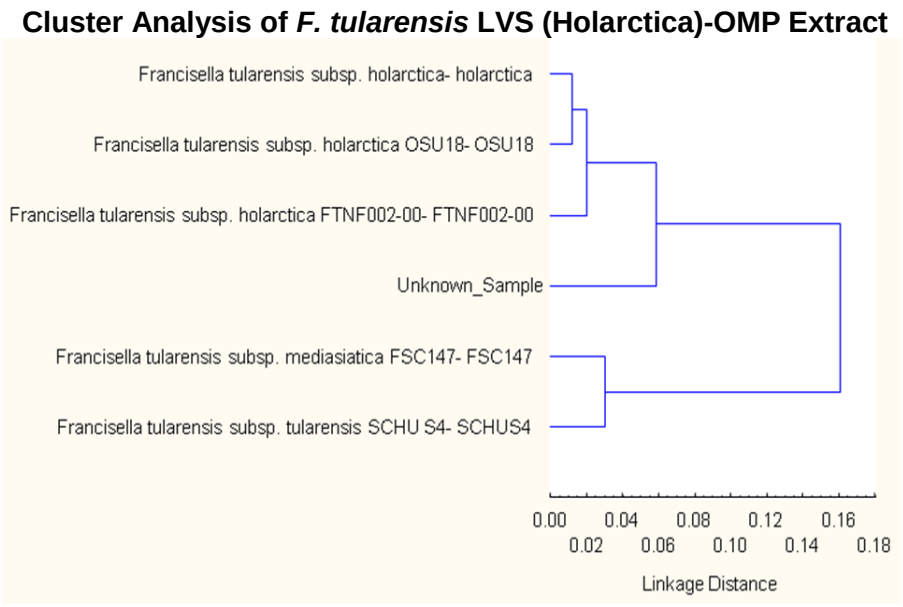



Figure 7. Near-neighbor classification of non-pathogenic *F. tularensis* LVS using OMPs.

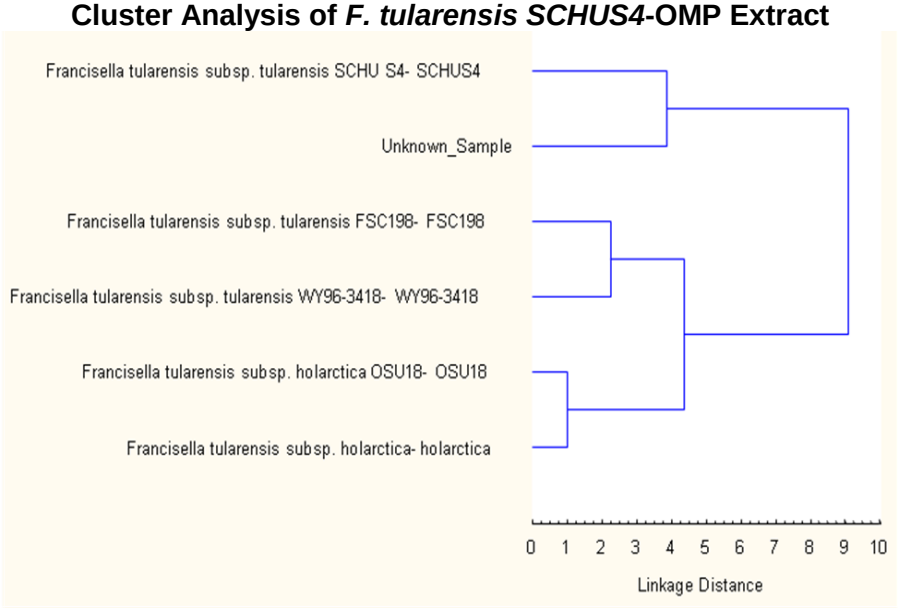



Figure 8. Near-neighbor classification of pathogenic *F. tularensis* Schu S4 using OMPs.

Table 1 shows the comparison of the strain unique proteins for *F. tularensis* Schu S4 and LVS identified from the different cellular extracts. The number of unique proteins identified differed between the pathogenic and non-pathogenic *F. tularensis* strains, with the pathogenic strain having a relatively lower number of unique proteins than that of the non-pathogenic *F. tularensis*. 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 from OMPs versus whole cell lysate for both pathogenic versus non-pathogenic strains, and non-pathogenic strain LVS had a slightly higher number of unique proteins from OMPs.

Table 1. Unique Proteins for *F. tularensis* Schu S4 and *F. tularensis* LVS from Whole Cell Lysates vs. OMPs

| ***F. tularensis* Schu S4 Unique Proteins** | |
|---|---|
| **Whole Cell Lysates** | **OMPs** |
| • Type IV pili fiber building block protein<br>• Hypothetical protein FTH_1232<br>• Outer membrane protein<br>• Hypothetical protein FTT_1334c<br>• Chaperonin GroEL | • Type IV pili fiber building block protein<br>• Molecular chaperone DnaK<br>• 3-oxoacyl-[acyl-carrier-protein] synthase II<br>• Lipoprotein<br>• Chaperonin GroEL<br>• Heat shock protein 90<br>• Outer membrane protein |
| ***F. tularensis* LVS Unique Proteins** | |
| **Whole Cell Lysates** | **OMPs** |
| • Chaperonin GroEL<br>• Hypothetical protein FTL_0617<br>• AhpC/TSA family protein<br>• Heat shock protein DnaK<br>• Hypothetical protein FTL_0572<br>• Hypothetical protein FTL_1110<br>• Hypothetical protein FTL_1494<br>• Macrophage infectivity potentiator, fragment<br>• Outer membrane protein<br>• Co-chaperonin GroES | • Outer membrane associated protein<br>• Chaperonin GroEL<br>• Hypothetical protein FTL_0617<br>• Succinyl-CoA synthetase subunit beta<br>• Intracellular growth locus, subunit C<br>• Sigma-54 modulation protein<br>• Outer membrane protein<br>• AhpC/TSA family protein<br>• heat shock protein DnaK<br>• 3-oxoacyl-[acyl-carrier-protein] synthase II<br>• Lipoprotein<br>• Chaperonin GroEL<br>• Co-chaperonin GroES<br>• DNA-directed RNA polymerase subunit beta<br>• Intracellular growth locus, subunit C |

### 3.2 Differentiation of pathogenic vs. non-pathogenic *B. pseudomallei* strains using whole cell lysates and OMPs

Pathogenic *B. pseudomallei* 1026B and non-pathogenic *B. thailandensis* E264 were analyzed by proteomic mass spectrometry for identification using OMPs and whole cell lysates. Figures 9 through 12 show 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 the figures, correct identification of each strain was achieved.

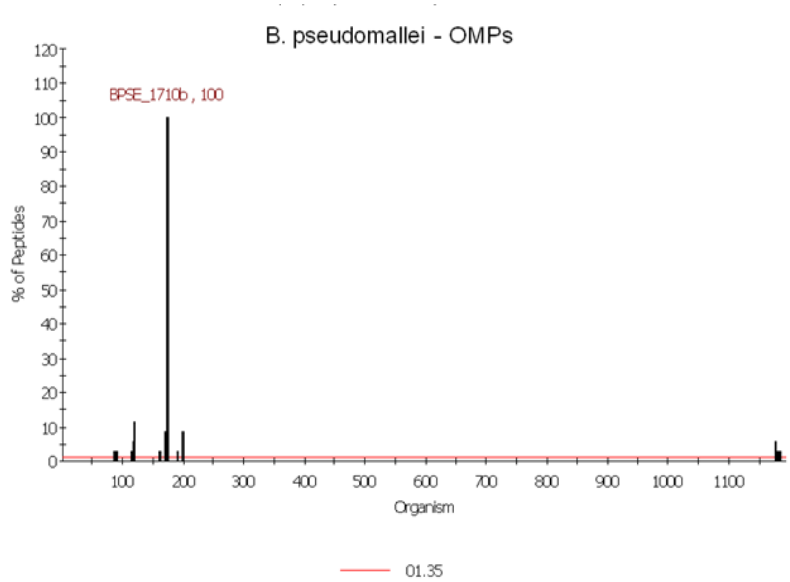



Figure 9. Bacterial differentiation of *B. pseudomallei* 1026B using OMPs. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

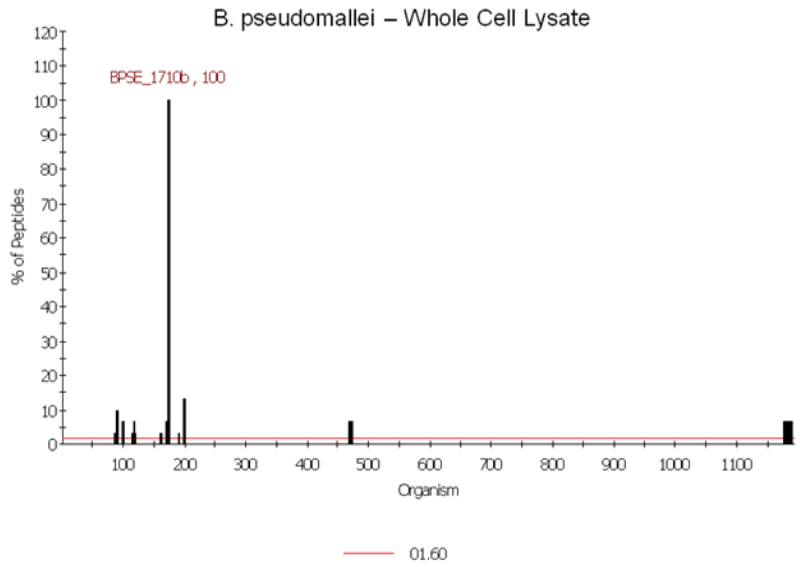



Figure 10. Bacterial differentiation of *B. pseudomallei* 1026B using whole cell lysates. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

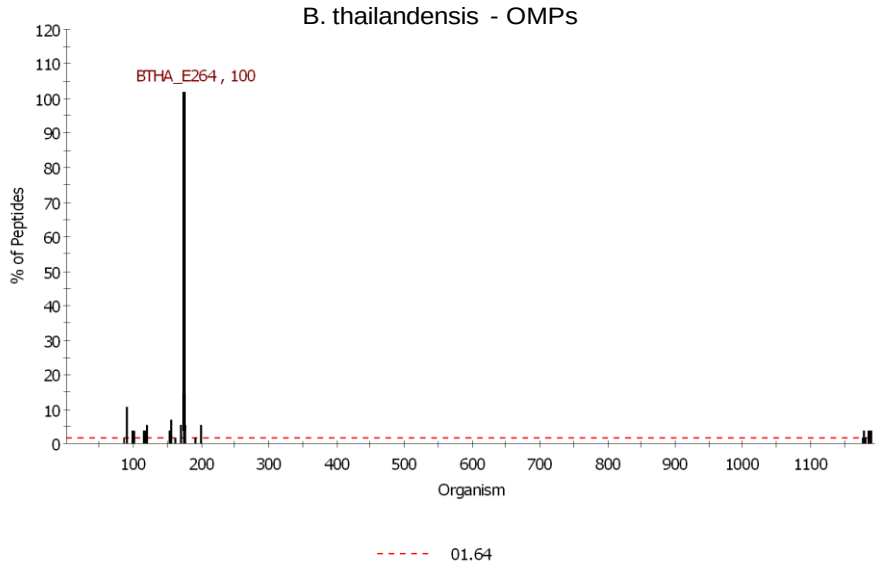



Figure 11. Bacterial differentiation of *B. thailandensis* E264 using OMPs. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

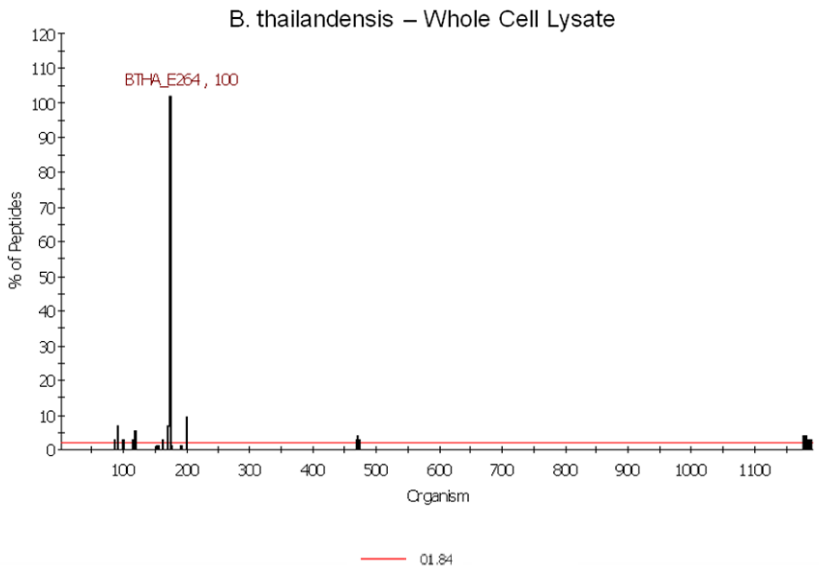



Figure 12. Bacterial differentiation of *B. thailandensis* E264 using whole cell lysates. X-axis represents bacterium proteome and Y-axis represents number of unique peptides at 95% confidence level. The horizontal line is the threshold under which peptides identified are considered statistically non-significant.

The near neighbor analysis, using Euclidean distance linkage approach, for OMPs of pathogenic *B. pseudomallei* 1026B showed that the identified unique set of proteins had the closest match with the employed *B. pseudomallei* strain. Therefore, correct identification to the strain level was achieved (Figure 13).

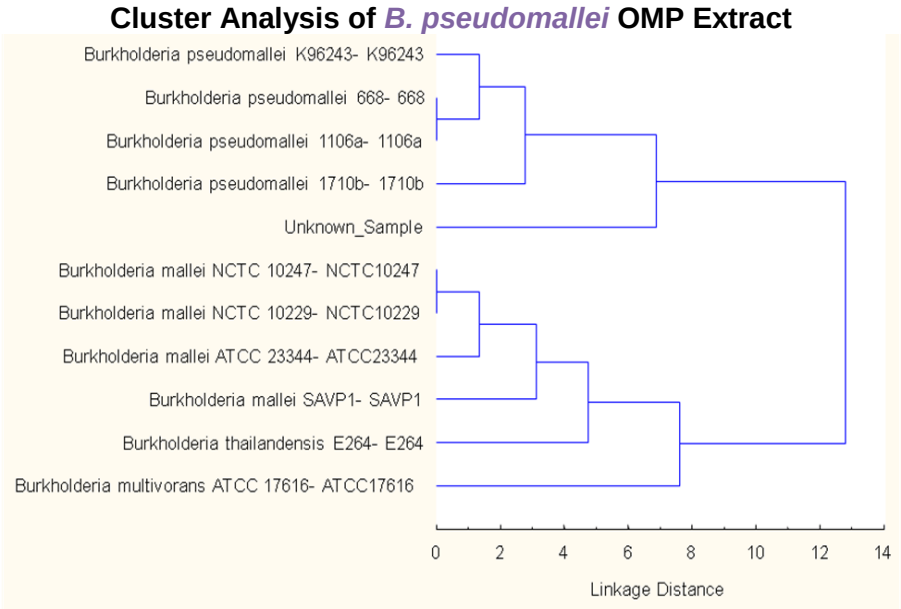



Figure 13. Near-neighbor classification of pathogenic *B. pseudomallei* 1026B using OMPs.

Table 2 shows the comparison of the strain unique proteins for *B. pseudomallei* and *B. thailandensis* identified from the different cellular extracts. The number of unique proteins identified differed between the pathogenic and non-pathogenic strains, with the pathogenic strain having a relatively lower number of unique proteins than that of non-pathogenic *B. thailandensis*. 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 higher for non-pathogenic *B. thailandensis* from OMPs extracts than whole cell lysate.

Table 2. Unique Proteins for *B. pseudomallei* and *B. thailandensis* from Whole Cell Lysate vs. OMPs

| ***B. pseudomallei* 1026 B Unique Proteins** | |
|---|---|
| **Whole Cell Lysates** | **OMPs** |
| • Pyochelin synthetase<br>• Hypothetical protein BPSL2599<br>• Putative cytochrome C precursor-related protein | • ABC transport system, ATP-binding protein<br>• Hypothetical protein BPSS2201<br>• Glutamate/aspartate periplasmic binding protein precursor<br>• Hypothetical protein BURPS1710b_2354<br>• Hypothetical protein BURPS1710b_2705 |
| ***B. thailandensis* E264 Unique Proteins** | |
| **Whole Cell Lysates** | **OMPs** |
| • Putative lipoprotein<br>• Glucan 1,4- a-glucosidase<br>• Glycosy hydrolase family protein<br>• Putative lipoprotein<br>• Glutamate/aspartate ABC transporter, periplasmic glutamate/aspartate-binding protein<br>• Putative lipoprotein<br>• RND efflux system outer membrane lipoprotein<br>• Putative lipoprotein<br>• Sensor histidine kinase, putative<br>• Putative lipoprotein<br>• Chaperonin GroEL | • Stress response protein<br>• Serine protease, MucD<br>• Putative lipoprotein<br>• Hypothetical protein BTH_I0089<br>• Glutamate/aspartate ABC transporter, periplasmic glutamate/aspartate-binding protein<br>• Hypothetical protein BTH_II2182<br>• Putative lipoprotein<br>• Gamma-glutamyltransferase<br>• Putative lipoprotein<br>• Transport-associated domain-containing protein<br>• Chaperonin GroEL<br>• Amino acid ABC transporter, periplasmic amino acid-binding protein, putative<br>• Dihydrolipoamide succinyltransferase<br>• Hypothetical protein BTH_II2183<br>• Hypothetical protein BTH_I0090<br>• Hypothetical protein BTH_I2265<br>• Hypothetical protein BTH_I1688<br>• NADPH-dependent FMN reductase domain-containing protein<br>• OmpA family protein<br>• Amino acid ABC transporter, periplasmic amino acid-binding protein<br>• Translation initiation factor IF-2<br>• Ribonuclease E<br>• Alcohol dehydrogenase, zinc-containing |

## 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. Although OMPs offered enhanced differentiation capability, whole cell lysates did provide discrimination and are an appropriate option for the differentiation of gram-positive bacterial strains. Overall, an extension of this project to include a wider investigation of other relevant pathogenic bacteria 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 2010 U. S. Army Edgewood Chemical Biological Center's In-house Laboratory Independent Research program.

## RESOURCES

[1] **Dworzanski, J.P., S.V. Deshpande, R. Chen, R.E. Jabbour, A.P. Snyder, C.H. Wick, and L. Li.** 2006. Mass spectrometry-based proteomics combined with bioinformatic tools for bacterial classification. J. Proteome Res. **5:** 76.

# Ricin toxicity in BALB/c 3T3 cells: peptide biomarkers of exposure

Vicky L. H. Bevilacqua[a], Janna S. Madren-Whalley[a], Samir Deshpande[b],
Rabih E. Jabbour[c], Kyle Jones[d], Lisa M. Reilly[e], Jeffrey S. Rice[f]
[a]ECBC, Reseach and Technology Directorate, Aberdeen Proving Ground, MD, USA 21010
[b]Science and Technology Corporation, 500 Edgewood Road, Edgewood, MD, USA, 21040
[c]Science Applications International Corporation, P.O. Box 68, Gunpowder, MD, USA, 21010
[d]NuSep, 340 Commerce Blvd, Bogart, GA, USA, 30622
[e]Bethany College, 219 Richardson Hall, Dept of Physical Science, Bethany, WV, USA 26032
[f]Elona Biotechnologies Inc, 1040 Sierra Drive, Greenwood, IN, USA, 46143

## ABSTRACT

This report is required for the U.S. Army Edgewood Chemical Biological Center In-House Laboratory Independent Research project "Cell Toxicity by NMR and Mass Spectrometry." Nuclear magnetic resonance (NMR) and/or mass spectrometry (MS) in combination with chemometrics analysis can yield physiological details in the form of metabolome or proteome biomarker information. We have analyzed trypsinized cellular protein extracts from BALB/c 3T3 murine fibroblasts dosed with vehicle control and with three ricin concentrations. At the time of sample collection (48 hours post dosing), only attached cells were harvested. Adjustment was made for cell concentration in preparing samples for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Analysis of the LC-MS/MS data to determine peptide sequences, followed by statistical filtering, has shown that at least one 3T3 peptide biomarker exists for each biological cell group (vehicle control and three ricin dose levels). These peptides correspond to particular proteins expressed by the 3T3 cells. Identification of the relevant proteins, followed by metabolic pathway analysis to determine the pathways in which those proteins are involved, may reveal the survival mechanism for cells dosed at low levels, which in turn will provide insight into the related 3T3 physiological changes required to overcome ricin dosing.

**Keywords:** 3T3 murine fibroblasts, cell toxicity, liquid chromatography, mass spectrometry, LC-MS

## 1. INTRODUCTION

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. Nuclear magnetic resonance (NMR) and/or mass spectrometry (MS) in combination with chemometrics analysis can yield physiological details in the form of metabolome or proteome biomarker information. We hypothesize that protein biomarkers in the form of dose dependent differences in protein expression occur upon ricin exposure, and that these can be identified using liquid-chromatography tandem mass spectrometry (LC-MS/MS) based proteomics methods. Here, we report on a study to explore this hypothesis for LC-MS/MS with ricin dosed BALB/c 3T3 murine fibroblasts.

## 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 h 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 0-4ºC. The dialyzed ricin concentration

was determined at 25ºC by ultraviolet (UV) absorbance using 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 spectrophotometer (NanoDrop® Technologies, Inc., Thermo Fisher Scientific, Waltham, MA). For the JASCO method, 15 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 $E^{0.1\%}_{280nm} = 1.4$ [1]. 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 $E^{0.1\%}_{280nm} = 1.4$ [1]. 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/c 3T3 murine fibroblasts

The fibroblasts (CCL-163 American Type Culture Collection [ATCC], Manassas, VA) were passaged a minimum of three times after thawing prior to testing. Cell culture flasks (75 $cm^2$) were seeded at 8 x $10^3$ cells/$cm^2$ and maintained in culture at 37ºC in a humidified atmosphere of 5% $CO_2$ in air for 24 h prior to treatment. Exposures were performed 24 ± 2 h 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 (NRU) assay [2,3]. 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 or without ricin. We will refer to the four groups of cells treated identically (VC and 3 ricin dose levels) as biological cell groups. After ricin exposure, the flasks were incubated at 37ºC in a humidified atmosphere of 5% $CO_2$ in air for another 48 h.

Cell harvesting was carried out 48 hours ± 0.5 hours 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 37ºC Routine Culture Medium (RCM) containing serum. The cells were removed from the flasks and centrifuged for 5 min at 1200 rpm. The supernatant was then discarded, and the pellet was resuspended in 4 ml of 37ºC PBS. A 700-µL sample was removed for cell counts and viability analysis employing a ViCell-XR (Beckman Coulter, Brea, CA). The remaining cells were centrifuged for 5 min at 1200 rpm. The supernatant was discarded. The cell pellet was resuspended in 1 ml of 37ºC PBS and then frozen at -80ºC.

### 2.3 Liquid-chromatography/mass spectrometry sample preparation

The cell samples were thawed and lysed by ultrasonication (25 sec on, 5 sec off, 4 min total) using a Branson Digital Sonifier® (Danbury, CT). The lysate was centrifuged at 14,000 rpm for 20 min at 10ºC using a Beckman GS-15R centrifuge and F2402H Rotor. The supernatant was transferred to a Microcon® YM-3 filter unit (Millipore, Billerica, MA) and centrifuged at 14,000 rpm and 10ºC (500 µL for 30 min x 2, total volume of 1.00 ml). The filtrate was stored at -20ºC for other analyses. The cellular proteins in the retentate were denatured overnight at 37ºC with 300 µL of 7.2 M urea and 3 µg/ml dithiothreitol in 100 mM ammonium bicarbonate (ABC). The urea was removed by centrifugation (14,000 rpm, 30 min, RT) and the retentate was 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) 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 were digested at 37ºC for 7 hours with 5 µL sequencing grade trypsin (Product # 511A, Promega, Madison, WI) in 10 µL acetonitrile and 235 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 -20ºC until analysis.

### 2.4 Liquid-chromatography/mass spectrometry experiments

A dilution scheme was prepared for the tryptic peptide samples based on the total cell counts in the 700-µL samples (see section 2.2), and on a final desired total volume of 40 µL for each MS sample. A dilution factor was calculated for each tryptic peptide sample by dividing the total cell count for that sample into the total cell count for the sample having to the highest total cell count. The diluent was a solution of 95% $H_2O$ and 5% acetonitrile.

Of the 40 µL dilution volume prepared for each MS sample, 27 µL was expected to be injected. However, the instrument could only accommodate a 25 µL maximum injection volume. Thus, 25 µL of each MS sample was injected into a Thermo Electron Corporation Finnigan Surveyor HPLC (Thermo Scientific, Waltham, MA), and the peptides were separated using a 0.1 x 150 mm $C_{18}$ Hypersil GOLD KAPPA column (5 Å particle size, 175 Å pore size, Fisher Scientific International, Pittsburgh, PA) with a linear gradient of 20-80% organic phase (100% acetonitrile, 0.1% formic acid). The column was connected to a Finnigan LTQ tandem ion trap MS 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.

The MS samples were analyzed in a semi-random order, with six groups back-to-back and each group containing four samples (one from each ricin dose level). In addition, at least three different replicate sets were represented per group. For example, the third group included peptide samples ran in the following order: $IC_{50}$, replicate set 3; $IC_{20}$, replicate set 5; VC replicate set 2; $IC_{80}$, replicate set 2. The mass spectrometer operator was "blind" to sample identities.

### 2.5 Proteomics analysis

A 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, accessed July 27, 2010). For this task, an in-house PERL (http://www.activatestate.com/Products/ActivePerl, accessed July 27, 2010) program was used to automatically download annotated human, bovine, rat, and mouse 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 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 raw results obtained from SEQUEST were used as input into ProteoIQ (NuSep, Bogart, GA) for label-free statistical filtering using the spectral count method [4] to identify those peptides having sequences corresponding to amino acid stretches in known proteins. The data was analyzed using the following parameters: 1) peptide length = 5 amino acids, 2) minimum Xcorr = 1, 3) minimum peptide probability = 0.05, and 4) minimum protein probability = 0.5. No charged state filter was applied (all charged states used), and no decoy data base was applied.

## 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 peptide sample is shown in Figure 1.

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

| Biological Cell Group | Ricin Concentration (µg/ml) | # of Replicate Samples | Average Total Cell Count in 700 µL Suspension |
|---|---|---|---|
| Vehicle Control Cells | ---0--- | 6 | $1.37 \times 10^{6}$ |
| Cells dosed w/ ricin at $IC_{20}$ | $9.54 \times 10^{-04}$ | 6 | $0.78 \times 10^{6}$ |
| Cells dosed w/ ricin at $IC_{50}$ | $2.15 \times 10^{-03}$ | 6 | $0.41 \times 10^{6}$ |
| Cells dosed w/ ricin at $IC_{80}$ | $5.97 \times 10^{-03}$ | 6 | $0.15 \times 10^{6}$ |
| **Total** | ---- | 24 | |

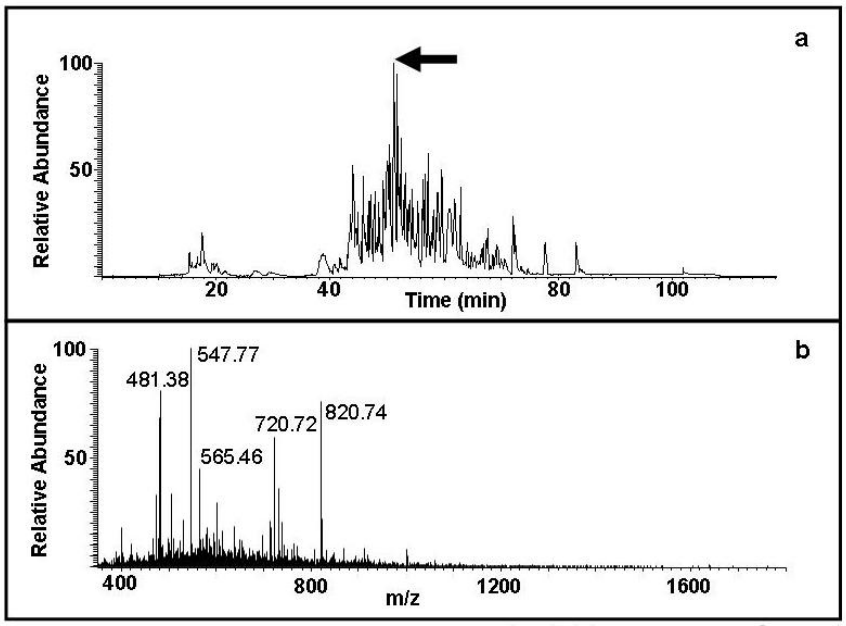



Figure 1. Liquid Chromatography-Mass Spectrometry Data from Extracted Cellular Proteins. a) Total ion chromatogram (where the arrow marks the peak having a 51-min retention time). b) Full MS spectrum for LC peak with 51-min retention time.

Statistical filtering of the MS data (SEQUEST analysis followed by ProteoIQ analysis) identifies amino acid sequences ("hits") that correspond to observed m/z values for the peptides in a sample and provides a score and probability for each peptide. The score, calculated by SEQUEST, indicates how well the observed spectrum matches the theoretical spectrum for a given peptide sequence. The probability, calculated by ProteoIQ, indicates the likelihood that the peptide sequence is correctly determined based on the observed mass data. Using a minimum peptide probability of 0.05, more than 14,000 peptides total were identified for the tryptic peptide samples from the biological cell groups (vehicle control and three ricin dose levels). Table 2 contains the results for some representative peptide sequences.

Table 2. Statistical Filtering Results for Representative Peptides

| Experimental Relative Molecular Mass | Calculated Relative Molecular Mass | Charge[a] | Score[b] | Probability | Peptide Sequence | Total Intensity | MS Scan # | Biological Cell Group[c] | Replicate |
|---|---|---|---|---|---|---|---|---|---|
| 1464.541 | 1462.815 | 2 | 4.78 | 0.95 | IYVDDGLISLQVK | 57490.9 | 4254.4254_2 | VC | rep1 |
| 1567.444 | 1566.780 | 2 | 4.01 | 1.00 | ITPSYVAFTPEGER | 54813 | 3518.3518_2 | IC50 | rep1 |
| 1516.900 | 1515.749 | 3 | 3.94 | 0.22 | IWHHTFYNELR | 627427.4 | 3384.3384_3 | VC | rep5 |
| 1321.153 | 1319.703 | 2 | 3.92 | 0.90 | IMNTFSVVPSPK | 32203.8 | 3671.3671_2 | IC20 | rep2 |
| 1171.348 | 1169.714 | 2 | 3.90 | 0.93 | ILLAELEQLK | 167745.2 | 4186.4186_2 | IC20 | rep4 |
| 1244.451 | 1243.653 | 2 | 3.65 | 1.00 | IQLVEEELDR | 212888.5 | 3517.3517_2 | IC80 | rep5 |
| 1269.171 | 1267.672 | 3 | 3.39 | 0.11 | ISMPDVDLHLK | 104355.8 | 3936.3936_3 | VC | rep3 |
| 1259.407 | 1258.671 | 2 | 3.32 | 0.95 | ISMPDIDLNLK | 132932 | 4171.4171_2 | IC20 | rep6 |
| 1150.122 | 1148.579 | 2 | 3.08 | 0.76 | ITESEEVVSR | 18406.2 | 1985.1985_2 | IC80 | rep6 |
| 1275.388 | 1274.666 | 2 | 3.06 | 0.83 | ISMPDIDLNLK | 66842.3 | 3692.3692_2 | IC20 | rep5 |
| 1032.269 | 1031.498 | 2 | 3.01 | 1.00 | IYFMAGSSR | 15457.5 | 3104.3104_2 | IC20 | rep2 |
| 1198.183 | 1197.582 | 2 | 2.94 | 1.00 | ITLDNAYMEK | 102791.8 | 3230.3230_2 | IC50 | rep2 |
| 1317.295 | 1315.672 | 3 | 2.93 | 0.05 | ISMPDFDLHLK | 14825.9 | 4176.4176_3 | IC80 | rep4 |
| 1138.436 | 1136.599 | 2 | 2.89 | 0.69 | IYGISFPDPK | 51655 | 3916.3916_2 | IC80 | rep4 |
| 1214.381 | 1213.577 | 2 | 2.87 | 0.99 | ITLDNAYMEK | 6298.7 | 2919.2919_2 | IC80 | rep3 |
| 1391.869 | 1390.794 | 2 | 2.86 | 0.33 | IPYDAKTIQTIK | 19170.9 | 3374.3374_2 | IC80 | rep4 |
| 1284.264 | 1283.666 | 2 | 2.86 | 0.99 | ISMPDVDLHLK | 27798.5 | 3342.3342_2 | IC50 | rep1 |
| 948.442 | 947.498 | 2 | 2.79 | 0.97 | ISQLEMAR | 117440.6 | 2554.2554_2 | IC80 | rep4 |
| 1070.929 | 1069.571 | 3 | 2.78 | 0.06 | ISMPDLHLK | 63578.6 | 3018.3018_3 | IC20 | rep3 |
| 1541.671 | 1540.840 | 3 | 2.76 | 0.14 | ISMPDIDLNLKGPK | 25018.1 | 4019.4019_3 | IC50 | rep3 |
| 1157.456 | 1155.565 | 2 | 2.66 | 0.33 | IYNMEMARK | 31892.3 | 3658.3658_2 | IC80 | rep1 |
| 1075.286 | 1074.554 | 2 | 2.53 | 0.98 | ITITNDQNR | 35892.6 | 1681.1681_2 | VC | rep6 |
| 1074.050 | 1073.584 | 2 | 2.51 | 0.84 | IQEAGTEVVK | 58600.9 | 2048.2048_2 | VC | rep1 |
| 1088.240 | 1086.631 | 2 | 2.47 | 0.50 | ITQVLHFTK | 6852.2 | 3078.3078_2 | VC | rep2 |
| 1259.741 | 1258.671 | 2 | 2.44 | 0.39 | ISMPEVDLNLK | 7492.1 | 4101.4101_2 | IC20 | rep1 |
| 1018.244 | 1017.569 | 2 | 2.43 | 0.99 | ITITNDKGR | 12178.3 | 1560.1560_2 | VC | rep3 |
| 1284.715 | 1285.628 | 3 | 2.38 | 0.40 | ISMPDVDLHMK | 10488.7 | 5664.5664_3 | VC | rep3 |
| 1293.169 | 1292.656 | 2 | 2.35 | 0.99 | ISMPDFDLNLK | 8644.8 | 4400.4400_2 | VC | rep3 |
| 964.766 | 963.493 | 2 | 2.26 | 0.30 | ISQLEMAR | 3243.8 | 1692.1692_2 | IC20 | rep1 |

[a]The score indicates how well the observed spectrum matches the theoretical spectrum for a given peptide sequence.
[b]The probability indicates the likelihood that the peptide sequence is correctly determined based on the observed mass data.
[c]VC = vehicle control.

Further filtering was carried out with a minimum peptide probability of 0.95, resulting in ~3600 hits. The ~3600 hits represent 968 different peptide sequences that were identified in one or more spectra from the 24 samples analyzed. Eleven percent of the peptides were observed at all dose levels, and we focused on these peptides for this report. These peptides are expected to be mapped to proteins expressed by all biological cell groups. Table 3 contains a spectral count summary for the peptides having a total spectral count of 24 or higher. The total spectral count value represents the total number of times the peptide was observed for all samples. For a given biological cell group, a peptide having a spectral count considerably higher or lower for that biological cell group as compared to other biological cell groups would represent a biomarker peptide.

Table 3. Spectral Count Summary for Peptides with Total Spectral Count of 24 or Higher

| Peptide Sequence | Total Spectral Count | Spectral Count | | | | Spectral Count % of Total | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | VC | $IC_{20}$ | $IC_{50}$ | $IC_{80}$ | VC | $IC_{20}$ | $IC_{50}$ | $IC_{80}$ |
| LVNELTEFAK | 116 | 11 | 19 | 33 | 53 | 9 | 16 | 28 | 46 |
| AGFAGDDAPR | 115 | 43 | 42 | 10 | 20 | 37 | 37 | 9 | 17 |
| YLYEIAR | 89 | 13 | 17 | 20 | 39 | 15 | 19 | 22 | 44 |
| HLVDEPQNLIK | 62 | 9 | 12 | 16 | 25 | 15 | 19 | 26 | 40 |
| YALYDATYETK | 44 | 9 | 12 | 11 | 12 | 20 | 27 | 25 | 27 |
| AEGPEVDVNLPK | 43 | 9 | 14 | 10 | 10 | 21 | 33 | 23 | 23 |
| QEYDESGPSIVHR | 42 | 13 | 14 | 9 | 6 | 31 | 33 | 21 | 14 |
| IQLVEEELDR | 39 | 7 | 12 | 9 | 11 | 18 | 31 | 23 | 28 |
| TLINILTER | 39 | 8 | 12 | 11 | 8 | 21 | 31 | 28 | 21 |
| RHPEYAVSVLLR | 38 | 8 | 7 | 8 | 15 | 21 | 18 | 21 | 39 |
| VSFELFADKVPK | 38 | 17 | 11 | 4 | 6 | 45 | 29 | 11 | 16 |
| IWHHTFYNELR | 37 | 21 | 8 | 3 | 5 | 57 | 22 | 8 | 14 |
| LGEYGFQNALIVR | 37 | 10 | 9 | 8 | 10 | 27 | 24 | 22 | 27 |
| VKGDVDVSLPK | 37 | 10 | 9 | 10 | 8 | 27 | 24 | 27 | 22 |
| VEIIANDQGNR | 36 | 9 | 10 | 10 | 7 | 25 | 28 | 28 | 19 |
| EGMNIVEAMER | 35 | 8 | 7 | 10 | 10 | 23 | 20 | 29 | 29 |
| HVGDLGNVTAGK | 34 | 8 | 9 | 8 | 9 | 24 | 26 | 24 | 26 |
| VSFELFADK | 34 | 8 | 11 | 10 | 5 | 24 | 32 | 29 | 15 |
| TLMNLGGLAVAR | 32 | 10 | 10 | 7 | 5 | 31 | 31 | 22 | 16 |
| QTALVELLK | 31 | 6 | 8 | 10 | 7 | 19 | 26 | 32 | 23 |
| DAGTIAGLNVLR | 29 | 4 | 8 | 11 | 6 | 14 | 28 | 38 | 21 |
| QLEDELVSLQK | 28 | 7 | 9 | 8 | 4 | 25 | 32 | 29 | 14 |
| APIIAVTR | 27 | 8 | 9 | 6 | 4 | 30 | 33 | 22 | 15 |
| IGFPWSEIR | 27 | 9 | 11 | 5 | 2 | 33 | 41 | 19 | 7 |
| LVTDLTK | 27 | 4 | 4 | 6 | 13 | 15 | 15 | 22 | 48 |
| GDVDVSLPK | 26 | 6 | 6 | 6 | 8 | 23 | 23 | 23 | 31 |
| KQTALVELLK | 24 | 1 | 1 | 10 | 12 | 4 | 4 | 42 | 50 |
| NDLAVVDVR | 24 | 5 | 6 | 6 | 7 | 21 | 25 | 25 | 29 |
| SGEIEPVSVK | 24 | 5 | 7 | 8 | 4 | 21 | 29 | 33 | 17 |
| SIDDLEEK | 24 | 9 | 6 | 5 | 4 | 38 | 25 | 21 | 17 |
| VLQATVVAVGSGGK | 24 | 4 | 7 | 6 | 7 | 17 | 29 | 25 | 29 |

Yellow highlight = spectral count percentage for the peptide is at least 10 percentage points higher for a particular biological cell group than for the other biological cell groups.

In Table 3, for a given peptide, the spectral count percentage is highlighted if it is at least 10 percentage points higher for a particular biological cell group than for the other biological cell groups. The highlighted peptides thus represent peptide biomarkers. At least one peptide biomarker is noted for each biological cell group. This analysis indicates the high likelihood of success in identifying protein biomarkers to which these peptides correspond. Completion of protein identification and metabolic pathway mapping will reveal the protein biomarkers and the biochemical pathways affected, respectfully. The resulting information will further our understanding of the physiological changes that occur in BALB/c 3T3 cells with ricin dose.

## 4. CONCLUSIONS

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. At the time of sample collection (48 hours post dosing) only attached cells were harvested. Adjustment was made for cell concentration in preparing samples for liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis. Analysis of the LC-MS/MS data, to determine peptide sequences followed by statistical filtering, has shown that at least one 3T3 peptide biomarker exists for each biological cell group (vehicle control and three ricin dose levels). These peptides correspond to particular proteins expressed by the 3T3 cells. Identification of the relevant proteins, followed by metabolic pathway analysis to determine the pathways in which those proteins are involved, may reveal the survival mechanism for cells dosed at low levels, which in turn will provide insight into the related 3T3 physiological changes required to overcome ricin dosing.

## ACKNOWLEDGEMENTS

The authors acknowledge the In-House Laboratory Independent Research program for funding, Alan Zulich and Dr. A. Way Fountain for administrative support, James A. Atwood, III for scheduling of software demonstrations, and NuSep for demonstration version of ProteoIQ software.

## RESOURCES

[1] Shimoda, T.; Funatsu, G. Binding of Lactose and Galactose to Native and Iodinated Ricin D. *Agric. Biol. Chem.* **1985**, *49*, 2125-2130.

[2] Bevilacqua, V.L.H.; Rice, J.S.; Madren-Whalley, J.S.; Reilly, L.M.; Rogers, T.J.; Schenning, A.M.; Smith, P.B.; Morrissey, K.M. *Ricin and Staphylococcal enterotoxin B Fate in Water Matrices*; ECBC-TR-702; U.S. Army Edgewood Chemical Biological Center: Aberdeen Proving Ground, MD, **2009**; UNCLASSIFIED Report (AD-A508 555).

[3] http://iccvam.niehs.nih.gov/methods/acutetox/invitrocyto/invcyt_proto.htm (accessed November 18, 2009).

[4] Washburn, M.P.; Wolters, D.; Yates, J.R. III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology.* **2001**, *19*,242-247.