9-8-2001

Proceedings of the Thirty-first Annual Biochemical Engineering Symposium

Larry E. Erickson
Kansas State University

Follow this and additional works at: http://lib.dr.iastate.edu/bce_proceedings

Part of the Biochemical and Biomolecular Engineering Commons

Recommended Citation
http://lib.dr.iastate.edu/bce_proceedings/27

This Book is brought to you for free and open access by the Chemical and Biological Engineering at Iowa State University Digital Repository. It has been accepted for inclusion in Biochemical Engineering Symposium Proceedings by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Proceedings of the
Thirty-first Annual Biochemical Engineering Symposium
September 8, 2001

Larry E. Erickson
Editor

Department of Chemical Engineering
Kansas State University
Manhattan, Kansas 66506
Proceedings of the Thirty-First Annual
Biochemical Engineering Symposium

The Thirty-First Annual Biochemical Engineering Symposium was held on September 7 and 8, 2001, at Kansas State University. The program included 10 oral presentations and 3 posters; however the paper by Boyack and Gilcrease was not presented because the presenter was ill and unable to come. Some of the papers describe work that is in progress while others describe completed projects. Many of the authors intend to submit their work for publication elsewhere in a more complete form. A listing of those who attended is given below. The activities began on Friday evening with an indoor picnic because of rain and wind.

List of Participants

Iowa State University: Jeffrey Gahan, Tony Hill, Matt Kipper, Alain Laederach, Chandrika Mulakala, Balaji Narasimhan, Peter Reilly, Jackie Shanks, Murali Subramanian, Melissa Swarts, and Ganesh Sriram

University of Oklahoma: Kemi Harris, Ulli Nollert, Elizabeth Phan, Omid Tawackoli, Tony Tran, Regina Visser, Heath Williams, and Marci Williams

University of Kansas: Wendy Sun

Kansas State University: Akkawit Aimdilokwong, Sigifredo Castro, Naveen Chennubhotla, Sasivimon Chittrakorn, Kenneth Dokken, Larry E. Erickson, L.T. Fan, Stevin Gehrke, Sharon Hagan, Brian Lindsay, Sameer Khaitan, and Kaila Young
31st Annual Biochemical Engineering Symposium
Fiedler Hall Auditorium
Saturday, September 8, 2001

8:30 a.m. Introduction and Welcome


9:00 TNT Transformation by Plants: Role of Hydroxylamines, Murali Subramanian and Jacqueline V. Shanks, Iowa State Univ.

9:25 Plant Uptake and Transformation of Benzotriazoles, Sigifredo Castro, Lawrence C. Davis, and Larry E. Erickson, Kansas State Univ.

9:50 Break to Put Up Posters and View Posters

10:20 Biological Sulfate Reduction in Alkaline Waters for the Reprocessing of Trona Tailings Ponds, Leigh Ann Boyack and Pat Gilcrease, Univ. of Wyoming

10:45 Microbial Degradation of Methyl Benzotriazole, Kaila Young, Lawrence C. Davis, and Larry E. Erickson, Kansas State Univ.

11:10 New Insights into the Catalytic Mechanism of the Family 47 Class I Alpha-1,2 Mannosidases by Computational Docking of Mannosyl Substrates, Chandrika Mulakala and Peter J. Reilly, Iowa State Univ.

11:35 Lunch and Poster Viewing

1:00 p.m. A Mathematical Model for Carbon Bond Labeling NMR Experiments: Analytical Solutions and Sensitivity Analysis for the Effect of Reaction Reversibilities on Estimated Fluxes, Ganesh Sriram and Jacqueline V. Shanks, Iowa State Univ.

1:25 Platelet Derived Nitric Oxide (NO) Inhibits Thrombus Formation on Collagen: The Role of Insulin, R. H. Williams and M.U. Nollert, Univ. of Oklahoma

1:50 Molecular Mechanics Calculations to Quantify Segmental Interactions in Bioerodible Polyanhydrides: Consequences for Drug Delivery, Matt Kipper and Balaji Narasimhan, Iowa State Univ.

2:15 Poster Viewing

2:30 Three-Dimensional Hydrophobic Cluster Analysis: Use of a Virtual Environment to Display Protein Cluster Information, Tony Hill, Alain Laederach, and Peter J. Reilly, Iowa State Univ.

Poster Presentations in Durland 129


2. The Use of Patterned Films for Directional Optic Nerve Growth, Jeffrey C. Gahan, Jennifer A. Behr, Surya K. Mallapragada, and Srdija Jeftinija, Iowa State Univ.

3. Multiple Sequence Alignment and Phylogenetic Analysis of a Family of Beta-Glycosidases, Tony Hill, Alain Laederach, and Peter J. Reilly, Iowa State Univ.
TABLE OF CONTENTS

Bioreduction of Chromium(VI) by Inactivated Medicago Sativa (Alfalfa) Biomass
Kenneth M. Dokken, Jorge L. Gardea-Torresdeyer, Kirk J. Tiemann, Jason G. Parsons,
and Gerardo Gamez .................................................... 1

TNT Transformation by Plants: Role of Hydroxylamines in the Pathway
Murali Subramanian and Jacqueline V. Shanks .................................. 11

Plant Uptake and Transformation of Benzotriazoles
Sigifredo Castro, Lawrence C. Davis, and Larry E. Erickson ...................... 21

Microbial Degradation of 5-Methyl Benzotriazole
Kaila Young, Larry Erickson, Lawrence Davis, and Sigifredo Castro Diaz ........... 29

Understanding Protein Structure-Function Relationships in Family 47 α-1,2-Mannosidases through Computational Docking of Ligands
Chandrika Mulakala and Peter J. Reilly ...................................... 35

A Mathematical Model for Carbon Bond Labeling Experiments: Analytical Solutions and Sensitivity Analysis for the Effect of Reaction Reversibilities on Estimated Fluxes
Ganesh Sriram and Jacqueline V. Shanks ..................................... 45

Platelet Derived Nitric Oxide (NO) Inhibits Thrombus Formation: The Role of Insulin
R.H. Williams and M. U. Nollert ............................................ 55

Molecular Mechanics Calculations to Quantify Segmental Interactions in Bioerodible Polyanhydrides: Consequences for Drug Delivery
Matt Kipper and Balaji Narasimhan ......................................... 65

Three-Dimensional Hydrophobic Cluster Analysis: The Use of a Virtual Environment for Protein Sequence Analysis: HELIX v0.3
Anthony D. Hill, Alain Laederach, and Peter J. Reilly .......................... 73

Mechanical Performance of Elastin-Mimetic Hydrogels
Éder D. Oliveira, Sharon A. Hagan, and Stevin H. Gehrke ........................ 76

Multiple Sequence Alignment and Phylogenetic Analysis of Family 1 β-Glycosidases
Anthony D. Hill, Alain Laederach, and Peter J. Reilly .......................... 82
BIOREDUCTION OF CHROMIUM(VI) BY INACTIVATED MEDICAGO SATIVA (ALFALFA) BIOMASS


*Present address: Department of Biochemistry, Kansas State University, Manhattan, Kansas 66506
**Department of Chemistry, University of Texas at El Paso, El Paso, Texas 79968.
*Corresponding author phone: (915) 747-5359; fax: (915) 747-5748; e-mail: jgardea@utep.edu

Abstract

Previous studies have demonstrated the ability of Medicago sativa (alfalfa) to adsorb chromium(III) ions in aqueous solution. However, the use alfalfa to remove chromium(VI) from aqueous solution has not been considered. Thus, batch studies were conducted to determine the behavior of chromium(VI) binding to alfalfa biomass. Optimum chromium(VI) binding was observed at pH 2.0 with increased binding over a 48 hour period. Modification experiments using esterified and hydrolyzed alfalfa biomass were carried out to determine the participation of carboxyl groups in chromium(VI) binding. The unmodified, hydrolyzed, and esterified alfalfa biomasses showed binding capacities of 5.9, 10.4, and 0.0 mg Cr(VI) per gram of biomass, respectively. Thus providing evidence that carboxyl groups may play a role in chromium(VI) binding. X-ray absorption spectroscopic studies were used to characterize chromium(VI) adsorption by alfalfa biomass through X-ray Absorption Near Edge Structure (XANES) and Extended X-ray Absorption Near Edge Structure (EXAFS). The results of the batch and XAS studies indicate that alfalfa biomass bioreduces chromium(VI) to chromium(III) and subsequently binds chromium(III) predominantly through carboxyl groups.

Introduction

Chromium(VI) is commonly found in wastewater from metal finishing, refractory, and metallurgical industries [1-3]. It is also commonly used as a pigment and a corrosion inhibitor. The increased demand for these products has led to a rise in wastewater production and contamination of natural waters. Chromium(VI) is the most toxic form of chromium and is highly soluble and mobile in the aquatic environment. The most common method for remediating chromium(VI) involves the reduction of chromium(VI) to chromium(III) by chemical or electrochemical means, followed by precipitation, and then filtration [3]. Other conventional technologies include adsorption by activated carbon or strongly basic anion exchange resins [4-8]. However, these techniques are extremely costly, utilize harsh chemicals in their fabrication, are energy and time extensive, and are subject to fouling in hard water conditions. Bioremediation has been the dominant alternative method of choice for the past twenty years. Losi et al. conducted a study on a wide variety of microorganisms for their potential to uptake chromium(VI) [3]. Live plants can adsorb chromium(VI) from contaminated waters.
However, live systems require constant maintenance, nutritional supplementation and can fall prey to the toxicological effects of chromium(VI). This presents a need for the production of techniques that can overcome these problems.

The use of inactivated (dead) plant or algal biosorbents and industrial by-products, such as sawdust and fly ash, has presently generated much interest due to their cost effectiveness, low maintenance, and high durability as compared to live systems [11-16]. Alfalfa (Medicago sativa) possesses these much wanted characteristics and has shown the potential to adsorb chromium(III) and other heavy metals at high levels in previous studies [17-19]. It is known that some plants have the ability to bioreduce chromium(VI) to chromium(III), and it is known that plants possess functional groups with reducing properties [10,11,20-22]. Alfalfa biomass may possess these reducing groups and may be able to function as a bioreduction/adsorption system for chromium(VI) [23].

**Experimental**

**Alfalfa Collection**

The alfalfa plant tissues used in this study were acquired from controlled agricultural field studies at New Mexico State University in Las Cruces, New Mexico. The species of alfalfa (Medicago sativa) used was the Malone variety. The plants were removed from the soil and washed thoroughly using deionized water to remove any soil or debris. The roots were separated from the shoots and oven dried at 90°C for a week. The samples were ground to pass through a 100-mesh sieve using a Wiley mill. The shoot biomass was utilized for this study due to its ease in agricultural availability and considerable adsorption capacity for aqueous heavy metal ions [17, 18].

**pH Profile for Chromium(VI) Binding**

Batch laboratory techniques were used for pH studies as previously reported by Gardea-Torresdey et al. [17,18]. Samples of unmodified alfalfa biomass was placed into deionized water at a concentration of 5 mg of biomass per mL. The pH of the suspensions were adjusted to pH 2.0 and allowed to equilibrate, and 2-mL aliquots were transferred to three 5-mL plastic tubes. The pH was then adjusted to 3.0, 4.0, 5.0, 6.0, and 7.0, and 2-mL aliquots of the suspension at each pH were transferred to three clean 5-mL tubes. The suspensions were centrifuged at 3000 rpm for 5 min. A 0.1 mM chromium(VI) solution was prepared from potassium dichromate (K$_2$Cr$_2$O$_7$) salt and the pH was adjusted to 2.0, 3.0, 4.0, 5.0, 6.0, and 7.0. At each pH, 2 mL of chromium(VI) solution were added to its respective pH biomass pellet. Each pH had a control chromium(VI) solution to compare for metal analyses. All of the samples were equilibrated for one hour on a rocker. The samples were then centrifuged for five minutes at 3000 rpm. The supernatants were transferred to clean tubes and kept for metal analyses by flame atomic absorption spectroscopy (FAAS). Experiments were conducted in triplicate for quality control purposes.

**Time Dependent Study for Chromium(VI) Binding**

The time dependent study was performed according to studies previously performed [19]. The protocol utilized an alfalfa biomass concentration of 5 mg per ml of solution. A 0.3mM Cr(VI) solution at pH 2.0 was prepared. The time intervals used in
this study were: 5, 10, 15, 20, 25, 30, 45, 60, 90, 120, 360, 720, 1440, and 2880 minutes. The alfalfa biomass was reacted with the Cr(VI) solution for the aforementioned time intervals. After the indicated time intervals, the samples were centrifuged for five minutes at 3000 rpm. The final pHs were recorded and the chromium concentration of the supernatants were determined by FAAS.

**Esterification of Alfalfa Biomass**

The esterification of the alfalfa biomass was conducted as previously reported [19]. Nine grams of Malone shoot biomass was suspended in acidic methanol solution containing 633 mL of 99.9% methanol and 5.4 mL of concentrated hydrochloric acid (HCl). The solution was continuously stirred for eight hours at 60°C. After eight hours, the alfalfa was removed from the heat and allowed to cool to room temperature. The alfalfa was pelleted by centrifugation at 3000 rpm for five minutes. The pelleted alfalfa biomass was washed three times with deionized water to stop the esterification reaction and remove any excess acidic methanol. The esterified alfalfa biomass was lyophilized overnight in a Labconco freeze dryer.

**Hydrolyzation of Alfalfa Biomass**

Hydrolyzation of the alfalfa biomass was performed according to previously described protocol [19]. Nine grams of Malone shoot biomass were reacted with 100 mL of 0.1 M sodium hydroxide (NaOH) for one hour. The biomass was then pelleted by centrifugation at 3000 rpm for five minutes and washed three times with deionized water. The hydrolyzed alfalfa was lyophilized overnight in a Labconco freeze dryer.

**Chromium(VI) Binding Capacities**

Batch laboratory methods utilized in previous studies conducted by Gardea-Torresdey et al. were used to determine the chromium(VI) binding capacity of the unmodified and modified alfalfa biomasses [17]. 100 mg of each sample were washed twice with 0.1 M HCl and twice with deionized water to remove any debris or soluble compounds. The washings were collected and weighed to account for any biomass or resin loss. The washed samples were resuspended into 10 mL of deionized water, and the pH was adjusted to pH 2.0. A 0.3mM chromium solution was prepared at pH 2.0 from the potassium chromate (K₂Cr₂O₇) salt. The samples were reacted for ten fifteen minute intervals with the chromium(VI) solution or until saturation was achieved. The samples were then centrifuged at 3000 rpm for five minutes. The supernatants were placed into clean tubes and analyzed for chromium concentration using FAAS.

**Flame Atomic Absorption Spectroscopy (FAAS)**

A Perkin Elmer model 3110 Atomic Absorption Spectrometer was used to determine the chromium content in the batch pH and capacity studies. The analytical wavelength used was 359.8nm with a slit width of 0.7nm. The chromium lamp current was 25mA. An impact bead was used to improve sensitivity. A calibration curve with a correlation coefficient of 0.98 or better was obtained. Known standards were used to check the instrument response. The samples were analyzed three times and the mean value and relative standard deviation were determined. In order to work within the linear
range, some samples were diluted using 0.1M hydrochloric acid. The amount of accumulated chromium by the alfalfa biomass was determined by mass balance.

**X-ray Absorption Spectroscopy (XAS)**

The X-ray Absorption spectroscopic studies were conducted at Stanford Synchrotron Laboratories (SSRL) at beamline 7-3. The Cr K edges (5989 eV) were collected using standard operating conditions of 3 GeV and 60-100 mA [12]. The chromium(VI) loaded biomass samples were run at 20K using a liquid helium cryostat to reduce any dampening or thermal disorder. Fluorescence spectra measurements of the biomass samples were acquired using a Canberra 13-element array germanium detector. The transmission measurements for the model compounds were taken using argon filled ionization chambers. The model compounds, potassium dichromate (K$_2$Cr$_2$O$_7$) and chromium(III) phosphate (CrPO$_4$), were ground and diluted using boron nitride. Samples of unmodified alfalfa biomass and Diaion CRB02 amino, Diaion WTO1S carboxyl, and cellulose phosphate resins (Supelco, Belafonte, PA) were saturated with a 1000 ppm Cr(VI) solution[19mM K$_2$Cr$_2$O$_7$] prior to analysis at pH 2.0 [24]. The samples and model compounds were packed into aluminum sample holders with x-ray transparent Kapton tape with a 1 mm path length. The packed samples were measured as solids. A Si(220) double crystal monochromator with an entrance slit of 1 mm was used to conduct the measurements. The monochromator was detuned by 50% in order to reject higher order harmonics. All spectra were calibrated against the edge position (5989 eV) of a chromium foil internal standard. To improve the signal to noise ratio, the scans of XANES spectra were averaged.

**Results and Discussion**

**Batch Studies**

The binding trend of chromium(VI) with respect to pH can be observed in Figure 1. It can be seen that the optimum chromium(VI) binding pH is 2.0 which is similar to others studies with seaweed and water hyacinth which display optimum chromium(VI) binding at acidic pH [10,11]. Gardea-Torresdey et al. also observed good binding at low pH for chromium(VI) with oat byproducts [25]. A low binding pH suggests the involvement of positively charged functional groups such as amino groups to provide protons for a bioreduction process described by Kratochvil et al [11]. These investigators proposed that the adsorption of chromium(VI) is similar to that of a weak anion exchange resin. At low pH, protons are available to bind to the chromate. Then protonated chromate can oxidize the alfalfa biomass and be reduced to chromium(III) ions. These chromium(III) ions may then adsorb to the alfalfa biomass similar to a cation exchange method. However, it is evident that the bioreduction and adsorption process consists of several steps. Further indication that a stepwise process is occurring can be seen in Figure 2 which depicts the amount of chromium bound with respect to time. The amount of chromium bound to the alfalfa biomass increases over a 48 hour period from 30% to almost 60%.

Table 1 shows the chromium(VI) binding capacities for unmodified and modified alfalfa biomasses. It can be observed that the hydrolyzed alfalfa produced a 46% enhancement in chromium(VI) binding from 5.9 to 10.4 mg chromium per gram of alfalfa biomass. Chromium(VI) adsorption was not observed for the esterified biomass.
These results indicate that carboxyl groups are playing a vital role in the adsorption/reduction of chromium(VI) by the alfalfa biomass. Similar results were observed by Gardea-Torresdey et al. for the trivalent species of chromium using unmodified and modified alfalfa biomasses [19]. This suggests that chromium(VI) is binding to the alfalfa biomass as chromium(III).

**X-ray Absorption Spectroscopy (XAS)**

Figure 3 depicts the XANES spectra for model compounds potassium dichromate and chromium(III) phosphate and chromium(VI) reacted with the alfalfa biomass at pH 2.0. It can be observed that the XANES spectrum for the chromium(VI) reacted with alfalfa biomass resembles that of the XANES spectrum for the chromium(III) phosphate. The difference between the XANES spectra for trivalent and hexavalent chromium can be seen by the prominent pre-edge for chromium(VI) at about 6 keV. This pre-edge is due to a bound-state 1s to 3d transition which is forbidden for octahedral Cr(III)O₆ but allowed for noncentrosymmetric tetrahedral Cr(VI)O₄ molecules[26]. The empty 3d orbitals of Cr(VI) allow for 1s to 3d transitions and therefore enhance the intensity of the pre-edge peak. The almost identical XANES spectra for the chromium(III) phosphate model compound and the alfalfa biomass reacted with chromium(VI) indicate that the chromium is binding to the biomass as chromium(III) and further provides evidence that a bioreduction process has taken place. Small pre-edge features can be observed for the chromium(III) phosphate and the chromium(VI) reacted with the alfalfa biomass at about 6 keV. The pre-edges for chromium(III) are at slightly higher energy positions than chromium(VI). This phenomenon can be explained in terms of the characteristics of the Cr-CO bond [27,28]. Figure 3 and Table 1 provide evidence that chromium(VI) is bioreduced to chromium(III) and subsequently binds to the alfalfa biomass by way of an oxygen ligand, which is most likely a carboxyl ligand as indicated by the batch modification studies.

To determine which functional group(s) may be capable of bioreducing chromium(VI) to chromium(III), several ion exchange resins were reacted with chromium(VI) at pH 2.0. Figure 4 depicts the XANES spectra for chromium(VI) reacted with amino, phosphate, and carboxyl ion exchange resins. The carboxyl and phosphate resins share similar spectra because no chromium was bound to either of them yielding spectra without an edge. However, the spectrum for the amino resin reacted with chromium(VI) resembles that of a chromium(III) spectrum. This possibly indicates that amino groups may bioreduce chromium(VI) to chromium(III). This is possible because at a low pH of 2.0 the amino groups are protonated and thus available to bind the chromium(VI) oxyanions (which are negatively charged) and then reduce chromium(VI) to chromium(III). This follows the proposed anion exchange mechanism described earlier by Kratochvil et al [11].

**Conclusions**

The ability of the alfalfa biomass to bioreduce and adsorb chromium(VI) ions from aqueous solution was observed through the batch and XAS studies. The bioreduction/adsorption process occurs mainly at pH of 2.0 and increases over time. This
indicates that chromium(VI) adsorption is a multistep process consisting of the initial binding of chromium(VI), bioreduction from chromium(VI) to chromium(III), and binding of chromium(III). The first step requires a positively charged functional group, most likely an amino group, to bind chromium(VI) and donate protons for the bioreduction process. Next, bioreduction of chromium(VI) to chromium(III) by amino groups on the surface of the biomass may occur. After the bioreduction has taken place, the chromium binds to the alfalfa biomass in the trivalent state. The chromium binds to the alfalfa biomass by way of oxygen ligands as observed by the small pre-edge for Cr-CO coordination. The modification studies indicated that hydrolysis enhanced chromium(VI) binding while esterification yielded no binding of chromium(VI). This suggests that the carboxyl group is the predominant oxygen ligand involved in the process. Alfalfa biomass has the potential to be an environmentally safe, cost-effective system for removal of chromium(VI) from contaminated waters.

References


Figure 1: Effect of solution pH on the adsorption of chromium(VI) by inactivated alfalfa biomass.

Figure 2: Percent of chromium bound by inactivated alfalfa biomass at different intervals ranging from 0 min to 48 hours.
Table 1: Chromium(VI) Binding Capacities for Unmodified and Modified Alfalfa Biomasses

<table>
<thead>
<tr>
<th>Type of Biomass</th>
<th>mg Chromium/gram biomass</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmodified</td>
<td>5.9±1.1</td>
</tr>
<tr>
<td>Hydrolyzed</td>
<td>10.4±1.4</td>
</tr>
<tr>
<td>Esterified</td>
<td>0.0±0.1</td>
</tr>
</tbody>
</table>

Figure 3: XANES spectra for chromium model compounds and chromium(VI) reacted with the alfalfa biomass. Potassium dichromate (K₂Cr₂O₇) (—), Chromium(III) phosphate (CrPO₄) (—–), Chromium(VI) reacted with the alfalfa biomass at pH 2.0 (***).
Figure 4: XANES spectra of ion exchange resins reacted with chromium(VI) at pH 2.0. Diaion CRB02 Amino resin (—), Diaion WTO1S Carboxyl resin (— —), Cellulose phosphate (***).
TNT Transformation by Plants: Role of Hydroxylamines in the Pathway
Murali Subramanian and Prof. Jacqueline V Shanks
Chemical Engineering, Iowa State University, Ames, IA 50011

Introduction
Phytoremediation is the science of using plants to clean up pollutant wastes. It offers many advantages over conventional methods of cleaning up soil wastes, such as incineration, since it is much less expensive (EPA 2000) and more ecologically friendly. 2,4,6-trinitrotoluene, TNT, a widely used explosive has contaminated around 40 army land sites in the US, and many more worldwide (Spain 2000). Clean up of such sites before the deleterious effects of TNT in the food chain manifest themselves is important (Nagel et al. 1999, Palazzo and Leggett 1986, Won et al. 1976, Yinon 1990).

TNT transformation by plants has been sufficiently researched to generate a hypothetical pathway for its transformation. A ‘Green Liver’ model (Bhadra et al. 1999a, Bhadra et al. 1999b, Hughes et al. 1997, Sandermann 1994, Vanderford et al. 1997) has been proposed and verified for TNT transformation. The TNT transformative experiments were performed with Catharanthus roseus hairy roots and Myriophyllum aquaticum, aquatic plants. Under this model, TNT is initially reduced or oxidized, and then conjugated with plant molecules, and finally the conjugates polymerize to form inextricable bound residues, covalently bonded to the plant. A hypothetical TNT transformation pathway has been proposed (figure 1) (Burken et al. 2000), which encompasses all these stages. In the hypothetical TNT-transformation pathway hydroxylamines occupy a central position, and are acted on by different enzymes to form varying metabolites. The types of metabolites formed possibly include reductive, conjugative, oxidative and diazoxy compounds. In addition, hydroxylamines are highly unstable in the presence of oxygen and in aqueous media (Ahmad and Hughes 2000, Burken et al. 2000, Wang, C. Y. et al. 2000). These factors make the detection of hydroxylamines a challenge in the aqueous, aquatic media that conventional phytoremediation experiments are carried out in.

Hydroxylamines are the central focus of this paper, since their presence in TNT metabolism in plants has been speculated on but never proved. It was attempted to identify hydroxylamines in the C. roseus system, prove its importance in the pathway, and develop a stabilization scheme for it. Confirming the presence of hydroxylamines, and realizing their role in TNT-metabolism is very significant, given both the toxic nature of hydroxylamines (Fu 1990), and the fact that they are the most reactive compounds in the pathway. Until now, hydroxylamines have not been positively identified in the C. roseus system, although theoretical evidence points to its formation. The explanation for the non-observance of hydroxylamines lie in probably in the fact that their instability in aerobic and aqueous media make them undetectable by conventional HPLC. Stabilizing the hydroxylamines and positively identifying them in the Catharanthus roseus system will help in further understanding of TNT metabolism by plants.
Figure 1: Hypothetical TNT transformative pathway. Adapted from Burken et al, 2001.
Hydroxylamine Feeding Studies

Motivation:
To understand the role and position of hydroxylamines in the pathway, a feeding study was performed, where 2-hydroxylamino-4,6-dinitrotoluene was fed to two-week-old C. roseus roots, in their late exponential phase, at a concentration of 0.02 mM. As a comparison, 0.05mM 2-amino-4,6-dinitrotoluene was fed to another flask of two-week-old C. roseus roots. Temporal product analysis was done until all the metabolites formed disappeared using a PDA equipped HPLC with a C8 column.

Results and Discussion:
2-hydroxylamino-4,6-dinitrotoluene fed flasks showed the formation of two types of conjugates, previously identified as 2A-1 and TNT-1 (Bhadra et al. 1999b, Wayment et al. 1999), whereas the 2-amino-4,6-dinitrotoluene fed flask produced only the 2A-1 conjugate. These results demonstrated the fact that hydroxylamines do not need to be reduced to conjugate, and plant organic molecules can add either directly to the hydroxylamine group, or can be added to the amine group. Hence, this experiment verifies the position of the hydroxylamines in the pathway (figure 1), and proves their versatility. It also proves that there may exist many conjugative schemes in the plant, depending on the substrate.

Stability of hydroxylamines

Motivation:
The poor stabilities of hydroxylamines in aqueous systems were probed by carrying out stability runs under different pH conditions and monitoring the hydroxylamine concentrations in the system. A normal pH (pH approx 7) and a pH <1 were used, the lower pH being introduced to investigate the stability of hydroxylamines under oxidative conditions.

Results and Discussion:
Figure 2 has the concentration profile of 2-hydroxylamino-4,6-dinitrotoluene in water, while figure 3 has the same for the 4-hydroxylamine isomer. Large degradation amounts and pH dependence were observed for both the hydroxylamines. By 40 hours, 50 percent of the 2HADNT had disappeared, and 80 percent of the 4HADNT was missing, at normal pH. Conditions of pH < 1 had a destabilizing effect on the hydroxylamines, quite contrary to what was expected. This could be due to either faults in the experimental set-up or execution, or could imply that the hydroxylamines under low pH conditions in aqueous media get oxidized by the electron accepting water. Water can behave as a weak acid or base; hence under low pH conditions it probably behaves as an acid. The hydroxylamines could be oxidized back to their nitroso functionalities. No degradation products were observed, indicating that the hydroxylamines did not reduce to their diamines, or any other identifiable metabolite, thereby also proving that all metabolites seen in the C. roseus system were obtained by enzymatic transformations.
While these results appear to contrast with those published by Wang et al, who showed good stability for hydroxylamines under aqueous conditions, the one important difference is starting concentration of hydroxylamines. While they used 50 mg/L concentrations, we used maximum concentrations of 5 mg/L in an effort to mimic our in-
vivo systems. Hence, at such low concentrations, even small amounts of degradation are reflected as large percentage drops in the amount remaining. These experiments showed the need for good stabilizing procedure for hydroxylamine detection, since both isomers showed different effects to lowered pH.

High TNT Concentration Experiments

Motivation:
Since hydroxylamines occupy a central position in the TNT-metabolism pathway, they are reacted on by a number of enzymes, and hence have very fleeting lives in the \emph{C. roseus} system. In an effort to prolong their presence, high concentration TNT experiments were performed on late-exponential phase hairy roots. The effect expected was two-fold; the high concentrations of TNT would cause a higher flux through the pathway, and therefore higher concentration of hydroxylamines, and the hence, their presence may be observed for longer. In addition, the toxicity caused by high concentrations of TNT may inhibit efficient transformation of the TNT, and hence the entire pathway may slow down in rate. This would imply that the rate of degradation of the hydroxylamines would be slowed down too, and hence an HPLC analysis may prove its presence.

Late exponential phase sterile hairy roots were kept in 250 ml flasks, and the volume of the media in them was equalized to 50 ml, using Gamborg’s B-5 media, with sucrose. Three different levels of TNT were added to the flasks, ranging from 0.35 to 0.45 mM. Sterile TNT solution in methanol was used. Samples were taken periodically for immediate analysis via HPLC.

Results and Discussion:
Both 2-hydroxylamino, and its 4-hydroxylamino isomer were seen in the extracellular media. This is the first positive identification of these compounds in a \emph{C. roseus} system, and their appearance might be contingent on the deceleration of the pathway. Figure 4 has the concentration of 2-hydroxylamine formed during the high concentration experiment, while Figure 5 has the tabular values for 4-hydroxylamine production. As the experiment progressed the roots began to look visibly stressed, as their color changed from pale white to a pinkish brown. This indicated the toxicity effect of the high TNT concentrations on the roots. The transient TNT profile indicated an uptake of TNT by the roots, although in a less efficient manner, since the rate of uptake was slower. This indicated that the transformative pathway as a whole was progressing slower. 4-hydroxylamino-2,6-DNT was seen in brief quantities for small periods of time, with all of it disappearing within 4 hours. 2-hydroxylamino-4,6-DNT, by contrast, displayed more stability and present for at-least 9 hours before falling below detection levels. The transient profile of its concentration shows a rise in concentration, followed by a rapid decline in levels.
16

12

0.37 mM (80 mg/L) TNT

0.47 mM (100 mg/L) TNT

0.33 mM (70 mg/L) TNT

Time (hours)

Figure 4: Transient extracellular concentrations of 2-hydroxylamine in the high-TNT concentration experiment.

<table>
<thead>
<tr>
<th>Initial TNT</th>
<th>Maximum 4HA</th>
<th>Hours since TNT addition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.33 mM</td>
<td>3.6</td>
<td>1</td>
</tr>
<tr>
<td>0.37 mM</td>
<td>7.85</td>
<td>4.5</td>
</tr>
<tr>
<td>0.47 mM</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 5: Extracellular 4-hydroxylamine levels in the high-TNT concentration experiment.

Amongst the other metabolites formed were the usual monoamines and a couple of the conjugates. These metabolites were formed at low-levels (below 5 mg/L), and this seems indicative of the fact that enzymes transforming these compounds were performing at their usual rate, and were not significantly affected by the high TNT concentration. Since hydroxylamine presence was proven in the transformation of TNT, its stabilization was next attempted in order to standardize its identification in phytoremediation systems.

**Derivatization of hydroxylamines**

**Motivation:** Optimization of a derivatization procedure for hydroxylamine stability (Wang, C. and Hughes 1998) was done, in an attempt to prevent the natural degradation of hydroxylamines in in-vitro conditions. Derivatization stabilizes the hydroxylamine group, thereby making it more detectable by conventional HPLC schemes. The resultant products, a 2,6-dinitro-4-(N)-acetoxyacetamido toluene (figure 6), or its 2,4-isomer, are most stable than the parent hydroxylamines and hence application of this procedure was expected to improve the identification of hydroxylamines in the *C. roseus* system.

The modified-derivatization procedure starts off with acidiying the sample containing the hydroxylamine to a pH < 1, since the lower pH implies oxidative conditions, and hence the potential for the natural conversion of the hydroxylamines to
their diazooxy and nitroso functionalities reduces. Methanol is added to aqueous samples to further increase the stability of the hydroxylamines and then acetic anhydride is added slowly drop-wise. The entire mixture is chilled in an ice-bath, and the products are allowed to stand for 5 to 10 minutes, after which Sodium Acetate is added to quench the reaction. Sodium Acetate also acts to stabilize the product formed by, probably, reacting with any of the unreacted acetic anhydride.

![Diagram of derivatization process]

**Figure 6: Derivatization of 4-hydroxylamine.**

**Optimization of Derivatization:** Since the derivatization procedure had quite a few parameters, and was potentially going to be used to quantify hydroxylamines in the in-vivo *C. roseus* system, it was going to be necessary to first optimize the procedure to obtain conditions of maximum yield and stability of the hydroxylamines. Towards this end, optimization with respect to amount of acetic anhydride, sodium acetate, and delay time (time gap between addition of acetic anhydride and sodium acetate) were performed. Optimization with respect to the amount of water the derivatization procedure could successfully handle, and the minimum amount of methanol required are important parameters for which derivatization runs need to be performed. Water content in the system is especially important since all the in-vivo samples are aqueous. Important: All optimization runs were carried out using 2,6-diamino-4-hydroxylamino since the UV-vis spectra for the corresponding derivatization product was obtained from Dr. Chaunyue Wang in the Hughes lab at Rice Univ.

**Acetic Anhydride Optimization:** Since Acetic Anhydride is the primary reactant with the hydroxylamines, the derivatization was optimized with the amount of Acetic Anhydride added. 100, 500, 1000, 2000 and 5000 microliters of Acetic Anhydride was added to the cocktail of 4HADNT, methanol and water, acidified to pH<1. The amounts of these components were 75 microliters (0.1 mg/ml concentration), 0.1 ml, 0.1 ml and 25 microliters respectively. 500 microliters of Sodium Acetate was added to stop the reaction after 5 minutes of reaction time. When two distinct aqueous and organic phases were visible, the sample for HPLC analysis was taken from the organic phase. A transient concentration measurement of the derivatization product was also done, to show the stability of the derivatized product, and to show its rate of formation.
**Sodium Acetate Optimization:** A procedure similar to the adopted for Acetic Anhydride optimization was used. Three different amounts of sodium acetate were used to stop the reaction- 250, 1000 and 2000 microliters. 50 microliters 4HADNT, 0.1 ml Methanol, 0.1 ml water, 1 ml Acetic Anhydride and 25 microliters of HCl were used in the derivatization reactions.

**Results and Discussion:**
The derivatized products when analyzed using the HPLC (C8 column, 78:22, Water: IPA- mobile phase) showed two distinct peaks- one at 18 minutes, and the other at 41 minutes. In addition, a few minor peaks were seen, but the levels were very low, and hence they were not analyzed. The product eluting at 18 minutes was identified, using UV-vis match, as 4(N)-acetoxyacetamido-2,6-DNT, the derivatized product of interest. The other peak (41 minutes) was an unidentified byproduct; it appeared to be a modification of 4-hydroxylamino-2,6-DNT. The transient concentrations of both products are presented in figure 7. The highest concentration attained is that of the 3000 microliter Acetic Anhydride sample, after 50 hours. The general trend seems to indicate that higher amounts of Acetic Anhydride will result in higher amounts of derivatized product being formed, and the amount formed increases with time, probably reaching a maximum by 50 hours. The rationale for assuming that the maximum amount is formed in 50 hours is that the concentration of the chief by-product formed falls to zero by 50 hours, in the case of 2000 and 3000 microliters Acetic Anhydride samples. Therefore, for Acetic Anhydride, the most suitable option is probably a large amount, in excess, and a gestation period of several hours. Comparing the concentration profiles of the by-product formed and the derivatization product reveals the fall of concentration of the formed is accompanied in a rise in concentration of the latter, and vice-versa. This indicates that an equilibrium exists between the two products, and the by-product, with time eventually reacts/degrades to become the (N)-acetoxyacetamido substituent. This is a significant finding, since it implies that the presence of the by-product means that all the hydroxylamines have not fully reacted, and more reaction time is required.

As in the case of Acetic Anhydride optimization, the amount of product formed increases with the amount of Sodium Acetate added to the system. The transient concentration profiles for the derivatization product show an increase in amount of product formed with passage of time, and a corresponding decline in concentration of by-product with time. This is again indicative of the fact that the by-product is in equilibrium with the derivatization product, and in time gets completely converted to the derivatization product. Therefore, using high amounts of sodium acetate (2 ml) and a reaction time of around 35 hours will probably give best results for the derivatization procedure. The two optimization runs of the derivatization procedure seem to indicate non-limiting, high amounts of acetic anhydride and sodium acetate are probably the best option. Also, a reaction time of around 35 hours is optimum for completion of reaction.

**Conclusions and Future Directions**
The role of hydroxylamines in the TNT transformative pathway was demonstrated by the feeding studies, where hydroxylamine fed roots produced two kinds of conjugates, as opposed to amino-fed roots that produced only one kind of conjugate. In addition, hydroxylamines were identified in high concentration TNT experiments, thereby further confirming their role in the TNT metabolism. The poor stability of hydroxylamines was shown in aqueous studies where they disappeared within 40 hours. An optimized
Figure 7: Transient concentration of 4-acetoxyacetamido-2,6-DNT and the byproduct in optimization w.r.t Acetic Anhydride.
derivatization procedure was developed to stabilize the hydroxylamines, which is expected to be useful in further in-vivo TNT metabolism studies.

References:
Plant Uptake and Transformation of Benzotriazoles

Sigifredo Castro (1), Lawrence C. Davis (2), and Larry E. Erickson (1).

KANSAS STATE UNIVERSITY

(1) Department of Chemical Engineering, KSU, Manhattan, KS 66506, Phone: 785-532-5584, Fax: 785-532-7372. (2) Department of Biochemistry, KSU, Manhattan, KS 66506, Phone: 785-532-6121, Fax: 785-535-7278.

ABSTRACT

Benzotriazoles are used in three main applications: to prevent the corrosion of metals, to stabilize plastics against UV decomposition, and to improve photographic characteristics of films. Recently the 1-hydroxy derivative is being considered as an alternative laccase mediator in biopulping for paper production. The majority of benzotriazoles produced go into anticorrosion applications, such as aircraft deicing fluids, automobile antifreeze solutions, and recirculating water systems. Benzotriazoles are stable compounds, toxic to microorganisms and plants, irritant to humans, and potentially carcinogenic. The release of these compounds to the environment represents an important environmental concern. Given that the structure of benzotriazoles is similar to some of the components of the lignin fraction in plants, we have studied the potential of using plants to remediate soil contaminated with these compounds. A hydroponic culture of young sunflowers (Helianthus annuus) under continuous lighting, and constantly fed at different concentrations of benzotriazole (BT) or hydroxy benzotriazole (HBT), was designed to determine the toxicity threshold and to gain a better understanding of the process by which the plants interact with these compounds, in an effort to elucidate the driving force of the process, determine the kinetics, and identify relevant parameters.

INTRODUCTION

The high stability of the chemical structure derived from the condensation of a benzene ring with a three nitrogen ring (USEPA, 1977) is one of the reasons for the use of benzotriazoles as corrosion inhibitors in aircraft deicing formulations (ADFs); however, the collected but not properly treated and the uncollected runoff generated from deicing processes at airports, represent a continuous source of benzotriazole-enriched waste that is undesirably persistent in the environment. Benzotriazole and its derivatives (except for 5-methyl-benzotriazole, 5-MBT) have been classified as resistant to bacterial degradation (Rollinson and Callery, 1986, Castro et al., 2000, Young et al., 2001). The presence of benzotriazole derivatives has been reported in monitoring wells at some airports (Cancilla et al., 1998), because of their significant water solubility and non-reactive character (USEPA, 1977). Because of the known toxicity of benzotriazoles to microorganisms, aquatic fauna, and plants (Pillard, 1995, Castro et al., 2000), and potential carcinogenicity to mammals (NCI, 1978), suitable strategies for the environmental management of benzotriazole waste should be found.

We have proposed (Castro et al., 2000) the use of plants for the in situ immobilization of benzotriazole (BT) and its 4 and 5 methyl derivatives (5-methyl-
benzotriazole and tolyltriazole (TT), which is a mixture of the 4 and 5-methyl benzotriazoles). Three factors provide the rationale for using vegetation as a strategy for triazole-waste remediation. First, triazoles are structurally similar to some substructures of the lignin in plants (see Figure 1). Second, the relatively small molecular size, their appreciable water solubility and most important, their LogK\textsubscript{OW} values (See Table 1), put them in the appropriate range for compounds to be mobile in the xylem and immobile in the phloem of the plants vascular system (Narayanan et al., 1999). Third, triazoles have been shown to be susceptible to react with peroxidases under specific conditions (Fenton reaction, in horseradish plant culture, or in cultures of the fungus Phanerochaete chrysosporium, Wu et al., 1998). The mechanism by which higher plants strengthen their cell walls, known as lignification, is probably a free-radical catalyzed polymerization of methoxylated aromatic alcohols with laccase and lignin peroxidases as the catalysts (Buchanan et al., 2000). It is reasonable to think that benzotriazoles may be reactive in plants where lignin is being synthesized and be covalently bonded into plant lignin. However, as it will be shown in this paper, triazoles are toxic to plants at concentrations in the liquid phase surrounding the roots of about 75 mg/L of triazole. Higher concentrations severely damage the root system and plants are dramatically stunted or killed. Nevertheless, not only can plants grow in properly fertilized solutions with triazole concentrations lower than 75 mg/L, but they also can make BT, 5-MBT and TT disappear from the contaminated solution to be irreversibly bound to the plant material (Castro et al., 2000). Another benzotriazole derivative, 1-hydroxy-benzotriazole (HBT), with a value of LogK\textsubscript{OW} less than 1 (See Table 1), shows different interactions with plants compared to the BT, 5-MBT and TT. The exact mechanism of triazole phytotoxicity and their ultimate fate within the plants remain unknown. In an effort to understand the process by which benzotriazoles can be taken up and transformed by plants, a hydroponic experiment with sunflowers (Helianthus annuus) was designed and is described in the following section. Given that 5-MBT is suspected to be degraded by bacteria (Young et al., 2001), only BT and HBT were studied.

Table 1. Some properties of benzotriazoles selected for study

<table>
<thead>
<tr>
<th>Compound</th>
<th>Solubility in water (wt % at 25°C)</th>
<th>Log octanol-water partition coefficient (Log K\textsubscript{OW})</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Experimental</td>
</tr>
<tr>
<td>1-H-benzotriazole</td>
<td>1.98\textsuperscript{a}</td>
<td>1.44\textsuperscript{b}</td>
</tr>
<tr>
<td>1-Hydroxy-benzotriazole</td>
<td>12\textsuperscript{c}</td>
<td>N/A</td>
</tr>
<tr>
<td>5-methyl-benzotriazole</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Tolyltriazole</td>
<td>0.55\textsuperscript{a}</td>
<td>N/A</td>
</tr>
</tbody>
</table>

\textsuperscript{a} From USEPA, 1977
\textsuperscript{b} From SRC, 2001
\textsuperscript{c} From Advanced ChemTech, 2001
Figure 1: Chemical structure of benzotriazole and some derivatives

MATERIALS AND METHODS

Pure benzotriazole (BT), and 1-hydroxy-benzotriazole (HBT) were purchased from Aldrich / Sigma Chemical Co. The compounds were kept as aqueous stock solutions (2 g/L) for later dilution to treat plants and/or prepare calibration standards for chromatography. Stored solutions appeared to be stable for at least two years. For separation and quantification of triazoles, detection at 275 nm was used with liquid chromatography on a Hamilton PRP-1 Column (150 x 4.1 mm). The eluent consisted of a solution of 1 mg/L 5-MBT and 1mM potassium phosphate monobasic in methanol (50-70 Vol.%) + water at a flow rate of 1.0 mL/min. The low level of MBT in the eluent allowed a uniform distribution of the injected sample throughout the column, which in turn resulted in a stable baseline in the chromatographic chart and more reproducible results with greater sensitivity. The phosphate was used to buffer the solution at a neutral pH. The methanol level was adjusted to optimize resolution of compounds of interest in each experiment. The detection limit of triazole was 2 mg/L at a range of 0.2 absorbance units for a full scale in the UV detector.

For the growth of plants, lighting was continuous with 40-watt cool white fluorescent light tubes (about 10 four-foot tubes per m²) at 25°C. The light energy available for photosynthetic activity fell in the range of 150 – 200 μmol/m², measured with a quantum meter (Apogee instruments).

Twenty sunflower seedlings that had been kept in nutrient-enriched moist vermiculite for fifteen days were transferred to 50 mL glass bottles with Hoagland’s 1X solution. The bottles were placed in a rack in such a way that light was prevented from
reaching most part of the solution to avoid the growth of photosynthetic algae. The plants were held by an open cell polyurethane foam assuring that only the roots were in contact with the water phase. After a four-day period for adaptation in the liquid system, the solution was changed to triazole-contaminated solution, where they were kept for 4 more days. Four of the plants were kept as controls, and two groups of eight were treated with either BT or HBT at concentration levels of 10, 20, 30, 50, 75, 100, 150, and 200 mg/L. Their initial fresh weights varied from 2.45 to 3.37 g right before starting the treatment and the triazole level for each plant was randomly assigned. The solution taken up by the plants was measured and brought back to the original 50 mL volume daily with fresh triazole-amended Hoagland's solution, after which a 1 mL sample was taken for HPLC analysis. The plant fresh weight was also measured daily.

RESULTS

With respect to water uptake, control plants showed an increase in the volume taken up daily since the plants were growing (see Figure 2). The same occurred for plants treated at low concentrations of triazole. For plants treated with benzotriazole, concentrations from 50 to 75 mg/L seemed to have affected the water uptake, but for concentrations of 200 mg/L, progressively diminished daily water consumption was observed (see Figure 2).

In the case of hydroxy benzotriazole, concentrations equal or less than 75 mg/L did not seem to severely affect the water uptake; however, at a concentration of 100 mg/L and higher, the water uptake was dramatically diminished the next day (see Figure 3).
Figure 3: Water uptake for sunflowers before and after treatment with hydroxy benzotriazole. Error bars show the standard deviation for the average of 4 control plants. Dotted lines show the data after treatment started.

To confirm the observations from the water uptake, the percentage change in the fresh weight during the four days of treatment was calculated as shown in Equation 1:

\[
\%\text{Fresh Weight Change} = \frac{\text{Final weight} - \text{Initial weight}}{\text{Initial weight}} \times 100\%
\]

Equation 1

Figure 4 shows that the percent change in fresh weight for plants treated with benzotriazole was in the lower range for the control plants for concentration between 20 and 75 mg/L and started to be less than the controls with a concentration of 100 mg/L. The effect seemed progressive, being more pronounced as the initial concentration was higher. For the hydroxy benzotriazole it seemed that the plants were able to tolerate and grow as the control plants at concentrations less than 75 mg/L; however, the plant treated with a concentration of 100 mg/L immediately showed a dramatic effect, similar to the plants with concentrations of 150 and 200 mg/L. Thus, it was concluded that a concentration of 75 mg/L of any of the triazoles was the maximum level tolerable by the plant without a severe effect in the water uptake and growth of the plant. Concentrations of benzotriazole higher than 75 mg/L affect the plant progressively, whereas concentrations of hydroxy benzotriazole higher than 75 mg/L may quickly lead to plant death.
Figure 4: Percent fresh weight change for young sunflowers treated with BT and HBT for 4 days (initial fresh weights varied from 2.45 to 3.37 g).

With respect to the interactions of triazole with plants, the amount that was "transformed" by the plant between sampling periods could be estimated by a mass balance, as shown in Equation 2, since the concentrations at the beginning and at the end of each sampling period and the volume and concentration of the solution added were known.

\[ \text{Initial amount} + \text{Amount added} - \text{Amount transformed} = \text{Final amount} \]  
**Equation 2**

In terms of the known data, Equation 2 will take the form of Equation 3:

\[ C_i V + C_s (V - SU) - \text{Amount transformed} = C_{i+1} V \]  
**Equation 3**

Where \( C_i \) is the initial concentration for interval 1, \( V \) is the total volume of solution, \( SU \) is the volume of solution taken up by the plant, \( C_s \) is the concentration of the solution added, and \( C_{i+1} \) is the initial concentration for interval \( i+1 \).

Thus, the amount of triazole transformed can be found for each interval; the total amount lost for the 4 days of the experiment will be the sum of the amounts lost during
each interval. Figure 5 shows the total amount transformed for each case for both treatments at all levels; it was clear that the plants were taking up the benzotriazole and the higher the initial concentration the more triazole was being transformed. However, the hydroxy benzotriazole-treated plants did not show the same behavior, and a particular trend could not be identified. The hydroxy benzotriazole seemed to be transformed, but the rate for the process was slower than that one for the benzotriazole.

![Figure 5: Total amount of triazole (mg) that was transformed by sunflowers treated with BT or HBT for 4 days](image)

**CONCLUSIONS**

A concentration less than 75 mg/L of BT seems to affect slightly the water uptake and growth rate of young sunflowers. A toxic effect, corresponding to plant weakening, stunted growth and reduction in water uptake is observed when the input BT concentration is greater than 75 mg/L; the effect is highly toxic at 200 mg/L. In the case of HBT, concentrations equal or less than 75 mg/L did not seem to severely affect the plants; concentrations equal or greater than 100 mg/L of HBT impact the plant severely rather quickly. HBT does not seem to be phytotransformed at the same rate as the BT and it does not seem to be taken up with the water coming into the plant, i.e., it may be excluded by the plant or phytotransformed very slowly. The differences in the phytotransformation of these compounds could be because plants can interact with them through different enzymatic pathways. Parameters such as LogKow for these compounds
are different and may be correlated with the observed phenomena; thus LogK\textsubscript{ow} could be used in future research to predict plant response to different triazole compounds.

ACKNOWLEDGMENTS

This research was partially supported by the U.S. E.P.A. and the U.S. Air Force under assistance agreements R-819653, R-825549, and R-825550 to the Great Plains Rocky Mountain Hazardous Substance Research Center for regions 7 and 8 under projects 94-27 and 98-3. It has not been submitted to the EPA for peer review and, therefore, may not necessarily reflect views of the agency and no official endorsement should be inferred. The Center for Hazardous Substance Research also provided partial funding.

REFERENCES

Abstract

Methyl benzotriazole (MBz) is used for several industrial applications and this contaminant is currently under study as a candidate for removal by phytoremediation. However, recent observation has shown that 5-MBz degradation occurs in the liquid remaining in the test vessels after the removal of the plants used for experimentation. Additionally, 5-MBz degradation has been observed in solutions containing samples of soil that had been previously exposed to the contaminant. Subsequent purification and separation of the microbial consortium present in the liquid and soil media led to the isolation of several morphologically distinct organisms believed to be responsible for the observed degradation. However, cultures of the separated organisms showed no ability to degrade the 5-MBz even when exposed to a variety of growth media and conditions. It was hypothesized that the organisms may exist in a symbiotic relationship so that 5-MBz degradation cannot occur in the absence of a bacterial consortium. Further experimentation will be conducted to investigate conditions under which the isolated organisms can degrade the contaminant. Work has also been done to elucidate the pathway by which the microbes degrade the 5-MBz. When both toluene and 5-MBz are present in solution containing the mixed culture, the degradation of toluene, which has a structure similar to a segment of the 5-MBz molecule, occurs at a slower rate than the rate observed when only toluene is present. This suggests that both substances may be degraded by similar mechanisms or that 5-MBz may inhibit toluene biodegradation.

Background

Federal Aviation Administration (FAA) requires that, for the safety of passengers, airplanes be subject to rigorous deicing standards in the winter months. To comply with the regulations set by the FAA, airports across the country routinely use massive amounts of deicing solution each year to remove ice from plane wings before takeoff, so that the plane is not dangerous weighted. However, the deicing solutions generally contain compounds that can potentially cause corrosion damage to the plane. Therefore, a mixture of 4- and 5-methyl benzotriazole (MBz, Figure 1), a corrosion inhibitor, is commonly added as a preventative measure. Another common use for MBz derivatives is protection of plastics against ultraviolet decomposition, as MBz is highly UV stable. As indicated by the UV stability of the compound, MBz is highly persistent in the environment and has shown toxic effects to plants and other organisms. Studies have found that prolonged exposure to MBz caused paralysis/central nervous system disorders in mice, and the compound is believed to inhibit protein synthesis (USEPA, 1977). Therefore, studies have been underway for several years in the search for an effective

$$\text{Figure 1. 5-Methyl benzotriazole (MBz).}$$
removal technique for the contaminant. A literature search revealed that degradation of MBz by microbes had not been reported (Rollinson, 1986). Instead, phytoremediation was under study as a means to remove the MBz contamination (Castro, et al., 2000).

Experimentation and Results

Many phytoremediation studies are done under hydroponic conditions, where the plants are grown in solution, without soil, so that sampling can be easily done (Castro, et al., 2001). In January of 2000, 5-MBz degradation was observed in the absence of vegetation in solutions previously used for these hydroponic studies. Subsequent 5-MBz feeds were also degraded in the absence of plants (Figure 2) and the growth of bacterial colonies was noted in the solution. The reader should note that the sharp increase of 5-MBz concentration on Day 3 corresponds to a “reloading” of the sample with additional 5-MBz so that a degradation rate could be estimated. The work here applies only to the 5-methyl derivatives. The 4-methyl form is not metabolized by the bacteria described herein (unpublished observation).

Bacterial Isolation

In the effort to isolate and determine the identity of the 5-MBz degrading organisms, streak plates were prepared using Hoagland’s Solution, the nitrogen-rich fertilizer used for hydroponic studies, as a base. Half of the prepared plates contained varying concentrations of 5-MBz in the agar support as a feed source for the organisms. The rest of the feed plates were simply an agar base with a crystal of 5-MBz placed in the center, so that the 5-MBz could diffuse through the agar and a zone of optimum 5-MBz concentration could be observed. After 72 hours, bacterial growth appeared on the plates in the form of 2 distinct colony morphologies: (1) small, clear and drop like and (2) large, yellow smears with blurred edges. Gram staining revealed that both colony types were gram (+). Zones of inhibition and optimum 5-MBz concentration were evident around the solid 5-MBz. The two colonies also grew on nitrogen-free media, which raises the point that 5-MBz may serve as a nitrogen source. However, in this case, the nitrogen source may have been the agar support. Excessive growth after only 24 hours was observed on plates that were rich in nutrients, indicating that 5-MBz is not a good sole source of energy.
In order to ascertain plausible sources of the 5-MBz-degrading organisms, soil from planted zones—channels that had been planted with alfalfa and cylinders that had been planted with grass—that had been treated previously with 5-MBz was removed and suspended in a fertilizer/5-MBz solution. A degradation of 98% of the initial 5-MBz concentration was noted after only 48 hours in all samples except the soil that had never been treated with 5-MBz (Figure 3). Note that no bacterial degradation of 5-MBz occurred in the solutions containing soil samples that had not been exposed to the contaminant, which indicates that the degradation is caused by an inducible enzyme or an increased population of microbes selected for MBz use. The morphology of isolated soil bacteria from these samples—small, clear, and drop like colonies—was similar to that of colony type (1) isolated from original hydroponics liquid.

![Figure 3. Degradation of 5-MBz by ubiquitous soil bacteria found in planted zones previously treated with MBz: Cylinder 1 (Δ), Cylinder 2 (x, dashed line), Cylinder 3 (*, dashed line), Channel 1 (*), Channel 2 (+), Soil untreated with MBz (o, dashed line).](image)

The isolated bacteria from the hydroponic solution and the soil samples were grown under a variety of conditions and on an assortment of media in the presence of 5-MBz. Semi-solid agar, liquid media, and flooded plates in which a layer of 5-MBz-containing solution was laid over a lawn of bacteria were all tried, however, even after several months, no degradation of the 5-MBz in the cultures was ever observed. Even recombining several isolated bacterial colonies into a single liquid culture did not lead to degradation of the 5-MBz. It was hypothesized that the oxygen or nutrient needs for the cultures were not being met. In original liquid culture, the bacteria appeared to exist in suspended “clumps”, where the observed flocculations possibly allowed for nutrients to be available in the necessary concentrations. There is also the possibility that a consortium is necessary for degradation of the 5-MBz. The ineffectiveness of recombining the bacteria indicates that, after several doubling times in the absence of other members of the symbiotic relationship, the bacteria may lose the ability to break down the 5-MBz even when they are again in the presence of other bacteria of the consortium. Alternatively, the effective organisms might not be culturable under the tested conditions. Although further experimentation will be conducted to investigate conditions under which the isolated organisms can degrade the contaminant, perhaps isolated colonies will never be able to perform the desired degradation.
Kinetics Studies

In addition to experimentation on the isolated bacterial colonies, work has also been done to ascertain the pathway by which the original consortium degrades the 5-MBz. Because of the similarity in molecular structure between toluene and 5-MBz, the kinetics of the degradation of both species were studied under the hypothesis that the mechanism of 5-MBz breakdown is related to that of the toluene degradation. In addition, the bacteria seem unable to degrade the MBz-related compound benzotriazole, which lacks the methyl group on the benzene ring. These kinetics experiments involved exposing the bacterial culture to both substrates and evaluating the effects of the presence of each substrate on the degradation of the other. Bottles were prepared with bacterial culture and 4.7 μmoles toluene, for each sample, with varying concentrations of 5-MBz. Each solution was diluted to either 11.5 or 15 mL, so that the head space in each bottle was 52.5 or 49 mL. The bottles were subsequently placed on a shaker. At the highest concentration of 5-MBz, about 90 mg/L, there were approximately four μmoles of 5-MBz per μmole of toluene in the solution phase; at the lowest concentration of 5-MBz, 3 mg/L, there were approximately 0.15 μmoles of 5-MBz per μmole of toluene. The concentrations of each component were measured over 20 hours (Figures 4 and 5). The reader should note that the scales on the y-axes are not consistently delineated from 0–100%.

Figure 4. Bacterial degradation of toluene in the presence of 5-MBz. The initial concentrations of 5-MBz in solution are: 86 mg/L (x), 28 mg/L (+, dashed line), 8.6 mg/L (Δ), 2.8 mg/L (*, dashed line), and 0 mg/L (o). The initial toluene molar concentration is approximately 0.18 mmolar in the liquid phase.

Inspection of Figure 4 reveals that 5-MBz in large concentrations inhibits the consumption of toluene as a substrate. This may mean that 5-MBz is degraded by the same mechanism as that for toluene, where competitive inhibition may be occurring between the two substrates.
Figure 5. Bacterial degradation of 5-MBz in the presence of toluene. The concentrations of 5-MBz in solution are: 86 mg/L (x), 28 mg/L (+, dashed line), 8.6 mg/L (Δ), and 2.8 mg/L (*, dashed line). The initial toluene molar concentration is approximately 0.18 mmolar in the liquid phase.

Figure 5 illustrates that low levels of 5-MBz in the presence of toluene are not rapidly degraded, indicating that the bacterial consortium simply feeds on the component in the highest concentration.

In order to determine if the presence of toluene affects the degradation of similar concentrations of 5-MBz, the original culture was exposed to two conditions: 5-MBz + toluene and 5-MBz alone. The concentrations of toluene and 5-MBz were measured over 50 hours (Figure 6).

Figure 6. Bacterial degradation of MBz in the absence (Δ, dashed line) and presence (o) of 0.15 mmolar equivalents of toluene. The initial MBz concentration was 50 mg/L.
From inspection of Figure 6, the degradation of 5-MBz in the presence of toluene was slightly faster than in the absence of toluene. Repetition of this experiment confirms these results. The addition of amounts of toluene from 0.5 to 3.0 times the molar concentration of the 5-MBz induces slightly faster degradation of 5-MBz.

Conclusions and Future Work

From the two sets of kinetic experiments and recent duplicate experiments, it appears that the bacterial consortium can degrade either toluene or 5-MBz—and degrades preferentially the component in the higher concentration. This leads to the possible conclusion that 5-MBz may be consumed by the same mechanism as toluene, where toluene di-oxygenase is the enzyme responsible for toluene degradation. Measurements of the increasing CO$_2$ levels over time in the headspace of samples exposed to 5-MBz, toluene, or both, indicate that the compounds serve as substrates and are at least partially mineralized.

A continuation of the kinetic studies described here is planned. Growth studies are also underway to determine the doubling time of the bacterial consortium in the presence of 5-MBz and toluene.

Acknowledgments

This research was partially supported by the USEPA and the U.S. Air Force under assistance agreements R-819653, R-825549, and R-825550 to the Great Plains-Rocky Mountain Hazardous Substance Research Center for regions 7 and 8 under projects 94-27 and 98-3. It has not been submitted to the EPA for peer review and, therefore, may not necessarily reflect the views of the agency and no official endorsement should be inferred. The Center for Hazardous Substance Research also provided partial funding.

References


34
Understanding Protein Structure-Function Relationships in Family 47 α-1,2-Mannosidases through Computational Docking of Ligands

Chandrika Mulakala and Peter J. Reilly*

Department of Chemical Engineering, Iowa State University, Ames, Iowa

ABSTRACT

Family 47 α-1,2-mannosidases are crucial enzymes involved in N-glycan maturation in the endoplasmic reticulum and Golgi apparatus of eukaryotic cells. High-resolution crystal structures of the human and yeast endoplasmic reticulum α-1,2-mannosidases have been recently determined, the former complexed with the inhibitors 1-deoxymannojirimycin and kifunensine, both of which bind in its active site in the unusual $1C_4$ conformation. However, unambiguous identification of the catalytic proton donor and nucleophile involved in glycoside bond hydrolysis was not possible from this structural information. In this work, α-D-galactose, α-D-glucose, and α-D-mannose were computationally docked in the active site in the energetically stable $4C_1$ conformation as well as in the $1C_4$ conformation to compare their interaction energetics. From these docked structures, a model for substrate and conformer selectivity based on the dimensions of the active site was proposed. α-D-Galactopyranosyl-(1→2)-α-D-mannopyranose, α-D-glucopyranosyl-(1→2)-α-D-mannopyranose, and α-D-mannopyranosyl-(1→2)-α-D-mannopyranose were also docked into the active site with their nonreducing-end residues in the $1C_4$ and $E_4$ (representing the transition state) conformations. Based on the docked structure of α-D-mannopyranosyl-$E_4$-(1→2)-α-D-mannopyranose, the catalytic acid and base are Glu132 and Glu435, respectively.

INTRODUCTION

Class I α-1,2-mannosidases (EC 3.2.1.113), which form glycosyl hydrolase Family 47,¹ are crucial enzymes in N-glycan synthesis in eukaryotic organisms. Glycoprotein synthesis through the N-glycan pathway begins with the transfer of Glc₄Man₉GlcNAc₂, a preformed mannose-rich oligosaccharide, from dolichyl phosphate to freshly synthesized polypeptides in the endoplasmic reticulum (ER).²⁻⁴ α-Glucosidases and α-1,2-mannosidases of the ER trim Glc₄Man₉GlcNAc₂ to form Man₅GlcNAc₂, which is transported to the Golgi apparatus for further processing. Subsequent trimming by the Golgi Class I α-1,2-mannosidases results in Man₅GlcNAc₂ formation, which is necessary for the maturation of the N-glycan to hybrid and complex oligosaccharides.

The crystal structure of the catalytic domain of ER Class I α-1,2-mannosidase of Saccharomyces cerevisiae has been recently determined.⁵ It has an unusual ($\alpha,\alpha$)$_7$-barrel structure, with an N-glycan (Fig. 1) from one molecule extending into the barrel of the adjacent symmetry-related molecule, interacting with the putative enzyme active site. A C-terminal β-hairpin protrudes into the center of the barrel from one side, plugging it. The other side of the barrel has a ~25 Å-diameter funnel-shaped cavity that decreases to ~5 Å at the funnel tube, which is also plugged by a Ca²⁺ ion. Site-directed mutagenesis of Arg273, located in the funnel neck in the yeast enzyme, to Leu273 (all residue numbering henceforth is based on the yeast enzyme) allowed the enzyme to cleave all four α-1,2-linked mannosyl residues, rather than just the single residue of the middle arm of Man₅GlcNAc₂.⁶

A crystal structure of the catalytic domain of human ER Class I α-1,2-mannosidase, both with and without the potent inhibitors 1-deoxymannojirimycin (DMJ) and kifunensine (KIF) (Fig. 2), has also been recently determined.⁷ Both inhibitors bind to the enzyme at the base of its
active site, with the Ca\(^{2+}\) ion coordinating and stabilizing O2 and O3 hydroxyls of the six-membered rings of both inhibitors in their \(^1\)C\(_4\) conformations. The overall structures of the yeast and human enzymes are essentially the same. Although the amino acid sequences of the two enzymes are no more than 35% similar, the positions of the amino acid residues that make up the active site in the two crystal structures are also practically identical.

This structural data, when combined with the distances of the catalytic acid and base from the glycosidic bond needed for catalysis,\(^8\) suggest that the only candidates as catalytic proton donor and nucleophile are Glu132, Asp275, and Glu435.\(^5,7\) Due to the absence of a terminal middle-arm mannosyl residue in the crystal structure, however, the two catalytic residues could not be clearly identified. The \(^1\)C\(_4\) conformation of the inhibitors in the \(-1\) catalytic site in the human enzyme suggests that the terminal mannosyl residue would also adopt the \(^1\)C\(_4\) conformation. Given this ring pucker, Glu132 would then have to be the catalytic acid for the enzyme to invert product conformation, and hence Asp275 or Glu435 would be the catalytic base.\(^5,7\) Since Glu132 is too far away from the glycosidic oxygen atom for direct attack, the water molecule W195 would have to mediate proton donation, suggesting an unusual catalytic mechanism for this enzyme.\(^7\)

We have used automated docking of ligands into the active sites of glycosyl hydrolases and the carbohydrate-binding site of surfactant protein D\(^9,15\) with AutoDock\(^16,18\) to predict different bound ligand conformations. Therefore we thought that it would be possible to use computation...
to supplement the available experimental knowledge of the active-site function and hydrolysis mechanism of Family 47 α-1,2-mannosidases. In the present study, DMJ and KIF in the $^{1}C_{4}$ conformation were first docked to validate that AutoDock would yield the same bound structures as found by x-ray crystallography. Next DMJ, 1-deoxynojirimycin (DNJ), α-D-galactopyranose (Gal), α-D-glucopyranose (Glc), and α-D-mannopyranose (Man) (Fig. 2) were docked both in the $^{4}C_{1}$ conformation, favored by these molecules in solution, as well as in the $^{1}C_{4}$ conformation to compare the binding energetics of the two conformations with the active site. To understand the nature of transition-state stabilization in α-1,2-mannosidase, the disaccharides α-D-galactopyranosyl-(1→2)-α-D-mannopyranose (Gal-1,2-Man), α-D-glucopyranosyl-(1→2)-α-D-mannopyranose (Glc-1,2-Man), and α-D-mannopyranosyl-(1→2)-α-D-mannopyranose (Man-1,2-Man) (Fig. 2) were docked into the active site. Each of these ligands was docked two ways, with the non-reducing-end residue either in the $^{4}C_{1}$ or the $E_{4}$ (representing the transition state) conformations.

**COMPUTATIONAL METHODS**

**Automated Docking**

AutoDock employs a Lamarckian genetic algorithm (LGA) to perform a search of the conformational space of the ligand. In the AutoDock implementation of the genetic algorithm, the genes are a string of real values representing the three cartesian coordinates for the ligand translation, four variables for the quaternion defining the ligand orientation, and one real value for each ligand torsion, in that order.

The LGA uses a local search algorithm from Solis and Wets (SW) after the global search performed by the genetic algorithm, and the local search results are inherited by the offspring. This search algorithm is adaptive; the step size is modified depending upon the recent energy history. User-defined numbers of consecutive increases or decreases in the energies cause the step size to be doubled or halved, respectively. A slightly modified version of the SW method that allows different step sizes for different genes has also been implemented in AutoDock.

The yeast α-1,2-mannosidase crystal structure (PDB 1DL2) was chosen for this docking study. All hydrogen atoms in both protein and ligands were explicitly modeled, with polar hydrogen atoms being assigned Lennard-Jones 12-10 hydrogen bonding parameters and nonpolar hydrogen atoms being assigned 12-6 parameters. Hydrogen atoms were added to structure 1DL2 followed by a small minimization run to optimize their positions using CHARMM. All water molecules were removed while docking. Partial charges were assigned to the protein atoms using all-atom charges of the AMBER force field. Atomic solvation parameters and fragmental atom volumes were added using the AddSol program provided in the AutoDock 3.06 suite.

In the testing phase of the study, DMJ and KIF structures were first isolated from the crystal coordinates of human α-1,2-mannosidase. Hydrogen atoms were added using Babel and partial charges for inhibitors and mannosyl substrates were generated using MOPAC. Rotatable ligand bonds were defined using the AutoTors module of AutoDock.

The grid maps for van der Waals and electrostatic energies were prepared using AutoGrid version 3.0, with 101 X 101 X 101 points spaced at 0.375-Å distances. The grid was centered on the Ca$^{2+}$ ion at the base of the active site. AMBER force-field parameters were used for evaluating nonbonded interaction energies. The Ca$^{2+}$ parameters were the same as those used by Allen et al. Electrostatic interactions were evaluated using a distance-dependent dielectric constant to model solvent effects.

For the LGA, the size of the initial random population was 50 individuals, the maximal number of energy evaluations was $1.5 \times 10^6$, the maximal number of generations was 80, the number
of top individuals that survived into the next generation, the elitism, was 1, the probability that a gene would undergo a random change was 0.02, the crossover probability was 0.80, and the average of the worst energy was calculated over a window of ten generations.

The pseudo-SW method was used for local searches. There were a maximum of 300 iterations per local search, the probability of performing a local search on an individual was 1.0, the maximal number of consecutive successes or failures before doubling or halving the step size of the local search was 4, and the lower bound on the step size, 0.01, was the termination criteria for the local search. A total of 100 dockings were performed in each docking run. In analyzing the docked conformations, the clustering tolerance of the root mean square positional deviation was 1.0 Å. The crystal coordinates of DMJ and KIF were references for their docking. For disaccharide dockings, crystal coordinates of DMJ and M7 (Fig. 1) served that purpose.

The involvement of two crystal-structure water molecules, W54 and W195, in the catalytic mechanism for nucleophilic attack and mediation of proton donation, respectively, has been suggested. \(^7\) It is likely that these waters would be displaced upon substrate entry, so they were optimized using the local search algorithm of AutoDock. The minimization parameters were the same as those used for the LGA local search.

The main aim of this docking study was to identify the catalytic residues by determining the bound conformation of a mannosyl residue in the tube of the active-site funnel, so a local search of the conformational space inside this tube was needed. However, a global search was expected to yield interesting results also. Therefore, to suit our specific docking needs we increased the local search character of the LGA by making the probability of local search 1.0, and reduced the global search character by keeping the maximal number of generations over which the genetic algorithm is looped to 80. Also, to reduce the search of meaningless conformational space, the initial conformation of the docked ligand was placed in the active-site funnel tube by superimposing it on the DMJ or KIF crystal coordinates.

All docking jobs were run on an SGI Origin 2000 with a 300-MHz MIPS R12000 processor and 1 GB of memory running IRIX 6.5.

RESULTS

Automated Docking

Validation of the docking procedure with inhibitors

The first step was to validate our method by docking DMJ and KIF in the active site of the yeast crystal structure to check for agreement with the observed structures. Since the two enzyme active sites are practically the same, human α-1,2-mannosidase was superimposed on the yeast structure to obtain the corresponding positions of KIF and DMJ in the yeast enzyme, and these coordinates were used to compare the docked structures. The lowest-energy docked structure of KIF-\(^1\)C\(_4\) (KIF in the \(^1\)C\(_4\) conformation) has a final docked energy of \(-107.0\) kcal/mol; it docked with a root mean square deviation (RMSD) of 0.72 Å to the crystal structure (Table I). Similarly, docking of DMJ-\(^1\)C\(_4\) gave a final docked energy of \(-95.0\) kcal/mol and an RMSD of 0.62 Å from the crystal structure (Table I). The agreement was sufficiently good to proceed with the docking of other molecules into the enzyme active site.

DNJ, a Glc analogue, does not inhibit α-1,2-mannosidase. \(^7\) As a negative control, DNJ-\(^1\)C\(_4\) was docked into the enzyme active site to compare its docking energy with that of DMJ. The lowest-energy docked structure of DNJ-\(^1\)C\(_4\) has a total docking energy of \(-88.5\) kcal/mol with a 1.31-Å RMSD with the crystal structure of DMJ-\(^1\)C\(_4\). This energy is higher than the docking
energy of DMJ-\(^1\)C\(_4\) and apparently is not sufficient to compensate for the loss of entropy and solvation energy incurred by it upon binding, i.e. its free energy of binding is not negative.

**Monosaccharide docking simulations**

The lowest-energy docked structure of Man-\(^1\)C\(_4\), with an energy of -93.0 kcal/mol, does not dock in the active-site funnel tube. Instead, it docks just outside the funnel neck away from M7. This structure hydrogen-bonds Asp61, Trp82, Arg136, and Glu497. Of these, only Arg136 is conserved, implying that this binding site is not unique across the \(\alpha\)-1,2-mannosidase family. However, the second lowest-energy cluster docks very close to the crystal-structure DMJ-\(^1\)C\(_4\), with a RSMD of 0.82 Å and an energy of -84.3 kcal/mol (Table I). The lowest-energy cluster of Man-\(^4\)C\(_1\) docks close to M7 with an energy of -94.6 kcal/mol. The second lowest-energy cluster docks in the funnel tube with an energy of -88.5 kcal/mol and a RMSD of 1.87. Glc and Gal, C2 and C4 epimers of Man, respectively, bind with higher energies than DMJ-\(^1\)C\(_4\).

**Disaccharide docking simulations**

The \(E_4\) conformation of Man-1,2-Man docks with a lower energy than the \(^1\)C\(_4\) conformation; this is not the case in the docking simulations of Gal-1,2-Man and Glc-1,2-Man (Table I). Only Man-\(E_4\)-1,2-Man can closely overlay the crystal structures of both DMJ and M7 in the active site, as is evident from their low RMSD values (Table I).

**TABLE I. Inhibitor, Monosaccharide, and Disaccharide Docking**

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Conformation</th>
<th>No. of major clusters</th>
<th>Lowest energy (kcal/mol)</th>
<th>RMSD of lowest energy (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMJ</td>
<td>(^1)C(_4)</td>
<td>2</td>
<td>1</td>
<td>-95.00</td>
</tr>
<tr>
<td>DMJ</td>
<td>(^4)C(_1)</td>
<td>9</td>
<td>6</td>
<td>-77.70</td>
</tr>
<tr>
<td>DNJ</td>
<td>(^1)C(_4)</td>
<td>4</td>
<td>1</td>
<td>-88.48</td>
</tr>
<tr>
<td>DNJ</td>
<td>(^4)C(_1)</td>
<td>6</td>
<td>1</td>
<td>-91.10</td>
</tr>
<tr>
<td>Kif</td>
<td>(^1)C(_4)</td>
<td>1</td>
<td>1</td>
<td>-107.00</td>
</tr>
<tr>
<td>Gal</td>
<td>(^1)C(_4)</td>
<td>5</td>
<td>1</td>
<td>-86.13</td>
</tr>
<tr>
<td>Gal</td>
<td>(^4)C(_1)</td>
<td>4</td>
<td>2</td>
<td>-91.71</td>
</tr>
<tr>
<td>Glc</td>
<td>(^1)C(_4)</td>
<td>6</td>
<td>1</td>
<td>-85.44</td>
</tr>
<tr>
<td>Glc</td>
<td>(^4)C(_1)</td>
<td>5</td>
<td>2</td>
<td>-80.41</td>
</tr>
<tr>
<td>Man</td>
<td>(^1)C(_4)</td>
<td>5</td>
<td>2</td>
<td>-84.25</td>
</tr>
<tr>
<td>Man</td>
<td>(^4)C(_1)</td>
<td>9</td>
<td>2</td>
<td>-88.51</td>
</tr>
<tr>
<td>Gal-1,2-Man</td>
<td>(^1)C(_4)</td>
<td>3</td>
<td>1</td>
<td>-141.00</td>
</tr>
<tr>
<td>Gal-1,2-Man</td>
<td>(E_4)</td>
<td>3</td>
<td>1</td>
<td>-137.40</td>
</tr>
<tr>
<td>Glc-1,2-Man</td>
<td>(^1)C(_4)</td>
<td>3</td>
<td>1</td>
<td>-146.79</td>
</tr>
<tr>
<td>Glc-1,2-Man</td>
<td>(E_4)</td>
<td>4</td>
<td>1</td>
<td>-137.24</td>
</tr>
<tr>
<td>Man-1,2-Man</td>
<td>(^1)C(_4)</td>
<td>4</td>
<td>1</td>
<td>-136.75</td>
</tr>
<tr>
<td>Man-1,2-Man</td>
<td>(E_4)</td>
<td>3</td>
<td>1</td>
<td>-153.03</td>
</tr>
</tbody>
</table>

* Lower-energy clusters that dock outside the active-site funnel tube are ranked but not reported.

\(^1\)A RMSD cluster tolerance, based on the crystallographic coordinates of DMJ (inhibitors and monosaccharides) or DMJ and M7 (disaccharides).
DISCUSSION

Automated Docking of Inhibitors and Monosaccharides

The very negative docking energies of DMJ-1C4 and KIF-1C4 explain their inhibitory action. The total docked energy in AutoDock is the sum of the total nonbonded intermolecular interaction energy between every ligand atom and the macro­ molecule and the nonbonded intramolecular energy of the ligand. An interesting point to be noted in these docked structures is the very high interaction energies of the C2 and C3 hydroxyl groups of the inhibitors with the enzyme (Fig. 3). In the lowest-energy docked structure of DMJ, they contribute -25.0 kcal/mol, almost 27.5% of the total interaction energy and 26.3% of the total docking energy of DMJ with the α-1,2-mannosidase. Equivalent values with KIF are -22.7 kcal/mol, 22.7%, and 21.2%. Both these hydroxyl groups are coordinated by the Ca²⁺ ion present at the base of the active-site funnel tube. Thus, the Ca²⁺ ion plays a very important role in stabilizing the energetically unfavorable 1C4 conformations of these inhibitors. Another point to be noted is that the activation energy for the transformation of mannose from the 4C1 to the 1C4 conformation is 12–14 kcal/mol, thus making the docked 1C4 conformation more stable. DNJ-1C4 and DNJ-4C1 dock with higher energies than the docked energy of DMJ-1C4, explaining why DNJ is not an inhibitor.

Enzymes function by stabilizing the transition state; however, for efficient function the active site should not favorably bind the product. The difference in the binding energies of DMJ-1C4 and Man-1C4 clearly demonstrates how enzymes achieve this kind of specificity. DMJ-1C4 binds the active site with very negative interaction energy, whereas Man-1C4 does not. The extra hydroxyl group at the mannosyl C1 causes a steric hindrance that is responsible for the reduced interaction energy. The enzyme active site is optimized to bind the E4 conformation of -1Man. Man-E4 in solution has an apparent free energy of ~10 kcal/mol greater than that of the Man-1C4. Before bond breakage the interactions provided by the rest of the substrate stabilize the E4 conformation and, as suggested earlier, might even induce the strain necessary for its deformation. After the bond breaks, the cleaved mannose probably leaves the funnel tube in an effort to relax back into an energetically more favorable conformation.

Man₉GlcNAc₂, the hydrolysis product of Family 47 ER α-1,2-mannosidases, is a necessary signal for degradation of misfolded protein. α-1,2-Mannosidase inhibitors block degradation of misfolded proteins, and it has been suggested that this slow-acting ER α-1,2-mannosidase (Kₘ = 0.5 mM, kₗ = 12 s⁻¹ for the S. cerevisiae enzyme) may work as a timer for glycoprotein degradation. This ER α-1,2-mannosidase therefore is the first enzyme in the N-glycan synthesis pathway that binds the newly synthesized protein in fully/nearly folded form in the cell under normal conditions.

The difference in apparent free energy between Man-1C4 and Man-4C1 in solution is 4.38 kcal/mol, which implies that the Man-1C4 solution concentration is <1%. Since α-1,2-mannosidase binds the 1C4 conformation, the corrected Kₘ for this enzyme would therefore be >0.005
mM and \( k_{cat}/K_M \) would be \( >2.4 \times 10^6 \text{M}^{-1}\text{s}^{-1} \), giving this enzyme very high affinity for its substrate. Also, the relaxation time for the conformational change between the two chair forms is of the order of microseconds.\(^{31}\) Protein folding itself takes milliseconds to minutes. The rate-limiting step, therefore, seems to be the diffusion of the bulky substrate to the \( \alpha\)1,2-mannosidase active site. Also, the lower concentration of the \( ^1\text{C}_4 \) conformation might aid in retention. Misfolded proteins would have greater resistance to diffusion and would therefore be retained in the ER longer as the ER resident chaperonins aid the folding process. The compact folded form of the enzyme would be able to move more easily to the ER \( \alpha\)1,2-mannosidase active site where its middle-arm mannose is trimmed. The MansGlcNAc\textsubscript{2} formed by hydrolysis by this enzyme makes it amenable to be marked for degradation by the ER glucosyltransferases.\(^{28}\) However, if correctly folded, due to its lower diffusion resistance compared to the unfolded form, it can migrate quickly enough out to the ER to avoid being marked for degradation. In this scenario, the retention time of the MansGlcNAc\textsubscript{2}-protein seems to be the most important step in ER-associated protein degradation.

There is also the possibility that the transformation of the terminal mannosyl residue from the \( ^4\text{C}_1 \) to the \( ^1\text{C}_4 \) conformation is mediated by interactions with the enzyme, but this seems less likely considering the high activation energy for this transformation.

### Transition State

Ring flattening at the C1 position of DMJ, leading to the \( E_4 \) conformation, was suggested by Vallée \textit{et al.}\(^7\) so that it could form an \( \alpha\)1,2 bond with M7. This would also be consistent with the absolute stereochemical requirement of the coplanarity of C1, C2, C5, and O5 atoms of the nonreducing-end residue necessary for forming the oxocarbenium-ion-like transition state.\(^5\) The disaccharides Gal-1,2-Man, Glc-1,2-Man, and Man-1,2-Man were docked with the nonreducing-end residues in their \( ^1\text{C}_4 \) and \( E_4 \) conformations. In each case the reducing-end mannosyl residue was in the relaxed \( ^4\text{C}_1 \) conformation. The disaccharide with the nonreducing end in the \( E_4 \) conformation would therefore mimic the substrate transition state.

The apparent free energy of Man-\( E_4 \) in solution is \(-10 \text{kcal/mol} \) higher that that of Man-\( ^1\text{C}_4 \).\(^{26}\) Therefore, for transition-state stabilization to occur, the decrease in energy upon binding should overcome the difference in energy between the two conformations. The lower energy of the docked Man-\( E_4 \)-1,2-Man compared to that of Man-\( ^1\text{C}_4 \)-1,2-Man clearly establishes that docked Man-\( E_4 \)-1,2-Man well approximates the enzyme-transition state complex. Energies of the two other disaccharide pairs are higher for their \( E_4 \) than for their \( ^1\text{C}_4 \) conformations. This implies that even if Gal-1,2-Man and Glc-1,2-Man penetrate the active site, their \( ^1\text{C}_4 \) conformations will not deform, and hence there will be no substrate activation for hydrolysis.

An examination of the interaction energies of the various substrate hydroxyl groups indicates that three, O3 and O6 of \(-1\text{Man} \) and O4 of \(+1\text{Man} \), form strong hydrogen bonds with the enzyme (interaction energies of 12.0, 12.2, and 13.8 \text{kcal/mol}, respectively). This implies that the interactions provided by the hydroxyl groups further away from the scissile bond also contribute significantly to transition-state stabilization. These bonds might also help generate the torque necessary for \(-1\text{Man} \) deformation.

![Fig. 4. Stabilizing interactions of \(+1\text{Man} \) (white). Conserved residues interacting with \(+1\text{Man} \) (gray) and M7 (black). Hydrogen bonds and hydrophobic interactions are shown as lines.](image-url)
The conserved residues of the active-site neck that bind +1Man are shown in Fig. 4. Conserved residue Asp275 hydrogen-bonds O3 and O4 of +1Man, thus securely anchoring the substrate for the reaction to occur. These hydrogen bonds cannot form if -1Man is linked by α-1,3, α-1,4, or α-1,6 bonds to +1Man, thus explaining its specificity for α-1,2-linked mannosyl residues. This might explain why the D275N mutant of the S. cerevisiae enzyme has very low α-1,2-mannosidase activity compared to the wild-type enzyme.25

Hydrophobic packing interactions between conserved aromatic residue Phe131 and the C4, C5, and C6 atoms of +1Man and between conserved Leu338 and the C2 and C3 atoms of +1Man also seem to have an important role in stabilizing +1Man (Fig. 4). For a negative-binding free energy, the interaction energy between the ligand and enzyme should compensate for entropy loss and solvation energy loss due to binding. The active site, being rich in acidic amino acid residues, can replace the hydrogen bonding network supplied by water. Binding to the enzyme active site also supplies a hydrophobic environment for the nonpolar carbon atoms of the ligand, and this could be key in causing the negative binding free energy.

Catalytic Mechanism

Identification of the catalytic acid/proton donor and base/nucleophile from the human and yeast mannosidase crystal structures was somewhat ambiguous.5,7 Docking of Man-4-1,2-Man and the optimal positions of W54 and W195 after minimization clearly establish their identities: Glu132 is the catalytic proton donor and Glu435 is the catalytic base (Fig. 5). Upon optimization, W54 is 0.7 Å and W195 is 0.8 Å from their corresponding crystal coordinates. As suggested earlier,5 W195 appears to mediate proton donation, since Glu132 is not within hydrogen-bonding distance of the O2 atom of M7. Another possibility could be the movement of Glu132 within hydrogen-bonding distance of the O2 atom. However, this does not seem likely because Glu132 is anchored in place by a salt bridge with Arg136 that restricts its motion. The water molecule activated by the nucleophile is W54, the only crystal-structure water that is located close enough to the C1 of -1Man for nucleophilic attack. Asp275 is not within hydrogen-bonding distance of W54 and hence is ruled out as the catalytic base (Fig. 5). W54 also is coordinated by the Ca2+ ion, hence Ca2+ has a more direct role in catalysis than stabilizing the 1C4 conformation of -1Man.

CONCLUSIONS

The detailed energetic information supplied by AutoDock significantly adds to the structural information obtained by x-ray crystallography and NMR. This information, upon careful interpretation, can contribute to understanding biomolecular recognition mechanisms. It is evident that, in the post-genomic era, when protein structural information will be available at genomic scale, detailed study of interactions at the atomic level would be invaluable for developing methods to determine a protein’s structure from its function.

ACKNOWLEDGMENTS

The authors are grateful to Pedro M. Coutinho for suggesting this project.
REFERENCES


25. Lipari F, Herscovics A. Calcium binding to the Class I α-1,2-mannosidase from Saccharomyces cerevisiae occurs outside the EF hand motif. Biochemistry 1999;38:1111–1118.
A Mathematical Model for Carbon Bond Labeling Experiments: Analytical Solutions and Sensitivity Analysis for the Effect of Reaction Reversibilities on Estimated Fluxes

Ganesh Sriram and Jacqueline V. Shanks
Department of Chemical Engineering
2114, Sweeney Hall, Iowa State University, Ames, IA 50011

Summary
Carbon labeling experiments provide valuable information toward the quantification of intracellular metabolic fluxes. A class of $^{13}$C labeling experiments, known as carbon bond labeling experiments, provides information on the extent of coupling between adjacent carbon atoms (i.e. 'bond integrities' of carbon-carbon bonds) in metabolite molecules. From these, it is possible to calculate the abundances of bondmers (molecules of the same metabolite having different C–C bond integrities). The bondmer abundances can be processed and translated to valuable metabolic flux maps. In this paper, we have developed a mathematical model based on metabolite and bondmer balances, for describing bondmer abundances of metabolites in the glycolysis and pentose phosphate pathways as a function of the flux split ratio at the glucose-6-phosphate branchpoint, and the reversibilities of the reactions catalyzed by hexose isomerase and transketolases. We show that in this case, analytical solutions can be obtained for groups of bondmers. Also, we report a sensitivity analysis (based on these solutions), of the effect of the reversibilities on the bondmer distributions.

Introduction
Metabolic flux analysis is the quantification of all steady state intracellular metabolic fluxes (i.e. biochemical reaction rates), in a metabolic pathway network, and has long been identified as a valuable computational tool in metabolic engineering (Stephanopoulos and Vallino 1991, Stephanopoulos 1994, 1999). Metabolic flux maps, resulting from these analyses, provide a measure of the degree of engagement of various pathways in cellular metabolism (Stephanopoulos et al. 1998) and thus provide a description of the physiological state of the cell (Schmidt et al. 1999). Therefore, metabolic flux analysis of genetic variants can be useful in deriving information about the effect of the genetic variation on the biochemistry of the cell (Bailey 1991, Nielsen 1998).

Metabolic flux analysis is typically performed by writing balances for intracellular metabolites in matrix form, by measuring the rates of changes in concentration of certain metabolites (such as substrate, extracellular products), and then solving the matrix equation obtained (Vallino and Stephanopoulos 1993). However, in large and biochemical networks (such as those of many eukaryotes) that are topologically complex as well as possibly compartmented, there is an inherent lack of measurable metabolites related to the intracellular flux distribution (Schmidt et al. 1999). This problem can be overcome by using isotopically labeled substrates, and subsequent analysis by nuclear magnetic resonance (NMR) spectroscopy or mass spectroscopy (MS). NMR and MS provide flux distributions at branchpoints in the network, which can be used as additional contraints in the solution of the aforementioned matrix equation.

1Author for correspondence, E-mail jshanks@iastate.edu
NMR (and MS) experiments involve the feeding of a isotopically labeled ($^{13}\text{C}$) substrate, and relating the fate of the label to the fluxes in the pathways. The ‘fate’ of the label could be (a) the extent of $^{13}\text{C}$ enrichment of carbon atoms in a metabolite molecule (if a non-uniformly labeled $^{13}\text{C}$ substrate is used) or (b) the extent of coupling between adjacent carbon atoms in a molecule (or ‘bond integrities’ of carbon–carbon bonds in the molecule), if a uniformly labeled $^{13}\text{C}$ substrate is used (Schmidt et al. 1999).

In this paper, we concentrate on category (b), also known as carbon bond labeling experiments (Szyperski 1995). Using bond labeling experiments to detect bond-integrities of the amino acids (from the cellular protein of an organism), is a particularly quick, efficient and powerful method of obtaining metabolic flux maps of primary and intermediary metabolism (Szyperski 1995, Sauer et al. 1997, 1999). This is because the carbon skeletons of the amino acids reflect the carbon skeletons of key precursors spread out over primary and intermediary metabolism (Szyperski 1998), which in turn, are related to the metabolic flux distribution. Further, the experimental aspect of this methodology is relatively straightforward. Additionally, the amino acids appearing on a 2-D HSQC NMR spectrum are segregated enough to allow easy quantitation of carbon–carbon bond integrities.

We have developed a mathematical model that uses metabolite and bondmer balances, to describe the bondmer distributions (see ‘Mathematical Model’ section for definition) of the amino acid precursors, as a function of the fluxes and reaction reversibility extents in glycolysis and the pentose phosphate pathway. Prior to this work, Klapa et al. (1999) have applied the concept of metabolite and isotopomer balances to develop a model for fluxes in the TCA cycle, in a carbon atom enrichment experiment (category (a) of experiments, above). Since this work focuses on carbon bond labeling experiments (category (b), a completely different experiment type), our work is novel. Also, we have concentrated on the glycolysis and pentose phosphate pathway reactions (i.e. the section of primary metabolism before pyruvate and acetyl CoA), whereas Klapa et al. have focused on the TCA cycle, which is the section of primary metabolism after pyruvate and acetyl CoA.

Mathematical model for bondmer distributions

Definitions

Before we proceed to describe the development of the model, we formalize the definitions of bond integrity and bondmer, and elucidate the method we have used to assign numbers to bondmers.

Bond integrity. The bond integrity of a covalent bond between two carbon atoms in a metabolite molecule, is a property indicating if those two carbon atoms originated from the same substrate molecule, or not. For a single molecule, bond integrity takes the binary values 1 and 0 (1 if the two carbons originated from the same substrate molecule, and 0 if they did not). For a pool of molecules of the same metabolite, the bond integrity of a certain bond is statistical mean of the bond integrities of that bond in all the metabolite molecules of that pool.

In other words, bond integrity is the probability that the carbon atoms connected by that bond originated from the same source molecule. Thus the bond integrity of a C–C bond in a metabolite pool can take fractional values in the range $[0, 1]$.

Bondmer. Bondmers (bond-isomers) are molecules of the same metabolite which have different bond integrities for different carbon–carbon bonds. A bondmer pool is a collection of all metabolite
Table 1: Illustration of bond integrity, bondmer and bondmer number

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Structure</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>Bondmer number</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1 2-3</td>
<td>0</td>
<td>1</td>
<td>$M_3$</td>
</tr>
<tr>
<td>B</td>
<td>1-2 3</td>
<td>1</td>
<td>0</td>
<td>$M_2$</td>
</tr>
<tr>
<td>C</td>
<td>1-2-3</td>
<td>1</td>
<td>1</td>
<td>$M_4$</td>
</tr>
<tr>
<td>D</td>
<td>1 2-3</td>
<td>0</td>
<td>1</td>
<td>$M_3$</td>
</tr>
<tr>
<td>E</td>
<td>1 2-3</td>
<td>0</td>
<td>1</td>
<td>$M_3$</td>
</tr>
<tr>
<td>F</td>
<td>1-2 3</td>
<td>1</td>
<td>0</td>
<td>$M_2$</td>
</tr>
<tr>
<td>G</td>
<td>1 2 3</td>
<td>0</td>
<td>0</td>
<td>$M_1$</td>
</tr>
<tr>
<td>H</td>
<td>1-2-3</td>
<td>1</td>
<td>1</td>
<td>$M_4$</td>
</tr>
</tbody>
</table>

Mean 0.50 0.63

<table>
<thead>
<tr>
<th>Bondmer</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_1$</td>
<td>$1/8 = 0.125$</td>
</tr>
<tr>
<td>$M_2$</td>
<td>$2/8 = 0.250$</td>
</tr>
<tr>
<td>$M_3$</td>
<td>$3/8 = 0.375$</td>
</tr>
<tr>
<td>$M_4$</td>
<td>$2/8 = 0.250$</td>
</tr>
</tbody>
</table>

Bonds of the same metabolite which have the same bond integrity for each bond. Thus bondmer pools are subsets of metabolite pools.

**Bondmer numbering** The number of bondmers for a given metabolite may sometimes be quite large, and it is therefore necessary to systematically assign numbers to them. We have used the following procedure: Consider a metabolite $M$ having an $n$-carbon skeleton, $12345\ldots n$. We number the C–C bonds in the molecule from left to right, and let $b_i$ be the integrity of the $i$th bond. Thus, $k = \Sigma_{i=1}^{n-1} 2^{i-1}b_i + 1$. Since we are considering a single molecule, its bond integrities will be either 0 or 1. Thus the number calculated above would be an integer. We then designate the abundance of this bondmer pool as $M_k$. (Note that the ‘1’ is added for convenience, so that the lowest-numbered bondmer $M_1$).

**Example**

Consider an experiment where glucose is the sole carbon substrate, and a 3-carbon metabolite, $M$, is formed. Assume, for simplicity, that its metabolite pool consists of the 8 molecules shown in Table 1. Their bond integrities $b_1$ and $b_2$, and the bondmer pool that they belong to, are shown alongside. The overall bond integrities of the C1–C2 and C2–C3 bonds for this metabolite are shown in the last row.

In order to translate NMR data to bondmer abundances, probability expressions, such as those derived by Szyperksi (1995) can be used. Additional details on the same will be available in our forthcoming publication (Sriram and Shanks 2002).
Figure 1: Glycolytic and pentose phosphate pathways, and flux distributions therein

Development of Model
The glycolytic and pentose phosphate pathways, and their flux distributions have been shown in Figure 1. All fluxes are relative to the input glucose flux, which has been taken to be 1. The flux into the pentose phosphate pathway has been designated as $3z$. All other fluxes can be described as a function of $z$, as shown in the figure. This can be calculated using metabolite balances (Sriram and Shanks 2002). Note that the amino acids whose bond integrities/bondmer distributions can be determined using NMR, have been shown inside ovals, and are linked to their metabolite precursors.

Case 1: All reactions are completely irreversible
To exemplify the use of bond integrities in obtaining metabolic flux maps, we have considered the evaluation of the flux through the pentose phosphate pathway, using the bondmer distribution of glyceraldehyde-3-phosphate (GAP). (Note that this information can be experimentally obtained by conducting a bond labeling NMR experiment on the amino acids Ala and Phe.)

Using information on the carbon skeleton rearrangement of the metabolites during the glycolytic/pentose phosphate pathway reaction, [such as that reported by Follstad and Stephanopoulos (1998), and shown in Figure 2], we enumerated all possible bondmers that would be formed in this reaction network (if the reactions were irreversible). As depicted in Figure 3, the only metabolites in this network that form multiple bondmers are fructose-6-phosphate (F6P or F) and glyceraldehyde-3-phosphate (GAP or T), the bondmers being:

$F_{18}: 1-2 3 4 5-6$
$F_{26}: 1-2 3 4-5-6$
$F_{30}: 1-2 3-4-5-6$
$F_{32}: 1-2-3-4-5-6$
Figure 2: Carbon atom transitions in pentose phosphate pathway reactions

\[ \begin{align*}
\text{F}6P & \rightarrow 2 \text{T}3P \\
\text{G}6P & \rightarrow 2 \text{F}6P \\
\text{G}6P & \rightarrow \text{F}6P
\end{align*} \]

\[ \begin{align*}
\text{P}5P & + \text{P}5P \rightarrow \text{F}6P + \text{P}5P \\
\text{F}6P & \rightarrow \text{S}7P \\
\text{S}7P & + \text{F}6P \rightarrow \text{P}5P + \text{S}7P
\end{align*} \]

\[ \begin{align*}
\text{G}6P & \rightarrow \text{P}5P + \text{CO}_2 \\
\text{T}3P & \rightarrow \text{E}4P + \text{F}6P \\
\text{T}3P & \rightarrow \text{F}6P + \text{E}4P
\end{align*} \]

To obtain these abundances as a function of \( z \), input-output balances can be written for the F6P and G6P bondmers, based on the bondmer formation/depletion depicted in Figure 3:

**F6P balance:**

\[
\frac{d}{dt} F = \frac{d}{dt} \begin{bmatrix} F_{18} \\ F_{26} \\ F_{30} \\ F_{32} \end{bmatrix} = \begin{bmatrix} T_3 & T_4 & 0 & 0 \\ T_4 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} + (1 - 3z) \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} - (1 - z) \begin{bmatrix} F_{18} \\ F_{26} \\ F_{30} \\ F_{32} \end{bmatrix}
\]

**GAP balance:**

\[
\frac{d}{dt} \text{GAP} = \frac{d}{dt} \begin{bmatrix} T_3 \\ T_4 \end{bmatrix} = \begin{bmatrix} 2(1 - z)F_{26} + (1 - z)F_{18} + (1 - z)F_{30} \\ 2(1 - z)F_{32} + (1 - z)F_{18} + (1 - z)F_{30} \end{bmatrix} + z \begin{bmatrix} 0 \\ 0 \end{bmatrix} + z \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\]
Figure 3: Enumeration of bondmers in the glycolysis and pentose phosphate pathways when all reactions are irreversible

Furthermore, metabolic and isotopic steady can be assumed, so that the rates of accumulation of the bondmers can be set to zero:

\[
\frac{dF}{dt} = 0
\]

\[
\frac{dT}{dt} = 0
\]

Thus, from the F6P balance, we have

\[
(1 - z) \begin{bmatrix} F_{18} \\ F_{25} \\ F_{30} \\ F_{32} \end{bmatrix} = z \begin{bmatrix} T_3 \\ T_4 \end{bmatrix} + (1 - 3z) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + z \begin{bmatrix} 0 \\ 1 \end{bmatrix}
\]

Combining this with the GAP balance gives

\[
2 \begin{bmatrix} T_3 \\ T_4 \end{bmatrix} = \begin{bmatrix} 2(1 - 3z) + zT_4 + z + 2z \\ 2zT_3 + zT_4 + z \end{bmatrix}
\]

This gives 2 simultaneous equations in \( T_3 \) and \( T_4 \), on solving which we obtain

\[
T_3 = \frac{2z}{2 - z}
\]
and

\[ T_4 = \frac{2 - 3z}{2 - z} \]  \hspace{1cm} (6)

Eqs. 5 and 6 provide the bondmer distribution of GAP in terms of the relative flux through the pentose phosphate pathway (3z). This demonstrates how the bondmer abundances, obtained from NMR data, can be used to evaluate metabolic fluxes.

Case 2: \textit{pgi, tktA} and \textit{tktB} reactions are reversible

Effect of \textit{tktA} reversibility

Reversible biochemical reactions can lead to a redistribution of the $^{13}$C label, and lead to label fates substantially different from the irreversible reactions scenario examined in the above section. We examined the effects of reversibilities of individual reactions in the glycolysis/PPP pathways. The first of the reactions examined was transketolase (\textit{tktA}) reaction. Szyperksi (1995) suggested the reversibility of this reaction to be the cause of formation of bondmers of pentose phosphate (P5P) with $b_2 = 0$, a class of bondmers not encountered in the previous section, where the reversibility of \textit{tktA} was not considered. Here, we present a systematic analysis of effect the reversibility on the abundance of such bondmers.

The reaction being examined is

\[ P5P + P5P \xrightarrow{\text{tktA}} S7P + GAP \]

Here, $z$ is the net flux of the reaction (from Figure 1), and $r$ is the flux of the backward reaction. The reaction reversibility extent, $e_A$, is defined as the ratio of the backflux to the total flux,

\[ e_A = \frac{r}{r + z} \]  \hspace{1cm} (7)

From the carbon atom transitions shown in Figure 2, it can be inferred that a pentose phosphate (P5P) molecule with unbroken C–C bonds will form the bondmers $S_{62}$ and $T_4$, of $S7P$ and GAP, respectively,

\[ P_{16} + P_{16} \xrightarrow{r \pm z} S_{62} + T_4 \]  \hspace{1cm} (8)

If the reaction is reversible, these bondmers will react to form a pentose phosphate molecule with a broken C2-C3 bond,

\[ S_{62} + T_4 \xrightarrow{r} P_{16} + P_{14} \]  \hspace{1cm} (9)

The $P_{14}$ molecules in the pentose phosphate pool can now participate in the forward reaction, either by themselves or with the $P_{16}$ molecules, giving $S_{54}$ and $S_{62}$ molecules,

\[ P_{14} + P_{14} \xrightarrow{r \pm z} S_{54} + T_4 \]  \hspace{1cm} (10)

\[ P_{14} + P_{16} \xrightarrow{r \pm z} \frac{1}{2} S_{54} + \frac{1}{2} S_{62} + T_4 \]  \hspace{1cm} (11)

\[ \]
The $S_{54}$ molecules can undergo the reverse reaction to give only $P_{14}$ molecules,

$$S_{54} + T_4 \xrightarrow{r} P_{14} + P_{14} \quad (13)$$

This completes the enumeration of the bondmers formed as a result of $tktA$ reversibility. We can now write bondmer balances in vector form, thus,

$$\frac{dP}{dt} = \frac{d}{dt} \begin{bmatrix} P_{14} \\ P_{16} \end{bmatrix} = 3z \begin{bmatrix} 0 \\ 1 \end{bmatrix} + r \begin{bmatrix} -P_{16} \\ P_{16} \end{bmatrix} - z \begin{bmatrix} P_{14}(P_{14} + P_{16}) \\ P_{15}(P_{14} + P_{16}) \end{bmatrix} - 2z \begin{bmatrix} P_{14}(P_{14} + P_{16}) \\ P_{16}(P_{14} + P_{16}) \end{bmatrix} \quad (14)$$

Since $P_{14}$ and $P_{16}$ are the only bondmers of P5P in this system, we have $P_{14} + P_{16} = 1$, and therefore, at steady state,

$$0 = -3zP_{14} + rP_{16} \quad (15)$$

so that $P_{14}/P_{16} = 3z/r$, and using the fact that $P_{14} + P_{16} = 1$, the P5P bondmer abundances are

$$P_{14} = \frac{r}{3z + r} = \frac{e_A}{3 - 2e_A}$$
$$P_{16} = \frac{3z}{3 + r} = \frac{3 - 3e_A}{3 - 2e_A}$$

Here, $P_{14}$ represents the class of P5P bondmers which have $b_2 = 0$. The above equations provide the relationship between their abundance and the extent of reversibility of the $tktA$ reaction.

**Effect of $pgi$ reversibility**

We analyzed the effect of $pgi$ reversibility, by a procedure similar to that in the previous two sections: enumerating bondmers, writing and solving bondmer balances. The number of bondmers of in this case is substantially larger than in the foregoing cases, and is not elaborated here for lack of space. Details can be obtained in Sriram and Shanks (2002). Table 2 lists the abundances of selected bondmers (occasionnally groups of bondmers), when the $pgi$ reaction is reversible. The reversibility extent of $pgi$ is denoted as $e$. Two different sets of results have been presented: (1) $pgi$ partially reversible ($0 < e < 1$) and (2) the special case of $pgi$ being completely reversible ($e = 1$), which results in complete scrambling of the G6P and F6P bondmer pools.

**Combined effect of $pgi$ and $tktA$ and reversibilities**

We analyzed the combined effects of reversibilities of $pgi$ and $tktA$ reactions, in a manner similar to that above. The results for selected bondmers are presented in Table 3. Here, the definitions of $e$ and $e_A$ are the same as in the above sections.

**Sensitivity Analysis**

Based on the analytical solutions obtained for the abundances of various bondmers, we conducted sensitivity analyses for the effect of various reaction reversibilities on the evaluated fluxes. Although the presentation of detailed analysis is not possible here, we report a summary of the analysis results: the bondmer distribution of GAP is insensitive to the reversibility of $tktA$ ($e_A$) and relatively much
<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Bondmer</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAP</td>
<td>$T_1 + T_3$</td>
<td>1 2 3 + 1 2-3</td>
</tr>
<tr>
<td></td>
<td>$T_4$</td>
<td>1-2-3</td>
</tr>
<tr>
<td>Pen</td>
<td>$P_1 + P_9$</td>
<td>1 2 3 4 5 + 1 2 3 4-5</td>
</tr>
<tr>
<td></td>
<td>$P_{13}$</td>
<td>1 2-3-4-5</td>
</tr>
<tr>
<td></td>
<td>$P_{15}$</td>
<td>1 2-3-4-5</td>
</tr>
<tr>
<td></td>
<td>$P_{16}$</td>
<td>1-2-3-4-5</td>
</tr>
<tr>
<td>E4P</td>
<td>$E_1 + E_5$</td>
<td>1 2 3 4 + 1 2 3-4</td>
</tr>
<tr>
<td></td>
<td>$E_7$</td>
<td>1 2-3-4</td>
</tr>
<tr>
<td></td>
<td>$E_8$</td>
<td>1-2-3-4</td>
</tr>
<tr>
<td>$p$ = &amp;frac2z(1-ez)/(2-z)(1+2ez) &amp;frac2z(1-z)/(2-z)(1+2z) &amp;frac2z(1-ez)/(2-z)(1+2ez) &amp;frac2z(1-z)/(2-z)(1+2z)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q = 1 - p$</td>
<td>&amp;frac2z(1-z)/(2-z)(1+2z) &amp;frac2z(1-ez)/(2-z)(1+2ez) &amp;frac2z(1-z)/(2-z)(1+2z) &amp;frac2z(1-ez)/(2-z)(1+2ez)</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: Effect of *pgi* reversibility on bondmer distributions

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Bondmer</th>
<th>Abundance</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAP</td>
<td>$G_1 + G_3$</td>
<td>1 2 3 + 1 2-3</td>
</tr>
<tr>
<td>Pen</td>
<td>$P_1 + P_9$</td>
<td>1 2 3 4 5 + 1 2 3 4-5</td>
</tr>
<tr>
<td></td>
<td>$P_{15}$</td>
<td>1 2-3-4-5</td>
</tr>
<tr>
<td></td>
<td>$P_{16}$</td>
<td>1-2-3-4-5</td>
</tr>
<tr>
<td>$p$ = &amp;frac2z(1-ez)(2e_Aez+e_A+ez)/(1+2ez)[(2-z)</td>
<td>(1+ez)+2e_A(2+ez)]-6e_A</td>
<td>&amp;frac2z(1-ez)(2e_Aez+e_A+ez)/(1+2ez)[(2-z)</td>
</tr>
<tr>
<td>$q = 1 - p$</td>
<td>&amp;frac2z(1-ez)(2e_Aez+e_A+ez)/(1+2ez)[(2-z)</td>
<td>(1+ez)+2e_A(2+ez)]-6e_A</td>
</tr>
</tbody>
</table>

Table 3: Combined effect of *pgi* and *tktA* reversibilities
sensitive to the reversibility of pgI (e). On the other hand, the bondmer distribution of the P5P pool is highly sensitive to the reversibility of tktA (eA), and relatively less sensitive to the reversibility of pgI (e). Also, in a carbon bond labeling experiment, it is crucial to accurately quantify the bondmer abundances of P5P (from histidine) and E4P (from the aromatic skeleton of phenylalanine and tyrosine). Information from the modeling and sensitivity analysis is expected to be useful in obtaining how significantly each measurement in a carbon bond labeling experiment contributes to the estimated fluxes, and is therefore expected to facilitate better experimental design, as well as faster and more accurate computation of fluxes.

References


Platelet Derived Nitric Oxide (NO) Inhibits Thrombus Formation: The Role of Insulin

R. H. Williams and M. U. Nollert
Department of Chemical Engineering and Materials Science
University of Oklahoma, Norman, OK 73019

ABSTRACT

Platelet adhesion was simulated an in vitro parallel plate flow chamber using blood anticoagulated with 20 U/mL HEPARIN. Blood was perfused over collagen type III protein coated slides (8 mg/mL) on the surface of the chamber for 5-10 minutes over a physiological range of shear (100 – 2000 s\(^{-1}\)). After 5 minutes the % platelet coverage decreased from 74 (+/- 9) % at 100 s\(^{-1}\) to 12 (+/- 8) % at 2000 s\(^{-1}\). The effect of insulin on platelet coverage was investigated by perfusing blood with different additives through a parallel plate flow chamber for 5 minutes at physiological shear (100, 1000, and 2000 s\(^{-1}\)). The following samples were tested: control (no additive), LNAME, which is an inhibitor of NO production, insulin, a possible promoter of NO synthesis, and LNAME/insulin. At 100 s\(^{-1}\), there wasn't a statistical difference in % coverage. However, at 1000 s\(^{-1}\), the % coverage between insulin and the other samples was statistically significant. After 2 minutes the % coverage for insulin was 26.43 (+/- 6.11) %. However, % coverage for all other samples ranged from 50.37 (+/- 8.47) to 60.33 (+/- 9.15). At 2000 s\(^{-1}\), perfusion of LNAME-containing samples resulted in statistically more % coverage than non-LNAME samples. After 5 minutes, samples without LNAME ranged from 10.36 (+/- 2.41) to 11.00 (+/- 2.15) %, compared to samples containing LNAME that ranged from 15.52 (+/- 1.88) to 21.98 (+/- 2.49) %. This research presents the first evidence that platelet adhesion decreases strongly with shear rate in an in vitro model that closely simulates arterial shear flow. The minimal effect of insulin on platelet adhesion at low shear rates (~100 s\(^{-1}\)) implies that insulin plays a minor role in capillary and venal clotting processes. However, there is evidence to suggest that insulin inhibits thrombus formation through an insulin-dependent NO production pathway at medium shear rates (~1000 s\(^{-1}\)). At high shear rates (~2000 s\(^{-1}\)) insulin has a small effect on % coverage due to shear-induced, optimal NO production by platelets.

INTRODUCTION

Recent studies have suggested that diabetics may have a larger rate of macrovascular thrombus (clot) formation than normal individuals. Current evidence strongly suggests that thrombus formation is an aggregating factor in CD. Therefore, it is hypothesized that the link between diabetes and CD may be through enhanced thrombus formation. Thrombus formation is initiated at a site of vascular injury, where the protective endothelial cell layer lining the blood vessels has been removed. This exposes extracellular matrix (ECM) proteins lining the vessel wall, such as collagen types I and II (figure 1). The ECM proteins utilize arginine-glycine-aspartagine (RGD) amino acid sequences as sites for firm attachment of platelets, which are the primary cellular types involved in thrombus formation.

Platelets are anucleic and are directly derived from megakaryocytes. The approximate diameter is 3 μm, the mean lifetime in body (MLB) range is 8 to 14 days, and the typical adult platelet concentration is 310,000/μL. Platelets normally remain in a "ground" or "resting" state...
as they move through the vasculature. However, during an injury platelets are very responsive to
external stimuli sent from an injury site. 19,20,18 This makes them essential in the primary immune
response (hemostasis) but also can be an aggravating factor in arterial vessel occlusion/
pathological thrombosis. There are adhesive glycoproteins on the surface of platelets, localized
on microvilli extending from the platelet membrane.

These glycoproteins are generally termed integrins due to their ability to integrate entire
cell groups into organized structures and to regulate metabolic processes (i.e. wound healing). 19,20,21,22,23,24 Integrins are present on the surface of all cells and are responsible for cell
attachment and organization (i.e. recognition of like cells) into tissues, organs, etc. 24 The two
platelet integrins important in wound healing are the GPIb and αIbβ3 integrins (figure
1). 19,20,22,24 GPIb integrins are comprised of a single chain, while αIbβ3 integrins are
heterodimeric. 22 Both types of integrins promote adhesion/thrombus formation to an injury site.
However, each integrin responds differently to physiological shear and promotes adhesion
through different mechanisms of action.

At low shear rates (γ ~ 100 s⁻¹), the GPIb integrin can bond directly to the RGD site. It can
also bond to the plasma derived von willebrand factor (vWF). 19,20,22 Due to the rigidity of the
bond formed between the GPIb integrin and the RGD site on the collagen, it is extremely
difficult for a permanent bond to form at high shear. The force of blood flowing through the
vessel exceeds the mechanical arm strength of the GPIb/RGD bond. Therefore, at higher shear
rates (> 100 s⁻¹) vWF is required for permanent adhesion to occur between platelets and
collagen. 20,22 This is due to the multivalency of vWF, which allows more bonding between RGD
sites and GPIb integrins and promotes higher overall GPIb/RGD bond strength. This results in a
longer residence time of platelets on the collagen surface, allowing for activation of αIbβ3
integrins and platelet recruitment. 20

After the GPIb-vWF-RGD complex is formed, the calcium (Ca²⁺) channel present in the
plasma membrane opens, allowing Ca²⁺ to diffuse through the membrane. The increased Ca²⁺
concentration results in an αIbβ3 integrin conformational change (figure 1). During the
conformational change the αIb and β3 chains link to become "activated" (attain a bonding
conformation). 20,22 After the αIbβ3 integrins become active, platelets can crosslink to each other
through fibrinogen and vWF proteins, resulting in thrombus formation. There are several agonist
that inhibit thrombus formation by blocking Ca²⁺-dependent αIbβ3 activation and crosslinking. 25
There is evidence to suggest that physiological insulin acts as an agonist that affects thrombus
formation through this mechanism. 6,7

A receptor site on platelets for insulin has been characterized. A recent study has also
demonstrated that elevated levels of insulin have resulted in reduced thrombus formation to
collagen type I. The most widely accepted pathway for insulin-dependent inhibition is the
NO/cGMP pathway. 25 When insulin bonds to the receptor, a signal is sent to endothelial nitric
oxide synthase (eNOS) located within the platelets (figure 2). The eNOS, using NADPH as the
cofactor, produces nitric oxide (NO) through the conversion of L-arginine into L-citrilline. The
NO is used by phosphodiesterase III (PDE III) to synthesize cyclic adenosine monophosphate
(cAMP) and by cyclic guanosylate cyclase (cGase) to produce cyclic guanosylate
monophosphate (cGMP). Several studies have demonstrated that elevated cAMP and cGMP
levels result in decreased thrombus formation.

In the vasculature, insulin also has an affect on endothelial cells. 7 Further, endothelial
cells possess an insulin receptor and produce NO through an almost identical mechanism as
platelets. Endothelial cells also produce much greater levels of NO than platelets. It is still
unclear what effect insulin has on the NO production of platelets in the presence of endothelial cell monolayers.

In this study, the extent of platelet coverage to collagen type III at different shear rates will be quantified with an in vitro system. This will be done to determine the effect of shear rate on initial platelet adhesion through GPIb integrin receptors and it will be shown that platelet adhesion decreases substantially at high shear rates (1500-2500 s⁻¹). These results disagree with several previous studies that have suggested increased thrombus formation at higher shear rates.²²

The relationship between platelet coverage and insulin-mediated NO production will also be quantified. It will be demonstrated that at medium shear rates (1000 s⁻¹) insulin has a substantial effect on thrombus formation through platelet derived NO production. However, at high shear rates (2000 s⁻¹) insulin has a reduced effect due to shear-induced NO production. This is the first study to indicate that shear may play a direct role in NO production.²²

MATERIALS AND METHODS

Materials: HEPARIN, HEPES, NaCl, collagen type III from calf skin, bovine serum albumin (BSA), L-Nitro Arginine Methyl Ester (L-NAME), insulin from porcine pancreas, and isoproterenol are obtained from Sigma. Mepacrine (quinacrine) is obtained from ICN Biomedicals. Glass cover slips (24x50 mm) are obtained from Fisher Scientific. BAECs are obtained from the labs of Roger E. McEver, Oklahoma Medical Research Foundation (OMRF), Oklahoma City, OK 73012. HUVECS are obtained from Norman Regional Hospital, Norman, OK 73070.

Preparation of Glass Coverslips: Glass coverslips are soaked in 69.6% HNO₃ for 12 hours. Each coverslip is then washed with 20 mL nanopure H₂O and 10 mL 95% ethanol. After washing is complete, the coverslips are then placed in a 95% ethanol bath for at least 12 hours before use.

Protein Coating of Coverslips: Collagen type III solution is prepared at .8 mg/mL by dissolving 24 mg in 30 mL of 17 mM acetic acid (pH 2.6). The .1% BSA solution is prepared at .1% by diluting .33 mL of 33% BSA with 99.67 mL of 10 mM HEPES/115 mM NaCl buffer (pH 7.4). Coverslips are removed from a 95% ethanol bath and allowed to air dry (approximately 15 min). Half of each coverslip is coated with collagen and allowed to incubate for 4 hours in a humidified environment (80-90%) at room temperature (~24 °C). After incubation, each slide is rinsed with 10-15 mL of HEPES buffer solution and soaked in .1% BSA for at least 2 hours.

Platelet Preparation: Blood collected from healthy donors (30-60 mL, depending on shear rate and length of experiment) is anticoagulated with HEPARIN to a final concentration of 20 U/mL. The fluorescent label mepacrine is then added to a final concentration of 10 μM and allowed to incubate for 10 min.

Flow Experiments: The fluid mechanical environment of the vasculature is modeled with an in vitro parallel plate flow chamber as has been previously characterized. Due to the wide range of shear rates studied, different flow regime dimensions are utilized. These dimensions are determined by the width of the gasket opening. At 100 and 500 s⁻¹, .013 x 1.31 and .013 x .50 cm are the dimensions used, respectively. At 1000 - 2500 s⁻¹, the dimensions .013 x .25 cm are
used. Also, the trials for each experiment are conducted in random order for each donor. This is done to average the effects of increasing time-dependent platelet activity on platelet coverage.

Shear Dependent Study: The effect of shear rate on platelet adhesion is determined at 6 shear rates (100, 500, 1000, 1500, 2000, and 2500 s\(^{-1}\)) with 5 donors. This range of shear rates simulates those found in a healthy individual. In each trial, the blood is perfused for 6 minutes.

Insulin Effect on NO Production Study: In some studies, 500 nM insulin is added to the anticoagulated blood 10 minutes before the start of an experiment. In other studies, L-NAME, an inhibitor of platelet NO synthesis is added at a concentration of 200 μM for 20 minutes before the start of an experiment. In still other studies, both L-NAME and insulin are added as described above, except in sequential order.

Image Analysis: For all studies, the rate of thrombus formation is recorded on S-VHS analog media. The extent of mural thrombus formation on the surface is quantified by digital image analysis performed on an SGI Indy workstation running the ISEE® image analysis software by Inovision. A background image is acquired after blood begins flowing over the surface but before adhesion of any platelets. This image is subtracted from subsequent images. Platelets are identified based on their size and intensity using adjustable parameters. Images are acquired and analyzed every 10-30 seconds over the course of each experiment, which lasts from 5-10 minutes.

RESULTS AND DISCUSSION

Preliminary results have been obtained for the shear rate and insulin dependence of thrombus formation. These results provide evidence that the effect of insulin on thrombus formation is shear rate dependent and that platelets produce enough NO to inhibit adhesion to a collagen type III surface.

Thrombus formation is a function of shear rate

Platelets, obtained from healthy donors, abundantly adhere and aggregate on collagen-coated surfaces, while only a small number of platelets attach to albumin-coated surfaces. Single platelets adhere to the collagen surface before formation of mural thrombi. The thrombi increase in size in chains parallel to the direction of flow. Platelets adhere at a slightly higher rate on collagen at medium shear rates (1000 s\(^{-1}\)), compared to lower shear rates (100 s\(^{-1}\)). The rate of adhesion is much smaller at higher shear rates (2000 s\(^{-1}\)). However, as time elapses (5-10 minutes), the numbers of adherent platelets at lower and higher shear rates are comparable. Figure 3 shows a sequence of images representing platelet accumulation at a shear rate of 1000 s\(^{-1}\). There is no platelet adhesion onto the left hand side of each image since this portion of the coverslip is coated with 1% BSA to simulate the non-adhesive nature of intact endothelium. The right hand side of the coverslip is coated with type III collagen to simulate the exposed subendothelial matrix proteins at a site of vascular injury. The flow is from left to right. The extent of mural thrombus formation is quantified by digital image analysis as described in detail in the previous section. Figure 4 demonstrates the extent of mural thrombus formation expressed as the percentage of the collagen coated surface that is covered by platelets as a function of time for five levels of shear rate within the physiological range for arterial and venous vascular beds. The
results are the average of five experiments with five different donors. Some results are similar to
those obtained by previous investigators at lower shear rates. However, at shear rates of 1500
and 2000 s$^{-1}$ the rate of platelet accumulation on the surface is lower than several published
results.\textsuperscript{17,19,24,26} This demonstrates the important role of shear rate in determining the rate of
mural thrombus formation. At low shear rates, thrombus formation rate is limited by the rate of
transport of platelets to the surface. At sufficiently high shear rates, fluid mechanical drag on
individual platelets and platelet aggregates is high. This prevents attachment to the surface or
promotes detachment of the cells from the surface, which reduces the rate of platelet deposition
on the surface.

**Insulin modulation of thrombus formation from healthy donors**

Previous studies have shown that insulin causes platelets to produce nitric oxide, which in
turn causes elevated intracellular levels of cAMP, inhibiting platelet activation.\textsuperscript{34,41} Insulin is
added, at a concentration of 500 nM, to blood collected from healthy donors and blood is then
perfused through a parallel plate flow chamber to determine the extent of mural thrombus
formation. This insulin concentration is much higher than that normally found in the blood, but
is used to induce a maximal response. The results are shown in figures 5, 6, and 7 where the
extent of thrombus formation, expressed as \% coverage of the injured area, is plotted as a
function of time.

At the two lower shear rates (100 and 1000 s$^{-1}$), the addition of insulin causes a
statistically significant reduction in the formation of mural thrombus. The effect is much larger
at 1000 s$^{-1}$ than at venous levels of shear rate. These differences disappear in the presence of L-
NAME (an inhibitor of eNOS), suggesting that platelet NO production in response to the added
insulin is responsible for the differences seen in mural thrombus formation. L-NAME alone has
no effect on mural thrombus formation at 100 and 1000 s$^{-1}$. These results are consistent with the
current hypothesis that platelet adhesion to collagen under flow is primarily mediated by GPIb
and $\alpha_{IIb}\beta_3$ integrins. As mentioned earlier, constitutively expressed GPIb is responsible for most
platelet adhesion to the surface at low shear rates. At higher levels of shear rate, GPIb mediates
only transient adhesion while $\alpha_{IIb}\beta_3$ is required for stable adhesion.\textsuperscript{20,41} However, $\alpha_{IIb}\beta_3$
exists on the platelet membrane in an inactive, or low affinity, form. Therefore, $\alpha_{IIb}\beta_3$ integrins must be
activated to become fully functional.\textsuperscript{20,41} It is anticipated that increased nitric oxide production
and subsequent inhibition of platelet activation only affects $\alpha_{IIb}\beta_3$-mediated adhesion and this is
precisely what is observed. At 1000 s$^{-1}$, a large effect of insulin-mediated nitric oxide production
is observed, while at 100 s$^{-1}$ only a modest effect is observed.

At the highest shear rate studied (2000 s$^{-1}$), added insulin does not have a significant
effect on the formation of mural thrombi. However, a curious phenomenon is observed that
includes L-NAME but not insulin (figure 7). These cases are anticipated to be identical to the
controls that do not have any additives. Instead, it is found that adding L-NAME causes an
increase in mural thrombus formation in the absence of insulin. This suggests that at these high
levels of shear rate, platelets are producing NO and that NO is inhibiting platelet aggregation on
the surface. These results are the first to suggest that shear stress induces nitric oxide production
in platelets and that the NO is produced in a sufficient amount to affect platelet function, even in
whole blood. Also, a number of studies have demonstrated that shear stress, by itself, can
activate platelets and promote aggregation and activation.\textsuperscript{24,27,28} Perhaps, shear stress-induced
signal transduction in platelets promotes activation of eNOS. However, it is uncertain whether

59
any studies demonstrate that shear stress induces platelets to produce inhibitory substances, such as NO.

**SUMMARY**

- Diabetes is a metabolic disorder that affects the way that glucose is broken down in the body. In type I diabetes, insulin-producing islet β cells are destroyed. In type II diabetes, insulin is not produced in sufficient quantity to facilitate sufficient glucose uptake by the cells. A major complication for both types of diabetes is cardiovascular disease (CD). It is hypothesized that the fluctuation in physiological insulin concentration may be a major link between diabetes and CD onset.

- Platelets and endothelial cells possess an insulin receptor, which has been shown to upregulate eNOS production of NO by a similar mechanism. This modulates conformational changes in heterodimeric integrins, such as αIIbβ3 types.

- Insulin is hypothesized to have a greater effect on inhibition of thrombus formation as arterial shear rate increases. At low shear rates, GPIb single-chain integrins are responsible for adhesion to the RGD sites of ECM proteins, such as collagen type III. However, at higher shear rates (>100 s⁻¹) the platelet αIIbβ3 integrins are required for thrombus formation. Thus, insulin effect increases as αIIbβ3 integrins play more of a role in thrombus formation.

- Insulin has a negligible effect on thrombus formation at high shear rates (2000 s⁻¹). This may be due to near-optimum NO synthesis induced by high physiological levels of shear.

**FIGURES**

Figure 1: Model of platelet adhesion to collagen type III surface. At high shear rates, platelets require vWF to mediate adhesion. Platelet-platelet binding through αIIbβ3 integrins is mediated through multivalent fibrinogen and vWF proteins.
Figure 2: Model for insulin inhibition of platelet aggregation. The insulin binds to a receptor on the surface of platelets, which signals the eNOS to produce more NO, thus resulting in inhibition of Ca^{2+} influx.

Figure 3: Images showing development of platelet thrombi (1000 s^{-1}). These images are representative of 5 separate experiments.
Figure 4: Shear dependence of platelet coverage. At higher shear rates, there are decreasing rates of coverage due to more mechanical stress on GPIb-substrate bonds or less contact time for permanent bonds to form.

Figure 5: Insulin mediated NO production at 100 s⁻¹. At low shear, insulin has a small effect on platelet adhesion since GPIb integrins can support bonding directly to the substrate without recruitment of other platelets through inside-out signaling mechanisms.
Figure 6: Insulin mediated NO production at 1000 s\(^{-1}\). At medium shear, insulin promotes inhibition of platelet coverage through eNOS production of NO. This is indicated by a higher rate of platelet coverage from samples preincubated with eNOS-inhibitor L-NAME.

Figure 7: Insulin mediated NO production at 2000 s\(^{-1}\). Insulin has a minimal effect on eNOS production of NO due to near-maximum platelet NO production at high shear. This is indicated by the high rates of coverage for LNAME samples, as compared to control samples.
REFERENCES


64
Introduction

Predicting polymer-polymer miscibility and polymer-solute miscibility (where the solute is a low molecular weight molecule, such as a drug) is the goal of a large body of research [1, 2]. An application of such research is the design of polymeric controlled release devices for drug delivery. Here, the phase behavior of the polymer-drug system is important for two reasons. First, for multicomponent polymeric systems (i.e. blends and copolymers) a single homogenous phase or multiple phases may be desirable. It is important to know under what conditions the system will phase separate. Second, in microphase-separated polymer systems, dissolved drug thermodynamically partitions into a preferred phase. Drug partitioning can be used to increase drug solubility in the polymer, tailor drug release profiles, and stabilize macromolecular drugs. Thus, predicting how dissolved drugs will partition into a phase-separated system and how the drug may alter the phase behavior of the system are also important.

Phase separation is a thermodynamically driven process that occurs (when the temperature and pressure and composition of a system are fixed) when the Gibbs’ free energy can be lowered by separating into two equilibrium phases. Thus, many models for phase behavior are based on predictions of the Gibbs’ free energy of mixing. The Flory-Huggins lattice model is one of the most widely used because of its simplicity [3-5]. More recent models are extensions or modifications to the Flory-Huggins model. This model assumes that the segments of polymer chains are packed in a lattice, and that the two components of the system have segments of approximately equal volume. The Flory-Huggins model equation for the Gibbs’ free energy of mixing for binary polymer systems is

\[
\Delta G_{\text{mix}}^{\text{RT}} = \frac{1}{RT} \left( \frac{\varphi_A}{N_A} \right) \ln \varphi_A + \left( \frac{\varphi_B}{N_B} \right) \ln \varphi_B + \chi \varphi_A \varphi_B
\]

The Gibbs’ free energy here has units of energy per mole of volume segments. R is the ideal gas constant and T is the system temperature. \( \varphi_A \) and \( \varphi_B \) represent the volume fractions of components A and B respectively. \( N_A \) and \( N_B \) are the number of moles of volume segments of A and B. The first two terms on the right-hand side of Equation 1 represent the combinatorial entropy associated with packing the given number of each type of volume segment in the lattice. The third term on the right-hand side of Equation 1 represents the enthalpic part of the Gibbs’
free energy of mixing. \( \chi \) is the Flory-Huggins interaction parameter. Since the entropic part of the Gibbs' free energy is relatively small (the number of attainable conformations being greatly restricted by chain connectivity) and always favors mixing, the enthalpic part, dictated by \( \chi \), is very important. If \( \chi \) is known for a system, the phase behavior can be predicted.

If volume changes on mixing are neglected, then \( \chi \) accounts only for the energy of mixing and is given by the equation

\[
\chi = z' \frac{\Delta \bar{E}_{\text{mix}}}{RT}
\]

(2)

Here, the energy of mixing is scaled as the Gibbs' free energy is in Equation 1. The \( z' \) in Equation 2 is the effective coordination number of the lattice, the number of nearest neighbor sites for each lattice site, excluding sites occupied by immediate members of the reference segment's own chain. The interaction energy can be obtained from [2]

\[
\Delta \bar{E}_{\text{mix}} = \bar{E}_{AB} - 0.5 \left( \bar{E}_{AA} + \bar{E}_{BB} \right)
\]

(3)

The energy terms on the right-hand side of Equation 3 are the pairwise interaction energies associated with pairs of volume segments indicated by the subscripts.

Pairwise interaction energies of the type used in Equation 3 can be obtained from molecular mechanics calculations on pairs of volume segments in vacuo. Monte Carlo simulations can be used to generate a large number of initial structures, whose local structures are allowed to relax until convergence criteria are reached. Likely conformations are selected based on the energy of the final conformations [2].

The simplicity of the Flory-Huggins model is obtained at the expense of making some significant assumptions about the microscopic structure of the polymer. These assumptions include

- no non-combinatorial entropic effects contribute to the Gibbs' free energy of mixing
- there is no volume change on mixing
- the volume segments are approximately equal in size
- the volume segments are packed in an on-lattice configuration

These assumptions lead to the conclusion that the effective coordination number for polymers is always 4 when packed in a cubic lattice. Each volume segment has six nearest neighbor segments in a cubic lattice, two of which are immediate members of the same chain and therefore don't contribute to the energy of mixing.

We extend the applicability of the Flory-Huggins model to a polymer system of interest for controlled drug delivery applications wherein the polymer segments have unequal volumes. The polymer system of interest is blends of copolymers of sebacic acid, SA, and 1,6-bis-p-(carboxyphenoxy)hexane, CPH, which has been shown to microphase separate at temperatures of interest for certain copolymer compositions. The merits of this system for controlled release of drugs and the microphase separation are discussed elsewhere [6].
The repeat units of this system are shown in Figure 1. In this system the ratio of the radii of the volume segments is about 1.6 to 1 (ratio of volumes is about 4.1 to 1), so the assumption that the volume segments are equal is not valid. Also, on-lattice packing for an amorphous system becomes difficult to rationalize if the volume segments are not approximately equal in size. Furthermore, it is conceivable that significant volume changes on mixing can occur as small segments may occupy the spaces between large segments that would be otherwise inaccessible in a phase-separated system. Finally, it is noted that if the volume of the system is not constant, the simple combinatorial entropy term must be modified to account for the change in volume. Thus, in order to account for volume segments of unequal size, all of the assumptions aforementioned inherent to the Flory-Huggins model must be relaxed. This, in turn leads to a composition dependence for $z'$ and thus for $\chi$.

Several approaches are available in the literature to model thermodynamics in polymer systems. For instance, Case et al. use atomistic simulations on polymer fragments [1]. This technique is accurate for the interaction energies, but the coordination number is left as an adjustable parameter, to which the model is quite sensitive. The work by Fan and coworkers uses a Monte Carlo packing algorithm to find $z'$ [2]. This algorithm randomly packs monomer units around a reference monomer. Chain connectivity is accounted for by putting a 'dummy' atom at the two ends of each monomer to exclude volume that would otherwise be occupied by the immediate members of the monomer's own chain. The potential for a composition dependence of $z'$ is not included since the model is restricted to polymers and solvents with approximately equal sizes for the volume fragments. In this model, no objective convergence criterion is established to determine when a monomer is 'jammed', i.e. the value of $z'$ when no more monomers can be packed around the reference monomer.

This work presents a Monte Carlo packing algorithm that can be used to calculate $z'$ for a binary blend of copolymers of unequally sized volume segments when the ratio of the segment radii and the composition is known. The purpose of the model is to investigate the dependence of $z'$ on the composition variables (both copolymer and blend composition) and the ratio of the segment radii. The volume segments are modeled as mutually attracting hard spheres and are packed in trimer units around a reference trimer. Since the spheres are mutually attracting, periodic...
boundary conditions are introduced to prevent the system from gravitating towards the center. Results are presented for homopolymer blends of polyCPH and polySA.

Model algorithm

The packing algorithm first constructs the reference trimer by randomly packing two monomer units on the surface of the reference monomer. These first two monomer units are selected based on the composition of the copolymer to which the reference monomer belongs and are members of the reference monomer's chain. They will not contribute to the coordination number, but they will exclude volume and participate in attractive interactions.

To determine the orientation of the first monomer in the trimer, two random numbers are generated and used to find a vector in space that defines the location of the first monomer on the surface of the reference monomer according to

\[
\begin{align*}
\theta &= 2\pi r_1 \\
\phi &= 1 - 2r_2
\end{align*}
\] (4)

Here \( r_1 \) and \( r_2 \) are random numbers between 0 and 1, and \( \theta \) and \( \phi \) are angles defining the direction of the vector in space. A similar technique is used to pack the second monomer. Trimer units are then randomly packed around the reference monomer by first packing the center unit, in contact with the reference monomer, then packing the adjacent two monomers in the newly packed monomer's chain on its surface. The monomers constituting each trimer unit are randomly selected based on the composition variables of the system.

Each time a monomer is packed the distance to all previously packed monomers is checked. If the newly packed monomer is within the interaction distance to another monomer, it is attracted and translates toward the attracting monomer until their surfaces just touch. If the new monomer overlaps with a previously packed one, then a new random vector is generated and it is re-packed. When a new monomer fails to pack after a specified number of attempts, \( \tau \), the reference monomer is assumed to be 'jammed' and \( z' \) for this reference monomer is computed as the number of monomers within the interaction distance of the reference monomer (excluding the first two monomers packed). This process is iterated over a specified number of reference monomers and the mean \( z' \) is calculated for the system.

This algorithm introduces three additional parameters to the model: N, the total number of reference monomers packed around to obtain \( z' \); L, the characteristic length of a unit cell; and \( \tau \) the number of attempts to pack each monomer. The sensitivity to N can be dealt with by choosing N sufficiently high. If doubling N results in a change in the result by more than a few percent, N is too small. L must be sufficiently small that the density of the system is approximately uniform, but sufficiently large that monomers are not attracted by their images in the next cell. The most important parameter is \( \tau \). To find \( z' \) when the system is truly jammed, \( \tau \) must be very large. To obviate the need to perform the packing algorithm with very large \( \tau \), data
were collected for \( \tau \) varying over three orders of magnitude and fit to a power law given in Equation 5.

\[
z'_{2\tau} - z' = K \tau^{-k}
\]  

(5)

Here \( z'_{2\tau} \) represents the value of \( z' \) obtained when \( \tau \) is twice as large as the value used to find \( z' \). The values of \( K \) and \( k \) are obtained from the fit and \( z' \) at very large \( \tau \) is computed such that the solution is converged to a sufficient tolerance (0.0001). This value is taken to be the 'jammed' \( z' \).

**Results and Discussion**

An example set of data for \( z' \) is shown in Figure 2 along with the fit to Equation 5. (The example shown is for a blend with a polyCPH mole fraction of 0.8). This procedure was performed for several blend compositions to obtain the composition dependence of \( z' \).

![Figure 2](image)

*Figure 2. Example of procedure used to compute the value of the constants \( K \) and \( k \) in Equation 5. The data shown is for a blend of polyCPH and polySA with a mole fraction of polyCPH of 0.8.*

Figure 3 shows the composition dependence of \( z' \) for a homopolymer blend of CPH and SA. \( F \) is the mole fraction of CPH in the system. The values of \( z' \) for the homopolymers are approximately equal, as is expected. When \( F \) is close to 0, most of the monomer units are SA. When packing around an SA monomer, the occasional randomly selected CPH excludes volume and reduces the coordination number for the SA reference monomer. When the reference monomer is CPH, the higher surface area, coupled with the fact that most of the monomers selected to be packed are SA results in an increase in the coordination number by maybe one or two. However when a CPH is chosen to be packed, it may not find room, though there is space for an additional SA. The CPH reference monomer would be jammed. Furthermore, when
computing the overall value of $z'$, the value is weighted with respect to the overall blend composition, so the $z'$ for the SA units dominates. Thus, the value for $z'$ is slightly lower than for the homopolymers.

\[ y = 1.6004x^2 - 1.4957x + 5.3642 \]
\[ R^2 = 0.8335 \]

Figure 3. Results for the dependence of $z'$ at infinite $\tau$ on the mole fraction of polyCPH in a blend of polyCPH and polySA.

When $F$ is close to 1, for the CPH reference monomers, the occasional SA monomer excludes a smaller volume than the more probable CPH monomer but does not occupy so small a space that an additional CPH monomer unit can be packed. When the reference monomer is SA, the available surface area is small and most of the monomers packed are large CPH molecules which greatly reduces the value of $z'$. The overall $z'$ for the system is again reduced from that of the homopolymers. There appears to be a minimum in the value of $z'$ at a blend composition near 50:50. The variation in the predicted $z'$ is about ten percent. This corresponds to a ten percent variation in the prediction for the critical temperature from the Flory-Huggins model, which is significant. Additionally, variations in the value of $\chi$ with composition are observed experimentally for many polymer systems. This model postulates a possible origin of such a compositional dependence.

Preliminary molecular mechanics simulations have also been performed using the InsightII software package from Accelrys Inc. to obtain the interaction energies. A small number (2500) of each type of pair-wise interactions (A-A, A-B, and B-B) have been simulated. The energies obtained were weighted with a Boltzmann distribution of the form [2]

\[ p(E_{i+1}) = \begin{cases} 1 & E_{i+1} < E_i \\ \frac{\Delta E}{e^{\frac{\Delta E}{kT}}} & E_{i+1} > E_i \end{cases} \]
This distribution results in a temperature dependence for the interaction energies, where \( k \) is the Boltzman constant and \( p \) is the probability that the energy state \( E_{i+1} \) exists, based on the value of the previous existing energy state \( E_i \). A Monte Carlo algorithm selects the existing energy states for each temperature and computes an average interaction energy for that temperature. Linear temperature dependencies for the pair-wise interaction energies were obtained over the temperature range of interest. When combined using Equation 3, a linear function of temperature for the energy of mixing is obtained. Substituting this energy of mixing into Equation 2 results in a function for \( \chi \) with temperature and composition dependence and two constants, \( A \) and \( B \), of the form

\[
\chi(T, \varphi_A) = \frac{z'(\varphi_A)}{R} \left( \frac{A + B}{T} \right)
\]  

(7)

In the form shown in Equation 7, \( \chi \) decreases as temperature increases and predicts an upper critical solubility temperature (UCST) for the phase diagram. Thus, as temperature is increased, the phase-separated mixtures become miscible. The numerical values of the constants \( A \) and \( B \) are not quantitatively accurate. When performing this type of atomistic level simulation, quantitative results are not expected. Rather, as previously stated, the purpose of the model is to provide insight into the mechanisms that dominate the phase behavior of the polymer system of interest. The prediction of the shape of the phase diagram will be used to design experiments that can yield quantitative values for the constants.

**Conclusions**

The Flory-Huggins model for binary polymer systems has been modified to extend its applicability to a polymer system wherein the sizes of the polymer volume segments are unequal. The assumptions of no volume change on mixing and on-lattice packing are also relaxed. These modifications result in a composition dependence for the coordination number, \( z' \), and ultimately a composition dependence for the Flory-Huggins interaction parameter \( \chi \). Atomistic level molecular mechanics simulations were performed to predict the temperature dependence of the energy of mixing so that a functional form for \( \chi \) could be derived. The results provide insight into the phase behavior of a polymer system that will ultimately aid in the design of controlled drug delivery devices.

**References**


Three-Dimensional Hydrophobic Cluster Analysis: The Use of a Virtual Environment for Protein Sequence Analysis: HELIX v0.3

Anthony D. Hill, Alain Laederach, and Peter J. Reilly
Department of Chemical Engineering
Iowa State University
Ames, IA 50011-2230

Abstract

Hydrophobic cluster analysis (HCA) allows simple, structurally important sequence information to be viewed and analyzed across species. Traditionally, HCA is performed by computationally wrapping an amino acid sequence around an α-helix, then cutting and flattening the sequence. The resulting pattern of hydrophobic residues produces a fingerprint for the protein structure. These hydrophobic fingerprints can be compared across species to find likely structural patterns. This is a study in the use of a virtual environment to display HCA data. The amino acid sequence is wrapped around an α-helix and displayed in three dimensions. The helix appears before the users and can be placed alongside other three-dimensional helices. The use of an additional dimension in the presentation of the hydrophobic fingerprint allows for an additional human sense to be used in recognizing and matching patterns.

Introduction and Methods

Many methods exist to display hydrophobic cluster analysis (HCA) in a virtual environment. The methods range from using an HCA preprocessor to generate a three-dimensional metafile to be displayed by a dedicated virtual environment rendering program to writing a custom graphics processor to display HCA data. For the purposes of this project, we decided that the best option was to write a program to generate spatial coordinates needed to display HCA data on the fly and to use a library to interact with the virtual environment.

The Iowa State University (ISU) Virtual Reality Applications Center (VRAC) has developed a library of C++ classes, VRJuggler, to interface with virtual environments (Bierbaum, 2000). VRJuggler allows the programmer to use abstracted input and to display objects in a program’s code. At run-time, VRJuggler attaches these abstracted inputs and displays to real input and display devices. VRJuggler then handles all the complications of managing each input and display device. This approach has two major benefits: It allows the programmer to focus on the program code unique to that program’s particular task, and it allows for a program to be written once (and for one set of input and display devices) and then run using a variety of devices (from conventional monitors and mice to three-dimensional tracking gloves and headsets). It is for these benefits that VRJuggler was used in the development of HELIX.

Although VRJuggler handles management of input and display devices, it does not actually render three-dimensional coordinates into displayed objects. To accomplish this task, another library, OpenGL, was used. OpenGL is a cross-platform three-dimensional graphics display library developed by Silicon Graphics (Segal and Akeley, 2001). Three-dimensional graphics hardware accelerators are commonplace on Windows, Macintosh, and a wide variety of UNIX computers. This three-dimensional graphics hardware allows for complex three-dimensional scenes to be rendered from simple three-dimensional coordinates and primitives quickly. It is because of the wide availability of OpenGL accelerators and its cross-platform nature that OpenGL was used to display the three-dimensional graphics.
In its current implementation, HELIX v0.3 reads a text file containing the amino acid sequences of interest in FASTA format. Each sequence is read and its amino acids are displayed as spheres in a right-handed spiral, with 3.6 amino acids per turn, or one amino acid residue each $100^\circ$. Each sequence's spiral is oriented vertically, with each subsequent sequence displayed to the right of the previous one. This display geometry was chosen because HCA aligns amino acid residues in this fashion into an $\alpha$-helix.

Amino acid residues are displayed as either large yellow spheres (hydrophobic residues) or small grey spheres (all other residues). In considering a display color scheme, several options were examined. The first option was to color each residue according to a large number of properties, with each range of the color spectrum representing different properties, such as acidity, hydrophobicity, and ability to bond with other amino acids (Taylor, 1997). This scheme proved too complex to actually pick out patterns visually, as all the patterns were covered up in the noise of a rainbow column of spheres several stories tall. Therefore we decided to reduce the information shown to only hydrophobicity. Using smaller darker spheres for the non-hydrophobic residues virtually made these residues disappear from notice, leaving only the hydrophobic patterns visible. So, we determined that the yellow-grey, large-small sphere color scheme enabled the user to find the same hydrophobic pattern across several species easily.

A wireless mouse with a motion tracker is used as an input wand. This wand is used to select amino acid sequences, rotate their representative columns, and move the columns vertically to align homologous regions of amino acids.

Results

At this point, HELIX v0.3 has been tested by only a small number of people. All this testing has taken place in the C6 virtual environment in the ISU VRAC. Sample input files of six to ten amino acid sequences from a family of $\beta$-glycosidases have been used to test the qualitative ability of researchers to find homologous hydrophobic amino acid patterns. At this point, the results are promising; the researchers were able to quickly find homologous hydrophobic patterns within the enzymes' primary sequences. Figures 1-2 show two screen captures of HELIX v0.3.

![Figure 1. Screenshot of HELIX v0.3 displaying two aligned enzyme sequences.](image1)

![Figure 2. Screenshot of HELIX v0.3 displaying twelve nonaligned enzyme sequences.](image2)
Future Work

The goal of any computational research tool is to allow researchers to discover new information. This project is no different; it seeks to allow researchers to better use their abilities of spatial pattern recognition and thought to analyze protein sequence and structure. HELIX is far from actually implementing all of the techniques that can be used to aid researchers. Currently computers can do quite a bit of homology detection through multiple sequence alignment, domain recognition, and motif recognition. Additionally, more senses can be used to detect homology: colors, shapes, and sound. As development of HELIX continues, these features will be added.

If HELIX is to be of more value than simply a novel way to view sequence analysis, it must enable the researcher to use this new information. A very likely way of using this information is to query the online protein databases SwissProt and GenBank in a BLAST-like manner, ranking similarity based on three-dimensional HCA patterns, rather than primary sequence. This could allow similarly folding domains to be found between sequences which are not as closely related over the whole protein.

Additionally, some further user interface refinements must be made. Users must be able to mark a hydrophobic pattern and preserve its relationship to similar patterns in other sequences if any sense is to be made of more than one hydrophobic pattern in a set of sequences. Along with this ability, a convenient method to save and restore a homology searching session is needed.

Conclusions

Although HELIX v0.3 lacks many features that could make it a serious tool for research, it shows a great deal of promise. HELIX uses a virtual environment to display protein sequence information in a manner that allows a researcher to use more of his or her senses to analyze information. It already is able to help a researcher find homologous patterns without guidance. With the abilities to automate and guide the homology search and BLAST protein databases for similar patterns, HELIX should be able to explore territories in sequence analysis that other sequence analysis programs miss.

Acknowledgments

The authors gratefully acknowledge the support of a Carver Trust Grant for this work, as well as Greg R. Luecke for access to the Virtual Reality Applications Center.

References


MECHANICAL PERFORMANCE OF ELASTIN-MIMETIC HYDROGELS

Éder D. Oliveira¹, Sharon A. Hagan² and Stevin H. Gehrke²

¹Department of Chemical Engineering, Universidade Federal de Minas Gerais, Belo Horizonte, MG - 30 160 - 030 BRASIL
²Department of Chemical Engineering, 105 Durland Hall, Kansas State University, Manhattan, KS 66506 USA

Abstract

Elastin-mimetic and polyamine hydrogels were used to investigate the effect of polypeptide secondary structure on shear modulus as a function of polymer volume fraction. Although the elastin-mimetic protein undergoes changes in secondary structure as a function of temperature, the crosslinked material behaves as an ideal rubber under all conditions, as do most synthetic gels (e.g. poly(allylamine)). This stands in contrast with results from poly(α,L-lysine) hydrogels, where the change in secondary structure with pH altered the dependence of the modulus on polymer volume fraction.

Introduction

Protein-based polymers, polymers of amino acids, pose several advantages over synthetic polymers. The diversity of monomers spans twenty natural occurring amino acids as well as amino acids with modified side chains. Amino acid polymers can be synthesized through two pathways, chemically and with recombinant DNA, both having specific advantages and disadvantages. Using either synthetic pathway, precise control of polymer properties such as sequence, stereochemistry and chain length is easily achieved [1]. One of the most promising advantages of protein-based polymers is their ability to mimic the properties of naturally occurring proteins such as structure or function (e.g. biological recognition sites). Potential high biocompatibility, low cost bioproduction, and renewable raw materials are all advantages that are environmentally friendly [1]. One class of highly studied protein-based polymers is modeled after a recurring repeat sequence of native elastin [1-6]. Elastin-mimetic proteins have the potential to exhibit enhanced elastic properties and durability.

Hydrogels, water-swellable crosslinked polymer networks, made from elastin-mimetic proteins hold great promise as a biomaterial for a wide variety of biomedical applications, including responsive drug delivery devices, catheter coatings, tissue engineering matrices, and regeneration, replacement, or repair of vascular tissue [7-10]. Such materials possess excellent biocompatibility, as shown in at least eleven different tests [7]. However, in general, poor mechanical properties are frequently the biggest limitation of hydrogels. Polypeptide gels can possess a variety of secondary structures (both rigid and flexible) not commonly seen in synthetic polymers; therefore, it has been hypothesized that the mechanical properties of polypeptide gels may be different from (and possibly superior to) gels of synthetic polymers [8,10]. Polypeptide gels also have the advantage of easily varying the secondary structure through manipulation of the amino acid sequence [7].

Responsive hydrogels shrink and swell with environmental stimuli such as pH, temperature and ionic strength [11]. These same stimuli can also induce secondary structure transitions in polypeptides. Figure 1 depicts the pH-dependent swelling of a poly(α,L-lysine) gel in aqueous solution. The uncrosslinked polymer exhibits a secondary structure transition (random coil to β-sheet) for pH between 10 and 11 for this polymer concentration [11]. This secondary structure transition corresponds to the pH transition that the poly(α,L-lysine) gel exhibits in the swelling curve. Poly(allylamine) gels exhibit the same pH dependence with a transition between pH 6 and pH 8, but its free polymer does not have a structural transition.
Figure 1: pH-dependent swelling of poly(α,L-lysine) gels in sodium chloride solution (pH < 5.0, pH > 11.4; I=0.05) or sodium p-phenolsulfonate buffer solution (5.0 ≤ pH ≤ 11.4; I=0.05). The gels are pH-responsive and show no hysteresis upon swelling and shrinking.

Theoretically, it has been predicted that the shear modulus of gels with flexible chains and flexible crosslinks (as in almost all synthetic polymer gels), varies with polymer volume fraction $\phi_2$ as $G \sim \phi_2^{1/3}$. In contrast, gels of rigid rods connected by flexible crosslinks should have a much steeper dependence of modulus on the polymer volume fraction: $G \sim \phi_2^{3/2}$ [12]. In fact, the log-log plot in Figure 2 shows that for a crosslinked gel of poly(α,L-lysine) under conditions of a random coil configuration $G \sim \phi_2^{0.33\pm0.06}$, while under conditions of probable β-sheet configuration $G \sim \phi_2^{1.50\pm0.2}$, matching theoretical predictions within experimental uncertainty [10]. In other words, poly(α,L-lysine) gels behave as flexible networks below their transition pH and behave as rigid networks with flexible junctions above their transition pH [10]. The poly(α,L-lysine) gels with the organized secondary structure also had moduli several times greater than the poly(α,L-lysine) gels with random coil configuration of the same polymer volume fraction. Poly(allylamine) gels, a similar polymer gel, behave as flexible networks, $G \sim \phi_2^{0.31\pm0.01}$, for all pH values (results not shown). Therefore, the modulus changes of poly(α,L-lysine) gels cannot be attributed solely to the effect of pH or polymer volume fraction. Elastin-mimetic polymers have been shown to undergo a transition from a random coil configuration to an organized secondary structure (β-spiral) upon an increase in temperature [7]. The hypothesis for this work is that for elastin-mimetic gels, this transition might be accompanied by a modulus shift analogous to poly(α,L-lysine) gels.
Experimental Methods

The elastin-mimetic protein used in these experiments was synthesized by genetic engineering of *E. coli*, microbial protein expression, and purification to yield an artificial protein with a molecular weight near 90 kDa [9]. The protein sequence [(VPGVG)$_4$(VPGKG)$_{40}$]$_n$, where V is valine, P is proline, G is glycine, and K is lysine, was modeled after a recurring sequence in native elastin (VPGVG) with the crosslinkable group lysine substituted in to every 25th position. After the expression of the protein in *E. coli*, the cells were disrupted using sonification and the protein was released into solution. The DNA was removed by precipitation with polyethyleneimine. Making use of the inverse temperature transition, the protein was salted out of the remaining solution using 1M sodium chloride at 37°C. The protein was recovered by redissolving the precipitate at 4°C in phosphate buffered saline. The temperature cycle was repeated to further purify the protein. After the second temperature cycle, the resulting solution was desalted using a centrifugal filter with a molecular weight cut-off of 300 kDa. The desalted solution was then freeze dried to yield a white powder form of the protein.

Elastin-mimetic hydrogels have been synthesized in several ways in the literature [7-9]. The goal of this synthetic procedure is to produce gels using a well-defined precursor polymer and a well-controlled crosslinking scheme. For this study, elastin-mimetic gels were synthesized by crosslinking the lysine groups of [(VPGVG)$_4$(VPGKG)$_{40}$]$_n$ by a condensation reaction with disuccinimidyl suberate in anhydrous dimethylacetamide. Gel disks were prepared by combining appropriate amounts of 0.1 g/mL polymer solution and 0.038 g/mL crosslinker solution (molar ratio of polymer amino groups to crosslinker was 2). Gelation occurred at 3°C in ~10 minutes. After 24 hours, the gel disks were

Figure 2: Shear modulus dependence of poly(α-L-lysine) gels on volume fraction of polymer. The gels behave as flexible networks below their transition at pH 11, and the gels behave as rigid networks with flexible junction above their transition pH.
removed from the mold and placed in a 10 v/v% solution of ethylenediamine in anhydrous dimethylsulfoxide for 24 hours at room temperature to terminate any unreacted N-hydroxysuccinimidyl groups remaining inside the gel with an amino group.

Poly(α,L-lysine) and poly(allylamine) hydrogels were synthesized with a similar procedure, but they were crosslinked by a base-catalyzed addition reaction with N,N-methylene-bis-acrylamide and without the subsequent termination reaction. Swelling curves were obtained from mass measurements after equilibration in aqueous solution for 24 hours at each pH or temperature. Shear moduli were obtained from uniaxial compression experiments.

Results and Discussion

The linearity of Figure 3 for a plot of stress vs. strain indicates ideal elastic behavior [13]. Ideal behavior is further verified through the calculation of the Mooney-Rivlin constants C2, which is a direct measure of the departure of a network from ideality [13]. The Mooney-Rivlin constants are virtually zero for elastin-mimetic gels at all conditions of temperature, pH, polymer volume fraction, and crosslink density tested.

![Figure 3: Stress-strain isotherm of elastin-mimetic gel obtained at pH=11.0, T=15°C, and m_{eq}/m_{dry}=1.96, where G = shear modulus, σ = nominal stress, and α = ratio of the deformed length, \( \ell \), to the undeformed length, \( \ell_0 \). Linearity of the plot indicates ideal elastic behavior.](image)

Similar to the pH-dependence of poly(α,L-lysine) gels, elastin-mimetic gels exhibit a temperature-dependence, contracting with increasing temperature, as demonstrated in Figure 4. The slight increase in swelling with reduction in pH seen at low temperatures indicates the presence of some ionic species, either unreacted lysine groups or singly-reacted crosslinker groups terminated with ethylenediamine. The uncrosslinked polymer is proposed to undergo a secondary structure transition (random coil to β-spiral) around 27°C.
Figure 4: Temperature- and pH-dependent swelling of elastin-mimetic gels in unbuffered sodium hydroxide or sodium chloride solution (open symbols represent a second trial at pH=11). The gels are temperature responsive; the slight pH-dependence shows that some free amine groups exist.

After equilibration at 20°C or 40°C, the polymer volume fraction of the elastin mimetic gels was reduced to varying levels by partial dehydration in order to test the theoretical predictions of modulus dependence on polymer volume fraction [12]. This is tested in Figure 5 for elastin-mimetic gels above and below the transition temperature, where the reduced shear modulus is used to eliminate the dependence of modulus on temperature [13-14]. The reduced modulus is defined as the ratio of the measured modulus to the limiting modulus of a perfect elastic dry network prepared in bulk conditions, \( \rho_TRT \), where \( \rho_T \) is the theoretical crosslink density of the network [14]. The log-log plot shows that the elastin-mimetic gels behave as flexible networks both above and below the transition temperature, despite the proposed secondary structure change. Thermoelasticity experiments also indicate that elastin-mimetic gels behave as typical rubbers at all temperatures [14].

Summary and Conclusions

Elastin-mimetic gels behave as flexible networks and display ideal rubber elastic behavior at all pH, temperature, and polymer volume fractions tested. This is similar behavior to poly(allylamine) gels, but in contrast to the poly(\( \alpha, L \)-lysine) gels which exhibit a change in network behavior upon crossing the structural transition pH of the polymer. It was shown that the rigid secondary structure of the poly(\( \alpha, L \)-lysine) polymer can improve the mechanical properties of protein-based gels, but the \( \beta \)-spiral structure proposed for the high temperature configuration of the elastin-mimetic polymer appears to lack the rigidity necessary to alter the mechanical performance of its gel. Despite the differences in protein-based gel performance, it is evident that protein-based hydrogels display elastic behavior distinct from that observed in conventional vinyl polymer gels.
Figure 5: Dependence of reduced shear modulus on polymer volume fraction for elastin-mimetic gels above and below the transition temperature (at pH=11). A single least squares line (solid line) fits all the data points with an exponent that closely matches the theoretical scaling law (dashed line) for flexible networks [12].

References

Acknowledgements
This work was supported by NASA Microgravity Biotechnology Program Grant 97-HEDS-02-082, and a fellowship to EDO from Conselho Nacional Desenvolvimento Científico e Tecnológico – CNPq/BRASIL. Dr. Vincent P. Conticello, Emory University, Atlanta GA, provided the biosynthetic elastin, bacteria clones, and helpful advice.
Multiple Sequence Alignment and Phylogenetic Analysis of Family 1 $\beta$-Glycosidases

Anthony D. Hill, Alain Laederach, and Peter J. Reilly
Department of Chemical Engineering
Iowa State University
Ames, IA 50011-2230

Abstract

$\beta$-Glycosidases are important enzymes in the digestion of $\beta$-linked glycosidic polymers, primarily in the breakdown of plant cell-wall tissue. A rough multiple sequence analysis of glycoside hydrolases divides the known sequences of $\beta$-glycosidases into two families, Family 1 and Family 3. These families are differentiated by sequence and structure, with Family 3 consisting of $\beta$-glycosidases primarily from fungi and prokaryotic organisms. Family 1 contains enzymes of many substrate specificities: glucosidases, mannosidases, galactosidases, and even a myrosinase. This study further analyzes the $\beta$-glycosidases in Family 1 using multiple sequence and phylogenetic analysis. Accurate subgrouping of the enzymes within Family 1 allows representative three-dimensional structures to be chosen so that automated docking can model the interaction of each subfamily with multiple $\beta$-linked substrates.

Methods

A multiple sequence alignment of functionally related enzymes is useful for two reasons. First, it would be expected that the functionally important amino acid motifs would be strongly conserved across multiple species. Finding regions of high homology helps to highlight potentially important amino acids. Second, the multiple sequence alignment is used to guide the construction of a phylogenetic tree, which shows how the each amino acid sequence is related to each other aligned sequence. These relationships can be used to group enzymes into sequence-related subfamilies.

As it is published, the amino acid sequence of an enzyme is added to online sequence databases like Swiss-Prot (Gasteiger et al., 2001) and GenBank (Benson et al., 2000). These sequences can be downloaded to a personal computer or workstation for use. The sequences used in this analysis are a subset of the sequences categorized as Family 1 $\beta$-glycosidases (Henrissat, 1991). CAZy is an online database of carbohydrate enzymes categorized by sequence similarity (Coutinho and Henrissat, 1999). From CAZy, the Family 1 $\beta$-glycosidases were screened so that fragments, open reading frames, and large precursors were not included. Using only the sequences of known functional enzymes eliminated unrelated variation in the multiple sequence alignment. The remaining sequences were downloaded from both Swiss-Prot and GenBank as linked to by CAZy (Coutinho and Henrissat, 2001).

The method for determining an optimal alignment of two sequences according to a scoring function is described completely by a computational algorithm (Smith and Waterman, 1981). This method is well known and can be computed in time proportional to the product of the number of letters in each sequence. Extrapolating Smith and Waterman's algorithm to a multiple sequence alignment would again require time proportional to the product of the number of letters in each sequence. It is easily seen that this results in an exponential growth in the time required to compute an alignment of multiple sequences. Such a global solution is unreasonably computationally demanding, so heuristics to solve multiple sequence alignments have been developed.
and are available in several released programs. One such program is Clustal W and its graphical interface Clustal X (Thompson et al., 1994). These programs were used to automate the process of finding an optimal alignment.

It is important to note that any method of solving an optimal sequence alignment requires a scoring function. This defines the metric by which the quality of alignment is judged. To align protein sequences, this metric is most easily defined as the probability of mutation from one sequence to another. The total probability of mutation is obviously the product of the probability of mutation at each amino acid. For simply comparing optimal metrics and when desiring smaller amounts of memory to be used to store the metric, the sum of the probability of mutation at each amino acid can be used. A substitution matrix simply defines the probability of mutation from one amino acid to another, as well as the probability of insertion or deletion of amino acids. Such a matrix has been determined theoretically (Kimura, 1980), but for practical purposes a matrix determined by a broad sample of nature is more accurate (Jones et al., 1992).

In conjunction with manual alignment of the known catalytic residues using the program SeaView (Galtier et al., 1996), Clustal W was used along with the Gonnet PAM 250 substitution matrix (Gonnet et al., 1992) and the following programs (Table 1) to find the multiple sequence alignment used for sequence and phylogenetic analysis.

Table 1. Program descriptions.

<table>
<thead>
<tr>
<th>Program</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proml</td>
<td>Computes a maximum likelihood phylogenetic tree for each of the bootstrapped multiple sequence alignments.</td>
</tr>
<tr>
<td>Protdist</td>
<td>Computes the alignment score between each pair sequences in an alignment.</td>
</tr>
<tr>
<td>Seqboot</td>
<td>Creates many sets of the multiple sequence alignment, each with random permutations in the amino acid sequence, simulating data samples from the population. This process is known as bootstrapping.</td>
</tr>
<tr>
<td>Drawtree</td>
<td>Outputs the graphical representation of the tree.</td>
</tr>
<tr>
<td>Consense</td>
<td>Finds a consensus tree from each of the trees created by passing the bootstrapped multiple sequence alignment through Proml.</td>
</tr>
</tbody>
</table>

Phylogenetic trees are computed as the most likely evolutionary path. This is accomplished by guessing the most likely parent sequences that could mutate into a pair of the child sequences that are observed today. Each pair of parent sequences in turn has its parent sequence guessed, and so on until only one parent sequence remains. The sequences of these parent sequences and their relation to the individual child sequences are optimized so that the probability of the series of mutations occurring is maximized.

One such method for computing this tree is called maximum likelihood. This attempts to find a phylogenetic tree that displays the property its name implies: maximum likelihood of occurring. The algorithm does so by using tabulated amino acid transition probabilities and calculating branch occurrence probabilities to find the probability of the phylogenetic tree in question occurring. This overall probability is then maximized to find the maximum likelihood tree (Felsenstein and Churchill, 1996).

PHYLIP is a collection of programs that implement the computer algorithms necessary to construct a phylogenetic tree from a multiple sequence alignment (Felsenstein, 1993). These programs are listed in Table 2.
The multiple sequence alignment produced by Clustal X was used as input to Seqboot. The relevant parameters used are also listed in Table 1:

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence, Morph, Rest., Gene</td>
<td>Molecular</td>
<td>The input sequences are molecular sequences.</td>
</tr>
<tr>
<td>Freq?</td>
<td>Sequence</td>
<td></td>
</tr>
<tr>
<td>Bootstrap, Jackknife, or Permute</td>
<td>Bootstrap</td>
<td>The purpose of running seqboot is to bootstrap the sequences</td>
</tr>
<tr>
<td>Block size</td>
<td>1</td>
<td>The frequency of site change appears independent.</td>
</tr>
<tr>
<td>How many replicates?</td>
<td>100</td>
<td>A data pool of 100 samples is beginning to get statistically significant,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>without being computationally impossible</td>
</tr>
</tbody>
</table>

Seqboot simulates having taken more samples from the population of sequences. It achieves this result by creating new data sets from an input data set by the method of bootstrapping. Bootstrapping samples \( N \) characters at random with replacement from the data set, so that the new data set has the same number of characters as the original, only some characters have been removed and others have been duplicated. Bootstrapping is statistically typical of sampling new data sets (Künsch, 1989). The parameters used are listed in Table 2.

The multiple sequence alignment was bootstrapped to 100 data sets. For each data set, the maximum likelihood tree was found using Proml (Table 3). These trees were then used as input to Consense (Table 4), which found the majority consensus tree representing all data sets. By using a consensus tree, only inner nodes that are statistically distinct are differentiated. Two nodes that are statistically close enough to be the same are replaced by one node with several branches, more closely modeling the behavior found in nature.

In addition to finding a phylogenetic tree, the bootstrapped data set was analyzed by Protdist (Table 5) to find the pairwise distance between each species. The distances over the species set were averaged using Excel and the standard deviations are reported as well. The resulting matrix lists the distance between each sequence as an expected percent of amino acids changed from sequence to sequence.

The majority consensus tree file was used as input to Drawtree.

The resulting PostScript file was edited in Adobe Illustrator to make the labels of the sequences more complete and meaningful and to add color behind the tree to denote the different subfamilies. Division into subfamilies was a qualitative judgment to find the natural breaks between them. This process is easily aided by finding a common trait among each member of a subfamily.

**Results**

The multiple sequence alignment results in the homologous residues being aligned. The catalytic portions of the alignment are pictured in Figure 1. It can be seen there that the catalytic acid, a glutamic acid, and is fully conserved throughout the species, except for the myrosinases, where a glutamine is aligned in place of a glutamic acid. From the literature previously published, it is not apparent which residue is the proton donor in myrosinases.

As shown in Figure 1, the catalytic base is fully conserved, with the exception of the third A. thaliana \( \beta \)-glucosidase. It appears that the DNA sequence that was translated has an error resul-
Table 3. Proml parameters.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance Matrix</td>
<td>JTT</td>
<td>The Jones-Thompson-Taylor distance matrix is a matrix based on the observation of a very large set of amino acid sequences. The XX distance matrix is based on a smaller set of amino acid sequences, while the XX distance matrix is based purely on theoretical probabilities.</td>
</tr>
<tr>
<td>Speedier, but rougher analysis</td>
<td>No, not rough</td>
<td>This option uses more CPU cycles, but produces a more accurate tree.</td>
</tr>
<tr>
<td>Global rearrangements</td>
<td>Yes</td>
<td>After the tree is created by building it one sequence at a time, this option tells the program to remove each sequence and re-add it optimally, refining the tree even more.</td>
</tr>
<tr>
<td>Randomize input order of sequences</td>
<td>Yes, $4n + 1, 1$</td>
<td>This adds the sequences in a random order. The $4n + 1$ random number seed starts the random number generator. The seed should be different for each data set, so that the same random progression is not always followed, and should be of the form $4n + 1$, because other numbers quickly collapse to the same random number progression. One jumble is sufficient, given the rearrangements.</td>
</tr>
<tr>
<td>Write out trees onto a tree file</td>
<td>Yes</td>
<td>The tree file is used as input to consense and eventually to produce the phylogenetic tree of interest.</td>
</tr>
</tbody>
</table>

Table 4. Consense parameters.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consensus Type</td>
<td>Majority rule (extended)</td>
<td>This will create a tree containing each species, yet will be consistent with the majority of the trees in the bootstrapped data set.</td>
</tr>
</tbody>
</table>

Table 5. Protdist parameters.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance Matrix</td>
<td>JTT</td>
<td>The Jones-Thompson-Taylor distance matrix is a matrix based on the observation of a very large set of amino acid sequences. The XX distance matrix is based on a smaller set of amino acid sequences, while the XX distance matrix is based purely on theoretical probabilities.</td>
</tr>
</tbody>
</table>
ting in a translation to glutamine rather than glutamic acid.

Overall, the sequence shows very high homology; none of the sequences differing by more than 3.6901 PAMs. The distance between each amino acid sequence can be seen in Figures 2-5. However, in spite of the high homology between sequences, the actual enzymes demonstrate a wide variety of substrate specificities.

Figure 6, the phylogenetic tree of Family 1, demonstrates that sequentially the enzymes are more related by taxonomic family than by substrate specificity. However, within each taxonomic subfamily, smaller subfamilies based on substrate specificity can be seen. Figure 6 has been pattern-coded to show the subfamilies apparent within Family 1. These subfamilies were identified qualitatively, noting distinct breaks within the tree. Accordingly, subfamilies from four sources are observed: one from plants and fungi, a second mostly from thermophilic bacteria, a third from bacteria primarily specific to β-glucosyl polymers, and a fourth from bacteria that are primarily specific to β-phosphoglycosyl polymers. Within each subfamily the variations in specificity are also color-coded, showing the subfamilies within subfamilies.

Within the plant subfamily (white dots), a small grouping of myrosinases (dark grey) is seen. It is interesting to note that while this subgrouping has a different specificity, and is accordingly grouped together, it resides in the midst of the entire plant subfamily. This phylogenetic placement indicates that much stronger sequence similarity exists between the myrosinases and other plant enzymes than between β-glucosidases from different species subfamilies.

It can be seen that there exists a fair amount of sequence difference between enzymes specific to β-phosphoglycosyl polymers and those specific to other β-glycoses. The β-phosphoglycosyl-specific bacterial subfamily (pyramids) lies opposite the β-glycosyl-specific bacterial subfamily (light grey) on the phylogenetic tree. The separation in the phylogenetic tree indicates a strong sequence divergence to enable specificity towards β-phosphoglycosyl polymers. Smaller groupings can be seen within the β-phosphoglycosyl-specific bacterial subfamily. The enzymes specific to P-β-galactose (white) and P-β-glucose (pyramids) have distinct sequence differences, as indicated by their separation into subgroups.

Another important feature of the phylogenetic tree is the separation of enzymes from thermophilic bacteria from other bacteria. An entire subfamily of thermophilic bacteria is indicated (dark grey) and a thermophilic subgroup is seen in the β-glycosyl-specific bacterial subfamily (swirled dots). The distinction between thermophilic and non-thermophilic bacteria indicates that significant sequence and therefore structural differences exist that enable the enzymes to be more thermostable. Although it can be seen that the enzymes are sequentially more related by taxonomy than by specificity, it is not a far stretch to represent each subfamily with one of the available three-dimensional structures found by crystallization. The subfamily clustering by substrate specificity indicates there is also a need to use several crystal structures to represent not just taxonomic differences, but also substrate specificity differences between enzymes.

**Future Work**

Obviously, more thorough methods must be used to determine the causes of substrate specificity among the enzymes. One such method is computational docking. For computational docking, three-dimensional crystal structures must be used. As the phylogenetic analysis shows, the available crystal structures represent Family 1 well.

The available crystal structures will be used to dock β-glycoside di-mers into the active site. The resulting conformations should allow the determination of residues and motifs important to substrate specificity.
Figure 1. Multiple sequence alignment around the catalytic acid (left) and catalytic base (right). The residue numbers correspond to positions within the full multiple sequence alignment, displayed in Appendix A. Unless otherwise noted, enzymes are β-glucosidases.
**Figure 2.** Plant subfamily distance matrix.

**Figure 3.** P-β-galactose and P-β-glucose bacteria subfamily distance matrix.
Figure 4. β-Glucosidase bacterial subfamily distance matrix.

Figure 5. Thermophilic bacteria subfamily distance matrix.
Figure 6. Phylogenetic tree of Family 1 β-glycosidases. Asterisks denote β-glucosidases. Crystal structures exist for *Sulfolobus solfataricus* β-galactosidase, *Lactococcus lactis* 6-P-β-galactosidase, *Sinapis alba* myrosinase, and *Trifolium repens, Zea mays, Streptomyces sp., Bacillus circulans*, *Bacillus polymyxa, Thermoplasma aggregans*, and *Pyrococcus furiosus* β-glucosidases.
Conclusions

Family 1 β-glycosidases exhibit a variety of specificities. In order to understand the elements causing specificity, more research than simple sequence and phylogenetic analyses must be performed. However, the current analyses support the use of the available three-dimensional crystal structures to computationally dock the various substrates into each enzymes active site. Each of the four subfamilies, and more importantly, five of the seven subfamily clusters are represented by crystal structures.

References


